

The distribution and spread of emerging human infectious diseases

A thesis submitted for the degree of *Doctor of Philosophy*

Moritz U G Kraemer*

Hertford College, University of Oxford, United Kingdom

Trinity Term 2016

* kramer.moritz@gmail.com

Abstract

The distribution and spread of emerging human infectious diseases

Moritz U G Kraemer
Doctor of Philosophy
University of Oxford, United Kingdom, Trinity 2016

Despite many successes in the control of human infectious diseases they continue to pose a considerable risk to human health. The global distributions of pathogens are driven by ecological factors and the magnitude and extent of transmission are influenced by the dynamics of human behaviour.

Viruses such as chikungnunya virus, Zika virus, dengue virus, and Ebola virus have recently expanded geographically. However, prior to their expansion there was little quantitative evidence available to identify locations that may be susceptible to transmission and to evaluate the likelihood of virus introduction to such locations. In this thesis statistical modelling techniques were applied with the aim of understanding infectious disease ecology, determining the main drivers of disease occurrence, and predicting the magnitude and regional spread of an outbreak in real-time.

My results provide estimates of the populations now living in areas with possible transmission of chikungunya virus, show that the seasonal dynamics of Zika infection coupled with data on international travel, can better predict the arrival of the virus into new locations. Analyses of regional outbreaks of viruses including Ebola virus in West Africa and Yellow fever virus in Angola and the Democratic Republic Congo, show that patterns of human mobility strongly predict the real-time spread of disease. Further, I demonstrate that the impact of human movement varies considerably depending on the time of the outbreak (expanding versus declining phase) and the country of interest.

The results and conclusions of these studies are discussed in the context of improving our understanding of infectious disease dynamics and of informing public health policies, interventions, and control efforts.

Statement of contribution and associated publications

The work presented in this thesis is comprised of articles that have either been published in peer reviewed academic journals or are in submission. The articles are of broad nature and represent collaborative efforts from members of the Spatial Ecology and Epidemiology Group (SEEG) and other international collaborators from institutions across the world. While full author contributions are provided at the end of each article, find below a summary of my own contributions to each chapter.

Introduction

Parts of my introduction were published in similar form in a paper accepted and published in final form which I lead as first author.

Kraemer, M. U. G., S. I. Hay, D. M. Pigott, D. L. Smith, G. R. W. Wint, and N. Golding. 2016. Progress and challenges in infectious disease cartography. Trends in Parasitology 32:19–29.

Chapter 1 – The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus.*

This chapter consists of two publications which I am first author on. Both publications are peer reviewed and published in its final form. I oversaw all aspects of the work. I conceived the research (with SIH and NG), generated and oversaw the geo-positioning of the occurrence points (with MES, KAD, AQNM), standardised the datasets from a variety of data sources including international entomological surveys, and implemented the modelling (with NG), interpreted the results (with SIH) and wrote the first draft of the manuscript, which was subsequently edited and approved by all authors.

Kraemer, M. U. G., M. Sinka, K. Duda, A. Mylne, F. Shearer, C. Barker, C. Moore, R. Carvalho, G. Coelho, W. Van Bortel, G. Hendrickx, F. Schaffner, I. Elyazar, H. Teng, O. Brady, J. Messina, D. Pigott, T. Scott, D. Smith, G. Wint, N. Golding, and S. Hay. 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. eLife 4:e08347.

Kraemer, M. U. G., M. E. Sinka, K. A. Duda, A. Mylne, F. M. Shearer, O. J. Brady, J. P. Messina, C. M. Barker, C. G. Moore, R. G. Carvalho, G. E. Coelho, W. Van Bortel, G.

Hendrickx, F. Schaffner, G. R. W. Wint, I. R. F. Elyazar, H.-J. Teng, and S. I. Hay. 2015. The global compendium of *Aedes aegypti* and *Ae . albopictus* occurrence. Scientific Data 2:150035.

Chapter 2 – The global distribution and environmental suitability for Chikungunya virus.

This chapter consists of a publication that is accepted for publication and published in its final form. In this paper I am joint first author (together with EON), solely responsible for the modelling, and designing the evidence consensus layer (with OJB, DMP). I analysed and interpreted the data (with ENO). In addition I submitted the paper for publication and was responsible for the revision process (with EON).

Nsoesie, E. O.*, **M. U. G. Kraemer***, N. Golding, D. M. Pigott, O. J. Brady, C. L. Moyes, M. A. Johansson, P. W. Gething, R. Velayudhan, K. Khan, S. I. Hay, and J.S. Brownstein. 2015. Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. Eurosurveillance 21:pii=30234.

*joint first authors

Chapter 3 – The introduction, suitability and spread of Zika virus.

This chapter consists of two publications that are accepted and published in final form. In one paper I was joint first author (with IIB and OJB), responsible for the modelling (with OJB), interpretation of results (with IIB, KK, OJB) and writing of the manuscript (with IIB, OJB). In the second paper I am significant contributor to the paper as a second author. I was responsible in conception and design of the study (with JPM, OJB, SIH), solely responsible for the modelling and contributed significantly to the interpretation of the results and writing (with JPM, OJB, SIH).

Bogoch, I. I.*, O. J. Brady*, **M. U. G. Kraemer***, M. German, M. I. Creatore, S. Brent, A. Watts, S. I. Hay, and K. Khan. 2016. Potential for Zika virus introduction and transmission in resource limited countries in Africa and Asia-Pacific: a modelling study. Lancet Infectious Diseases: dx.doi.org/10.1016/S1473-3099(16)30270-5.

Messina, J. P., **M. U. G. Kraemer**, O. J. Brady, D. M. Pigott, F. M. Shearer, D. J. Weiss, N. Golding, C. W. Ruktanonchai, K. Khan, A. J. Tatem, T. Jaenisch, C. J. L. Murray, F. Marinho, T. W. Scott, and S. I. Hay. 2016. Mapping global environmental suitability for Zika virus. eLife:10.7554/eLife.15272.

*joint first authors

Chapter 4 – Human mobility and dengue transmission in Pakistan.

This paper has been accepted and published in final form. I was overseeing all aspects of the work including data collection, conception and design of the study (with RCR), modelling and interpretation of results (with RCR), as well as writing and revising the manuscript (with RCR).

Kraemer, M. U. G., T. A. Perkins, D. A. T. Cummings, R. Zakar, S. I. Hay, D. L. Smith, and R. C. Reiner. 2015. Big city, small world: density, contact rates, and transmission of dengue across Pakistan. Journal of the Royal Society Interface 12:20150468.

Chapter 5 – The impact of human mobility on the transmission dynamics of Ebola virus disease in West Africa.

This paper is revised and resubmitted to a scientific journal. I was responsible for all aspects of the work, including conception and design (with SIH, RCR, NG), data collection (with NG), analysis and interpretation (with RCR). I wrote the first draft of the manuscript which was subsequently approved by all authors.

Kraemer, M. U. G., N. Golding, D. Bisanzio, S. Bhatt, D. M. Pigott, O. J. Brady, N. R. Faria, O. G. Pybus, D. L. Smith, A. J. Tatem, S. I. Hay, and R. C. J. Reiner. 2016. Predicting the geographic spread of the 2014-2016 West Africa Ebola virus disease outbreak..

Chapter 6 – The spatial spread of a Yellow fever virus outbreak in Central Africa.

This paper is currently under initial submission to a scientific journal. I was responsible for all aspects of the work. I conceived the study (with SC), collected the data (with BN, MAJ, HS), analysed the data (with SC, RCR), implemented the modelling (with RCR), interpreted the results (with SC) and wrote the first draft of the manuscript which was subsequently approved by all authors.

Kraemer, M. U. G., N. R. Faria, R. C. J. Reiner, N. Golding, B. Nikolay, M. A. Johansson, H. Salje, G. R. W. Wint, F. M. Shearer, D. Bisanzio, H. H. Nax, B. S. R. Pradelski, N. R. Murphy,

J. S. Brownstein, A. J. Tatem, T. de Oliveira, D. L. Smith, O. G. Pybus, S. I. Hay, and S. Cauchemez. Spread of Yellow Fever Virus outbreak in Angola and the Democratic Republic Congo 2015-2016.

In addition to the articles above there are a number of other published articles that complement the topics discussed in this thesis. All of the articles in their final form have been added as appendices to this document.

This article models the temperature suitability of both *Aedes* mosquito species in relation to dengue virus transmission.

Brady, O. J., N. Golding, D. M. Pigott, **M. U. G. Kraemer**, J. P. Messina, R. C. Reiner, T. W. Scott, D. L. Smith, P. W. Gething, and S. I. Hay. 2014. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. Parasites & vectors 7:338.

This article is the first assessment of the origin of Zika virus in Brazil of the 2015-2016 outbreak.

Faria, N. R.*, S. Azevedo*, **M. U. G. Kraemer,** R. Souza, M. S. Cunha, S. C. Hill, J. Thézé, M. B. Bonsall, T. A. Bowden, I. M. Rocco, J. S. Nogueira, A. Y. Maeda, F. Giseli, F. Luiz, D. L. Macedo, A. Suzuki, S. G. Rodrigues, A. C. Ribeiro, B. T. Nunes, D. Barbosa, D. A. Medeiros, D. Sueli, G. Rodrigues, L. N. Queiroz, E. Vieira, D. F. Henriques, E. Salbe, C. S. De Oliveira, L. C. Martins, H. B. Vasconcelos, L. M. Neves, D. D. B. Simith, J. P. Messina, L. Abade, J. Lourenço, L. C. Junior, M. M. De Lima, M. Giovanetti, S. I. Hay, R. S. De Oliveira, S. Lemos, L. F. De Oliveira, C. Pereira, S. De Lima, S. Patroca, M. De Vasconcelos, L. Franco, J. F. Cardoso, and J. Lídio. 2016. Zika virus in the Americas: Early epidemiological and genetic findings. Science 352:345–349.

*joint first authors

In this commentary we model the likely invasion of Zika virus in the Americas that originate from Brazil based on underlying environmental suitability defined in chapter 1.

Bogoch, I. I., O. J. Brady, **M. U. G. Kraemer**, M. German, M. I. Creatore, M. A. Kulkarni, J. S. Brownstein, S. R. Mekaru, S. I. Hay, E. Groot, A. Watts, and K. Khan. 2016. Anticipating the international spread of Zika virus from Brazil. The Lancet 387:335–336.

This modelling work uses outputs from Chapter 1 to assess the cumulative number of people infected with Zika virus in the Americas after the first wave epidemic.

Perkins, T. A., A. S. Siraj, C. W. Ruktanonchai, **M. U. G. Kraemer**, and A. J. Tatem. 2016. Model-based projections of Zika virus infections in childbearing women in the Americas. Nature Microbiology 16126.

This article builds upon the *Aedes* suitability surfaces discussed in Chapter 1 and discusses the risk of spread of Zika virus in Europe.

Rocklöv, J.*, M. Brandon*, B. Sudre*, M. German*, **M. U. G. Kraemer,** O. Brady, I. I. Bogoch, J. Liu-helmersson, A. Wilder-smith, J. C. Semenza, M. Ong, K. Kaasik, and K. Khan. 2016. Assessing Seasonal Risks for the Introduction and Mosquito-borne Spread of Zika Virus in Europe. EBIOM 9:2050-256.

*joint first authors

This paper combines phylogenetic, mapping and epidemiological analysis to estimate the likely locations of spread of Chikungunya virus in Brazil.

Nunes, M. R. T., N. R. Faria, J. M. de Vasconcelos, N. Golding, **M. U. G. Kraemer**, L. F. de Oliveira, R. D. S. D. S. Azevedo, D. E. A. da Silva, E. V. P. da Silva, S. P. da Silva, V. L. de Carvalho, G. E. Coelho, A. C. R. Cruz, S. G. Rodrigues, J. L. da Silva Gonçalves Vianez, B. T. D. Nunes, J. F. Cardoso, R. B. Tesh, S. I. Hay, O. G. Pybus, and P. F. da Costa Vasconcelos. 2015. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Medicine 13:102.

This paper is a review that discusses the many different approaches the risk of dengue has been projected in the past.

Messina, J. P., O. J. Brady, N. Golding, D. M. Pigott, **M. U. G. Kraemer,** T. W. Scott, G. R. W. Wint, D. L. Smith, and S. I. Hay. 2015. The many projected futures of dengue. Nature Reviews Microbiology 13 (4), 2030-9.

Further, there is a set of publications that have been published and are related to my thesis but do not represent results that can be directly related to the content above.

Pigott, D. M., N. Golding, A. Mylne, Z. Huang, A. J. Henry, D. J. Weiss, J. Oliver, **M. U. G. Kraemer**, D. L. Smith, C. L. Moyes, S. Bhatt, P. W. Gething, P. W. Horby, I. I. Bogoch, J. S. Brownstein, S. R. Mekaru, A. J. Tatem, and S. I. Hay. 2014. Mapping the zoonotic niche of Ebola virus disease in Africa. eLife 3:e04395.

Mylne, A., O. J. Brady, Z. Huang, D. M. Pigott, N. Golding, **M. U. G. Kraemer**, and S. I. Hay. 2014. A comprehensive database of the geographic spread of past human Ebola outbreaks. Scientific Data 1:140042.

Pigott, D. M., A. I. Millear, L. Earl, C. Morozoff, B. A. Han, F. M. Shearer, D. J. Weiss, O. J. Brady, **M. U. G. Kraemer**, C. L. Moyes, S. Bhatt, P. W. Gething, N. Golding, and S. I. Hay. 2016. Updates to the zoonotic niche map of Ebola virus disease in Africa. eLife 5:e16412.

Pigott, D. M., N. Golding, A. Mylne, Z. Huang, D. J. Weiss, O. J. Brady, **M. U. G. Kraemer**, and S. I. Hay. 2015. Mapping the zoonotic niche of Marburg virus disease in Africa. Transactions of the Royal Society of Tropical Medicine and Hygiene 109:366–378.

Pigott, D. M., and **M. U.G. Kraemer.** 2014. Enhancing infectious disease mapping with open access resources. Eurosurveillance 19:pii=20989.

As Moritz's supervisor, I hereby certify that the statement of contribution listed above is a fair representation of his work.



Professor Simon I Hay
September 1st 2016

Table of contents

Abstract **II**

Statement of Contribution and Associated Publications **III**

Table of Contents **XI**

Acknowledgements **XII**

Introduction.....1

Chapter 1: The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*.....20

Chapter 2: The global distribution and environmental suitability for chikungunya virus.....47

Chapter 3: The introduction, suitability, and spread of Zika virus.....60

Chapter 4: Human mobility and dengue transmission in Pakistan.....89

Chapter 5: The impact of human mobility on the transmission dynamics of Ebola virus disease in West Africa.....99

Chapter 6: The spatial spread of a Yellow fever virus outbreak in Central Africa.....122

Discussion.....144

A.1. Appendix to Introduction.....158

A.2. Appendix to Chapter 1.....170

A.3.1 Appendix to Chapter 3.1.....175

A.3.2 Appendix to Chapter 3.2.....214

A.4. Appendix to Chapter 4.....220

A.5. Appendix to Chapter 5.....228

A.6. Appendix to Chapter 6.....247

A.7. Additional Publications.....262

Acknowledgements

First , *Mum, Dad* and *Nora*, I cannot thank you enough for your love and support throughout the last 27 years.

I want to thank my scientific advisors *Simon Hay* and *David Smith*, as well as *Nick Golding* and *Bobby Reiner* for their advice and encouragement over the past three years. Thanks for all the important oversights you provided, the long lasting scientific collaborations and friendships you have opened up, and all that came along with it.

David, Oli, Janey, Sam, Kate, Donal et al. sharing an office with me wasn't always easy; thank you for the good time in the Oxford Zoology Department. *Nuno* and *Oliver Pybus*, thank you for adopting me into the phylogenetic community over the past months, I have learned a lot during my short but important time in your lab.

I want to also thank the German Academic Exchange Service (DAAD), Studienstiftung des Deutschen Volkes, Hertford College Oxford and the Wellcome Trust for their financial support and opportunities they provided.

The Cave men, *Jony* and *Peder* for many discussions and grass court tennis games, *Tobi, Lukas R* and *Lukas KR* for their long friendships, and *Philipp* and *Oliwia* for being such great housemates.

Introduction

Quantitatively mapping the spatial distributions of infectious diseases is key both to investigating their epidemiology and identifying populations at risk of infection. Important advances in data quality and methodologies have allowed for much better investigation of disease risk and its association with environmental factors. To understand infectious disease dynamics in outbreak situations requires quantification of a number of other factors that may influence the expected rise and subsequent decline in cases reported. Among these, incorporating seasonal and human behavioural processes in disease epidemiology remains particularly challenging. For example, connectivity among human populations, a key driver of pathogen dispersal, has increased sharply over the last century. However, recent years have also seen an increasing availability of data derived from mobile phones and other dynamic data sources, improving our ability to track and quantify these processes.

Recent Trends in Infectious Disease Cartography

Infectious disease cartography (or infectious disease mapping) describes the quantitative spatial prediction of disease transmission risk. It has been facilitated by the rapid increases in computing power over the past three decades and the increasing availability of spatially continuous data on environmental risk factors for diseases, such as derived from satellite imagery. High-resolution spatial data representing both environmental and socioeconomic predictors of disease transmission [1,2] and improvements in statistical modelling [3,4] have therefore increased our ability to characterize these patterns more accurately whilst also incorporating uncertainty [5].

Maps have become increasingly important in assessing the spatial distribution of disease and are now powerful tools for addressing major public health policies. The visual nature of maps is especially helpful in identifying spatial heterogeneity of infectious disease risk. Whereas traditional epidemiological modelling has focussed on transmission parameters and dynamics and the effect of interventions (the 'what'), mapping such risk spatially largely helps to pinpoint

areas 'where' targeted interventions should be focussed. Today, novel approaches are being developed to collate information from a variety of sources, produce risk maps and disseminate them effectively to scientists, public health professionals, and policy makers in real time [6].

Below I review current approaches in disease mapping, how they can complement analyses of infectious disease dynamics and integrate aspects of human behavioural changes that must become an integral part of such modelling efforts.

Approaches for Mapping Infectious Disease

Historically disease mapping was confined to simply drawing case numbers on physical maps. Over the past decades the world has been mapped out digitally and information about disease outbreaks are recorded from numerous data sources. Major approaches to disease mapping today involve deterministic, correlative, and geo-statistical modelling techniques (see details in Figure 1 and Table 1). For policy makers to translate these maps into action, levels of confidence must be assigned to them which is often challenging.

Rationale for environmental disease mapping

Environmental conditions are a strong predictor of risk for many infectious diseases of humans. The link is particularly strong for some vector-borne diseases, whose distributions are driven by the conditions influencing disease vector abundance and pathogen replication rate in the insect host [7,8]. Major vector-borne diseases such as dengue (DENV) [9], chikungunya (CHIKV) [10], and malaria [11] are therefore closely associated with patterns in rainfall [12], temperature [13] and associated determinants such as deforestation [14] and biodiversity [15]. These relationships can be harnessed in order to produce global maps of the likely distributions of both vector and disease. Approaches range from simplified deterministic approaches (mathematical formulation of relationship) to statistical and geo-spatial models that each focus on different aspects of the interaction of factors influencing disease (Table 1).

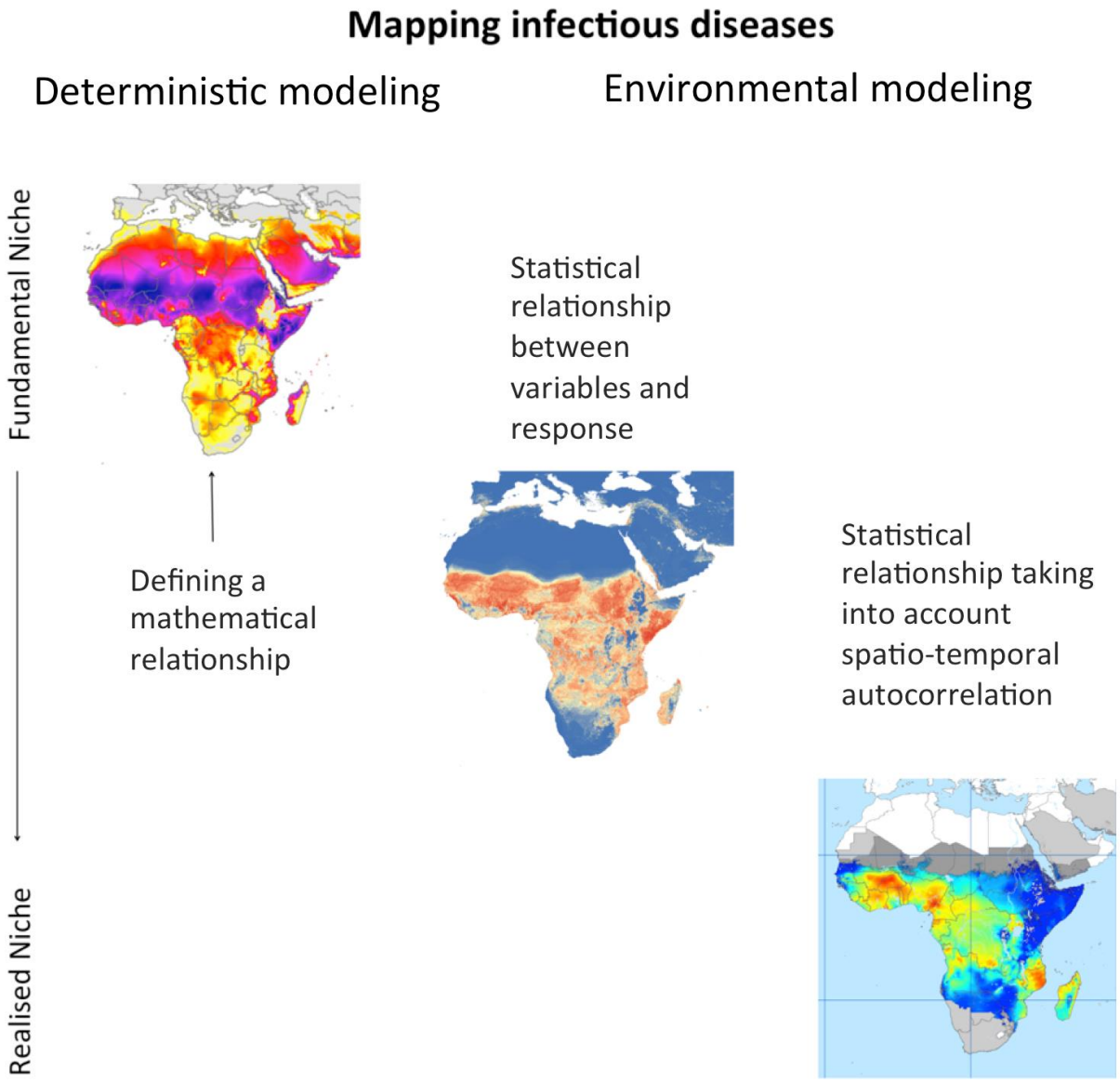


Figure 1. Approaches to mapping infectious diseases; from deterministic modelling to geostatistical modelling. Deterministic mapping can be seen as the representation of the fundamental niche of a species or pathogen, *i.e.* areas where the temperature is suitable for transmission as shown in top left map [16]. Environmental modelling techniques incorporate a wider range of covariates that can influence the distribution of disease, such as vegetation coverage and urbanisation shown here for *Aedes aegypti* [17]. In geostatistical approaches the true distribution (realised niche) of disease is further approximated including spatial autocorrelation and the effect of disease control and intervention as shown in bottom right map for malaria [18]. Figure adopted from [19].

Table 1. Environmental Modelling in Infectious Disease Cartography - Main Features and Limitations.

Modelling approach	Key features	Data requirements	Examples	Limitations	Refs
Deterministic modelling	Identification of key thermal and hydrological parameters (derived from field and laboratory experiments) Key parameters are subsequently extrapolated to regions where there is no data available.	Laboratory and or field estimates of relationship between survival of vector and/or pathogen.	Diseases with strict relationship between a single factor and disease or vector presence Malaria, dengue, West Nile virus, and Zika virus.	Singles out environmental factors.	16,20,21
Statistical environmental modelling	Extends deterministic modelling by including multiple environmental predictors to map a species spatial distribution.	Set of environmental and socio-economic cofounders.	This mapping approach can capture spatial patterns of vector borne diseases such as malaria and dengue but also Ebola and influenza.	It fails to include the spatial autocorrelation	22–24
Geostatistical modelling	Major advantage is the ability to model the spatial autocorrelation.	Large set of space-time disease prevalence and incidence data.	Many advances in the field have been made in malaria and schistosomiasis modelling. Detailed information about intervention strategies and their effect over time.	This approach is largely limited by the availability of data.	25

Mapping the socioeconomic correlates of disease

Historically, research focus has been on tying spatial models of disease to climatic factors. The spatial distributions of many important deadly diseases, however, are likely driven significantly by socioeconomic risk factors, such as population density, economic development and associated determinants such as sanitation infrastructure and access to health facilities [26]. The burden of a disease in any given location also crucially depends on the number of people experiencing such conditions [27], and therefore understanding the specific demographics of these populations enables more accurate assessments of disease burden and appropriate interventions [28,29]. All of these factors are highly spatially variable and are rapidly changing [30]. Since 1950 the global population has tripled, the proportion of people living in urban areas has increased from 29% to 53%, and the number of people living in extreme poverty has dropped by 50% since the 1990s (http://esa.un.org/unpd/wpp). These socioeconomic changes will continue over the coming decades and reliably mapping disease risk both today and into the future will depend on the availability of spatial estimates of these factors, particularly in countries that will likely undergo the greatest shifts (such as China, India, and Brazil). The advent and continued development of remote sensing technology has enabled the high-resolution global mapping of environmental conditions such as temperature and rainfall. In contrast, accurately quantifying spatial variation in key socioeconomic disease risk factors is much more difficult.

Mapping settlements and populations

For disease mapping it is crucial to understand the distribution of populations to understand the potential of humans to be infected by a particular pathogen. Human population mapping projects, such as the Global Rural-Urban Mapping Project (GRUMP) (http://sedac.ciesin.columbia.edu) and WorldPop (http://www.worldpop.org.uk) aim to develop global maps of the locations and sizes of urban areas and other human settlements. Typically data to inform these maps are

obtained from national censuses and surveys at specific locations, though these data have a coarse spatial resolution and for many of the most resource-poor parts of the world may be many years out of date. As with environmental mapping, remote sensing can be used to quantify related variables such as night-time lights, peri-domestic habitats, and roofs [31,32]. However, disentangling population density and economic prosperity makes robust mapping of these factors a very difficult task.

Several recent technological advances offer an opportunity to drastically increase the accuracy of spatial estimates of population density. The rapidly falling cost of cell phones and increases in global coverage have led to high levels of use in even the most resource-poor settings (http://www.itu.int). By triangulating individual cell phones, it may be possible in the near future to accurately map settlements, estimate spatial variation in population densities and to assess how these change through time [33]. An ever-increasing availability of high-resolution satellite imagery in conjunction with tools for online collaboration has led to mapping projects such as Open StreetMap developing highly detailed and openly available maps of built up areas and road networks [34]. This information offers much increased precision and accuracy for developing global maps of population densities [35]. As with the census data used to estimate population densities, coverage of these maps tends to be poorer in resource limited locations. Building on the success of emergency mapping drives in the wake of natural disasters (http://hot.openstreetmap.org) [36], public health and humanitarian agencies have recently sought to improve mapping in resource poor areas *via* the Missing Maps Project, which aims to provide detailed crowd-sourced cartographic information for the most vulnerable and disease-prone parts of the world within the next two years (http://www.msf.org.uk/missing-maps-project). In addition there is novel work trying to infer economic variables based on mobile phone data and night time lights which offers interesting new insights into local poverty levels [37,38].

Infectious Diseases and Human and Trade Networks

In tandem with the unprecedented increases in human population over the past 50 years, connectivity among human populations has increased rapidly. Increasing urbanization, economic growth, improvements in transport infrastructure and changing trade and immigration policies [30,39,40] have led to populations in each successive generation becoming vastly more mobile than the last, with regular international travel now the norm for a significant fraction of the global population.

For many zoonotic and vector-borne diseases, and for all directly-transmitted infections, introduction of the disease into currently non-endemic locations is strongly influenced by human movement and trade [41–44]. Similarly, local dynamic epidemic patterns in endemic regions are often driven by human movements of infected people on several spatial and temporal scales [26,45]. Until now it has been difficult for public health policy makers to make sense of disease spread during a specific outbreak, mainly due the difficulty in accurately accounting for human movement in real time.

For many vector-borne diseases, only a limited number of species are able to act as capable vectors and the disease can therefore only be transmitted where they are present [17]. In these instances the introduction of the vector into a new region is a prerequisite for disease establishment [46]. The relatively low dispersal ability of most disease vectors means that long-distance vector spread is typically driven by human population connectivity, such as via international shipping [47]. Similarly, international trade in livestock has led to the introduction of infected animals into new areas [48], and human settlement expansion has increased the likelihood of animal to human infections for zoonotic diseases [49].

Diseases which were once restricted to discrete and isolated regions can now pose a risk across the world [50–53]. Accounting for the impact of human mobility on the spatial distribution of disease risk is now a critical challenge in adapting infectious disease cartography to the modern age.

Quantifying connectivity

Incorporation of connectivity into predictive disease mapping requires quantification of the degree to which populations are connected across spatial scales [54]. There is a large body of work that is trying to understand the general movement patterns of human populations, their specific attributes and limitations at different spatial and temporal scales (Table 2). One common limiting factor of the diverse data sources that capture human mobility is that they are often confined to political boundaries. With increasing international travel and concern over international spread of disease, quantifying how people move across borders is of paramount importance. Such international movement patterns may be assessed using long-term migration data [55,56], but it is unclear how representative this is of shorter-term movements, such as international travel for trade or to visit relatives. Data on short-term border crossings might be available for small geographical areas. Synthesising these two sensibly could provide useful tools for future cross-country disease modelling. Though with no single 'gold-standard' mobility metric, and methodological differences in obtaining data at different scales, this is a daunting task. In this era of Big Data, individual-based metadata are increasingly available and stored by large private and public institutions. However, the collection and use of individual-based information is associated with serious privacy concerns, which need to be considered carefully [57]. To date, no common guidelines exist for dealing with user re-identifiability using geolocated data of simply anonymized data sources [58]. Developing and implementing such guidelines will enable the wider use of these invaluable data, whilst safeguarding the rights of individuals.

Not only human movements, but also the movements of goods have significantly contributed to the international spread of the dengue and chikungunya virus vectors *Aedes aegypti* and *Aedes albopictus* [59–61]. Quantifying the international movement of goods transported between ports remains challenging as this information is of high economic value and therefore not readily shared with researchers and policy makers. Similarly, continental-scale trade of live animals has been linked to the emergence of highly pathogenic avian influenza (HPAI) in Asia and quantifying

the extent of such trade, especially in Asia is challenging [23,62]. In addition, natural calamities such as floods may contribute to large scale displacement of animal populations and can subsequently cause the introduction of novel pathogens in new populations.

Incorporating connectivity into disease cartography

The impact of human connectivity in disease transmission and spread has been the subject of much contemporary epidemiological research. For example, theoretical and empirical models of network contagion processes have been applied to interpret the global spread of pathogens via air travel [41]. Metapopulation models of disease transmission, building on those developed in ecology, have been applied to understand the local-scale spread of vaccine-preventable diseases and guide the effective deployment of resources [63,64]. Spatially-explicit simulation modelling has been applied to evaluate the likely impacts of differing control scenarios in the face of disease outbreaks and their associated economic costs [65,66]. Similarly, statistical models have been applied to reconstruct past transmission networks after disease introductions have taken place [67]. Improving the rapid assessment of outbreak dynamics and accurate modelling of number of infected individuals can directly impact public health policy interventions in affected and surrounding areas such as done in the Ebola response [68].

Whilst spatial spread of diseases has been widely studied, the predominant focus of this research has been on epidemiological inference, rather than the generation of predictive maps. The integration of this existing body of research and the methods it has developed into risk mapping offers the potential to broaden the scope and utility of disease maps. For example, if maps could identify likely paths of epidemics before they emerge, public health officials could implement responses much more timely.

The integration of human connectivity data into disease cartography and public health policy planning has been applied in the context of malaria where long-term human migration data has been used to identify groups of countries among which malaria cases are most likely to be imported [69]. Moreover, at a finer spatial scale, human movement data has been used to map

regions acting as sources and sinks for malaria importations and this information has been combined with environmental risk mapping to provide rapidly-updated risk maps as a tool for malaria control in countries seeking to eliminate the disease [70]. This research has demonstrated the need for mobility to be integrated in guiding elimination strategies, as they can only be successful if both local cases are reduced along with the threat of new introductions from endemic regions.

These examples illustrate both the potential and the feasibility of integrating human connectivity into disease cartography more widely. Malaria is among the most comprehensively studied of the major infectious diseases, and we have a detailed understanding of disease transmission processes, as well as systematic monitoring of disease prevalence in specific locations. In contrast, the vast majority of the world's infectious diseases are more poorly understood and reliable data on their distributions is scarcer pointing towards prioritising the mapping of these particular diseases [71]. Challenges remain also in understanding how connectivity and its interaction with other epidemiological processes evolves over the different phases of an outbreak (growth phase vs. decline in number of cases) and there is debate over how such phases can best be represented in epidemiological models. Another important challenge will be understanding not only how human populations move in general but identifying specific group behaviours that are more likely to contract and disseminate the pathogen of interest (e.g. poor rural communities in Southeast Asia that have very different patterns of movement but high incidence of malaria). This is particularly important given that policy makers can only deploy limited resources when attempting to control and eliminate disease.

Whilst the scientific community works to increase this epidemiological knowledge base, the development of flexible statistical approaches to incorporate connectivity into disease modelling could enable improvements over existing risk mapping in the near future.

Table 2. Approaches to Quantifying and Mapping Human Mobility at a Range of Spatial and Temporal Scales.

Approach	Key features	Spatial and temporal scales	Data sources	Limitations	References
Long-term international migration census	Describing longer term (permanent) population dynamics allowing for understanding genetic diversity.	Country wide estimates; 5-10 years.	United Nations migration database	Heavily relying on large-scale studies. Often incomplete. Poor spatial and temporal resolution.	[55]
Within country census migration data	Within country population dynamics that might arise from economic and social transformations.	Typically at district level; 1, 5 and 10 years.	Integrated Public Use Microdata Series (IPUMS)	Captures only a small fraction of the population.	[56]
Flight networks	Allows for rapid assessment of international spread of infectious diseases.	Global coverage on ticketed commercial flight connections.	International Air Transport Association	Captures only flight connection but not the subsequent travel patterns within each of the countries.	[41]
Commuting networks	Key to understand regular human behavioural patterns allowing to predict city-to-city and country wide transmission.	Daily commuting patterns often on a country level.	National Census Surveys	Restricted to high-income countries with well documented national statistics.	[72]
Cell-phone data	Very useful as it captures daily/seasonal human mobility patterns.	Individual user specific patterns of movement	Mobile phone providers and mobile	Only captures a fraction of the population,	[73]

		at varying temporal resolution.	app companies.	often high income groups. Limited to national borders.	
Logging devices such as GPS	Highly accurate assessment of human contact networks.	High spatial and temporal resolution.	Individually funded projects where GPS devices are rolled out to the study group.	Cost and time intensive. Confined to small study groups.	[54]

Significance and aim of this thesis

With improvements in living standards, hygiene, drug therapy and widespread deployment of medical advances such as vaccinations, in the 1970s there was the widespread belief that the fraction of mortality and morbidity due to infectious disease would continue to dramatically decline [74]. Instead however, the last few decades have shown that as the world continues to grow in population and become more connected, infectious diseases still remain a challenge, such as exemplified by the Ebola outbreak in West Africa and the current Zika epidemic, as well as the continued globalisation of many other pathogens (DENV, CHIKV). The complex interplay of environmental and human behavioural changes result in a changing context of pathogen-host interactions, requiring new and varied approaches to manage them.

In this thesis I leverage a variety of techniques to map the distribution and chart the spread of a number of these pathogens, assessing both environmental aspects of these diseases and vectors, as well as the dynamic nature that determine the severity of regional outbreaks.

Chapter 1 applies ecological niche models to derive the global distribution of the two main vectors of DENV, CHIKV, ZIKV, and YFV, namely *Ae. aegypti* and *Ae. albopictus.* It also includes a data publication that describes the process of building the largest and most comprehensive database

to date on both species. Both vectors have been expanding rapidly over the past decades and are now present in all continents. These maps represent a current estimate of the limits and intensity of their distribution and have been widely used to inform policies to implement vector surveillance.

By using *Aedes* suitability surfaces in addition to other environmental variables I derive the global environmental suitability of chikungunya virus (CHIKV) in addition to compiling comprehensive evidence for its reported presence and absence globally. CHIKV distribution is characterized by its rapid expansion globally which was the aim of this chapter. These maps comprise the first global estimates that show current evidence of its distribution in addition to identifying locations where it may spread next.

Zika virus (ZIKV) has spread to the new world between May and December 2013 and is now spreading within the Americas at unprecedented speed [75]. Chapter 3 consist of two publications; the first defining the potential for spread of ZIKV globally again making use of outputs from the *Aedes* suitability maps presented earlier; and the second analysing the seasonal variation in risk of introduction and local spread from regions that at time of analysis reported active ZIKV transmission using commercial passenger flight data.

In Chapter 4, I investigate the drivers of the regional emergence of dengue in Pakistan at epidemiologically relevant spatial and temporal scales. I analysed weekly data on dengue hospital admissions using time-series models in combination with weekly time-varying environmental data trying to understand the local dynamics of an emerging disease in a city vs. a rural context. Evidence from this chapter has direct implications for control and prevention in an outbreak situation.

The emergence and spread of Ebola virus disease in West Africa constituted a 'perfect storm' as described by Peter Piot that presumably originated from a single spillover event (from the animal reservoir to the human population) [76]. The subsequent rapid geographic expansion was believed to be driven by human movements [77]. Here I draw upon models developed in previous chapters,

but include a spatially explicit aspect that can explain the spread of Ebola in West Africa during different phases of the outbreak (the expanding vs. the contracting phase) and identify spatio-temporal heterogeneity of risk. This allowed to directly infer the relative roles of country specific vs. between country movements to the spread of the pathogen in the region.

The first urban YFV outbreak since the 1980s is currently affecting Central Africa. In the last Chapter, I combine *Aedes* suitability maps with human mobility data to derive estimates of the risk of introduction (spread) for the ongoing outbreak in Central Africa in real-time. Such modelling outputs are important to identify where and when vaccine delivery should be deployed, especially as vaccine stocks are limited.

Concluding Remarks

Technological advances over the past decade have transformed infectious disease cartography from providing visualisations of spatial data and qualitative descriptions of disease risk into a framework for providing robust, quantified predictive maps to guide public health policy. The rapid pace of global change over the same time period has brought socio-economic factors and global connectivity to the forefront as factors governing global disease risk. Infectious disease cartography has begun to adapt to these changing priorities, integrating new data sources and types of analysis into a unified cartographic framework. Particularly, data on disease control efforts need to be rigorously incorporated in disease mapping to better inform public health policy makers on their intervention strategies.

The world is increasingly connected and now more than ever diseases can spread rapidly across the globe. Simultaneously, a number of the world's major infectious diseases, such as cholera, are being brought under control and being locally eliminated [78,79]. Keeping track of the ever-changing global landscape of disease risk is crucial for limiting disease outbreaks and guiding control strategies.

Future work may be targeted towards an open data policy to allow reproducibility and increase transparency [80]. Moving from static disease mapping to continuously updated maps of contemporary disease risk poses significant technical challenges, but will enable the rapid translation of disease data into public health policy.

References

1. Nordhaus, W. D. Geography and macroeconomics: new data and new findings. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3510–7 (2006).
2. Tatem, A. J. Mapping population and pathogen movements. *Int. Health* **6,** 5–11 (2014).
3. Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecogeography* **29,** 129–51 (2006).
4. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning*. (Springer, 2009).
5. Lindström, T., Tildesley, M. & Webb, C. A bayesian ensemble approach for epidemiological projections. *PLOS Comput. Biol.* **11,** e1004187 (2015).
6. Hay, S. I., George, D. B., Moyes, C. L. & Brownstein, J. S. Big data opportunities for global infectious disease surveillance. *PLoS Med.* **10,** e1001413 (2013).
7. Lounibos, L. P. Invasion by insect vectors of human disease. *Annu. Rev. Entomol.* **47,** 233–66 (2002).
8. Tsetsarkin, K. a & Weaver, S. C. Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence. *PLoS Pathog.* **7,** e1002412 (2011).
9. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–7 (2013).
10. Nsoesie, E. O. *et al.* Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. *Eurosurveillance* **21,** pii=30234 (2015).
11. Hay, S. I. *et al.* A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med.* **6,** e1000048 (2009).
12. Hii, Y. L., Zhu, H., Ng, N., Ng, L. C. & Rocklöv, J. Forecast of dengue incidence using temperature and rainfall. *PLoS Negl. Trop. Dis.* **6,** e1908 (2012).
13. Weiss, D. J. *et al.* Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malar. J.* **13,** 171 (2014).
14. Hahn, M. B., Gangnon, R. E., Barcellos, C., Asner, G. P. & Patz, J. a. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLoS One* **9,** e85725 (2014).
15. Cunningham, A. a, Dobson, A. P. & Hudson, P. J. Disease invasion: impacts on biodiversity and human health. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367,** 2804–6 (2012).
16. Brady, O. J. *et al.* Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission. *Parasit. Vectors* **7,** 338 (2014).
17. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* **4,** e08347 (2015).
18. Gething, P. W. *et al.* A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar. J.* **10,** 378 (2011).
19. Kraemer, M. U. G. *et al.* Progress and challenges in infectious disease cartography. *Trends Parasitol.* **32,** 19–29 (2016).
20. Brady, O. J. *et al.* Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. *Parasit. Vectors* **6,** 351 (2013).
21. Gething, P. W. *et al.* Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax. *Parasit. Vectors* **4,** 92 (2011).

22. Pigott, D. M. *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* **3,** e04395 (2014).

23. Gilbert, M. *et al.* Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat. Commun.* **5,** 4116 (2014).

24. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–7 (2013).

25. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526,** 207–11 (2015).

26. Gong, P. *et al.* Urbanisation and health in China. *Lancet* **379,** 843–52 (2012).

27. Tatem, A. J. Mapping the denominator: spatial demography in the measurement of progress. *Int. Health* **6,** 153–5 (2014).

28. Tatem, A. J. *et al.* Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *Int. J. Health Geogr.* **13,** 1–11 (2014).

29. Murray, C. J. L. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2197–223 (2012).

30. Alirol, E., Getaz, L., Stoll, B., Chappuis, F. & Loutan, L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* **11,** 131–41 (2011).

31. Bharti, N. *et al.* Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334,** 1424–7 (2011).

32. Sutton, P., Roberts, D., Elvidge, C. & Baugh, K. Census from Heaven: An estimate of the global human population using night-time satellite imagery. *Int. J. Remote Sens.* **22,** 3061–3076 (2001).

33. Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **111,** 15888–93 (2014).

34. Bakillah, M., Liang, S., Mobasheri, A., Jokar Arsanjani, J. & Zipf, A. Fine-resolution population mapping using OpenStreetMap points-of-interest. *Int. J. Geogr. Inf. Sci.* **28,** 1940–63 (2014).

35. Linard, C., Gilbert, M., Snow, R. W., Noor, A. M. & Tatem, A. J. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* **7,** e31743 (2012).

36. Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & von Schreeb, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med.* **8,** e1001083 (2011).

37. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350,** 1073–76 (2015).

38. Jean, N. *et al.* Combining satellite imagery and machine learning to predict poverty. *Science* **353,** 790–4 (2016).

39. Huang, Z., Wu, X., Garcia, A. J., Fik, T. J. & Tatem, A. J. An open-access modeled passenger flow matrix for the global air network in 2010. *PLoS One* **8,** (2013).

40. Lambin, E. F. & Meyfroidt, P. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 3465–72 (2011).

41. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342,** 1337–42 (2013).

42. Nunes, M. R. T. *et al.* Air travel is associated with intracontinental spread of dengue virus serotypes 1-3 in Brazil. *PLoS Negl. Trop. Dis.* **8,** e2769 (2014).

43. Gubler, D. J. Dengue, urbanization and globalization: the unholy trinity of the 21st

century. *Trop. Med. Health* **39,** 3–11 (2011).

44. Guagliardo, S. A. *et al.* Patterns of geographic expansion of Aedes aegypti in the Peruvian Amazon. *PLoS Negl. Trop. Dis.* **8,** e3033 (2014).
45. Horby, P. W., Pfeiffer, D. & Oshitani, H. Prospects for emerging infections in east and southeast Asia 10 years after severe acute respiratory syndrome. *Emerg. Infect. Dis.* **19,** 853–60 (2013).
46. Schaffner, F., Medlock, J. M. & Van Bortel, W. Public health significance of invasive mosquitoes in Europe. *Clin. Microbiol. Infect.* **19,** 685–92 (2013).
47. Tatem, A. J., Hay, S. I. & Rogers, D. J. Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6242–7 (2006).
48. Kilpatrick, A. M. & Randolph, S. E. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **380,** 1946–55 (2012).
49. Moyes, C. L. *et al.* Defining the geographical range of the Plasmodium knowlesi reservoir. *PLoS Negl. Trop. Dis.* **8,** e2780 (2014).
50. Bogoch, I. I. *et al.* Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* **6736,** 1–7 (2014).
51. Khan, K. *et al.* Potential for the International Spread of Middle East Respiratory Syndrome in Association with Mass Gatherings in Saudi Arabia. *PLOS Curr. Outbreaks* (2013). doi:10.1371/currents.outbreaks.a7b70897ac2fa4f79b59f90d24c860b8.Abstract
52. Khan, K. *et al.* Assessing the origin of and potential for international spread of Chikungunya virus from the Caribbean. *PLOS Curr. Outbreaks* **Jun 6,** (2014).
53. Smith, K. F., Sax, D. F., Gaines, S. D., Guernier, V. & Guegan, J.-F. Globalization of human infectious diseases. *Ecology* **88,** 1903–10 (2007).
54. Stoddard, S. T. *et al.* House-to-house human movement drives dengue virus transmission. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 994–9 (2013).
55. Abel, G. J. & Sander, N. Quantifying global international migration flows. *Science* **343,** 1520–2 (2014).
56. Sorichetta, A. *et al.* Mapping internal connectivity through human migration in malaria endemic countries. *Sci. Data* **3,** 160066 (2016).
57. Vayena, E., Salathé, M., Madoff, L. C. & Brownstein, J. S. Ethical Challenges of Big Data in Public Health. *PLOS Comput. Biol.* **11,** e1003904 (2015).
58. Montjoye, Y. De, Radaelli, L. & Singh, V. K. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* **347,** 356–9 (2015).
59. Weaver, S. C. & Lecuit, M. Chikungunya virus and the global spread of a mosquito-borne disease. *N. Engl. J. Med.* **372,** 1231–9 (2015).
60. Hawley, W. A., Reiter, P., Copeland, R. S., Pumpuni, C. B. & Craig, G. B. Aedes albopictus in North America: probable introduction in used tires from northern Asia. *Science* **236,** 1114–6 (1987).
61. Tatem, A. J., Rogers, D. J. & Hay, S. I. Global transport networks and infectious disease spread. *Adv. Parasitol.* **62,** 293–343 (2006).
62. Kilpatrick, a M. *et al.* Predicting the global spread of H5N1 avian influenza. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 19368–19373 (2006).
63. Metcalf, C. J. E. *et al.* Seven challenges in modeling vaccine preventable diseases. *Epidemics* 3–7 (2014). doi:10.1016/j.epidem.2014.08.004

64. Lessler, J. & Grenfell, B. T. Measurement of vaccine-derived immunity: how do we use all the data ? *Expert Rev. Vaccines* **11,** 747–749 (2012).

65. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437,** 209–14 (2005).

66. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442,** 448–52 (2006).

67. Cauchemez, S. *et al.* Local and regional spread of chikungunya fever in the Americas. *Eurosurveillance* **19,** 20854 (2014).

68. WHO Ebola Response Team. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. *N. Engl. J. Med.* (2014). doi:10.1056/NEJMoa1411100

69. Tatem, A. J. & Smith, D. L. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12222–7 (2010).

70. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338,** 267–70 (2012).

71. Hay, S. I. *et al.* Global mapping of infectious disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120250 (2013).

72. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484,** 96–100 (2012).

73. Wesolowski, A. *et al.* Commentary: Containing the Ebola outbreak – the potential and challenge of mobile network data. *PLOS Curr. Outbreaks* **Sep 29,** (2014).

74. WHO (World Health Organisation). *Global strategy for health for all by the year 2000*. (1981).

75. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).

76. Dudas, G. & Rambaut, A. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. 1–11 (2014). doi:10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.Abstract

77. Merler, S. *et al.* Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *Lancet Infect. Dis.* **15,** 204–211 (2015).

78. Chiyaka, C. *et al.* The Stability of Malaria Elimination. *Science* **339,** 909–10 (2013).

79. Andrus, J. K., Solorzano, C. C., de Oliveira, L., Danovaro-Holliday, M. C. & de Quadros, C. a. Strengthening surveillance: confronting infectious diseases in developing countries. *Vaccine* **29 Suppl 4,** D126-30 (2011).

80. Pigott, D. M. & Kraemer, M. U. Enhancing infectious disease mapping with open access resources. *Eurosurveillance* **19,** pii=20989 (2014).

Chapter 1 – The global distribution of the arbovirus vector *Aedes aegypti* and *Aedes albopictus*

This first research chapter consist of two publications; one modelling the predicted environmental distribution of *Aedes aegypti* and *Ae. albopictus*; the second describing the methods and sources used to comprise database for both vectors. This chapter lays the ground work for subsequent research chapters that derive the global distribution of chikungunya virus (Chapter 2), Zika virus seasonal distribution (Chapter 3) and dynamics of dengue (Chapter 4) and yellow fever (Chapter 6). Both publications in this chapter are published in final form in *eLife* and *Nature Scientific Data*, respectively.

eLIFE elifesciences.org

RESEARCH ARTICLE

The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*

Moritz UG Kraemer[1]*, Marianne E Sinka[1], Kirsten A Duda[1], Adrian QN Mylne[2], Freya M Shearer[2], Christopher M Barker[3], Chester G Moore[4], Roberta G Carvalho[5], Giovanini E Coelho[5], Wim Van Bortel[6], Guy Hendrickx[7], Francis Schaffner[7], Iqbal RF Elyazar[8], Hwa-Jen Teng[9], Oliver J Brady[2], Jane P Messina[1], David M Pigott[1,2], Thomas W Scott[10,11], David L Smith[1,10,12], GR William Wint[13], Nick Golding[2], Simon I Hay[2,10,14]*

[1]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom; [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; [3]Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, Davis, United States; [4]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, United States; [5]National Dengue Control Program, Ministry of Health, Brasilia, Brazil; [6]European Centre for Disease Prevention and Control, Stockholm, Sweden; [7]Avia-GIS, Zoersel, Belgium; [8]Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia; [9]Center for Research, Diagnostics and Vaccine Development, Centers for Disease Control, Taipei, Taiwan; [10]Fogarty International Center, National Institutes of Health, Bethesda, United States; [11]Department of Entomology and Nematology, University of California, Davis, Davis, United States; [12]Sanaria Institute for Global Health and Tropical Medicine, Rockville, United States; [13]Environmental Research Group Oxford, Department of Zoology, University of Oxford, Oxford, United Kingdom; [14]Institute for Health Metrics and Evaluation, University of Washington, Seattle, United States

***For correspondence:** moritz.kraemer@zoo.ox.ac.uk (MUGK); simon.hay@well.ox.ac.uk (SIH)

Competing interests: See page 12

Funding: See page 12

Received: 26 April 2015
Accepted: 18 June 2015
Published: 30 June 2015

Reviewing editor: Mark Jit, London School of Hygiene & Tropical Medicine, and Public Health England, United Kingdom

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Abstract Dengue and chikungunya are increasing global public health concerns due to their rapid geographical spread and increasing disease burden. Knowledge of the contemporary distribution of their shared vectors, *Aedes aegypti* and *Aedes albopictus* remains incomplete and is complicated by an ongoing range expansion fuelled by increased global trade and travel. Mapping the global distribution of these vectors and the geographical determinants of their ranges is essential for public health planning. Here we compile the largest contemporary database for both species and pair it with relevant environmental variables predicting their global distribution. We show *Aedes* distributions to be the widest ever recorded; now extensive in all continents, including North America and Europe. These maps will help define the spatial limits of current autochthonous transmission of dengue and chikungunya viruses. It is only with this kind of rigorous entomological baseline that we can hope to project future health impacts of these viruses.
DOI: 10.7554/eLife.08347.001

Introduction

The mosquitoes *Aedes aegypti* [= *Stegomyia aegypti*] and *Aedes albopictus* [= *Stegomyia albopicta*] (***Reinert et al., 2009***) are vectors of several globally important arboviruses, including dengue virus (DENV) (***Simmons et al., 2012***), yellow fever virus (***Jentes et al., 2011***), and chikungunya virus (CHIKV) (***Leparc-Goffart et al., 2014***). The public health impact of DENV and CHIKV has increased

eLife digest Mosquitoes spread many disease-causing viruses and parasites between people and other animals, including viral infections such as dengue and chikungunya. Both infections cause high fevers often accompanied with excruciating joint pain or other flu-like symptoms. Dengue and chikungunya have become growing public health problems over the last fifty years. Today about half of the world's population is at risk of dengue infection, while chikungunya outbreaks, which were previously limited to Africa and Asia, have recently been reported in the Caribbean, South America and Europe.

The dengue and chikungunya viruses are transmitted between people by two species of mosquitoes called *Aedes aegypti* and *Ae. albopictus*. Therefore it is important to work out where these mosquito species are found around the globe to identify the areas at risk. It is also important to predict where these species could become established if they were introduced, in order to identify areas that could become at risk in the future.

Kraemer et al. now provide updated predictions about the distribution of these two mosquito species around the globe. These predictions are based upon the most up-to-date data on the known locations of the species combined with information on environmental conditions across the globe. The updated maps show that these *Aedes* mosquitoes are now found across all continents, including North America and Europe.

Aedes albopictus mosquitoes in particular are rapidly expanding their territory around the globe. Kraemer et al. used their new maps to show that, unlike in the United States, many of the areas in Europe and China that could support this mosquito species do not yet appear to have been colonized.

These findings provide a map of the distribution of both species as it stands at the moment. Further work is now needed to better understand which factors are contributing to the rapid expansion of these mosquitoes' range and what might be done to control this spread.

DOI: 10.7554/eLife.08347.002

dramatically over the last 50 years, with both diseases spreading to new geographic locations and increasing in incidence within their range (**Weaver, 2014**). The remaining burden of vaccine-preventable yellow fever is similarly likely to be dramatically underestimated (**Garske et al., 2014**). DENV, with a nearly ubiquitous distribution in the tropics and more recently introduced to Europe (**ECDC, 2014**; **Schaffner and Mathis, 2014**), is the most prevalent human arboviral infection causing 100 million apparent annual infections world-wide with almost half of the world's population at risk of infection (**Brady et al., 2012**; **Bhatt et al., 2013**). CHIKV recently received considerable public health attention due to the outbreaks in Réunion in 2005–2006 (225,000 infections) (**Borgherini et al., 2007**), Italy in 2007 (205 infections) (**Rezza et al., 2007**), and France in 2010 and 2014 (2 and 11 locally transmitted cases, respectively) (**La Ruche et al., 2010**; **Grandadam et al., 2011**; **Paty et al., 2014**) as well as its recent invasion into the Americas with over 1 million cases recorded to date (**Cauchemez et al., 2014**; **Johansson et al., 2014**; **Morens and Fauci, 2014**). Increases in distribution and intensity of transmission are compounded by the lack of commercially available antivirals or vaccines for either disease (**Simmons et al., 2012**; **Roy et al., 2014**), although new therapeutics and vaccines are in development (**McArthur et al., 2013**; **Powers, 2014**; **Villar et al., 2015**). Similarly, while yellow fever infections have been on the decline due to extensive vector control and an effective vaccine developed more than 70 years ago, it still causes a significant disease burden in Africa and South America (**Poland et al., 1981**; **World Health Organization, 1990**; **Garske et al., 2014**). Given the public health impact of these diseases and their rapid global spread, understanding the current and future distribution, and determining the geographic limits of transmission and transmission intensity, will enable more efficient planning for disease control (**Carrington and Simmons, 2014**; **Semenza et al., 2014**; **Messina et al., 2015**). Because these diseases can only persist where their mosquito vectors, *Ae. aegypti* and *Ae. albopictus* are present, understanding the distributions of these two species underpins this strategy.

The global expansion of these arboviruses was preceded by the global spread of their vectors (**Charrel et al., 2014**). *Ae. aegypti* originated in Africa where its ancestral form was a zoophilic treehole mosquito named *Ae. aegypti formosus* (**Brown et al., 2014**). The domestic form *Ae. aegypti*

is genetically distinct with discrete geographic niches (**Brown et al., 2011**). It was hypothesised that due to harsh conditions coupled with the onset of the slave trade, *Ae. aegypti* were introduced into the New World from Africa, from where it subsequently spread globally to tropical and sub-tropical regions of the world (**Brown et al., 2014**). *Ae. albopictus,* originally a zoophilic forest species from Asia, spread to islands in the Indian and Pacific Oceans (**Delatte et al., 2009**). During the 1980s it rapidly expanded its range to Europe, the United States and Brazil (**Medlock et al., 2012**; **Carvalho et al., 2014**). Today both *Ae. aegypti* and *Ae. albopictus* are present in most Asian cities and large parts of the Americas (**Lambrechts et al., 2011**). *Ae. aegypti* feed almost exclusively on humans in daylight hours and typically rest indoors (**Scott and Takken, 2012**). In contrast *Ae. albopictus* is usually exophagic and bites humans and animals opportunistically (**Paupy et al., 2009**) but has also been shown to exhibit strongly anthropophilic behavior similar to *Ae. aegypti* in specific contexts (**Ponlawat and Harrington, 2005**; **Delatte et al., 2010**).

A number of previous studies have mapped the global or regional distributions of *Ae. aegypti* and *Ae. albopictus* by focusing on different aspects of their ecology. The majority examined the impacts of climatic conditions, often with an exclusive focus on temperature. **Kobayashi et al. (2002)** and **Nawrocki and Hawley (1987)** used results from laboratory studies to identify potential limits of establishment in Japan and Asia suggesting a minimum mean temperature in the coldest months of −2˚C and −5˚C respectively limits their distribution. **Brady et al. (2013)** extended that work by modeling the adult survival of both species under laboratory and field conditions, indicating that *Ae. albopictus* has higher survival rates than *Ae. aegypti*, though adults of the latter can tolerate a wider range of temperatures. Applying these results to global temperature data, **Brady et al. (2014)** produced maps indicating areas where the temperature is suitable for these vectors to persist. Whilst temperature is clearly a crucial factor constraining the distribution of the two species, these results alone are not sufficient to discriminate between areas where the species can and cannot persist. Other studies went further using statistical models, predicting the distributions of both species (though particularly *Ae. albopictus*) using a broader range of climatic variables including precipitation (**Benedict et al., 2007**; **Medley, 2010**; **Fischer et al., 2011**; **Caminade et al., 2012**; **Khormi and Kumar 2014**; **Campbell et al., 2015**).

Whilst these studies incorporated several generic climatic factors to predict the current and future distribution of the species, we were able to integrate a bespoke species-specific temperature suitability covariate and account for anthropogenic factors that are known to influence *Ae. aegypti* and *Ae. albopictus* distributions (**Reiter et al., 2003**). Both species are container-inhabiting but differ in their behaviour and biology so that they occupy different niches (**Eisen and Moore, 2013**). A few local studies showed, however, that local spread of *Ae. albopictus* and declining *Ae. aegypti* populations might be linked to inter-species competition (**O'Meara et al., 1995**; **Daugherty et al., 2000**; **Juliano et al., 2007**) and/or non-reciprocal cross-species inseminations (**Bargielowski et al., 2013**). Socio-economic factors affecting the distribution of the *Aedes* mosquitoes other than the use of containers to store water, include the use of air-conditioning, housing quality, and the rate of urbanisation (**Ramos et al., 2008**; **Aström et al., 2012**). In addition to exclusively focusing on meteorological factors in determining the spatial extent of the *Aedes* mosquitoes, many models used small sets of input occurrence data, which were biased towards particular countries with well-developed surveillance systems, such as, Brazil and Taiwan (**Benedict et al., 2007**; **Medley, 2010**; **Fischer et al., 2011**; **Campbell et al., 2015**).

In this context, we set out to model the global distribution of these two important vector species, compiling the most comprehensive occurrence dataset to date from published literature and national entomological surveys. To overcome previous modelling limitations, a probabilistic species distribution model using Boosted Regression Trees (BRT) was produced for each vector. Our models combine environmental and, for the first time, land-cover variables to predict the global distribution of both species at high spatial resolution. Importantly, the models quantify prediction uncertainty and aim at identifying key contributing factors and inter-species differences in their environmental niches.

Results

In total, data collection yielded 19,930 and 22,137 spatially unique occurrence records for *Ae. aegypti* and *Ae. albopictus* respectively, which were used to train the distribution models. This includes up-to date records from national entomological surveys from Brazil and Taiwan for both species (**Carvalho et al., 2014**; **Yang et al., 2014**). For *Ae. aegypti,* >60% of all occurrence records are from Asia and

Oceania, 35% are from the Americas and only 575 unique occurrences are available for Africa and Europe (**Table 1a**). Similarly for *Ae. albopictus*, most of the occurrences are from Asia (75%), 23% are from the Americas and only 542 records are available from Europe and Africa (**Table 1b**). For each continent the top 10 countries in terms of occurrences recorded are shown for both species (**Table 1**). The geographic distribution of the occurrence records is the widest ever recorded with particularly high spatial and temporal resolution in Taiwan and Brazil for both species and in the United States for *Ae. albopictus*. All occurrence data have been made openly available through an online data repository to ensure consistency and reproducibility (**Pigott and Kraemer, 2014**; **Kraemer et al., 2015a**).

Maps showing the predicted global distribution for *Ae. aegypti* and *Ae. albopictus* are presented in **Figures 1**, **2**, respectively. The distributions of the two species differ markedly in a number of places. *Ae. aegypti* is predicted to occur primarily in the tropics and sub-tropics, with concentrations in northern Brazil and southeast Asia including all of India, but with relatively few areas of suitability in Europe (only Spain and Greece) and temperate North America. In Australia, however, *Ae. aegypti* shows a wider geographic distribution than *Ae. albopictus*, which is confined to the east coast, largely reflecting the known historic distribution of *Ae. aegypti*. By contrast, the distribution of *Ae. albopictus* extends into southern Europe (**Figure 3A**), northern China, southern Brazil, northern United States (3b), and Japan. Again, this reflects the current and historic distribution of *Ae. albopictus* and the ability of the species to tolerate lower temperatures (**Tsuda and Takagi, 2001**; **Lounibos et al., 2002**; **Thomas et al., 2012**; **Brady et al., 2014**).

Table 1. The geographic distribution of spatially unique occurrence records for the Americas, Europe/Africa, and Asia/Oceania

	Country	Occurrences		Country	Occurrences		Country	Occurrences
Ae. aegypti								
Americas	Brazil	5,044	Europe/ Africa	Senegal	112	Asia/ Oceania	Taiwan	9,490
	USA	436		Cameroon	55		Indonesia	603
	Mexico	411		Kenya	52		Thailand	495
	Cuba	177		United Republic of Tanzania	44		India	423
	Argentina	170		Côte d'Ivoire	40		Australia	282
	Trinidad and Tobago	152		Nigeria	35		Viet Nam	223
	Venezuela	130		Madagascar	28		Malaysia	112
	Colombia	128		Gabon	27		Singapore	44
	Puerto Rico	120		Mayotte	20		Philippines	36
	Peru	89		Sierra Leone	20		Cambodia	29
Ae. albopictus								
Americas	Brazil	3,441	Europe/ Africa	Italy	203	Asia/ Oceania	Taiwan	15,339
	USA	1,594		Madagascar	58		Malaysia	186
	Mexico	50		Cameroon	42		Indonesia	161
	Cayman Islands	15		France	37		India	150
	Haiti	13		Gabon	27		Japan	97
	Guatemala	12		Albania	22		Thailand	82
	Venezuela	7		Mayotte	21		Singapore	44
	Colombia	3		Greece	18		Lao People's Democratic Republic	26
	Cuba	3		Israel	17		Philippines	22
	Puerto Rico	3		Lebanon	15		Viet Nam	18

Top 10 countries in terms of occurrence records for each continent are shown for *Ae. aegypti* (a) and *Ae. albopictus* (b).

DOI: 10.7554/eLife.08347.003

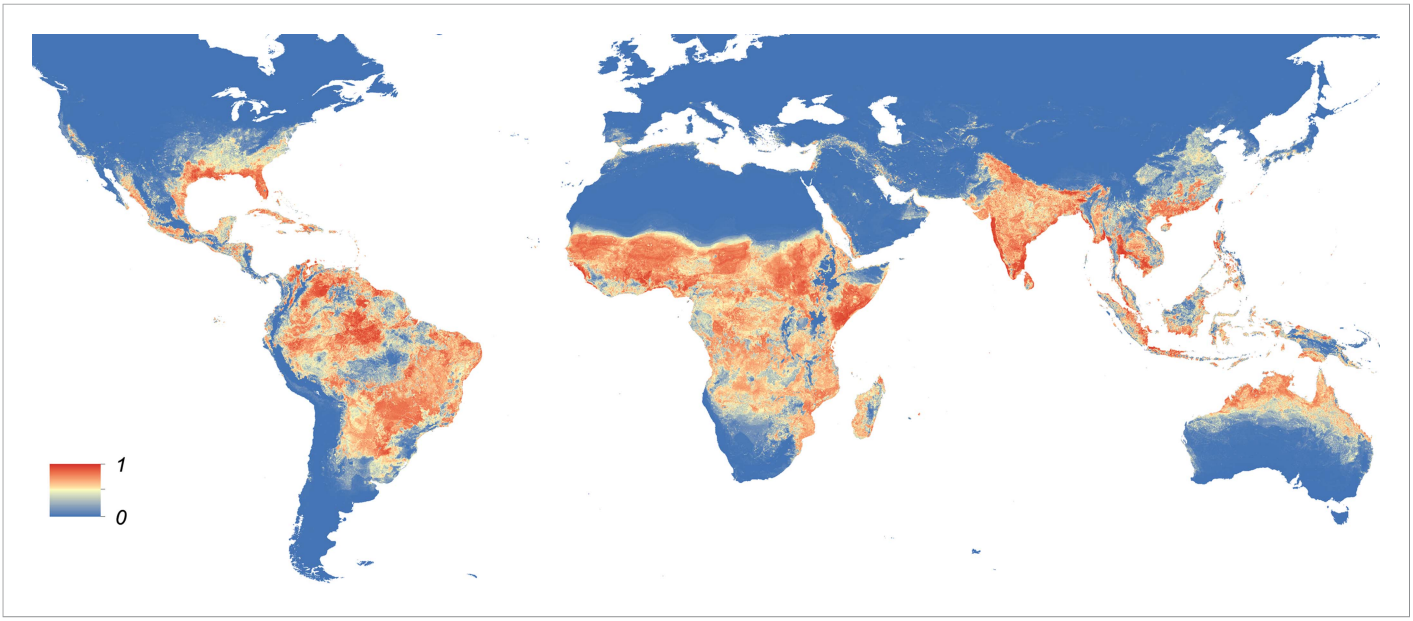


Figure 1. Global map of the predicted distribution of *Ae. aegypti*. The map depicts the probability of occurrence (from 0 blue to 1 red) at a spatial resolution of 5 km × 5 km.

DOI: 10.7554/eLife.08347.004

The following figure supplements are available for figure 1:

Figure supplement 1. Effect plots of covariates used in this study showing the marginal effect of each covariate on probability of presence for *Ae. aegypti* (1) and *Ae. albopictus* (2): enhanced vegetation index (EVI) annual mean (**A**); Enhanced vegetation index—range (**B**); annual monthly maximum precipitation (**C**); annual monthly minimum precipitation (**D**); temperature suitability (**E**); urban areas (**F**); peri-urban areas (**G**).

DOI: 10.7554/eLife.08347.005

Figure supplement 2. Set of covariate layers used to predict the ecological niche of *Ae. aegypti* and *Ae. albopictus* described in detail in the 'Materials and methods' section; (**A**) enhanced vegetation index (EVI) annual mean, (**B**) EVI annual range, (**C**) annual monthly maximum precipitation, (**D**) annual monthly minimum precipitation, (**E**) temperature suitability for *Ae. albopictus*, (**F**) temperature suitability for *Ae. aegypti*, (**G**) rural, peri-urban and urban classification layer.

DOI: 10.7554/eLife.08347.006

Figure supplement 3. Visualization of pixel level uncertainty calculated using the upper and lower bounds of the 95% confidence intervals associated with the prediction maps for *Ae. aegypti* (**A**) and *Ae. albopictus* (**B**).

DOI: 10.7554/eLife.08347.007

Figure supplement 4. The distribution of the occurrence database for *Ae. aegypti* (**A**) and *Ae. albopictus* (**B**) plotted on the underlying prediction surface.

DOI: 10.7554/eLife.08347.008

In Europe, the predicted potential distribution of *Ae. albopictus* contains most of the known occurrence points, but suitability is also predicted in Portugal and the west of Spain, and in much of south-eastern Europe and the Balkans, where the species has yet to be reported. Similarly, in China *Ae. albopictus* has yet to be reported from much of the area predicted to be environmentally suitable. By contrast, in the United States the species has been reported from almost all of the predicted suitable areas, with the exception of a small band of predicted suitability on the western slope of the Sierra Nevada. Due to the relatively sparse reporting from Africa it remains uncertain whether areas predicted to be highly suitable are already infested or have yet to be colonized by the species. *Ae. albopictus* for example has only been reported from some West African countries (Nigeria, Cameroon, Gabon, the Central African Republic, Congo, Côte d'Ivoire) and Madagascar, and South Africa (as well as some islands in the Indian Ocean). The distribution of *Ae. aegypti* in Africa seems to be much wider, with reports of species occurrence in over 30 countries.

For both species, the most important predictor was temperature. Temperature suitability indices had high relative influence statistics for both species; this variable was selected in approximately half of regression tree decisions for *Ae. aegypti* (54.9%, CI = 53.7–56%) and *Ae. albopictus* (44.3%, CI = 42.7–45.6%). The full definition of a relative influence statistic is given in the 'Materials and methods' section under the heading *Predictive performance and relative influence of covariates*.

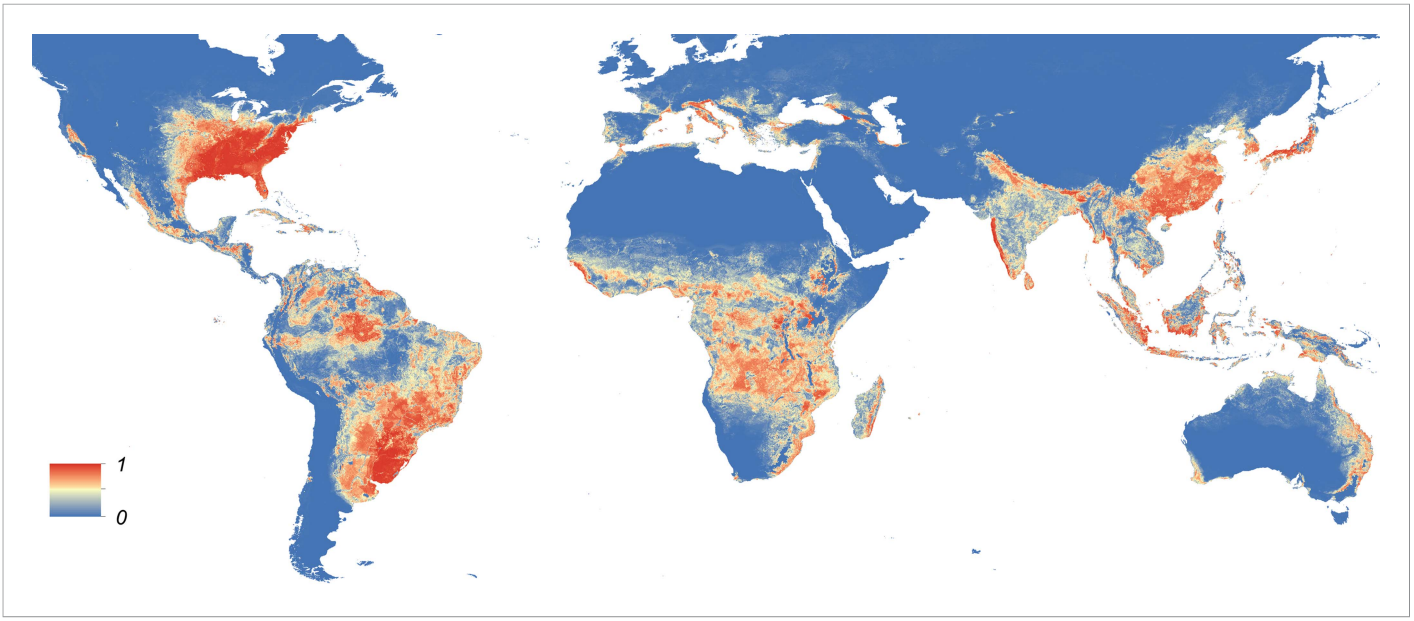


Figure 2. Global map of the predicted distribution of *Ae. albopictus*. The map depicts the probability of occurrence (from 0 blue to 1 red) at a spatial resolution of 5 km × 5 km.
DOI: 10.7554/eLife.08347.009

Precipitation and vegetation indices made up the remainder of predictors. Urban land cover made very little contribution to either model (***Table 2***). Model evaluation statistics under cross-validation were high (AUC: 0.87 and 0.9 respectively) for both model ensembles, indicating high predictive performance of the model. Effect plots for each covariate are shown in ***Figure 1—figure supplement 2***. Maps of uncertainty associated with these predictions are presented in ***Figure 1—figure supplement 3***.

Discussion

By combining the most comprehensive dataset of occurrence records with an advanced modelling approach and a bespoke set of environmental and land-cover correlates, we have produced contemporary high-resolution probability of occurrence maps for *Ae. aegypti* and *Ae. albopictus*, two of the most important disease vectors globally. Dengue and chikungunya, pathogens transmitted by these vectors and rapidly expanding in their distributions, are increasingly prominent in public health agendas and pose significant health threats to humans (***Staples et al., 2009***; ***Gardner et al., 2012***; ***Bhatt et al., 2013***; ***Weaver and Lecuit, 2015***). In common with previous work to map the global distributions of the dominant vectors of malaria (***Sinka et al., 2010a***, ***2010b***, ***2011***), the maps will improve efforts to understand the spatial epidemiology of associated arboviruses, and to predict how these could change in the future. Specifically, these maps may be used to prioritize surveillance for these vector species and the diseases caused by the viruses they transmit in areas where disease and entomological reporting remains poor. For example, in parts of Asia and Africa where there is a mismatch between predicted environmental suitability and reported occurrences, these maps could be used to determine whether the vector has yet to fill its niche or if it is present but has not been reported due to limited entomological surveillance. They may also be used to identify areas where the species could persist but has yet to be reported, in order to proactively prevent vector establishment.

The relative contributions of each of the environmental covariates to the global models concur with our theoretical and experimental understanding of each species' biology. Both species' distributions are highly dependent on the limiting factor temperature places on survival of the adult mosquitoes and on the gonotrophic cycle (***Brady et al., 2013***) (***Table 2***). The inclusion of a bespoke temperature suitability index (***Brady et al., 2014***), both in defining the pseudo-absences and as a covariate, allowed us to capture both geographic and temporal variations in the species-specific effects of temperature in a single variable, leading to improved predictive skill of the models. As both

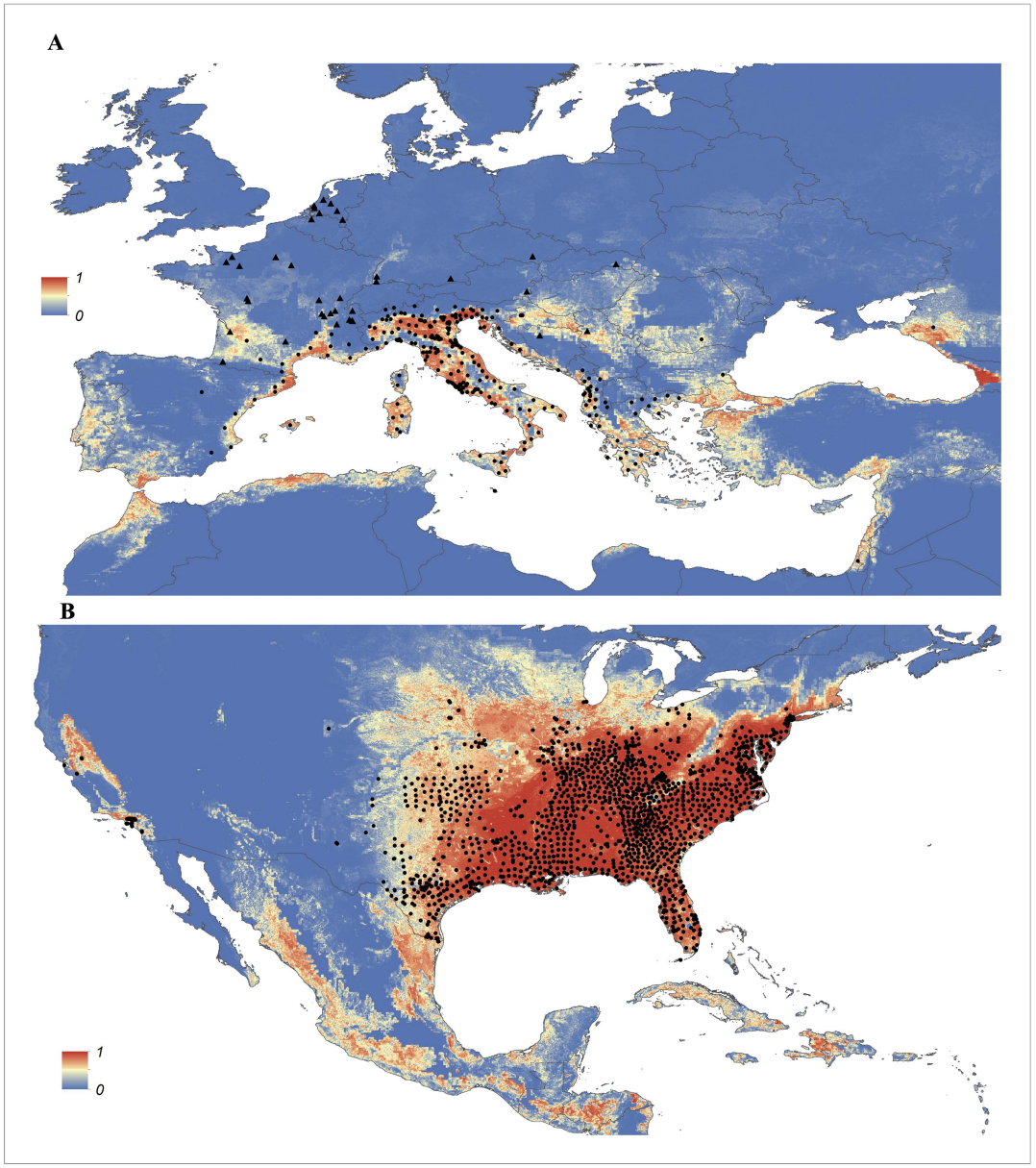


Figure 3. Predicted probability of occurrence of *Ae. albopictus* in Europe (**A**) and the United States (**B**), regions in which *Ae. albopictus* is rapidly expanding its range. Points represent known occurrences (transient [triangles] or established [circles]) until the end of 2013.

DOI: 10.7554/eLife.08347.010

Ae. aegypti and *Ae. albopictus* lay their eggs in small water-filled containers (***Morrison et al., 2004***), it is encouraging that precipitation also has a strong influence on the model's predictions. The stronger influence of minimum precipitation for *Ae. albopictus* than for *Ae. aegypti* (16.1% vs 9.1%, ***Table 2***) may reflect the former species' preference for non-domestic juvenile habitats, which are solely reliant on filling via precipitation. By contrast, *Ae. aegypti* primarily inhabits domestic water-holding containers (***Scott et al., 2000***) that are maintained in low-precipitation environments by water storage activities. The greater importance of enhanced vegetation index (EVI) for *Ae. albopictus* than for *Ae. aegypti* (15.3% vs 12.1%, ***Table 2***) also supports the hypothesis that *Ae. albopictus* tends to prefer non-domestic juvenile sites (***Morrison et al., 2004***). This does not, however, rule out the possibility that the two species can overlap. Additional finer scale studies need to be conducted to investigate if competitive exclusion for hosts and/or habitat occurs between *Ae. aegypti* and *Ae. albopictus*. The effect of urbanicity was surprisingly low for both species (2% and 1.1% for *Ae. albopictus* and

Table 2. Relative contribution of environmental covariates predicting the global distribution of *Ae. aegypti* and *Ae. albopictus*

	Mean contribution *Ae. aegypti* (%)	95% confidence interval *Ae. aegypti* (%)	Mean contribution *Ae. albopictus* (%)	95% confidence interval *Ae. albopictus* (%)
Temperature suitability	54.9	53.7–56	44.3	42.7–45.6
Maximum precipitation	13.6	12.6–14.6	13.9	12.7–14.9
Enhanced vegetation index (mean)	12.1	11.3–12.9	15.3	14.5–16.3
Minimum precipitation	9.1	8.5–10	16.1	15.2–16.9
Enhanced vegetation index (range)	8.3	7.7–9	9.1	8.3–10.1
Urbanicity	2	1.3–2.4	1.1	0.7–1.7

DOI: 10.7554/eLife.08347.011

Ae. aegypti, respectively). As both species have been shown to inhabit a wide variety of urban and peri-urban settings with various degrees of intensity (***Powell and Tabachnick, 2013***; ***Li et al., 2014***), it is likely that the simple urban/rural distinction of our urbanicity covariate did not sufficiently capture this variation and instead continuous covariates such as EVI allow to better distinguish the respective habitat types and were thus chosen more frequently by the model. Incorporating a larger set of covariates allowed us to investigate not only the effect of temperature on survival but for additional variance as shown in the relative influence plots (***Figure 1—figure supplement 1***). Future *Aedes* species distribution models could be improved by including a comprehensive global covariate that distinguishes human settlements using complex satellite imagery processing tools (***Schneider, 2012***).

Our maps are based on covariates where each 5 km × 5 km pixel represents yearly mean average values. We therefore produce maps that represent the long-term average distribution of both species. However, this does not allow us to directly infer seasonal patterns of distributions which might be of importance on the periphery of the species distributions. With a more temporally resolved dataset it may be possible to capture the effects of intra-annual seasonality on the species' distributions. Adding mechanistic determinants, such as survival, have previously been used to combine seasonal patterns with global distribution maps (***Johansson et al., 2014***). To make best use of the comprehensive set of data collected, we construct models and maps at a global scale, allowing the model to share information across the whole spectrum of environmental regions. However, given the scale at which this study was performed, there is always the possibility that variation in microclimate or local adaptive strategies of both species may have a significant impact in some locations.

Previous studies have discussed the risk of pathogen importation and autochthonous transmission of DENV and CHIKV in Europe and the Americas without comprehensively accounting for the distribution of the vectors (***Bogoch et al., 2014***; ***Schaffner and Mathis, 2014***). These freely available vector distributions maps (http://goo.gl/Zl2P7J) can now be used as covariates to refine these studies and to generate high-resolution maps of the risk of possible local DENV and CHIKV transmission in currently non-endemic settings. Such maps would be useful for prioritizing surveillance in areas where there is a risk of disease importation. This will be especially important in areas where sporadic cases of related viruses have been reported, such as Europe, the United States, Argentina, and China (***Rezza et al., 2007***; ***Otero and Solari, 2010***; ***Wu et al., 2010***; ***Johansson, 2015***).

Both *Ae. aegypti* and *Ae. albopictus* have a history of global expansion associated with trade and travel (***Tatem et al., 2006***; ***Brown et al., 2014***; ***Gloria-Soria et al., 2014***). Introductions of the species over long distances and between continents has been associated with international trade routes via shipping and overland spread driven by human movement and transport routes, both facilitated by the endophilic behavior of the two species (***Nawrocki and Hawley, 1987***; ***Tatem et al., 2006***; ***Hofhuis et al., 2009***). The global spread of the associated pathogens has undoubtedly been a consequence of increasing global connectedness. As these processes continue and the world becomes increasingly connected and urbanized, risk of importation and subsequent autochthonous transmission of DENV and CHIKV will continue to increase (***Allwinn et al., 2008***; ***Tomasello and Schlagenhauf, 2013***; ***Khan et al., 2014***; ***Messina et al., 2015***). The true distribution of both species is influenced by a variety of factors, not just the ones presented here. Nevertheless, this study represents an important baseline for further refinements. For instance, our maps can be used to indicate areas

where the species are likely to become established if introduced. Accurately predicting the future distributions of these species will also require model-based estimates of the rate at which these species colonize new areas. Such predictions can be informed by human and trade mobility patterns between endemic and non-endemic regions as well as data on the past spread of the vectors. Improving our ability to predict rates of vector importation will therefore be crucial to inferring future risk (**Seebens et al., 2013**).

Previous studies have provided crucial information on genetic variation both within and between populations of these two vector species (**Brown et al., 2011**). As the volume of georeferenced information on the population genetics of *Ae. aegypti* and *Ae. albopictus* increases, the potential to incorporate this information into mapping analyses to understand the current and future distribution of disease risk also increases. Phylogeographic analyses offer a unique way to infer the recent patterns of vector spread and to identify the major routes of importation (**Allicock et al., 2012**). This information is crucial to inform models that predict the risk of vector introductions.

Phylogenetic information could also be used to inform future iterations of the species distribution models used here by enabling the model to characterize and map environmental suitability for different vector subspecies. This could be particularly useful in the case of *Ae. albopictus* where genetic variation is known to underlie the ability to undergo diapause and therefore to overwinter in colder locations (**Takumi et al., 2009**). Mapping the distributions of distinct genetic subgroups could also improve our understanding of the complex interactions between mosquito vector populations and virus strains and how this relates to spatial variation in transmission intensity (**Tsetsarkin et al., 2007**; **Vazeille et al., 2007**; **Tsetsarkin and Weaver, 2011**; **Zouache et al., 2014**).

The maps presented comprise a contemporary estimate of the current and potential future distribution of *Ae. aegypti* and *Ae. albopictus*. As more occurrence data become available, these maps can be refined to incorporate recent importation and establishment events and corresponding improvements in predictions. By disseminating both the occurrence data and the predictive maps on an open-access basis we hope to facilitate both the future development of these maps and their uptake by the global public health community.

Materials and methods

A BRT modelling approach was applied to derive probabilistic global environmental risk maps for *Ae. aegypti* and *Ae. albopictus*. BRT models are machine-learning model ensembles commonly used in species distribution modelling (SDM) and show strong predictive performance due to their ability to handle complex non-linear relationships between probability of species occurrence and multiple environmental correlates (**Elith et al., 2006**, **2008**). Our model required the following sets of input data in order to make accurate predictions of the distribution of these two species: (i) a temperature suitability mask defining the fundamental limits of both species; (ii) a globally comprehensive dataset of geo-positioned occurrence points for both species; (iii) appropriate land-cover and environmental covariate datasets that help explain the current distribution of the species; and (iv) a set of species absence records that further refine the species range and reduce sampling bias. Details regarding the specific attributes of the model and data generation are outlined below and maps of each of the covariates are shown in ***Figure 1—figure supplement 2***.

Temperature suitability mask

While the niche of a species is determined by a host of environmental, ecological and socio-economic factors of unknown influence and interaction strength, it is possible to exclude parts of the niche if the direct effects of one factor on a step rate-limiting to population persistence are well known. One such example for mosquito population persistence is whether temperature permits adult females to survive long enough to complete their first gonotrophic cycle and thus oviposit. Both adult female longevity and length of first gonotrophic cycle are temperature dependent. Combining these two relationships with a dynamic population-level simulation, **Brady et al. (2013**, **2014)** evaluated the thermal limits to persistence of *Ae. aegypti* and *Ae. albopictus* populations on a global scale. The binary outputs of this model are used as a mask to sample pseudo-absence points in locations known to be unsuitable–thereby informing the statistical model using mechanistic model outputs. The temperature suitability index developed by Brady et al. is also used in a continuous variable form (i.e., the relative number of ovipositions of parous females permitted by temperature) as a covariate in the BRT model.

Occurrence records

The database used for this study contains information on the known global occurrences of the adults, pupae, larvae or eggs of *Ae. aegypti* and *Ae. albopictus* globally from 1960–2014. We included data from a variety of sources, including (1) published literature and (2) primary and unpublished occurrence data from national and international entomological surveys. To our knowledge this is the largest, most comprehensive global dataset for both *Ae. aegypti* and *Ae. albopictus.* Confirmed *Aedes* occurrences were entered in the database after a comprehensive literature search using methods described elsewhere (**Kraemer et al., 2015a**; **Kraemer et al., 2015b**; http://dx.doi.org/10.5061/dryad.47v3c). In short, this included extracting all available location (latitude and longitude) information from the relevant articles, primarily using Google Maps (http://www.google.com/maps) so that it matched the spatial resolution of our covariate datasets of approximately 5 km × 5 km. Primary and unpublished data sources were obtained from Brazil, Europe, Indonesia, Taiwan, and the United States. After consolidating all data into two large databases for each species, independently they underwent spatial and temporal standardization. An occurrence record was defined as a single occurrence at a given unique location within one calendar year. This was important to avoid over-representation in regions where multiple surveys per year were performed, such as Taiwan or Brazil. To ensure the accuracy of the data we overlaid the geolocated occurrence points with a raster that distinguished land from water. Any records that were positioned outside the land area were subsequently removed. In total we assembled 19,930 and 22,137 occurrence records for *Ae. aegypti* and *Ae. albopictus* respectively. The distribution of occurrence points are plotted in ***Figure 1—figure supplement 4***.

Land-cover and environmental variables

The distribution of both species considered in this study are known to be influenced by environmental factors such as temperature and demographic factors such as urbanisation (**Lounibos, 2002**; **Brown et al., 2014**). Global gridded maps of such variables are becoming ever more available and have been commonly applied in SDM and disease mapping (**Hijmans et al., 2005**; **Hay et al., 2006**; **Gething et al., 2011**; **Bhatt et al., 2013**; **Pigott et al., 2014a**, **2014b**). The rationale for the inclusion of each variable we used is described below.

EVI

Survival of *Ae. aegypti* and *Ae. albopictus* is highly dependent on temperature and water availability (**Luz et al., 2008**). EVI measures vegetation canopy greenness and can be used as a proxy for soil surface-level moisture that are associated with the availability of mosquito larval development sites (**Estallo et al., 2008**; **Nihei et al., 2014**). Eggs and adults require moisture to survive, with low dry season moisture levels affecting adult mortality (**Sota and Mogi, 1992**; **Russell et al., 2001**). Vegetation canopy cover reduces evaporation and wind speed in the sub-canopy, which protects mosquito development sites (**Linthicum, 1999**; **Fuller et al., 2009**; **Hahn et al., 2014**). We used range and mean values of MODIS EVI after processing through a gap-filling algorithm described elsewhere (**Weiss et al., 2014**).

Precipitation

The principal larval habitats of both species are man-made containers that are used for water storage or accumulate rain (**Morrison et al., 2004**). Some local studies have shown that there is a relationship between precipitation and vector abundance (**Scott et al., 2000**; **Romero-Vivas and Falconar, 2005**). To account for the availability of water-filled containers a maximum and minimum annual precipitation layer was extracted from the WorldClim database and projected for the year 2015 (http://www.worldclim.org).

Urbanisation

Ae. aegypti adults are highly domesticated mosquitoes feeding almost exclusively on humans (**Bargielowski et al., 2013**), larvae develop preferentially in artificial containers in close association with human habitation, often in urban settings (**Lounibos, 2002**; **Honório et al., 2003**; **Brown et al., 2011**, **2014**; **Powell and Tabachnick, 2013**). *Ae. albopictus* are more commonly found in rural and peri-urban settings, feeding readily on a variety of mammalian and avian species, although *Ae. albopictus* shows similar larval development behavior in artificial containers (**Reiter, 2001**; **Gratz, 2004**; **Juliano and Philip Lounibos, 2005**; **Li et al., 2014**). To account for differences in urban, peri-urban and rural environments we built a categorical variable by supplementing the projected 2010 Global Rural Urban Mapping Project (GRUMP) urban and rural categories with land-cover classes using night-time light satellite imagery and population density, using the most up-to-date national censuses available to the

smallest available administrative unit available (**Balk et al., 2006**). A gridded surface of 5 km × 5 km cells was generated with each pixel representing either urban, peri-urban, or rural areas.

Modelling approach

BRT models consistently outperform other species distribution models such as maximum entropy (Maxent), GARP, and BIOCLIM in their predictive performance (**Elith et al., 2006**; **Leathwick et al., 2006**). BRT combines the strengths of regression trees (i.e., the omission of irrelevant variables and the ability to model complex interactions) with machine learning techniques (i.e., the building of an ensemble of models that approximate the true response surface [**Elith and Leathwick, 2009**]). To prevent overfitting, the model used a penalized forward stepwise search and cross-validation method to identify the optimal number of decision trees (**Elith et al., 2008**). Modelling was performed using the gbm, dismo, raster and seegSDM R packages using the R v 3.1.1 environment (**Ridgeway, 2013**; **Golding, 2014**; **Hijmans, 2014**; **R Core Team, 2014**).

Removing sample selection bias

Pseudo-absence (also referred to as background) records provide a sample of the set of conditions available to the species in the region rather than actual absences (**Phillips et al., 2009**). These records are needed because true absences are generally unavailable in large composite datasets such as the one used in this study. To account for reporting bias in presence data, a common problem with presence-only SDM, which if not accounted for can lead to biases in the resulting predictions, we follow **Phillips et al. (2009)** in sampling pseudo-absence points according to the same reporting bias likely to be present in occurrence records (namely spatial variation in reporting of mosquito occurrence). Firstly, we selected 10,000 occurrence records of *Aedes* species from the Global Biodiversity Information Facility (http://www.gbif.org), omitting all records of *Ae. aegypti* and *Ae. albopictus*. This dataset is intended to reflect biases in mosquito reporting in areas which are suitable for *Aedes* mosquitoes. Secondly, to reflect areas where habitats are biologically not suitable for *Aedes* occurrence we sampled an additional 10,000 pseudo-absence points at random locations, with sampling probability greatest in areas that the biologically-based temperature suitability index predicted to be unsuitable. Thirdly, sampling of occurrence points was also biased towards oversampled regions such as Brazil and Taiwan in which there were a large number of reported occurrence records due to the inclusion of results of large national entomological surveys (**Table 1**). Therefore, we weighted occurrence records from these locations so that the density of occurrence records per country matched the density of all other records globally by dividing the number of occurrence points by the size of the respective countries.

Modelling

An ensemble BRT was constructed using 120 sub-models to derive uncertainty distributions of the prediction map. Each of the 120 sub-models was fitted to a separate bootstrap resampling of the dataset and used to generate a probability map for each individual species on a 5 km × 5 km resolution. The mean of these 120 sub-models was used as the final *Aedes* risk maps. Pixel based uncertainty was estimated by calculating the 95% confidence interval from the 120 sub-models.

Predictive performance and relative influence of covariates

The variables used as land-cover and environmental correlates used in this study are quantified based on their relative influence (0–100) on explaining the variance in the models calculated as the sum of the number of times a particular variable is selected for splitting the decision tree, weighted by the squared improvement to the overall model averaged over all trees (**Friedman, 2001**; **Friedman and Meulman, 2003**). Note that in a BRT, non-informative predictors are largely ignored (**Elith et al., 2008**). Predictive performance of each sub-model was evaluated using the area under curve (AUC) statistic calculated as the mean AUC for each of the ten cross-validation folds evaluated against the other 90% of the data under the pairwise distance sampling procedure of **Hijmans (2012)**. The overall predictive accuracy of the model was measured as the mean and standard deviation of these AUCs across all 120 sub-models (**Merckx et al., 2010**; **Hijmans, 2012**).

Acknowledgements

MUGK is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. MES is funded by a project grant from the Bill & Melinda Gates Foundation as part of the VecNet consortium (http://vecnet.org). FMS is funded by the Rhodes Trust. CMB acknowledges funding from

the U.S. National Aeronautics and Space Administration (#NNX15AF36G). CGM was funded in part by contract N01-A1-25489 from the NIH/National Institute of Allergy and Infectious Diseases. IRFE is funded by the Wellcome Trust (#099872). WVB, GH, and FS acknowledge funding from VBORNET, an ECDC funded project (contract number ECDC/09/018). OJB is funded by a BBSRC studentship. JPM is funded by, and SIH, GRWW, TWS and OJB acknowledge the support of the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, European Commission seventh Framework Programme (#21803) http://www.idams.eu). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. DMP is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford. TWS acknowledges funding from the Bill & Melinda Gates Foundation (#OPP52250), the Innovative Vector Control Consortium, and the NIH (R01-AI069341, R01-AI091980, R01-GM08322, and 1P01AI098670). TWS, CMB, DLS and SIH also acknowledge funding support from the Research and Policy in Infectious Diseases Dynamics (RAPIDD) program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. NG is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), which also supports KAD and AQNM.

We want to thank Ralph E Harbach for his comments on the final manuscript. We thank Dr Roseli La Corte dos Santos, Departamento de Morfologia, Universidade Federal de Sergipe for providing useful guidance, the Ministry of Health of Brazil municipality secretaries and Centers for Disease Control of Taiwan for providing mosquito occurrence data and Bimandra Djaafara and Karin Dian Lestari of EOCRU for their work on geo-referencing the Indonesia *Aedes* datasets. We also want to thank the VBORNET & TigerMaps network funded by ECDC, Stockholm, and all their contributors (a detailed list is given in ***Supplementary file 1***) for releasing their vector distribution data. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Additional information

Competing interests

SIH: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference	Author
Studienstiftung des Deutschen Volkes		Moritz UG Kraemer
Bill and Melinda Gates Foundation	#OPP1053338	Nick Golding
Wellcome Trust	#095066	Kirsten A Duda, Adrian QN Mylne, Simon I Hay
European Centre for Disease Prevention and Control	ECDC/09/018	Wim Van Bortel, Guy Hendrickx, Francis Schaffner
European Commission Directorate-General for Research and Innovation	#21803	Oliver J Brady, Jane P Messina
Biotechnology and Biological Sciences Research Council (BBSRC)		Oliver J Brady
National Aeronautics and Space Administration (NASA)	#NNX15AF36G	Christopher M Barker
National Institutes of Health (NIH)	RAPIDD program	Christopher M Barker, Thomas W Scott, David L Smith, Simon I Hay
National Institutes of Health (NIH)	R01-AI069341	Thomas W Scott
National Institutes of Health (NIH)	R01-AI091980	Thomas W Scott

Funder	Grant reference	Author
National Institutes of Health (NIH)	R01-GM08322	Thomas W Scott
National Institutes of Health (NIH)	N01-A1-25489	Chester G Moore
Bill and Melinda Gates Foundation	#OPP52250	Thomas W Scott
Sir Richard Southwood Graduate Scholarship		David M Pigott
Wellcome Trust	Vecnet	Marianne E Sinka
Wellcome Trust	#099872	Iqbal RF Elyazar
The Rhodes Trust		Freya M Shearer

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MUGK, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; MES, FMS, Acquisition of data, Drafting or revising the article; KAD, AQNM, Acquisition of data; CMB, CGM, RGC, GEC, WVB, FS, IRFE, H-JT, Drafting or revising the article, Contributed unpublished essential data or reagents; GH, Contributed unpublished essential data or reagents; OJB, NG, SIH, Conception and design, Analysis and interpretation of data, Drafting or revising the article; JPM, DMP, DLS, Analysis and interpretation of data, Drafting or revising the article; TWS, Conception and design, Drafting or revising the article; GRWW, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents

Author ORCIDs

Christopher M Barker, http://orcid.org/0000-0002-7941-346X
Francis Schaffner, http://orcid.org/0000-0001-9166-7617
David M Pigott, http://orcid.org/0000-0002-6731-4034
David L Smith, http://orcid.org/0000-0003-4367-3849
Nick Golding, http://orcid.org/0000-0001-8916-5570
Simon I Hay, http://orcid.org/0000-0002-0611-7272

Additional files

Supplementary file

• Supplementary file 1. List of contributors and their affiliation from TigerMaps & VBORNET for *Ae. albopictus* presence records in Europe.
DOI: 10.7554/eLife.08347.012

Major dataset

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI	2015	Data from: The global compendium of Aedes aegypti and Ae. albopictus occurrence	http://dx.doi.org/10.5061/dryad.47v3c	Available at Dryad Digital Repository under a CC0 Public Domain Dedication.

References

Allicock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, Mueller BA, Suchard MA, Foster JE, Rambaut A, Carrington CV. 2012. Phylogeography and population dynamics of dengue viruses in the Americas. *Molecular Biology and Evolution* **29**:1533–1543. doi: 10.1093/molbev/msr320.

Allwinn R, Hofknecht N, Doerr HW. 2008. Dengue in travellers is still underestimated. *Intervirology* **51**:96–100. doi: 10.1159/000131667.

Aström C, Rocklöv J, Hales S, Béguin A, Louis V, Sauerborn R. 2012. Potential distribution of dengue fever under scenarios of climate change and economic development. *Ecohealth* **9**:448–454. doi: 10.1007/s10393-012-0808-0.

Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. 2006. Determining global population distribution: methods, applications and data. *Advances in Parasitology* **62**:119–156. doi: 10.1016/S0065-308X(05)62004-0.

Bargielowski IE, Lounibos LP, Carrasquilla MC. 2013. Evolution of resistance to satyrization through reproductive character displacement in populations of invasive dengue vectors. *Proceedings of the National Academy of Sciences of USA* **110**:2888–2892. doi: 10.1073/pnas.1219599110.

Benedict MQ, Levine RS, Hawley WA, Lounibos LP. 2007. Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. *Vector Borne and Zoonotic Diseases* **7**:76–85. doi: 10.1089/vbz.2006.0562.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. 2013. The global distribution and burden of dengue. *Nature* **496**:504–507. doi: 10.1038/nature12060.

Bogoch II, Creatore MI, Cetron MS, Brownstein JS, Pesik N, Miniota J, Tam T, Hu W, Nicolucci A, Ahmed S, Yoon JW, Berry I, Hay SI, Anema A, Tatem AJ, MacFadden D, German M, Khan K. 2014. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* **6736**:1–7. doi: 10.1016/S0140-6736(14)61828-6.

Borgherini G, Poubeau P, Staikowsky F, Lory M, Le Moullec N, Becquart JP, Wengling C, Michault A, Paganin F. 2007. Outbreak of chikungunya on Reunion Island: early clinical and laboratory features in 157 adult patients. *Clinical Infectious Diseases* **44**:1401–1407. doi: 10.1086/517537.

Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, Moyes CL, Farlow AW, Scott TW, Hay SI. 2012. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLOS Neglected Tropical Diseases* **6**:e1760. doi: 10.1371/journal.pntd.0001760.

Brady OJ, Golding N, Pigott DM, Kraemer MU, Messina JP, Reiner RC Jr, Scott TW, Smith DL, Gething PW, Hay SI. 2014. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasites & Vectors* **7**:338. doi: 10.1186/1756-3305-7-338.

Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, Delatte H, Grech MG, Leisnham PT, Maciel-de-Freitas R, Styer LM, Smith DL, Scott TW, Gething PW, Hay SI. 2013. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites & Vectors* **6**:351. doi: 10.1186/1756-3305-6-351.

Brown JE, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, Zhao H, Caccone A, Powell JR. 2014. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution* **68**:514–525. doi: 10.1111/evo.12281.

Brown JE, McBride CS, Johnson P, Ritchie S, Paupy C, Bossin H, Lutomiah J, Fernandez-Salas I, Ponlawat A, Cornel AJ, Black WC IV, Gorrochotegui-Escalante N, Urdaneta-Marquez L, Sylla M, Slotman M, Murray KO, Walker C, Powell JR. 2011. Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. *Proceedings. Biological Sciences/The Royal Society* **278**:2446–2454. doi: 10.1098/rspb.2010.2469.

Caminade C, Medlock JM, Ducheyne E, McIntyre KM, Leach S, Baylis M, Morse AP. 2012. Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. *Journal of the Royal Society, Interface/The Royal Society* **9**:2708–2717. doi: 10.1098/rsif.2012.0138.

Campbell LP, Luther C, Moo-Llanes D, Ramsey JM, Danis-Lozano R, Peterson AT. 2015. Climate change influences on global vector distributions for dengue and chikungunya viruses. *Philosophical Transactions of the Royal Society B* **370**. doi: 10.1098/rstb.2014.0135.

Carrington LB, Simmons CP. 2014. Human to mosquito transmission of dengue viruses. *Frontiers in Immunology* **5**:1–8. doi: 10.3389/fimmu.2014.00290.

Carvalho RG, Lourenço-de-Oliveira R, Braga IA. 2014. Updating the geographical distribution and frequency of *Aedes albopictus* in Brazil with remarks regarding its range in the Americas. *Memórias do Instituto Oswaldo Cruz* **109**:787–796. doi: 10.1590/0074-0276140304.

Cauchemez S, Ledrans M, Poletto C, Quenel P, de Valk H, Colizza V, Boëlle PY. 2014. Local and regional spread of chikungunya fever in the Americas. *Euro Surveillance* **19**:20854. doi: 10.2807/1560-7917.ES2014.19.28.20854.

Charrel R, Leparc-Goffart I, Gallian P, de Lamballerie X. 2014. Globalization of chikungunya: 10 years to invade the world. *Clinical Microbiology and Infection* **20**:662–663. doi: 10.1111/1469-0691.12694.

Daugherty MP, Alto BW, Juliano SA. 2000. Invertebrate carcasses as a resource for competing *Aedes albopictus* and *Aedes aegypti* (Diptera: *Culicidae*). *Journal of Medical Entomology* **37**:364–372. doi: 10.1093/jmedent/37.3.364.

Delatte AH, Gimonneau G, Triboire A, Fontenille D, Delatte H. 2009. Influence of temperature on immature development, survival, longevity, fecundity, and gonotrophic cycles of *Aedes albopictus*, vector of chikungunya and dengue in the Indian Ocean. *Journal of Medical Entomology* **46**:33–41. doi: 10.1603/033.046.0105.

Delatte H, Desvars A, Bouétard A, Bord S, Gimonneau G, Vourc'h G, Fontenille D. 2010. Blood-feeding behavior of *Aedes albopictus*, a vector of chikungunya on La Reunion. *Vector-Borne and Zoonotic Diseases* **10**:249–258. doi: 10.1089/vbz.2009.0026.

ECDC. 2014. *Dengue outbreak in Madeira, Portugal, March 2013.* Stockholm: European Centre for Disease Prevention and Control.

Eisen L, Moore CG. 2013. *Aedes (Stegomyia) aegypti* in the continental United States: a vector at the cool margin of its geographic range. *Journal of Medical Entomology* **50**:467–478. doi: 10.1603/ME12245.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Ric K. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecogeography* **29**:129–151. doi: 10.1111/j.2006.0906-7590.04596.x.

Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**:677–697. doi: 10.1146/annurev.ecolsys.110308.120159.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *The Journal of Animal Ecology* **77**:802–813. doi: 10.1111/j.1365-2656.2008.01390.x.

Estallo EL, Lamfri MA, Scavuzzo CM, Almeida FL, Introini MV, Zaidenberg M, Almirón WR. 2008. Models for predicting *Aedes aegypti* larval indices based on satellite images and climatic variables. *Journal of American Mosquito Control Association* **24**:368–376. doi: 10.2987/5705.1.

Fischer D, Thomas SM, Niemitz F, Reineking B, Beierkuhnlein C. 2011. Projection of climatic suitability for *Aedes albopictus Skuse (Culicidae)* in Europe under climate change conditions. *Global and Planetary Change* **78**:54–64. doi: 10.1016/j.gloplacha.2011.05.008.

Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* **29**:1189–1232. doi: 10.1214/aos/1013203451.

Friedman JH, Meulman JJ. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* **22**:1365–1381. doi: 10.1002/sim.1501.

Fuller DO, Troyo A, Beier JC. 2009. El Niño Southern Oscillation and vegetation dynamics as predictors of dengue fecer cases in Costa Rica. *Environmental Research Letters* **4**:140111–140118. doi: 10.1088/1748-9326/4/1/014011.

Gardner LM, Fajardo D, Waller ST, Wang O, Sarkar S. 2012. A predictive spatial model to quantify the risk of air-travel-associated dengue importation into the United States and Europe. *Journal of Tropical Medicine* **2012**:103679. doi: 10.1155/2012/103679.

Garske T, Van Kerkhove MD, Yactayo S, Ronveaux O, Lewis RF, Staples JE, Perea W, Ferguson NM, Yellow Fever Expert Committee. 2014. Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLOS Medicine* **11**:e1001638. doi: 10.1371/journal.pmed.1001638.

Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI. 2011. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites & Vectors* **4**:92. doi: 10.1186/1756-3305-4-92.

Gloria-Soria A, Brown JE, Kramer V, Hardstone Yoshimizu M, Powell JR. 2014. Origin of the dengue fever mosquito, *Aedes aegypti*, in California. *PLOS Neglected Tropical Diseases* **8**:e3029. doi: 10.1371/journal.pntd.0003029.

Golding N. 2014. Streamlined functions for species distribution modelling in the seeg research group. R Packag. Version 0.1-3.

Grandadam M, Caro V, Plumet S, Thiberge JM, Souarès Y, Failloux AB, Tolou HJ, Budelot M, Cosserat D, Leparc-Goffart I, Desprès P. 2011. Chikungunya virus, southeastern France. *Emerging Infectious Diseases* **17**:910–913. doi: 10.3201/eid1705.101873.

Gratz NG. 2004. Critical review of the vector status of *Aedes albopictus*. *Medical and Veterinary Entomology* **18**:215–227. doi: 10.1111/j.0269-283X.2004.00513.x.

Hahn MB, Gangnon RE, Barcellos C, Asner GP, Patz JA. 2014. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLOS ONE* **9**:e85725. doi: 10.1371/journal.pone.0085725.

Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. 2006. Global environmental data for mapping infectious disease distribution. *Advances in Parasitology* **62**:37–77. doi: 10.1016/S0065-308X(05)62002-7.

Hijmans RJ. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93**:679–688. doi: 10.1890/11-0826.1.

Hijmans RJ. 2014. Geographic data analysis and modeling. R Packag. Version 2.3-12.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965–1978. doi: 10.1002/joc.1276.

Hofhuis A, Reimerink J, Reusken C, Scholte EJ, Boer AD, Takken W, Koopmans M. 2009. The hidden passenger of lucky bamboo: do imported *Aedes albopictus* mosquitoes cause dengue virus transmission in the Netherlands? *Vector Borne and Zoonotic Diseases* **9**:217–220. doi: 10.1089/vbz.2008.0071.

Honório NA, Silva WDC, Leite PJ, Gonçalves JM, Lounibos LP, Lourenço-de-Oliveira R. 2003. Dispersal of *Aedes aegypti* and *Aedes albopictus (Diptera: Culicidae)* in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Memórias do Instituto Oswaldo Cruz* **98**:191–198. doi: 10.1590/S0074-02762003000200005.

Jentes ES, Poumerol G, Gershman MD, Hill DR, Lemarchand J, Lewis RF, Staples JE, Tomori O, Wilder-Smith A, Monath TP, Informal WHO Working Group on Geographic Risk for Yellow Fever. 2011. The revised global yellow fever risk map and recommendations for vaccination, 2010: consensus of the informal WHO Working Group on Geographic Risk for Yellow Fever. *The Lancet. Infectious Diseases* **11**:622–632. doi: 10.1016/S1473-3099(11)70147-5.

Johansson MA. 2015. Chikungunya on the move. *Trends in Parasitology* **31**:43–45. doi: 10.1016/j.pt.2014.12.008.

Johansson MA, Powers AM, Pesik N, Cohen NJ, Staples JE. 2014. Nowcasting the spread of chikungunya virus in the Americas. *PLOS ONE* **9**:e104915. doi: 10.1371/journal.pone.0104915.

Juliano SA, Lounibos LP, Meara GFO. 2007. A field test for competitive effects of *Aedes albopictus* on *A. aegypti* in South Florida: differences between sites of coexistence and exclusion? *Oecologia* **139**:583–593. doi: 10.1007/s00442-004-1532-4.

Juliano SA, Philip Lounibos L. 2005. Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecology Letters* **8**:558–574. doi: 10.1111/j.1461-0248.2005.00755.

Khan K, Bogoch I, Brownstein JS, Miniota J, Nicolucci A, Hu W, Nsoesie EO, Cetron M, Creatore MI, German M, Wilder-Smith A. 2014. Assessing the origin of and potential for international spread of Chikungunya virus from the Caribbean. *PLOS Currents* **6**. doi: 10.1371/currents.outbreaks.2134a0a7bf37fd8d388181539fea2da5.

Khormi HM, Kumar L. 2014. Climate change and the potential global distribution of *Aedes aegypti*: spatial modelling using geographical information system and CLIMEX. *Geospatial Health* **8**:405–415. doi: 10.4081/gh.2014.29.

Kobayashi AM, Nihei N, Kurihara T. 2002. Analysis of northern distribution of *Aedes albopictus (Diptera: Culicidae)* in Japan by geographical information system. *Journal of Medical Entomology* **39**:4–11. doi: 10.1603/0022-2585-39.1.4.

Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H-J, Hay SI. 2015a. The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Scientific Data* **2**:150035. doi: 10.1038/sdata.2015.35.

Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI. 2015b. Data from: The global compendium of Aedes aegypti and Ae. albopictus occurrence. *Dryad Digital Repository*. doi: 10.5061/dryad.47v3c.

La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, Lenglet A, Jourdain F, Leparc-Goffart I, Charlet F, Ollier L, Mantey K, Mollet T, Fournier JP, Torrents R, Leitmeyer K, Hilairet P, Zeller H, Van Bortel W, Dejour-Salamanca D, Grandadam M, Gastellu-Etchegorry M. 2010. First two autochthonous dengue virus infections in metropolitan France, September 2010. *Euro Surveillance* **15**:19676.

Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD. 2011. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proceedings of the National Academy of Sciences of USA* **108**:7460–7465. doi: 10.1073/pnas.1101377108.

Leathwick J, Elith J, Francis M, Hastie T, Taylor P. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* **321**:267–281. doi: 10.3354/meps321267.

Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. 2014. Chikungunya in the Americas. *Lancet* **383**:514. doi: 10.1016/S0140-6736(14)60185-9.

Li Y, Kamara F, Zhou G, Puthiyakunnon S, Li C, Liu Y, Zhou Y, Yao L, Yan G, Chen XG. 2014. Urbanization increases *Aedes albopictus* larval habitats and accelerates mosquito development and survivorship. *PLOS Neglected Tropical Diseases* **8**:e3301. doi: 10.1371/journal.pntd.0003301.

Linthicum KJ. 1999. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* **285**:397–400. doi: 10.1126/science.285.5426.397.

Lounibos LP. 2002. Invasion by insect vectors of human disease. *Annual Review of Entomology* **47**:233–266. doi: 10.1146/annurev.ento.47.091201.145206.

Lounibos LP, Suárez S, Menéndez Z, Nishimura N, Escher RL, O'Connell SM, Rey JR. 2002. Does temperature affect the outcome of larval competition between *Aedes aegypti* and *Aedes albopictus*? *Journal of Vector Ecology* **27**:86–95.

Luz C, Tai MHH, Santos AH, Silva HHG. 2008. Impact of moisture on survival of *Aedes aegypti* eggs and ovicidal activity of Metarhizium anisopliae under laboratory conditions. *Memorias do Instituto Oswaldo Cruz* **103**:214–215. doi: 10.1590/S0074-02762008000200016.

McArthur MA, Sztein MB, Edelman R. 2013. Dengue vaccines: recent developments, ongoing challenges and current candidates. *Expert Review of Vaccines* **12**:933–953. doi: 10.1586/14760584.2013.815412.

Medley KA. 2010. Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (*Culicidae*), revealed by reciprocal distribution models. *Global Ecology and Biogeography* **19**:122–133. doi: 10.1111/j.1466-8238.2009.00497.x.

Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, Van Bortel W. 2012. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne and Zoonotic Diseases* **12**:435–447. doi: 10.1089/vbz.2011.0814.

Merckx B, Steyaert M, Vanreusel A, Vincx M, Vanaverbeke J. 2010. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecological Modelling* **222**:588–597. doi: 10.1016/j.ecolmodel.2010.11.016.

Messina JP, Brady OJ, Golding N, Pigott DM, Kraemer MUG, Scott TW, Wint GR, Smith DL, Hay SI. 2015. The many projected futures of dengue. *Nature Reviews. Microbiology* **13**:230–239. doi: 10.1038/nrmicro3430.

Morens DM, Fauci AS. 2014. Chikungunya at the door—deja vu all over again? *The New England Journal of Medicine* **371**:885–887. doi: 10.1056/NEJMp1408509.

Morrison AC, Gray K, Getis A, Astete H, Sihuincha M, Focks D, Watts D, Stancil JD, Olson JG, Blair P, Scott TW. 2004. Temporal and geographic patterns of *Aedes aegypti* (Diptera: *Culicidae*) production in Iquitos, Peru. *Journal of Medical Entomology* **41**:1123–1142. doi: 10.1603/0022-2585-41.6.1123.

Nawrocki SJ, Hawley WA. 1987. Estimation of the northern limits of distribution of *Aedes albopictus* in North America. *Journal of the American Mosquito Control Association* **3**:314–317.

Nihei N, Komagata O, Mochizuki K, Kobayashi M. 2014. Geospatial analysis of invasion of the Asian tiger mosquito *Aedes albopictus*: competition with *Aedes japonicus japonicus* in its northern limit area in Japan. *Geospatial Health* **8**:417–427. doi: 10.4081/gh.2014.30.

O'Meara GF, Evans LF, Gettman AD, Cuda JP. 1995. Spread of *Aedes albopictus* and decline of *Ae. aegypti (Diptera: Culicidae)* in Florida. *Journal of Medical Entomology* **32**:554–562. doi: 10.1093/jmedent/32.4.554.

Otero M, Solari HG. 2010. Stochastic eco-epidemiological model of dengue disease transmission by *Aedes aegypti* mosquito. *Mathematical Biosciences* **223**:32–46. doi: 10.1016/j.mbs.2009.10.005.

Paty MC, Six C, Charlet F, Cochet A, Wiegandt A, Chappert JL, Dejour-Salamanca D, Guinard A, Soler P, Servas V, Vivier-Darrigol M, Ledrans M, Debruyne M, Schaal O, Jeannin C, Helynck B, Leparc-Goffart I, Coignard B. 2014. Large number of imported chikungunya cases in mainland France, 2014: a challenge for surveillance and response. *Euro Surveillance* **19**:20856. doi: 10.2807/1560-7917.ES2014.19.28.20856.

Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D. 2009. *Aedes albopictus*, an arbovirus vector: from the darkness to the light. *Microbes and Infection* **11**:1177–1185. doi: 10.1016/j.micinf.2009.05.005.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution model: implications for background and pseudo-absence data. *Ecological Applications* **19**:181–197. doi: 10.1890/07-2153.1.

Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld CC, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI. 2014a. Global distribution maps of the leishmaniases. *eLife* **3**:e02851. doi: 10.7554/eLife.02851.

Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, Brady OJ, Kraemer MU, Smith DL, Moyes CL, Bhatt S, Gething PW, Horby PW, Bogoch II, Brownstein JS, Mekaru SR, Tatem AJ, Khan K, Hay SI. 2014b. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**:e04395. doi: 10.7554/eLife.04395.

Pigott DM, Kraemer MU. 2014. Enhancing infectious disease mapping with open access resources. *Euro Surveillance* **19**.

Poland JD, Calisher CH, Monath TP, Downs WG, Murphy K. 1981. Persistence of neutralizing antibody 30-35 years after immunization with 17D yellow fever vaccine. *Bulletin of the World Health Organization* **59**:895–900.

Ponlawat A, Harrington LC. 2005. Blood feeding patterns of *Aedes aegypti* and *Aedes albopictus* in Thailand. *Journal of Medical Entomology* **42**:844–849. doi: 10.1093/jmedent/42.5.844.

Powell JR, Tabachnick WJ. 2013. History of domestication and spread of *Aedes aegypti*—a review. *Memorias do Instituto Oswaldo Cruz* **108**(Suppl 1):11–17. doi: 10.1590/0074-0276130395.

Powers AM. 2014. Chikungunya virus control: is a vaccine on the horizon? *Lancet* **384**:2008–2009. doi: 10.1016/S0140-6736(14)61290-3.

R Core Team. 2014. *R: a language and environment for computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramos MM, Mohammed H, Zielinski-Gutierrez E, Hayden MH, Lopez JLR, Fournier M, Trujillo AR, Burton R, Brunkard JM, Anaya-Lopez L, Banicki AA, Morales PK, Smith B, Muñoz JL, Waterman SH, Dengue Serosurvey Working Group. 2008. Epidemic dengue and dengue hemorrhagic fever at the Texas-Mexico border: results of a household-based seroepidemiologic survey, December 2005. *The American Journal of Tropical Medicine and Hygiene* **78**:364–369.

Reinert JF, Harbach RE, Kitching IJ. 2009. Phylogeny and classification of tribe Aedini (Diptera: *Culicidae*). *Zoological Journal of the Linnean Society* **157**:700–794. doi: 10.1111/j.1096-3642.2009.00570.x.

Reiter P. 2001. Climate change and mosquito-borne disease. *Environmental Health Perspectives* **109**:141–161. doi: 10.1289/ehp.01109s1141.

Reiter P, Lathrop S, Bunning M, Biggerstaff B, Singer D, Tiwari T, Baber L, Amador M, Thirion J, Hayes J, Seca C, Mendez J, Ramirez B, Robinson J, Rawlings J, Vorndam V, Waterman S, Gubler D, Clark G, Hayes E. 2003. Texas lifestyle limits transmission of dengue virus. *Emerging Infectious Diseases* **9**:86–89. doi: 10.3201/eid0901.020220.

Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli AC, Panning M, Cordioli P, Fortuna C, Boros S, Magurano F, Silvi G, Angelini P, Dottori M, Ciufolini MG, Majori GC, Cassone A, CHIKV Study Group. 2007. Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet* **370**:1840–1846. doi: 10.1016/S0140-6736(07)61779-6.

Ridgeway G. 2013. Generalized boosted regression models. R Packag. Version 2.1.

Romero-Vivas CME, Falconar AKI. 2005. Investigation of relationships between *Aedes aegypti* egg, larvae, pupae, and adult density indices where their main breeding sites were located indoors. *Journal of the American Mosquito Control Association* **21**:15–21. doi: 10.2987/8756-971X(2005)21[15:IORBAA]2.0.CO;2.

Roy CJ, Adams AP, Wang E, Plante K, Gorchakov R, Seymour RL, Vinet-Oliphant H, Weaver SC. 2014. Chikungunya vaccine candidate is highly attenuated and protects nonhuman primates against telemetrically-monitored disease following a single dose. *Journal of Infectious Diseases* **209**:1891–1899. doi: 10.1093/infdis/jiu014.

Russell ABM, Kay BH, Shipton W. 2001. Survival of *Aedes aegypti (Diptera: Culicidae)* eggs in surface and subterranean breeding sites during the northern Queensland dry season. *Journal of Medical Entomology* **38**:441–445. doi: 10.1603/0022-2585-38.3.441.

Schaffner F, Mathis A. 2014. Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *The Lancet Infectious Diseases* **14**:1271–1280. doi: 10.1016/S1473-3099(14)70834-5.

Schneider A. 2012. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment* **124**:689–704. doi: 10.1016/j.rse.2012.06.006.

Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Gary G, Strickman D, Kittayapong P, Edman JD. 2000. Longitudinal studies of *Aedes aegypti* (Diptera: *Culicidae*) in Thailand and Puerto Rico: Blood feeding frequency. *Journal of Medical Entomology* **37**:89–101. doi: 10.1603/0022-2585-37.1.89.

Scott TW, Takken W. 2012. Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission. *Trends in Parasitology* **28**:114–121. doi: 10.1016/j.pt.2012.01.001.

Seebens H, Gastner MT, Blasius B. 2013. The risk of marine bioinvasion caused by global shipping. *Ecology Letters* **16**:782–790. doi: 10.1111/ele.12111.

Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, Suk JE, Van Bortel W, Khan K. 2014. International dispersal of dengue through air travel: importation risk for Europe. *PLOS Neglected Tropical Diseases* **8**:e3278. doi: 10.1371/journal.pntd.0003278.

Simmons CP, Farrar JJ, Chau NVV, Wills B. 2012. Dengue. *The New England Journal of Medicine* **366**:1423–1432. doi: 10.1056/NEJMra1110265.

Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IR, Kabaria CW, Harbach RE, Hay SI. 2011. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors* **4**:89. doi: 10.1186/1756-3305-4-89.

Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Okara RM, Van Boeckel T, Godfray HC, Harbach RE, Hay SI. 2010a. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors* **3**:117. doi: 10.1186/1756-3305-3-117.

Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, Van Boeckel T, Kabaria CW, Harbach RE, Hay SI. 2010b. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors* **3**:72. doi: 10.1186/1756-3305-3-72.

Sota T, Mogi M. 1992. Interspecific variation in desiccation survival time of *Aedes (Stegomyia)* mosquito eggs is correlated with habitat and egg size. *Oecologia* **90**:353–358. doi: 10.1007/BF00317691.

Staples JE, Breiman RF, Powers AM. 2009. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clinical Infectious Diseases* **49**:942–948. doi: 10.1086/605496.

Takumi K, Scholte EJ, Braks M, Reusken C, Avenell D, Medlock JM. 2009. Introduction, scenarios for establishment and seasonal activity of *Aedes albopictus* in The Netherlands. *Vector Borne and Zoonotic Diseases* **9**:191–196. doi: 10.1089/vbz.2008.0038.

Tatem AJ, Hay SI, Rogers DJ. 2006. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of USA* **103**:6242–6247. doi: 10.1073/pnas.0508391103.

Thomas S, Obermayr U, Fischer D, Kreyling J, Beierkuhnlein C. 2012. Low-temperature threshold for egg survival of a post-diapause and non-diapause European aedine strain, *Aedes albopictus* (Diptera: *Culicidae*). *Parasites & Vectors* **5**:100. doi: 10.1186/1756-3305-5-100.

Tomasello D, Schlagenhauf P. 2013. Chikungunya and dengue autochthonous cases in Europe, 2007–2012. *Travel Medicine and Infectious Disease* **11**:274–284. doi: 10.1016/j.tmaid.2013.07.006.

Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. 2007. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLOS Pathogens* **3**:e201. doi: 10.1371/journal.ppat.0030201.

Tsetsarkin KA, Weaver SC. 2011. Sequential adaptive mutations enhance efficient vector switching by chikungunya virus and its epidemic emergence. *PLOS Pathogens* **7**:e1002412. doi: 10.1371/journal.ppat.1002412.

Tsuda Y, Takagi M. 2001. Survivial and development of *Aedes aegypti* and *Aedes albopictus* (Diptera: *Culicidae*) Larvae under a seasonally changing environment in Nagasaki, Japan. *Environmental Entomology* **30**:855–860. doi: 10.1603/0046-225X-30.5.855.

Vazeille M, Moutailler S, Coudrier D, Rousseaux C, Khun H, Huerre M, Thiria J, Dehecq JS, Fontenille D, Schuffenecker I, Despres P, Failloux AB. 2007. Two chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, *Aedes albopictus*. *PLOS ONE* **2**:e1168. doi: 10.1371/journal.pone.0001168.

Villar L, Dayan GH, Arredondo-García JL, Rivera DM, Cunha R, Deseda C, Reynales H, Costa MS, Morales-Ramírez JO, Carrasquilla G, Rey LC, Dietze R, Luz K, Rivas E, Miranda Montoya MC, Cortés Supelano M, Zambrano B, Langevin E, Boaz M, Tornieporth N, Saville M, Noriega F, CYD15 Study Group. 2015. Efficacy of a Tetravalent dengue vaccine in Children in Latin america. *The New England Journal of Medicine* **372**:113–123. doi: 10.1056/NEJMoa1411037.

Weaver SC. 2014. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health. *PLOS Neglected Tropical Diseases* **8**:e2921. doi: 10.1371/journal.pntd.0002921.

Weaver SC, Lecuit M. 2015. Chikungunya virus and the global spread of a mosquito-borne disease. *The New England Journal of Medicine* **372**:1231–1239. doi: 10.1056/NEJMra1406035.

Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. 2014. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* **98**:106–118. doi: 10.1016/j.isprsjprs.2014.10.001.

World Health Organization. 1990. Yellow fever in 1988. *Weekly Epidemiological Record* **65**:213–219.

Wu JY, Lun ZR, James AA, Chen XG. 2010. Dengue fever in Mainland China. *The American Journal of Tropical Medicine and Hygiene* **83**:664–671. doi: 10.4269/ajtmh.2010.09-0755.

Yang CF, Hou JN, Chen TH, Chen WJ. 2014. Discriminable roles of *Aedes aegypti* and *Aedes albopictus* in establishment of dengue outbreaks in Taiwan. *Acta Tropica* **130**:17–23. doi: 10.1016/j.actatropica.2013.10.013.

Zouache K, Fontaine A, Vega-rua A, Mousson L, Thiberge J, Lourenco-de-oliveira R, Caro V, Lambrechts L, Failloux AB. 2014. Three-way interactions between mosquito population, viral strain and temperature underlying chikungunya virus transmission potential. *Proceedings. Biological Sciences/The Royal Society* **281**. doi: 10.1098/rspb.2014.1078.

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Risk Factors

» Dengue Virus

» Ecological Epidemiology

» Entomology

The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence

Moritz U.G. Kraemer[1], Marianne E. Sinka[2,3], Kirsten A. Duda[1], Adrian Mylne[2,3], Freya M. Shearer[2,3], Oliver J. Brady[2,3], Jane P. Messina[1], Christopher M. Barker[4,5,6], Chester G. Moore[7], Roberta G. Carvalho[8], Giovanini E. Coelho[8], Wim Van Bortel[9], Guy Hendrickx[10], Francis Schaffner[10], G.R. William Wint[1,11], Iqbal R.F. Elyazar[12], Hwa-Jen Teng[13] & Simon I. Hay[1,6]

Received: 30 March 2015

Accepted: 23 June 2015

Published: 7 July 2015

Aedes aegypti and *Ae. albopictus* are the main vectors transmitting dengue and chikungunya viruses. Despite being pathogens of global public health importance, knowledge of their vectors' global distribution remains patchy and sparse. A global geographic database of known occurrences of *Ae. aegypti* and *Ae. albopictus* between 1960 and 2014 was compiled. Herein we present the database, which comprises occurrence data linked to point or polygon locations, derived from peer-reviewed literature and unpublished studies including national entomological surveys and expert networks. We describe all data collection processes, as well as geo-positioning methods, database management and quality-control procedures. This is the first comprehensive global database of *Ae. aegypti* and *Ae. albopictus* occurrence, consisting of 19,930 and 22,137 geo-positioned occurrence records respectively. Both datasets can be used for a variety of mapping and spatial analyses of the vectors and, by inference, the diseases they transmit.

Design Type(s)	observation design • epidemiological study • data integration objective
Measurement Type(s)	viral host epidemiology
Technology Type(s)	data collection method
Factor Type(s)	
Sample Characteristic(s)	Aedes aegypti • anthropogenic habitat • Aedes albopictus

[1]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2]Wellcome Trust Centre for Human Genetics,University of Oxford, Oxford, UK. [3]Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA. [4]Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA, USA. [5]Center for Vectorborne Diseases, University of California, Davis, CA, USA. [6]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. [7]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, USA. [8]National Dengue Control Program, Ministry of Health, Brasilia, DF, Brazil. [9]European Centre for Disease Prevention and Control, Stockholm, Sweden. [10]Avia-GIS, Zoersel, Belgium. [11]Environmental Research Group Oxford Ltd, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK. [12]Eijkman-Oxford Clinical Research Unit, Jakarta, Indonesia. [13]Center for Research, Diagnostics and Vaccine Development, Centers for Disease Control, Taipei, Taiwan (ROC). Correspondence and requests for materials should be addressed to S.I.H. (email: simon.i.hay@gmail.com).

Background & Summary

Aedes aegypti [=*Stegomyia aegypti*[1]] and *Ae. albopictus* [=*Stegomyia albopicta*[1]] are disease vectors for many important viral human diseases such as dengue, chikungunya and yellow fever[2–4]. Dengue is the most prevalent human arboviral infection causing approximately 100 million apparent annual infections with almost half of the world's population at risk[5]. Dengue transmission now occurs in over 120 countries[6], mostly in the tropics and sub-tropics. Chikungunya, another arthropod-borne virus, has caused over 2.5 million infections over the past decade and has more recently been spreading in the Americas and emerging in Europe, posing new challenges to health systems as it spreads into new areas, infecting naïve populations and consequently causing large outbreaks[7–10]. The disease burden of yellow fever was significantly reduced due to large-scale vaccination programs in the twentieth century but current estimates of 51,000–380,000 severe cases in Africa per year point to the continuing difficulty in fully controlling this virus[11]. As a result, there is growing interest in describing the global geographic distribution of both vector species to better understand the risk of transmission of these viruses.

Aedes aegypti is a predominantly urban vector, utilising the abundance of artificial containers as larval sites and feeding almost exclusively on humans[12]. *Aedes albopictus* can more often be found in peri-urban and rural environments, feeding readily on a variety of mammalian (including humans) and avian species[13].

Aedes mosquito surveys are performed to better understand ecological and epidemiological aspects of the vectors as well as to assist disease surveillance and control[14–16]. Surveillance of *Aedes* can involve; (i) systematic household surveys that involve searching water-filled containers for larvae and pupae, (ii) the use of backpack aspirators and suction traps baited with a chemical lure and/or CO_2 for the collection of adult mosquitoes or, (iii) using ovitraps placed strategically around a neighbourhood to collect mosquito eggs that can then be reared back in the laboratory for morphological identification or directly processed for molecular identification[17–19].

The database described here contains information on the known global occurrences of the adults, pupae, larvae or eggs of *Ae. aegypti* and *Ae. albopictus* globally from 1960 and 2014.

By including data from a variety of sources we were able to create the largest currently available standardised up-to-date global dataset for both *Ae. aegypti* and *Ae. albopictus* (Fig. 1), containing 42,067 geo-positioned occurrences.

Methods

Data collection

PubMed (http://www.ncbi.nlm.nih.gov/nlmcatalog/journals) was searched using the term '*Aedes*' OR '*aegypti*' OR '*albopictus*' for the years 1960 to 2013. The Medical Subject Headings (MeSH) term technology used in the PubMed citation archive ensured all pseudonyms were automatically included (http://www.nlm.nih.gov/mesh) in the searches. The same process was repeated for ISI Web of Science (http://wok.mimas.ac.uk) and ProMED (http://www.promedmail.org). The searches were last updated on 15th November 2013. No language restrictions were placed on these searches; however, only those citations with a full title and abstract were retrieved. This resulted in a collection of 8,597 references, of which 2,804 unique articles were identified from their abstracts as potentially containing useable location data. In-house language skills allowed processing of all English, French, Portuguese and Spanish articles. Confirmed *Aedes* occurrences within these articles were entered into the database. Occurrences were classified as confirmed when the article clearly stated the presence of the vector at a specific time in a specific location. This includes transient populations, i.e., found in ports or only during the summer months. Only for Europe were we able to include information of transient versus established populations using expert opinions. Laboratory studies were included if the mosquito/larvae were collected from the wild specifically for the purpose of the study. Occurrences were recorded separately for both species. More specific information about 'sub-species' or 'genetic characteristics' for example, were recorded where available but not included in the final database. This information can be obtained from the authors upon request. Data Citation 1 lists full references for each published record in the database.

In addition to the data directly sourced from published literature, primary and unpublished occurrence data from national entomological surveys were obtained through contact with administrators of these surveys when possible (Fig. 1). This includes primary data for *Ae. albopictus* provided from an earlier published article by Carvalho *et al.*[20] Collections have been part of the Levantamento Rapido de Indice para *Aedes aegypti* (LIRAa) in Brazil and are described in full elsewhere[20,21]. Similarly, *Ae. aegypti* primary data with geographic locations were also provided by an entomological survey directly from the Ministry of Health of Brazil for 2013. All occurrences for Brazil were classified as polygons (see geo-positioning methods) as they represent surveys conducted in Brazilian municipalities with their respective centroids being used as geographic information in the database. Primary and unpublished occurrence points for both species were provided by Elyazar (2014) for Indonesia. Moore and Barker (2014) provided summary data from the United States, collected between 1985–2014 from published and unpublished sources, including geographic location at the USA county administrative level which were included as polygons in the database. In addition, Teng (2014) added occurrence records for both species based on national entomological surveys performed in Taiwan between 2004–2013. European *Ae. albopictus* records were derived from the European Centre for Disease Prevention and Control (ECDC) funded TIGERMAPS

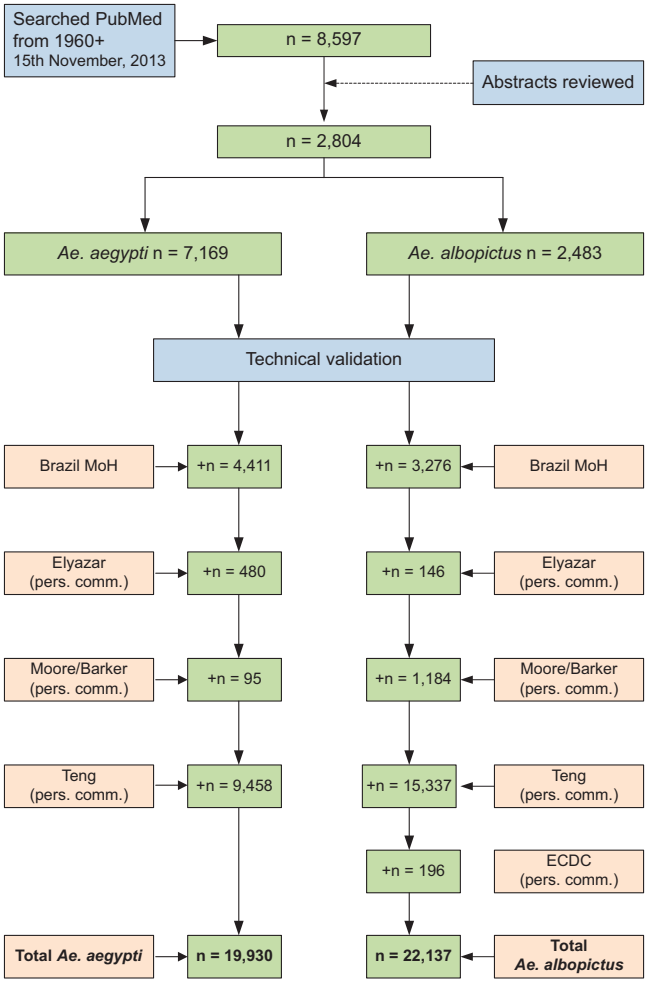


Figure 1. The numbers of occurrences for *Ae. aegypti and Ae. albopictus* by source.

and VBORNET datasets provided by Schaffner and Hendrickx (2013). Data were based on Nomenclature of territorial units for statistics 3 (NUTS3) level for Europe from 1979–2013. NUTS3 centroids were calculated in ArcGIS, occurrences classified as polygons and added to the database.

Geo-positioning of data from published sources

All available location information was extracted for each occurrence from the relevant primary research article. The site name was used together with all contextual information provided about the site position to determine its latitudinal and longitudinal coordinates using Google Maps (https://www.maps.google.co.uk), Google Earth (http://www.google.co.uk/intl/en_uk/earth), or other online geo-positional databases including Geonames (http://www.geonames.org), Fallingrain (http://www.fallingrain.com/world/index.html) or as a last resort, using simple Google searches. Place names are often duplicated within a country, so contextual information was used to ensure the right site was selected. When the site name was not found, information from the text, was also used to scan sites in the approximate area to check for alternate spelling of the site name. If the study site could be geo-positioned to a specific latitude and longitude within a 5 km × 5 km pixel, it was termed a 'point location'. For each occurrence that could not be assigned a single 5 km × 5 km pixel, e.g., a large city, the occurrence was entered as a polygon data type. Polygon occurrences were subsequently classified based on the polygon size they correspond to as either between 5–10 km^2, 10–25 km^2, 25–100 km^2 or >100 km^2. All locations were then linked to administrative units as recognised by the FAO Global Administrative Unit Layer (GAUL) system[22]. This initial database then underwent spatial and temporal standardisation and finally technical validation.

Occurrence database management: spatial and temporal standardisation

As the database was compiled from many different sources and several institutions, it was first necessary to standardise the data entries such that identical locations which may have been geo-positioned slightly differently were given the same unique identifier. Point records were given the same unique identifier if they lay within the same 5 km × 5 km pixel within a global grid. Finally, any record associated with a

polygon measuring larger than 111 km × 111 km at the equator (1 degree) was removed from the database ($n = 475$, $n = 54$ for *Ae. aegypti* and *Ae. albopictus* respectively).

Similarly, it was necessary to temporally standardise the database to avoid duplicates. We chose to define a single occurrence at a given unique location (as identified above) within one calendar year. This was particularly important for oversampled regions that undergo multiple yearly surveys such as Taiwan and involved a procedure which: (i) disaggregated any records which were in the same location but spanning multiple years into individual records for each respective year and then (ii) aggregated all records with the same unique location identifier and occurring within the same year to form a single occurrence record. This led to 1,112 and 370 records being removed for *Ae. aegypti* and *Ae. albopictus,* respectively.

Data Records

This database is publicly available online as a comma-delimited file for both species independently for ease of use and the ability to import it into a variety of software programs (Data Citation 1). Each of the rows represents a single occurrence record (one or more *Aedes* cases in the same unique location within a single calendar year). The fields contained in the database are as follows:

1. **VECTOR:** Identifying the species; *Ae. aegypti* or *Ae. albopictus*
2. **OCCURRENCE_ID:** Unique identifier for each occurrence in the database after temporal and spatial standardisation.
3. **SOURCE_TYPE:** Published literature or unpublished sources with reference ID that corresponds to the full list of references in Data Citation 1.
4. **LOCATION_TYPE:** Whether the record represents a point or a polygon location.
5. **POLYGON_ADMIN:** Admin level or polygon size which the record represents when the location type is a polygon. –999 when the location type is a point (5 km × 5 km).
6. **X:** The longitudinal coordinate of the point or polygon centroid (WGS1984 Datum).
7. **Y:** The latitudinal coordinate of the point or polygon centroid (WGS1984 Datum).
8. **YEAR:** The year of the occurrence.
9. **COUNTRY:** The name of the country within which the occurrence lies.
10. **COUNTRY_ID:** ISO alpha-3 country codes.
11. **GAUL_AD0:** The country-level global administrative unit layer (GAUL) code (see http://www.fao.org/geonetwork) which identifies the Admin-0 polygon within which any smaller polygons and points lie.
12. **STATUS:** Established versus transient populations.

Technical Validation

The following procedures were carried out on the final database to ensure the accuracy and validity of the occurrence records.

1. A raster distinguishing land from water[22] was created at a 5 km × 5 km resolution and was used to ensure all occurrences were positioned on a valid land pixel ($n = 95$ and $n = 64$ records were removed for *Ae. aegypti* and *Ae. albopictus* respectively).
2. We cross-validated all of the unique occurrence locations against temperature-based *Aedes* population persistence metrics developed by Brady *et al.*[23] In brief, this classification was determined by modelling the effect of temperature on adult *Ae. aegypti* and *Ae. albopictus* survival and length of first gonotrophic cycle, the interaction of which determines whether the population can persist. Population

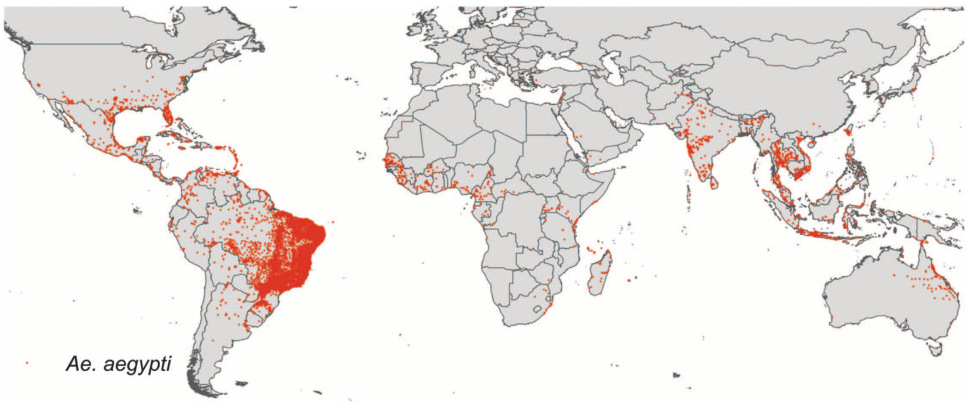


Figure 2. Map of occurrence points for *Ae. aegypti*.

persistence was then predicted on a global scale using interpolated meteorological data[24]. Occurrences that fell outside this range were re-checked to ensure the quality of the occurrence records.

The result is a database consisting of 19,930 and 22,137 geo-positioned occurrences in total worldwide for *Ae. aegypti* and *Ae. albopictus* respectively, broken down by region, location type and source type in Fig. 1. In Figs 2 and 3 the global geographic distribution of both species is displayed.

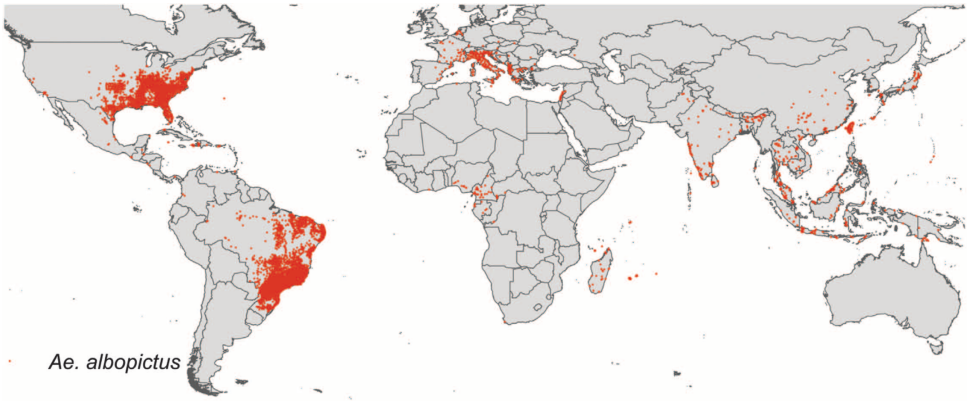


Figure 3. Map of occurrence points for *Ae. albopictus*.

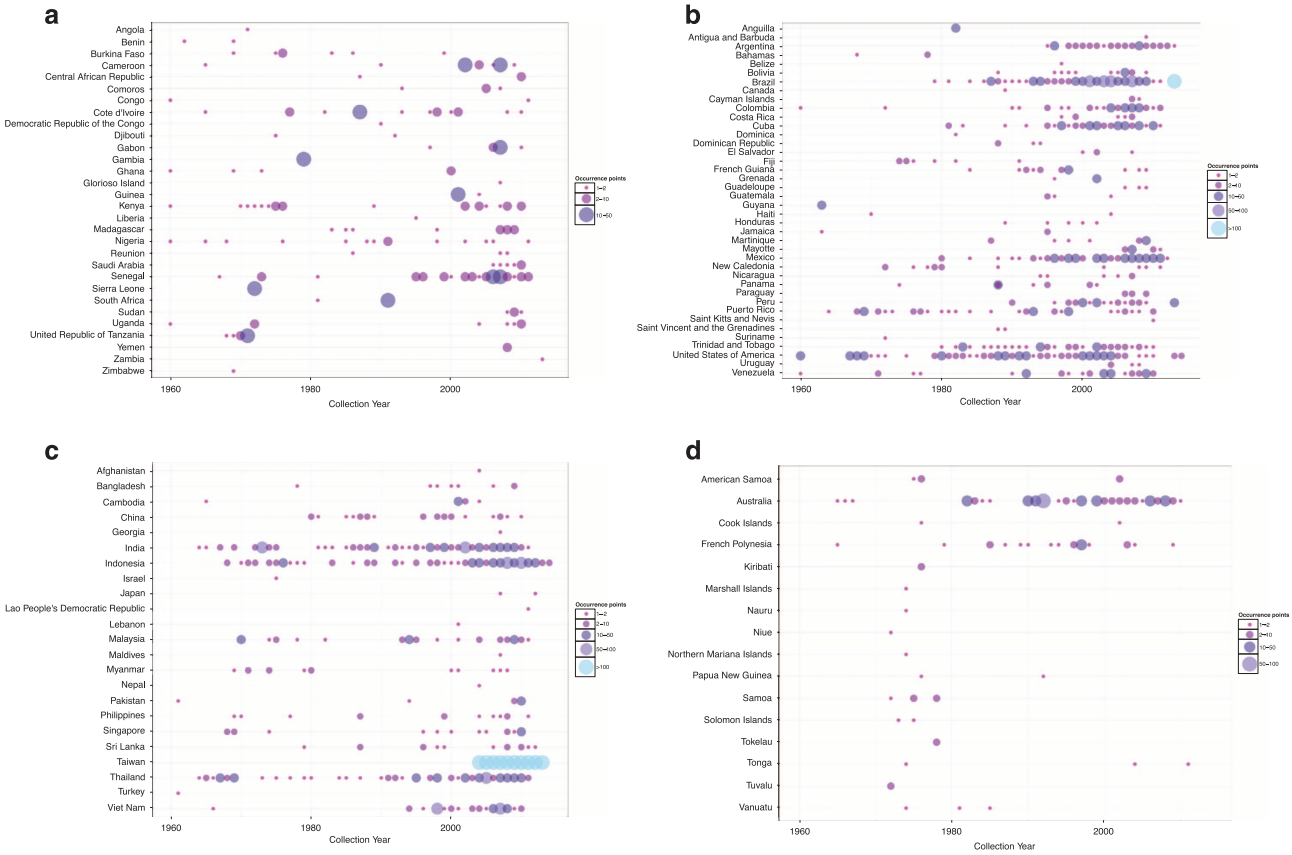


Figure 4. The numbers of occurrences for *Ae. aegypti* with locations per year separated by region, with panels (**a**–**d**) representing regions. Colours indicate the number of mosquitoes recorded in each year with the size of the circles scaled respectively.

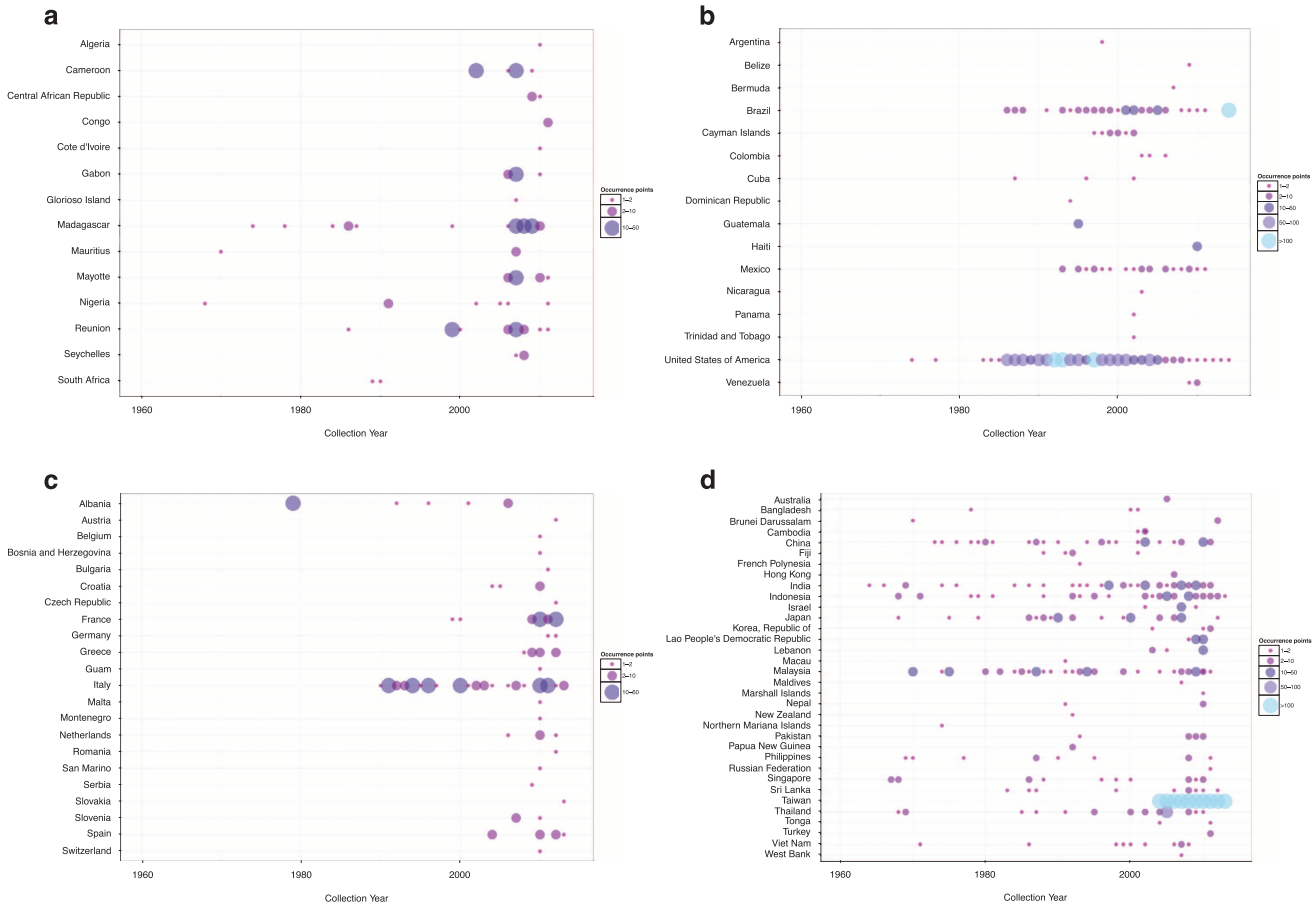


Figure 5. The numbers of occurrences for *Ae. albopictus* with locations per year separated by region, with panels (**a**–**d**) representing regions. Colours indicate the number of mosquitoes recorded in each year with the size of the circles scaled respectively.

Figs 4 and 5 show occurrence records by year and region for both *Ae. aegypti* and *Ae. albopictus* respsectively. The increase in occurrence records since 2013 is largely attributable to the LIRAa data from Brazil.

Usage Notes

The dataset described here can be used to investigate the spatial and temporal patterns of *Aedes* distribution at multiple scales and resolutions. As *Ae. aegypti* and *Ae. albopictus* are invasive species, spreading to new areas via shipping routes and human movement[25–27], this dataset could improve predictions of locations at high-risk for importation[25]. This dataset can also be used to contribute to modelling areas at risk for dengue[28] and chikungunya[9] especially in areas in Europe[16,28] and the USA[29,30]. We aimed at building a comprehensive set of data based on occurrences ever recorded globally including their respective dates to allow researchers as well as policy makers to filter the dataset based on their respective research questions.

This dataset was first used in an ecological niche modelling framework along with a set of environmental covariates to map the global distribution of each species[32]. A generic code to produce the global risk maps is openly available as an R software package 'seegSDM' from GitHub (https://github.com/SEEG-Oxford/seegSDM). Such maps can help to guide vector surveillance efforts in countries where the distribution of both species is not well-known, but which are at high risk for importation of related viruses.

Regional biases in density of occurrence records are apparent and may be due to differences in the amount of regular surveillance, differences in the number of published studies and availability of routinely collected data. Use on a global scale, however, would need to take into account geographical sampling bias as done in Kraemer *et al.* using similarly biased background points in a presence-only niche modelling approach[31,32]. The method for accounting for sampling bias, however, might vary depending on the research question asked and methodology applied in subsequent analyses.

References

1. Reinert, J. F., Harbach, R. E. & Kitching, I. J. Phylogeny and classification of tribe *Aedini (Diptera: Culicidae)*. *Zool. J. Linn. Soc.* **157,** 700–794 (2009).
2. Hay, S. I. *et al.* Global mapping of infectious disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci* **368,** 20120250 (2013).
3. Juliano, S. a. & Philip Lounibos, L. Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecol. Lett.* **8,** 558–574 (2005).
4. Morrison, A. C., Zielinski-Gutierrez, E., Scott, T. W. & Rosenberg, R. Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti*. *PLoS Med.* **5,** e68 (2008).
5. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
6. Brady, O. J. *et al.* Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6,** e1760 (2012).
7. Staples, J. E. & Fischer, M. Chikungunya virus in the Americas—what a vectorborne pathogen can do. *N. Engl. J. Med.* **371,** 887–889 (2014).
8. Sharp, T. M. *et al.* Chikungunya Cases Identified Through Passive Surveillance and Household Investigations—Puerto Rico, May 5—August 12, 2014. *MMWR Morb Mortal Wkly Rep.* **63,** 500–501 (2014).
9. Powers, A. M. Chikungunya virus control: is a vaccine on the horizon? *Lancet* **384,** 2008–2009 (2014).
10. Schaffner, F., Medlock, J. M. & Van Bortel, W. Public health significance of invasive mosquitoes in Europe. *Clin. Microbiol. Infect.* **19,** 685–692 (2013).
11. Garske, T. *et al.* Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med.* **11,** e1001638 (2014).
12. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of *Aedes aegypti*—a review. *Mem. Inst. Oswaldo Cruz* **108** (Suppl.), 11–17 (2013).
13. Delatte, H. *et al.* Geographic distribution and developmental sites of *Aedes albopictus (Diptera: Culicidae)* during a Chikungunya epidemic event. *Vector borne zoonotic Dis.* **8,** 25–34 (2008).
14. Zouache, K. *et al.* Three-way interactions between mosquito population, viral strain and temperature underlying chikungunya virus transmission potential. *Proc. R. Soc. Biol. Sci.* **281,** 20141078 (2014).
15. Gubler, D. J. Prevention and Control of *Aedes aegypti*-borne Diseases: Lesson Learned from Past Successes and Failures. *AsPac J. Mol. Biol. Biotechnol.* **19,** 111–114 (2011).
16. Schaffner, F. & Mathis, A. Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *Lancet. Infect. Dis.* **14,** 1271–1280 (2014).
17. Honório, N. A. *et al.* Dispersal of *Aedes aegypti* and *Aedes albopictus (Diptera: Culicidae)* in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* **98,** 191–198 (2003).
18. Schaffner, F. *et al.* Development of guidelines for the surveillance of invasive mosquitoes in Europe. *Parasit. Vectors* **6,** 209 (2013).
19. Schaffner, F., Kaufmann, C., Pflüger, V. & Mathis, A. Rapid protein profiling facilitates surveillance of invasive mosquito species. *Parasit. Vectors* **7,** 142 (2014).
20. Carvalho, R. G., Lourenço-de-Oliveira, R. & Braga, I. A. Updating the geographical distribution and frequency of *Aedes albopictus* in Brazil with remarks regarding its range in the Americas. *Mem. Inst. Oswaldo Cruz* **109,** 787–796 (2014).
21. Coelho, G. E. Challenges in the control of *Aedes aegypti*. *Rev. Inst. Med. Trop. Sao Paolo* **54,** S13–S14 (2012).
22. Food and Agricultural Organization of the United Nations. The Global Administrative Unit Layers (GAUL): Technical Aspects. *Food Agric. Organ. United Nations, EC-FAO Food Secur. Program.* (2008).
23. Brady, O. J. *et al.* Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit. Vectors* **7,** 338 (2014).
24. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25,** 1965–1978 (2005).
25. Tatem, A. J., Hay, S. I. & Rogers, D. J. Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. USA* **103,** 6242–6247 (2006).
26. Hawley, W. A., Reiter, P., Copeland, R. S., Pumpuni, C. B. & Craig, G. B. *Aedes albopictus* in North America: probable introduction in used tires from northern Asia. *Science* **236,** 1114–1116 (1987).
27. Hofhuis, A. *et al.* The hidden passenger of lucky bamboo: do imported *Aedes albopictus* mosquitoes cause dengue virus transmission in the Netherlands? *Vector borne zoonotic Dis* **9,** 217–220 (2009).
28. Semenza, J. C. *et al.* International dispersal of dengue through air travel: importation risk for Europe. *PLoS Negl. Trop. Dis.* **8,** e3278 (2014).
29. Morens, D. M. & Fauci, A. S. Chikungunya at the door—deja vu all over again? *N. Engl. J. Med.* **371,** 885–887 (2014).
30. Khan, K. *et al.* Assessing the origin of and potential for international spread of Chikungunya virus from the Caribbean. *PLOS Curr. Outbreaks* Jun 6, doi: 10.1371/currents.outbreaks.2134a0a7b"f37fd8d388181539fea2da5 (2014).
31. Phillips, S. J. *et al.* Sample selection bias and presence-only distribution model: implications for background and pseudo-absence data. *Ecol. Appl.* **19,** 181–197 (2009).
32. Kraemer, M. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife.* **4,** e08347 (2015).

Data Citation

1. Kraemer, M. U. G. *et al. Dryad Digital Repository* http://dx.doi.org/10.5061/dryad.47v3c (2015).

Acknowledgements

M.U.G.K. is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. M.E.S. is funded by a project grant from the Bill and Melinda Gates Foundation via the VecNet consortium (http://vecnet.org). O.J.B. is funded by a BBSRC studentship. J.P.M. and G.R.W.W. are funded by the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, European Commission 7th Framework Programme (21803), http:www.idams.eu, Publication #28). I.R.F.E. is funded by the Wellcome Trust (#B9RZGS0). VBORNET is an ECDC funded project (contract number ECDC/09/018). S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) which also supports A.M. and K.A.D. S.I.H. and C.M.D. also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. FMS is funded by the Rhodes Trust.

We thank Dr Roseli La Corte dos Santos, Departamento de Morfologia, Universidade Federal de Sergipe for providing useful guidance, the Ministry of Health of Brazil for providing additional data, Bimandra Djaafara and Karin Dian Lestari of EOCRU for their work on geo-positioning data from Indonesia. We thank all contributors to the Europe *Ae. albopictus* dataset (a detailed list in Data Citation 1).

Author Contributions

M.U.G.K. drafted the manuscript with editorial input from J.P.M. and S.I.H. and approval from all authors. M.E.S. coordinated and compiled the data collection. K.A.D., A.M., M.U.G.K. and F.M.S. compiled the data records. M.U.G.K. performed database standardisation and technical validation. C.M.B., C.G.M., R.G.C., G.E.C., I.R.F.E., H.J.T., W.V.B., F.S., G.H., O.J.B. and G.R.W.W. provided additional data and geo-positioning. S.I.H. conceived the database design and advised on standardisation and validation procedures.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kraemer, M. U. G. *et al.* The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Sci. Data* 2:150035 doi: 10.1038/sdata.2015.35 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0

Metadata associated with this Data Descriptor is available at http://www.nature.com/sdata/ and is released under the CC0 waiver to maximize reuse.

Chapter 2 – The global distribution and environmental suitability of chikungunya virus

---

This chapter analyses the global distribution and environmental suitability of chikungunya virus (CHIKV) and provides estimates of their population living in areas where transmission may occur. By using modelling outputs from chapter 1, this chapter directly builds upon evidence developed in this thesis. This work has been published in *Eurosurveillance* and is included here in its final form.

Research article

Global distribution and environmental suitability for chikungunya virus, 1952 to 2015

EO Nsoesie [1 2 3 4], MU Kraemer [2 5], N Golding [6 7], DM Pigott [4 6], OJ Brady [6], CL Moyes [6], MA Johansson [8 9], PW Gething [10], R Velayudhan [11], K Khan [12], SI Hay [4 6 13], JS Brownstein [13 13 14]

1. Children's Hospital Informatics Program, Boston Children's Hospital, Boston, United States
2. These authors contributed equally to this manuscript as first authors
3. Department of Pediatrics, Harvard Medical School, Boston, United States
4. Institute of Health Metrics and Evaluation, University of Washington, Seattle, United States
5. Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, United Kingdom
6. Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, United Kingdom
7. Department of BioScience, University of Melbourne, Australia
8. Centers for Disease Control and Prevention, San Juan, Puerto Rico
9. Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, United States
10. Malaria Atlas Project, Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, United Kingdom
11. World Health Organization, Geneva, Switzerland
12. Li Ka Shing Knowledge Institute, Division of Infectious Diseases, St Michael's Hospital, Toronto, Canada
13. These authors contributed equally to the manuscript as last authors
14. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Correspondence: Elaine O. Nsoesie (en22@uw.edu), Moritz U.G. Kraemer (moritz.kraemer@zoo.ox.ac.uk)

Citation style for this article:
Nsoesie EO, Kraemer MU, Golding N, Pigott DM, Brady OJ, Moyes CL, Johansson MA, Gething PW, Velayudhan R, Khan K, Hay SI, Brownstein JS. Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. Euro Surveill. 2016;21(20):pii=30234. DOI: http://dx.doi.org/10.2807/1560-7917.ES.2016.21.20.30234

Article submitted on 21 September 2015 / accepted on 20 March 2016 / published on 20 March 2016

Chikungunya fever is an acute febrile illness caused by the chikungunya virus (CHIKV), which is transmitted to humans by *Aedes* mosquitoes. Although chikungunya fever is rarely fatal, patients can experience debilitating symptoms that last from months to years. Here we comprehensively assess the global distribution of chikungunya and produce high-resolution maps, using an established modelling framework that combines a comprehensive occurrence database with bespoke environmental correlates, including up-to-date *Aedes* distribution maps. This enables estimation of the current total population-at-risk of CHIKV transmission and identification of areas where the virus may spread to in the future. We identified 94 countries with good evidence for current CHIKV presence and a set of countries in the New and Old World with potential for future CHIKV establishment, demonstrated by high environmental suitability for transmission and in some cases previous sporadic reports. *Aedes aegypti* presence was identified as one of the major contributing factors to CHIKV transmission but significant geographical heterogeneity exists. We estimated 1.3 billion people are living in areas at-risk of CHIKV transmission. These maps provide a baseline for identifying areas where prevention and control efforts should be prioritised and can be used to guide estimation of the global burden of CHIKV.

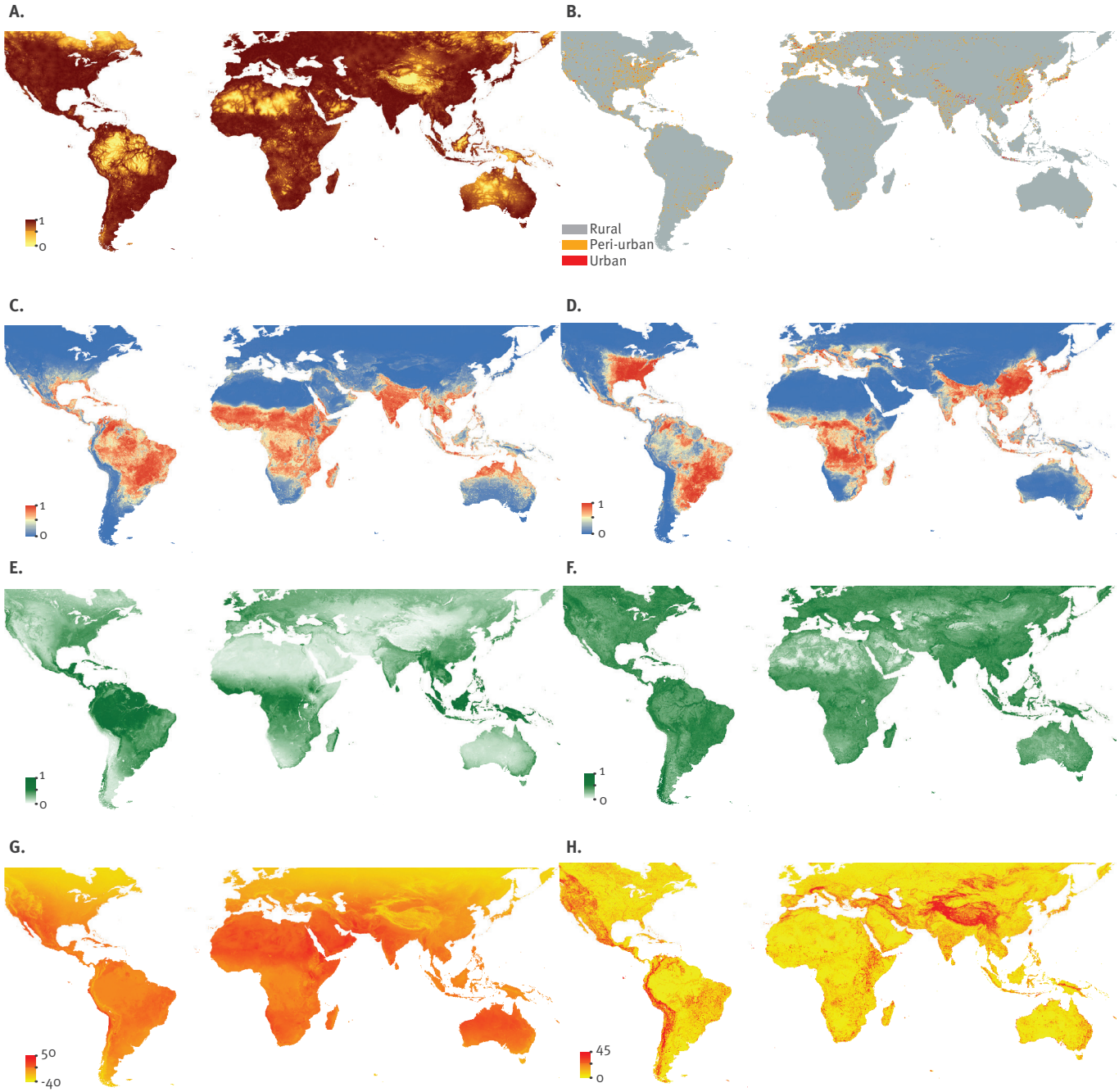
Introduction

Recent emergence and re-emergence of chikungunya virus (CHIKV) in several regions globally underscores the importance of implementing and strengthening surveillance systems for rapid and accurate case identification [1]. CHIKV causes an acute febrile illness with severe arthralgia [2]. It is a mosquito-borne pathogen of the genus *Alphavirus* and transmitted to and between humans by *Aedes* mosquitoes with four circulating genotypes [3,4]. In the eastern part of Africa CHIKV is sustained in an enzootic, sylvatic cycle that involve arboreal mosquito vectors and non-human primates [5]. Clinical manifestations of the disease usually include fever, rash and arthralgia, which are similar to dengue symptoms, potentially leading to misdiagnosis and underreporting in the absence of specific laboratory diagnostic testing [6]. CHIKV and dengue also share the same primary vector *Ae. aegypti* and occasional co-infection has been reported [7]. Although it has been shown that also *Ae. albopictus* can transmit the virus between humans such outbreaks have been more localised [8]. Mortality from CHIKV infection is rare, infected persons can experience prolonged joint pain and arthritis over a period of months to years. There are also no antiviral agents or treatments for chikungunya fever (CHIK), but candidate vaccines are under development [9].

In 2013, the first autochthonous transmission of CHIKV in the western hemisphere was reported on Saint

Figure 1

Environmental covariates used in this study to predict the global environmental suitability of chikungunya virus transmission

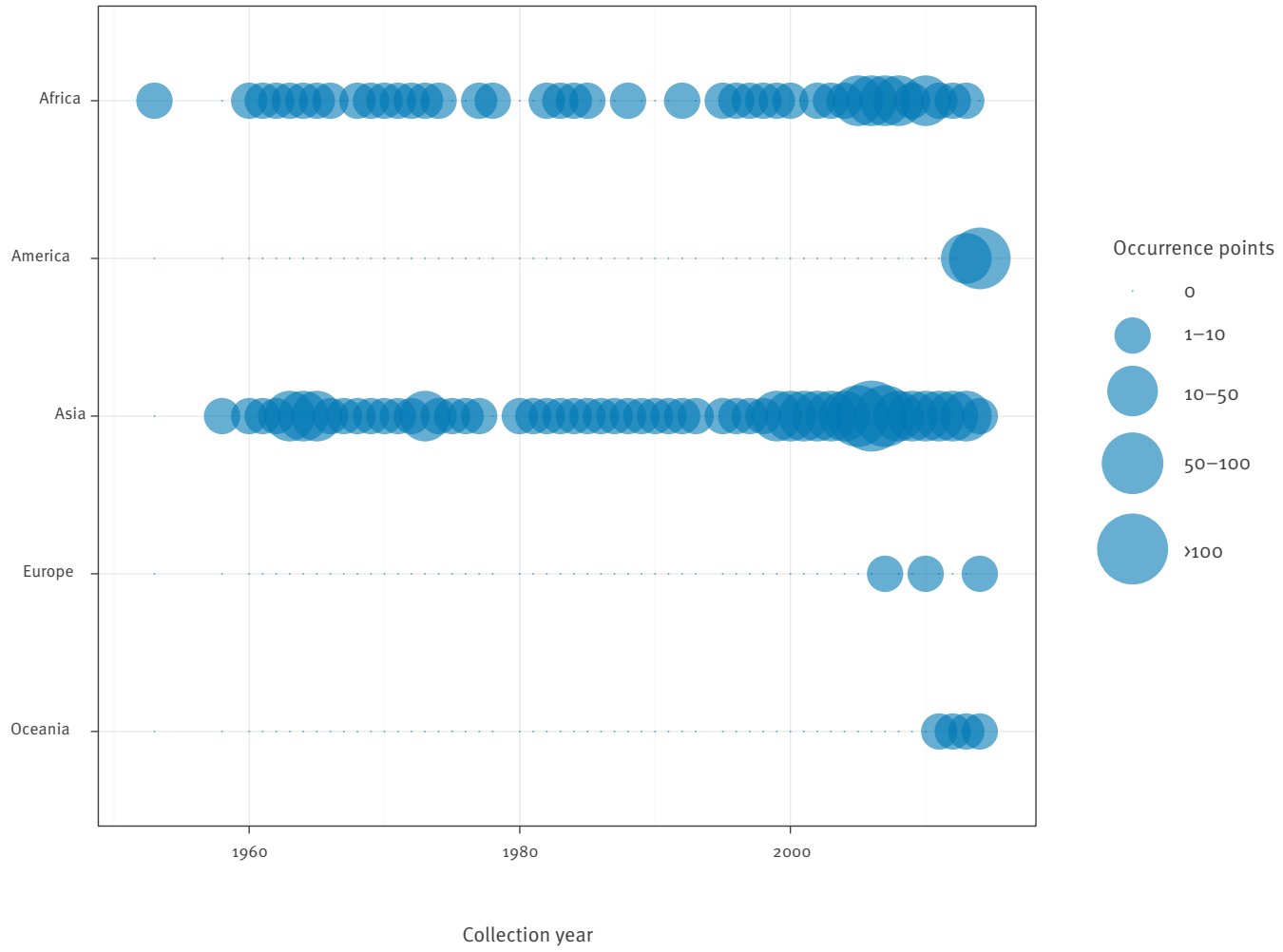


EVI: enhanced vegetation index; LST: land surface temperature.

A. Urban accessibility, with brown representing high relative accessibility to urban areas and yellow representing rural isolated areas; B. Urban, peri-urban and rural areas; C. Aedes aegypti suitability,with red representing high environmental suitability for the vector and blue representing low suitability; D. Ae. albopictus suitability, with red representing high environmental suitability for the vector and blue representing low suitability; E. EVI mean values, whereby dark green represents areas with year round vegetation growth and light green represents areas with low moisture levels; F. EVI range; G. LST mean values, whereby orange represents high temperatures and yellow lower relative temperatures; H. LST range, with orange representing high variation throughout an average year and yellow representing little annual variation.

Figure 2

Distribution per continent of chikungunya virus transmission occurrence points from peer-reviewed evidence and health organisations, 1952–May 2015



Martin Island in the Caribbean. The outbreak subsequently spread to several Caribbean islands and the wider Americas (including many countries in Central and northern South America), reaching over one million suspected and confirmed cases by December 2014 [10] due to the presence of a large naïve population and competent vectors. Despite an increasing global spatial distribution and clinical significance, detailed spatial information of the global geographical extent of CHIKV remains sparse. Previous work has mostly focused on assessment of CHIKV presence at a country level [1]. A comprehensive assessment of CHIKV transmission records globally is critical to identify knowledge gaps and regions where preventive and targeted control efforts should be prioritised. Additionally, this information can be used to guide estimation of the global clinical burden of CHIKV, which is currently unknown. Studies on the prediction of future trends and possibility of transmission in unaffected regions would also benefit from a thorough evaluation of the contemporary distribution and mapping of the extent of CHIKV on a fine spatial scale.

To assess the global distribution of CHIKV, we adopt established disease mapping approaches [11] previously used for dengue [12,13], Zika [14], and the leishmaniases [15]. We developed a comprehensive database of unique locations where CHIKV transmission has been reported and an evidence consensus score – a summary statistic of evidence for CHIKV presence or absence using a weighted scoring system [13]. We then couple these data with a comprehensive set of environmental correlates and contemporary *Aedes* distribution maps in a species distribution model to generate high-resolution risk maps (5km x 5km), representing environmental suitability for CHIKV transmission and to derive population-at-risk estimates. These maps and estimates can be refined as new data on disease and vector range expansion become available.

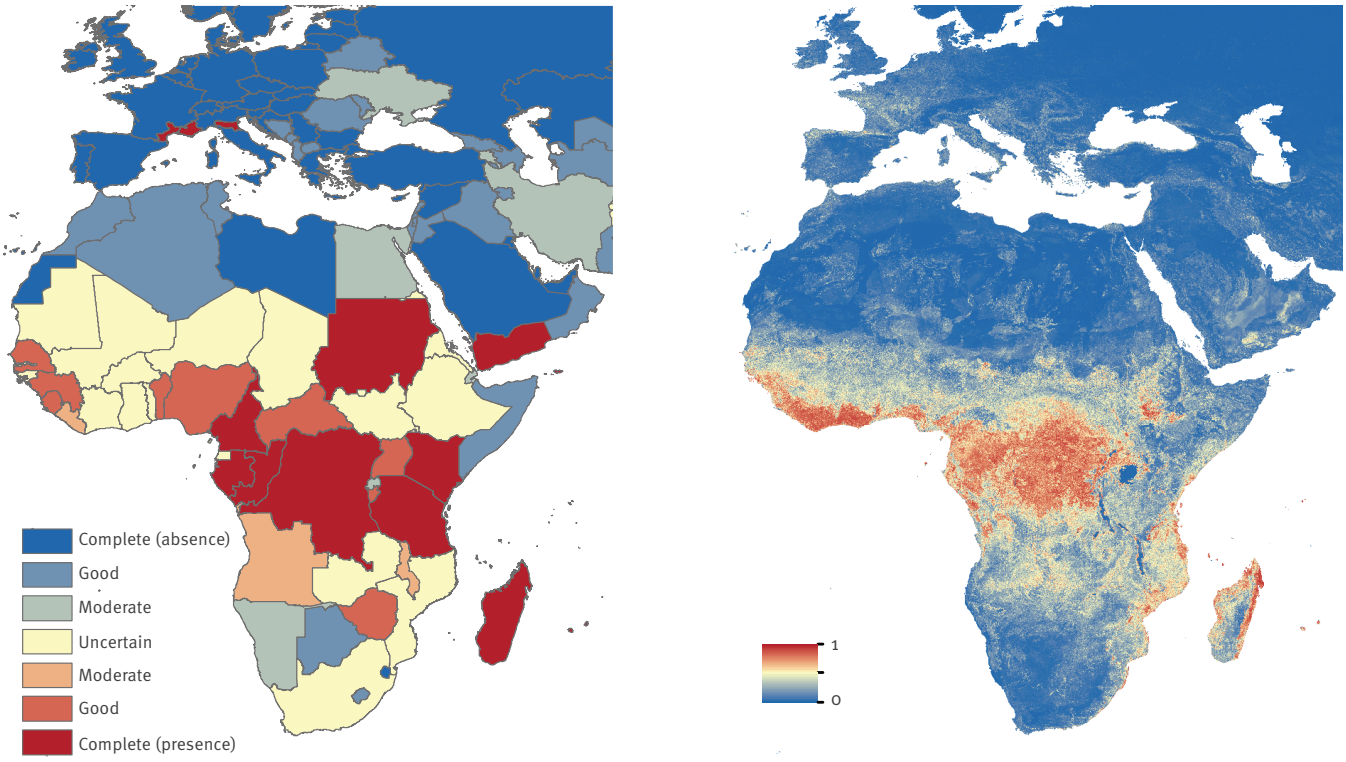
Methods

Occurrence records

Data on CHIKV transmission events were obtained from peer-reviewed literature, health organisations, and supplementary evidence such as, data on vector presence. A comprehensive review of peer-reviewed

Figure 3

Evidence consensus (a) and predicted environmental suitability (b) for chikungunya virus in Africa and Europe, as of May 2015



The colour legend for the evidence consensus ranges from red to blue, representing complete consensus on presence to complete consensus on absence. Indeterminate consensus is in yellow. Evidence consensus is presented at Admin1 level for Italy and France, and Admin0 level for all other countries. The predicted environmental suitability map is presented in similar colours with 1 and 0 representing most and least suitability for chikungunya virus, respectively.

literature was initially conducted on 4 April 2014 in PubMed using the search term 'chikungunya'. The data were subsequently updated every month until 7 May 2015. A total of 2,120 articles were identified. Abstracts of all articles were read to identify those that possibly included geographical information of clinical cases, epidemiological records or outbreaks of CHIKV. Information was extracted on: (i) date of occurrence, (ii) type of event (autochthonous vs imported cases), (iii) number of cases, and (iv) the geographical location (latitude, longitude and administrative level) following a standard protocol [16]. If people were diagnosed after returning from travel but the location of infection could be identified unambiguously due to information on the travel history, the data were retained. Of the 2,120 articles, 339 were used, resulting in 463 unique occurrence points.

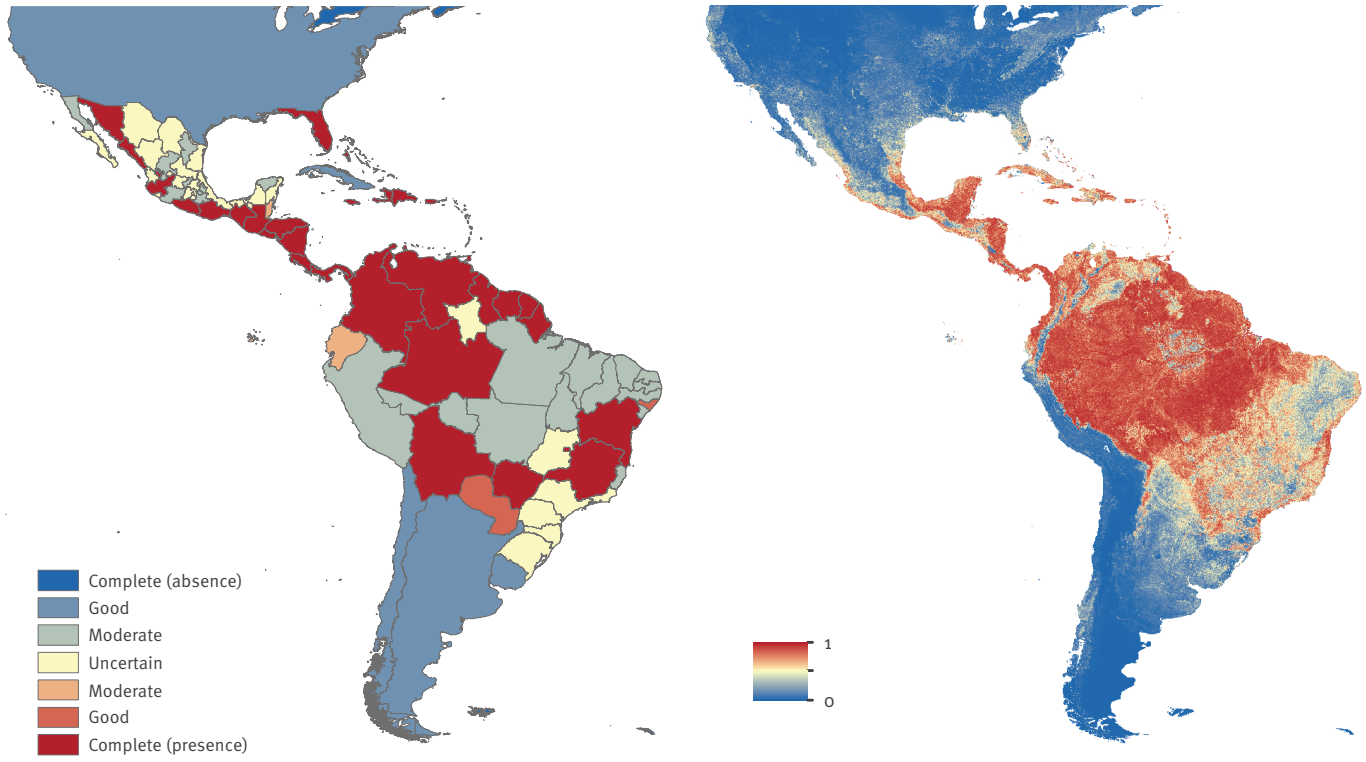
We also consulted health organisation webpages, including the World Health Organization (WHO), Pan American Health Organization (PAHO), Global Infectious Disease and Epidemiology Network (GIDEON) and the United States (US) Centers for Disease Control and Prevention (CDC). Data on recent and ongoing outbreaks in the Americas were obtained from weekly PAHO epidemiological reports. Since PAHO is a WHO affiliate organisation, PAHO and WHO data were combined to represent a single health organisation data source resulting in 65 unique occurrence points.

Finally, 600 occurrence points based on reports of CHIKV outbreaks were obtained from the online surveillance system, HealthMap (http://www.healthmap.org/en).

Each location was then classified as precise (e.g. a town) or an administrative unit and the latter were linked to the appropriate polygon information from the Global Administrative Unit Layers dataset (http://www.gadm.org). Each occurrence point was attributed to a specific location (latitude and longitude) based on the level of detail presented in the data source. We used Google Maps (https://www.google.co.uk/maps) to support the geopositioning. The database subsequently underwent temporal and spatial standardisation using a standard protocol described elsewhere [16]. We removed any duplicate records so that the resulting database contained only one occurrence record per spatial entity. Further, any records reported from the above mentioned data sources that fell outside land borders (e.g. due to wrong geopositioning of the original source) were removed. Occurrence records

Figure 4

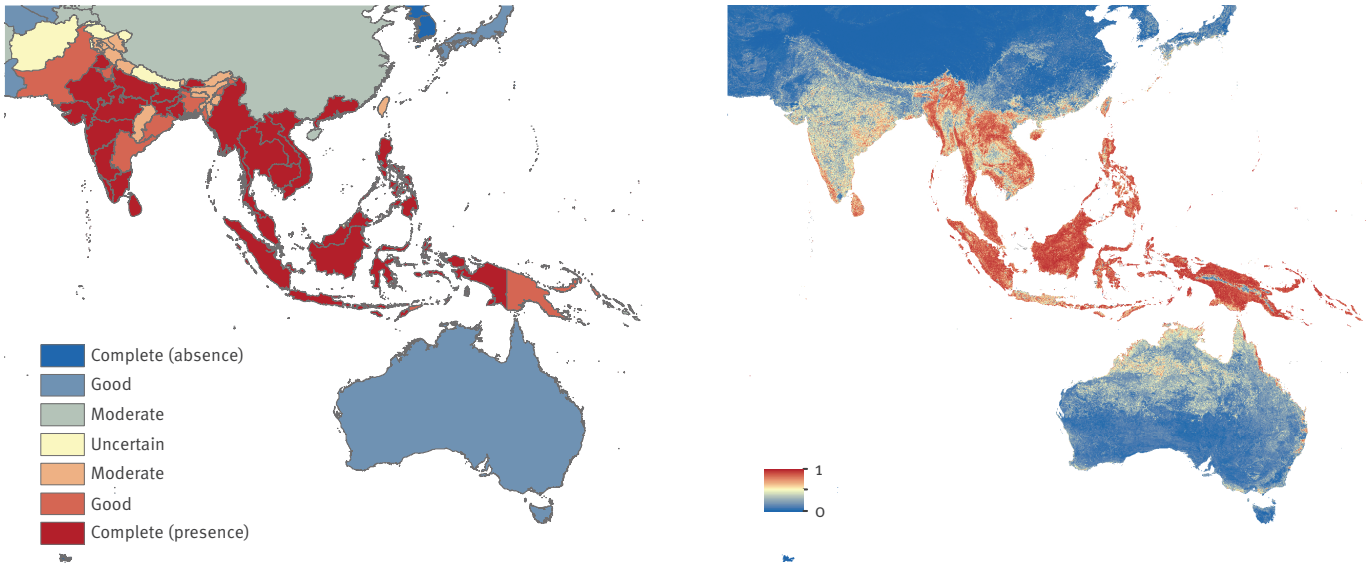
Evidence consensus (a) and predicted environmental suitability (b) for chikungunya virus in the Americas, as of May 2015



The colour legend for the evidence consensus ranges from red to blue representing complete consensus on presence to complete consensus on absence. Indeterminate consensus is in yellow. Evidence consensus is presented at Admin1 level for United States, Mexico, Argentina and Brazil, and Admino level for all other countries. The predicted environmental suitability map is presented in similar colours with 1 and 0 representing most and least suitability for chikungunya virus, respectively.

Figure 5

Evidence consensus (a) and predicted environmental suitability (b) for chikungunya virus in Asia and Oceania, as of May 2015



The colour legend for the evidence consensus ranges from red to blue representing complete consensus on presence to complete consensus on absence. Indeterminate consensus is in yellow. Evidence consensus is presented at Admin1 level for India and China, and country level for all other countries. The predicted environmental suitability map is presented in similar colours with 1 and 0 representing most and least suitability for chikungunya virus, respectively.

Table 1

Summary of evidence used for chikungunya virus transmission consensus scoring

Evidence category	Score
Health organisation status	
Number of health organisations[a] reporting epidemics or local transmission of chikungunya virus	
3	+6/-6
2	+3/-3
Peer reviewed evidence	
Date of chikungunya virus transmission occurrence	
2005–2015	3
1997–2004	2
Pre-1997	1
Diagnostic procedure	
Polymerase chain reaction (PCR)	3
IgM/IgG-based enzyme-linked immunosorbent assay (ELISA) and haemaglutination inhibition	2
No specified test	1
Combination score[b]	
3 types[b]	3
2 types[b]	2
1 types[b]	1
Case data or health expenditure	
Case data: most recent outbreak	
0–7 years	9
7–14 years	6
14–21 years	3
28–35 years	-3
≥35 years	-6
If no case data: health expenditure in 2014	
HE <100 USD + sporadic cases	6
HE <100 USD + sporadic cases	3
100 USD ≤HE <500 USD	-3
HE ≥500 USD	-9
Supplementary evidence[c]	
Number of evidence types[c]	
4 types[c]	6
3 types[c]	4
2 types[c]	2

[a] The three health organisations considered included (i) the United States Centers for Disease Control and Prevention (CDC), (ii) the Global Infectious Disease and Epidemiology Network (GIDEON) and (iii) the World Health Organization (WHO) and Pan American Health Organization (PAHO) taken together.

[b] The types of evidences include (i) reports from health organisations, (ii) the date of chikungunya virus transmission occurrence, and (iii) the diagnostic procedure.

[c] The types of evidences include (i) mosquito presence, (ii) dengue presence, (iii) travel advisories, and (iv) HealthMap.

were then cross-checked manually and for each occurrence point, the probability of occurrence for both vectors was extracted (Figure 1c and d) and records deemed implausible (probability of occurrence for both vectors individually fell below 70%) [17] were removed. The cut-off at 70% was chosen based on a qualitative interpretation of the *Ae. aegypti* and *Ae. albopictus* probability of occurrence maps, where >90% of the geographical spaces with non-zero probabilities were recorded between vector presence 70% to 100%.

Evidence consensus

The evidence consensus is a summary statistic representing confidence in the presence or absence of CHIKV for a given political region. We calculated evidence consensus scores at a national level for all countries apart from Argentina, Brazil, China, France, India, Italy, Mexico, and the US, where sufficient detailed epidemiological information was available to distinguish within-country epidemiological, environmental, and economic variability. Smaller countries were not considered for subnational assessment.

Table 2

Relative contribution of each covariate in explaining the global environmental suitability of chikungunya virus and 95% confidence interval, 1952–May 2015

Variable	Mean (95% confidence interval)
Aedes aegypti	46.16 (43.45–48.74)
EVI mean and range	33.37 (29.71–36.94)
Urban accessibility	7.42 (6.52–8.50)
Ae. albopictus	6.70 (5.78–7.60)
LST mean and range	5.98 (4.83–6.90)
Urbanisation	0.30 (0.15–0.70)

EVI: enhanced vegetation index; LST: land surface temperature.

Each evidence category was scored independently and category weights applied to reflect the level of detail each category provides: health organisation status (max. score 6), peer-reviewed evidence (max. 9), case data (max. 9) and supplementary evidence (max. 6). For example, if all three health organisations considered (i.e. PAHO/WHO, GIDEON and CDC) agreed on presence/absence of CHIKV, the country was assigned a score of +6/-6. If two health organisations agreed on presence/absence, the country was assigned a score of +3/-3. For each country, we identified peer-reviewed evidence to confirm CHIKV presence based on seroprevalence studies, polymerase chain reaction (PCR) confirmation, case reports and local outbreaks. Each published study was scored independently based on date of disease occurrence and the average was taken and added to a score of accuracy, which was assigned based on diagnostic procedure. The case data category was scored based on the most recent outbreak year and pre-defined inter-epidemic period intervals to account for herd immunity [18,19]. In the absence of case data, healthcare expenditure (HE) was used as a proxy to indicate confidence in a country's ability to detect and report CHIKV cases and to supplement this category [13]. In addition, countries for which HE data were unavailable were assigned low HE, and overseas territories were assigned the same HE as their parent nations.

For areas with conflicting reports from the input sources and recent geographical expansion (e.g. Mexico and Brazil), database records were cross-checked against expert knowledge and recent national reports. Supplementary evidence was extracted from peer-reviewed literature on mosquito presence, dengue presence, travel advisories (http://www.nathnac.org/travel), and the HealthMap database (http://www.healthmap.org) and scored based on number of evidence types. The HealthMap database contains data extracted from online news sources and alerts from official public health sources (e.g. ministries of public health) and was available from 2006 to 2014. We also used data on the global evidence consensus on dengue transmission from Brady et al. [13] to supplement data on the presence of other arboviruses for each country. See Table 1 for a summary and Brady et al. [13] for additional details on the scoring process.

The scores from all previously described evidence categories were summed, divided by the maximum potential score, 30 and multiplied by 100. The evidence consensus was then mapped based on seven equidistant categories: complete evidence for presence (›71.42 to ≤100) or absence (›-100 to ≤-71.42), good evidence for presence (›42.84 to ≤71.42) or absence (›-71.42 to ≤-42.84], moderate evidence for presence (›14.27 to ≤42.84) or absence (›-42.84 to ≤-14.27), and indeterminate evidence (›-14.27 to ≤14.27).

Environmental risk maps

A boosted regression tree (BRT) modelling approach was applied to derive probabilistic global environmental risk maps for CHIKV. BRT models combine the strengths of regression trees with boosting and are frequently used in species distribution modelling (SDM) [20] with the overall aim to identify areas that exhibit similar conditions to areas where the disease has been reported. These statistical models heavily depend on the input occurrence database. We included the following globally available environmental and socioeconomic covariates in our models: (i) probabilistic global niche maps for *Ae. aegypti* and *Ae. albopictus* [17]; (ii) land surface temperature (LST); (iii) a categorical differentiation in urban, peri-urban and rural environments, as well as their accessibility; and (iv) enhanced vegetation index (EVI).

Aedes aegypti and *Aedes albopictus* suitability

We constructed probabilistic global *Aedes* risk maps based on the most comprehensive globally representative occurrence dataset for both *Ae. aegypti* and *Ae. albopictus* species and a set of contemporary environmental covariates explaining their global distribution (Figure 1c and d) [17,21].

Land surface temperature (LST)

Temperature is one of the most important determinants of *Aedes* survival [22]. We used MODIS daytime LST after processing through a gap-filling algorithm to account for variation in temperature globally [23] (Figure 1g and h). Additionally, transmission of CHIKV is dependent on the persistent presence of the disease vectors, *Ae. aegypti* and *Ae. albopictus* mosquitoes. Both adult female longevity and length of first gonotrophic cycle are temperature dependent. We used the combined temperature suitability of population persistence for both species to identify areas where by proxy we expect the geographical limits of transmission of CHIKV to be located [13,15]. The binary outputs of this model were used as a mask to exclude the placement of 10,000 pseudo-absences inside their physiologically plausible range.

Enhanced vegetation index (EVI)

Water availability is another important factor for CHIKV transmission, as it has been shown, for example, that mosquitoes may cluster around houses if containers with standing water due to precipitation are present [24]. EVI measures vegetation canopy greenness and can be used as a proxy for soil surface-level moisture that are associated with the presence of vector larval development sites [25]. We used range and mean values of MODIS EVI after processing through a gap-filling algorithm described in Weiss et al. [23] (Figure 1e and f).

Urbanisation

To account for differences in urban, peri-urban and rural environments we built a categorical variable by supplementing the projected 2010 Global Rural Urban Mapping Project (GRUMP) urban rural categories with land-cover classes using night-time light satellite imagery and population density based on the most up-to-date national censuses available at the smallest available administrative unit [26]. A gridded surface of 5 km x 5 km cells was generated with each pixel representing urban, peri-urban, or rural areas (Figure 1b).

Urban accessibility

Human movement defines the spatial scale of transmission of most communicable diseases. To account for areas that are highly accessible, we used a friction surface that shows pixel level travel time through the respective region of interest [27] (Figure 1a).

We then applied an ensemble BRT using 180 submodels to derive probabilistic estimates and underlying uncertainty at a 5 km x 5 km resolution. We also estimated populations-at-risk of CHIKV transmission. For each occurrence record we identified the maximum risk probability of transmission within a buffer of 10 km around the occurrence record to derive a threshold value of probability of occurrence that should be included in the population-at-risk estimate. We then converted the continuous surface of transmission risk into a binary at-risk/not-at-risk classification. Finally, we extracted the population living in areas of potential transmission using a global population surface.

Results

The data abstraction yielded 528 occurrence points from peer-reviewed evidence and health organisations. The distribution of occurrence points by region is shown in Figure 2. Individually, WHO (+PAHO), CDC and GIDEON implicated 60, 96, and 70 countries with reported CHIKV transmission, respectively. However, based on evidence from all available sources, we identified 94 countries with good or better evidence consensus on CHIKV presence from which 47% were in Asia and Africa and 44% in the Americas. Of the 94 countries, at total of 24, 41, 20, two and seven were located in Africa, the Americas, Asia, Europe and Oceania, respectively.

High probability of *Ae. aegypti* occurrence was the strongest predictor for environmental suitability with a mean contribution of 46% (Table 2). Based on our environmental suitability predictions, we estimate that 1.3 billion people are living in areas of potential CHIKV transmission.

Global consensus on CHIKV presence or absence based on the evidence consensus scoring system and predictions of environmental suitability for CHIKV are further presented in for Africa and Europe, the Americas, and Asia and Oceania, respectively. The evidence ranges from complete presence to complete absence.

Africa and Europe

Most of the African countries with reliable evidence on presence and absence were located in Central and North Africa, respectively. The strong evidence for CHIKV presence in Central Africa was aided by in-country seroprevalence studies such as in Cameroon and Gabon and confirmed imported cases to non-endemic regions [28,29]. There was insufficient data to determine presence or absence for some countries in Africa such as Ghana, Ethiopia, and Mozambique, due to a general lack of data and poor surveillance. South Africa, for example, had a reported outbreak in 1977 [30], but there have been no new reports of outbreaks or seroprevalence studies. Most countries in Central and parts of West Africa with strong evidence on CHIKV presence also demonstrated high environmental suitability for CHIKV (Figure 3b). We estimated that over 240 million people live in areas at risk for CHIKV transmission in Africa.

Autochthonous transmission of CHIKV in Europe has been reported in Ravenna, northern Italy in 2007 and in the south-eastern French city of Fréjus in 2010 and Montpellier, southern France in 2014 (Figure 3a). Prior to these cases, CHIKV in Europe was limited to travel-related cases. Genomic micro-evolution of CHIKV in Reunion facilitated transmission by *Ae. albopictus* [31] in strains belonging to the so called Indian Ocean lineage (IOL) which increased the likelihood of CHIKV outbreaks in Europe. We predicted moderate environmental suitability for parts of France, Spain, Germany and Italy (Figure 3b), which is in agreement with previous projections of increased risk of CHIKV for France, northern Italy and the Pannonian Basin [32].

Americas

Systematic reporting of confirmed cases to PAHO provided well-documented evidence for CHIKV occurrence in the Americas. As of 13 March 2015, 45 countries had reported local transmission of CHIKV, while the evidence consensus identified 90% of countries and island states with good or better consensus (Figure 4a). Laboratory confirmation of autochthonous cases made a major contribution in defining the South and Central America's higher consensus cluster on presence. The evidence strength varied across states for Brazil and Mexico. Brazil had strong evidence on presence in

south-central states; indeterminate evidence for several southern states and Amazonas state in the north; and moderate evidence in central and northern states. In contrast, the environmental suitability map predicted that coastal Brazil and areas in the Amazon were highly suitable for CHIKV transmission. Peru and states in Brazil where evidence currently suggest absence, demonstrated a moderate to high suitability for CHIKV (Figure 4b).

Our models identified *Ae. aegypti* as the strongest predictor of CHIKV suitability, which is the primary vector implicated in the outbreak in most of the Caribbean and the Americas [33,34]. Based on the environmental suitability predictions, we estimated that over 260 million people live in areas at risk for CHIKV transmission in the Americas.

Asia and Oceania

CHIKV presence is well documented for most countries and islands in south and south-east Asia. Persistent vector populations and naïve populations contributed to notable pre-2000 CHIKV outbreaks in India, Myanmar/Burma, Philippines, Sri Lanka, Thailand and Vietnam [35,36]. In the last decade, outbreaks have been reported in India, Indian Ocean islands, Indonesia, Sri Lanka, and Thailand, following a 41-year period of quiescence [37,38]. Good to better consensus on CHIKV presence was reported for 80% of Asian countries and clusters of complete consensus were observed ranging from Myanmar/Burma to Vietnam for south Asia and Indonesia to Papua New Guinea for south-east Asian islands (Figure 5a). Complete consensus on CHIKV presence was noted for Guangdong province, China, where an outbreak was reported in 2010 [39]. With the exception of Chhattisgarh in Central India and smaller states in the north-east and north-west, most Indian states had good consensus on CHIKV presence. There was insufficient data on presence or absence of CHIKV in Nepal. Despite reports of large outbreaks in neighbouring India and indication of possible local transmission, there are no published national seroprevalence surveys for Nepal and CHIKV cases are possibly misdiagnosed as typhoid or other bacterial or viral fevers [40].

The environmental suitability map strongly complemented the evidence consensus with high risk of CHIKV transmission predicted for most south-east Asia islands and parts of south-east Asia, and moderate risk predicted for India (Figure 5b). Predictions of suitability for CHIKV transmission in China were patchy, with high risk predicted for regions where local CHIKV transmission has been observed. High prevalence has been noted for other arboviral diseases in south-east coastal regions, and northern and inland regions [41,42], signalling suitability and risk for CHIKV transmission. We estimated that over 270 million people live in areas at risk for CHIKV in Asia.

Discussion

We present the global distribution and predict environmental suitability for CHIKV transmission. In this analysis we combined extensive evidence available at multiple geographical scales and integrated them into a well-established disease mapping approach. We identified areas particularly in Africa where CHIKV status is uncertain (e.g. Ethiopia, Mozambique) although there is high environmental suitability, presence of disease vectors and reported outbreaks in neighbouring countries such as Tanzania. We also predicted high environmental suitability for some regions but observed moderate evidence for absence of local transmission, such as in the Amazon basin. The former could be attributable to inadequate surveillance or misdiagnosis, while the latter could be explained by the recent emergence of CHIKV and the endemicity of dengue in these regions. Our comprehensive assessment of these evidence gaps enables the identification of areas where surveillance, seroprevalence studies, and vector control measures are increasingly important [43].

Despite the growing economic and public health impact of CHIKV outbreaks, there have not been any thorough assessments of knowledge gaps or high-resolution distribution maps of the disease. This is the first comprehensive assessment of the current spatial extent and environmental suitability of CHIKV transmission globally. We considered extensive available evidence from 1952 when CHIKV was first identified in Tanzania. We observed the following: (i) CHIKV evidence is dynamic, re-emerging and due to its high attack rate likely to cause outbreaks with significant timelags; (ii) misclassification and delay in reporting may influence the perception of current vs. former presence; (iii) while there is strong evidence for some areas (e.g. Asia), data on CHIKV occurrence in many parts of Africa remain sparse.

Similar to dengue [44], CHIK appears to be an underrecognised problem in Africa. Our environmental suitability maps indicate moderate to strong risk for CHIKV transmission in most of West, Central and East African countries. However, there is insufficient data to determine presence or absence for several countries in West Africa and parts of East Africa. Seroprevalence surveys are needed for countries with indeterminate status.

Additionally, co-occurrence and comorbidity of CHIKV and other arboviruses have been reported, which stresses the importance for better diagnostic tests. Furthermore, since dengue and CHIKV share the same primary vectors, similar estimates for population-at-risk would be expected. Population at risk of dengue virus transmission has been estimated at 3.97 billion people [13], which is far more than the 1.3 billion predicted for CHIKV. However, the estimates for dengue were derived on a national level, while our estimates are at a 5 km x 5 km spatial scale.

Improvements, such as the identification of the primary vector for each occurrence record can be made as new data become available. Particularly in Africa and Micronesia it is hypothesised that other mosquito vectors (*Ae. afticanus, Ae. luteocephalus, Ae. furcifer-taylori* and *Ae. hensilli*) are responsible for transmitting CHIKV between humans [45]. Data on their geographical distribution however remains sparse and could not be integrated in the current analysis. Furthermore, since the maps are derived using long-term average covariate layers, they identify areas that are at risk of transmission at any time during the year but do not consider explicitly how seasonality or human mobility affects transmission; these factors influence the short and long-term dynamics of CHIKV transmission. In addition, there are a number of other factors that may influence the probability of infection that relate to social aspects, such as, housing quality, poverty, susceptibility of the population, and the use of air-conditioning. Such factors were not included in our study since no reliable data are available on a global scale.

Furthermore, there are some challenges in developing a global model to predict environmental suitability for CHIKV transmission. For instance, although there have been reports of major outbreaks of CHIKV in India, the environmental suitability models predicted moderate environmental suitability for CHIKV in these regions. The spatial variability observed for India is similar to the *Ae. albopictus* distribution map [17]. This suggests that our modelling approach strongly mirrors *Ae. albopictus* distribution in India, which has been implicated in recent outbreaks, but fails to capture the distribution of other covariates. Similarly, in Europe where there have been documented outbreaks of CHIKV in France, and Italy, risk appears to be low in our prediction. Modelling the likely distribution of CHIKV based on *Ae. albopictus* which is well distributed in southern Europe could help refine our approach. However, as transmission of CHIKV is relatively low in Europe, our results indicate the likely lower relative global contribution of *Ae. albopictus* compared with *Ae. aegypti*. Additionally, our maps indicate high transmission risk in small island states that have experienced large outbreaks in the last decade. However, herd immunity was not incorporated into the environmental risk models, implying that small countries or islands with (›50% attack rates) might have a lower risk of CHIKV outbreaks despite environmental suitability. These observations highlight the need for region or country-specific modelling, which would be most beneficial for local public health policy decisions. These maps can however be used as baseline layers for future projections of CHIKV and compared with global maps of dengue [12].

Despite these limitations, the usefulness of these maps cannot be understated, especially given the CHIKV's currently expanding range in the Americas and elsewhere. By synchronising the evidence consensus, environmental suitability prediction maps and population at-risk estimates, areas most in need of CHIKV surveillance resources were identified. It is equally important that these maps are updated as soon as new information becomes available, such as recent reports of local CHIKV transmission in states of Brazil, and Peru and this will be achieved by incorporating the results of this study into the Atlas of Baseline Risk Assessment for Infectious Diseases (ABRAID, at www.abraid.ox.ac.uk).

Acknowledgements

We thank Maria Devine for proofreading.

EON is supported by funding from the National Institute of Environmental Health Sciences of the National Institutes of Health (#K01ES025438). MUGK receives funding from the International research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS; European Commission 7th Framework Programme (21893)). NG is supported by a University of Melbourne McKenzie fellowship. DMP is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford. OJB is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). CLM is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1093011). MAJ is supported by the Models of Infectious Disease Agent Study program (cooperative agreement #1U54GM088558). PWG is a Career Development Fellow (#K00669X) jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DfID) under the MRC/DfID Concordat agreement and receives support from the Bill & Melinda Gates Foundation (#OPP1068048, #OPP1106023). KK is funded by the Canadian Institutes for Health Research. SIH received a grant from the Research for Health in Humanitarian Crises (R2HC) Programme, managed by ELRHA (#13468) which also supported NG & MUGK. The Research for Health in Humanitarian Crises (R2HC) programme aims to improve health outcomes by strengthening the evidence base for public health interventions in humanitarian crises. Visit www.elrha.org/work/r2hc for more information. The £8 million R2HC programme is funded equally by the Wellcome Trust and DFID, with Enhancing Learning and Research for Humanitarian Assistance (ELRHA) overseeing the programme's execution and management. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), and grants from the Bill & Melinda Gates Foundation (OPP1119467, OPP1093011, OPP1106023 and OPP1132415). JSB acknowledges funding from NIH National Library of Medicine (#R01LM010812-05) and from the Bill & Melinda Gates Foundation (#OPP1093011). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Conflict of interest

None declared.

Authors' contributions

Elaine Nsoesie and Moritz Kraemer performed data collection and analysis, and drafted the manuscript. Catherine Moyes reviewed data for accuracy. Michael Johansson served as the expert reviewer. Nick Golding, Oliver Brady and David Pigott advised on methods. John Brownstein and Simon Hay provided resources. All authors contributed to the writing and editing of the manuscript.

References

1. Weaver SC, Lecuit M. Chikungunya virus and the global spread of a mosquito-borne disease.N Engl J Med. 2015;372(13):1231-9. DOI: 10.1056/NEJMra1406035 PMID: 25806915
2. Ross RW. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic.J Hyg (Lond). 1956;54(2):177-91. DOI: 10.1017/S0022172400044442 PMID: 13346078
3. Powers AM, Logue CH. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus.J Gen Virol. 2007;88(Pt 9):2363-77. DOI: 10.1099/vir.0.82858-0 PMID: 17698645
4. Weaver SC. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health.PLoS Negl Trop Dis. 2014;8(6):e2921. DOI: 10.1371/journal.pntd.0002921 PMID: 24967777
5. Volk SM, Chen R, Tsetsarkin KA, Adams AP, Garcia TI, Sall AA, et al. Genome-scale phylogenetic analyses of chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. J Virol. 2010;84(13):6497-504. DOI: 10.1128/JVI.01603-09 PMID: 20410280
6. Staples JE, Breiman RF, Powers AM. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. Clin Infect Dis. 2009;49(6):942-8. DOI: 10.1086/605496 PMID: 19663604
7. Chahar HS, Bharaj P, Dar L, Guleria R, Kabra SK, Broor S. Co-infections with chikungunya virus and dengue virus in Delhi, India.Emerg Infect Dis. 2009;15(7):1077-80. DOI: 10.3201/eid1507.080638 PMID: 19624923
8. Weaver SC. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health.PLoS Negl Trop Dis. 2014;8(6):e2921. DOI: 10.1371/journal.pntd.0002921 PMID: 24967777
9. Chang LJ, Dowd KA, Mendoza FH, Saunders JG, Sitar S, Plummer SH, et al. , VRC 311 Study Team. Safety and tolerability of chikungunya virus-like particle vaccine in healthy adults: a phase 1 dose-escalation trial.Lancet. 2014;384(9959):2046-52. DOI: 10.1016/S0140-6736(14)61185-5 PMID: 25132507
10. Johansson MA. Chikungunya on the move.Trends Parasitol. 2015;31(2):43-5. DOI: 10.1016/j.pt.2014.12.008 PMID: 25649340
11. Kraemer MUG, Hay SI, Pigott DM, Smith DL, Wint GRW, Golding N. Progress and Challenges in Infectious Disease Cartography.Trends Parasitol. 2016;32(1):19-29. DOI: 10.1016/j.pt.2015.09.006 PMID: 26604163
12. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504-7. DOI: 10.1038/nature12060 PMID: 23563266
13. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012;6(8):e1760. DOI: http://dx.doi.org/10.1371/journal.pntd.0001760
14. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. eLife. 2016;5:e15272. DOI: 10.7554/eLife.15272 PMID: 27090089
15. Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, et al. Global distribution maps of the leishmaniases. eLife. 2014;3. DOI: 10.7554/eLife.02851 PMID: 24972829
16. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Sci Data. 2014;1:140004. DOI: 10.1038/sdata.2014.4 PMID: 25977762
17. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. eLife. 2015;4:e08347. DOI: 10.7554/eLife.08347 PMID: 26126267
18. Gérardin P, Guernier V, Perrau J, Fianu A, Le Roux K, Grivard P, et al. Estimating Chikungunya prevalence in La Réunion Island outbreak by serosurveys: two methods for two critical times of the epidemic. BMC Infect Dis. 2008;8(1):99. DOI: 10.1186/1471-2334-8-99 PMID: 18662384
19. Nitatpattana N, Kanjanopas K, Yoksan S, Satimai W, Vongba N, Langdatsuwan S, et al. Long-term persistence of Chikungunya virus neutralizing antibodies in human populations of North Eastern Thailand. Virol J. 2014;11(1):183. DOI: 10.1186/1743-422X-11-183 PMID: 25330992
20. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl. 2009;19(1):181-97. DOI: 10.1890/07-2153.1 PMID: 19323182
21. Kraemer MU, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, et al. The global compendium of Aedes aegypti and Ae. albopictus occurrence. Sci Data. 2015;2:150035. DOI: 10.1038/sdata.2015.35 PMID: 26175912
22. Brady OJ, Golding N, Pigott DM, Kraemer MU, Messina JP, Reiner RC, et al. Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission. Parasit Vectors. 2014;7(1):338. DOI: 10.1186/1756-3305-7-338 PMID: 25052008
23. Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. An effective approach for gap-filling continental scale remotely sensed time-series.ISPRS J Photogramm Remote Sens. 2014;98:106-18. DOI: 10.1016/j.isprsjprs.2014.10.001 PMID: 25642100
24. Luz C, Tai MHH, Santos AH, Silva HHG. Impact of moisture on survival of Aedes aegypti eggs and ovicidal activity of Metarhizium anisopliae under laboratory conditions.Mem Inst Oswaldo Cruz. 2008;103(2):214-5. DOI: 10.1590/S0074-02762008000200016 PMID: 18425277
25. Estallo EL, Lamfri MA, Scavuzzo CM, Almeida FF, Introini MV, Zaidenberg M, et al. Models for predicting Aedes aegypti larval indices based on satellite images and climatic variables. J Am Mosq Control Assoc. 2008;24(3):368-76. DOI: 10.2987/5705.1 PMID: 18939688
26. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. Determining global population distribution: methods, applications and data.Adv Parasitol. 2006;62:119-56. DOI: 10.1016/S0065-308X(05)62004-0 PMID: 16647969
27. Uchida H, Nelson A. Agglomeration index: Towards a New Measure of Urban Concentration. Background paper for the World Bank's World Development Report. Washington D.C.: World Bank; 2009.
28. Demanou M, Antonio-Nkondjio C, Ngapana E, Rousset D, Paupy C, Manuguerra J-C, et al. Chikungunya outbreak in a rural area of Western Cameroon in 2006: A retrospective serological and entomological survey. BMC Res Notes. 2010;3(1):128. DOI: 10.1186/1756-0500-3-128 PMID: 20444282
29. Peyrefitte CN, Bessaud M, Pastorino BAM, Gravier P, Plumet S, Merle OL, et al. Circulation of Chikungunya virus in Gabon, 2006-2007. J Med Virol. 2008;80(3):430-3. DOI: 10.1002/jmv.21090 PMID: 18205212
30. Sudeep AB, Parashar D. Chikungunya: an overview.J Biosci. 2008;33(4):443-9. DOI: 10.1007/s12038-008-0063-2 PMID: 19208970
31. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney M-C, et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006;3(7):e263. DOI: 10.1371/journal.pmed.0030263 PMID: 16700631
32. Fischer D, Thomas SM, Suk JE, Sudre B, Hess A, Tjaden NB, et al. Climate change effects on Chikungunya transmission in Europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. Int J Health Geogr. 2013;12(1):51. DOI: 10.1186/1476-072X-12-51 PMID: 24219507
33. Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. Chikungunya in the Americas.Lancet. 2014;383(9916):514. DOI: 10.1016/S0140-6736(14)60185-9 PMID: 24506907
34. Van Bortel W, Dorleans F, Rosine J, Blateau A, Rousset D, Matheus S, et al. Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. Euro Surveill. 2014;19(13):20759. DOI: 10.2807/1560-7917.ES2014.19.13.20759 PMID: 24721539
35. Pulmanausahakul R, Roytrakul S, Auewarakul P, Smith DR. Chikungunya in Southeast Asia: understanding the emergence and finding solutions.Int J Infect Dis. 2011;15(10):e671-6. DOI: 10.1016/j.ijid.2011.06.002 PMID: 21775183
36. Lanciotti RS, Valadere AM, . Transcontinental movement of Asian genotype chikungunya virus. Emerg Infect Dis. 2014;20(8):1400-2. DOI: 10.3201/eid2008.140268 PMID: 25076384
37. Azami NAM, Salleh SA, Shah SA, Neoh HM, Othman Z, Zakaria SZS, et al. Emergence of chikungunya seropositivity in healthy Malaysian adults residing in outbreak-free locations: chikungunya seroprevalence results from the Malaysian Cohort. BMC Infect Dis. 2013;13(1):67. DOI: 10.1186/1471-2334-13-67 PMID: 23379541
38. Pulmanausahakul R, Roytrakul S, Auewarakul P, Smith DR. Chikungunya in Southeast Asia: understanding the emergence and finding solutions.Int J Infect Dis. 2011;15(10):e671-6. DOI: 10.1016/j.ijid.2011.06.002 PMID: 21775183
39. Wu D, Wu J, Zhang Q, Zhong H, Ke C, Deng X, et al. Chikungunya outbreak in Guangdong Province, China, 2010. Emerg Infect Dis. 2012;18(3):493-5. DOI: 10.3201/eid1803.110034 PMID: 22377135

40. Pun SB, Bastola A, Shah R. First report of Chikungunya virus infection in Nepal.J Infect Dev Ctries. 2014;8(6):790-2. DOI: 10.3855/jidc.3701 PMID: 24916880
41. Wang J, Zhang H, Sun X, Fu S, Wang H, Feng Y, et al. Distribution of mosquitoes and mosquito-borne arboviruses in Yunnan Province near the China-Myanmar-Laos border. Am J Trop Med Hyg. 2011;84(5):738-46. DOI: 10.4269/ajtmh.2011.10-0294 PMID: 21540383
42. Wu J-Y, Lun Z-R, James AA, Chen X-G. Dengue Fever in mainland China.Am J Trop Med Hyg. 2010;83(3):664-71. DOI: 10.4269/ajtmh.2010.09-0755 PMID: 20810836
43. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med. 2015;13(102):102. DOI: 10.1186/s12916-015-0348-x PMID: 25976325
44. Jaenisch T, Junghanss T, Wills B, Brady OJ, Eckerle I, Farlow A, et al. , Dengue in Africa Study Group. Dengue expansion in Africa-not recognized or not happening?Emerg Infect Dis. 2014;20(10):e140487. DOI: 10.3201/eid2010.140487 PMID: 25271370
45. Ledermann JP, Guillaumot L, Yug L, Saweyog SC, Tided M, Machieng P, et al. Aedes hensilli as a potential vector of Chikungunya and Zika viruses. PLoS Negl Trop Dis. 2014;8(10):e3188. DOI: 10.1371/journal.pntd.0003188 PMID: 25299181

License and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made.

This article is copyright of the European Centre for Disease Prevention and Control, 2016.

Chapter 3 – The introduction, suitability, and spread of Zika virus

---

To evaluate the dynamics of Zika virus (ZIKV) spread globally this chapter draws upon Aedes *suitability* surfaces from chapter 1 and derives estimates of ZIKV introduction potential globally in each month of the year separately. It consists of two publications; one evaluating the ecological landscape of potential transmission per month based on *Aedes* maps; and a second one that discusses the specific ZIKV average mean suitability. Both publications are included below in final form as published in *The Lancet Infectious Diseases* and *eLife*, respectively.

Articles

Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study

Isaac I Bogoch, Oliver J Brady*, Moritz U G Kraemer*, Matthew German, Maria I Creatore, Shannon Brent, Alexander G Watts, Simon I Hay, Manisha A Kulkarni, John S Brownstein, Kamran Khan*

Summary

Background

o support public health readiness, we aim to identify regions and times where the potential health, economic, and social effects from Zika virus are greatest, focusing on resource-limited countries in Africa and the Asia-Pacific region.

avivirus transmission, using data from the International Air Transport Association, entomological observations from Zika's primary vector species, and climate conditions using WorldClim. We overlaid monthly flows of airline travellers arriving to Africa and the Asia-Pacific region from areas of the Americas suitable for year-round transmission of Zika virus with monthly maps of climatic suitability for mosquito-borne transmission of Zika virus within Africa and the Asia-Pacific region.

c region where the presence of competent mosquito vectors and suitable climatic conditions could support local transmission of Zika virus. Countries with large volumes of travellers arriving from Zika virus-affected areas of the Americas and large populations at risk of mosquito-borne Zika virus infection include India (67 422 travellers arriving per year; 1·2 billion residents in potential Zika transmission areas), China (238 415 travellers; 242 million residents), Indonesia (13 865 travellers; 197 million residents), Philippines (35 635 travellers; 70 million residents), and Thailand (29 241 travellers; 59 million residents).

c region are vulnerable to Zika virus. Strategic use of available health and human resources is essential to prevent or mitigate the health, economic, and social consequences of Zika virus, especially in resource-limited countries.

Lancet Infect Dis 2016

Medicine, Division of Infectious Diseases, University of Toronto, Toronto, ON, Canada (I I Bogoch MD, K Khan MD); **Divisions of General Internal Medicine and Infectious Diseases, University Health Network, Toronto, ON, Canada** (I I Bogoch); **Centre for the Mathematical Modelling of Infectious Diseases and Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK** (O J Brady DPhil); **Spatial Ecology and Epidemiology Group (SEEG), Department of Zoology, University of Oxford, UK** (M U G Kraemer BSc); **Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada** (M German MSc, M I Creatore PhD, A G Watts PhD, K Khan); **Dalla Lana School of Public Health, University of Toronto, ON, Canada** (M I Creatore, S Brent MPH); **Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA** (Prof S I Hay DSc); **Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK** (Prof S I Hay); **School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, ON, Canada** (M A Kulkarni PhD); **and Children's Hospital Boston, Harvard Medical School, Boston, MA, USA** (Prof J S Brownstein PhD)

Correspondence to:
Dr Kamran Khan,
St Michael's Hospital, Toronto, ON M5B 1W8, Canada
khank@smh.ca

Introduction

Zika virus is a flavivirus that was first isolated in 1947 from a sentinal rhesus monkey in a Ugandan forest,[1] with the first human cases identified 5 years later in Nigeria.[2] Humans become infected with Zika virus mainly through the bites of several species of aedes mosquitoes, including *Aedes aegypti* and presumably *Aedes albopictus*.[3] Eradication of Zika virus poses substantial challenges because of its sylvatic transmission cycle between aedes mosquitoes and non-human primates.[4–8] Although Zika virus is known to have circulated in parts of Africa and the Asia-Pacific region,[1,9,10] a series of epidemics during the past decade have transported this virus eastward across islands in the Pacific Ocean.[11,12] On May 15, 2015, the Program for Monitoring Emerging Diseases (ProMED-mail) published a report confirming locally acquired cases of Zika virus in several northeastern Brazilian states, marking the first time that the virus is known to have spread within the Americas.[13–15] On Feb 1, 2016, WHO declared this epidemic to be a global Public Health Emergency of International Concern, mainly because of its rapid spread between countries and its vertical transmission of human infection resulting in birth defects, including but not limited to microcephaly[16–19] and an association with Guillain-Barré syndrome.[20] As of June 27, 2016, autochthonous transmission of Zika virus has been confirmed in 40 countries and territories within the Americas.[21]

As the epidemic expands its range, increasing numbers of travellers are transporting Zika virus to distant regions across the world.[22] If competent mosquito vectors become infected from these travellers in areas where environmental conditions are conducive to the virus's spread, new epidemics could occur, subject to the presence of an immunologically susceptible human population.[23] During warmer months across the northern hemisphere when mosquito vectors are most abundant, active, and capable of transmitting arboviruses to humans, the risk of new epidemics appearing outside of the Americas is expected to peak. To support public health readiness, we aim to identify regions and times where the potential

www.thelancet.com/infection **Published online September 1, 2016 http://dx.doi.org/10.1016/S1473-3099(16)30270-5**

Research in context

the potential for the international spread of the Asian strain of Zika virus from Brazil to other parts of the Americas, and from the Latin American and Caribbean region to Europe. We searched PubMed for studies about the international spread of Zika virus published between June 1, 2015, and June 30, 2016. Our search terms included "Zika", "Zika virus", "air travel", mosquito", and "*Aedes*". Additionally, we searched ProMED-mail for all reports of "Zika" and "Zika virus" over the same timeframe. No previous studies have modelled the potential for importation to, and local transmission of Zika virus within, Africa and the Asia-Pacific region

o our knowledge, our study is the first to investigate the potential spread of Zika virus from the Americas into resource-limited areas of Africa and the Asia-Pacific region. This study provides timely data to help decision-makers to optimise use of their finite public health resources by assessing seasonal variability in the risks of Zika virus importation alongside seasonal ecological suitability for autochthonous transmission.

Implications of all the available evidence

Across Africa and the Asia-Pacific region, an estimated 2·6 billion people live in areas that have competent mosquito vectors and suitable climatic conditions for the introduction and local transmission of Zika virus. Our results show the locations and times in Africa and the Asia-Pacific region in which risks of Zika virus importation and autochthonous transmission coincide. Our findings can support time-sensitive public health decisions at local, national, and international levels, particularly as they relate to risks from Zika virus in resource-limited countries.

health, economic, and social effects of Zika virus are greatest, focusing on resource-limited countries in Africa and the Asia-Pacific region.

Methods

Overview

To identify the areas most susceptible to Zika virus in Africa and the Asia-Pacific region (ie, all of Asia and all Pacific Islands), we performed and integrated a series of analyses. These included monthly flows of airline travellers departing from regions of the Americas in which conditions are suitable for year-round local transmission of Zika virus (as a proxy for where departing travellers are most likely to become infected with Zika virus during the current outbreak); monthly climatic suitability models for autochthonous transmission of Zika virus in Africa and the Asia-Pacific region, conditional on the predicted occurrence of competent aedes mosquito vectors; estimates of the number of people living in countries in Africa and the Asia-Pacific region where the potential for mosquito-borne transmission of Zika virus exists; and health expenditure per capita as a proxy of countries' capacity to detect and effectively respond to Zika virus importation and domestic epidemic transmission.

See Online for appendix

Geographical range of Zika virus in the Americas

Given the high proportion of asymptomatic or subclinical Zika virus infections, combined with the limited capacity for diagnosis in some parts of the Americas, substantial under-reporting or delayed reporting of Zika virus infections is widely presumed. Hence, we assumed that by July, 2016, more than 1 year after the first cases of Zika virus were confirmed in Brazil, the virus was active across all geographical areas in the Americas conducive to year-round transmission. To delineate this geographical range, we combined continuous distribution maps of *A aegypti* and *A albopictus*,[24] regarding both as competent vectors, and converted the combined data into binary presence–absence maps. The combined geographical range was defined by the minimal extent that incorporated 90% of all known independently observed mosquito occurrences. We applied a model of dengue virus intra-mosquito dynamics[25] to this combined map to assess whether the temperature profile for a given geographical range would enable adult female aedes mosquitoes to survive for long enough for dengue virus to complete its extrinsic incubation period (ie, to enable a newly infected mosquito to survive for long enough to become infectious).[25] We used the dengue virus model to represent competence of Zika virus transmission because of the paucity of data on the extrinsic incubation period for Zika virus and how it varies in response to temperature or across aedes mosquito species. Evidence from previous studies shows similar extrinsic incubation periods across related flaviviruses.[26] We judged this approach to be reasonable, because evidence suggests that dengue virus and Zika virus share many common epidemiological features, including the same vector species, rate of epidemic progression, and general latitudinal and seasonal limits in South America (appendix pp 1–11). We then paired this seasonal model with high-resolution empirical global temperature data[27] to predict the number of days throughout a typical year (average year profile 1950–2000) in which local transmission of Zika virus would be possible.[28] These data were then used to map areas where climatic conditions are conducive to year-round transmission of Zika virus via *A aegypti* or *A albopictus*. We subsequently refined this geographical range by excluding the USA, since at the time of analysis, there were no locally acquired mosquito-borne cases of disease and because the transmission of another emerging arbovirus

 www.thelancet.com/infection Published online September 1, 2016 http://dx.doi.org/10.1016/S1473-3099(16)30270-5

transmitted by aedes mosquitoes—chikungunya—was absent or limited in 2014–15 (ie, the USA reported only 12 confirmed cases in Florida in 2014). We refer to this geographical extent as the ecological niche for Zika virus in the Americas.

Translocation of Zika virus out of the Americas

We established a 50 km buffer zone around the ecological niche for Zika virus in the Americas, to accommodate the potential movement of individuals travelling from areas of Zika virus activity to nearby commercial airports. For the 689 cities with one or more commercial airports falling within this zone, we analysed worldwide airline ticket sales and flight itinerary data from the International Air Transport Association (IATA) between Dec 1, 2014, and Nov 30, 2015. These itineraries included data on the initial airport of embarkation, final airport destination, and where applicable, connecting airports, representing an estimated 90% of all passenger trips on commercial flights worldwide (the remaining 10% of passenger data are estimates modelled by the IATA). We then mapped the monthly final destinations of all travellers departing from airports within these buffered zones for airports in Africa and the Asia-Pacific region.

Transmission of Zika virus within Africa and Asia-Pacific

Overview

Current evidence suggests that *A aegypti* and *A albopictus* are both potentially capable of transmitting Zika virus.[29] However, the relative competence for each mosquito vector is not well understood. To address this uncertainty, we created three scenarios to reflect suitability for autochthonous Zika virus transmission in Africa and the Asia-Pacific region. We assumed that Zika virus suitability would either be the same as the global geographical range for dengue[30] (scenario 1; most specific assumption); include the global geographical range for dengue, but also include areas in which *A aegypti* occurrences are predicted and where conducive climatic conditions exist for autochthonous Zika virus transmission (scenario 2; intermediate assumption); or include the global geographical range of dengue, but also include areas in which *A aegypti* and *A albopictus* occurrences are predicted and where conducive climatic conditions exist for Zika virus autochthonous transmission (scenario 3; most sensitive assumption).

Zika seasonal suitability maps

To account for seasonal variation in the geographical range of Zika virus suitability, we produced maps for each month of the year to delineate our three suitability scenarios (figures 1 and 2; appendix pp 1–11). We used mosquito species distribution models for *A aegypti* and *A albopictus*[24] (originally fitted to annual data and covariates) to make monthly predictions by use of new monthly covariates for temperature-persistence suitability,[25] relative humidity, and precipitation. Non-dynamic model covariates (eg, urbanisation) were assumed to be constant

[23] This approach allowed us to account for relative changes in suitability throughout the year and between regions. The continuous seasonal maps were converted into binary range maps for each species. Final monthly vector maps show predictions of areas with high likelihood for observation or detection of mosquito populations, which were assumed to be sufficiently abundant to enable transmission of disease to humans.

To add spatiotemporal resolution to our predicted Zika virus transmission zones, we used an arbovirus incubation period model[25,31] to identify regions and times where temperature would permit Zika virus transmission to humans (ie, mosquito survival exceeds the extrinsic incubation period). Pixels (areas of 5 km²) for which temperature did not meet extrinsic incubation period criteria for at least 21 days of the month (ie, assuming that Zika virus could be sustained in human populations during days of absence in mosquitoes) were masked out of the final predicted transmission zones.

Populations at risk in Africa and the Asia-Pacific region

After defining our three suitability zones for autochthonous Zika virus transmission for each month of the year, we calculated the number of people living in each zone by extracting population data from LandScan (Oak Ridge, TN, USA), a satellite-based sensor dataset that estimates ambient population density worldwide in 1 km² grids. To link these datasets, we resampled our Zika virus suitability zone maps (originally at 5 km² resolution) to 1 km² pixels by use of a nearest-neighbour sampling algorithm and aligned them with national boundaries to extract population values. For each country, we estimated the number of people at risk of mosquito-borne Zika virus exposure during the month in which the geographical range for Zika virus suitability was greatest. We did not make any assumptions about the degree of existing Zika virus immunity in populations in Africa and the Asia-Pacific region, despite sporadic and historical reports of Zika virus cases from the continent.[23] As a proxy for health care and public health capacity, we ranked countries by health expenditures per capita (in 2014 US$) as reported by the World Bank.[31]

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

70% of travellers departing from the ecological niche of Zika virus in the Americas for destinations in Africa and the Asia-Pacific region arrived in ten countries: China (238 415

travellers per year), Japan (179 926 travellers per year), Israel (106 365 travellers per year), Australia (96 430 travellers per year), Turkey (90 632 travellers per year), Angola (88 048 travellers per year), South Korea (87 768 travellers per year), United Arab Emirates (70 848 travellers per year), India (67 422 travellers per year), and South Africa (59 318 travellers per year [appendix pp 94–122]).

Worldwide, an estimated 2·6 billion people live in areas of Africa and the Asia-Pacific region where the presence of competent mosquito vectors and suitable climatic conditions could support local transmission of Zika virus. According to our most conservative scenario (scenario 1), populations living within the geographical range for Zika virus (during the month where the geographical range was broadest) were highest in India (1·2 billion people), China (242 million), Indonesia (197 million), Nigeria (179 million), Pakistan (168 million), and Bangladesh (163 million [appendix pp 94–22]). Geographical variability in the three suitability zones for autochthonous Zika virus transmission on a quarterly and monthly basis is shown for Africa (figure 1; appendix pp 12–24); and the Asia-Pacific region (figure 2; appendix pp 25–37).

On the basis of the volume of travellers arriving from airports within the ecological niche of Zika virus in the Americas, the resident population at risk for Zika virus exposure during the month of its broadest activity, and

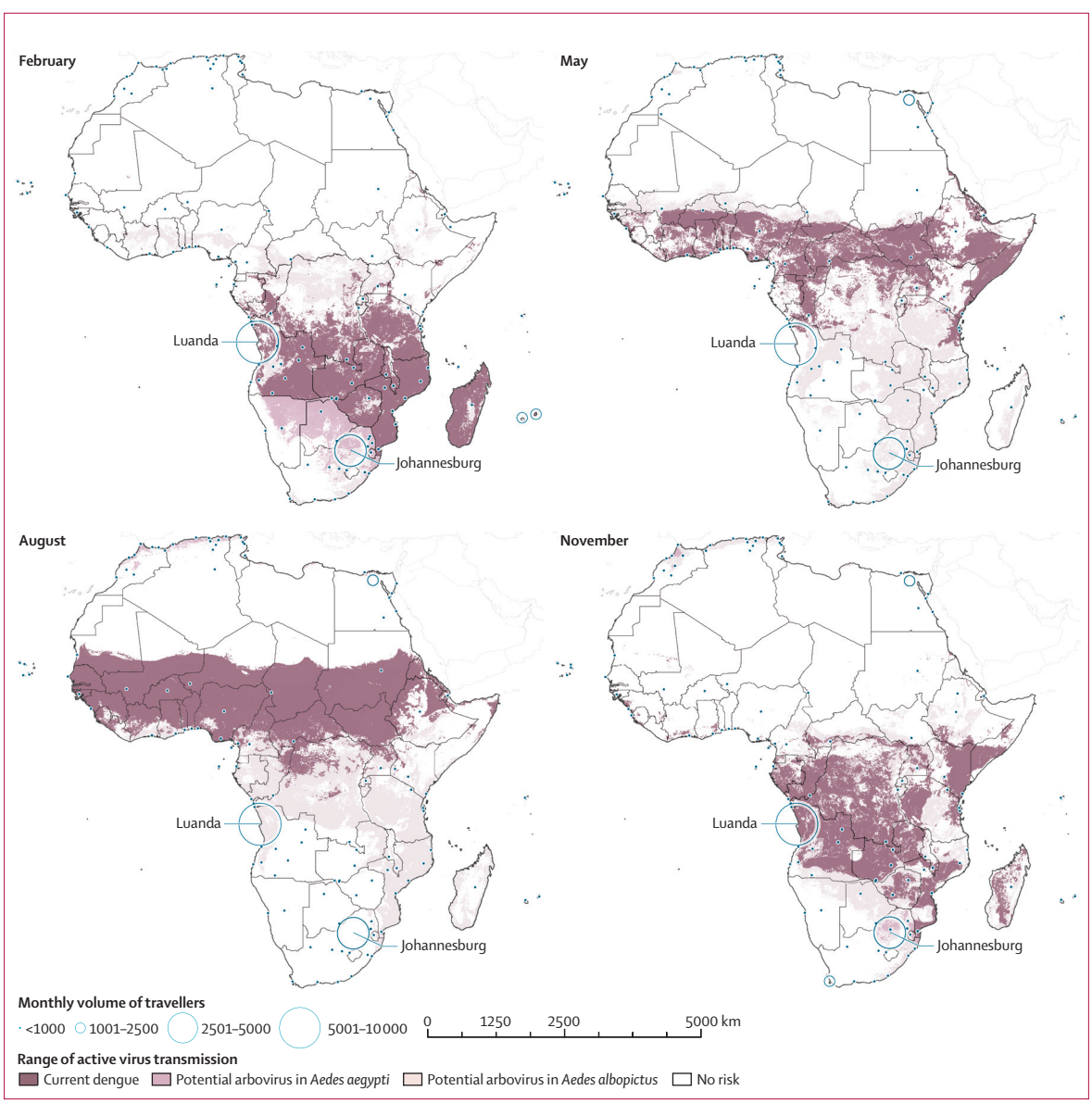


Figure 1: **Seasonal geographical suitability for Zika virus transmission in Africa and seasonal volume of airline travellers arriving from the Americas**
Monthly maps are shown for Africa. Travellers arriving from the Americas refers to travellers originating from regions of Latin America and the Caribbean that are suitable for year-round transmission of Zika virus.

www.thelancet.com/infection **Published online September 1, 2016** **http://dx.doi.org/10.1016/S1473-3099(16)30270-5**

health expenditures per capita, we found that India, the Philippines, Indonesia, Nigeria, Vietnam, Pakistan, and Bangladesh have some of the highest expected risks for Zika virus importation and population health impact (figure 3). Although China receives the greatest number of travellers of all countries in Africa or the Asia-Pacific region and has a large population predicted to be living within the geographical range of dengue virus, it spends significantly more on health than do many of the highest risk countries. Whereas countries such as Tanzania, Ethiopia, Mozambique, and the Democratic Republic of Congo have smaller populations at risk of Zika virus exposure, each has a moderately high volume of travellers arriving from the Americas and low health expenditures (table).

Discussion

By integrating time-dependent analyses on the potential for international dispersion of Zika virus and environmental suitability for autochthonous transmission, we

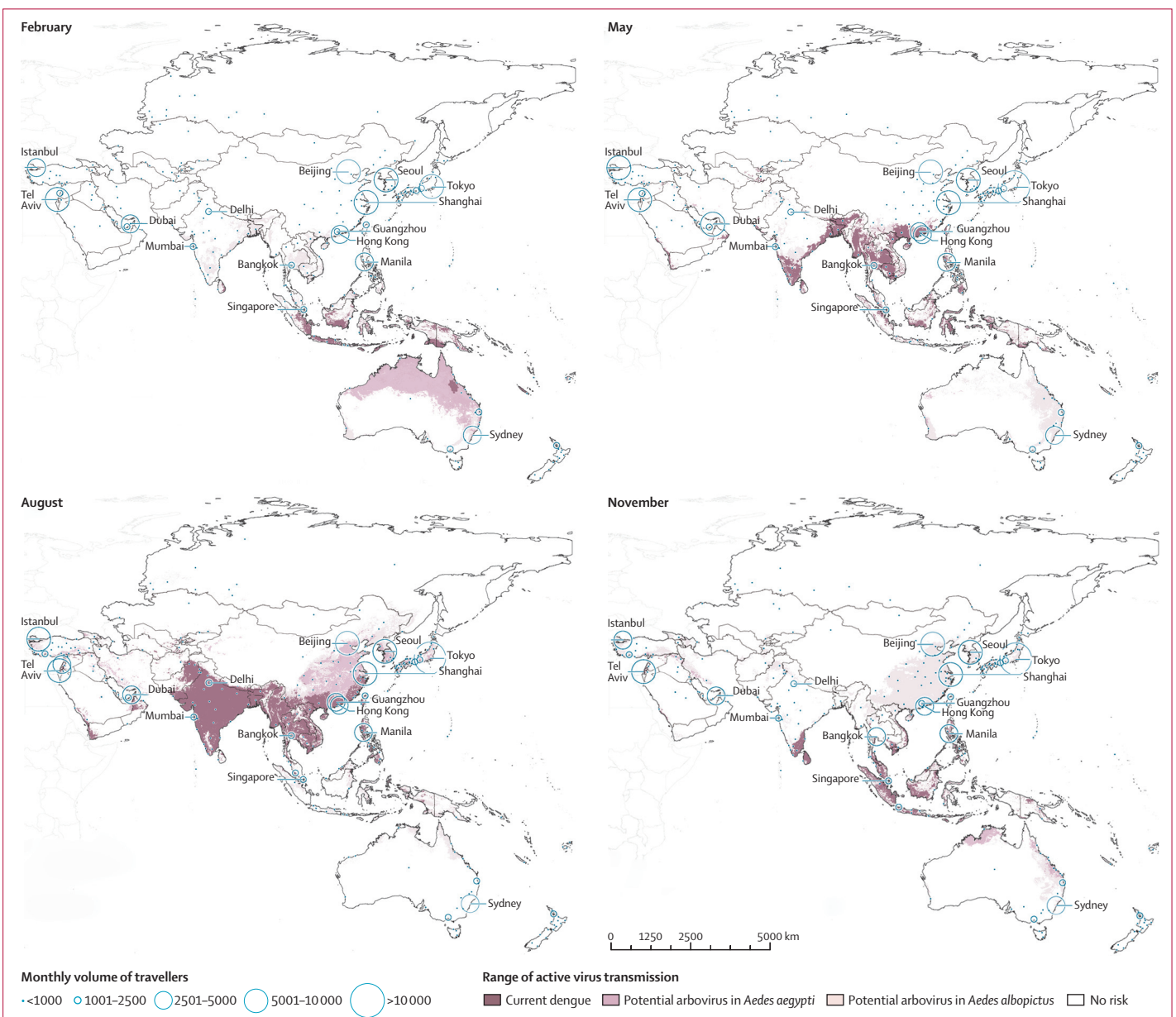


Figure 2: **Seasonal geographical suitability for Zika virus transmission in Asia-Pacific and seasonal volume of airline travellers arriving from the Americas**
Monthly maps are shown for Asia-Pacific region. Travellers arriving from the Americas refers to travellers originating from regions of Latin America and the Caribbean that are suitable for year-round transmission of Zika virus.

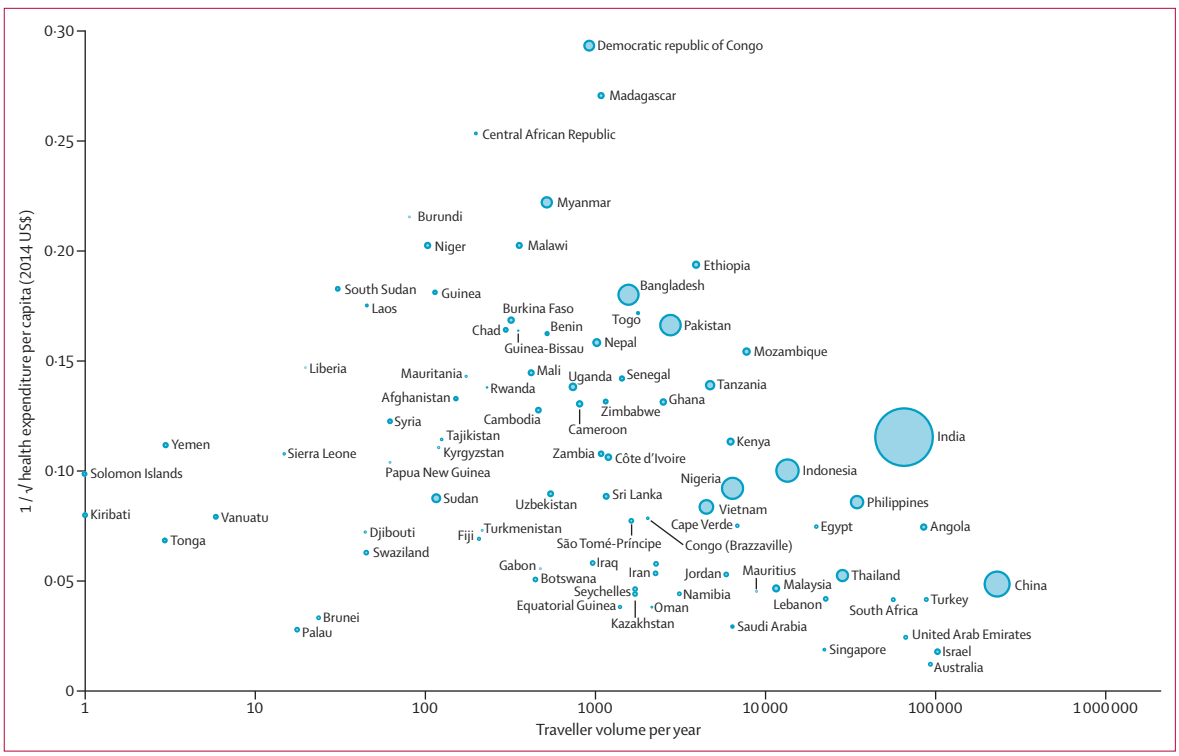


Figure 3: **Volume of airline travellers arriving from the Americas by peak resident population at risk of Zika virus exposure and health expenditures per capita**
The x-axis shows log of the annual volume of travellers arriving from airports within the ecological niche of Zika virus in the Americas, y-axis shows the transformation of health expenditures per capita (square root of 1/health expenditures per capita), and the size of circles shows the populations residing within the geographical range of Zika virus during the month of its broadest activity. Travellers arriving from the Americas refers to travellers originating from regions of Latin America and the Caribbean that are suitable for year-round transmission of Zika virus. Due to highly skewed distribution of values, volume of airline travellers was plotted on a logarithmic scale. Population at risk of Zika virus exposure refers to resident populations living in areas suitable for Zika virus transmission during the month when the geographical range of suitability is broadest. Health expenditures are measured in 2014 US$; because of highly skewed distribution of values, we transformed health expenditures per capita for visualisation purposes.

aimed to identify where and when populations would be most susceptible to the health, economic, and social effects of Zika virus. In India alone, an estimated 1·2 billion people are susceptible to Zika virus exposure at the time of peak seasonal risk. In China, an estimated 242 million people live in southern regions of the country where large dengue virus epidemics have occurred previously,[32,33] and hence where Zika virus has the potential to circulate locally. However, the predicted suitability range for autochthonous transmission by aedes mosquitoes extends into northern China, where collectively, more than one billion people reside.[24,28] In Africa, Angola receives the largest number of travellers from Zika virus-affected countries in the Americas, probably because of Angola's cultural and historical ties to Brazil. In Angola, there is an historical and ongoing epidemic of yellow fever,[34] an arbovirus also transmitted by *A aegypti*, that is already straining the public health system. The epidemiological significance of these ties has already been shown in the context of the exportation of chikungunya from Angola to Feira de Santana, Brazil, which has since affected many countries across the Americas.[35]

Although our analysis emphasises the potential for autochthonous transmission of Zika virus via aedes mosquitoes, sexual transmission via travellers returning from Zika virus-affected areas is now well documented.[36] Given the current geographical range of Zika virus within the Americas, many travellers returning from affected areas would benefit from health education to prevent sexual transmission. For example, since Zika virus RNA has been detected in semen 93 days after symptom onset,[37] the US Centers for Disease Control and Prevention have issued recommendations to help to guide individuals in the prevention of sexual transmission of Zika virus.[21] Although these recommendations could change as new evidence becomes available, our analysis of traveller flows from the Americas emphasises how populations in specific destinations in Africa and the Asia-Pacific region could benefit from intensified public health education.

Our findings must be considered in the context of a number of assumptions and data limitations. First, we regarded *A albopictus* as a competent vector for Zika virus, including it in one of our three analytical scenarios, even though understanding of its capacity to transmit Zika virus is still evolving.[3] Furthermore, we did not account for other *Aedes* species such as *Aedes africanus* and *Aedes hensilli*,[1,8] which might be competent vectors

 www.thelancet.com/infection Published online September 1, 2016 http://dx.doi.org/10.1016/S1473-3099(16)30270-5

	Country	Month of peak exposure to Zika virus	Population at risk during peak exposure month (millions)			Traveller volume at peak exposure* (number of travellers)	Health expenditure per capita (US$), 2014	Health expenditure per capita rank (of 104 countries), 2014	Alignment of peak travel and peak exposure†
			Scenario 1 (most conservative scenario)	Scenario 2	Scenario 3 (least conservative scenario)				
1	India	August	1181·6	1181·6	1208·3	5430	75	71	No
2	China	July	241·8	969·7	1134·9	19 266	419·7	27	Yes
3	Indonesia	December	196·9	196·9	227·6	1012	99·4	62	No
4	Nigeria	July	178·5	178·5	181·5	359	117·5	59	No
5	Pakistan	August	168·0	168·0	172·5	236	36·2	90	No
6	Bangladesh	July	162·9	162·9	162·9	106	30·8	94	No
7	Vietnam	September	82·7	82·7	85·6	481	142·4	54	Yes
8	Philippines	July	70·2	70·2	80·9	3597	135·2	55	Yes
9	Thailand	September	59·3	59·3	60·5	1464	360·4	31	No
10	Myanmar	June	50·9	50·9	53·4	47	20·3	101	No
11	Democratic Republic of Congo	November	47·3	47·3	62·1	105	11·6	104	Yes
12	Tanzania	March	35·9	35·9	42·9	567	51·7	80	Yes
13	Sudan	August	32·9	32·9	33·0	<10	129·8	56	No
14	Nepal	July	26·5	26·5	30·2	50	39·9	86	No
15	Uganda	April	24·6	24·6	30·4	39	52·3	79	No
16	Mozambique	February	23·4	23·4	23·7	568	42	85	No
17	Ethiopia	July	22·4	22·4	44·5	350	26·6	97	No
18	Taiwan	July	22·0	22·0	22·3	1259	No data	No data	No
19	Kenya	April	21·7	21·7	28·1	457	77·7	69	No
20	Malaysia	November	21·7	21·7	26·9	927	455·8	26	No

Travellers arriving from the Americas refers to travellers originating from regions of Latin America and the Caribbean that are suitable for year-round transmission of Zika virus. Resident population at risk of Zika virus exposure refers to resident populations living in areas suitable for Zika virus transmission during the month when the geographical range of suitability is broadest. Scenario 1 is Zika virus extent equal to dengue extent; scenario 2 is Zika virus extent equal to dengue extent plus *Aedes aegypti* extent; and scenario 3 is Zika virus extent equal to dengue extent plus *A aegypti* plus *Aedes albopictus* extents. *Based on the month when the size of the population exposed to dengue virus was predicted to be greatest. †Based on alignment of month of peak exposure to Zika virus (based on dengue extent) and months with peak travel volumes (defined as top quartile).

Table: **Countries in Africa and Asia-Pacific ranked by monthly volume of airline travellers arriving from the Americas and peak resident population at risk of Zika virus exposure**

for Zika virus in particular geographical areas. We also assumed that the extrinsic incubation period for Zika virus in aedes mosquitoes was similar to that of dengue virus, since the extrinsic incubation periods for other related arboviruses have been similar.[26] In our estimates of the public health and health-care capacity of countries, we assumed that health expenditures per capita would be a reasonable proxy, recognising that it will not address variability in resources within countries and might not adequately capture the many dimensions that comprise a country's ability to respond to imported or locally transmitted Zika virus.

As an emerging epidemic currently concentrated in the Americas, we were unable to validate the outputs of our analysis for the Africa and Asia-Pacific regions. Given that an estimated 80% of all Zika virus infections are asymptomatic, and because resource-limited countries in Africa and the Asia-Pacific region might have sub-optimum surveillance capacity to readily detect Zika virus infections, our validation efforts were retrospective and focused on the Americas. For example, the predicted arrival of Zika virus into Brazil in 2013 coincided with a surge in the volume of airline travellers arriving into Brazil from countries with reported Zika virus activity.[15] Although not formally assessed, the observed spread of Zika virus across the Americas has largely aligned with the findings of a modelling study that used a similar methodological approach to the one used in this analysis.[38] At the time of writing, the only country in Africa to have confirmed locally acquired infections with the Asian strain of Zika virus (identical to the strain circulating in the Americas) is Cape Verde. In our analysis, we noted that Cape Verde received more than 7000 travellers from Zika virus-affected countries in 2015 (ranked 27th of 118 countries), including direct flights from northeastern Brazil. Further validation efforts will become possible as more data on the observed spread of Zika virus become available.

Finally, the population health consequences of imported Zika virus from the Americas will depend heavily on underlying levels of immunity to Zika virus in Africa and the Asia-Pacific region. Although sporadic cases of Zika virus have been reported in both continents,[23] the breadth and extent of previous infection with Zika virus remains unknown. To fill this gap in knowledge, serological surveys in Africa and Asia-Pacific are needed. Furthermore, the degree and duration of protective immunity stemming from infection with Zika virus is also unknown. This point is especially relevant when considering whether previous infection with the African strain of Zika virus confers any protective benefit against infections with the Asian strain of the virus.

As the Zika virus epidemic in the Americas intensifies and expands, hundreds, and possibly thousands, of infected travellers are now transporting the virus to distant regions of the world. Given the broad global range of aedes mosquitoes and the arrival of summer in the northern hemisphere, these translocation events could catalyse new Zika virus epidemics, in much the same manner that the epidemic in Brazil began. The potential for epidemics to occur in parts of Africa and the Asia-Pacific region is particularly worrying given the vast numbers of people who are potentially susceptible to Zika virus and are living in environments where health and human resources to prevent, detect, and respond to epidemics are limited. Our findings could offer valuable information to support time-sensitive public health decision-making at local, national, and international levels.

Contributors

IIB, OJB, MUGK, and KK contributed to the idea for the study. IIB, OJB, MUGK, MG, MIC, and KK contributed to the design of the study. IIB, OJB, MUGK, MG, MIC, SB, AGW, SIH, MAK, JSB, and KK contributed to the data analysis and IIB, OJB, MUGK, MG, MIC, SIH, MAK, JSB, and KK contributed to the data interpretation. IIB, OJB, MUGK, MG, MIC, and KK contributed to the writing of this manuscript.

Declaration of interests

KK is the founder of BlueDot, a social benefit corporation that models global infectious disease threats. MIC, MG, SB, and AGW have received employment income from BlueDot, and IIB has consulted to BlueDot. All other authors declare no competing interests.

Acknowledgments

This study was supported in part by the Canadian Institutes of Health Research and the US Centers for Disease Control and Prevention (BioMosaic program). SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), and grants from the Bill & Melinda Gates Foundation (OPP1119467, which also supports OJB, OPP1106023 and OPP1093011). MUGK receives funding from the International research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS; European Commission 7th Framework Programme [21893]). The views and opinions expressed in this publication are those of the authors and are not necessarily endorsed by the funding agencies.

References

1 Haddow AJ, Williams MC, Woodall JP, Simpson DI, Goma LK. Twelve isolation isolations of zika virus from *Aedes (stegomyia) africanus (theobald)* taken in and above a Uganda forest. *Bull World Health Organ* 1964; **31:** 57–69.

2 Macnamara FN. Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans R Soc Trop Med Hyg* 1954; **48:** 139–45.

3 Chouin-Carneiro T, Vega-Rua A, Vazeille M, et al. Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika virus. *PLoS Negl Trop Dis* 2016; **10:** e0004543.

4 Marchette NJ, Garcia R, Rudnick A. Isolation of Zika virus from *Aedes aegypti* mosquitoes in Malaysia. *Am J Trop Med Hyg* 1969; **18:** 411–15.

5 Olson JG, Ksiazek TG, Suhandiman, Triwibowo. Zika virus, a cause of fever in Central Java, Indonesia. *Trans R Soc Trop Med Hyg* 1981; **75:** 389–93.

6 Li MI, Wong PSJ, Ng LC, Tan CH. Oral susceptibility of Singapore *Aedes (Stegomyia) aegypti (Linnaeus)* to Zika virus. *PLoS Negl Trop Dis* 2012; **6:** e1792.

7 Wong P-SJ, Li MI, Chong C-S, Ng L-C, Tan C-H. *Aedes (Stegomyia) albopictus (Skuse)*: a potential vector of Zika virus in Singapore. *PLoS Negl Trop Dis* 2013; **7:** e2348.

8 Ledermann JP, Guillaumot L, Yug L, et al. *Aedes hensilli* as a potential vector of Chikungunya and Zika viruses. *PLoS Negl Trop Dis* 2014; **8:** e3188.

9 Fagbami AH. Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State. *J Hyg (Lond)* 1979; **83:** 213–19.

10 Diallo D, Sall AA, Diagne CT, et al. Zika virus emergence in mosquitoes in southeastern Senegal, 2011. *PLoS One* 2014; **9:** e109442.

11 Duffy MR, Chen T-H, Hancock WT, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 2009; **360:** 2536–43.

12 Roth A, Mercier A, Lepers C, et al. Concurrent outbreaks of dengue, chikungunya and Zika virus infections—an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012–2014. *Euro Surveill* 2014; **19:** Article 1.

13 ProMED-mail. Undiagnosed illness—Brazil (02): Zika virus conf . ProMED-mail 2015; 15 May: 20150515.3364149. www.promedmail.org. Accessed June 1, 2016.

14 Faria NR, Azevedo RdSdS, Kraemer MUG, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science* 2016; **352:** 345–49.

15 Brasil P, Pereira JP, Raja Gabaglia C, et al. Zika virus infection in pregnant women in Rio de Janeiro—preliminary report. *N Engl J Med* 2016; published online March 4. DOI:10.1056/NEJMoa1602412.

16 Cauchemez S, Besnard M, Bompard P, et al. Association between Zika virus and microcephaly in French Polynesia, 2013–15: a retrospective study. *Lancet* 2016; **387:** 2125–32.

17 Calvet G, Aguiar RS, Melo ASO, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis* 2016; **16:** 653–60.

18 Mlakar J, Korva M, Tul N, et al. Zika virus associated with microcephaly. *N Engl J Med* 2016; **374:** 951–58.

19 Cao-Lormeau V-M, Blake A, Mons S, et al. Guillain-Barré syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* 2016; **387:** 1531–39.

20 US Centers for Disease Control and Prevention. Zika virus. www.cdc.gov/zika (accessed June 1, 2016).

21 Armstrong P, Hennessey M, Adams M, et al. Travel-associated Zika virus disease cases among U.S. residents—United States, January 2015–February 2016. *MMWR Morb Mortal Wkly Rep* 2016; **65:** 286–89.

22 Messina JP, Kraemer MU, Brady OJ, et al. Mapping global environmental suitability for Zika virus. *Elife* 2016; **5:** e15272.

23 Kraemer MU, Sinka ME, Duda KA, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae albopictus*. *Elife* 2015; **4:** e08347.

24 Brady OJ, Golding N, Pigott DM, et al. Global temperature constraints on *Aedes aegypti* and *Ae albopictus* persistence and competence for dengue virus transmission. *Parasit Vectors* 2014; **7:** 338.

25 Lambrechts L, Paaijmans KP, Fansiri T, et al. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proc Natl Acad Sci USA* 2011; **108:** 7460–65.

26 WordlClim. WorldClim global climate data, Data for current conditions. http://www.worldclim.org/current (accessed Feb 26, 2016).

27 Kraemer MUG, Sinka ME, Duda KA, et al. The global compendium of *Aedes aegypti* and *Ae albopictus* occurrence. *Sci Data* 2015; **2:** 150035.

28 Grard G, Caron M, Mombo IM, et al. Zika virus in Gabon (Central Africa)—2007: a new threat from *Aedes albopictus*? *PLoS Negl Trop Dis* 2014; **8:** e2681.

29 Bhatt S, Gething PW, Brady OJ, et al. The global distribution and burden of dengue. *Nature* 2013; **496:** 504–07.

30 Brady OJ, Johansson MA, Guerra CA, et al. Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasit Vectors* 2013; **6:** 351.

31 World Bank. Health expenditure per capita (current US$). http://data.worldbank.org/indicator/SH.XPD.PCAP (accessed March 3, 2016).

32 Zhang F-C, Zhao H, Li L-H, et al. Severe dengue outbreak in Yunnan, China, 2013. *Int J Infect Dis* 2014; **27:** 4–6.

33 Huang XY, Ma HX, Wang HF, et al. Outbreak of dengue fever in central China, 2013. *Biomed Environ Sci* 2014; **27:** 894–97.

34 World Health Organization. Yellow fever—Angola. www.who.int/csr/don/13-april-2016-yellow-fever-angola/en (accessed April 13, 2016).

35 Nunes MRT, Faria NR, de Vasconcelos JM, et al. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med* 2015; **13:** 102.

36 D'Ortenzio E, Matheron S, de Lamballerie X, et al. Evidence of sexual transmission of Zika virus. *N Engl J Med* 2016; **374:** 2195–98.

37 Mansuy JM, Pasquier C, Daudin M, et al. Zika virus in semen of a patient returning from a non-epidemic area. *Lancet Infect Dis*. 2016; **16:** 894–95.

38 Bogoch II, Brady OJ, Kraemer MUG, et al. Anticipating the international spread of Zika virus from Brazil. *Lancet* 2016; **387:** 335–36.

eLIFE elifesciences.org

RESEARCH ARTICLE

Mapping global environmental suitability for Zika virus

Jane P Messina[1]*, Moritz UG Kraemer[1], Oliver J Brady[2], David M Pigott[2,3], Freya M Shearer[2], Daniel J Weiss[1], Nick Golding[4], Corrine W Ruktanonchai[5], Peter W Gething[1], Emily Cohn[6], John S Brownstein[6], Kamran Khan[7,8], Andrew J Tatem[5,9], Thomas Jaenisch[10,11], Christopher JL Murray[3], Fatima Marinho[12], Thomas W Scott[13], Simon I Hay[2,3]*

[1]Department of Zoology, University of Oxford, Oxford, United Kingdom; [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; [3]Institute for Health Metrics and Evaluation, University of Washington, Seattle, United States; [4]Department of BioSciences, University of Melbourne, Parkville, United Kingdom; [5]WorldPop project, Department of Geography and Environment, University of Southampton, Southampton, United Kingdom; [6]Boston Children's Hospital, Harvard Medical School, Boston, United Kingdom; [7]Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Canada; [8]Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada; [9]Flowminder Foundation, Stockholm, Sweden; [10]Section Clinical Tropical Medicine, Department for Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany; [11]German Centre for Infection Research (DZIF), Heidelberg partner site, Heidelberg, Germany; [12]Secretariat of Health Surveillance, Ministry of Health Brazil, Brasilia, Brazil; [13]Department of Entomology and Nematology, University of California Davis, Davis, United States

***For correspondence:** jane.messina@zoo.ox.ac.uk (JPM); sihay@uw.edu (SIH)

Competing interest: See page 13

Funding: See page 13

Received: 15 February 2016
Accepted: 10 April 2016
Published: 19 April 2016

Reviewing editor: Mark Jit, London School of Hygiene & Tropical Medicine, and Public Health England, United Kingdom

© Copyright Messina et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract Zika virus was discovered in Uganda in 1947 and is transmitted by Aedes mosquitoes, which also act as vectors for dengue and chikungunya viruses throughout much of the tropical world. In 2007, an outbreak in the Federated States of Micronesia sparked public health concern. In 2013, the virus began to spread across other parts of Oceania and in 2015, a large outbreak in Latin America began in Brazil. Possible associations with microcephaly and Guillain-Barré syndrome observed in this outbreak have raised concerns about continued global spread of Zika virus, prompting its declaration as a Public Health Emergency of International Concern by the World Health Organization. We conducted species distribution modelling to map environmental suitability for Zika. We show a large portion of tropical and sub-tropical regions globally have suitable environmental conditions with over 2.17 billion people inhabiting these areas.
DOI: 10.7554/eLife.15272.001

Introduction

Zika virus (ZIKV) is an emerging arbovirus carried by mosquitoes of the genus *Aedes* (***Musso et al., 2014***). Although discovered in Uganda in 1947 (***Dick et al., 1952***; ***Dick, 1953***) ZIKV was only known to cause sporadic infections in humans in Africa and Asia until 2007 (***Lanciotti et al., 2008***), when it caused a large outbreak of symptomatic cases on Yap island in the Federated States of Micronesia (FSM), followed by another in French Polynesia in 2013–14 and subsequent spread across Oceania (***Musso et al., 2015a***). In the 2007 Yap island outbreak, it was estimated that approximately 20% of ZIKV cases were symptomatic. While indigenous transmission of ZIKV to humans was reported for

eLife digest Zika virus is transmitted between humans by mosquitoes. The majority of infections cause mild flu-like symptoms, but neurological complications in adults and infants have been found in recent outbreaks.

Although it was discovered in Uganda in 1947, Zika only caused sporadic infections in humans until 2007, when it caused a large outbreak in the Federated States of Micronesia. The virus later spread across Oceania, was first reported in Brazil in 2015 and has since rapidly spread across Latin America. This has led many people to question how far it will continue to spread. There was therefore a need to define the areas where the virus could be transmitted, including the human populations that might be risk in these areas.

Messina et al. have now mapped the areas that provide conditions that are highly suitable for the spread of the Zika virus. These areas occur in many tropical and sub-tropical regions around the globe. The largest areas of risk in the Americas lie in Brazil, Colombia and Venezuela. Although Zika has yet to be reported in the USA, a large portion of the southeast region from Texas through to Florida is highly suitable for transmission. Much of sub-Saharan Africa (where several sporadic cases have been reported since the 1950s) also presents an environment that is highly suitable for the Zika virus. While no cases have yet been reported in India, a large portion of the subcontinent is also suitable for Zika transmission.

Over 2 billion people live in Zika-suitable areas globally, and in the Americas alone, over 5.4 million births occurred in 2015 within such areas. It is important, however, to recognize that not all individuals living in suitable areas will necessarily be exposed to Zika.

We still lack a great deal of basic epidemiological information about Zika. More needs to be known about the species of mosquito that spreads the disease and how the Zika virus interacts with related viruses such as dengue. As such information becomes available and clinical cases become routinely diagnosed, the global evidence base will be strengthened, which will improve the accuracy of future maps.

DOI: 10.7554/eLife.15272.002

the first time in Latin America in 2015 (**Zanluca et al., 2015**; **World Health Organisation, 2015**), recent phylogeographic research estimates that the virus was introduced into the region between May and December 2013 (**Faria et al., 2016**). This recent rapid spread has led to concern that the virus is following a similar pattern of global expansion to that of dengue and chikungunya (**Musso et al., 2015a**).

ZIKV has been isolated from 19 different *Aedes* species (**Haddow et al., 2012**; **Grard et al., 2014**), but virus has been most frequently found in *Ae. aegypti* (**Monlun et al., 1992**; **Marchette et al., 1969**; **Smithburn, 1954**; **Pond, 1963**; **Faye et al., 2008**; **Foy et al., 2011b**; **WHO Collaborating Center for Reference and Research on Arboviruses and Hemorrhagic Fever Viruses: Annual Report, 1999**). These studies were based upon ancestral African strains of ZIKV, but the current rapid spread of ZIKV in Latin America is indicative of this highly efficient arbovirus vector (**Marcondes and Ximenes, 2015**). The relatively recent global spread of *Ae. albopictus* (**Benedict et al., 2007**; **Kraemer et al., 2015c**) and the rarity of ZIKV isolations from wild mosquitoes may also partially explain the lower frequency of isolations from *Ae. albopictus* populations. Whilst virus transmission by *Ae. albopictus* and other minor vector species has normally resulted in only a small number of cases (**Kutsuna et al., 2015**; **Roiz et al., 2015**), these vectors do pose the threat of limited transmission (**Grard et al., 2014**). The wide geographic distribution of *Ae. albopictus* combined with the frequent virus introduction *via* viraemic travellers (**McCarthy 2016**; **Bogoch et al., 2016**; **Morrison et al., 2008**; **Scott and Takken, 2012**), means the risk for ZIKV infection *via* this vector must therefore also be considered in ZIKV mapping.

The fact that ZIKV reporting was limited to a few small areas in Africa and Asia until 2007 means that global risk mapping has not, until recently, been a priority (**Pigott et al., 2015b**). Recent associations with Guillain-Barré syndrome in adults and microcephaly in infants born to ZIKV-infected mothers (**World Health Organisation, 2015**; **Martines et al., 2016**) have revealed that ZIKV could lead to more severe complications than the mild rash and flu-like symptoms that characterize the

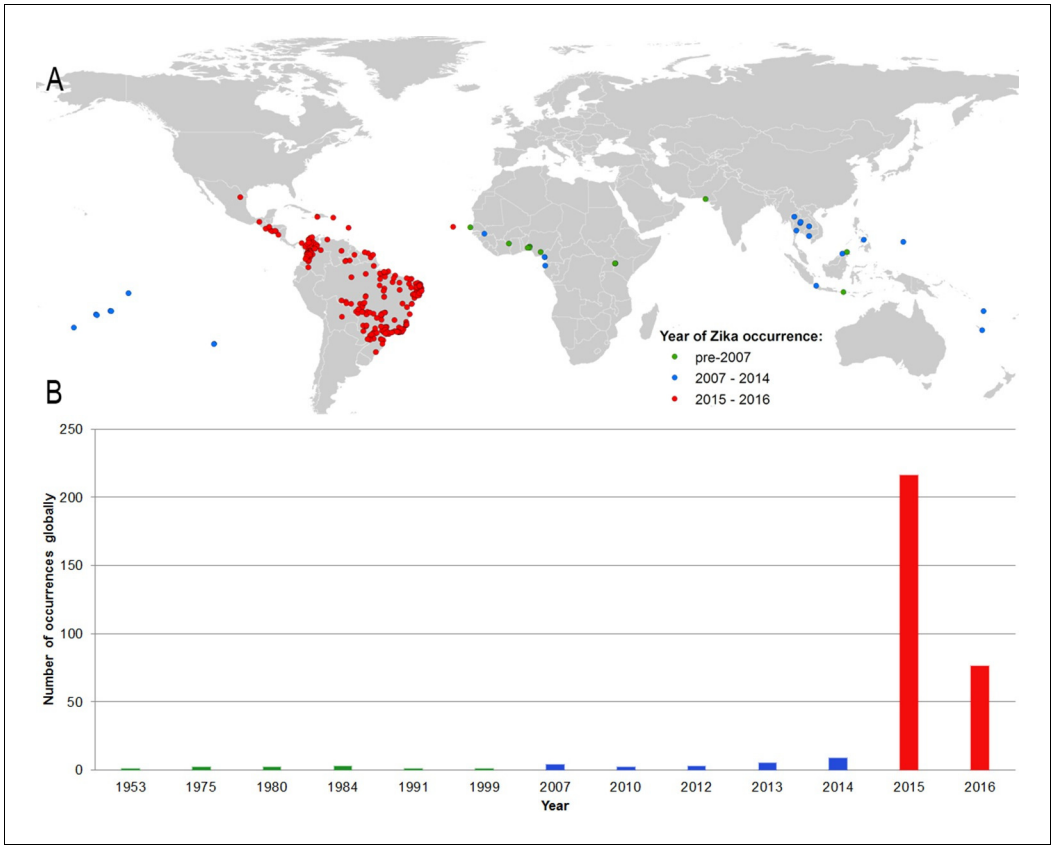


Figure 1. (A) Map showing the distribution of the final set of 323 ZIKV occurrence locations entered into the ensemble Boosted Regression Tree modelling procedure. Locations are classified by year of occurrence to show those which took place (i) prior to the 2007 outbreak in Federated States of Micronesia; (ii) between 2007–2014; and (iii) during the 2015–2016 outbreak; (B) the total number of locations reporting symptomatic ZIKV occurrence in humans globally over time.

DOI: 10.7554/eLife.15272.003

The following figure supplement is available for figure 1:

Figure supplement 1. Maps of all covariates entered into the 300 BRT models.

DOI: 10.7554/eLife.15272.004

majority of symptomatic cases (***Gatherer and Kohl, 2016***). Considering these potentially severe complications and the rapid expansion of ZIKV into previously unaffected areas, the global public health community needs information about those areas that are environmentally suitable for transmission of ZIKV to humans. Being a closely related flavivirus to DENV, there is furthermore the potential for antigen-based diagnostic tests to exhibit cross-reactivity when IgM ELISA is used for rapid diagnosis. Although ZIKV-specific serologic assays are being developed by the U.S. Centers for Disease Control, currently the only method of confirming ZIKV infection is by using PCR on acute specimens (Lanciotti et al., 2008, ***Faye et al., 2008***). Awareness of suitability for transmission is essential if proper detection methods are to be employed.

In this paper, we use species distribution modelling techniques that have been useful for mapping other vector-borne diseases such as dengue (***Bhatt et al., 2013***), Leishmaniasis (***Pigott et al., 2014b***), and Crimean-Congo Haemorrhagic Fever (***Messina et al., 2015b***) to map environmental suitability for ZIKV. The environmental niche of a disease can be identified according to a combination of environmental conditions supporting its presence in a particular location, with statistical modelling then allowing this niche to be described quantitatively (***Kraemer et al., 2016***). Niche modelling uses records of known disease occurrence alongside hypothesized environmental covariates to predict suitability for disease transmission in regions where it has yet to be reported (***Elith and Leathwick, 2009***). Contemporary high spatial-resolution global data representing a

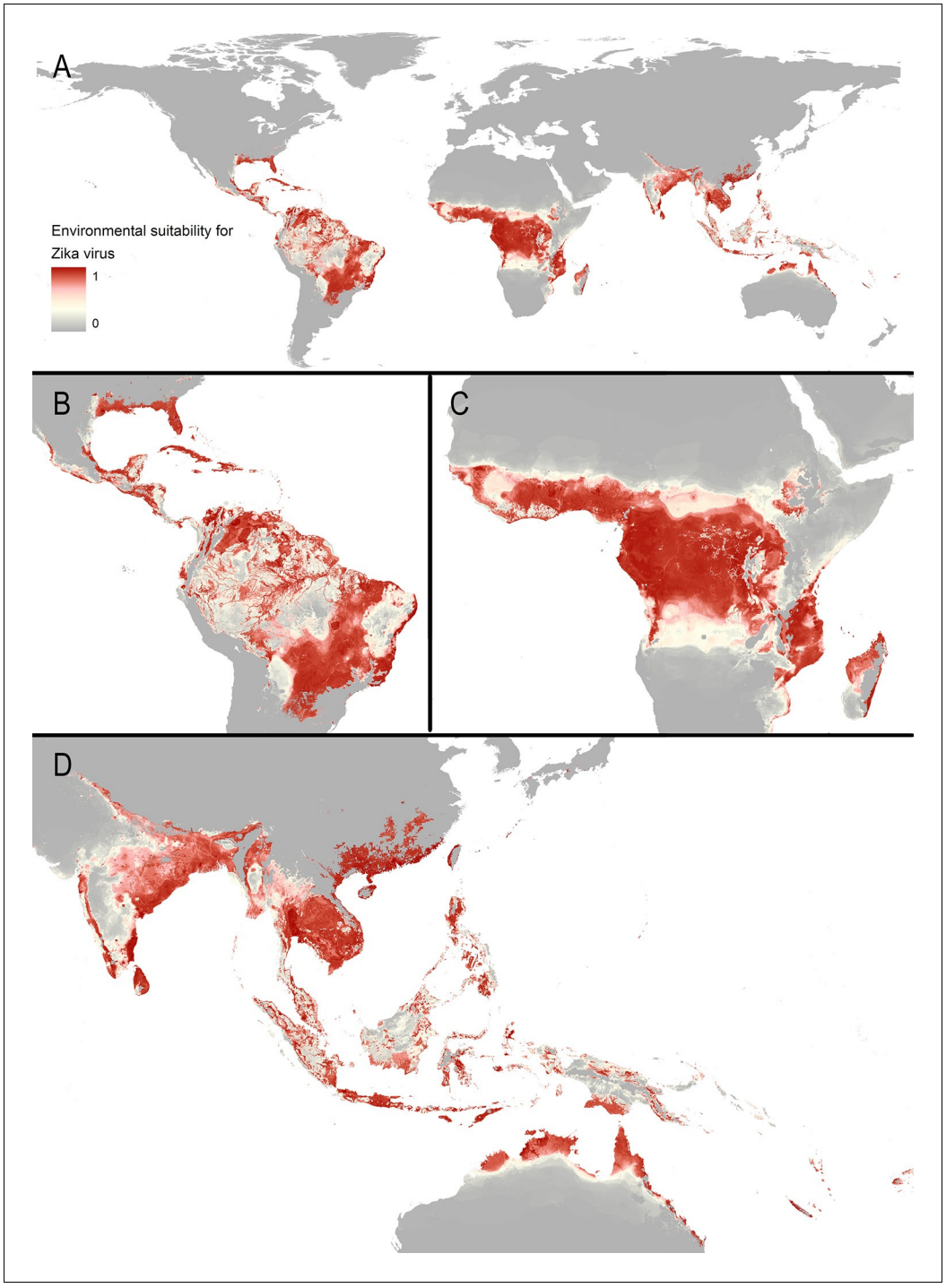


Figure 2. Maps of (**A**) global environmental suitability for ZIKV, ranging from 0 (grey) to 1 (red), showing greater detail for (**B**) the Americas, (**C**) Africa, and (**D**) Asia and Oceania.

DOI: 10.7554/eLife.15272.005

The following figure supplements are available for figure 2:

Figure supplement 1. Uncertainty around Zika suitability predictions displayed in main manuscript – ***Figure 2***, ranging from less than 0.01 (very little uncertainty) to 0.94 (greatest uncertainty).

DOI: 10.7554/eLife.15272.006

Figure supplement 2. Effect plots for each covariate entered into the ensemble of 300 BRT models.

DOI: 10.7554/eLife.15272.007

Figure 2 continued on next page

Figure 2 continued

Figure supplement 3. Environmental suitability for Zika virus transmission to humans, not taking into account temperature suitability for dengue *via Aedes albopictus*.

DOI: 10.7554/eLife.15272.008

Figure supplement 4. Map showing areas predicted to have greater dengue suitability (from ***Bhatt et al., 2013***, Nature) vs those which are predicted to have greater Zika suitability in the current study.

DOI: 10.7554/eLife.15272.009

variety of environmental conditions allows for these predictions to be made at a global scale (***Hay et al., 2006***).

Results

Figure 1A shows the locations of the 323 standardized occurrence records in the final dataset, classified by the following date ranges: (i) up until 2006 (before the outbreak in FSM); (ii) between 2007 (the year of the FSM outbreak) and 2014; and (iii) since 2015, the first reporting of ZIKV in the Americas. This map is accompanied by the graph in ***Figure 1B***, showing the number of reported occurrence locations globally by year. These figures highlight the more sporadic nature of reporting until recent years, with the majority of occurrences in the dataset (63%) coming from the recent 2015–2016 outbreak in Latin America.

The final map that resulted from the mean of 300 ensemble Boosted Regression Tree (BRT) models is shown in ***Figure 2A*** (with greater detail shown for each region in ***Figures 2B–D***). ***Figure 2—figure supplement 1*** shows the distribution of uncertainty based upon the upper and lower prediction quantiles from the 300 models. We restricted our models to make predictions only within areas where i) mosquito vectors (in this case *Ae. aegypti*) were able to persist and ii) where temperature was sufficient for arboviral replication within the mosquito. The former of these was calculated by taking the *Ae. aegypti* probability of occurrence (***Kraemer et al., 2015c***) value that incorporated 90% of all known occurrences (***Kraemer et al., 2015b***) (giving a threshold value of 0.8 and greater) while the latter was evaluated using a mechanistic mosquito model (***Brady et al., 2013***; ***2014***), which identified regions where arboviral transmission could be sustained for at least 355 days (one year

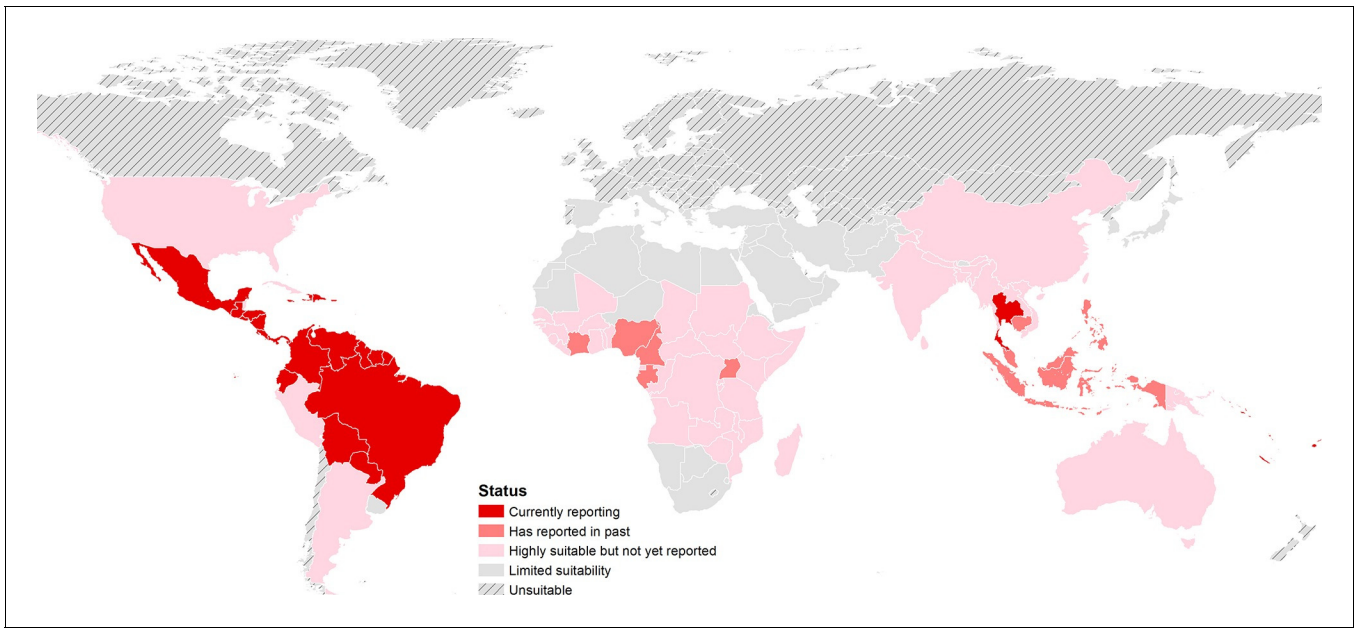


Figure 3. Status of ZIKV reporting as of 2016 by country, showing countries that are highly environmentally suitable (having a suitable area of more than 10,000 square kilometres) but which have not yet reported symptomatic cases of ZIKV in humans. 'Currently reporting' countries are those having reported cases since 2015.

DOI: 10.7554/eLife.15272.010

minus the human incubation period) in an average year. ***Figure 3*** is a country-level map distinguishing between those countries that are currently reporting ZIKV, those which have reported ZIKV in the past, those which have highly suitable areas for transmission, and those which are unsuitable. Our models predicted high levels of risk for ZIKV in many areas within the tropical and sub-tropical zones. Large portions of the Americas are suitable for transmission, with the largest areas of risk occurring in Brazil, followed by Colombia and Venezuela, all of which have reported high numbers of cases in the 2015–2016 outbreak. In Brazil, where the highest numbers of ZIKV are reported in the ongoing epidemic, the coastal cities in the south as well as large areas of the north are identified to have the highest environmental suitability of ZIKV. The central region of Brazil, on the other hand, has low population densities and smaller mosquito populations, which is reflected in the relatively low suitability for ZIKV transmission seen in the map. Although ZIKV has yet to be reported in the USA, a large portion of the southeast region of the country, including much of Texas through to Florida, is also highly suitable for transmission. Potential risk for ZIKV transmission is high in much of sub-Saharan Africa, with continuous suitability in the Democratic Republic of Congo and surrounding areas and several sporadic case reports in western sub-Saharan countries since the 1950s. Although no symptomatic cases have yet been reported in India, a large portion of this country is at potential risk for ZIKV transmission (over 2 million square kilometres), with environmental suitability extending from its northwest regions through to Bangladesh and Myanmar. The Indochina region, southeast China, and Indonesia all have large areas of environmental suitability as well, extending into Oceania. While only representing less than ten percent of Australia's total land area, the area shown to be suitable for ZIKV transmission in its northernmost regions is considerable (comprising nearly 250,000 square kilometres).

Our models showed ZIKV risk to be particularly influenced by annual cumulative precipitation, contributing 65.0% to the variation in the ensemble of models. The next most important predictor in the model was temperature suitability for DENV transmission *via Ae. albopictus,* contributing 14.6%. These are followed by urban extents (8.3%), temperature suitability for DENV *via Ae. aegypti* (5.7%), the Enhanced Vegetation Index (EVI; 3.8%), and minimum relative humidity (2.5%). Effect plots for each covariate are provided in ***Figure 2—figure supplement 2***. Validation statistics indicated high predictive performance of the BRT ensemble mean map evaluated in a 10-fold cross-validation procedure, with area under the receiver operating characteristic (AUC) of 0.829 (± 0.121 SD). Due to

Table 1. Population living in areas suitable for ZIKV transmission within each major world region and top four countries contributing to these populations at risk.

Region/Country	Population living in areas suitable for ZIKV transmission (millions)
Africa	452.58
Nigeria	111.97
Democratic Republic of the Congo	68.95
Uganda	33.43
United Republic of Tanzania	22.70
Americas	298.36
Brazil	120.65
Mexico	32.22
Colombia	29.54
Venezuela	22.22
Asia	1,422.13
India	413.19
Indonesia	226.04
China	213.84
Bangladesh	133.29
World	2,173.27

DOI: 10.7554/eLife.15272.011

the uncertainty about *Ae. albopictus* as a competent vector for ZIKV, we also provide results for an ensemble of models which did not include temperature suitability for dengue *via* this mosquito species in ***Figure 2—figure supplement 3***.

A threshold environmental suitability value of 0.397 in our final map was determined to incorporate 90% of all ZIKV occurrence locations. This was used to classify each 5 km x 5 km pixel on our final map as suitable or unsuitable for ZIKV transmission to humans. Using high-resolution global population estimates (***WorldPop, 2015***; ***SEDAC, 2015***), we summed the populations living in Zika-suitable areas and have identified 2.17 billion people globally living within areas that are environmentally suitable for ZIKV transmission. ***Table 1*** shows a breakdown of this figure by major world region, also showing the top four contributing countries to the potential population at risk. Asia has the most people living in areas that are suitable for ZIKV transmission at 1.42 billion, accounted for in large part by those living in India. In Africa, roughly 453 million people are living in areas suitable for ZIKV transmission, the largest proportion of which live in Nigeria. In the Americas, more than 298 million people live in ZIKV-suitable transmission zones, with approximately 40 percent of these people living in Brazil. Within the majority of environmentally suitable areas for ZIKV in the Americas, prolonged year-round transmission is possible. Southern Brazil and Argentina, however, are more likely to see transmission interrupted throughout the year, as is the case with the USA should autochthonous ZIKV transmission occur there. Using high-resolution data on births for the year 2015 (***WorldPop, 2015***), we also estimate that 5.42 million births will occur in the Americas over the next year within areas and times of environmental suitability for ZIKV transmission.

Discussion

A large number of viruses (circa 219) are known to be pathogenic (***Woolhouse et al., 2012***). Of the 53 species of *Flavivirus*, 19 are reported to have caused illness in humans (***ICTV, 2014***). Some flaviviruses, such as DENV, YFV, Japanese encephalitis virus, and West Nile virus, are widespread, causing many thousands of infections each year. The remainder, however, have been recognized as being pathogenic to humans for decades, but have highly focal reported distributions and are only minor contributors to mortality and disability globally (***Hay et al., 2013***; ***Murray et al., 2015***). As a result, many are of relatively low priority when research and policy interest are considered (***Pigott et al., 2015b***). The recent spread of ZIKV across the globe highlights the need to reassess our consideration of these other flaviviruses, to gain a better understanding of the factors driving their spread and the potential for geographic expansion beyond their currently limited geographical extents.

Environmental suitability for virus transmission in an area does not necessarily mean that it will arrive and/or establish in that location. Arboviral infections in particular are dependent on a variety of non-environmental factors, with their movement having historically been largely attributed to human mobility from travel, trade, and migration, which introduce the viruses to places where mosquito vectors are already present (***Murray et al., 2013***; ***Weaver and Reisen, 2010***; ***Nunes et al., 2015***; ***Gubler and Clark, 1995***). The identification of locations with permissible environments for transmission of emerging diseases like ZIKV is crucial, as importation could give rise to subsequent autochthonous cases in these locations (***Hennessey et al., 2016***; ***Zanluca et al., 2015***). In order to identify places potentially receptive for ZIKV, we assembled the first comprehensive spatial dataset for ZIKV occurrence in humans and compiled a comprehensive set of high-resolution environmental covariates. We then used these data to implement a species distribution modelling approach (***Elith and Leathwick, 2009***) that has proven useful for mapping other vector-borne diseases (***Bhatt et al., 2013***; ***Pigott et al., 2014a***; ***Mylne et al., 2015***; ***Messina et al., 2015b***), allowing us to make inferences about environmental suitability for ZIKV transmission in areas where it has yet to be reported or where we are less certain about its presence. How the ongoing epidemic unfolds in terms of case numbers (or incidence) will depend on a range of other factors such as local transmission dynamics, herd immunity, patterns of contact among mosquitoes and infectious and susceptible humans (***Stoddard et al., 2013***), and mosquito-to-human ratios as recently shown for dengue (***Kraemer et al., 2015a***) and chikungunya (***Salje et al., 2016***).

Globally, we predict that over 2.17 billion people live in areas that are environmentally suitable for ZIKV transmission. We also estimate the number of births occurring in the Americas only, as it is the region for which the most accurate high-resolution population data on births exists (***Tatem et al., 2014***; ***Sorichetta et al., 2015***) and because it is the focus of an ongoing outbreak,

which is the largest recorded thus far. In the Americas alone, an estimated 5.42 million births occurred in 2015 within areas and at times that are suitable for ZIKV transmission. It is important to recognize that not all individuals will be exposed to ZIKV. Like with other flaviviruses, a ZIKV outbreak may be temporally and spatially sporadic and, even in the most receptive environments, is unlikely that all of the population will be infected. Furthermore, increasing herd immunity of this likely sterilizing infection will rapidly reduce the size of the susceptible population at risk for infection in subsequent years (**Dick et al., 1952**) and work is ongoing to predict the likely infection dynamics after establishment. Instead, the estimates are intended as indicators of the total number of individuals or births that may require protection during the first wave of the outbreak. Specifically, these populations should be the focus of efforts to increase awareness and provide guidelines for mitigating personal risk of infection. In future analyses, our estimates could be extended to include ZIKV incidence and the virus' effect on incidence of associated conditions such as Guillain-Barré syndrome and microcephaly. Before appropriately caveated estimates can be generated, however, more information is needed regarding: (i) the background rate of these conditions due to other causes; (ii) how risk may vary throughout the course of a pregnancy; (iii) the proportion of the population exposed during outbreaks; and (iv) whether or not immunity acquired through a mother's prior exposure is protective.

For all arboviral diseases, public health education about reducing populations and avoiding contact with mosquito vectors is required in at-risk areas. Specific to ZIKV is the risk of microcephaly in newborns, which has led public health agencies to issue warnings for women who are currently or planning on becoming pregnant in areas suspected to have ongoing ZIKV transmission and the declaration of a Public Health Emergency of International Concern (**Heymann et al., 2016**). Due to the sensitive nature and implications of these warnings, it is important that levels of risk are rigorously estimated, validated, and updated. Transmission of related arboviral diseases still occurs in many areas we defined as at-risk for ZIKV, which highlights the need for improved vector control outcomes, particularly those targeting *Ae. aegypti*. Predicted levels of risk for ZIKV transmission are potentially helpful for prioritized allocation of vector control resources, as well as for differential diagnosis and, if a vaccine becomes available, delivery efforts. It should be noted that instances of ZIKV sexual transmission have been reported (**Patino-Barbosa et al., 2015**; **Musso et al., 2015b**; **Foy et al., 2011a**). We did not incorporate secondary modes of transmission into the models we described here, but our map can help inform future discussions about the potential impact of this mode of transmission as its relative importance becomes better understood.

A great deal of basic epidemiological information specific to ZIKV is lacking. As a result, information must be leveraged from our knowledge about transmission of related arboviruses. Previous work has focused on mapping other vector borne diseases that share much of the ecology of Zika, such as DENV (**Bhatt et al., 2013**) and CHIKV, as well as for its primary vectors, *Ae. aegypti* and *Ae. albopictus* (**Kraemer et al., 2015c**). For this reason, temperature suitability for dengue (**Brady et al., 2013**, **2014**) was entered into the models due to the greater number of field and laboratory studies available for parameterising this metric for DENV. Until more studies related to vector competence and temperature constraints on ZIKV transmission to humans are conducted, this is the most accurate indicator of arboviral disease transmission *via Aedes* mosquitoes currently available. Indeed, all other covariates in our models could equally be applied to mapping DENV and CHIKV, and ZIKV-specific refinements to modelling covariates will be possible as the disease continues to expand to allow for improvements in future iterations of the map. The relatively smaller amount of occurrence data available for ZIKV (especially prior to recent outbreaks) means that this dataset should also be updated with new information as necessary, leading to a stronger global evidence base and improved accuracy of future maps. Better understanding of ZIKV transmission dynamics will eventually allow for further cartographic refinements to be made, such as the differentiation between endemic- and epidemic-prone areas. Still, all covariates included in the current study have been updated and refined since (**Bhatt et al., 2013**), and when combined with the most extensive occurrence database available for ZIKV, the resulting map we present here is currently the most accurate depiction of the distribution of environmental suitability for ZIKV. A map highlighting differences in predicted suitability for both diseases is provided in ***Figure 2—figure supplement 4***.

Conclusion

In this study, we produced the first global high spatial-resolution map of environmental suitability for ZIKV transmission to humans using an assembly of known records of ZIKV occurrence and environmental covariates in a species distribution modelling framework. While it is clear that much remains to be understood about ZIKV, this first map serves as a baseline for understanding the change in the geographical distribution of this globally emerging arboviral disease. Knowledge of the potential distribution can encourage more vigilant surveillance in both humans and *Aedes* mosquito populations, as well as help in the allocation of limited resources for disease prevention. Public health awareness campaigns and advice for mitigation of individual risk can also be focused in the areas we have predicted to be highly suitable for ZIKV transmission, particularly during the first wave of infection in a population. The maps we have presented may also inform existing travel advisories for pregnant women and other travellers. The maps and underlying data are freely available online *via* figshare (http://www.figshare.com).

Materials and methods

To map environmental suitability for ZIKV transmission to humans, we applied a species distribution modelling approach to establish a multivariate empirical relationship between the probability of ZIKV occurrence and the environmental conditions in locations where the disease has been confirmed. We employed an ensemble boosted regression trees (BRT) methodology (***De'ath, 2007***; ***Elith et al., 2008***), which required the generation of: (i) a comprehensive compendium of known locations of disease occurrence in humans; (ii) a set of background points representing locations where ZIKV has not yet been reported; and (iii) a set of high-resolution globally gridded environmental and socioeconomic covariates hypothesised to affect ZIKV transmission. The resulting model produces a 5 x 5 km spatial-resolution global map of environmental suitability for ZIKV transmission to humans.

Assembly of the geo-referenced ZIKV occurrence dataset

Information about the locations of ZIKV occurrence in humans was extracted from peer-reviewed literature, case reports, and informal online sources following previously established protocols (***Kraemer et al., 2015b***; ***Messina et al., 2014***; ***2015a***). To collate the peer-reviewed dataset, literature searches were undertaken using PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and ISI Web of Science (http://www.webofknowledge.com) search engines using the search term 'Zika'. No language restrictions were placed on these searches; however, only those citations with a full title and abstract were retrieved, resulting in the review of 148 references ranging in publication dates between 1951 and 2015. In-house language skills allowed review of all English, French, Portuguese and Spanish articles for useable location information for human ZIKV occurrence. ProMED-mail (http://www.promedmail.org) was also searched using the term 'Zika', resulting in the review of 139 reports between 27 June 2007 and 18 January 2016. Additionally, the most current database of ZIKV case locations in Brazil was obtained directly from the Brazilian Ministry of Health. From all sources, only laboratory confirmation of symptomatic ZIKV infection in humans was entered into the dataset (mention of suspected cases was not entered). Serological evidence from healthy individuals could represent a past infection, with transmission potentially occurring in a different location to that where the individual currently resides (***Darwish et al., 1983***), or could be an artefact from possible cross-reactivity with a variety of different viruses (***Smithburn et al., 1954***). As a result, these less reliable diagnoses of ZIKV were excluded.

All available location information was extracted from each peer-reviewed article and ProMED case report. The site name was used together with all contextual information provided about the site to determine its latitudinal and longitudinal coordinates using Google Maps (https://www.maps.google.com). If the study site could be geo-positioned to a specific place, it was recorded as a point location. If the study site could only be identified at an administrative area level (e.g. province or district), it was recorded as a polygon along with an identifier of its administrative unit. If imported cases were reported with information on the site of infection, they were geo-positioned to this site; if imported cases were reported with no information about the site of infection, they were not entered into the dataset. Informal online data sources were collated automatically by the web-based

system HealthMap (http://www.healthmap.org) as described elsewhere (**Freifeld et al., 2008**). Alerts for ZIKV were obtained from HealthMap for the years 2014–2016, and then manually checked for validity. In total, usable location information was extracted from 110 sources. Information was also collected about the status of symptoms in each reported occurrence, distinguishing between those where symptomatic cases were being reported, versus those where only seroprevalence was detected in healthy individuals.

Due to the potential for multiple independent reports referring to the same cases temporal and spatial standardization was required, as we have described previously in detail for dengue mapping efforts (**Messina et al., 2014**). In brief, an occurrence was defined as a unique location with one or more confirmed cases of ZIKV occurring within one calendar year (the finest temporal resolution available across all records). Point locations were considered to be overlapping if they lay on the same 5 km x 5 km pixel, and polygon locations were identified by a unique administrative unit code. Furthermore, all polygons whose geographic area was greater than one square decimal degree (approximately 111 square kilometers at the equator) were removed from the dataset to avoid averaging covariate values over very large areas, and only those occurrences comprising symptomatic individuals were retained for modelling purposes to ensure an accurate location of infection. In total, the final occurrence dataset contained 323 unique occurrences to be entered into our BRT modelling procedure. A map of the final set of occurrence locations is provided as **Figure 1A**.

Generation of the background location dataset

Separate maps of the relative probability of occurrence of *Ae. aegypti* and *Ae. albopictus* (**Kraemer et al., 2015c**) were used to compute a combined metric of the relative probability of vector occurrence, by taking the maximum value from the two layers for all 5 km x 5 km gridded cells globally. The inverse of this combined-*Aedes* occurrence probability layer (higher values indicating greater certainty of absence) was then used to draw a biased sample of 10,000 background locations. As such, a greater number of background points were sampled in areas where we are more certain that *Ae. aegypti* or *Ae. albopictus* do not occur, and therefore where ZIKV is less likely to be transmitted to humans. While it has been demonstrated that predictive accuracy from presence-background species distribution models can be improved by biasing background record locations toward areas with greatest reporting probabilities (**Phillips et al., 2009**), information on possible reporting biases, or proxies of such spatial bias, are currently unavailable for ZIKV. These 10,000 background locations were combined with the standardized occurrence dataset to serve as comparison data locations in the BRT species distribution modelling procedure. The background locations were weighted such that their total sum was equal to the total number of occurrence locations (n=237; pseudo-absence weighting=0.0237), in order to aid in the discrimination capacity of the model (**Barbet-Massin et al., 2012**).

Explanatory covariates

A set of six covariates hypothesized to influence the global distribution of ZIKV transmission to humans were used in our models to establish an empirical relationship between ZIKV presence or absence and underlying environmental conditions. These six covariates included: (i) an index of temperature suitability for dengue transmission to humans *via Ae. aegypti;* (ii) temperature suitability for dengue transmission to humans *via Ae. albopictus*; (iii) minimum relative humidity; (iv) annual cumulative precipitation; (v) an enhanced vegetation index (EVI); and (vi) urban versus rural habitat type. The underlying hypothesis behind each of the covariates is discussed in more detail below, along with a description of data sources and any processing that was undertaken before entering these covariates into our models. Maps of each covariate layer are provided in the supplementary information in **Figure 1—figure supplement 1**.

Temperature suitability for dengue transmission to humans

via Ae. aegypti *or* Ae. albopictus: Temperature affects key physiological processes in *Aedes* mosquitoes, including age- and temperature-dependent adult female survival, as well as the duration of the extrinsic incubation period (EIP) of arboviruses and the length of the gonotrophic cycle (**Brady et al., 2013**). While these parameters have yet to be measured experimentally for ZIKV, they have been for the closely related DENV. We obtained temperature data from WorldClim v1.03 (http://www.

wordclim.org), which uses historic global meteorological station data from 1961–2005 to interpolate global climate surfaces. MARKSIM software (**Jones and Thornton, 2000**) was then used to apply the coefficients of 17 Global Climate Models (GCMs) to estimate temperature values for the year 2015. This enabled us to incorporate the quantified effects of temperature on DENV transmission into a cohort simulation model that analysed the cumulative effects of both diurnal and inter-seasonal changes in temperature on DENV transmission within an average year, both for *Ae. aegypti* and *Ae. albopictus* separately. The models were then applied to the 2015 temperature data for each 5 km x 5 km grid cell globally. This resulted in maps of temperature suitability for DENV transmission by either *Aedes* species ranging from 0 (no suitable days) to 1 (365 suitable days). These measures were then used as a proxy for temperature suitability for ZIKV transmission to humans.

Annual cumulative precipitation

Presence of static surface water in natural or man-made containers is a pre-requisite for *Aedes* oviposition and larval and pupal development. While fine-scale spatial and temporal heterogeneities have been observed between precipitation, vector abundance, and incidence of human DENV infections, there is evidence that areas with greater amounts of precipitation are generally associated with higher DENV infection risk (**Chandy et al., 2013**; **Chowell and Sanchez, 2006**; **Dom et al., 2013**; **Pinto et al., 2011**; **Restrepo et al., 2014**; **Sang et al., 2014**; **Sankari et al., 2012**; **Campbell et al., 2015**). Although studies that directly connect levels of precipitation to ZIKV transmission have yet to exist, we assumed for Zika a similar association of precipitation as closely related flaviruses. WorldClim v1.03 precipitation data and MARKSIM software were used as described above for temperature, to estimate annual cumulative precipitation for the year 2015 for each 5 km x 5 km grid cell globally.

Minimum relative humidity

Greater relative humidity has been found to promote DENV propagation in *Ae. aegypti* mosquitoes in several localized settings (**Colón-González et al., 2011**; **Thu et al., 1998**), and has also been found to be an important contributor when predicting DENV risk at a global scale (**Hales et al., 2002**). Therefore, we again assumed a similar association for ZIKV in the absence of any direct studies, and included the minimum annual relative humidity in our models as a potential limiting factor to ZIKV transmission. Relative humidity (RH) was calculated as a percent of saturation humidity, or the amount of water vapour required to saturate the air given a particular temperature, using the temperature data from WorldClim v1.03 described earlier. The saturation, or 'dew', point (T_{dew}) was calculated using a tabular relationship (**Linacre, 1977**). RH was then calculated as follows:

$$RH = \frac{V(T_x)}{V(T_{dew})} \times 100$$

Where $V(T_{dew}) = 611.21 \times \exp\left(17.502 \times \frac{\mathrm{T}}{240.97+\mathrm{T}}\right)$ and $V(T_x)$ is the humidity at the given temperature. We then extracted the minimum annual RH for each 5 km x 5 km pixel globally for the year 2015.

Enhanced vegetation index (EVI)

A close association has been shown between local moisture supply, vegetation canopy development, and abundance of mosquito reproduction (**Linthicum et al., 1999**), with previous studies highlighting the importance of moisture-related measures such as relative humidity to DENV occurrence (**Hales et al., 2002**). Although resistant to desiccation, both *Aedes* eggs and adults require moisture to survive (**Cox et al., 2007**; **Sota and Mogi, 1992**; **Reiskind and Lounibos, 2009**; **Costa et al., 2010**; **Luz et al., 2008**), with low dry season moisture levels substantially affecting *Aedes* mortality (**Russell et al., 2001**; **Trpis, 1972**; **Luz et al., 2008**). Vegetation canopy cover has previously been associated with higher *Aedes* larvae density (**Fuller et al., 2009**; **Troyo et al., 2009**; **Bisset Lazcano et al., 2006**; **Barrera et al., 2006**) by reducing evaporation from containers, decreasing sub-canopy wind speed, and protecting outdoor habitats from direct sunlight. To account for these factors, we included a 5 km x 5 km resolution measure of the EVI derived from NASA's Moderate Resolution Imaging Spectrometer (MODIS, Boulder, Colorado) imagery (**Wan et al., 2002**; **Lin, 2012**), summarized from gap-filled, 8-day, 1 km x 1 km resolution images

acquired globally for years 2000 through 2014 (**Weiss et al., 2014**) to produce a mean annual EVI layer. This mean EVI product is indicative of amount of photosynthesis taking place in the environment over the course of a year, which is positively correlated with the density of vegetation, and is thus a proxy for the level of moisture available given the relationship between precipitation and vegetative growth.

Urban versus rural habitat type

There is a well-established link between urban areas, some vector borne diseases, and their vectors. In particular, *Ae. aegypti* is found in close proximity to human dwellings often breeding in artificial containers (**Brown et al., 2011**; **Powell and Tabachnick, 2013**; **Kraemer et al., 2015c**). To identify the relationship between urbanisation and ZIKV presence we adapted probabilistic spatial modelling techniques to predict the spatial distribution of global urban extents at a 5 km x 5 km spatial resolution. We used urban growth rates from the United Nations Population Division (**Division, 2014**), paired with urban extents measured and tested by the Moderate Resolution Imaging Spectroradiometer Collection 5 (MODIS C5) land-cover product for Asia (**Schneider et al., 2009**; **2010**; **2015**). A set of spatial covariate datasets hypothesized to influence the spatial patterns of urban expansion was generated, including the time to travel from each 5 km x 5 km pixel to a major city (**Nelson, 2008**), the proportion of urbanised land within a buffer of 20 km, human population density (**Linard and Tatem, 2012**; **Stevens et al., 2015**; **Gaughan et al., 2013**), slope (**Becker et al., 2009**), and distance to water (**Arino et al., 2008**). A BRT modelling approach was then used to predict areas that would become urban in 2015 (**Linard et al., 2013**). Outputs were tested against a training dataset comprising points from Asia only, and showed good overall predictive performance (AUC=0.82). The output raster is a 5 km x 5 km gridded surface with urban (1) vs. rural (0) pixels.

Ensemble boosted regression trees approach

The boosted regression tree (BRT) modelling procedure combines regression trees with gradient boosting (**Friedman, 2001**). In this procedure, an initial regression tree is fitted and iteratively improved upon in a forward stagewise manner (boosting) by minimising the variation in the response not explained by the model at each iteration. It has been shown to fit complicated response functions efficiently, while guarding against over-fitting by use of extensive internal cross-validation. As such, this approach has been successfully employed in the past to map dengue and its *Aedes* mosquito vectors, as well as other vector-borne diseases (**Bhatt et al., 2013**; **Pigott et al., 2014b**; **Messina et al., 2015b**; **Kraemer et al., 2015c**). To increase the robustness of model predictions and quantify model uncertainty, we fitted an ensemble (**Araújo and New, 2007**) of 300 BRT models to separate bootstraps of the data. We then evaluated the central tendency as the mean across all 300 BRT models (**Bhatt et al., 2013**). Each of the 300 individual models was fitted using the gbm. step subroutine in the dismo package in the R statistical programming environment (**Elith et al., 2008**). All other tuning parameters of the algorithm were held at their default values (tree complexity= 4, learning rate= 0.005, bag fraction= 0.75, step size= 10, cross-validation folds=10). Each of the 300 models predicts environmental suitability on a continuous scale from 0 to 1, with a final prediction map then being generated by calculating the mean prediction across all models for each 5 km x 5 km pixel. Cross-validation was applied to each model, whereby ten subsets of the data comprising 10% of the presence and background observations were assessed based on their ability to predict the distribution of the other 90% of records using the mean area under the curve (AUC) statistic. This AUC value was then averaged across the ten sub-models and finally across all 300 models in the ensemble in order to derive an overall estimate of goodness-of-fit. Additionally, to avoid AUC inflation due to spatial sorting bias, a pairwise distance sampling procedure was used, resulting in a final AUC which is lower than would be returned by standard procedures but which gives a more realistic quantification of the model's ability to extrapolate predictions to new regions (**Wenger and Olden, 2012**). We restricted our models to make predictions only within areas where either *Ae. aegypti* probability of occurrence (**Kraemer et al., 2015c**) is more than 0.8 or temperature is conducive to transmission for at least 355 days in an average year. A second ensemble of 300 models was executed which did not take into account temperature suitability for dengue transmission *via Ae. albopictus*, due to the uncertainty of this species as a competent ZIKV vector. The results of this ensemble of models are provided in ***Figure 2—figure supplement 3***.

Population and births at risk

To calculate the number of people located in an area that is at any level of risk for ZIKV transmission, the global ZIKV environmental suitability map was combined with fine-scale global population surfaces (**SEDAC, 2015**; **WorldPop, 2015**). Firstly, the continuous ZIKV environmental suitability map (ranging from 0 to 1) was converted into a binary surface indicating whether there is any risk of transmission. To do this, we carried out a protocol previously used in (**Pigott et al., 2015a**), choosing a threshold environmental suitability value that encompasses 90% of the ZIKV occurrence point locations. This threshold cut-off of 90% was chosen (rather than 100%) to reflect potential errors or inaccurate locations in the occurrence point dataset. Every 5 km x 5 km pixel in the suitability map with a value above this threshold value was considered at risk for ZIKV transmission. Finally, to estimate the population at risk, we multiplied this binary ZIKV risk map by the global population counts (aligned and aggregated to the same 5 x 5 km grid) for the year 2015 and summed across all cells.

We next estimated the maximum number of births potentially affected by ZIKV in Latin America, as this region is the focus of the recent outbreak and the first to point to a possible association with microcephaly in newborn infants to mothers infected with ZIKV. In order to do this, we first identified the proportion of the year that is suitable for ZIKV transmission within areas that are predicted to be suitable in the binary ZIKV risk map. This proportion was derived from existing temperature suitability models (**Brady et al., 2013**; **2014**), which predict the total number of days within an average year that arbovirus transmission can be sustained in *Ae. aegypti*, assuming there is a local human reservoir of infection. While the intra-mosquito viral dynamics in this model were parameterised for dengue virus, the limited information currently available on other arboviruses suggests that their dynamics are similar (**Lambrechts et al., 2011**). Using the resulting 5 km x 5 km map showing the proportion of the year suitable for ZIKV transmission to humans, we then multiplied this by a map (also at a 5 km x 5 km resolution) of the number of births in the Americas for the year 2015, updated from (**Tatem et al., 2014**; **UNFPA, 2014**). The resulting map indicates the number of births in the Americas potentially at risk for ZIKV (for 2015), assuming ZIKV currently fully occupies its environmental niche and that births are evenly distributed throughout the year.

Acknowledgements

We thank the Secretariat of Health Surveillance, Ministry of Health of Brazil for providing access to the geographical coordinates of occurrence. JPM, MUGK and TJ receive, and OJB and SIH acknowledge funding from the International research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS; European Commission 7th Framework Programme (21893)). OJB and SIH are supported by the Bill & Melinda Gates Foundation (OPP1053338). SIH is also funded by a Senior Research Fellowship from the Wellcome Trust (095066), and grants from the Bill & Melinda Gates Foundation (OPP1119467, OPP1106023 and OPP1093011). DMP is also funded by the Bill & Melinda Gates Foundation (OPP1093011). DJW and PWG receive support from the Bill and Melinda Gates Foundation (OPP1068048, OPP1106023). NG is supported by a University of Melbourne McKenzie fellowship. CWR is funded through the University of Southampton's Economic and Social Research Council's Doctoral Training Centre. TWS is supported by grants from the National Institutes of Health (P01AI098670) and the Bill and Melinda Gates Foundation (OPP1081737). EC and JSB are supported by the National Library of Medicine of the National Institutes of Health (R01LM010812).

Additional information

Competing interests

SIH: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
European Commission	21893	Jane P Messina Moritz UG Kraemer Thomas Jaenisch

Bill and Melinda Gates Foundation	OPP1053338; OPP1119467; OPP1106023; OPP1093011; OPP1081737; OPP1068048	Oliver J Brady David M Pigott Daniel J Weiss Thomas W Scott Simon I Hay
Wellcome Trust	095066	Simon I Hay
University of Southampton	Economic and Social Research Council's Doctoral Training Centre	Corrine W Ruktanonchai
National Institutes of Health	P01AI098670	Thomas W Scott
University of Melbourne	McKenzie fellowship	Nick Golding

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

JPM, DMP, SIH, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; MUGK, OJB, NG, Conception and design, Analysis and interpretation of data, Drafting or revising the article; FMS, PWG, Conception and design, Acquisition of data, Drafting or revising the article; DJW, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; CWR, EC, AJT, Acquisition of data, Analysis and interpretation of data; JSB, CJLM, Acquisition of data, Drafting or revising the article; KK, TJ, TWS, Analysis and interpretation of data, Drafting or revising the article; FM, Acquisition of data, Contributed unpublished essential data or reagents

Author ORCIDs

Jane P Messina, http://orcid.org/0000-0001-7829-1272
Moritz UG Kraemer, http://orcid.org/0000-0001-8838-7147
Simon I Hay, http://orcid.org/0000-0002-0611-7272

Additional files

Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Jane P Messina, Freya M Shearer	2016	Global compendium of human Zika virus occurrence	http://dx.doi.org/10.6084/m9.figshare.2573629	Publicly available at figshare
Jane P Messina, Moritz UG Kraemer, Oliver J Brady, David M Pigott, Freya M Shearer, Daniel J Weiss, Nick Golding, Corrine W Ruktanonchai, Emily Cohn, John S Brownstein, Kamran Khan, Andrew J Tatem, Thomas Jaenisch, Thomas W Scott, Simon I Hay	2016	Environmental suitability for Zika virus transmission	http://dx.doi.org/10.6084/m9.figshare.2574298	Publicly available at figshare

References

Araújo MB, New M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* **22**:42–47. doi: 10.1016/j.tree.2006.09.010

Arino O, Bicheron P, Achard F, Latham J, Witt R, Weber JL. 2008. GLOBCOVER the most detailed portrait of earth. *ESA Bulletin*:24–31.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution* **3**:327–338. doi: 10.1111/j.2041-210X.2011.00172.x

Barrera R, Amador M, Clark GG. 2006. Use of the pupal survey technique for measuring Aedes aegypti (Diptera: Culicidae) productivity in Puerto Rico. *The American Journal of Tropical Medicine and Hygiene* **74**:290–302.

Becker JJ, Sandwell DT, Smith WHF, Braud J, Binder B, Depner J, Fabre D, Factor J, Ingalls S, Kim S-H, Ladner R, Marks K, Nelson S, Pharaoh A, Trimmer R, Von Rosenberg J, Wallace G, Weatherall P. 2009. Global bathymetry and elevation data at 30 arc seconds resolution: Srtm30_plus. *Marine Geodesy* **32**:355–371. doi: 10.1080/01490410903297766

Benedict MQ, Levine RS, Hawley WA, Lounibos LP. 2007. Spread of the tiger: Global risk of invasion by the mosquito aedes albopictus. *Vector Borne and Zoonotic Diseases* **7**:76–85. doi: 10.1089/vbz.2006.0562

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. 2013. The global distribution and burden of dengue. *Nature* **496**:504–507. doi: 10.1038/nature12060

Bisset Lazcano JA, Marquetti MC, Portillo R, Rodríguez MM, Suárez S, Leyva M. 2006. Ecological factors linked to the presence of Aedes aegypti larvae in highly infested areas of Playa, a municipality belonging to Ciudad de La Habana, Cuba. *Revista Panamericana de Salud Pública* **19**:379–384. doi: 10.1590/S1020-49892006000600003

Bogoch II, Brady OJ, Kraemer MU, German M, Creatore MI, Kulkarni MA, Brownstein JS, Mekaru SR, Hay SI, Groot E, Watts A, Khan K. 2016. Anticipating the international spread of zika virus from brazil. *Lancet* **387**:335–336. doi: 10.1016/S0140-6736(16)00080-5

Brady OJ, Golding N, Pigott DM, Kraemer MU, Messina JP, Reiner RC, Scott TW, Smith DL, Gething PW, Hay SI. 2014. Global temperature constraints on aedes aegypti and ae. albopictus persistence and competence for dengue virus transmission. *Parasites & Vectors* **7**:338. doi: 10.1186/1756-3305-7-338

Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, Delatte H, Grech MG, Leisnham PT, Maciel-de-Freitas R, Styer LM, Smith DL, Scott TW, Gething PW, Hay SI. 2013. Modelling adult aedes aegypti and aedes albopictus survival at different temperatures in laboratory and field settings. *Parasites & Vectors* **6**:351. doi: 10.1186/1756-3305-6-351

Brown JE, McBride CS, Johnson P, Ritchie S, Paupy C, Bossin H, Lutomiah J, Fernandez-Salas I, Ponlawat A, Cornel AJ, Black WC, Gorrochotegui-Escalante N, Urdaneta-Marquez L, Sylla M, Slotman M, Murray KO, Walker C, Powell JR. 2011. Worldwide patterns of genetic differentiation imply multiple 'domestications' of aedes aegypti, a major vector of human diseases. *Proceedings of the Biological Sciences* **278**:2446–2454. doi: 10.1098/rspb.2010.2469

Campbell KM, Haldeman K, Lehnig C, Munayco CV, Halsey ES, Laguna-Torres VA, Yagui M, Morrison AC, Lin CD, Scott TW. 2015. Weather regulates location, timing, and intensity of dengue virus transmission between humans and mosquitoes. *PLoS Neglected Tropical Diseases* **9**:e0003957. doi: 10.1371/journal.pntd.0003957

Chandy S, Ramanathan K, Manoharan A, Mathai D, Baruah K. 2013. Assessing effect of climate on the incidence of dengue in tamil nadu. *Indian Journal of Medical Microbiology* **31**:283–286. doi: 10.4103/0255-0857.115640

Chowell G, Sanchez F. 2006. Climate-based descriptive models of dengue fever: the 2002 epidemic in Colima, Mexico. *Journal of Environmental Health* **68**:40–44.

Colón-González FJ, Lake IR, Bentham G. 2011. Climate variability and dengue fever in warm and humid mexico. *The American Journal of Tropical Medicine and Hygiene* **84**:757–763. doi: 10.4269/ajtmh.2011.10-0609

Costa E, Santos EMM, Correia JC, Albuquerque CMR. 2010. Impact of small variations in temperature and humidity on the reproductive activity and survival of aedes aegypti (diptera, culicidae). *Rev Bras Entomol* **54**:488–493 . doi: 10.1590/s0085-56262010000300021

Cox J, Grillet ME, Ramos OM, Amador M, Barrera R. 2007. Habitat segregation of dengue vectors along an urban environmental gradient. *The American Journal of Tropical Medicine and Hygiene* **76**:820–826.

Darwish MA, Hoogstraal H, Roberts TJ, Ahmed IP, Omar F. 1983. A sero-epidemiological survey for certain arboviruses (togaviridae) in pakistan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77**:442–445 . doi: 10.1016/0035-9203(83)90106-2

De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**:243–251 . doi: 10.1890/0012-9658(2007)88[243:btfema]2.0.co;2

Dick GW, Kitchen SF, Haddow AJ. 1952. Zika virus. I. isolations and serological specificity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **46**:509–520. doi: 10.1016/0035-9203(52)90042-4

Dick GW. 1953. Yellow fever; a problem in epidemiology. *British Medical Bulletin* **9**:215–235.

Dom NC, Ahmad AH, Latif ZA, Ismail R, Pradhan B. 2013. Coupling of remote sensing data and environmental-related parameters for dengue transmission risk assessment in subang jaya, malaysia. *Geocarto International* **28**:258–272. doi: 10.1080/10106049.2012.696726

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *The Journal of Animal Ecology* **77**:802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Elith J, Leathwick JR. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annu Rev Ecol Evol S* **40**:677–697 . doi: 10.1146/annurev.ecolsys.110308.120159

Faria NR, Azevedo RS, Kraemer MU, Souza R, Cunha MS, Hill SC, Thézé J, Bonsall MB, Bowden TA, Rissanen I, Rocco IM, Nogueira JS, Maeda AY, Vasami FG, Macedo FL, Suzuki A, Rodrigues SG, Cruz AC, Nunes BT,

Medeiros DB, et al. 2016. Zika virus in the americas: Early epidemiological and genetic findings. *Science* **352**. doi: 10.1126/science.aaf5036

Faye O, Faye O, Dupressoir A, Weidmann M, Ndiaye M, Alpha Sall A. 2008. One-step RT-PCR for detection of zika virus. *Journal of Clinical Virology* **43**:96–101. doi: 10.1016/j.jcv.2008.05.005

Foy BD, Kobylinski KC, Foy JLC, Blitvich BJ, Travassos da Rosa A, Haddow AD, Lanciotti RS, Tesh RB. 2011a. Probable non–vector-borne transmission of zika virus, colorado, USA. *Emerging Infectious Diseases* **17**:880–882. doi: 10.3201/eid1705.101939

Foy BD, Kobylinski KC, Foy JLC, Blitvich BJ, Travassos da Rosa A, Haddow AD, Lanciotti RS, Tesh RB. 2011b. Probable non–vector-borne transmission of zika virus, colorado, USA. *Emerging Infectious Diseases* **17**:880–882. doi: 10.3201/eid1705.101939

Freifeld CC, Mandl KD, Reis BY, Brownstein JS. 2008. Healthmap: Global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association* **15**:150–157. doi: 10.1197/jamia.M2544

Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29**: 1189–1232. doi: 10.1214/aos/1013203451

Fuller DO, Troyo A, Beier JC. 2009. El niño southern oscillation and vegetation dynamics as predictors of dengue fever cases in costa rica. *Environmental Research Letters* **4**:014011. doi: 10.1088/1748-9326/4/1/014011

Gaughan AE, Stevens FR, Linard C, Jia P, Tatem AJ. 2013. High resolution population distribution maps for southeast asia in 2010 and 2015. *PloS One* **8**:e55882. doi: 10.1371/journal.pone.0055882

Gatherer D, Kohl A. 2016. Zika virus: A previously slow pandemic spreads rapidly through the americas. *The Journal of General Virology* **97**:269–342. doi: 10.1099/jgv.0.000381

Grard G, Caron M, Mombo IM, Nkoghe D, Mboui Ondo S, Jiolle D, Fontenille D, Paupy C, Leroy EM. 2014. Zika virus in gabon (central africa)–2007: A new threat from aedes albopictus? *PLoS Neglected Tropical Diseases* **8**: e2681. doi: 10.1371/journal.pntd.0002681

Gubler DJ, Clark GG. 1995. Dengue/dengue hemorrhagic fever: The emergence of a global health problem. *Emerging Infectious Diseases* **1**:55–57. doi: 10.3201/eid0102.952004

Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, Guzman H, Tesh RB, Weaver SC. 2012. Genetic characterization of zika virus strains: Geographic expansion of the asian lineage. *PLoS Neglected Tropical Diseases* **6**. doi: 10.1371/journal.pntd.0001477

Hales S, de Wet N, Maindonald J, Woodward A. 2002. Potential effect of population and climate changes on global distribution of dengue fever: An empirical model. *Lancet* **360**:830–834. doi: 10.1016/S0140-6736(02)09964-6

Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, Brownstein JS, Collier N, Myers MF, George DB, Gething PW. 2013. Global mapping of infectious disease. *Philosophical Transactions of the Royal Society B* **368**:20120250. doi: 10.1098/rstb.2012.0250

Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. 2006. Global environmental data for mapping infectious disease distribution. *Advances in Parasitology* **62**:37–77. doi: 10.1016/S0065-308X(05)62002-7

Hennessey M, Fischer M, Staples JE. 2016. Zika virus spreads to new areas - region of the americas, may 2015-january 2016. *Morb Mortal Wkly Rep* **65**:55–58 . doi: 10.1111/ajt.13743

Heymann DL, Hodgson A, Sall AA, Freedman DO, Staples JE, Althabe F, Baruah K, Mahmud G, Kandun N, Vasconcelos PFC, Bino S, Menon KU. 2016. Zika virus and microcephaly: Why is this situation a PHEIC? *The Lancet* **387**:719–721. doi: 10.1016/S0140-6736(16)00320-2

ICTV. 2014. *Virus taxonomy: 2014 release*. 1st Edition.

Jones PG, Thornton PK. 2000. Marksim: Software to generate daily weather data for Latin America and Africa. *Agronomy Journal* **92**:445–453. doi: 10.2134/agronj2000.923445x

Kraemer MU, Hay SI, Pigott DM, Smith DL, Wint GR, Golding N. 2016. Progress and challenges in infectious disease cartography. *Trends in Parasitology* **32**:19–29. doi: 10.1016/j.pt.2015.09.006

Kraemer MUG, Perkins TA, Cummings DAT, Zakar R, Hay SI, Smith DL, Reiner RC. 2015a. Big city, small world: Density, contact rates, and transmission of dengue across pakistan. *Journal of the Royal Society Interface* **12**: 20150468. doi: 10.1098/rsif.2015.0468

Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H-J, Hay SI. 2015b. The global compendium of aedes aegypti and ae. albopictus occurrence. *Scientific Data* **2**:150035. doi: 10.1038/sdata.2015.35

Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IRF, Teng H-J, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GRW, et al. 2015c. The global distribution of the arbovirus vectors aedes aegypti and ae. albopictus. *eLife* **4**:e08347. doi: 10.7554/eLife.08347

Kutsuna S, Kato Y, Moi ML, Kotaki A, Ota M, Shinohara K, Kobayashi T, Yamamoto K, Fujiya Y, Mawatari M, Sato T, Kunimatsu J, Takeshita N, Hayakawa K, Kanagawa S, Takasaki T, Ohmagari N. 2015. Autochthonous dengue fever, tokyo, japan, 2014. *Emerging Infectious Diseases* **21**:517. doi: 10.3201/eid2103.141662

Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, Thomas MB, Scott TW. 2011. Impact of daily temperature fluctuations on dengue virus transmission by aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America* **108**:7460–7465. doi: 10.1073/pnas.1101377108

Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, Stanfield SM, Duffy MR. 2008. Genetic and serologic properties of zika virus associated with an epidemic, yap state, micronesia, 2007. *Emerging Infectious Diseases* **14**:1232–1241. doi: 10.3201/eid1408.080287

Lin QH. 2012. Enhanced vegetation index using moderate resolution imaging spectroradiometers. *In 5th International Congress on Image and Signal Processing* :1043–1046 . doi: 10.1109/cisp.2012.6470008

Linacre ET. 1977. Simple formula for estimating evaporation rates in various climates using temperature data alone. *Agr Meteorol* **18**:409–424 . doi: 10.1016/0002-1571(77)90007-3

Linard C, Tatem AJ, Gilbert M. 2013. Modelling spatial patterns of urban growth in africa. *Applied Geography* **44**:23–32. doi: 10.1016/j.apgeog.2013.07.009

Linard C, Tatem AJ. 2012. Large-scale spatial population databases in infectious disease research. *Int J Health Geogr* **11**. doi: 10.1186/1476-072x-11-7

Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ. 1999. Climate and satellite indicators to forecast rift valley fever epidemics in kenya. *Science* **285**:397–400 . doi: 10.1126/science.285.5426.397

Luz C, Tai MH, Santos AH, Silva HH. 2008. Impact of moisture on survival of aedes aegypti eggs and ovicidal activity of metarhizium anisopliae under laboratory conditions. *Memo'Rias Do Instituto Oswaldo Cruz* **103**:214–215 . doi: 10.1590/s0074-02762008000200016

Marchette NJ, Garcia R, Rudnick A. 1969. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. *The American Journal of Tropical Medicine and Hygiene* **18**:411–415.

Marcondes CB, Ximenes MF. 2015. Zika virus in brazil and the danger of infestation by aedes (stegomyia) mosquitoes. *Revista Da Sociedade Brasileira De Medicina Tropical*. doi: 10.1590/0037-8682-0220-2015

Martines RB, Bhatnagar J, Keating MK, Silva-Flannery L, Muehlenbachs A, Gary J, Goldsmith C, Hale G, Ritter J, Rollin D, Shieh WJ, Luz KG, Ramos AM, Davi HP, Kleber de Oliveria W, Lanciotti R, Lambert A, Zaki S. 2016. Notes from the field: Evidence of zika virus infection in brain and placental tissues from two congenitally infected newborns and two fetal losses - brazil, 2015. *MMWR. Morbidity and Mortality Weekly Report* **65**:159–219. doi: 10.15585/mmwr.mm6506e1

McCarthy M. 2016. First US case of zika virus infection is identified in texas. *BMJ* **352**:i212. doi: 10.1136/bmj.i212

Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. 2014. A global compendium of human dengue virus occurrence. *Scientific Data* **1**:140004. doi: 10.1038/sdata.2014.4

Messina JP, Pigott DM, Duda KA, Brownstein JS, Myers MF, George DB, Hay SI. 2015a. A global compendium of human crimean-congo haemorrhagic fever virus occurrence. *Scientific Data* **2**:150016. doi: 10.1038/sdata.2015.16

Messina JP, Pigott DM, Golding N, Duda KA, Brownstein JS, Weiss DJ, Gibson H, Robinson TP, Gilbert M, William Wint GR, Nuttall PA, Gething PW, Myers MF, George DB, Hay SI. 2015b. The global distribution of crimean-congo hemorrhagic fever. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**:503–513. doi: 10.1093/trstmh/trv050

Monlun E, Zeller H, Le Guenno B, Traore-Lamizana M, Hervy J, Adam F, Ferrara L, Fontenille D, Sylla R, Mondo M. 1992. Surveillance of the circulation of arbovirus of medical interest in the region of eastern Senegal. *Bulletin de la Société de pathologie exotique* **86**:21–28.

Morrison AC, Zielinski-Gutierrez E, Scott TW, Rosenberg R. 2008. Defining challenges and proposing solutions for control of the virus vector aedes aegypti. *PLoS Medicine* **5**:e68. doi: 10.1371/journal.pmed.0050068

Murray CJ, Barber RM, Foreman KJ, Abbasoglu Ozgoren A, Abd-Allah F, Abera SF, Aboyans V, Abraham JP, Abubakar I, Abu-Raddad LJ, Abu-Rmeileh NM, Achoki T, Ackerman IN, Ademi Z, Adou AK, Adsuar JC, Afshin A, Agardh EE, Alam SS, Alasfoor D, et al. 2015. Global, regional, and national disability-adjusted life years (dalys) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: Quantifying the epidemiological transition. *Lancet* **386**:2145–2191. doi: 10.1016/S0140-6736(15)61340-X

Murray NE, Quam MB, Wilder-Smith A. 2013. Epidemiology of dengue: Past, present and future prospects. *Clinical Epidemiology* **5**:299–309. doi: 10.2147/CLEP.S34440

Musso D, Nilles EJ, Cao-Lormeau V-M. 2014. Rapid spread of emerging zika virus in the pacific area. *Clinical Microbiology and Infection* **20**:O595–O596. doi: 10.1111/1469-0691.12707

Musso D, Cao-Lormeau VM, Gubler DJ. 2015a. Zika virus: Following the path of dengue and chikungunya? *Lancet* **386**:243–247. doi: 10.1016/S0140-6736(15)61273-9

Musso D, Roche C, Robin E, Nhan T, Teissier A, Cao-Lormeau VM. 2015b. Potential sexual transmission of zika virus. *Emerging Infectious Diseases* **21**:359–361. doi: 10.3201/eid2102.141363

Mylne AQ, Pigott DM, Longbottom J, Shearer F, Duda KA, Messina JP, Weiss DJ, Moyes CL, Golding N, Hay SI. 2015. Mapping the zoonotic niche of lassa fever in africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**:483–492. doi: 10.1093/trstmh/trv047

Nelson A. 2008. *Travel Time to Major Cities: A Global Map of Accessibility*. Ispra, Italy: GEMU-JRCOTE Commission.

Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, Azevedo RS, da Silva DE, da Silva EV, da Silva SP, Carvalho VL, Coelho GE, Cruz AC, Rodrigues SG, Vianez JL, Nunes BT, Cardoso JF, Tesh RB, Hay SI, Pybus OG, et al. 2015. Emergence and potential for spread of chikungunya virus in brazil. *BMC Medicine* **13**. doi: 10.1186/s12916-015-0348-x

Patiño-Barbosa AM, Medina I, Gil-Restrepo AF, Rodriguez-Morales AJ. 2015. Zika: Another sexually transmitted infection? *Sexually Transmitted Infections* **91**. doi: 10.1136/sextrans-2015-052189

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* **19**:181–197 . doi: 10.1890/07-2153.1

Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld CC, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI. 2014b. Global distribution maps of the leishmaniases. *eLife* **3**. doi: 10.7554/eLife.02851

Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld CC, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI.;2014a Global distribution maps of the leishmaniases. *eLife* **3**. doi: 10.7554/eLife.02851

Pigott DM, Golding N, Mylne A, Huang Z, Weiss DJ, Brady OJ, Kraemer MU, Hay SI. 2015a. Mapping the zoonotic niche of marburg virus disease in africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**:366–378. doi: 10.1093/trstmh/trv024

Pigott DM, Howes RE, Wiebe A, Battle KE, Golding N, Gething PW, Dowell SF, Farag TH, Garcia AJ, Kimball AM, Krause LK, Smith CH, Brooker SJ, Kyu HH, Vos T, Murray CJL, Moyes CL, Hay SI. 2015b. Prioritising infectious disease mapping. *PLOS Neglected Tropical Diseases* **9**:e0003756. doi: 10.1371/journal.pntd.0003756

Pinto E, Coelho M, Oliver L, Massad E. 2011. The influence of climate variables on dengue in singapore. *International Journal of Environmental Health Research* **21**:415–426. doi: 10.1080/09603123.2011.572279

Pond WL. 1963. ARTHROPOD-borne virus antibodies in sera from residents of south-east asia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **57**:364–371 . doi: 10.1016/0035-9203(63)90100-7

Powell JR, Tabachnick WJ. 2013. History of domestication and spread of aedes aegypti–a review. *Memórias Do Instituto Oswaldo Cruz* **108 Suppl 1**:11–17. doi: 10.1590/0074-0276130395

Reiskind MH, Lounibos LP. 2009. Effects of intraspecific larval competition on adult longevity in the mosquitoes aedes aegypti and aedes albopictus. *Medical and Veterinary Entomology* **23**:62–68. doi: 10.1111/j.1365-2915.2008.00782.x

Restrepo AC, Baker P, Clements AC. 2014. National spatial and temporal patterns of notified dengue cases, colombia 2007-2010. *Tropical Medicine & International Health* **19**:863–871. doi: 10.1111/tmi.12325

Roiz D, Boussès P, Simard F, Paupy C, Fontenille D. 2015. Autochthonous chikungunya transmission and extreme climate events in southern france. *PLoS Neglected Tropical Diseases* **9**:e0003854. doi: 10.1371/journal.pntd.0003854

Russell BM, Kay BH, Shipton W. 2001. Survival of aedes aegypti (diptera: Culicidae) eggs in surface and subterranean breeding sites during the northern queensland dry season. *Journal of Medical Entomology* **38**:441–445 . doi: 10.1603/0022-2585-38.3.441

Salje H, Cauchemez S, Alera MT, Rodriguez-Barraquer I, Thaisomboonsuk B, Srikiatkhachorn A, Lago CB, Villa D, Klungthong C, Tac-An IA, Fernandez S, Velasco JM, Roque VG, Nisalak A, Macareo LR, Levy JW, Cummings D, Yoon IK. 2016. Reconstruction of 60 years of chikungunya epidemiology in the philippines demonstrates episodic and focal transmission. *The Journal of Infectious Diseases* **213**:604–610. doi: 10.1093/infdis/jiv470

Sang S, Yin W, Bi P, Zhang H, Wang C, Liu X, Chen B, Yang W, Liu Q. 2014. Predicting local dengue transmission in guangzhou, china, through the influence of imported cases, mosquito density and climate variability. *PloS One* **9**:e102755. doi: 10.1371/journal.pone.0102755

Sankari T, Hoti SL, Singh TB, Shanmugavel J. 2012. Outbreak of dengue virus serotype-2 (DENV-2) of Cambodian origin in Manipur, India - association with meteorological factors. *The Indian Journal of Medical Research* **136**:649–655.

Schneider A, Friedl MA, Potere D. 2009. A new map of global urban extent from MODIS satellite data. *Environmental Research Letters* **4**:044003. doi: 10.1088/1748-9326/4/4/044003

Schneider A, Friedl MA, Potere D. 2010. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment* **114**:1733–1746. doi: 10.1016/j.rse.2010.03.003

Schneider A, Mertes CM, Tatem AJ, Tan B, Sulla-Menashe D, Graves SJ, Patel NN, Horton JA, Gaughan AE, Rollo JT, Schelly IH, Stevens FR, Dastur A. 2015. A new urban landscape in east–southeast asia, 2000–2010. *Environmental Research Letters* **10**:034002. doi: 10.1088/1748-9326/10/3/034002

Scott TW, Takken W. 2012. Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission. *Trends in Parasitology* **28**:114–121. doi: 10.1016/j.pt.2012.01.001

Socio-economic Data and Applications Center. 2015. Gridded Population of the World, v4 (GPWv4).

Smithburn K. 1954a. Neutralizing antibodies against arthropod-borne viruses in the sera of long-time residents of Malaya and Borneo. *American Journal of Hygiene* **59**:157–163.

Smithburn KC, Taylor RM, Rizk F, Kader A. 1954b. Immunity to certain arthropod-borne viruses among indigenous residents of Egypt. *The American Journal of Tropical Medicine and Hygiene* **3**:9–18.

Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. High-resolution gridded population datasets for latin america and the caribbean in 2010, 2015, and 2020. *Scientific Data* **2**. doi: 10.1038/sdata.2015.45

Sota T, Mogi M. 1992. Interspecific variation in desiccation survival *time of* aedes (stegomyia) mosquito eggs is correlated with habitat and egg size. *Oecologia* **90**:353–358. doi: 10.1007/BF00317691

Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One* **10**:e0107042. doi: 10.1371/journal.pone.0107042

Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, Reiner RC, Vilcarromero S, Elder JP, Halsey ES, Kochel TJ, Kitron U, Scott TW. 2013. House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences of the United States of America* **110**:994–999. doi: 10.1073/pnas.1213349110

Tatem AJ, Campbell J, Guerra-Arias M, de Bernis L, Moran A, Matthews Z. 2014. Mapping for maternal and newborn health: The distributions of women of childbearing age, pregnancies and births. *International Journal of Health Geographics* **13**. doi: 10.1186/1476-072X-13-2

Thu HM, Aye KM, Thein S. 1998. The effect of temperature and humidity on dengue virus propagation in Aedes aegypti mosquitos. *The Southeast Asian Journal of Tropical Medicine and Public Health* **29**:280–284.

Troyo A, Fuller DO, Calderón-Arguedas O, Solano ME, Beier JC. 2009. Urban structure and dengue fever in puntarenas, costa rica. *Singapore Journal of Tropical Geography* **30**:265–282. doi: 10.1111/j.1467-9493.2009.00367.x

Trpis M. 1972. Dry season survival of Aedes aegypti eggs in various breeding sites in the Dar es Salaam area, Tanzania. *Bulletin of the World Health Organization* **47**.

United Nations. 2014. *World Urbanization Prospects: The 2014 Revision*. New York. http://esa.un.org/unpd/wup/Publications/Files/WUP2014-Highlights.pdf.

UNFPA. 2014. *The State of the World's Midwifery 2014: A Universal Pathway. A Woman's Right to Health*. 1 228 UNFPA http://www.unfpa.org/sowmy.

Wan ZM, Zhang YL, Zhan QC, Li ZL. 2002. The MODIS land-surface temperature products for regional environmental monitoring and global change studies. *Int Geosci Remote Se*:3683–3685 . doi: 10.1109/igarss.2002.1027290

Weaver SC, Reisen WK. 2010. Present and future arboviral threats. *Antiviral Research* **85**:328–345. doi: 10.1016/j.antiviral.2009.10.008

Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. 2014. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* **98**:106–118. doi: 10.1016/j.isprsjprs.2014.10.001

Wenger SJ, Olden JD. 2012. Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* **3**:260–267. doi: 10.1111/j.2041-210X.2011.00170.x

***WHO Collaborating Center for Reference and Research on Arboviruses and Hemorrhagic Fever Viruses: Annual Report*.** Dakar, Senegal 1999:143.

Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. 2012. Human viruses: Discovery and emergence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367**:2864–2871. doi: 10.1098/rstb.2011.0354

World Health Organisation. 2015. Zika virus infection – Brazil and Colombia. http://www.who.int/csr/don/21-october-2015-zika/en/.

WorldPop. 2015. *High Resolution Age-Structured Population Distribution Maps. U. O. S. GeoData Institute*.

Zanluca C, Melo VC, Mosimann AL, Santos GI, Santos CN, Luz K. 2015. First report of autochthonous transmission of zika virus in brazil. *MemoRias Do Instituto Oswaldo Cruz* **110**:569–572. doi: 10.1590/0074-02760150192

Chapter 4 – Human mobility and dengue transmission in Pakistan

---

This chapter analyses the spatio-temporal dynamics of a vector borne disease once it enters into the human population at a new location either *via* an infected traveller or an animal-human transmission. Such estimates are crucial as they provide information directly relevant for policy makers during an outbreak. This work is published in *The Journal of the Royal Society Interface* and included here in its final form.

INTERFACE

rsif.royalsocietypublishing.org

Research

Cite this article: Kraemer MUG, Perkins TA, Cummings DAT, Zakar R, Hay SI, Smith DL, Reiner Jr RC. 2015 Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12**: 20150468.
http://dx.doi.org/10.1098/rsif.2015.0468

Received: 25 May 2015
Accepted: 23 September 2015

Subject Areas:
computational biology

Keywords:
dengue, epidemiology, heterogeneity, mixing, spatial dynamics, spatial accessibility

Author for correspondence:
M. U. G. Kraemer
e-mail: moritz.kraemer@zoo.ox.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2015.0468 or via http://rsif.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Big city, small world: density, contact rates, and transmission of dengue across Pakistan

M. U. G. Kraemer[1], T. A. Perkins[2,3], D. A. T. Cummings[4], R. Zakar[5], S. I. Hay[3,6,7], D. L. Smith[1,3,8] and R. C. Reiner Jr[3,9]

[1]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[2]Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA
[3]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA
[4]Department of Epidemiology, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD 21205, USA
[5]Department of Public Health, University of Punjab, Lahore 54590, Pakistan
[6]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[7]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA
[8]Sanaria Institute for Global Health and Tropical Medicine, Rockville, MD 20850, USA
[9]Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN 47405, USA

Macroscopic descriptions of populations commonly assume that encounters between individuals are well mixed; i.e. each individual has an equal chance of coming into contact with any other individual. Relaxing this assumption can be challenging though, due to the difficulty of acquiring detailed knowledge about the non-random nature of encounters. Here, we fitted a mathematical model of dengue virus transmission to spatial time-series data from Pakistan and compared maximum-likelihood estimates of 'mixing parameters' when disaggregating data across an urban–rural gradient. We show that dynamics across this gradient are subject not only to differing transmission intensities but also to differing strengths of nonlinearity due to differences in mixing. Accounting for differences in mobility by incorporating two fine-scale, density-dependent covariate layers eliminates differences in mixing but results in a doubling of the estimated transmission potential of the large urban district of Lahore. We furthermore show that neglecting spatial variation in mixing can lead to substantial underestimates of the level of effort needed to control a pathogen with vaccines or other interventions. We complement this analysis with estimates of the relationships between dengue transmission intensity and other putative environmental drivers thereof.

1. Introduction

The transmission dynamics of all infectious diseases depend on a few basic but key determinants: the availability of susceptible and infectious hosts, contacts between them and the potential for transmission upon contact. Susceptibility is shaped primarily by historical patterns of transmission, the natural history of the pathogen, the host's immune response and host demography [1]. What constitutes an epidemiologically significant contact depends on the pathogen's mode of transmission [2], and structure in contact patterns can be influenced by transportation networks and the spatial scale of transmission [3], by host heterogeneities such as age [4], and dynamically in response to the pathogen's influence on host behaviour [5]. Whether transmission actually occurs during contact between susceptible and infectious hosts often depends heavily on environmental conditions [6–8]. Disentangling the relative roles of these factors in driving patterns of disease incidence and prevalence is a difficult but central

© 2015 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

pursuit in infectious disease epidemiology, and mathematical models that capture the biology of how these mechanisms interact represent an indispensible tool in this pursuit [9].

The time-series susceptible–infected–recovered (TSIR) model was developed by Finkenstädt & Grenfell [10] to offer an accurate and straightforward way to statistically connect mechanistic models of infectious disease transmission with time-series data. Among other features, TSIR models readily account for inhomogeneous mixing in a phenomenological way by allowing for rates of contact between susceptible and infectious hosts to depend on their densities nonlinearly. Although this is a simple feature that can be incorporated into any model based on mass-action assumptions—indeed, earlier applications pertained to inhomogeneous mixing in predator–prey systems [11]—the 'mixing parameters' that determine the extent of this nonlinearity have primarily been fitted to empirical data in applications of the TSIR model to measles, cholera, rubella and dengue [12–15]. Applied to discrete-time models such as the TSIR, mixing parameters also have an interpretation as corrections for approximating a truly continuous-time process with a discrete-time model [16,17]. In no application of the TSIR model to date has the potential for variation in these parameters been assessed, leaving the extent to which inhomogeneity of mixing varies across space and time as an open question in the study of infectious disease dynamics.

There are a number of reasons why mixing might vary in time or space. Seasonal variation in mixing might arise because of travel associated with labour [18], religious events [19] or vacation [20], or because of the timing of school openings in the case of influenza [21]. Spatial variation in mixing could arise because of cultural differences at geographical scales [3,22,23], because of variation in the density and quality of roads [24], or because of variation in human densities and myriad-associated factors [13,25]. For vector-borne diseases, variation in mixing is amplified even further by variation in vector densities [26], which effectively mediate contact between susceptible and infectious hosts.

Dengue is a mosquito-borne viral disease with a strong potential for spatial variation in mixing [27,28]. The dominant vectors of dengue viruses (*Aedes* spp.) thrive in areas where they are able to associate with humans, as humans provide not only a preferred source of blood but also water containers that the mosquitoes use for egg laying and for larval and pupal development [29]. Two additional aspects of *Aedes* ecology—limited dispersal distance of the mosquito [30] and daytime biting [31]—imply that human movement should be the primary means by which the viruses spread spatially [2]. Indeed, analyses of dengue transmission dynamics at a variety of scales have strongly supported this hypothesis [32–35]. To the extent that human movement in dense urban environments is more well mixed than elsewhere, there are likely to be differences in the extent of inhomogeneous mixing in peri-urban and rural areas. This is also presumably the case for directly transmitted pathogens, but with a potentially even stronger discrepancy for dengue due to the urban–rural gradient in mosquito densities.

To assess the potential for spatial variation in the inhomogeneity of mixing as it pertains dengue transmission, we performed an analysis of district-level time series of dengue transmission in the Punjab province of Pakistan using a TSIR model with separate mixing parameters for urban and rural districts. We likewise made estimates of the relationships between density-independent transmission potential and putative drivers thereof, such as temperature, to allow for the relative roles of extrinsic and intrinsic factors to be teased apart. Finally, we performed mathematical analyses of the fitted model to assess the significance of spatial variation in mixing inhomogeneity for how time-series data are interpreted and used to guide control efforts.

Table 1. Reported cases by year in Lahore and all other districts.

	2011	2012	2013	2014	total
Lahore	18 020	4013	11 516	1799	35 348
other	3162	500	2356	2790	8808

2. Material and methods

We obtained daily dengue case data aggregated at hospital level from Punjab province provided by the Health Department Punjab, Pakistan, between 2011 and 2014. In total, 47 156 suspected and confirmed dengue cases were reported in 109 hospitals. All hospitals were subsequently geo-located using 'Google maps' (http://www.maps.google.com) similar to methods described here [36]. Hospitals that could not be identified were removed from the database. The hospitals were then assigned to a district within Punjab, Pakistan by their spatial location. A total of 21 182 cases alone were reported from the year 2011, which affected almost the entire province. Many more cases occurred in Lahore (35 348) compared to all other districts (8808) (table 1). A breakdown per year and each province is provided in electronic supplementary material, table S1, and additional information about collection can be found in the electronic supplementary material, appendix. No information on dengue serotypes were available. However, the predominant serotype circulating in Punjab province, Pakistan is that of DENV-2 [37].

2.1. Covariate selection and processing

Environmental conditions are instrumental in defining the risk of transmission of dengue [28]. Transmission is limited by the availability of a competent disease vector. Due to a lack of resources and political instability, no comprehensive nationwide entomological surveys have been performed in Pakistan. Therefore, we use a probabilistic model to infer the probability of occurrence of *Aedes aegypti* and *Aedes albopictus* in Pakistan derived from a globally comprehensive dataset containing more than 20 000 records for each species (figure 1*a*,*b*). In short, a boosted regression tree model was applied that predicts a continuous spatial surface of mosquito occurrence from a fitted relationship between the distribution of these mosquitoes and their environmental drivers. A detailed description of the occurrence database and modelling outputs are available here [36,38,39]. Such model outputs have proved useful in identifying areas of risk of transmission of dengue as well as malaria [28,40,41]. Other important environmental conditions defining the risk of transmission of dengue are temperature, water availability and vegetation cover [42]. To account for such variation, raster layers of daytime land surface temperature were processed from the MOD11A2 satellite, gap-filled to remove missing values, and then averaged to a monthly temporal resolution for all 4 years [43]. The density of vegetation coverage has been shown to be associated with vector abundance [44]. Moreover, vegetation indices are useful proxies for precipitation and may be used to derive the presence of standing water buckets that are habitats for the *Aedes* mosquitos [45]. The same method was again applied to derive the enhanced vegetation index (EVI) from the MOD11A2 satellite to produce 16-day and monthly pixel-based estimates for 2011–2014

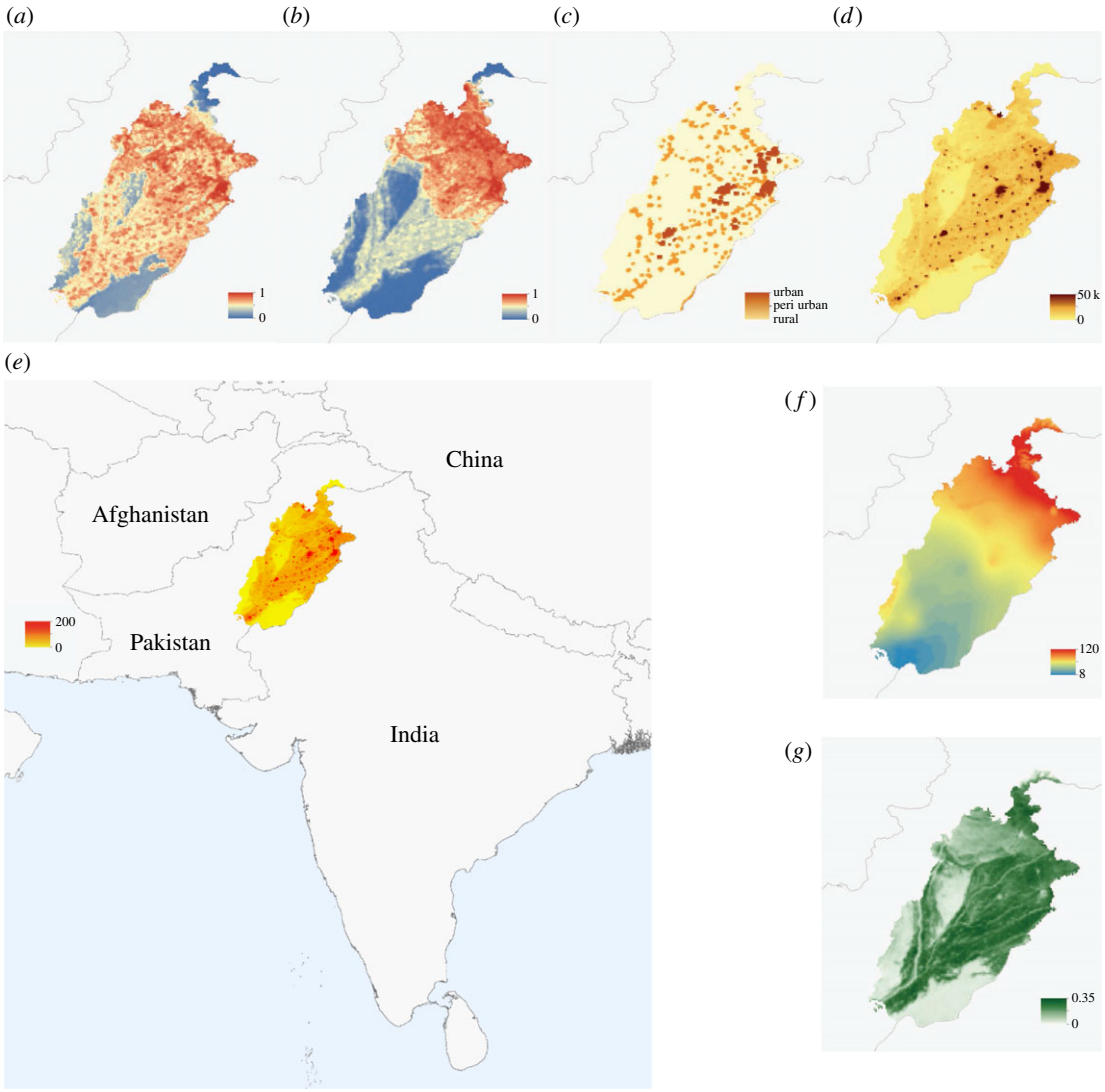


Figure 1. Covariates used in this study to derive environmental drivers of transmission. *Aedes aegypti* probability of occurrence (*a*); *A. albopictus* probability of occurrence (*b*); urbanicity (*c*); weighted urban accessibility (*d*); population density and study area (*e*); precipitation (*f*); EVI mean (*g*).

(figure 1*g*) [46]. Due to the inherent delay between rainfall and daily temperature influencing mosquito population dynamics and those mosquitoes contributing to an increase in DENV transmission, we consider both, the influence of the current temperature, vegetation indices and precipitation, data on current transmission as well as the values of those covariates the time step before (figure 1*f*).

We used population count estimates on a 100 m resolution that were subsequently aggregated up to match all other raster layers to a 5×5 km resolution for the year 2015 (http://www.worldpop.org) (figure 1*e*). In a follow-up analysis to our primary investigation into the climatological drivers of dengue transmission, we included several density-based covariates. We derived a weighted accessibility metric that includes both, population density and urban accessibility, a metric commonly used to derive relative movement patterns [24,47]. This map shows a friction surface, i.e. the time needed to travel through a specific pixel (figure 1*d*). We also used an urban, peri-urban and rural classification scheme to quantify patterns of urbanicity based on a globally available grid [48] (figure 1*c*). All covariates and case data were aggregated and averaged (where appropriate) to a district level.

2.2. Model

Following Finkenstädt & Grenfell [10], we assume a general transmission model of

$$I_{t,i} = \beta_{t,i} \cdot \frac{I_{t-1,i}^{\alpha_i}}{N_i} \cdot S_{t-1,i} \cdot \epsilon_{t,i}, \tag{2.1}$$

where $I_{t,i}$ is the number of infected and infectious individuals and $S_{t,i}$ the number of susceptible individuals, at time t in district i, N_i is the population of district i, and $\beta_{t,i}$ is the covariate driven contact rate. We assume each individual to be susceptible as the 2011 epidemic is the first large dengue outbreak. The mixing parameter for the ith district is given by α_i; when α_i is equal to 1, the population has homogeneous mixing where values less than one can either indicate inhomogeneous mixing or a need to correct for the discretization of the continuous-time process. $\beta_{t,i}$ was fitted using covariates shown in figure 1. Finally, the error terms $\epsilon_{t,i}$ are assumed to be independent and identically log-normally distributed random variables.

2.3. Model selection

The term $\beta_{t,i}$ is fit using generalized additive model regression [49–51]. The time-varying climatological covariates are all fit as a smooth spline, while all other covariates enter $\beta_{t,i}$ linearly. For example, if covariate X_1 and X_2 are time varying and X_3 and X_4 are temporally constant, then we fit β as

$$E[\beta(t)] = s[X_1(t)] + s[X_2(t)] + X_3 + X_4, \tag{2.2}$$

where s are smooths.

Additionally, unexplained seasonal variation is accounted for using a 12-month periodic smooth spline.

Model selection was performed using backwards selection. Two base models were investigated. First, a climate-only model

was created using only the climatological and environmental suitability covariates. Second, a 'full' model using the density-dependent covariates as well as the climatological covariates were combined into a single model which was then subjected to backwards selection. For both models, the mixing coefficient was initially set equal for each district and once a final model was arrived upon, the mixing coefficient for Lahore was allowed to vary separately from the other coefficients. All model fitting was conducted using R [52] and the 'mgcv' package [51]. Models are fit by maximizing the restricted maximum likelihood [53] to reduce bias and over-fitting of the smooth splines. The model source code and processing of covariates will be made available in line with previous projects [54].

2.4. Model analysis

To explore the potential significance of spatial variation in mixing parameters, we conducted an analysis to probe the inherent mathematical trade-off between the mixing parameter α and the density-independent transmission coefficient β. Specifically, to answer the question, what difference in local transmission would be necessary to account for a difference in mixing while achieving identical transmission dynamics. To explore this, we used equation (2.1) to establish:

$$\beta_1 \frac{{I_{1,t}}^{\alpha_1}}{N_1} = \beta_2 \frac{{I_{2,t}}^{\alpha_2}}{N_2}, \tag{2.3}$$

from which we obtained

$$\frac{\beta_1}{\beta_2} = I^{\alpha_2 - \alpha_1}. \tag{2.4}$$

We then examined how variation in I and $\alpha_2 - \alpha_1$ affected the left-hand side of equation (2.4) and likewise the critical proportion of the population to control in order to effect pathogen elimination, which, under our model, is $p_c = 1 - (1/\beta)$.

3. Results

3.1. Description of case distribution

The majority of cases were clustered in Lahore, the capital of Punjab province. Ongoing transmission appeared to be focal in three (Vehari, Rawalpindi and Lahore) districts and to have spread through infective 'sparks' to smaller more rural provinces.

3.2. Model selection

To disentangle the different aspects of dengue dynamics and their drivers, we used a model containing only the climatological covariates and performed backwards model selection until each covariate in the model was significant at a 5% level. This resulted in a model that explained 76.9% of the deviance and that had an adjusted R^2 of 0.746. Among the yearly averaged covariates, EVI and precipitation remained in the model, as well as the derived *A. albopictus* range map ($p = 8.21 \times 10^{-4}$, 0.01, and 3.9×10^{-5}, respectively). Interestingly, when we substituted the derived *A. aegypti* map for the *A. albopictus* map, the deviance explained changed very little to 76.8%. For climatological covariates that were fit as smooth splines, temperature, lagged temperature and EVI remained in the model (figure 2, p-value of 0.010, 0.030 and 0.030 with effective degrees of freedom 7.55, 5.47 and 1.83, respectively). There was a significant amount of periodic variation unexplained by the climatological covariates alone, as the 'seasonality' covariate was retained by the model selection algorithm (figure 2, $p = 0.0034$). The estimated median values for R_0 per district were clustered around 2 (mean = 2.1), and their geographical distribution indicated a clear trend towards districts with larger populations (figure 3). Finally, the mixing coefficient was significantly lower than 1 ($\alpha = 0.69$, 95% CI = (0.614, 0.771), $p = 1.6 \times 10^{-14}$).

To understand these differences, the final model was then compared to a nested model in which the coefficient for Lahore was allowed to vary independently of all other districts. Deviance explained increased to 77% and adjusted R^2 increased to 0.753. Further, the mixing coefficient for Lahore ($\alpha = 0.74$) was significantly larger than the mixing coefficient for the other districts ($\alpha = 0.59$, $p = 0.0068$) (electronic supplementary material, figure S1). The median R_0 for Lahore was estimated at 3.28, the highest among all districts.

To assess the extent to which the variation in mixing coefficients could be explained by other covariates, we considered the possibility that movement accounted for the differences in the mixing coefficients between Lahore and the other districts. The density-dependent covariates (described earlier) were then added to the full model and backwards selection was performed again. The resulting model explained 78.6% of the deviance, had an adjusted R^2 of 0.763 and was superior to the final climatological model based on AIC (699.23 versus 714.83). Yearly averaged EVI, normalized difference vegetation index and precipitation were all significant ($p = 8.7 \times 10^{-5}$, 0.00024 and 0.00028, respectively). Again, the derived *A. albopictus* map was significant ($p = 0.00816$). For climatological covariates fit as smooth splines, only temperature and lagged temperature were found to be significant (figure 2*b*, $p = 4.0 \times 10^{-5}$ and 0.0013, and effective degrees of freedom 7.61 and 4.81, respectively), and there was still a significant 'seasonality' effect (figure 2*b*, $p = 4.0 \times 10^{-7}$, effective degrees of freedom 4.48). The best-fit mixing coefficient was $\alpha = 0.58$, barely lower than the mixing coefficient for non-Lahore districts in the climatological model. The estimated median R_0 again clustered around 2 (mean = 1.8), and again the R_0 for Lahore was largest, but in this model it was considerably larger than in the climatological model (Lahore $R_0 = 7.82$, figure 2*b*). Full details of the best-fit model parameters are shown in electronic supplementary material, table S2–S5.

Two of the density-dependent covariates remained in the model: the urban map ($p = 0.01$) and the weighted access map ($p = 3 \times 10^{-5}$). When the nested model that allowed Lahore's mixing coefficient to vary was fitted, there was no significant difference between the two mixing coefficients ($p \approx 1$).

3.3. Model analysis

Given a difference in estimates of the mixing parameters between Lahore and elsewhere of 0.15, we analysed equation (2.3) to assess the bias in estimates of the transmission coefficient that would result from ignoring this extent of variation in the inhomogeneity of mixing displayed between two areas. For the purpose of *ceteris paribus* comparisons, we assumed equal force of infection but varied it across several orders of magnitude. Depending on the order of magnitude, estimates of transmission coefficients made if overestimating the mixing parameter by 0.15 could easily result in a two- to threefold underestimate in the transmission coefficient (figure 4). For realistic ranges of the transmission coefficient for dengue, and equivalently for the basic reproductive

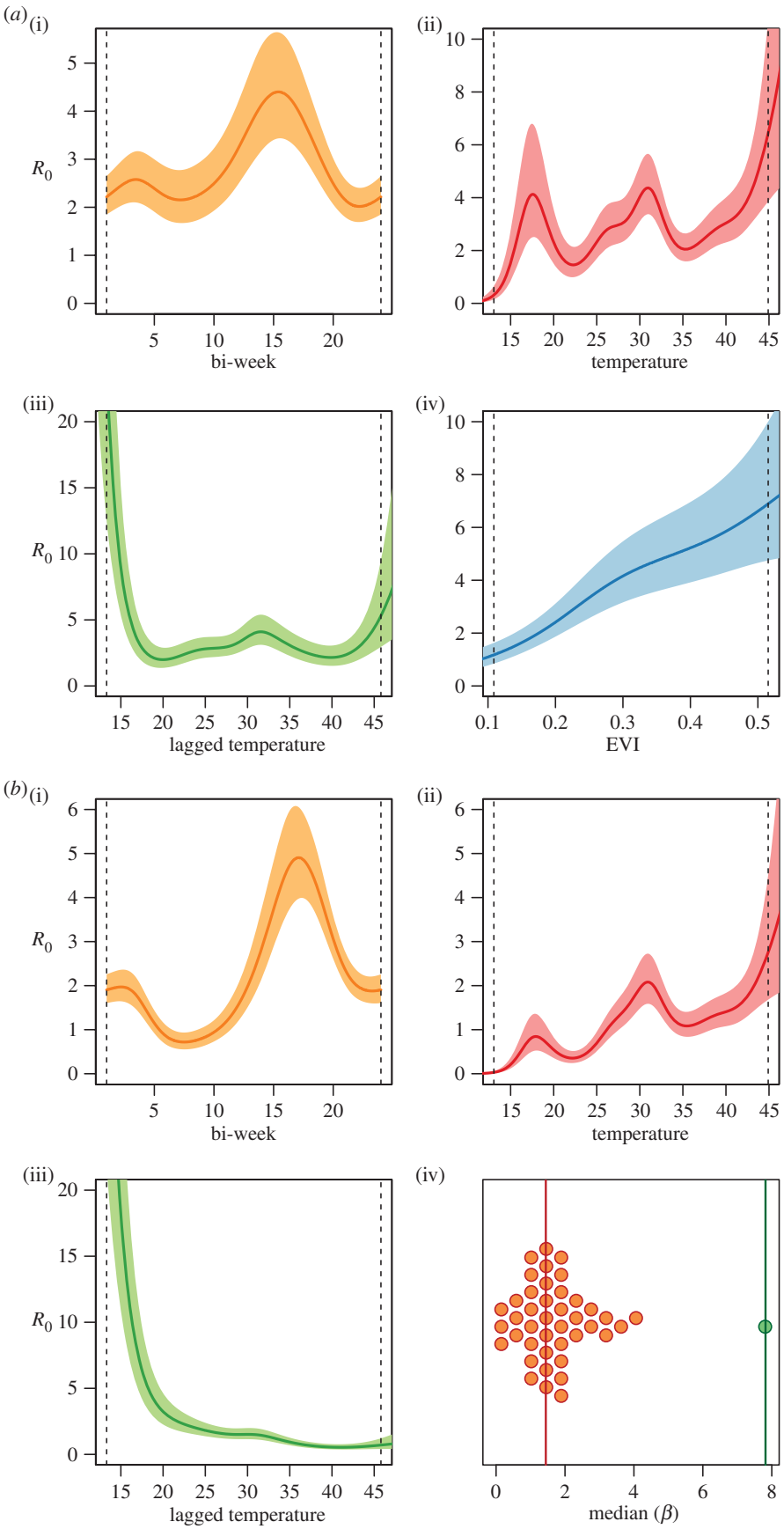


Figure 2. Model outputs using a backwards model selection procedure in the model using climatological variables (*a*, i–iv), and including the density-dependent variables (*b*, i–iv). Every subplot shows the predictions of the model for the indicated parameter carrying across the indicated range and every other parameter set to their mean. Figure (*b*, iv) shows the differences in the transmission coefficient from Lahore (green) and all other districts (red).

number R_0 under the present model, this extent of underestimation of R_0 could lead to underestimating the critical proportion of the population to which vaccines or other interventions must be applied by 20–30% (figure 5).

4. Discussion

Our results point to considerable spatial heterogeneity in the inhomogeneity of mixing and the strength of an associated

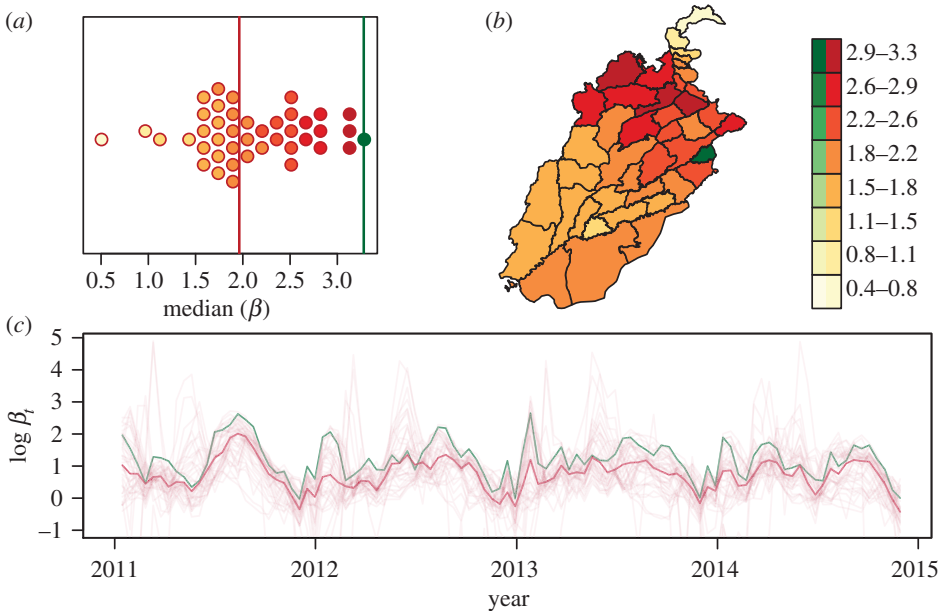


Figure 3. Average distribution of R_0 with green representing Lahore versus red all other districts (*a*), their geographical distribution (*b*) and over time in which the green line again is representing Lahore versus the red line representing all other districts (*c*). Outputs shown here correspond to the model without including density-related variables.

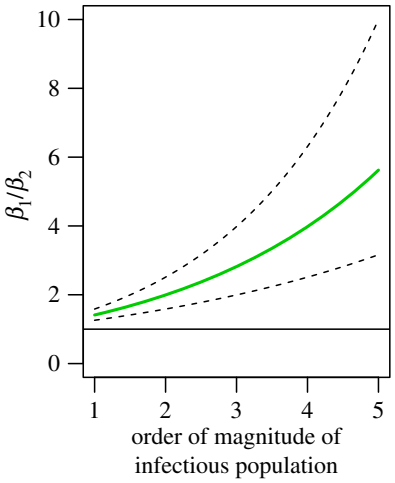


Figure 4. Ratio of betas (R_0) assuming equal force of infection and a difference in $\alpha_2 - \alpha_1$ of 0.2, 0.15 (green), and 0.1, from top to bottom. The straight line indicates a ratio of 1. Order of magnitude of the infectious population refers to number of infectious people in powers of 10; i.e. the range is 10–100 000.

nonlinearity in transmission along an urban–rural gradient. This regional variability in mixing has direct implications for estimates of the basic reproductive number of dengue in our study region and elsewhere. Although the potential for such bias in estimates of the basic reproductive number has been shown in a theoretical context [26,55], we provide quantitative estimates of the extent of this problem by interfacing models with a rich spatio-temporal dataset. Our results have implications for estimates of population-level parameters not only for dengue but also for other infectious diseases [13,56–60] and possibly even more broadly in ecology [11].

Our analysis revealed significant differences in the inhomogeneity of mixing between urban and rural settings and found that a population-weighted urban accessibility metric was able to account for differences in mixing between these settings. Mixing is presumably influenced directly by human behaviour and has been shown to be highly unpredictable, largely dependent on the local context and the spatial and temporal scale [61]. In this study, however, we could show that the density-dependent covariate we considered was able to capture the influence of these behavioural effects on a district level. Once differences in the inhomogeneity of mixing were accounted for, estimated R_0 values indicated considerably larger differences between transmission potential in Lahore and all other districts. Synchronizing more accurate geo-referenced data would allow for the assessment of the extent to which the relationship between 'mixing parameters' and urban accessibility is dependent on the spatial scale at which data are aggregated [26,62]. In the case of dengue, this has been limited specifically by the availability of high-resolution data [63]. Complementing such an analysis with measurement of social contact patterns could be important for exploring this relationship in even more detail [22,64,65] and could be informed by mathematical models that explored this relationship previously for other diseases [13,66]. Another encouraging result from our analysis was the finding that large-scale mosquito suitability surfaces helped capture the environmental determinants of dengue transmission [28].

Intervention strategies are contingent on both understanding key environmental drivers of transmission and the dynamics of ongoing human-to-human transmission, particularly in outbreak situations [67]. Environmental drivers such as seasonal fluctuations in rainfall, temperature, vegetation coverage or mosquito abundance will help guide surveillance and control efforts targeted mostly towards the mosquito vector and its ecology [68]. Once infection occurs, an important and unresolved question for dengue is how to best optimize the delivery of intervention strategies to reduce disease incidence, which is largely determined by R_0. Our analysis shows that the interaction between mixing parameters and force of infection has potentially large implications for optimizing targeted intervention, particularly in countries where transmission is high and resources are scarce [69]. In fact, this may be even more important in areas of low transmission where incidence appears to be

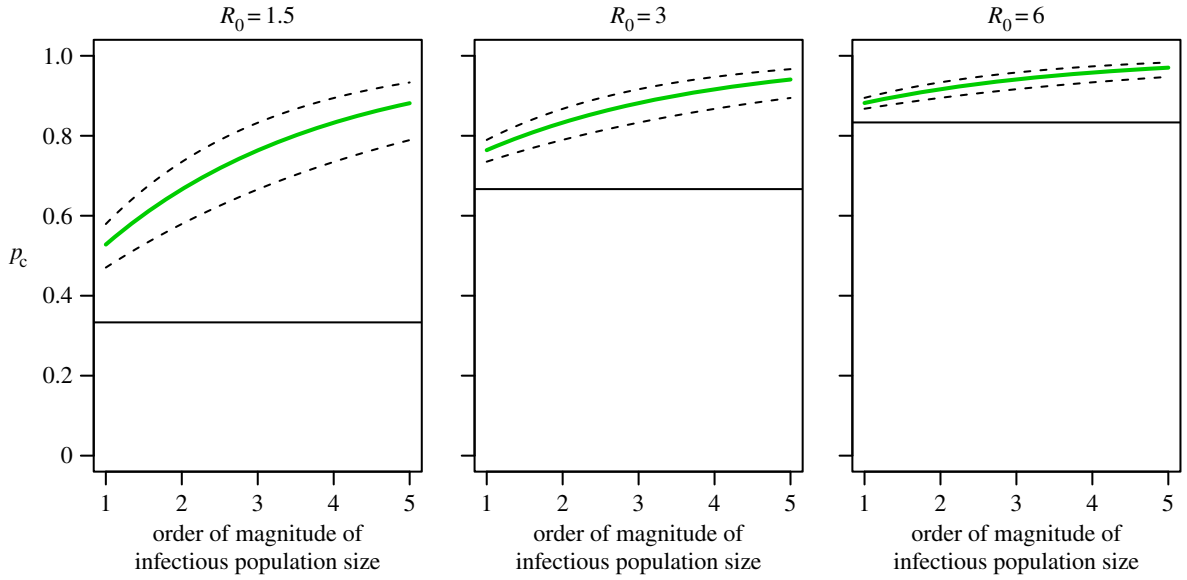


Figure 5. Critical proportion of the population to control in population 2 as a function of R_0 in population 1, the order of magnitude of the infectious numbers in each population, and a difference in $\alpha_2 - \alpha_1$ of 0.1, 0.15 (green) and 0.2. The straight line indicates the critical proportion assuming the α in each population are equal. Order of magnitude of the infectious population refers to number of infectious people in powers of 10; i.e. the range is 10–100 000.

more focal [70]. Again, however, more attention is needed to determine the spatial and temporal resolution of appropriate intervention strategies and the effects of key covariates and model parameters [62]. Empirical understanding of the spatial scale that is most appropriate for carrying out large-scale interventions remains unknown.

Once transmission has occurred in one place, understanding not only spatial heterogeneity in transmission dynamics but also their subsequent spread in mechanistic stochastic models would help to empirically determine the propagation of the disease [71]. Interest in spatial spread dynamics has risen with increasing importation of dengue into heretofore non-endemic areas due to travel and trade continentally and internationally [72]. Exploration of the case data in Pakistan that we analysed here suggests that the virus spreads along major transport routes from Lahore to Karachi and north to Rawalpindi. Using results presented here on mixing coefficients and environmental drivers will help pinpoint areas of major risk of importation more accurately, especially in the case of recurring epidemics. We explored the consequences of a spatially differentiated mixing coefficient in the context of transmission potential within this analysis. Using the fitted relationships of the environmental drivers of transmission and R_0 will enable future analyses and comparisons between diseases and geographical regions. In this context, it will be instrumental to integrate a variety of movement and social network models with the evidence presented here to infer more accurately how the geographical spread of dengue is determined.

Authors' contributions. M.U.G.K., R.C.R., T.A.P. and D.L.S. conceived and designed the experiment. R.C.R., T.A.P. and M.U.G.K. performed the experiments. R.C.R. and M.U.G.K. analysed the data. D.L.S., S.I.H., D.A.T.C. and Z.R. edited the manuscript. M.U.G.K., R.C.R. and T.A.P. wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. M.U.G.K. acknowledges funding from the German Academic Exchange Service (DAAD). T.A.P., S.I.H. and R.C.R. acknowledge the support from the RAPIDD programme of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (no. 095066) and a grant from the Bill & Melinda Gates Foundation (no. OPP1093011). Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements. The authors thank the WHO Punjab office, Health Department Punjab and PID for providing the epidemiological data, and all participants that helped collect the data.

References

1. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332. (doi:10.1126/science.1090727)
2. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, Elder JP, Scott TW. 2009 The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.* **3**, e481. (doi:10.1371/journal.pntd.0000481)
3. Brockmann D, Helbing D. 2013 The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342. (doi:10.1126/science.1245200)
4. Kilpatrick AM, Daszak P, Jones MJ, Marra PP, Kramer LD. 2006 Host heterogeneity dominates West Nile virus transmission. *Proc. R. Soc. B* **273**, 2327–2333. (doi:10.1098/rspb.2006.3575)
5. Fenichel EP *et al.* 2011 Adaptive human behavior in epidemiological models. *Proc. Natl Acad. Sci. USA* **108**, 6306–6311. (doi:10.1073/pnas.1011250108)
6. Weiss DJ, Bhatt S, Mappin B, Van Boeckel TP, Smith DL, Hay SI, Gething PW. 2014 Air temperature suitability for *Plasmodium falciparum* malaria transmission in Africa 2000–2012: a high-resolution spatiotemporal prediction. *Malar. J.* **13**, 171. (doi:10.1186/1475-2875-13-171)
7. Gilbert M *et al.* 2014 Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat. Commun.* **5**, 4116. (doi:10.1038/ncomms5116)
8. Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M. 2010 Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316. (doi:10.1371/journal.pbio.1000316)

9. Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
10. Finkenstädt BF, Grenfell BT. 2000 Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.* **49**, 187–205. (doi:10.1111/1467-9876.00187)
11. Pascual M, Levin SA. 1999 From individuals to population densities: searching for the scale of intermediate determinism. *Ecology* **80**, 2225–2236. (doi:10.1890/0012-9658(1999)080[2225:FITPDS]2.0.CO;2)
12. Reich NG *et al.* 2013 Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *J. R. Soc. Interface* **10**, 20130414. (doi:10.1098/rsif.2013.0414)
13. Bjørnstad ON, Finkenstädt BF, Grenfell BT. 2002 Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* **72**, 169–184. (doi:10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2)
14. Koelle K, Pascual M. 2004 Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* **163**, 901–913. (doi:10.1086/420798)
15. Metcalf CJE, Bjornstad ON, Ferrari MJ, Klepac P, Bharti N, Lopez-Gatell H, Grenfell BT. 2011 The epidemiology of rubella in Mexico: seasonality, stochasticity and regional variation. *Epidemiol. Infect.* **139**, 1029–1038. (doi:10.1017/S0950268810002165.The)
16. Liu WM, Hethcote HW, Levin SA. 1987 Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.* **25**, 359–380. (doi:10.1007/BF00277162)
17. Glass K, Xia Y, Grenfell BT. 2003 Interpreting time-series analyses for continuous-time biological models: measles as a case study. *J. Theor. Biol.* **223**, 19–25. (doi:10.1016/S0022-5193(03)00031-6)
18. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. 2011 Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334**, 1424–1427. (doi:10.1126/science.1210554)
19. Lessler J *et al.* 2014 Estimating potential incidence of MERS-CoV associated with Hajj pilgrims to Saudi Arabia, 24 November 2014, 1st edn (referring to: http://currents.plos.org/outbreaks/article/estimating-potential-incidence-of-mers-cov-associated-with-hajj-pilgrims-to-saudi-arabia-2014/). (doi:10.1371/currents.outbreaks.c5c9c9abd636164a9b6fd4dbda974369)
20. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ. 2014 Dynamic population mapping using mobile phone data. *Proc. Natl Acad. Sci. USA* **111**, 15 888–15 893. (doi:10.1073/pnas.1408439111)
21. Salathé M, Kazandjieva M, Woo J, Levis P, Feldman MW, Jones JH. 2010 A high-resolution human contact network for infectious disease transmission. *Proc. Natl Acad. Sci. USA* **107**, 22 020–22 025. (doi:10.1073/pnas.1009094108)
22. Vazquez-Prokopec GM *et al.* 2013 Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS ONE* **8**, e58802. (doi:10.1371/journal.pone.0058802)
23. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, Guan Y, Jiang CQ, Cummings DAT. 2014 Social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. B* **281**, 20140268. (doi:10.1098/rspb.2014.0268)
24. Background Paper for World Development Report 2009: Reshaping Economic Geography. Washington, DC: World Bank.
25. Simini F, González MC, Maritan A, Barabási A-L. 2012 A universal model for mobility and migration patterns. *Nature* **484**, 96–100. (doi:10.1038/nature10856)
26. Perkins TA, Scott TW, Le Menach A, Smith DL. 2013 Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput. Biol.* **9**, e1003327. (doi:10.1371/journal.pcbi.1003327)
27. Brady OJ *et al.* 2012 Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6**, e1760. (doi:10.1371/journal.pntd.0001760)
28. Bhatt S *et al.* 2013 The global distribution and burden of dengue. *Nature* **496**, 504–507. (doi:10.1038/nature12060)
29. Morrison AC, Zielinski-Gutierrez E, Scott TW, Rosenberg R. 2008 Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti*. *PLoS Med.* **5**, e68. (doi:10.1371/journal.pmed.0050068)
30. Harrington LC *et al.* 2005 Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *Am. J. Trop. Med. Hyg.* **72**, 209–220.
31. Akram W, Hafeez F, Ullah UN, Kim YK, Hussain A, Lee JJ. 2009 Seasonal distribution and species composition of daytime biting mosquitoes. *Entomol. Res.* **39**, 107–113. (doi:10.1111/j.1748-5967.2009.00204.x)
32. Allicock OM *et al.* 2012 Phylogeography and population dynamics of dengue viruses in the Americas. *Mol. Biol. Evol.* **29**, 1533–1543. (doi:10.1093/molbev/msr320)
33. Reiner RC, Stoddard ST, Scott TW. 2014 Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. *Epidemics* **6**, 30–36. (doi:10.1016/j.epidem.2013.12.003)
34. Stoddard ST *et al.* 2013 House-to-house human movement drives dengue virus transmission. *Proc. Natl Acad. Sci. USA* **110**, 994–999. (doi:10.1073/pnas.1213349110)
35. Bhoomiboonchoo P *et al.* 2014 The spatial dynamics of dengue virus in Kamphaeng Phet, Thailand. *PLoS Negl. Trop. Dis.* **8**, e3138. (doi:10.1371/journal.pntd.0003138)
36. Kraemer MUG *et al.* 2015 The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Sci. Data* **2**, 150035. (doi:10.1038/sdata.2015.35)
37. Khan MA *et al.* 2013 Emergence and diversification of dengue 2 cosmopolitan genotype in Pakistan, 2011. *PLoS ONE* **8**, e56391. (doi:10.1371/journal.pone.0056391)
38. Kraemer MUG *et al.* 2015 The global distribution of the arbovirus vectors *Aedes aegypti* and Ae. albopictus. *Elife* **4**, e08347. (doi:10.7554/eLife.08347)
39. Kraemer MUG *et al.* 2015 Data from the global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Dryad Digit. Repositry*. (doi:10.5061/dryad.47v3c)
40. Sinka ME *et al.* 2012 A global map of dominant malaria vectors. *Parasit. Vectors* **5**, 69. (doi:10.1186/1756-3305-5-69)
41. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, Tatem AJ, Hay SI. 2011 A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10**, 378. (doi:10.1186/1475-2875-10-378)
42. Messina JP, Brady OJ, Pigott DM, Golding N, Kraemer MUG, Scott TW, Wint GRW, Smith DL, Hay SI. 2015 The many projected futures of dengue. *Nat. Rev. Microbiol.* **13**, 230–239. (doi:10.1038/nrmicro3430)
43. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI, Gething PW. 2015 Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach. *Malar. J.* **14**, 1–18. (doi:10.1186/s12936-015-0574-x)
44. Eisen L, Lozano-Fuentes S. 2009 Use of mapping and spatial and space-time modeling approaches in operational control of *Aedes aegypti* and dengue. *PLoS Negl. Trop. Dis.* **3**, e411. (doi:10.1371/journal.pntd.0000411)
45. Maciel-de-Freitas R, Marques WA, Peres RC, Cunha SP, De Oliveira RL. 2007 Variation in *Aedes aegypti* (Diptera: Culicidae) container productivity in a slum and a suburban district of Rio de Janeiro during dry and wet seasons. *Mem. Inst. Oswaldo Cruz* **102**, 489–496. (doi:10.1590/S0074-02762007005 000056)
46. Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. 2014 An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogramm. Remote Sens.* **98**, 106–118. (doi:10.1016/j.isprsjprs.2014.10.001)
47. Tatem AJ, Hemelaar J, Gray RR, Salemi M. 2012 Spatial accessibility and the spread of HIV-1 subtypes and recombinants. *AIDS* **26**, 2351–2360. (doi:10.1097/QAD.0b013e328359a904)
48. CIESIN/IFPRI/WB/CIAT. Global Rural Urban Mapping Project (GRUMP): Gridded Population of the World, version 3. See http://sedac.ciesin.columbia.edu/gpw.
49. Hastie TJ, Tibshirani RJ. 1990 *Generalized additive models*, vol. 43. Chapman and Hall/CRC, CRC Press.
50. Dominici F, McDermott A, Zeger SL, Samet JM. 2002 On the use of generalized additive models in time-series studies of air pollution and health. *Am. J. Epidemiol.* **156**, 193–203. (doi:10.1093/aje/kwf062)
51. Wood SN. 2011 Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat.*

Soc. Ser. B Stat. Methodol. **73**, 3–36. (doi:10.1111/j.1467-9868.2010.00749.x)

52. R Core Team. 2014 *R: a language and environment for computing*. Vienna, Austria: the R Foundation for Statistical Computing.
53. Patterson HD, Thompson R. 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554. (doi:10.1093/biomet/58.3.545)
54. Pigott DM, Kraemer MU. 2014 Enhancing infectious disease mapping with open access resources. *Eurosurveillance* **19**, ppii: 20989.
55. Hu H, Nigmatulina K, Eckhoff P. 2013 The scaling of contact rates with population density for the infectious disease models. *Math. Biosci*. **244**, 125–134. (doi:10.1016/j.mbs.2013.04.013)
56. Bartlett MS. 1960 The critical community size for measles in the U.S. *J. R. Stat. Soc. A* **123**, 37–44. (doi:10.2307/2343186)
57. Keeling MJ, Eames KTD. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)
58. Kiss IZ, Green DM, Kao RR. 2008 The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *J. R. Soc. Interface* **5**, 791–799. (doi:10.1098/rsif.2007.1272)
59. Smith DL, Lucey B, Waller LA, Childs JE, Real LA. 2002 Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl Acad. Sci. USA* **99**, 3668–3672. (doi:10.1073/pnas.042400799)
60. Keeling MJ, Grenfell BT. 1997 Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65–67. (doi:10.1126/science.275.5296.65)
61. Yang Y, Herrera C, Eagle N, González MC. 2014 Limits of predictability in commuting flows in the absence of data for calibration. *Sci. Rep*. **4**, 5662. (doi:10.1038/srep05662)
62. Mills HL, Riley S. 2014 The spatial resolution of epidemic peaks. *PLoS Comput. Biol*. **10**, e1003561. (doi:10.1371/journal.pcbi.1003561)
63. Ruberto I, Marques E, Burke DS, Van Panhuis WG. 2015 The availability and consistency of dengue surveillance data provided online by the World Health Organization. *PLoS Negl. Trop. Dis*. **9**, e0003511. (doi:10.1371/journal.pntd.0003511)
64. Heesterbeek H *et al*. 2015 Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, paaa4339. (doi:10.1126/science.aaa4339)
65. Bauch CT, Galvani AP. 2013 Epidemiology. Social factors in epidemiology. *Science* **342**, 47–49. (doi:10.1126/science.1244492)
66. Reiner RC, King AA, Emch M, Yunus M, Faruque ASG, Pascual M. 2012 Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc. Natl Acad. Sci. USA* **109**, 2033–2036. (doi:10.1073/pnas.1108438109)
67. Perkins TA, Metcalf CJE, Grenfell BT, Tatem AJ. 2015 Estimating drivers of autochthonous transmission of chikungunya virus in its invasion of the Americas. *PLOS Currents Outbreaks*, 10 February 2015, 1st edn. (doi:10.1371/currents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc)
68. Johansson MA. 2015 Chikungunya on the move. *Trends Parasitol*. **31**, 43–45. (doi:10.1016/j.pt.2014.12.008)
69. Cesare MDi, Bhatti Z, Soofi SB, Fortunato L, Ezzati M, Bhutta ZA. 2015 Geographical and socioeconomic inequalities in women and children's nutritional status in Pakistan in 2011: an analysis of data from a nationally representative survey. *Lancet Glob. Health* **3**, e229–e239. (doi:10.1016/S2214-109X(15)70001-X)
70. Salje H *et al*. 2012 Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc. Natl Acad. Sci. USA* **109**, 9535–9538. (doi:10.1073/pnas.1120621109)
71. Grenfell BT, Bjùrnstad ON, Kappey J. 2001 Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723. (doi:10.1038/414716a)
72. Schaffner F, Mathis A. 2014 Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. *Lancet. Infect. Dis*. **14**, 1271–1280. (doi:10.1016/S1473-3099(14)70834-5)

Chapter 5 – The impact of human mobility on the transmission dynamics of Ebola virus disease in West Africa

---

Human mobility has a significant impact on the transmission dynamics of infectious diseases. Over the course of an outbreak the impact of such movements may change. Drawing upon previous modelling work in chapter 4, this work directly links the timing of the epidemic, growing vs. declining phase, to the movements of people. This work is submitted for publication and currently being revised for resubmission. Figures are at the end of the written text and additional files are included in the Appendix.

Predicting the geographic spread of the 2014-2016 West Africa Ebola virus disease outbreak

Kraemer, M.U.G., Golding, N., Bisanzio, D., Bhatt, S., Pigott, D.M., Brady, O.J., Faria, N.R., Pybus, O.G., Smith, D.L, Tatem, A.J., Hay, S.I., Reiner, R.C. Jr.

Abstract

Between 2014 and 2016 West Africa experienced the most geographically extensive outbreak of Ebola virus disease (EVD) recorded to date. It affected all districts in Sierra Leone and Liberia and most of Guinea, resulting in more than 11,301 deaths among 28,603 cases. The outbreak spread from its origin in Meliandou, Guinea by movements of infected individuals across the region. Understanding what drives the transmission dynamics at what time of this epidemic is important for planning responses to future outbreaks. We isolate the effect of human movement on the geographic diffusion of the outbreak and model the dynamics in the growth and decline phases to identify areas that were the main exporters and importers of disease. We identify considerable spatio-temporal heterogeneity in transmission within and between countries with implications to rapidly respond to future contagious disease outbreaks.

Introduction

The Ebola virus disease (EVD) epidemic in West Africa has now caused at least 28,000 infections and resulted in more than 11,000 deaths [1]. At the height of the outbreak in late 2014 the geographic extent of transmission was the widest ever recorded for Ebola virus, with cases reported in all districts in Sierra Leone (14/14) and Liberia (15/15) as well as in the majority of districts in Guinea (27/34) [1]. Phylogenetic analysis suggests that the outbreak caused by the Makona strain was triggered by a single cross-species transmission event from an animal reservoir near Meliandou, Guinea, with the subsequent outbreak sustained exclusively by human-to-human transmission [2–4]. The rapid geographical expansion of the epidemic stands in stark contrast to previous outbreaks of EVD [5]. It has been hypothesised that the complex interplay between increased urbanisation and human mobility through porous borders in West Africa contributed to the catastrophic nature of this outbreak [6–8]. These changes in human behavioural conditions in part led to the geographic spread that subsequently overwhelmed the countries' poorly equipped health systems and revealed the lack of coordinated rapid response to address this international health crises [9,10].

Transmission dynamics of directly transmitted diseases such as EVD depend on the contact patterns between infected and susceptible individuals [11]. However, the dispersal of a pathogen in space and time is limited structurally by the distribution and nature of transport infrastructure [12], which in turn are influenced by the spatial heterogeneity of economic opportunities [13], and can vary due to events such as vacations [14], growing seasons [15], and religious events [16]. Geographic heterogeneity in the connectivity of populations is therefore a key factor determining disease transmission rates and geographic spread at fine spatial scales [17]. In the context of EVD, identifying and quantifying the factors that drive case importation, and disentangling the relative roles of different aspects of mobility on disease transmission is central in defining strategies to contain and control disease transmission [18,19].

Detailed investigations on chains of transmission in Guinea have shown that continued unmonitored reintroductions into large urban centres and subsequent inter-urban transmission events led to the extensive geographical spread of the virus [20]. The majority of models attempting to predict the regional spread of EVD were limited to a single country, often did not include specific human movement metrics [21,22] and did not assess important characteristics such as the relative contribution of transmission from each district over time [23]. One study using data from Sierra Leone focussed exclusively on the timing of arrival of the disease but did not include any human mobility

data [24]. Other studies attempted to anticipate the risk of international spread of EVD *via* commercial air travel [25,26]. Furthermore, phylogenetic studies of EVD in Sierra Leone and Liberia indicate that despite inter-country spread during the early phase (December 2013 to mid-March 2014) of the outbreak, most virus transmission occurred locally during the contracting phase of the outbreak and within national borders [2,4,27,28]. Some of these changes may be explained by unofficial border closings, curfews, and restrictions on funeral gatherings [29].

Epidemiological theory predicts that contacts between infected and susceptible individuals should be more frequent in large, densely-populated urban areas than in smaller communities in rural areas [30]. Recent advances in the availability of high resolution data on human mobility [31,32], new formulations of mathematical models to represent disease-related movement patterns [33], and the integration of such models in disease transmission models [34] now provide a comprehensive set of tools to enable detailed investigation of the dynamic drivers of EVD transmission.

Here we use openly available data on human mobility, the spatial configuration of districts and EVD case counts to investigate the dynamics of the geographic spread of the 2014-2016 West Africa EVD outbreak. We assess quantitatively the relative contribution to transmission of a range of mobility metrics and how these contributions changed over the course of the outbreak. This analytical framework has been integrated in an open-source software pipeline to enable real-time predictive mapping of EVD using publically available data (https://github.com/SEEG-Oxford/ebola-spread). This software can be rapidly updated, applied to other pathogens, and is flexible enough to be tailored to baseline analyses and predictive mapping of future infectious disease outbreaks.

Results

Human mobility matrices and predictive power of cases two weeks ahead: To understand the relative role of human mobility on the transmission parameters of the current EVD epidemic in West Africa we used a model for the growth and declining periods (before and after week 42) and did not add any additional covariates (*e.g.*, control and containment effort as district specific covariates) to identify the spatio-temporal heterogeneity during the course of the epidemic and across countries (*Materials and Methods*). Splitting the data at week 42 resulted in a significantly better fit for covariate-free 'base' transmission models given by equation 2 with $\beta_{t,i}$ set to 1 ($p < 0.001$, LRT statistic $\chi_2^2 = 16.62$). We confirm the utility of our extrapolated mobility models given the strong relationship between fitted mobility models to the IPUMS migration data (Pearson's correlation coefficient = 0.88 (CI 0.87 – 0.89, $p < 0.001$), Spearman Rank Coefficient = 0.9, Figure S7, Table S6, *Supplementary Information*).

Relative contribution of districts to overall transmission: Using mobile phone derived human mobility metrics and an adjacency network (Figure 1) revealed that in the expanding phase of the outbreak (weeks 1 – 42) the relative contribution to transmission is highest around the origin of the outbreak in Meliandou, Guinea (ranked 1st), yet the top three ranked districts are all located in three different countries (Kailahun, Sierra Leone, ranked 2nd, and Lofa, Liberia, ranked 3rd) (Figure 2a, *Materials and Methods*). See the full list of ranks for all districts (k = 63) in Table S1. In the first half of the outbreak, the districts that exported the fewest cases were rural areas in Guinea; indicating low transmission as a result of low connectivity (Table S1, Figure 2a). In the second half of the outbreak after the epidemic had reached its maximum geographic extent (week 42 – 83), the epidemic centre transitioned to the big urban centres in all countries (Figure 2b). During that time, the corridor between Freetown in Sierra Leone and Conakry in Guinea was the major contributor to transmission in the region (Figure 2b). The role of Liberia as a source of the outbreak also continued to decline in the second half of the epidemic. A dynamic map showing the trend over the whole outbreak is shown in Video S1 and Figure S6.

Transmission before week 42: For the first half of the outbreak, the 'base' (covariate free non-spatial) model fits well to the data (R^2-adjusted=0.6435) with a relatively high mixing coefficient of $\alpha = 0.80616$ (see [30] for more discussion of mixing coefficients within TSIR models, *Materials and Methods*). Adding a set of human mobility matrices and applying backward model selection using AIC significantly improved the model ($p < 0.001$, LRT statistic $\chi_9^2 = 38.46$) with moderate change in

α (0.666) and minor improvement in R^2-adjusted (0.668). The retained covariates are listed in Table S3 and predictions with 95% confidence intervals are shown in Figure 3.

Starting with the structure of the reduced model, the coefficients of $\log \beta$ varied significantly among countries ($p < 0.001$, LRT statistics $\chi^2_{22} = 52.96$). Repeating the modelling effort for each country independently, we again find that the incorporation of human mobility significantly improved the model (with a slightly different resulting reduced model than the above; $p < 0.05$, 0.001, and 0.05 for Guinea, Liberia and Sierra Leone respectively).

Country specific results: To investigate whether the country specific mobility metrics can improve local predictions we repeated the analysis for each country separately (*i.e.*, first fitting a covariate free model and then using backward selection on connectivity matrices to understand the effect of such metrics in each country separately). This analysis was also undertaken to understand if the outbreak was driven by local rather than between country movements. For Guinea, the base model (*i.e.*, covariate-free non-spatial) fitted poorly (R^2-adjusted = 0.384). When starting with only Guinea epidemiological data and covariates and reducing that 'full' model, we arrived at an improved model (R^2-adjusted = 0.516, Table S3, Figure S1 for each district). However, using AIC, the Guinea-only model was strongly preferred to the base model (357.583 and 368.242, respectively, Table 1).

In Liberia, the 'base' model fitted well (R^2-adjusted = 0.76) and the Liberia-only reduced model that includes connectivity matrices improved the fit significantly (R^2-adjusted = 0.82, Table S4, Figure S2 for each district). Unlike R^2-adjusted, AIC showed more separation from the three models with the Liberia-only reduced model (*i.e.*, including human mobility being strongly preferred (288.15 versus 304.8 for the base model, Table 1). This indicates a strongly locally driven epidemic structure where most of the variation in transmission can be explained by the local patterns of human mobility.

In Sierra Leone, the 'base' model fitted moderately well (R^2-adjusted = 0.63) and again, country specific Sierra Leone connectivity model improved the fit (R^2-adjusted = 0.71, Table S5, Figure S3 for each district). As with the Liberia models, AIC showed separation for the Sierra Leone-only reduced model being strongly preferred (AIC=339.3 versus 350.2 for the base model, Table 1).

Transmission after week 42: For the 'post week 42' data, the base model fitted moderately well to the data (R^2-adjusted = 0.557) with a mixing coefficient of $\alpha = 0.75610$. Adding human mobility matrices and applying backward model selection using AIC again significantly improved the model

($p < 0.001$, LRT statistic = $\chi_8^2 = 56.$) with large change α (0.596) and a minor improvement in R^2-adjusted (0.593).

Unlike in the model using data from the first half of the epidemic, differences in the parameter values of the covariates across the three countries when considered as a group were not significant ($p < 0.1$, LRT statistic = $\chi_{20}^2 = 29.52$). Also unlike the first half of the data, when considering each country independently, adding human mobility did not significantly improve the model for Guinea ($p > 0.5$). For Liberia and Sierra Leone however, the addition of mobility covariates improved model fit based on either R^2 or likelihood (using LRT and AIC). However, in all circumstances, the R^2-adjusted values were lower than the best fits from the first half of the data (Supplementary Information).

Invasion process: There was considerable heterogeneity in the invasion process so models differed from week to week. However, several covariates were retained in the majority of the models. Of the 28 models (from week 10 to week 37 where most of the invasions occurred), the total number of cases in each district was retained in almost all of the final models (26/28). Two of the four covariates that were retained in 25 of the 28 models were not location dependent: the total number of cases in Guinea each week, and the weighted sum of all cases using the gravity model in West Africa. The other two covariates that were retained in 25 of the 28 models were different for each location: the gravity model weighted sum of all cases that were both in Guinea and direct neighbours (if there were any) and the radiation models weighted sum of all cases that were direct neighbours. This confirms that gravity and radiation models, both capturing a different set of patterns of human mobility are important in understanding the invasion process (*Materials and Methods*).

As with the transmission models, the predictive ability of the invasion model varied from country to country (Figure S4). In general predictions 'one week ahead' were difficult to estimate due to the time lag in reporting and underlying uncertainties inherent in the reporting of the disease. When the time-steps were increased from one week to one month, however, the invasion model made accurate predictions of the invasion process evaluated as the timing of predicted invasion (AUC = 0.697, Figure S5). This trend was visible for almost the entirety of the outbreak (Figure S4).

Discussion

Our analyses found that human mobility metrics derived from mobile phone data alone can explain a considerable proportion of observed local EVD transmission dynamics. Further, we showed that there were temporally varying sources and sinks of virus transmission as well as significant spatio-temporal variation in future risk of invasion to previously unaffected districts (Figure 2, Video S1). Investigating the relative contribution of transmission shows that the centre of epidemic spread shifted early in the epidemic from the origin of the outbreak to the highly populated corridor along the coast between Conakry, Guinea and Freetown, Sierra Leone. We clearly identified that during the expanding phase of the epidemic (weeks 1-42) the districts predominantly contributing to transmission were located around the origin of the outbreak in Meliandou, Guinea. Also, our findings support the hypothesis that the rapid progression of the outbreak in West Africa was fuelled by the initial introduction into Kailahun, Sierra Leone and Lofa, Liberia, both adjacent to Gueckedou but under different jurisdictions [35]. Moreover, we show that during the entire outbreak, only Nimba county, Liberia, was a main contributor to transmission that was located along the border to an unaffected country (in this case Côte d'Ivoire) (Figure 2). This result helps to explain why transmission was not observed in any of the bordering countries; many of the border districts were sinks of EVD transmission rather than exporters (Video S1, Table S1). Our study confirms previous mechanistic modelling results that the early epidemic spread in Liberia can be approximated by using assumptions about human mobility from an agent based model [22].

We show an ability to predict the explosive geographic spread of EVD into districts in the core affected countries. Further, our model is sensitive to changes in human population sizes so it can be adjusted as populations grow. In all three countries the centrality of large population hubs (Conakry, Freetown and Monrovia) was identified as having facilitated rapid spread (Figure 2b).

For Liberia and Sierra Leone, intra-country dynamics seem to be more important drivers of transmission dynamics, whereas for Guinea country-specific population fluxes only slightly improve estimates of EVD case predictions. These findings extend previous work that looked at Liberia and Sierra Leone in isolation [21]. Our study also shows that both gravity and radiation human mobility models should be used in conjunction as there is considerable heterogeneity from week to week during the epidemic that may be explained better by the different types of human movement that the two models capture (commuting vs. longer distance movements). As our models are

automated in a near-real time framework, it allows for their rapid update which can be used to inform public health prevention and response decisions today, and in future outbreak situations [31].

During the course of a catastrophic outbreak people's everyday behaviours change due to public awareness campaigns, government restrictions or illness [36]. Our model identifies the relative importance of distinct human mobility measures that have changed over the first and second half of the outbreak (Table 1). In the first half of the outbreak country specific metrics were important in governing how the virus initially spread, even in the absence of real-time human movement data. Once the disease had spread into almost all corners of the affected countries the country specific covariates were less important. Instead, overall movement dynamics within and between all countries were responsible for improving the models' performance and differences between countries seemed to be less important. This indicates that our dynamic model, in the absence of real-time movement data, was able to discriminate between the full range of covariates and capture aspects of changing behaviour that may have contributed to the decline in case numbers [37,38]. Much of this may also be explained by the full deployment of containment activities such as safe burials that prevented onward infection [10,29]. Future work may be conducted using near real-time movement data to infer the ability to predict these changes in human behaviour more directly [39]. It is unfortunate to note that contemporary mobile phone data for Guinea, Sierra Leone, and Liberia are still unavailable and human movement data from other surrounding countries is not openly accessible. Such availability could help analyse the likely nature of pathogen flow prior to an outbreak and thus improve surveillance and containment preparedness plans [40,41].

We further confirm country specific phylogenetic studies that concluded that virus spread in Sierra Leone primarily occurred within national borders by showing that country specific covariates for Sierra Leone significantly improved the model fit [27]. Genomic surveillance, including real-time portable genome sequencing [42], may be used to extend our modelling approach by helping to identify the origins of an outbreak [2], monitor the diversity of the circulating viruses [4,28], characterize signatures of host adaptation [27], and pinpoint the source and sink locations of circulating strains [27,28,43–45]. For now, genetic analyses are often limited by the sample size, comparatively slow time until collection during an outbreak, and poor spatial coverage. At this stage, our analyses do not address the EVD outbreak from a genetic perspective, but our results provide a baseline to which further genetic results may be directly compared and used. We contend that future work should benefit from a joint incorporation of epidemiological, spatial and genetic data to use a full set of evidence that can be derived from different sources [31,46–48].

Highly adaptive response systems are required to mitigate the impact of disease outbreaks [49]. For example, in the case of the EVD outbreak the detection system failed due to the lack of surveillance and laboratory equipment. This led to inherent delays of response as the outbreak spread undetected in its early stage [50]. Modelling exercises should be applied beyond describing the dynamics of past outbreaks, but must harness the potential of machine-enabled learning processes to afford lessons for how these observed and predicted dynamics can inform structural improvements to public health systems [51,52].

The identification of transmission sources and sinks has broad application in disease control to identify where treatment and prevention measures would be best implemented preventing the rapid geographic spread of a pathogen. This has been shown for other diseases using historical data, but the modelling techniques presented here allow for application of near real-time data for the control of an ongoing outbreak. The methodologies have the potential to be used by national and international public health institutions to plan and perform effective control and surveillance systems to limit the geographic extent and burden of future outbreaks in areas with high potential emergence of contagious viral haemorrhagic fevers, as well as other directly transmissible infectious diseases [53].

Materials and Methods

Epidemiological data: We obtained weekly case data on numbers of EVD cases (both probable and confirmed) from the World Health Organization (WHO) starting January 2014 ending week 32 (August 3rd to 10th) 2015 after which only sporadic cases occurred. Details about the Situation Reports used to collate these numbers are described elsewhere [35].

Connectivity models: As empirical data on human movement patterns were not available for the three most affected countries (Guinea, Liberia and Sierra Leone) we instead predicted human movements between each pair of districts using three distinct mathematical models, each reflecting a different aspect of human mobility (Figure 1a). The three models were the gravity model $T_{i,j} = k\frac{N_i^{\alpha}N_j^{\beta}}{d_{i,j}^{\gamma}}$, the radiation model $T_{i,j} = T_i\frac{N_iN_j}{(N_i+s_{i,j})(N_i+N_j+s_{i,j})}$, and an adjacency network, where total commuting is $T_{i,j}$ from district i, to j; N_i^{α} is the population in the origin and N_j^{β} in the destination district; $d_{i,j}^{\gamma}$ the distance between them, and $s_{i,j}$ the population in the radius between i and j. The gravity model assumes that relative flow between districts is a log-linear function of the population of the districts and the distance between them [33,54]. This model therefore emphasises the attractive power of large population centres. The radiation model also accounts for origin and destination population sizes and distances, but in addition considers the draw from other populations within the same radius [13]. This model therefore reflects likely patterns of commuting for work assuming every locality has a competing underlying attractiveness. Adjacency networks encode the number of district borders an individual would need to cross to move from one district to another. This metric thus reflects the impacts of national and sub-national borders on movements in the region. Each of these models have been shown to be useful depending on the local context to infer regular daily commuting patterns, longer term movements and the general population diffusion process [13,55,56]. We used all three metrics, as well as interactions between them, as terms in disease spread models [57–61].

Connectivity data: Both the gravity and radiation models have variable parameters that determine the patterns of movement that they predict. These parameters can be deduced by fitting the models against empirical data on human movements. Common data sources for training these models include census commuting or mobile phone data (call detail records) [31]. Since no such data were available for the West Africa region, we instead trained models against three high resolution,

openly available movement datasets derived from call detail records, representing movement between districts in France, Spain, and Portugal [59]. These movement matrices have been used successfully to predict movements in developing countries [62]. To show the ability to predict these movements with high-resolution mobile phone data, we compare our estimates to long-term migration data from the Integrated Public Use Microdata Series (IPUMS, https://international.ipums.org/international). These data represent 10% samples of the total population at the individual level from the national censuses conducted in Guinea, Liberia and Sierra Leone in 1996, 2008 and 2004, respectively. Census questions about where respondents lived a year ago were used to quantify migration flows across the period of the year before the census between administrative units at the finest level available. Further details can be found in Wesolowski et al. [54]and Sorichetta et al. [63]. They have previously been shown to be a good representation of shorter term movements [64]. A full list of correlation coefficients is provided in the Supplementary information.

Connectivity metrics: For each district in West Africa (n = 63), we determined the total human population size using gridded population estimates and we calculated the distance between the centroids of each pair of districts [65]. Gravity and radiation model parameters were fitted to the empirical data described above and applied to the three core affected countries using the movement R package [66]. National adjacency networks were computed using administrative boundary data from the GADM dataset (http://www.gadm.org). This adjacency matrix was then disaggregated into three binary connectivity matrices with connectivity degrees of one (*i.e.,* districts share a border), two (*i.e.,* districts share a common neighbour), three, and more than three.

Feature engineering: EVD transmission in the core-affected countries is likely to be influenced by human mobility at a variety of spatial and temporal scales, with different aspects of movement varying in importance through the course of the epidemic and among countries. Due to the paucity of human movement data in the region, it is not possible to directly quantify all aspects of mobility relevant to disease transmission. Instead, we carried out a feature engineering procedure to produce a large number of candidate covariates reflecting non linearities and interactions between the features and response [67]. We then applied statistical inference procedures that are able to select a sparse optimal set of features that minimize the generalisation error of the model. Whilst model generalizability and interpretation are generally aided by using a parsimonious set of covariates,

this approach reflected our main aim here, which was to maximise the predictive accuracy of the model.

Each of the resulting six fitted connectivity models was then used to predict human mobility between all districts in Guinea, Sierra Leone, and Liberia and for each country separately (Figure 1a). Each of the metrics was then weighted by different sets of adjacency (Figure 1b) and iterated through each week depending on the cases in each district (Figure 1c).

Disease model specifications: To model the effect of human mobility on the geographic spread and rate of transmission of EVD within and between the three core countries we used a two-stage model to characterise both geographic expansion (*i.e.,* introduction into previously unaffected districts) and the expected secondary cases arising from these introductions.

Invasion model: The invasion model estimates the probability $p_i(t)$ that one or more cases will be identified in previously disease-free district i at time t (with presence or absence of new cases indicated by $Y_i(t)$), as a function of the number of cases $C_{-i}(t-1)$ in all other districts in the previous time point which is chosen to be bi-weekly in our analysis, the product of corresponding values of the connectivity covariates $x_{j,-i}$ for each covariate j, regression coefficient b_j and a fixed intercept term a. This gives the following logistic regression model:

$$Y_i(t) \sim Bernoulli(p_i(t))$$

$$logit(p_i(t)) = a + \sum_j \sum_i b_j x_{j,-i} C_{-i}(t-1) \qquad (1)$$

Transmission model: For all districts reporting one or more cases, we assume a general bi-weekly transmission model following [68]:

$$I_{t,i} = \beta_{t,i} * \frac{I_{t-1,i}^{\alpha_i}}{N_i} * S_{t-1,i} * \epsilon_{t,i} \quad (2)$$

where $I_{t,i}$ is the number of infected and infectious individuals and $S_{t-1,i}$ the number of susceptible individuals, at time t in district i, N_i is the population of district i and $\beta_{t,i}$ is the covariate-driven mobility rate characterised by a linear combination of the mobility metrics described above (*i.e.,* $\beta_{t,i}$ has the same model structure as the final line in equation (1), although parameters are fitted independently). α_i represents an approximation to how well mixed the population is in district *i.* The error terms $\epsilon_{t,i}$ are independent, identically log-normally distributed random variables. We further assumed all individuals to be initially susceptible to infection. Before including a set of

connectivity matrices to test the effect of such inclusion we fit a covariate-free model that assumes no interaction between districts.

Calculation of source and sink districts: To pinpoint which district had the highest contribution to transmission in West Africa at each time-step (t, week) we calculate the relative weights of each district by converting equation (2) into a linear regression of the form:

$$y_{i,t} = \log \beta_i + \alpha * x_i + \log \epsilon_i \quad (3)$$

with

$$\log \beta_{i,t} = S_1 * X_{1j} + S_2 * X_{2j} + \cdots + S_k * X_{kj} \quad (4)$$

Where, for any district *i,* each X_{ij} is one of k district-specific covariate that combines how many cases there are in all the other districts weighted by a connectivity matrix. We can rearrange equation (4) so that $\log \beta_i$ for district *i* (*i* = 1, 2, 3, .. ,63) is a function the number of cases in every other district (X_j):

$$log\beta_i = \sum_{\substack{j=1 \\ j \neq i}}^{63} \gamma_{ij} x_j \quad (5)$$

Full details of how the relative contributions for each district are derived can be found in the supplementary information.

Invasion model evaluation: The invasion models estimate the probability of invasion for each district that has not already been invaded using all the data up to that time point. To assess the predictive performance of the invasion models under a realistic real-time scenario, we re-fitted the model for each week of the epidemic, each time using only the information available from the preceding weeks, up to and including the week in question. The week's model was then used to predict the probability of invasion in the following week or the following month in all districts currently disease-free. These predictions were then compared with the observed invasions in these districts on the following week. Rather than determining a fixed threshold probability value with which to evaluate the model's predictive power, we computed a receiver operating characteristic (ROC) curve to evaluate predictive power under all possible threshold values. The area under curve (AUC) represents the probability that a district drawn uniformly at random where an invasion is going to occur would receive a higher "probability of invasion" than a randomly drawn district where no invasion occurs. The closer the value is to 1, the better the predictive power of the model.

Transmission model selection: In general $\beta_{t,i}$ terms were fitted entering the covariates linearly following previous work [69–71]. All model fitting was conducted in R and model selection was conducted using the Akaike Information Criterion (AIC) and Likelihood Ratio Tests (LRT); with AIC used to identify reduced models and LRT to compare final models with simpler nested models [30]. To identify country specific effects the transmission model was conducted using two approaches. First, a single model was fitted for data from all countries and then reduced using backwards selection and AIC. Secondly, the reduced model structure was fitted independently for each country again using backwards selection and AIC.

Phases of the epidemic: Both the geographic extent of the outbreak, and the weekly numbers of reported cases peaked and then declined as the outbreak was brought under control. The extent of the epidemic was greatest in week 42 (October 13th to 19th) of 2014, four weeks after the peak in weekly numbers of cases (week 38, September 15th to 21st). In order to account for the impact of human mobility processes contributing to transmission during the epidemic's expansion and contraction, we split the data in two and separately analysed the first half of the epidemic (expanding phase) until week 42 of 2014 and the second half of the epidemic (contracting phase) from week 43 to 84. The invasion process model was applied only to the growth phase of the epidemic, whilst the transmission model was fitted separately to the first and second halves respectively.

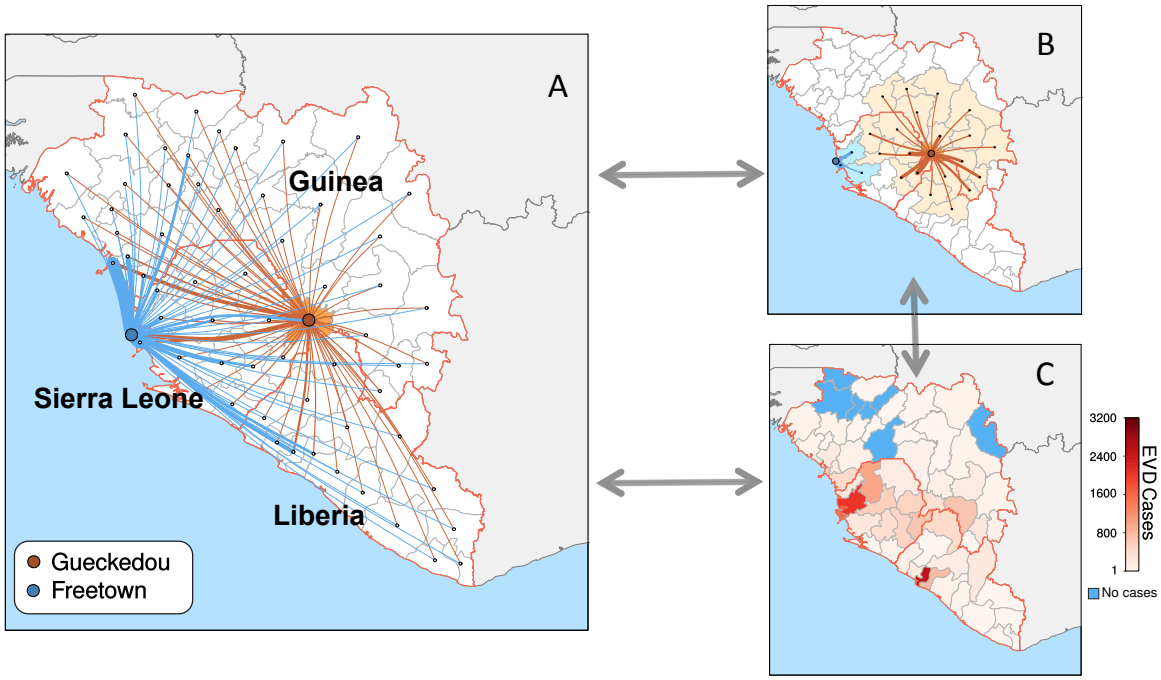


Figure 1: To account for different patterns in movement that might contribute to how the epidemic spread we constructed a comprehensive database that combines different attributes of movement. (A) shows the connections between Gueckedou, where the outbreak started and all other districts in the region using a gravity model. We further included Freetown to highlight the different strengths of connections that result from the pull of a large city. (B) shows second degree adjacent districts. (C) shows the total distribution of cases as of March 17th, 2016.

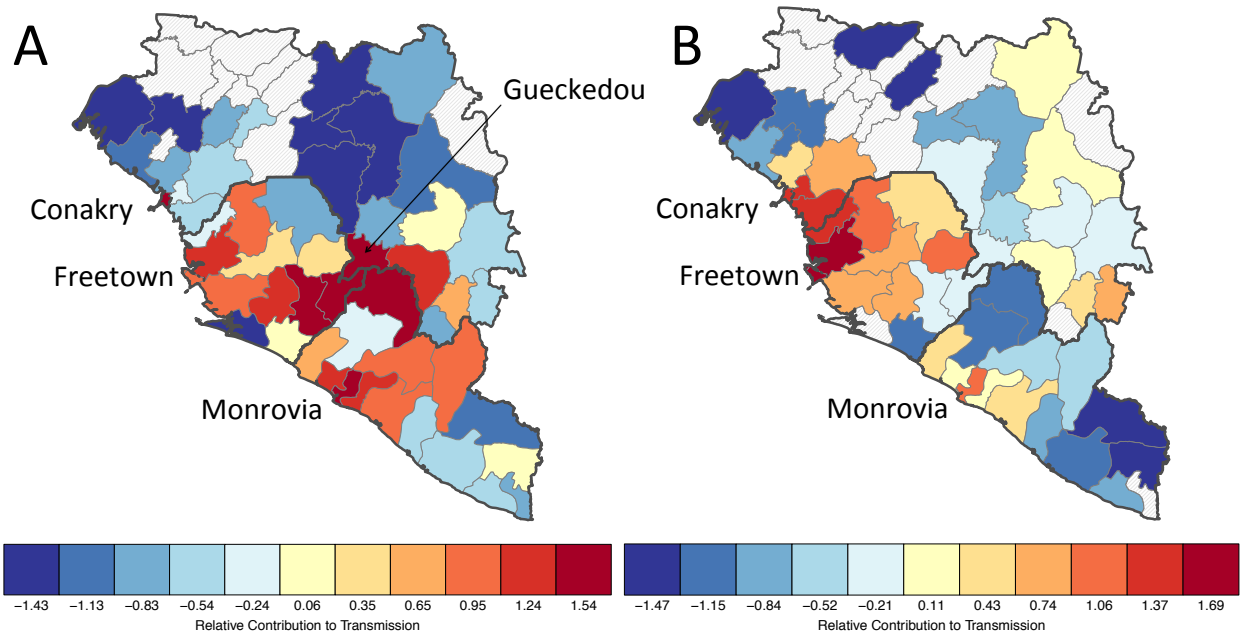


Figure 2: Relative contribution to transmission in the expanding phase of the outbreak in West Africa (week 1-42, panel A) and the second half of the outbreak (42 – 83, panel B). Red shows sources of transmission measured how much they contributed to transmission elsewhere. Blue shows districts that are contributing less to the spread of EVD.

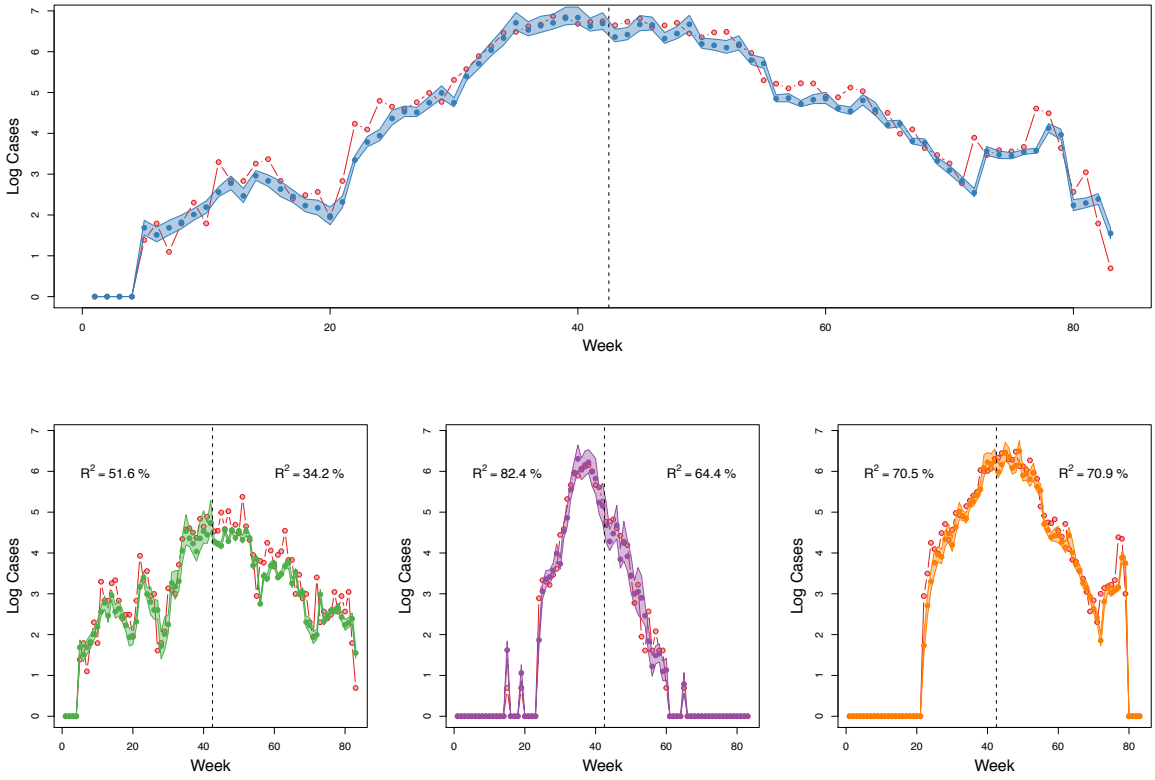


Figure 3: Observed (probable and confirmed) vs. two week ahead predicted transmission in all three core countries (top panel), in Guinea, Liberia, and Sierra Leone for both the expanding and the contracting phase of the epidemic, from left to right respectively. Red lines representing observed cases. 95% CI intervals are given for the predicted cases.

Table 1: Summary of modelling results (adjusted R^2 and Akaike Information Criterion) for the covariate free model for all countries and with human mobility. In addition the country specific results for the full model and country covariates are shown.

		Expanding Phase (Week 1 - 42)	
All countries		Covariate free	With human mobility
	R^2	0.6435	0.668
Guinea		Full model	Country specific model
	R^2	0.429	0.52
	AIC	368.242	357.583
Liberia			
	R^2	0.79	0.82
	AIC	297.67	288.15
Sierra Leone			
	R^2	0.65	0.71
	AIC	350.5	339.3

References

1. World Health Organization. Ebola Situation Report. 16 March 2016 (2016).
2. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345,** 1369–1372 (2014).
3. Schieffelin, J. S. *et al.* Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone. *N. Engl. J. Med.* **371,** 2092–2100 (2014).
4. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524,** 97–101 (2015).
5. Mylne, A. *et al.* A comprehensive database of the geographic spread of past human Ebola outbreaks. *Sci. Data* **1,** 140042 (2014).
6. Farrar, J. J. & Piot, P. The Ebola emergency — immediate action, ongoing strategy. *N. Engl. J. Med.* **371,** 1545–6 (2014).
7. Maganga, G. D. *et al.* Ebola Virus Disease in the Democratic Republic of Congo. *N. Engl. J. Med.* **371,** 2083–2091 (2014).
8. Piot, P. Ebola's perfect storm. *Science* **345,** 1221 (2014).
9. Gates, B. The next epidemic - lessons from Ebola. *N. Engl. J. Med.* **372,** 1381–1384 (2015).
10. Kucharski, A. J. *et al.* Measuring the impact of Ebola control measures in Sierra Leone. *Proc. Natl. Acad. Sci. USA* **112,** 14366–71 (2015).
11. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control*. (Oxford University Press, 1991).
12. Uchida, H. & Nelson, A. *Agglomeration index: towards a new measure of urban concentration*. (2008).
13. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484,** 96–100 (2012).
14. Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **111,** 15888–93 (2014).
15. Bharti, N. *et al.* Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334,** 1424–7 (2011).
16. Lessler, J. *et al.* Estimating potential incidence of MERS -CoV associated with Hajj pilgrims to Saudi Arabia, 2014. *PLOS Curr. Outbreaks* **November 2,** (2014).
17. Wesolowski, A. *et al.* Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc. Natl. Acad. Sci. USA* **112,** 11114–11119 (2015).
18. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437,** 209–14 (2005).
19. Longini, I. M. *et al.* Containing pandemic influenza at the source. *Science* **309,** 1083–7

(2005).

20. Faye, O. *et al.* Chains of transmission and control of Ebola virus disease in Conakry , Guinea , in 2014 : an observational study. *Lancet Infect. Dis.* **15,** 320–326 (2015).

21. Valdez, L. D., Aragão Rêgo, H. H., Stanley, H. E. & Braunstein, L. A. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. *Sci. Rep.* **5,** 12172 (2015).

22. Merler, S. *et al.* Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *Lancet Infect. Dis.* **15,** 204–211 (2015).

23. Chretien, J., Riley, S. & George, D. B. Mathematical modeling of the West Africa Ebola epidemic. *Elife* **4,** e09186 (2016).

24. Yang, W. *et al.* Transmission network of the 2014 – 2015 Ebola epidemic in Sierra Leone. *J. R. Soc. Interface* **12,** 1–9 (2015).

25. Bogoch, I. I. *et al.* Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* **6736,** 1–7 (2014).

26. Poletti, M. *et al.* Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveill.* **19,** pii=20936 (2014).

27. Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161,** 1516–1526 (2015).

28. Ladner, J. T. *et al.* Evolution and Spread of Ebola Virus in Liberia, 2014-2015. *Cell Host Microbe* **18,** 659–669 (2015).

29. Fang, L. *et al.* Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl. Acad. Sci. USA* PNAS Early Edition (2016). doi:10.1073/pnas.1518587113

30. Kraemer, M. U. G. *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12,** 20150468 (2015).

31. Kraemer, M. U. G. *et al.* Progress and challenges in infectious disease cartography. *Trends Parasitol.* **32,** 19–29 (2016).

32. Tatem, A. J. Mapping population and pathogen movements. *Int. Health* **6,** 5–11 (2014).

33. Wesolowski, A., O'Meara, W. P., Eagle, N., Tatem, A. J. & Buckee, C. O. Evaluating Spatial Interaction Models for Regional Mobility in Sub-Saharan Africa. *PLOS Comput. Biol.* **11,** e1004267 (2015).

34. Xia, Y., Bjørnstad, O. N. & Grenfell, B. T. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164,** 267–281 (2004).

35. WHO Ebola Response Team. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. *N. Engl. J. Med.* (2014). doi:10.1056/NEJMoa1411100

36. Fenichel, E. P. *et al.* Adaptive human behavior in epidemiological models. *Proc. Natl. Acad. Sci. USA* **108,** 6306–11 (2011).

37. Funk, S. & Knight, G. M. Ebola: the power of behaviour change. *Nature* **515,** 492–492 (2014).

38. Funk, S., Gilad, E., Watkins, C. & Jansen, V. a a. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 6872–7 (2009).

39. Wesolowski, A. *et al.* Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci. USA* **112,** 11887–11892 (2015).

40. Pigott, D. M. *et al.* Mapping the zoonotic niche of Marburg virus disease in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109,** 366–378 (2015).

41. Mylne, A. Q. N. *et al.* Mapping the zoonotic niche of Lassa fever in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109,** 483–492 (2015).

42. Rambaut, A. Phylogeography of 2014-2015 Ebola virus epidemic. *virological.org* at <http://virological.org/t/phylogeography-of-2014-2015-ebola-virus-epidemic/199>

43. du Plessis, L. & Stadler, T. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol.* **23,** 383–386 (2015).

44. Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance — the time is now. *Genome Biol.* **16,** 155 (2015).

45. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).

46. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303,** 327–32 (2004).

47. Pybus, O. G., Tatem, A. J. & Lemey, P. Virus evolution and transmission in an ever more connected world. *Proc. R. Soc. B Biol. Sci.* **282,** 20142878 (2015).

48. Currie, J., Grenfell, B. & Farrar, J. Beyond Ebola. *Science* **351,** 815–16 (2016).

49. Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* **8,** 32–39 (2012).

50. Gignoux, E. *et al.* Use of Capture–Recapture to Estimate Underreporting of Ebola Virus Disease, Montserrado County, Liberia. *Emerg. Infect. Dis.* **21,** 2265–2267 (2015).

51. Levin, S. A. *et al.* Social-ecological systems as complex adaptive systems: modeling and policy implications. *Environ. Dev. Econ.* **18,** 111–132 (2013).

52. May, R. M., Levin, S. A. & Sugihara, G. Complex systems: ecology for bankers. *Nature* **451,** 893–895 (2008).

53. Pigott, D. M. *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* **3,** e04395 (2014).

54. Wesolowski, A. *et al.* Commentary: Containing the Ebola outbreak – the potential and challenge of mobile network data. *PLOS Curr. Outbreaks* **Sep 29,** (2014).

55. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342,** 1337–42 (2013).

56. Jongejans, E. *et al.* A unifying gravity framework for dispersal. *Theor. Ecol.* **8,** 207–223 (2015).

57. Balcan, D., Colizza, V., Gonc, B. & Hu, H. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106,** 21484–89 (2009).

58. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338,** 267–70 (2012).

59. Tizzoni, M. *et al.* On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10,** e1003716 (2014).

60. Simini, F., Maritan, A. & Néda, Z. Human mobility in a continuum approach. *PLoS One* **8,** e60069 (2013).

61. Reiner, R. C. *et al.* Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc. Natl. Acad. Sci. USA* **109,** 2033–6 (2012).

62. Nunes, M. R. T. *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med.* **13,** 102 (2015).

63. Sorichetta, A. *et al.* Mapping internal connectivity through human migration in malaria endemic countries. *Sci. Data* **3,** 160066 (2016).

64. Ruktanonchai, N. W. *et al.* Census-derived migration data as a tool for informing malaria elimination policy. *Malar. J.* **15,** 273 (2016).

65. WorldPop project. WorldPop. at <http://worldpop.org.uk/>

66. Golding, N., Schofield, A. & Kraemer, M. U. G. Movement: Functions for the analysis of movement data in disease modelling and mapping. *R Packag. version 0.2* (2015).

67. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350,** 1073–76 (2015).

68. Finkenstädt, B. F. & Grenfell, B. T. Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.* **49,** 187–205 (2000).

69. Hastie, T. J. & Tibshirani, R. J. *Generalized additive models*. (CRC Press, 1990).

70. Dominici, F., McDermott, A., Zeger, S. L. & Samet, J. M. On the use of generalized additive models in time-series studies of air pollution and health. *Am. J. Epidemiol.* **156,** 193–203 (2002).

71. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73,** 3–36 (2011).

72. R Core Team. R: A language and environment for computing. Vienna, Austria. *R Found. Stat. Comput.* (2014).

Chapter 6 – The spatial spread of a Yellow fever virus outbreak in Central Africa

---

In this chapter the impact of both, demography, mobility, and ecology are analysed in an integrated framework. It draws largely on the *Aedes* maps from chapter 1 and extends along the lines of chapter 4 and 5 including explicitly human mobility metrics. By doing so the importance of both, ecological and demographic factors are revealed. This approach is exemplified using an urban Yellow fever outbreak in Central Africa in 2015 and 2016. This work is submitted for publication and is included here formatted according to journal guidelines. Additional information is available in the Appendix.

Spread of Yellow Fever Virus outbreak

in Angola and the Democratic Republic Congo 2015-2016

Kraemer, M.U.G., Faria, N.R., Reiner Jr., R.C., Golding, N., Nikolay, B., Johansson, M.A., Salje, H., Faye, O., Wint, G.R.W., Niedrig, M., Shearer, F.M., Hill, S.C., Bisanzio, D., Nax, H.H., Pradelski, B.S.R., Murphy, N.R., Bogoch, I.I., Khan, K., Brownstein, J.S., Tatem, A.J., de Oliveira, T., Smith, D.L., Sall, A., Pybus, O.G., Hay, S.I., Cauchemez, S.

Abstract

Background: Since late 2015, an epidemic of Yellow fever virus (YFV) has caused over 6,554 suspected cases in Angola and the Democratic Republic of Congo, including 387 deaths. We sought to understand the spatial spread of this YFV outbreak to optimise the use of the limited available vaccine stock.

Methods: We jointly analysed datasets describing the epidemic of YFV, vector suitability, human demography and mobility in Central Africa in order to understand and predict the expansion of YFV. We used a standard logistic model to infer the district YFV infection risk over the course of the epidemic in the region.

Findings: Early spread of YFV was characterized by fast exponential growth (doubling time of 5-7 days) and fast spatial expansion (49 districts reporting cases after only three months) from Luanda, the capital of Angola. Early invasion was positively correlated with high population density (0·52, 95% CI: 0·34, 0·66). The further away locations were from Luanda the later the invasion date (0·60, 95% CI: 0·52, 0·66). Districts with higher population densities also featured higher risks of sustained transmission. A model that captured human mobility and vector suitability successfully discriminated districts with high risk of invasion from others. If at the start of the epidemic sufficient vaccines had been available to target 50 out of 313 districts in the area, our model would have correctly identified 27 (84%) of the 32 districts that were eventually affected.

Interpretation: Our findings reveal the contributions of ecological and demographic factors to the ongoing spread of the YFV outbreak and provide estimates for where vaccines may be prioritised, although other constraints (e.g. vaccine supply and delivery) need to be accounted for before such insights may be translated into policy.

Funding: Wellcome Trust

Research in context

Evidence before this study: We searched the PubMed database for previous publications on any Yellow fever virus (YFV) outbreak in Angola and the Democratic Republic Congo (DRC), YFV outbreaks since beginning of 2015, and the spatial spread of YFV (search date 11/10/2016). Used search terms were: ("yellow fever"[MeSH Terms] OR "yellow fever"[All Fields]) AND (("angola"[MeSH Terms] OR "angola"[All Fields]) OR ("congo"[MeSH Terms] OR "congo"[All Fields] OR "DRC"[All Fields) OR (("epidemic"[All Fields] OR "epidemics"[MeSH Terms] OR "disease outbreaks"[MeSH Terms]) AND ("2015/01/01"[PDAT]:"3000/12/31"[PDAT])) OR "spatial"[All Fields]). Very limited empirical information is available for the current YFV outbreak in Angola and DRC. Previous work has described the identification of the outbreak, discussed vaccination coverage and the apparent vaccine shortage from a policy perspective, and highlighted the risk of international spread outside the region.

Added value of this study: To our knowledge this work is the first to investigate key epidemiological parameters of the outbreak and reveal the important ecological and demographic determinants that govern transmission and spread of the virus in the region. In addition, this study provides timely data and estimates to help decision-makers to deploy limited vaccine stock to areas at highest risk of pathogen spread.

Implications of all the available evidence: Across the region the spread of YFV was governed by high population density, including locations that were distant from the origin of the outbreak in Luanda. Transmission is also more likely to be sustained in areas where people live in close proximity. Human movements in and out of the capital cities of Angola and the DRC have led to YFV spread to almost all districts in Angola, and there is a continued threat the virus may spread throughout the DRC. Our model fits the expansion process of the pathogen well and allows for extrapolation into the future. Our results can be used to help policy makers prioritise areas to be targeted, especially in the context of finite public-health resources.

Introduction

Yellow fever virus (YFV) is a mosquito-borne flavivirus that causes infections among humans with symptoms ranging from mild non-specific illness to severe disease with jaundice, haemorrhage and death.[1] A single-dose vaccine has been rolled out since the 1940s and has helped to control and reduce YFV transmission dramatically.[2–4] Complete eradication however is prevented by the sylvatic cycle of the virus within which non-human primates act as primary hosts and *Aedes aegypti* mosquitoes are responsible for occasional transmission to humans.[5,6]

Since December 2015, a large YFV outbreak has been ongoing in Angola and the Democratic Republic Congo (DRC) with a total of 6,554 suspected cases of which 959 were confirmed and 387 deaths reported to the World Health Organization as of September 2nd 2016.[7] The response to such outbreaks relies largely on reactive vaccination campaigns and poses a number of strategic and logistical challenges. For example, in the current Central Africa outbreak, the global YFV vaccine emergency stockpile (6 million doses) was exhausted after the initial mass vaccination campaign.[7] In a context of limited resources, decisions about geographic areas that should be targeted first need to be informed by a detailed understanding of the determinants of YFV spatial spread and by predictions of where YFV is most likely to expand in the future.

Such assessments need to capture a multiplicity of factors. First, human mobility may facilitate the introduction of the pathogen into disease-free areas, a phenomenon that has been documented for other outbreaks in Africa.[8,9] Such regional spread of a disease is largely governed by underlying population structures and transport networks, as well as by individuals' economic, cultural and recreational activities.[10,11] Furthermore, in the context of a vector-borne disease such as YFV, the ecological landscape of the *Ae. aegypti* mosquito that transmits the virus between humans must also coincide with patterns of human movements for a successful viral transmission cycle to be established.[12,13] This ecological landscape has been shown to be strongly influenced by temperature, precipitation, humidity, vegetation coverage, and level of urbanisation.[14]

Here, we jointly analysed datasets describing the epidemic and spatial spread of YFV, vector suitability, human demography and mobility in Central Africa with a view to better understand, quantify and predict the spread of YFV in this region. By identifying areas that are at highest risk of YFV transmission, such a framework may inform the response to ongoing outbreaks.

Materials and Methods

Epidemiological data: Epidemiological data on confirmed and suspected YFV cases in the most affected countries, Angola and the DRC were obtained from the World Health Organization Situation Reports (http://www.afro.who.int/en/yellow-fever/sitreps.html) for the period December 2015 to August 2016. First and last dates of disease onset were extracted for each district in Angola (n = 163) and for each commune in the DRC (n = 150). Weekly numbers of cases in Angola were used. For the DRC, weekly case numbers were not available at the time of collection (August 2016). Locations were matched with administrative unit files available from the Database of Global Administrative Areas (http://www.gadm.org/).

Aedes aegypti data: To generate a district level ecological risk of disease transmission, estimated suitability surfaces for the primary urban vector *Ae. aegypti* were extracted at the lowest administrative level (n = 313).[14,15] These estimates combine climatic and socio-economic variables to obtain an estimate of vector suitability and have previously been used to estimate the time-varying risk of emergence and spread of arboviruses.[16,17]

Human movement data and models: We considered a total of eight connectivity metrics that capture different aspects of connectivity between two districts and have been shown to be useful to infer regular daily commuting patterns, longer term movements, and the general human diffusion process.[8,18] A first set of metrics included:

1. Distance: We calculated the great circle distance between the centroids of each pair of districts.[19].
2. Travel time distance: We calculated the travel time between each pair of districts. This metric reflects the underlying transport network that has been shown to reflect population mixing (Supplementary Information Figure S3).[16]

A second set of metrics was derived from standard models describing human mobility:

3. The gravity model assumes that relative flow between districts is a log-linear function of the population sizes of the districts and the distance between them (functional form provided in Supplementary Information).[20] This model therefore emphasises the attractive power of large population centres such as the capital cities Luanda and Kinshasa (Supplementary information Figure S2).

4. The radiation model additionally considers the draw from other populations within the same radius of the districts considered, as well as the population sizes and distance of the origin and destination locations (functional form provided in Supplementary Information).[18] This model therefore reflects recurrent workplace commuting, assuming that every locality has a competing underlying attractiveness.
5. Adjacency network metrics, which encode the number of district borders an individual would need to cross to move from one district to another. This metric thus reflects the impacts of national and sub-national borders on movements within the region. This adjacency matrix was then disaggregated into three binary connectivity matrices with
    a. connectivity degrees of one (*i.e.,* districts that share a border),
    b. connectivity degrees of two (*i.e.,* districts that share a common neighbour), and
    c. connectivity degrees of more than two.

A full list of variables used in this analysis is available in the Supplementary information (Table S1).

To calibrate the gravity and radiation models, we used aggregated and de-identified mobile phone-derived mobility estimates from October 2010 to September 2011 from Namibia at the constituency level. These data measure the proportion of time that unique SIM cards in each constituency spend in all other constituencies, and are described in detail in Ruktanonchai et al.[21] We used data from Namibia as it is directly neighboring the study area and has a similar per capita gross domestic product (GDP). Data from the study region was not available at the time of analysis. The models were then used to predict between district mobility for Angola and the DRC using the movement package in R (https://github.com/SEEG-Oxford/movement).[2221,27,28] National adjacency networks were computed using administrative boundary data from the GADM dataset (http://www.gadm.org). For each administrative region in Angola and the DRC (n = 313) we determined the total human population size using gridded population estimates

Change in the rate of spatial expansion over time: We tested the hypothesis that the rate of spatial expansion changed over time by introducing a time-varying binary variable that was equal to 0 before change point T and to 1 after T. In our baseline scenario, we assumed that the week of the change point corresponded to the rollout of mass vaccination in the region (week nine of 2016); but explored alternative scenarios in a sensitivity analysis.

A model for the geographical spread of YFV: Let $x_i(t)$ be the infection status of district i at time t (i.e., a binary variable takes the value 1 if there were infections that time step, and 0 otherwise). We

assume that districts were infected in all time-steps between the first and last reported case as described in the World Health Organization reports. We used a standard logistic model to characterize the probability district *j* will become infected at time t:

$$logit\left(P\left(x_j(t) = 1 | x_j(t-1) = 0\right)\right) = \beta_0 + \sum_{k=1}^{n} \beta_k Y_{j,t}^k + \epsilon$$

where Y^k corresponds to explanatory variable k and ϵ is an error distributed by the standard logistic distribution. Explanatory variables included in this analysis are: vector suitability and after/before rollout of vaccination. In addition, for each connectivity metric described above that quantifies the connectivity $A_{ij}^{(k)}$ between districts *i* and *j*, we derived the global force of infection exerted from all infected districts to *j*: $Y_{j,t}^k = \sum_i A_{ij}^{(n)} x_i(t-1)$. A total of 11 explanatory variables were therefore considered in this analysis.

Univariate and multivariable analyses: Our model for the geographic spread of YFV was used to investigate the contribution of single variables in an univariate analysis and then in a multivariable framework.

Model accuracy: To assess model accuracy, we calculated the weekly probability of invasion p_W predicted by the model for each district-week that had not yet reported cases. We then partitioned district-weeks into eight groups whose predicted probability was in the range [0%, 1%], [1%, 5%], [5%, 10%], [10%, 15%], [15%, 20%], [20%, 25%], [25%, 35%], [35%, 100%], respectively. For each group, we calculated the mean predicted probability and compared it with the proportion of district-weeks in the group where invasion effectively took place. The monthly probability of invasion can be calculated with the formula $p_M = 1 - (1 - p_W)^{4.33}$ since there are 4.33 weeks per month.

Predictive power and optimisation of prioritisation strategy: In a context of limited resources, we assessed how this analytical framework could have helped inform prioritisation of districts to be targeted for intervention at the time when such insight was most needed, *i.e.* on week nine of 2016 when the vaccination campaign started. We fitted the model to data available up until that week, then ran 1,000 simulations of the model to infer the probability of invasion in the following month for each district. In each simulation, a district may become infected in a certain week and then contribute to invasion risk in that particular simulation. We assumed that invaded districts would not become disease-free within the short simulation period. Assuming only *n* districts can be

targeted for intervention, we compared the performance of a prioritisation strategy that targeted the *n* districts with the highest predicted invasion probability versus a strategy that targeted *n* districts chosen at random.

Forward prediction: The same approach was used to predict the possible expansion of YFV one to two months ahead. We fitted the model to the entire epidemic from week 49 in 2015 to late August 2016 and then used the model parameters to simulate the geographic spread forward in time.

Epidemic doubling time and reproduction number: We estimated the exponential growth rate and the doubling time of the epidemic in the early phase when the number of cases was growing exponentially. We reconstructed the generation time distribution of YFV (*i.e.*, the time lag from the infection of a case to infection of subsequent cases in the chain of transmission) from prior knowledge of the natural history of YFV infection in humans and in mosquitoes. The reproduction number was then derived from standard formula linking it to the exponential growth rate and the generation time distribution.[24] Technical details are presented in the Supplementary Information.

Sensitivity to time-varying reporting rates: Our spatial model is based on a detailed analysis of the dates when districts reported their first YFV case. If reporting of cases increased during the course of the epidemic, the delay between the first infection and the first reported case should have shortened over time. We therefore performed a simulation study in which, for each district, the date of first infection was reconstructed from the date of first report under the assumption of an average delay of four weeks in the early phase of the epidemic (until week nine of 2016) and of one week in subsequent weeks.

Role of the funding source:

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The first YFV cases were reported from Luanda, the capital of Angola, in the first week of December 2015. Between week one and week five of 2016, the number of cases grew exponentially with a doubling time of 6 days (95% CI: 5-7 days) (Figure 1a). Under the assumptions that reporting remained stable during this period and that the generation time had a mean of 15 days and a standard deviation of 6 days (Supplementary Information), we estimated that the reproduction number of YFV was 4·8 (95% CI: 4·0, 5·6).

Although the epidemic was initially focussed in Luanda, it quickly expanded to other districts such as Belas, Lobito, and Huambo (Figure 1b and 2a). By the end of February, 49 districts had reported cases (Figure 1b) with the proportion of cases from Luanda dropping from 75% in early February 2016 to about 30% in April 2016 (Figure 1a).

In a simple univariate correlation analysis, we found that earlier timing of invasion of a district with YFV was correlated with greater population density (Pearson correlation 0·52, 95% CI: 0·34, 0·66, Figure 2a, Supplementary information, Figure S4). In addition, the further away from Luanda places were (by distance and by travel time) the longer it took for them to be invaded (Pearson correlation for distance 0·60, 95% CI: 0·52, 0·66; Pearson correlation for travel time 0·63, 95% CI: 0·56, 0·69; Supplementary information, Figure S5A and Figure S5B respectively). In contrast, early invasion was not correlated with *Ae. aegypti* suitability which is abundant in most of the study area.[14]

We also investigated the duration of persistence of the virus in a district after it was invaded. In a survival analysis, the probability that a district was still reporting cases at 17, 18 and 20 weeks after their first report was estimated to be 94% (95% CI: 0·84, 1), 88% (95% CI: 0·73, 1) and 50% (95% CI: 0·26, 0·98), respectively (Figure S4B). In a Cox model, we estimated that a log-unit increase in the population size of a district reduced the probability that case reports cease by 25% (Hazard Ratio = 0·74; 95% CI: 0·13, 0·92). The duration for which a district reported cases was not significantly associated with the date when it was first invaded.

We then fitted the full model to the data available from week 50 in 2015 (December) to week 35 in 2016 (August) (with all data available). The following terms were retained in the model: (i) the adjacency metric (more than two degrees away), (ii) the binary variable defining whether the invasion happened before or after vaccination rollout (week nine of 2016), (iii) the *Ae. aegypti* suitability, (iv) gravity and (v) radiation human mobility metrics. Assessing model accuracy, we

found excellent agreement between the probability of invasion per week per district, as predicted by the model, and the observed proportion invaded (Figure 3a). The model successfully discriminated those district-weeks with a high probability of invasion from other district-weeks. For example, 31 (0·31%) district-weeks had a predicted weekly probability of invasion greater than 25%, corresponding to an average per month probability of invasion of 77% (Figure 3a). In contrast, 8,754 (88·5%) district-weeks had a weekly invasion probability smaller than 1% (Figure 3a, Supplementary information Figure S6a-d). If we consider the expanding phase of the outbreak from late January to late March, then there was also a high correlation between the invasion of YFV and the invasion probability predicted by the model (0·78; 95% CI: 0·74, 0·82).

We found that spatial expansion slowed down significantly during the course of the epidemic as is apparent in Figure 1b. Assuming that the change occurred on week 9 of 2016 when vaccination was rolled out, the odds ratio for a district being invaded after the change point T vs before T was 0·948 (95% CI: 0·946, 0·949) (Table S3-4). However, the deviance explained was slightly higher under the assumption that the change occurred 2-3 weeks prior to vaccination roll out (33.6% vs 30.8%) (Table S4).

In the context of limited resources, insights concerning district-specific real-time invasion risks might help to inform the prioritisation of districts to be targeted for interventions such as vaccines and vector control. Consider, for example the situation in mid-February 2016 when the epidemic was still expanding (32 districts were invaded in the following month). If the 20, 30, and 50 districts with the highest probabilities of invasion had been targeted in mid-February (on the basis of data available until that week) then the number of targeted districts among the 32 districts that were in fact invaded in the following month would have been 13 (41%), 17 (54%), and 27 (84%) respectively (Figure 3b, Supplementary information Figure S7). The expected values of these statistics would drop to two (6%), three (9%), and six (19%), respectively, if prioritisation had been random.

Univariate models tended to perform worse than the full model when comparing the models predicted probability of invasion vs. the observed proportion of districts invaded (Figure S6 vs Figure 3a). However, it is encouraging to observe that simpler models that included only travel distance or neighbourhood effects were able to discriminate similarly well in terms of their ranking of locations that may become affected (Supplementary information, Figure S9).

We then used the full model to predict future spread of YFV in Angola and the DRC. In Figure 4, districts and communes currently infected are shown in dark red, and all other locations are shown with a shading that indicates the probability of invasion one or two months ahead. The most recent confirmed cases are from Kinshasa, DRC and future spread is predicted to be mostly along the road east to Kananga and south to the border with Angola where cases have been observed previously (Figure 4, Supplementary information Figure S2 and S3). The two locations in Angola with high predicted risk of introduction are Uige and Luanda (Figure 4). Rural regions in Angola and the DRC have low predicted probabilities of YFV introduction.

We explored the robustness of the variables retained in the final model to the time when the analysis was performed. The gravity metric was the variable selected most often (21 out of 24 weeks) compared to 17 times for the radiation metric and 19 times for the *Aedes* suitability surface (Supplementary information Table S2). However, there was heterogeneity as to when these variables were selected. For example, the gravity metric appeared to be important in the early phase of the outbreak and radiation metric in the transitioning phase. In a sensitivity analysis, we found that model parameters would remain almost identical and predictive accuracy very high if the delay from the first infection to the first reported case in a district had shortened from four to one week during the course of the epidemic (Supplementary information, Table S3, Figure S10).

Discussion

In the context of rapidly spreading and potentially fatal infectious disease epidemics, such as the current YFV outbreak in Central Africa for which vaccine stockpiles are limited, it is essential to determine which areas are at greatest risk of infection, in order to inform vaccine prioritization decisions. This requires an in-depth understanding of the determinants of disease spread. In this paper, we integrated and analysed diverse datasets, including the size and mobility of human populations, and detailed maps of vector suitability, in order to describe the epidemic and spatial spread of YFV. We found that the spatial spread of YFV was well explained by human mobility and vector suitability and that our approach could help discriminate districts with high risk of invasion from others.

We estimated that the reproduction number of YKV was 4.8, in line with other estimates available for YFV.[25] This suggests a critical vaccination coverage for YFV of the order of 80%. However, we cannot rule out the possibility that in the early stage of the epidemic, part of the rise in case counts was due to an increase in reporting, which could bias estimates of the reproduction number (and the critical vaccination coverage) upwards. Although we estimated the reproduction number from national data, it will be interesting to assess transmission dynamics at a refined local level.

We characterized a variety of factors that affect the spread of YFV in Central Africa. For example, we found that human mobility was an important predictor of the spatial expansion of YFV. The importance of large population centres in driving the expansion dynamics during the early stage of the epidemic was apparent in both the univariate and multivariate models, and the gravity metric played a key role early on (Figure 2a, Supplementary information). Our results from the univariate analysis also indicate that there was a considerable neighbourhood effect during this outbreak; *i.e.* places closer to the outbreak were more likely to be infected than those further away which was exemplified by the adjacency and travel distance metric performing better than other univariate models (Figure S6). The univariate models however perform poorly compared to the full model when comparing the predicted probability of invasion vs. the observed district invasion. This confirms reports that many of the cases in Kinshasa, DRC, are imported from Luanda, Angola. The inclusion of the spatial distribution of *Ae. aegypti* suitability substantially increased the fit of the full model, but this variable did not show significant associations with early YFV invasion (Supplementary information, Figure S6). Vector suitability is relatively constant across the region which may result in a lower predictive power of that variable (Supplementary information, Figure

S8).[14] More studies are needed to measure the mosquito per human ratio and quantify their spatial heterogeneity (Supplementary information, Figure S8). Our analysis also suggested that epidemics may last longer in areas that are more densely populated (Figure 2b, Supplementary information, Figure S2). This confirms estimates of other arboviruses such as dengue virus, where urban areas have been shown to have considerably higher values of the reproduction number due to a higher person to person contact rate.[16]

We found that the rate of spatial expansion of YFV declined in February 2016, which roughly coincides with the start of the vaccination campaign. Although the observation of such temporal association is interesting, our statistical framework does not allow to demonstrate a causal link or to quantify the proportion of the decline that was due to vaccination versus other causes. There is weak evidence in the data that spatial expansion may have started to decline even before vaccination started (Table S4). However, the current absence of data describing the vaccination campaign makes it difficult to perform a more detailed and definitive assessment of its impact on spread. Future research should aim to further characterize this impact from more detailed data documenting both case counts and vaccine distribution over space and time. It should assess its impact on both spatial expansion (i.e. ability of YFV to spread from district to district) but also on local transmission (i.e. ability of YFV to generate outbreaks within districts).

In addition to understanding the drivers of spread of YFV transmission, we tested our statistical model's predictive power in real-time. We found that during the expanding phase of the outbreak, the model could help to predict the invasion of YFV in the region and to determine which districts should be targeted for intervention (Figure 3b). This is a useful result in a context in which resources and vaccine stock are limited. We highlight the need to re-parameterise the model as the importance of different ecological and mobility factors may vary depending on the week of the epidemic. Fortunately, the rate of transmission of the outbreak in Angola and the DRC seems to be slowing. Nevertheless, our approach presents a flexible framework that can be readily updated and could therefore be used to predict the geographic spread of future epidemics of YFV or other related infectious diseases. However, to translate the insight generated by such a modelling approach into a concrete vaccination strategy, one would need to also account for complex logistic constraints concerning for example vaccine supply and delivery. These important issues are beyond the scope of this paper. Our estimates of characteristics of spatial spread may be affected if the reporting of cases varies spatially.

In recent years, mosquito-borne diseases such as those caused by dengue virus, chikungunya virus, and more recently Zika virus, have been geographically expanding. By contrast, until recently, YFV was thought to be on the decline thanks to the positive impact of vaccination campaigns. However, the ongoing transmission of YFV in Central Africa has demonstrated the continued threat that YFV poses to human populations, and the need for continued vaccination campaigns and close monitoring of sporadic outbreaks in rural regions. Previous work also showed that vaccination rates in the region analysed here are low and that there was no significant spatial heterogeneity between districts.[6] There is a wider context of increased urbanisation and human mobility in the region that may lead to future increases in the risk of emergence and rapid expansion of outbreaks such as observed here.

At this stage, due to the lack of available data, our analysis cannot address the YFV outbreak from a genetic perspective. However, our results provide a baseline to which future viral genetic results may be directly added. Genomic surveillance could be used to extend our results by, (i) helping to identify the origin of the outbreak[26], (ii) monitoring the diversity and spatial distribution of circulating viruses[27], and (iii) characterising signatures of host adaptation[28]. Two past outbreaks of YFV have been documented in Angola in 1971 and 1988, and a single genetic isolate indicates that the 1971 outbreak strain belongs to the East/Central African genotype. Although genetic data for the current outbreak in Central Africa have yet to be published, preliminary analysis of available YFV genomes from returning travellers suggest that the current outbreak also belongs to the East/Central African lineage.[29] Whether the virus has been circulating in a sylvatic cycle in rural areas, or whether there has been cryptic circulation in humans in the region for at least 28 years are questions that require further investigation.

Due to the lack of weekly case count data from outside the capitals of Angola and the DRC, we were restricted to modelling the potential for location specific YFV introduction and circulation, rather than predicting cumulative numbers of expected cases and the quantitative effect of intervention strategies and climatic variables such as rainfall and precipitation. We were also further restricted to using suspected and confirmed cases, as the fraction of cases that have been laboratory tested remained small.[30] Although our model performs well in predicting the expansion of YFV in the region estimates may be improved by obtaining country specific real-time mobility estimates from mobile phone data or other sources such as travel surveys. Such data could inform the key traveller groups that contribute most to the invasion process and may help target intervention strategies.

However, recent work has shown that similar movement patterns are observed across low-income countries.[31,32]

Beyond local spread, there is a risk of international spread of the virus via air travel, as seen by the recent YFV cases that have been detected in travellers returning from Angola to China.[29] In this study we refrained from extrapolating our results to regions other than the two that are currently at the core of the affected countries as evidence of overland cross-border transmissions has, thus far, only been reported between Angola and the DRC; frequent exchange along the major road from Uige, Angola to Kinshasa, DRC is documented (Figure 4). In addition, an Angolan district (Cabinda) is located entirely within the DRC, which results in substantial population flows through the western coastal corridor of the DRC. Spread to other neighbouring countries such as Namibia, Zambia or the Central African Republic may be possible but has not been documented as part of this outbreak.

The identification of locations expected to receive virus introductions has broad implications for disease control, specifically for prioritising where and when treatment and prevention measures would be best implemented in order to prevent subsequent rapid geographic spread of a pathogen.[33] The importance of these techniques using historical data has been shown for other diseases. However, the modelling techniques presented here allow the analysis of near real-time data to inform the control of an ongoing outbreak, a methodological advance applicable to other diseases.

Bibliography

1 Monath TP. Reviews Yellow fever: an update. *Lancet* 2001; **1**: 11–20.

2 Barrett AD. Yellow fever in Angola and beyond - the problem of vaccine supply and demand. *N Engl J Med* 2016. DOI:10.1056/NEJMp1606997.

3 Ellis BR, Barrett ADT. The enigma of yellow fever in East Africa. *Rev Med Virol* 2008; **18**: 331–46.

4 Barrett ADT, Higgs S. Yellow Fever: A Disease that Has Yet to be Conquered. *Annu Rev Entomol* 2007; **52**: 209–29.

5 Monath TP, Vasconcelos PFC. Yellow fever. *J Clin Virol* 2015; **64**: 160–73.

6 Garske T, Van Kerkhove MD, Yactayo S, *et al.* Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med* 2014; **11**: e1001638.

7 World Health Organization. Yellow fever in Angola. http://apps.who.int/iris/bitstream/10665/249603/1/yellowfeversitrep2Sept16-eng.pdf. 2016.

8 Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013; **342**: 1337–42.

9 Valdez LD, Aragão Rêgo HH, Stanley HE, Braunstein LA. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. *Sci Rep* 2015; **5**: 12172.

10 Lessler J, Rodriguez I, Derek B, *et al.* Estimating potential incidence of MERS -CoV associated with Hajj pilgrims to Saudi Arabia, 2014. *PLOS Curr Outbreaks* 2014; **November 2**. DOI:10.1371/currents.outbreaks.c5c9c9abd636164a9b6fd4dbda974369.

11 Faria NR, Azevedo S, Kraemer MUG, *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* 2016; **352**: 345–9.

12 Perkins TA, Scott TW, Le Menach A, Smith DL. Heterogeneity, mixing, and the spatial scales of mosquito-borne Ppathogen transmission. *PLoS Comput Biol* 2013; **9**: e1003327.

13 Nunes MRT, Faria NR, de Vasconcelos JM, *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med* 2015; **13**: 102.

14 Kraemer MUG, Sinka ME, Duda KA, *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* 2015; **4**: e08347.

15 Kraemer MUG, Sinka ME, Duda KA, *et al.* The global compendium of Aedes aegypti and Ae . albopictus occurrence. *Sci Data* 2015; **2**: 150035.

16 Kraemer MUG, Perkins TA, Cummings DAT, *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J R Soc Interface* 2015; **12**: 20150468.

17 Nsoesie EO, Kraemer MU, Golding N, *et al.* Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. *Eurosurveillance* 2015; **21**: pii=30234.

18 Simini F, González MC, Maritan A, Barabási A-L. A universal model for mobility and migration patterns. *Nature* 2012; **484**: 96–100.

19 WorldPop project. WorldPop. http://worldpop.org.uk/.

20 Wesolowski A, O'Meara WP, Eagle N, Tatem AJ, Buckee CO. Evaluating Spatial Interaction Models for Regional Mobility in Sub-Saharan Africa. *PLOS Comput Biol* 2015; **11**: e1004267.

21 Ruktanonchai NW, Deleenheer P, Tatem AJ, *et al.* Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. *PLOS Comput Biol* 2016; **12**: e1004846.

22 Golding N, Schofield A, Kraemer MUG. Movement: Functions for the analysis of movement data in disease modelling and mapping. *R Packag version 02* 2015.

23 Simini F, Maritan A, Néda Z. Human mobility in a continuum approach. *PLoS One* 2013; **8**: e60069.

24 Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B* 2007; **274**: 599–604.

25 Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Gallagher N, Marano N, Staples JE. Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *Am J Trop Med Hyg* 2012; **86**: 349–58.

26 Gire SK, Goba A, Andersen KG, *et al.* Genomic surveillance elucidates Ebola virus origin and

transmission during the 2014 outbreak. *Science* 2014; **345**: 1369–72.

27 Carroll MW, Matthews DA, Hiscox JA, *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* 2015; **524**: 97–101.

28 Park DJ, Dudas G, Wohl S, *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* 2015; **161**: 1516–26.

29 Chen Z, Liu L, Lv Y, *et al.* A fatal yellow fever virus infection in China: description and lessons. *Emerg Microbes Infect* 2016; **5**: e69.

30 Johansson MA, Vasconcelos PFC, Staples JE. The whole iceberg: Estimating the incidence of yellow fever virus infection from the number of severe cases. *Trans R Soc Trop Med Hyg* 2014; **108**: 482–7.

31 González MC, Hidalgo C a, Barabási A-L. Understanding individual human mobility patterns. *Nature* 2008; **453**: 779–82.

32 Bengtsson L, Gaudart J, Lu X, *et al.* Using mobile phone data to predict the spatial spread of cholera. *Sci Rep* 2015; **5**: 8923.

33 Kraemer MUG, Hay SI, Pigott DM, Smith DL, Wint GRW, Golding N. Progress and challenges in infectious disease cartography. *Trends Parasitol* 2016; **32**: 19–29.

Figures

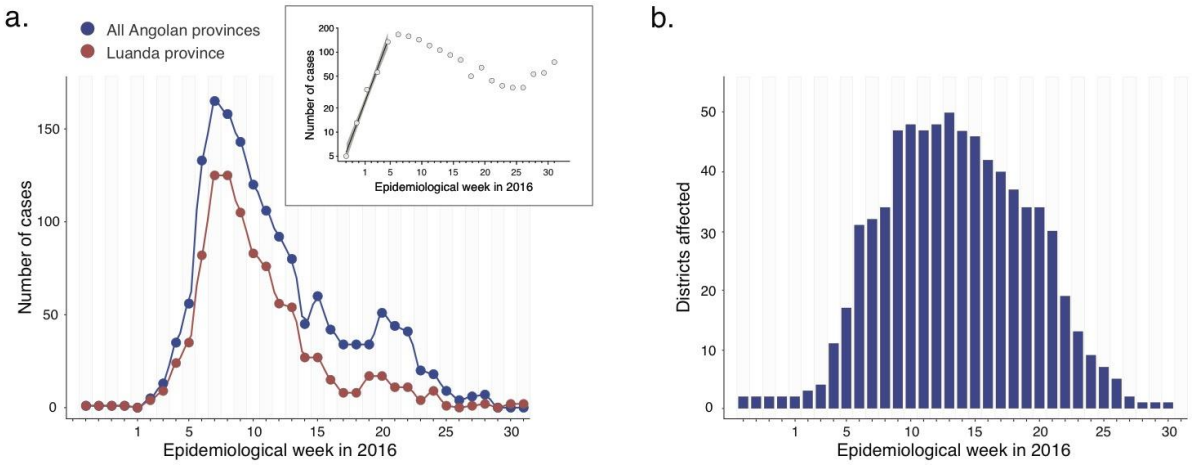


Figure 1: Number of cases in Angola and geographic spread of the epidemic.

Epidemic curve for suspected and confirmed cases in Angola (blue) and Luanda (red) from December 2015 and August 25th 2016. Inset: fit of exponential growth during the early phase of the epidemic. Panel b shows the number of districts affected during each week over the course of the outbreak.

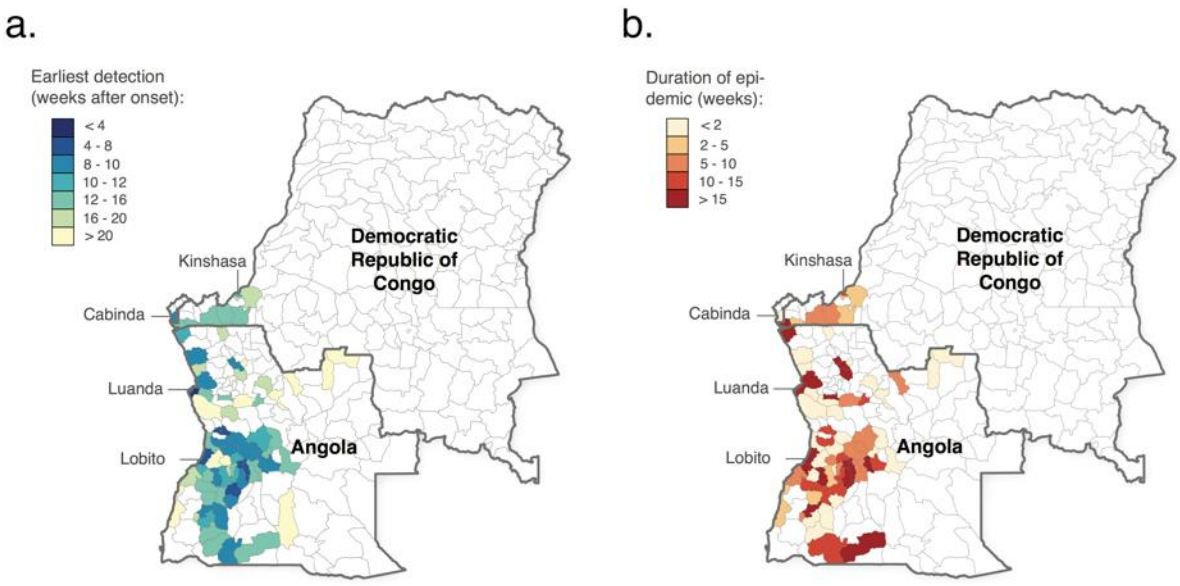


Figure 2: Maps showing the timing of introduction of Yellow fever virus in districts in Central Africa and the duration of infection.

The timing of introduction per district from the origin of the outbreak in Luanda, Angola (a). Dark blue to yellow showing the number of weeks until the first case was reported. Duration of transmission in weeks from yellow to red (b).

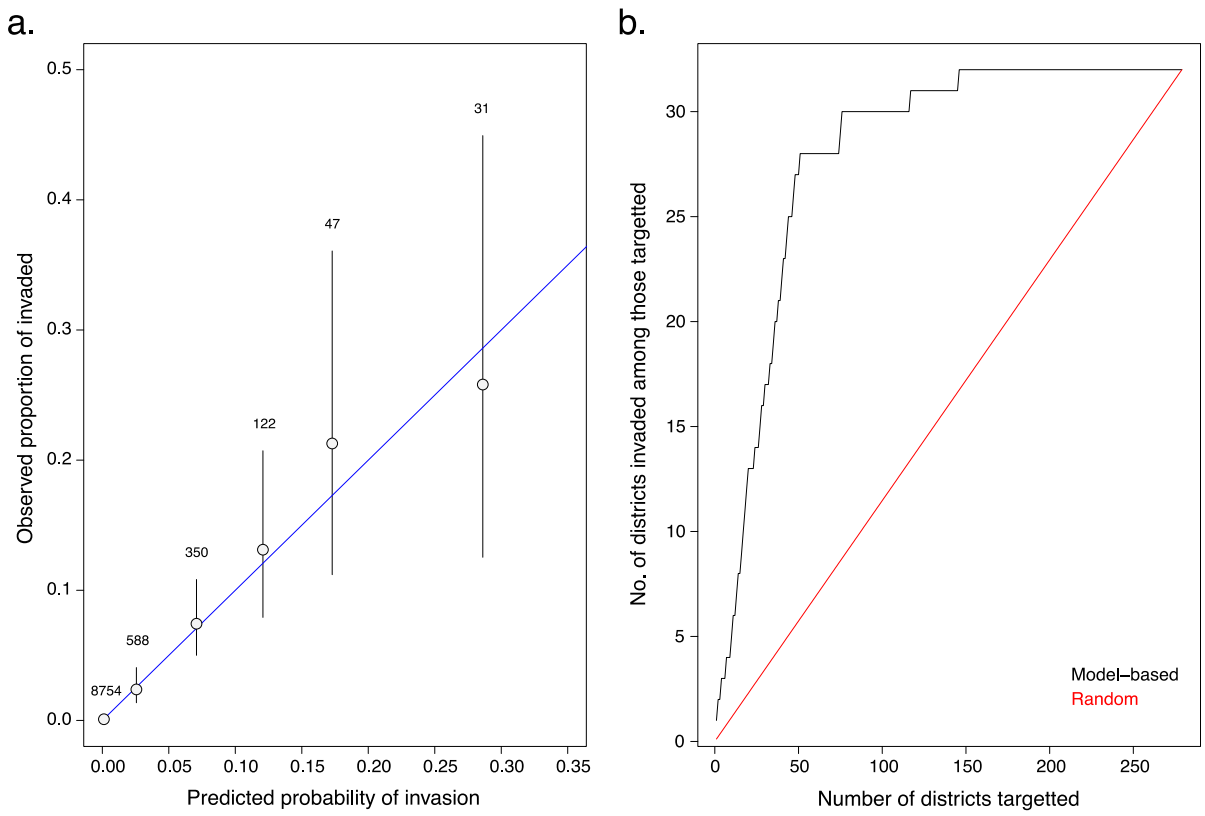


Figure 3: Model accuracy and real-time prediction of the Yellow fever virus invasion model

Model prediction accuracy comparing the predicted invasion probability from the geographic spread model versus. the observed proportion of districts that became invaded; numbers represent district weeks (a). Comparisons between district targeting based on real-time modelling analysis versus random targeting during the expansion phase of the outbreak between mid-March and mid-April where 32 districts were newly invaded (b).

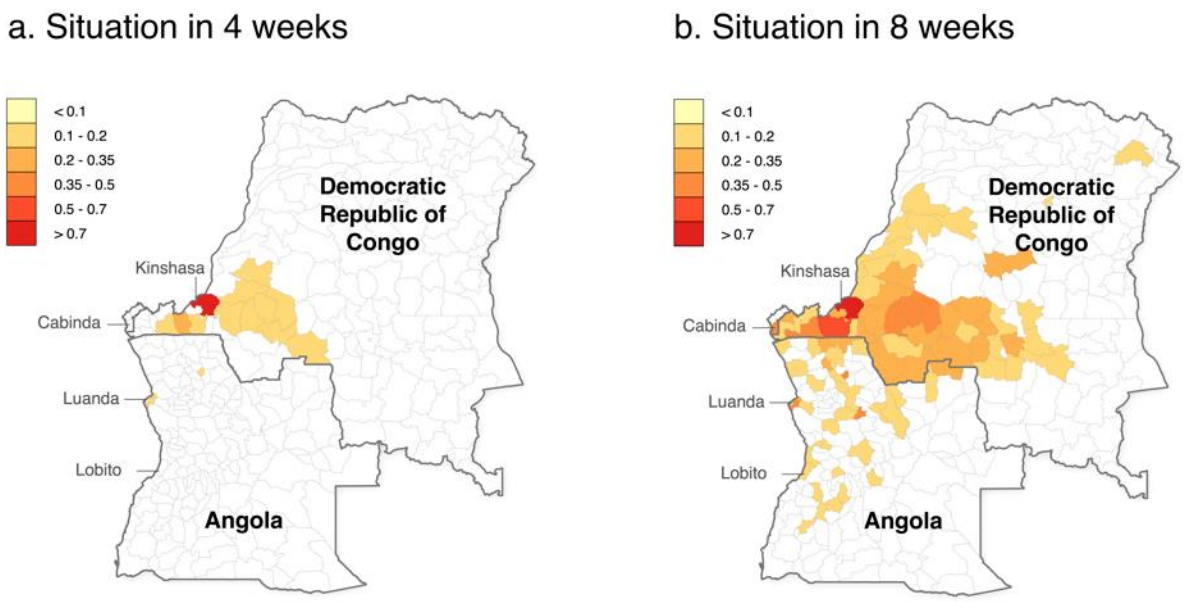


Figure 4: Model based predictions of Yellow fever virus spread in Central Africa.

Maps show model-based predictions for the invasion in Central Africa originating from Kinshasa (dark red), the location with the last reported cases one and two months ahead (a, b respectively).

Discussion

The aim of this thesis was to understand the epidemiology and spread of emerging infectious diseases. This includes predicting their geographic distribution and ecological landscape, their geographic spread and seasonal dynamics. I examined the distribution of the arbovirus vectors *Ae. aegypti* and *Ae. albopictus*, chikungunya virus (CHIKV), analysed the seasonal dynamics of Zika virus (ZIKV), the outbreak dynamics of dengue (DENV), and the spread of Ebola virus disease (EVD) and Yellow fever virus (YFV).

In summary, I predicted the global spatial distribution of *Ae. aegypti* and *Ae. albopictus* to further stratify the global landscape of arbovirus risk at high spatial resolution. Building upon this work improved maps and first estimates of the populations living in areas of CHIKV transmission were build. To further investigate the seasonal variation in transmission potential, maps of ZIKV during an ongoing outbreak were developed using an integrative time-dependent analysis that include estimates of the risk of dispersal by air passenger traffic foreshadowing the pathogens spread. Moving beyond large-scale patterns of spread, the local spatio-temporal dynamics of a large dengue outbreak was investigated using a classical epidemiological method characterizing the importance of urban-rural variation on dengue epidemiology. Building upon this modelling approach the dynamics of another emerging disease was investigated, namely EVD. Using a time-series SIR model the contribution of each district to transmission elsewhere was analysed explicitly accounting for human mobility. Human mobility was shown to help understand the patterns of spread in the early phase of the epidemic in West Africa retrospectively. Combining methodologies and insights from previous chapters, mathematical modelling techniques were applied to an urban YFV outbreak in Central Africa, using real-time data to understand the risk of invasion of the pathogen into surrounding areas and to guide vaccination roll-out.

In this discussion I will summarize key findings from each chapter of this thesis, highlight the strengths and limitations of the methodologies and data used, and contextualise the work in the broader literature of infectious disease mapping and epidemiological modelling.

Chapter summary

Mapping the global distribution of disease vectors has previously been an important driver to improve efforts to understand the spatial epidemiology of associated diseases such as malaria [1–3]. In my first chapter I present the most comprehensive dataset of occurrence records of *Ae. aegypti* and *Ae. albopictus* to date and use an advanced modelling approach with a bespoke set of environmental and land-cover correlates, to map the spatial distribution of both species at high resolution. The maps show that the distribution is the widest ever recorded and may be used to prioritize where surveillance of these vectors should be targeted. More specifically, where there is a large mismatch between predicted occurrence probability and reported presences of the vectors such as in large parts of Africa more research is needed to fill these gaps. As both species are currently expanding at a continental level it will be important to pair the presented maps with human and trade data to infer the likely future expansion of both species as they continue their march [4–6]. In addition, genetic information may be used to characterize the spread of the vectors as it allows to infer the relatedness of local subpopulations of both species [7,8].

After defining the spatial limits of the main vectors, I used these outputs to improve upon estimates of environmental risk of CHIKV globally. In chapter 2, I present the global distribution identified based on evidence consensus and identify areas suitable for local transmission at high spatial resolution. Since the invasion of CHIKV in the new world it was pivotal to identify areas that may experience local transmission and classify current evidence to highlight areas where systematic sampling could have large effects to fill in gaps. 1.3 billion people living in areas where local transmission may occur were identified. In addition, similarly to dengue, in many West African countries, there is lacking evidence of CHIKV but high environmental suitability. It was

encouraging to see that *Aedes* suitability maps improved our maps substantially and that there was a clear association between the presence of *Ae. aegypti* and CHIKV transmission (relative contribution >45%). The relative role of *Ae. albopictus* could not be investigated independently due to the lack of data but further work would allow more detailed analyses. Especially in areas such as Europe where there was a small CHIKV outbreak caused by *Ae. albopictus.*

The recent emergence of ZIKV, its introduction into the Americas, and rapid spread have introduced a novel challenge to determining the dynamic risk of arbovirus spread as it occurs at unprecedented speed [9]. Using real-time passenger flight data in combination with estimates of seasonal variation in environmental risk of ZIKV monthly estimates of the combined risk of local transmission and the importation from viremic travellers were highlighted. In India alone for example 1.2 billion people live in areas susceptible to ZIKV exposure at the peak time of seasonal risk. In Africa, Angola receives most of the travellers departing from countries that currently report transmission of ZIKV and has previously been shown to be a major hub of arbovirus exchange [10]. For example, CHIKV was previously introduced to Brazil from Angola and our analysis reveals that human mobility, and specifically passenger flight data may help us identify large scale patterns of continental arbovirus exchange [9,11]. This study needs to be seen with a number of limitations that are mostly related to the lack of information available at the time of analysis. First, the true number of ZIKV infections remains unclear as it cross-reacts with other flaviviruses such as dengue [12,13]. In addition, as the number of asymptotic infections is believed to be very high (80%), countries with ongoing transmission may not be reporting them as such, so the risk of importation may be underestimated [14,15]. Other *Aedes* species may be competent vectors of ZIKV but their relative contribution still needs to be explored [16].

Mapping out key locations where a disease may appear next is pivotal in addressing questions about prioritising surveillance. Understanding the dynamics of local spread and the main determinants that govern local explosion of cases however is important for decision makers to

implement response strategies [17]. In Chapter 4, I model the dynamics of an urban dengue outbreak in Pakistan by using a set of time-varying environmental covariates and socio-economic drivers of dengue incidence. Our results show considerable spatial heterogeneity in the mixing parameter (contact rate). This heterogeneity is strongly associated with the urban-to-rural gradient and can be directly explained by a population weighted urban-accessibility metric. By accounting for such differences in population mixing large implications in regards to prevention and control were revealed. These results constitute the first empirical evidence and were previously only shown in a theoretical context. In this analysis, each district is viewed as an independent spatial entity and movements between them were not accounted for which should be explored in future work [18].

To explore the relationship between the movements of people during a novel emerging infectious disease outbreak, in this case EVD, using a mechanistic-stochastic modelling framework. The pathogen spread rapidly throughout West Africa after its initial introduction from the animal reservoir to the human population [19]. In this chapter, I clearly identified the utility of including human movement metrics in time-series epidemiological models to reconstruct the pathway of EVD. In addition, I showed that there was temporally varying heterogeneity in sources and sinks of virus spread in the first half of the outbreak (expanding phase) and second half of the outbreak. Initially the districts around the origin of the outbreak contributed most to transmission (Gueckedou, Guinea) and later the highly populated districts in Sierra Leone and Guinea were the centres of transmission. I also confirmed that even in the absence of real time movement data from the region our model performed well which is encouraging for future modelling in outbreak situations in real-time. However, the model could have been improved if real-time movement data from the region were available as it could have captured adaptive human behavioural processes over the course of the outbreak.

Moving from reconstruction to near real-time modelling chapter 6 discussed key epidemiological parameters during the first urban YFV outbreak since the 1980s in Nigeria. In addition, I modelled the spread of the pathogen during its expansion in Central Africa [20]. This chapter uses a number of different modelling approaches that are aided in understanding the growth of the epidemic in Luanda, Angola, reconstructing the epidemics pathway using univariable but also more complex statistical techniques and predicts the likely future of the virus. We find that demographic variables are important to explaining the timing of invasion of YFV and duration of local transmission. *Aedes aegypti* suitability was not important in examining this trend but helped discriminate which locations may become invaded in real-time. Human mobility metrics derived from mobile phone data helped predict the invasion of YFV and with this integrated framework we hope to be able to help policy makers to deploy limited resources, especially in a situation with limited vaccine stocks. This study was limited by our ability to predict the number of expected cases of YFV in each locality as time-series incidence data was only available for a limited amount of locations. I also did not consider the international spread outside of Central Africa to areas with low vaccination coverage which could be explored the future.

In this thesis, I have used a number of statistical modelling techniques to infer a large variety of epidemiologically relevant quantities, from mapping a continuous landscape of risk of vector and arbovirus suitability, to the risk of invasion of a pathogen from known observed locations. A summary of the techniques deployed here are provided in Table 3.

Table 3: Summary of modelling techniques applied in this thesis, their input data sources, quantities, and uncertainty quantification.

	Inputs	**Data quantity**	**Analysis/modelling framework**	**Main outputs**	**Uncertainty**
Distribution of *Ae. aegypti* and *Ae. albopictus*	Geographically stratified occurrence records	19,930 and 22,137 respectively	Boosted Regression Tree (BRT) modelling framework	Predicted global probability of occurrence	Model 95% CI

Distribution of Chikungunya virus	1. Human disease occurrence data 2. Evidence consensus	528 occurrence records and variety of sources to define presence/absence	BRT modelling framework	1. Predicted global environmental suitability for chikungunya virus 2. Country wide evidence consensus stratifying presence/absence	Model 95% CI
Zika virus distribution	Human disease occurrence data	323 occurrence records	BRT modelling framework	Predicted global environmental suitability of Zika virus	Model 95% CI
Global spread of Zika virus	1. Mosquito maps from Chapter 1 2. IATA flight data	1. Seasonal variation in climatic variables 2. Flight records from 2014-2015	Integrative modelling framework combining BRT outputs overlaid with flight data	Number of passengers arriving in locations receptive to local transmission	Scenario based uncertainty (conservative – least conservative)
Epidemiological analysis of a dengue outbreak	Weekly hospital human dengue case data	44,156 observed cases over four years	Time-series SIR model fitted using a generalized additive regression model (GAM)	Geographic distribution of R_0 and effect of climatological variables over the range of values	95% CI
Predicting the spread of Ebola virus disease	Weekly incidence data per district between	2013-2015 incidence data per location	Time-series SIR model fitted using a GAM	Relative contribution of each district to transmission elsewhere	95% CI
Yellow fever spread in Central Africa	Weekly incidence data for Luanda (capital Angola) and location specific estimates of time of infection	Epidemiological data for each location and each week	SIR model, geographic spread model (GAM), survival analysis	Combined outputs of exponential growth phase, location specific risk of invasion, location specific parameters governing the spread and duration of the outbreak	95% CI
CI = confidence interval; BRT = Boosted Regression Tree model; TSIR = Time-series Susceptible Infected Recovered model; GAM = Generalized additive model					

Implications and future directions

Moving on from this thesis there is a number of improvements and further directions that will be discussed here. By integration of human mobility into disease mapping and thus introducing dynamic elements into previously static estimates, I provide novel tools that can be used for any

expanding infectious disease at multiple geographic scales. I exemplify how they can be used to assess global risk of spread as well as local dynamics. However there is a number of interesting future directions that may be explored in addition to the presented work.

Especially the increasing volumes of data from a variety of sources, and the technology to rapidly collate and organise them, offer enormous opportunities for improving disease surveillance and control in the near future [21]. For example, a number of digital resources for disease detection data have been established to collate data from online resources [21], including ProMED-mail [22], the Global Public Health Intelligence Network (GPHIN), HealthMap and BioCaster [21,23]. These online resources can fill gaps in public health reporting infrastructure and complement existing traditional surveillance systems of infectious diseases, particularly in areas where clinical surveillance capacity is underdeveloped and in disease outbreak situations where a rapid response is required [24,25].

In contrast with these advances in data assembly, predictive risk mapping has so far focused on the release of static maps of disease distribution over long time periods. This situation reflects the computational and time demands of creating and refining the predictive spatial models used for disease risk mapping. With advances in available computing power and improving computational methods the rate at which risk maps can be updated has increased to the point where dynamic, near-real-time risk maps are now a feasible prospect. Similarly to live data resources, systems which are able to rapidly and automatically acquire and verify disease occurrence data, generate predictive maps and disseminate these to policy makers offer a major advance on the current state of affairs [26,27]. This could be applied to chapter 1-3 to generate timely estimates of disease risk in near real-time. Rapidly translating contemporary epidemiological data into useful knowledge may facilitate control during disease outbreak situations, as well as to inform the general public [27,28].

Understanding population interconnectedness helps identify key individuals, risk groups or regions in a network that would benefit most from targeted interventions, *e.g.* through vaccination [29,30]. Pinpointing key centers of infection can then have a multiplying effect, reducing case numbers, preventing catastrophic outbreaks and reducing disease control implementation costs. Failure to account for human movements in previous public health campaigns has slowed eradication efforts for malaria and will continue to do so if no comprehensive mapping approach unifying economic, geographic and medical approaches is developed [31]. Particularly in countries with increasing economic activity (*i.e.* Southeast Asia) these new mapping strategies deployed here can inform their public health responses. This principle also extends to better control and monitoring of livestock production and the reduction of contact between domestic animals, humans and wildlife to reduce zoonotic disease emergence and spread [28,32,33].

Genomic surveillance, including real-time portable genome sequencing [34], may be used to extend our modelling approach by helping to identify the origins of the outbreak [19], as well as to monitor the diversity of the circulating viruses [35,36], to characterize signatures of host adaptation [37], and to pinpoint the source and sink locations of strains [9,36–39]. Such analyses have been successfully deployed in reconstructing the outbreak, however, when applied in conjunction with the methods discussed in chapter 4-6 they could inform the predictive pathway of epidemics with more accuracy [9].

Moving forward, individual diseases cannot be studied in isolation, but rather they should be assessed in the context of possible co-infections (*e.g.* tuberculosis and diabetes mellitus [40] or malaria and helminthiasis [41,42]). Quantifying the spatial patterns of the not only one disease but the composition of diseases that cause a significant burden to the population can ultimately help scale up intervention methods and reveal potential co-benefits to reduce the global burden of infectious diseases. This will become a critical issue for emerging economies such as China, Indonesia, Brazil and Mexico, where both infectious and chronic diseases are now occurring in

tandem [43,44]. For this purpose spatially explicit information about urban poverty, health care access and other socio-economic determinants of health are yet to be developed on a global scale to help identify populations at risk of a comprehensive set of diseases as well as their inter-linkages [45].

References

1. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit. Vectors* **5,** 69 (2012).
2. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **3,** 72 (2010).
3. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **4,** 89 (2011).
4. Tatem, A. J., Rogers, D. J. & Hay, S. I. Global transport networks and infectious disease spread. *Adv. Parasitol.* **62,** 293–343 (2006).
5. Hawley, W. A., Reiter, P., Copeland, R. S., Pumpuni, C. B. & Craig, G. B. Aedes albopictus in North America: probable introduction in used tires from northern Asia. *Science* **236,** 1114–6 (1987).
6. Kraemer, M. U. G. *et al.* Progress and challenges in infectious disease cartography. *Trends Parasitol.* **32,** 19–29 (2016).
7. Brown, J. E. *et al.* Worldwide patterns of genetic differentiation imply multiple 'domestications' of Aedes aegypti, a major vector of human diseases. *Proc. R. Soc. Biol. Sci.* **278,** 2446–54 (2011).
8. Allicock, O. M. *et al.* Phylogeography and population dynamics of dengue viruses in the Americas. *Mol. Biol. Evol.* **29,** 1533–43 (2012).
9. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352,** 345–349 (2016).
10. Nunes, M. R. T. *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med.* **13,** 102 (2015).
11. Bogoch, I. I. *et al.* Anticipating the international spread of Zika virus from Brazil. *Lancet* **387,** 335–336 (2016).
12. Priyamvada, L. *et al.* Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. *Proc. Natl. Acad. Sci. USA* **113,** 2–7 (2016).
13. Dejnirattisai, W. *et al.* Dengue virus sero-cross-reactivity drives antibody- dependent enhancement of infection with zika virus. *Nat. Immunol.* **17,** 1102–8 (2016).
14. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America. *Science* 10.1126/science.aag0219 (2016). doi:10.1126/science.aag0219
15. Lessler, J. *et al.* Assessing the global threat from Zika virus. *Science* 10.1126/science.aaf8160

(2016) (2016). doi:10.1126/science.aaf8160 (2016)

16. Musso, D., Nilles, E. J. & Cao-Lormeau, V. M. Rapid spread of emerging Zika virus in the Pacific area. *Clin. Microbiol. Infect.* **20,** O595–O596 (2014).
17. Heesterbeek, H. *et al.* Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347,** aaa4339 (2015).
18. Wesolowski, A. *et al.* Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci. USA* **112,** 11887–11892 (2015).
19. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345,** 1369–1372 (2014).
20. Garske, T. *et al.* Yellow Fever in Africa: estimating the burden of disease and impact of mass vaccination from outbreak and serological data. *PLoS Med.* **11,** e1001638 (2014).
21. Brownstein, J. S., Freifeld, C. & Madoff, L. Digital disease detection - harnessing the web for public health surveillance. *N. Engl. J. Med.* **360,** 2153–2157 (2009).
22. Madoff, L. C. ProMED-mail: an early warning system for emerging diseases. *Clin. Infect. Dis.* **39,** 227–232 (2004).
23. Collier, N. *et al.* BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* **24,** 2940–1 (2008).
24. Salathé, M. *et al.* Digital epidemiology. *PLoS Comput. Biol.* **8,** e1002616 (2012).
25. Chan, E. H., Sahai, V., Conrad, C. & Brownstein, J. S. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl. Trop. Dis.* **5,** e1206 (2011).
26. Broniatowski, D. A., Paul, M. J. & Dredze, M. Twitter: Big data opportunities. *Science* **345,** 148 (2014).
27. Hay, S. I., George, D. B., Moyes, C. L. & Brownstein, J. S. Big data opportunities for global infectious disease surveillance. *PLoS Med.* **10,** e1001413 (2013).
28. Pigott, D. M. *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* **3,** e04395 (2014).
29. Lazer, D., Kennedy, R., King, G. & Vespignani, A. Big data. The parable of Google Flu: traps in big data analysis. *Science* **343,** 1203–5 (2014).
30. Cauchemez, S. *et al.* Local and regional spread of chikungunya fever in the Americas.

Eurosurveillance **19,** 20854 (2014).

31. Tatem, A. J. & Smith, D. L. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12222–7 (2010).
32. Horby, P. W., Pfeiffer, D. & Oshitani, H. Prospects for emerging infections in east and southeast Asia 10 years after severe acute respiratory syndrome. *Emerg. Infect. Dis.* **19,** 853–60 (2013).
33. Kilpatrick, A. M. & Randolph, S. E. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **380,** 1946–55 (2012).
34. Rambaut, A. Phylogeography of 2014-2015 Ebola virus epidemic. *virological.org* at <http://virological.org/t/phylogeography-of-2014-2015-ebola-virus-epidemic/199>
35. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524,** 97–101 (2015).
36. Ladner, J. T. *et al.* Evolution and Spread of Ebola Virus in Liberia, 2014-2015. *Cell Host Microbe* **18,** 659–669 (2015).
37. Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161,** 1516–1526 (2015).
38. du Plessis, L. & Stadler, T. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol.* **23,** 383–386 (2015).
39. Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance — the time is now. *Genome Biol.* **16,** 155 (2015).
40. Jeon, C. Y. & Murray, M. B. Diabetes mellitus increases the risk of active tuberculosis: A systematic review of 13 observational studies. *PLoS Med.* **5,** 1091–101 (2008).
41. Brooker, S. *et al.* Epidemiology of Plasmodium-helminth co-infection in Africa: Populations at risk, potential impact on anemia, and prospects for combining control. *Am. J. Trop. Med. Hyg.* **77,** 88–98 (2007).
42. Mulu, A. *et al.* Epidemiological and clinical correlates of malaria-helminth co-infections in Southern Ethiopia. *Malar. J.* **12,** 227 (2013).
43. Bygbjerg, I. C. Double burden of noncommunicable and infectious diseases in developing countries. *Science* **337,** 1499–501 (2012).
44. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional and national levels of age-specific mortality and 240 causes of death, 1990-2013: A systematic analysis for the Global

Burden of Disease Study 2013. *Lancet* **385,** 117–71 (2015).

45. Tatem, A. J. Mapping population and pathogen movements. *Int. Health* **6,** 5–11 (2014).

Appendix

A.1. Appendix to Introduction

Trends in Parasitology

CellPress

Review

Progress and Challenges in Infectious Disease Cartography

Moritz U.G. Kraemer,[1,*] Simon I. Hay,[2,3,4] David M. Pigott,[1,2] David L. Smith,[1,4,5] G.R. William Wint,[1,6] and Nick Golding[2]

Quantitatively mapping the spatial distributions of infectious diseases is key to both investigating their epidemiology and identifying populations at risk of infection. Important advances in data quality and methodologies have allowed for better investigation of disease risk and its association with environmental factors. However, incorporating dynamic human behavioural processes in disease mapping remains challenging. For example, connectivity among human populations, a key driver of pathogen dispersal, has increased sharply over the past century, along with the availability of data derived from mobile phones and other dynamic data sources. Future work must be targeted towards the rapid updating and dissemination of appropriately designed disease maps to guide the public health community in reducing the global burden of infectious disease.

Trends

Infectious disease cartography is not only a powerful field of research in disease modelling, but also useful for communicating results to the public.

Our understanding of the interaction of environmental, socioeconomic, and behavioural factors, and disease presence has improved as more data become available.

Human connectivity is a central driver of disease spread globally as humans become more mobile.

Challenges remain in developing automated tools that allow the rapid detection and dissemination of disease maps, particularly in outbreak situations.

Recent Trends in Infectious Disease Cartography

Infectious disease cartography (or infectious disease mapping) describes the quantitative spatial prediction of disease transmission risk. It has been facilitated by the rapid increases in computing power over the past three decades and the increasing availability of spatially continuous data on environmental risk factors for diseases, such as satellite imagery. Therefore, high-resolution spatial data representing both environmental and socioeconomic predictors of disease transmission [1,2] and improvements in statistical modelling [3,4] have increased our ability to characterize these patterns more accurately while also incorporating uncertainty [5].

Maps have become increasingly important in assessing the spatial distribution of disease and are now powerful tools for addressing major public health policies. The visual nature of maps is especially helpful in identifying the spatial heterogeneity of infectious disease risk. Whereas traditional epidemiological modelling focussed on transmission parameters and dynamics and the effect of interventions (the 'what'), mapping such risk spatially largely helps to pinpoint areas 'where' targeted interventions should be focussed. Today, novel approaches are being developed to collate information from a variety of sources, produce risk maps, and disseminate them effectively to scientists, public health professionals, and policy makers in real time [6].

Here, we review existing disease-mapping approaches and show how, in a technologically advancing world, human behavioural changes must become an integral part of disease mapping globally.

Approaches for Mapping Infectious Disease

Historically, disease mapping was confined to simply drawing case numbers on physical maps. Over the past decades, the world has been mapped out digitally and information about disease

[1]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
[2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
[3]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA
[4]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA
[5]Sanaria Institute for Global Health and Tropical Medicine, Rockville, MD 20850, USA
[6]Environmental Research Group Oxford (ERGO), Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK

*Correspondence: moritz.kraemer@zoo.ox.ac.uk (M.U.G. Kraemer).

© 2015 Elsevier Ltd. All rights reserved.

Trends in Parasitology

CellPress

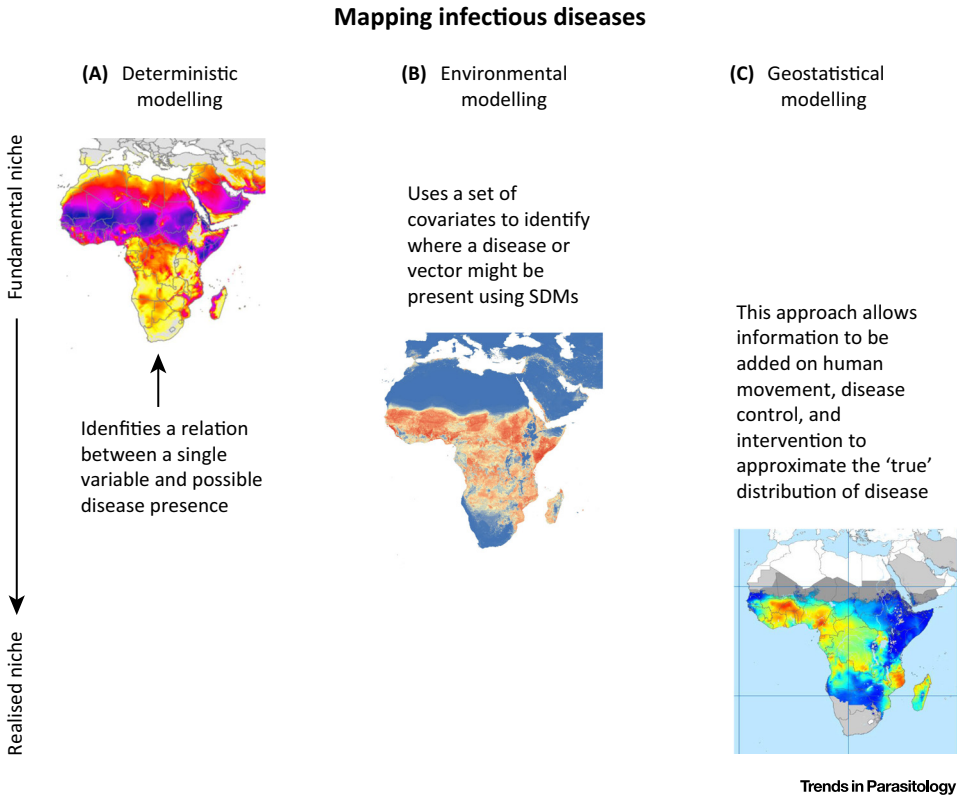


Figure 1. Approaches to Mapping Infectious Diseases: From Deterministic Modelling to Geostatistical Modelling. Deterministic mapping can be seen as the approximate representation of the fundamental niche of a species or pathogen, such as areas where the temperature is suitable for transmission (A) [85]. Environmental modelling techniques (B) incorporate a wider range of covariates that can influence the distribution of disease, such as vegetation coverage and urbanisation shown here for *Aedes aegypti* [33]. In geostatistical approaches (C), the true distribution (realised niche) of disease is approximated using a range of covariates, including disease control and intervention, as shown for malaria [86]. Abbreviation: SDM, species distribution model.

outbreaks is recorded from numerous data sources. Major approaches to disease mapping today involve deterministic, correlative, and geostatistical modelling techniques (see details in Figure 1 and Table 1). For policy makers to translate these maps into action, levels of confidence must be assigned to them, which is often challenging (Box 1).

Rationale for Environmental Disease Mapping

Environmental conditions are a strong predictor of risk for many infectious diseases of humans. The link is particularly strong for some vector-borne diseases, whose distributions are driven by the conditions influencing disease vector abundance and pathogen replication rate in the insect host [7,8]. Therefore, major vector-borne diseases, such as dengue (DENV) [9], chikungunya (CHIKV), and malaria [10], are closely associated with patterns in rainfall [11], temperature [12], and associated determinants, such as deforestation [13] and biodiversity [14]. These relations can be harnessed to produce global maps of the likely distributions of both vector and disease. Approaches range from simplified deterministic approaches to statistical and geospatial models that each focus on different aspects of the interaction of factors influencing disease (Table 1).

Mapping the Socioeconomic Correlates of Disease

Historically, the research focus has been on tying spatial models of disease to climatic factors. However, the spatial distributions of many important deadly diseases are likely driven

Trends in Parasitology

CellPress

Table 1. Environmental Modelling in Infectious Disease Cartography: Main Features and Limitations

Modelling Approach	Key Features	Data Requirements	Examples	Limitations	Refs
Deterministic modelling	Identification of key thermal and hydrological parameters (field and laboratory experiments); key parameters are subsequently extrapolated to regions where there are no data available	Laboratory and/or field estimates of relation between survival of vector and/or pathogen	Diseases with strict relation between a single environmental factor and disease or vector presence; e.g., malaria, dengue, West Nile virus, and Zika virus	Singles out environmental factors; fails to include human behaviour and socioeconomic factors	[85,87,88]
Statistical modelling	Extends deterministic modelling by including multiple environmental predictors to map the spatial distribution of a species; addressing complex mechanisms and interactions between environmental factors is possible	Presence and absence data of disease with detailed spatial information; set of environmental and socioeconomic cofounders	Can capture spatial patterns not only of vector-borne diseases, such as malaria and dengue, but also of Ebola and influenza	Control interventions and disease spread are often not accounted for	[9,48,71]
Geostatistical modelling	Accounts for local variation resulting from disease control and intervention; spatial surfaces can then be approximated with a richer set of statistical confidence	Large set of space-time disease prevalence and incidence data; detailed information about intervention strategies and their effect over time	Many advances in the field have been made in malaria and schistosomiasis modelling, among others	Largely limited by the availability of data; geographical spread of a virus or vector cannot be accounted for	[86,89]

significantly by socioeconomic risk factors, such as population density, economic development, and associated determinants, such as sanitation infrastructure and access to health facilities [15]. The burden of a disease in any given location also crucially depends on the number of humans experiencing such conditions [16] and, therefore, understanding the specific demographics of these populations enables more accurate assessments of disease burden and appropriate interventions [17,18]. All of these factors are highly spatially variable and are rapidly changing [19]. Since 1950, the global population has tripled, the proportion of humans living in urban areas has increased from 29% to 53%, and the number of humans living in extreme poverty has dropped by 50% since the 1990s[i]. These socioeconomic changes will continue over the coming decades and reliably mapping disease risk both today and into the future will depend on the availability of spatial estimates of these factors, particularly in countries that will likely undergo the greatest shifts (such as China, India, and Brazil). The advent and continued development of remote-sensing technology has enabled the high-resolution global mapping of environmental conditions, such as temperature and rainfall. By contrast, accurately quantifying spatial variation in key socioeconomic disease-risk factors is more difficult.

Mapping Settlements and Populations

For disease mapping, it is crucial to understand the distribution of populations to understand the potential of humans to be infected by a particular pathogen. Human population-mapping projects, such as the Global Rural-Urban Mapping Project (GRUMP)[ii] and WorldPop[iii], aim to develop global maps of the locations and sizes of urban areas and other human settlements. Typically, data to inform these maps are obtained from national censuses and surveys at specific locations, although these data have a coarse spatial resolution and, for many of the most resource-poor parts of the world, may be many years out of date. As with environmental

Trends in Parasitology

CellPress

Box 1. Uncertainty and Validation in Disease Mapping

Uncertainty

Uncertainty is a key element of disease mapping and is important for decision makers. While geostatistical approaches (Table 1, main text) often generate fully probabilistic predictions that can be used to quantify uncertainty, their interpretation often remains difficult. However, uncertainty maps are a great resource, identifying areas where future surveys would be most valuable. By contrast, species distribution models (SDM) less frequently consider uncertainty, given their model structure. SDMs are typically generated from a poorer set of input data, but uncertainty quantifications are equally important, especially given that SDMs have weaker predictive power.

There still is a significant gap between how policy makers make sense of uncertainty maps and how they are used. Researchers should aim at communicating with the public health policy community to improve the feedback loop between these two communities.

Validation

In general, it is fair to assume that maps derived from both geostatistical and environmental approaches are only as good as the model and data that are used to produce them. Therefore, it is important to fully understand the strengths and weaknesses of the modelling and mapping approaches in determining their utility. Common methods to evaluate the predictive power of models are cross-validation procedures [90]. Such procedures are useful if there is a high density of disease data that can be split in training and test data sets by randomly assigning points to either one and then interpolate risk in areas between sampling locations. Although this method is a step towards handling validation more rigorously, it often overestimates the ability of the model to extrapolate to new regions [91].

To assess the ability of the model to extrapolate to areas where no data are available, or to different time points in the future, a stratified validation procedure can be used. Gilbert *et al.* [48] used such an approach to assess the ability of an environmental model of avian influenza to predict risk of disease transmission in areas where the disease had yet to be introduced. Researchers should openly address the challenges of validating their maps and improve their maps as new data come in on a continuous basis.

mapping, remote sensing can be used to quantify related variables, such as nighttime lights, peridomestic habitats, and roofs [20,21]. However, disentangling population density and economic prosperity makes robust mapping of these factors difficult.

Several recent technological advances offer an opportunity to increase the accuracy of spatial estimates of population density. The rapidly falling cost of cell phones and increases in global coverage have led to high levels of use in even the most resource-poor settings[iv]. By triangulating individual cell phones, it may be possible in the near future to accurately map settlements, estimate spatial variation in population densities, and to assess how these change through time [22]. An ever-increasing availability of high-resolution satellite imagery in conjunction with tools for online collaboration has led to mapping projects, such as Open StreetMap, developing highly detailed and openly available maps of built-up areas and road networks [23]. This information offers much increased precision and accuracy for developing global maps of population densities [24]. As with the census data used to estimate population densities, coverage of these maps tends to be poorer in resource-limited locations. Building on the success of emergency mapping drives in the wake of natural disasters[v] [25], public health and humanitarian agencies have recently sought to improve mapping in resource-poor areas via the Missing Maps Project, which aims to provide detailed crowd-sourced cartographic information for the most vulnerable and disease-prone parts of the world within the next 2 years[vi].

Infectious Diseases and Global Connectivity

In tandem with the unprecedented increases in human population over the past 50 years, connectivity among human populations has increased rapidly. Increasing urbanization, economic growth, improvements in transport infrastructure, and changing trade and immigration policies [19,26,27] have led to populations in each successive generation becoming more mobile than the last, with regular international travel now the norm for a significant fraction of the global population.

For many zoonotic and vector-borne diseases, and for all directly transmitted infections, introduction of the disease into nonendemic locations is strongly influenced by human

movement and trade [28–31]. Similarly, local dynamic epidemic patterns in endemic regions are often driven by human movements of infected individuals on several spatial and temporal scales [15,32]. Until now, it has been difficult for public health policy makers to make sense of disease spread during a specific outbreak, mainly due to the difficulty in accurately accounting for human movement in real time.

For many vector-borne diseases, only a limited number of species are able to act as capable vectors and, therefore, the disease can only be transmitted where they are present [33]. In these instances, the introduction of the vector into a new region is a prerequisite for disease establishment [34]. The relatively low dispersal ability of most disease vectors means that long-distance vector spread is typically driven by human population connectivity, such as via international shipping [35]. Similarly, international trade in livestock has led to the introduction of infected animals into new areas [36], and human settlement expansion has increased the likelihood of animal to human infections for zoonotic diseases [37].

Diseases that were once restricted to discrete and isolated regions can now pose a risk across the world [38–41]. Accounting for the impact of human mobility on the spatial distribution of disease risk is now a critical challenge in adapting infectious disease cartography to the modern age.

Quantifying Connectivity

Incorporation of connectivity into predictive disease mapping requires quantification of the degree to which populations are connected across spatial scales [42]. There is a large body of work that is trying to understand the general movement patterns of human populations, their specific attributes, and limitations at different spatial and temporal scales (Table 2). One common limiting factor of the diverse data sources that capture human mobility is that they are often confined to political boundaries. With increasing international travel and concern over international spread of disease, quantifying how humans move across borders is of paramount importance. Such international movement patterns may be assessed using long-term migration data [43], but it is unclear how representative these are of shorter-term movements, such as international travel for trade or to visit relatives. Data on short-term border crossings might be available for small geographical areas. Synthesising these two sensibly could provide useful tools for future cross-country disease modelling. However, with no single 'gold-standard' mobility metric, and methodological differences in obtaining data at different scales, this remains a daunting task. In this era of Big Data, individual-based metadata are increasingly available and stored by large private and public institutions. However, the collection and use of individual-based information is associated with serious privacy concerns, which need to be considered carefully [44]. To date, no common guidelines yet exist for dealing with user re-identifiability using geolocated data of simply anonymised data sources [45]. Developing and implementing such guidelines will enable the wider use of these invaluable data, while safeguarding the rights of individuals.

Not only human movements, but also the movements of goods have significantly contributed to the international spread of the DENV and CHIKV vectors *Aedes aegypti* and *Aedes albopictus* [46,47]. Quantifying the international movement of goods transported between ports remains challenging because this information is of high economic value and, therefore, not readily shared with researchers and policy makers. Similarly, continental-scale trade of live animals has been linked to the emergence of highly pathogenic avian influenza (HPAI) in Asia and quantifying the extent of such trade, especially in Asia, is challenging [48,49].

Incorporating Connectivity into Disease Cartography

The impact of human connectivity in disease transmission and spread has been the subject of much contemporary epidemiological research. For example, theoretical and empirical models of

Table 2. Approaches to Quantifying and Mapping Human Mobility at a Range of Spatial and Temporal Scales

Approach	Key Features	Spatial and Temporal Scales	Data Sources	Limitations	Refs
Long-term international migration census	Describes longer-term population dynamics, allowing for understanding of genetic diversity	Country-wide estimates; 5–10 years	United Nations migration database	Heavily reliant on large-scale studies; often incomplete; poor spatial and temporal resolution	[43]
Within-country census migration data	Describes within-country population dynamics that might arise from economic and social transformations	Typically at district level; 1, 5, and 10 years	Integrated Public Use Microdata Series (IPUMS)	Captures only a small fraction of the population	[92]
Flight networks	Allows for rapid assessment of international spread of infectious diseases	Global coverage on ticketed commercial flight connections	International Air Transport Association	Captures only flight connection but not the subsequent travel patterns within each country	[28]
Commuting networks	Key to understand regular human behavioural patterns enabling predictions of city- and/or country-wide transmission	Daily commuting patterns often on a country level	National Census Surveys	Restricted to high-income countries with well-documented national statistics	[93]
Cell-phone data	Highly useful because they capture daily and/or seasonal human mobility patterns	Individual user-specific patterns of movement at varying temporal resolution	Mobile phone providers and mobile app companies	Only captures a fraction of the population; often high-income groups; limited to national borders	[63,94]
Logging devices, such as GPS	Highly accurate assessment of human contact networks	High spatial and temporal resolution	Individually funded projects where GPS devices are rolled out to the study group	Cost and time intensive; confined to small study groups	[42]

network contagion processes have been applied to interpret the global spread of pathogens via air travel [28]. Metapopulation models of disease transmission, building on those developed in ecology, have been applied to understand the local-scale spread of vaccine-preventable diseases and guide the effective deployment of resources [50,51]. Spatially explicit simulation modelling has been applied to evaluate the likely impacts of differing control scenarios in the face of disease outbreaks and their associated economic costs [52,53]. Similarly, statistical models have been applied to reconstruct past transmission networks after disease introductions have taken place [54]. Improving the rapid assessment of outbreak dynamics and accurate modelling of the number of infected individuals can directly impact public health policy interventions in affected and surrounding areas, as done during the Ebola response [55].

Falling costs of genetic sequencing technologies have lead to an explosion in the field of phylogenetics [56]. The application of these technologies to understanding the spread of species has led to the birth of the subdiscipline of phylogeography [57,58]. Phylogeographic methods have been particularly useful when applied to genetic disease surveillance data to understand the past spread of pathogens, and the role of connectivity in their dissemination [59,60]. Tying genetic data to contemporary outbreaks is still a relatively new but powerful approach [61].

While the spatial spread of diseases has been widely studied, the predominant focus of this research has been on epidemiological inference, rather than on the generation of predictive maps. The integration of this existing body of research and the methods it has developed into risk mapping offers the potential to broaden the scope and utility of disease maps. For example, if maps could identify likely paths of epidemics before they emerge, public health officials could implement responses in a more timely manner.

The integration of human connectivity data into disease cartography and public health policy planning has been applied in the context of malaria, where long-term human migration data have been used to identify groups of countries among which malaria cases are most likely to be imported [62]. Moreover, at a finer spatial scale, human movement data have been used to map regions acting as sources and sinks for malaria importations and this information has been combined with environmental risk mapping to provide rapidly updated risk maps as a tool for malaria control in countries seeking to eliminate the disease [63]. This research has demonstrated the need for mobility to be integrated in guiding elimination strategies, because such strategies can only be successful if local cases are reduced along with the threat of new introductions from endemic regions.

These examples illustrate both the potential and the feasibility of integrating human connectivity into disease cartography more widely. Malaria is among the most comprehensively studied of the major infectious diseases, and we have a detailed understanding of disease transmission processes, as well as systematic monitoring of disease prevalence in specific locations. By contrast, most of the world's infectious diseases are more poorly understood and reliable data on their distributions are scarcer, pointing towards prioritising the mapping of these particular diseases [64]. Challenges remain also in understanding how connectivity and its interaction with other epidemiological processes evolves over the different phases of an outbreak (growth phase versus decline in number of cases) and there is debate over how such phases can best be represented in epidemiological models. Another important challenge will be not only understanding how human populations move in general, but also identifying specific group behaviours that are more likely to contract and disseminate the pathogen of interest (e.g., poor rural communities in Southeast Asia that have different patterns of movement but high incidences of malaria). This is particularly important given that policy makers can only deploy limited resources when attempting to control and eliminate disease.

While the scientific community works to increase this epidemiological knowledge base, the development of flexible statistical approaches to incorporate connectivity into disease modelling could enable improvements over existing risk mapping in the near future.

New Frontiers in Mapping Diseases: Opportunities From Big Data

Increasing volumes of data from a variety of sources, and the technology to rapidly collate and organise them, offer enormous opportunities for improving disease surveillance and control in the near future [65]. Several digital resources for disease detection data have been established to collate data from online resources [65], including ProMED-mail [66], the Global Public Health Intelligence Network (GPHIN), HealthMap, and BioCaster [65,67]. These online resources can fill gaps in public health-reporting infrastructure and complement existing traditional surveillance systems of infectious diseases, particularly in areas where clinical surveillance capacity is underdeveloped and in disease outbreak situations where a rapid response is required [68,69].

In contrast with these advances in data assembly, predictive risk mapping has so far focussed on the release of static maps of disease distribution over long time periods. This situation reflects the computational and time demands of creating and refining the predictive spatial models used for disease-risk mapping. With advances in available computing power and improving

computational methods, the rate at which risk maps can be updated has increased to the point where dynamic, near-real-time risk maps are now a feasible prospect. Similarly to live data resources, systems that are able to rapidly and automatically acquire and verify disease occurrence data, generate predictive maps, and disseminate these to policy makers offer a major advance on the current state of affairs [6,70].

Such platforms offer the potential to rapidly translate contemporary epidemiological data into useful knowledge to facilitate control in disease outbreak situations, as well as to inform the general public [6,71].

Public Health Implications and Ways Forward

Spatial information about infectious diseases has broad implications for targeted public health interventions. For example, the targeted distribution of insecticide-treated mosquito nets (ITN) to key risk areas of high malaria prevalence has been shown to not only protect individuals from becoming infected, but also have wider implications in terms of spatial spillover effects [72,73]. Understanding population interconnectedness helps identify key individuals, risk groups, or regions in a network that would benefit most from targeted interventions, such as through vaccination [54,74]. Pinpointing key centres of infection can then have a multiplying effect, reducing case numbers, preventing catastrophic outbreaks, and reducing disease control implementation costs. Failure to account for human movements in previous public health campaigns has slowed eradication efforts for malaria and will continue to do so if no comprehensive mapping approach unifying economic, geographic, and medical approaches is developed [62]. Particularly in countries with increasing economic activity (i.e., Southeast Asia), these new mapping strategies can inform public health responses. This principle also extends to the better control and monitoring of livestock production and the reduction of contact between domestic animals, humans, and wildlife to reduce zoonotic disease emergence and spread [32,36,71].

Moving forward, individual diseases cannot be studied in isolation, but rather they should be assessed in the context of possible co-infections (e.g., TB and diabetes mellitus [75] or malaria and helminthiasis [76,77]). Quantifying the spatial patterns of the not only one disease but the composition of diseases that cause a significant burden to the population can ultimately help scale up intervention methods and reveal potential co-benefits to reduce the global burden of infectious diseases. This will become a critical issue for emerging economies such as China, Indonesia, Brazil, and Mexico, where both infectious and chronic diseases are now occurring in tandem [78,79]. For this purpose, spatially explicit information about urban poverty, healthcare access, and other socioeconomic determinants of health are yet to be developed on a global scale to help identify populations at risk of a comprehensive set of diseases as well as their interlinkages [2].

Particularly in the case of malaria, innovations in mapping have allowed the public health community to spatially target large ITN programs leading to significant reductions of infection and, in some cases, eradication [80,81]. Pinpointing the risk and burden of dengue globally has led vaccine developers to recognize the importance of this formerly poorly quantified disease [9]. Comprehensive mapping of infectious diseases will expose knowledge gaps and uncover new opportunities for translation of disease data into public health policies.

Concluding Remarks

Technological advances over the past decade have transformed infectious disease cartography from providing visualisations of spatial data and qualitative descriptions of disease risk into a framework for providing robust, quantified predictive maps to guide public health policy. The rapid pace of global change over the same time period has brought socioeconomic factors and

Outstanding Questions

How can researchers better communicate findings from disease cartography to public health officials and those in the field?

New data are becoming available at an ever-faster pace and with high volume: how can we better direct data acquisition towards the needs of researchers?

As data volume increases, what specific computational tools are necessary to allow rapid processing of disease and covariate data?

How can uncertainty be better integrated into disease-risk maps?

How can we best integrate human mobility data into a variety of disease models in a sensible way so that we can guide public health interventions and control?

global connectivity to the forefront as factors governing global disease risk (see Outstanding Questions). Infectious disease cartography has begun to adapt to these changing priorities, integrating new data sources and types of analysis into a unified cartographic framework. Particularly, data on disease control efforts need to be rigorously incorporated in disease mapping to better inform public health policy makers on their intervention strategies.

The world is increasingly connected and now more than ever diseases can spread rapidly across the globe. Simultaneously, a number of the world's major infectious diseases, such as cholera, are being brought under control and being locally eliminated [82,83]. Keeping track of the ever-changing global landscape of disease risk is crucial for limiting disease outbreaks and guiding control strategies.

Future work may be targeted towards an open data policy to allow reproducibility and increase transparency [84]. Moving from static disease mapping to continuously updated maps of contemporary disease risk poses significant technical challenges, but will enable the rapid translation of disease data into public health policy.

Acknowledgements

M.U.G.K. is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (#095066). S.I.H. and D.L.S. also acknowledge funding support from the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center (FIC), National Institutes of Health (NIH). S.I.H. and G.R.W.W. acknowledge funding from the International Research Consortium on Dengue Risk Assessment Management and Surveillance [IDAMS, European Commission 7th Framework Programme (#21803) http://www.idams.eu]. N.G. and D.M.P. are funded by a grant from the Bill and Melinda Gates Foundation (OPP1093011).

Disclaimer

Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Resources

[i] http://esa.un.org/unpd/wpp

[ii] http://sedac.ciesin.columbia.edu

[iii] www.worldpop.org.uk

[iv] www.itu.int

[v] http://hot.openstreetmap.org

[vi] www.msf.org.uk/missing-maps-project

References

1. Nordhaus, W.D. (2006) Geography and macroeconomics: new data and new findings. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3510–3517
2. Tatem, A.J. (2014) Mapping population and pathogen movements. *Int. Health* 6, 5–11
3. Elith, J. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecogeography* 29, 129–151
4. Hastie, T. *et al.* (2009) *The Elements of Statistical Learning.* (2nd edn), Springer
5. Lindström, T. *et al.* (2015) A bayesian ensemble approach for epidemiological projections. *PLoS Comput. Biol.* 11, e1004187
6. Hay, S.I. *et al.* (2013) Big data opportunities for global infectious disease surveillance. *PLoS Med.* 10, e1001413
7. Lounibos, L.P. (2002) Invasion by insect vectors of human disease. *Annu. Rev. Entomol.* 47, 233–266
8. Tsetsarkin, K.a. and Weaver, S.C. (2011) Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence. *PLoS Pathog.* 7, e1002412
9. Bhatt, S. *et al.* (2013) The global distribution and burden of dengue. *Nature* 496, 504–507
10. Hay, S.I. *et al.* (2009) A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med.* 6, e1000048
11. Hii, Y.L. *et al.* (2012) Forecast of dengue incidence using temperature and rainfall. *PLoS Negl. Trop. Dis.* 6, e1908
12. Weiss, D.J. *et al.* (2014) Air temperature suitability for *Plasmodium falciparum* malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malar. J.* 13, 171
13. Hahn, M.B. *et al.* (2014) Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLoS ONE* 9, e85725
14. Cunningham, A. and a *et al.* (2012) Disease invasion: impacts on biodiversity and human health. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 367, 2804–2806
15. Gong, P. *et al.* (2012) Urbanisation and health in China. *Lancet* 379, 843–852
16. Tatem, A.J. (2014) Mapping the denominator: spatial demography in the measurement of progress. *Int. Health* 6, 153–155
17. Tatem, A.J. *et al.* (2014) Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *Int. J. Health Geogr.* 13, 1–11
18. Murray, C.J.L. *et al.* (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic

analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2197–2223

19. Alirol, E. *et al.* (2011) Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* 11, 131–141
20. Bharti, N. *et al.* (2011) Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* 334, 1424–1427
21. Sutton, P. *et al.* (2001) Census from Heaven: an estimate of the global human population using night-time satellite imagery. *Int. J. Remote Sens.* 22, 3061–3076
22. Deville, P. *et al.* (2014) Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15888–15893
23. Bakillah, M. *et al.* (2014) Fine-resolution population mapping using OpenStreetMap points-of-interest. *Int. J. Geogr. Inf. Sci.* 28, 1940–1963
24. Linard, C. *et al.* (2012) Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE* 7, e31743
25. Bengtsson, L. *et al.* (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med.* 8, e1001083
26. Huang, Z. *et al.* (2013) An open-access modeled passenger flow matrix for the global air network in 2010. *PLoS ONE* 8, e64317
27. Lambin, E.F. and Meyfroidt, P. (2011) Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3465–3472
28. Brockmann, D. and Helbing, D. (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342, 1337–1342
29. Nunes, M.R.T. *et al.* (2014) Air travel is associated with intracontinental spread of dengue virus serotypes 1-3 in Brazil. *PLoS Negl. Trop. Dis.* 8, e2769
30. Gubler, D.J. (2011) Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Trop. Med. Health* 39, 3–11
31. Guagliardo, S.A. *et al.* (2014) Patterns of geographic expansion of *Aedes aegypti* in the Peruvian Amazon. *PLoS Negl. Trop. Dis.* 8, e3033
32. Horby, P.W. *et al.* (2013) Prospects for emerging infections in east and southeast Asia 10 years after severe acute respiratory syndrome. *Emerg. Infect. Dis.* 19, 853–860
33. Kraemer, M. *et al.* (2015) The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Elife* 4, e08347
34. Schaffner, F. *et al.* (2013) Public health significance of invasive mosquitoes in Europe. *Clin. Microbiol. Infect.* 19, 685–692
35. Tatem, A.J. *et al.* (2006) Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6242–6247
36. Kilpatrick, A.M. and Randolph, S.E. (2012) Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* 380, 1946–1955
37. Moyes, C.L. *et al.* (2014) Defining the geographical range of the *Plasmodium knowlesi* reservoir. *PLoS Negl. Trop. Dis.* 8, e2780
38. Bogoch, I.I. *et al.* (2014) Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* 385, 29–35
39. Khan, K. *et al.* (2013) Potential for the International Spread of Middle East Respiratory Syndrome in Association with Mass Gatherings in Saudi Arabia. *PLoS Curr.* Published online July 17, 2013. http://dx.doi.org/10.1371/currents.outbreaks.a7b70897ac2fa4f79b59f90d24c860b8
40. Khan, K. *et al.* (2014) Assessing the origin of and potential for international spread of Chikungunya virus from the Caribbean. *PLoS Curr.* Published online June 6, 2014. http://dx.doi.org/10.1371/currents.outbreaks.2134a0a7bf37fd8d388181539fea2da5
41. Smith, K.F. *et al.* (2007) Globalization of human infectious diseases. *Ecology* 88, 1903–1910
42. Stoddard, S.T. *et al.* (2013) House-to-house human movement drives dengue virus transmission. *Proc. Natl. Acad. Sci. U.S.A.* 110, 994–999
43. Abel, G.J. and Sander, N. (2014) Quantifying global international migration flows. *Science* 343, 1520–1522
44. Vayena, E. *et al.* (2015) Ethical challenges of big data in public health. *PLOS Comput. Biol.* 11, e1003904
45. Montjoye, Y. and De *et al.* (2015) Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science* 347, 356–359
46. Weaver, S.C. and Lecuit, M. (2015) Chikungunya virus and the global spread of a mosquito-borne disease. *N. Engl. J. Med.* 372, 1231–1239
47. Hawley, W.A. *et al.* (1987) *Aedes albopictus* in North America: probable introduction in used tires from northern Asia. *Science* 236, 1114–1116
48. Gilbert, M. *et al.* (2014) Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat. Commun.* 5, 4116
49. Kilpatrick, A.M. *et al.* (2006) Predicting the global spread of H5N1 avian influenza. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19368–19373
50. Metcalf, C.J.E. *et al.* (2014) Seven challenges in modeling vaccine preventable diseases. *Epidemics* 10, 11–15
51. Lessler, J. and Grenfell, B.T. (2012) Measurement of vaccine-derived immunity: how do we use all the data? *Expert Rev. Vaccines* 11, 747–749
52. Ferguson, N.M. *et al.* (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437, 209–214
53. Ferguson, N.M. *et al.* (2006) Strategies for mitigating an influenza pandemic. *Nature* 442, 448–452
54. Cauchemez, S. *et al.* (2014) Local and regional spread of chikungunya fever in the Americas. *Eurosurveillance* 19, 20854
55. WHO Ebola Response Team (2014) Ebola virus disease in West Africa: the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* 371, 1481–1495
56. McCormack, J.E. *et al.* (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66, 526–538
57. Pybus, O.G. *et al.* (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15066–15071
58. Frost, S.D.W. *et al.* (2015) Eight challenges in phylodynamic inference. *Epidemics* 10, 88–92
59. Lemey, P. *et al.* (2014) Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* 10, e1003932
60. Faria, N.R. *et al.* (2014) The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346, 56–61
61. Carroll, M.W. *et al.* (2015) Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature* 524, 97–101
62. Tatem, A.J. and Smith, D.L. (2010) International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12222–12227
63. Wesolowski, A. *et al.* (2012) Quantifying the impact of human mobility on malaria. *Science* 338, 267–270
64. Hay, S.I. *et al.* (2013) Global mapping of infectious disease. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20120250
65. Brownstein, J.S. *et al.* (2009) Digital disease detection - harnessing the web for public health surveillance. *N. Engl. J. Med.* 360, 2153–2157
66. Madoff, L.C. (2004) ProMED-mail: an early warning system for emerging diseases. *Clin. Infect. Dis.* 39, 227–232
67. Collier, N. *et al.* (2008) BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* 24, 2940–2941
68. Salathé, M. *et al.* (2012) Digital epidemiology. *PLoS Comput. Biol.* 8, e1002616
69. Chan, E.H. *et al.* (2011) Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl. Trop. Dis.* 5, e1206
70. Broniatowski, D.A. *et al.* (2014) Twitter: Big data opportunities. *Science* 345, 148
71. Pigott, D.M. *et al.* (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* 3, e04395
72. Bhattacharya, D. *et al.* (2013) Estimating the impacts of means-tested subsidies under treatment externalities with application to anti-malarial bednets. *NBER Working Pap. Ser.* 18833

Trends in Parasitology

73. Kim, D. and a *et al.* (2015) Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet* 386, 145–153
74. Lazer, D. *et al.* (2014) Big data. The parable of Google Flu: traps in big data analysis. *Science* 343, 1203–1205
75. Jeon, C.Y. and Murray, M.B. (2008) Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med.* 5, 1091–1101
76. Brooker, S. *et al.* (2007) Epidemiology of *Plasmodium*-helminth co-infection in Africa: populations at risk, potential impact on anemia, and prospects for combining control. *Am. J. Trop. Med. Hyg.* 77, 88–98
77. Mulu, A. *et al.* (2013) Epidemiological and clinical correlates of malaria-helminth co-infections in Southern Ethiopia. *Malar. J.* 12, 227
78. Bygbjerg, I.C. (2012) Double burden of noncommunicable and infectious diseases in developing countries. *Science* 337, 1499–1501
79. GBD 2013 Mortality and Causes of Death Collaborators (2015) Global, regional and national levels of age-specific mortality and 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 385, 117–171
80. Snow, R.W. *et al.* (2013) The malaria transition on the Arabian Peninsula: progress toward a malaria-free region between 1960–2010. *Adv. Parasitol.* 82, 205–251
81. Murray, C.J.L. *et al.* (2012) Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet* 379, 413–431
82. Chiyaka, C. *et al.* (2013) The stability of malaria elimination. *Science* 339, 909–910
83. Andrus, J.K. *et al.* (2011) Strengthening surveillance: confronting infectious diseases in developing countries. *Vaccine* 29 (Suppl. 4), D126–D130
84. Pigott, D.M. and Kraemer, M.U. (2014) Enhancing infectious disease mapping with open access resources. *Eurosurveillance* 19, 20989
85. Brady, O.J. *et al.* (2014) Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit. Vectors* 7, 338
86. Gething, P.W. *et al.* (2011) A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar. J.* 10, 378
87. Brady, O.J. *et al.* (2013) Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasit. Vectors* 6, 351
88. Gething, P.W. *et al.* (2011) Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit. Vectors* 4, 92
89. Lai, Y-S. *et al.* (2015) Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. *Lancet Infect. Dis.* 15, 927–940
90. Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* 36, 111–147
91. Purse, B.V. and Golding, N. (2015) Tracking the distribution and impacts of diseases with biological records and distribution modelling. *Biol. J. Linn. Soc.* 115, 664–677
92. Wesolowski, A. *et al.* (2014) Commentary: containing the ebola outbreak – the potential and challenge of mobile network data. *PLoS Curr.*
93. Tizzoni, M. *et al.* (2014) On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* 10, e1003716
94. Wesolowski, A. *et al.* (2013) The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* 10, 20120986

A.2. Appendix to Chapter 1

Figure 1 – figure supplement 1. Effect plots of covariates used in this study showing the marginal effect of each covariate on probability of for *Ae. aegypti* (1) and *Ae. albopictus* (2): Enhanced vegetation index - mean (a); Enhanced vegetation index - range (b); Maximum precipitation (c); Minimum precipitation (d); Temperature suitability (e); Urban areas (f); peri-urban areas (g).

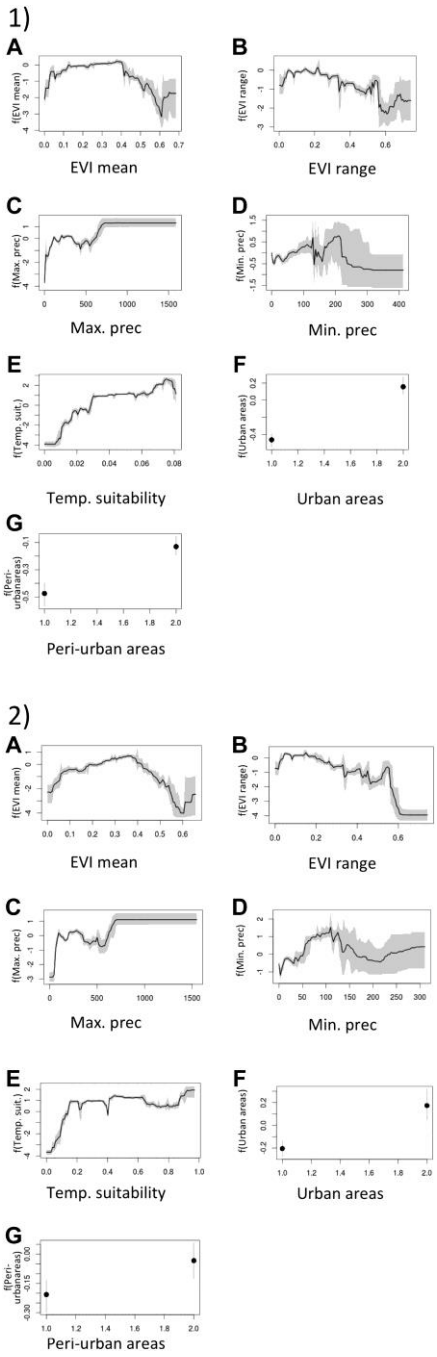


Figure 1 – figure supplement 2. Set of covariate layers used to predict the ecological niche of *Ae. aegypti* and *Ae. albopictus* described in detail in the Materials and section; a) enhanced vegetation index (EVI) annual mean, b) EVI annual range, c) annual monthly maximum precipitation, d) annual monthly minimum precipitation, e) temperature suitability for *Ae. albopictus,* f) temperature suitability for *Ae. aegypti,* g) rural, peri-urban and urban classification layer.

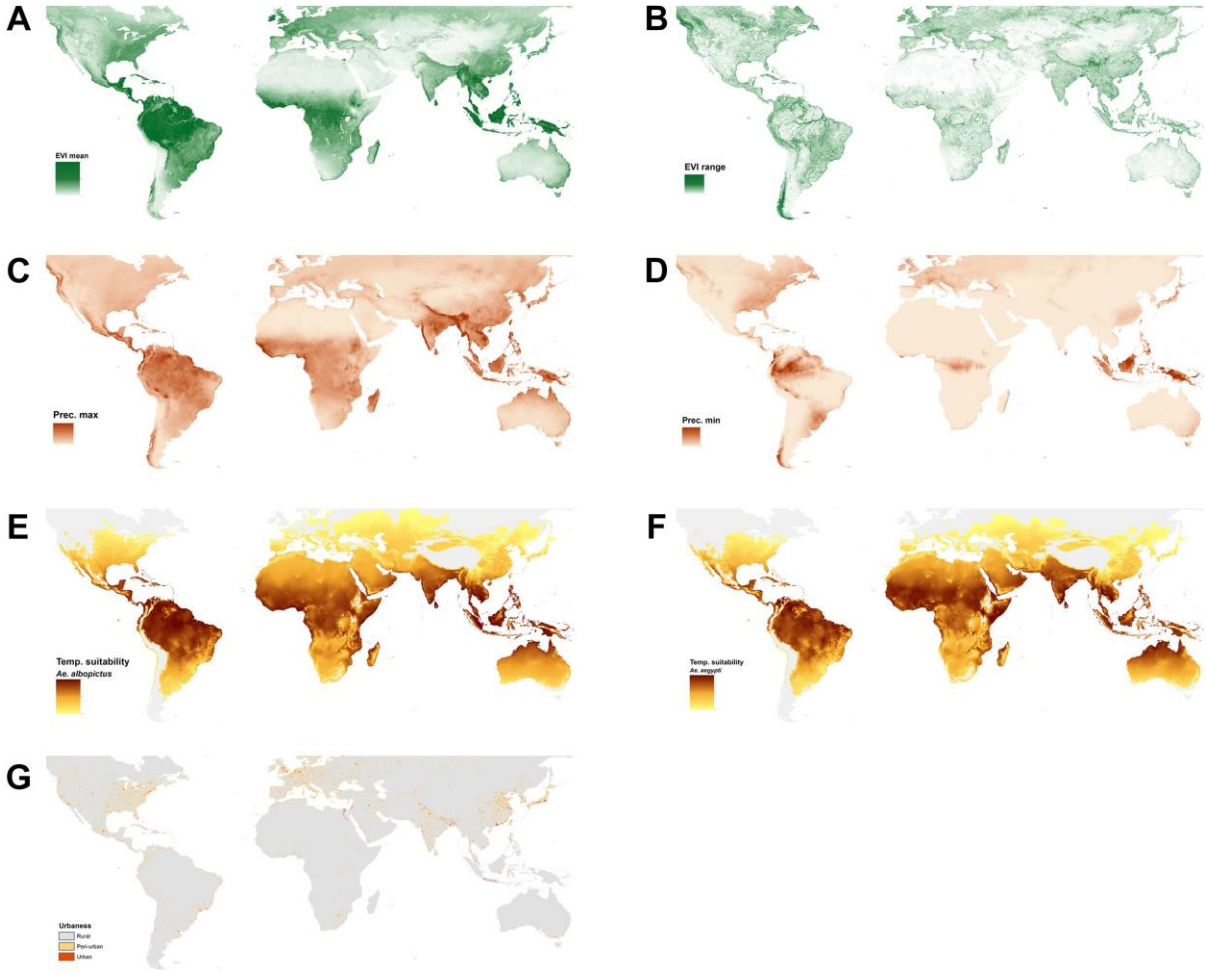


Figure 1 – figure supplement 3. Visualization of pixel level uncertainty calculated using the upper and lower bounds of the 95% confidence intervals associated with the prediction maps for *Ae. aegypti* (a) and *Ae. albopictus* (b).

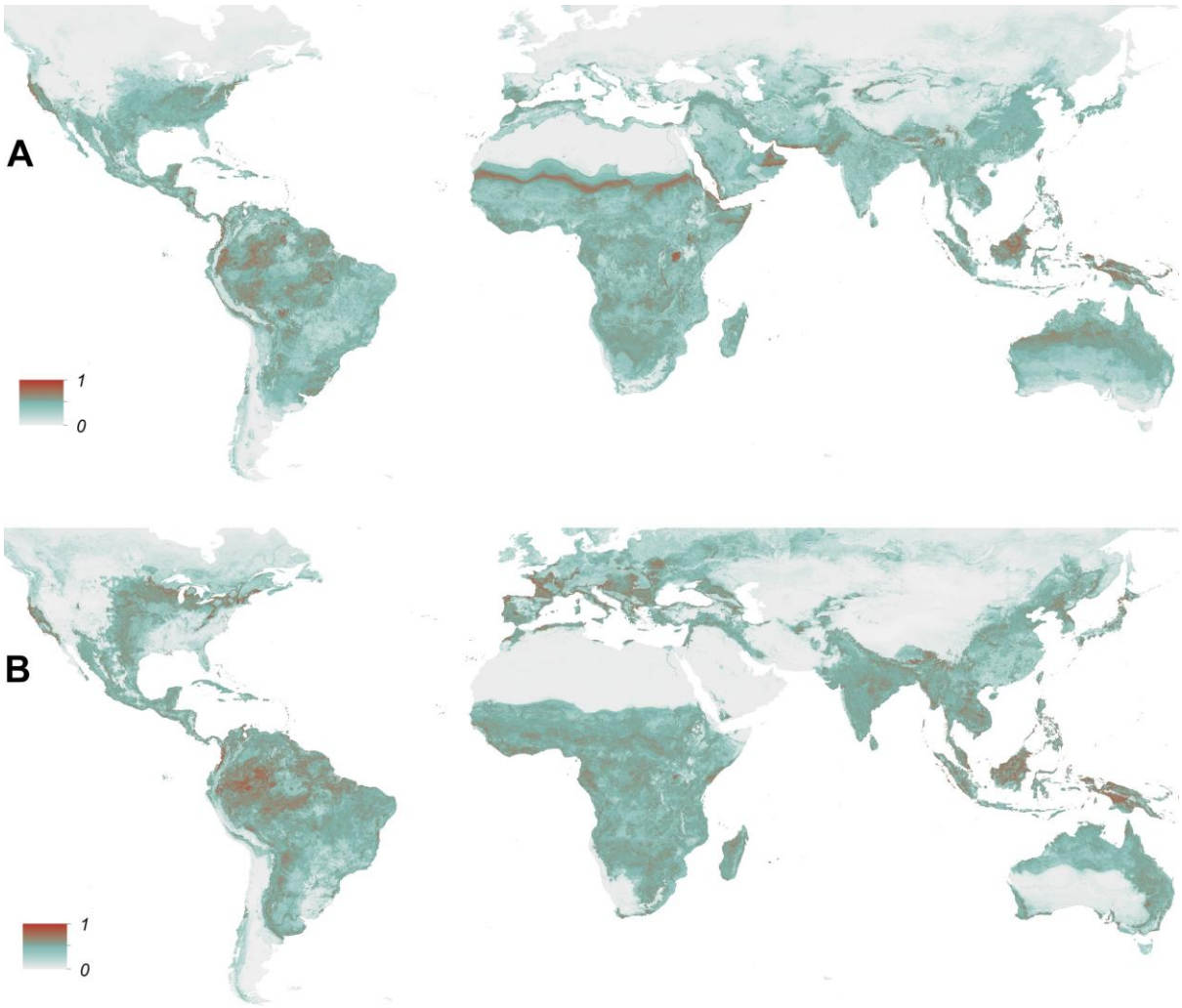
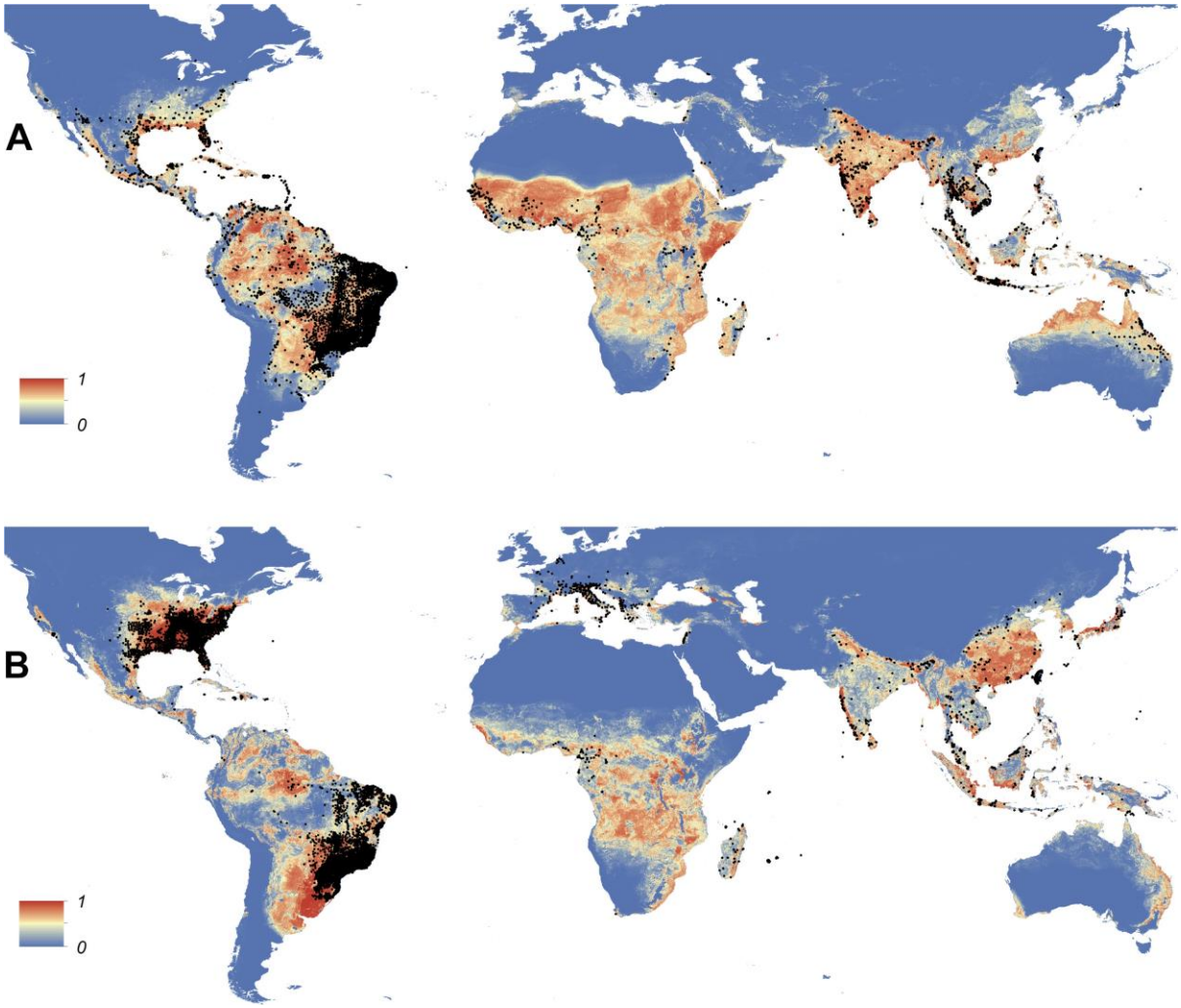


Figure 1 – figure supplement 4. The distribution of the occurrence database for Ae. aegypti (a) and Ae. albopictus (b) plotted on the underlying prediction surface.



A.3.1 Appendix to Chapter 3.1

THE LANCET Infectious Diseases

Supplementary webappendix

This webappendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Bogoch II, Brady OJ, Kraemer MUG, et al. Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *Lancet Infect Dis* 2016; published online Sept 1. http://dx.doi.org/10.1016/S1473-3099(16)30270-5.

Appendix 1: Technical Appendix

Appendix 1 (Technical Appendix)

1.1. Modeling the extent of the outbreak in the Americas

1.1.1. The distribution of the mosquito vectors of ZIKV in the Americas

1.1.2. The distribution of vector competence in the Americas

1.2. Predicting the extent and timing of risk of introduction in the Africa and Asia-Pacific

1.2.1. The distribution of the mosquito vectors of ZIKV in Africa and Asia-Pacific

1.2.2. The distribution of vector competence in Africa and Asia-Pacific

1.3. Model validation in Central America and the Caribbean

1.4. Prediction uncertainty maps

1.5. References

1.1. Modeling the extent of the outbreak in the Americas

Two separate approaches were taken to define geographic areas and times which present a risk of ZIKV export (described in 1.1., 1.1.1. and 1.1.2) from the Americas and ZIKV introduction (described in 1.2., 1.2.1. and 1.2.2) to Africa and Asia-Pacific. First we describe how we delineated areas in the Americas that pose a risk for ZIKV export in the over the year 2016.

1.1.1. The distribution of the mosquito vectors of ZIKV

Here we focus on two mosquito vectors of ZIKV; *Aedes aegypti and Ae. albopictus,* given they have been implicated in the current outbreak [1] and sufficient information is available on their global distribution [2]. The probability of detecting each of these mosquito species has previously been mapped on a global scale (see Kraemer at al. [3] for details). These continuous probability maps (ranging from 0-1) were converted to binary mosquito range maps (0 or 1) to delineate the absolute geographic extent of each species. To choose a suitable threshold for this binary conversion the probability of occurrence values in the final map were extracted for each pixel where each mosquito species has previously been reported from Kraemer at al. [2]. Plotting a histogram of these extracted values shows that in many areas where the mosquito species have been reported in the past there is a high predicted probability of occurrence (Fig. A1). There are also some data points with predicted probability of occurrence values, however, upon re-inspecting the dataset it was observed that these frequently fall in temperate areas where the species has only been observed transiently, often due to an abnormally warm year, for example. Our binary map aimed to represent long-term average conditions and we therefore want to choose a threshold value that excludes these transient events. To achieve this a statistical compromise was used and the threshold value was defined by the probability of occurrence value above which 90% of data points lie (red lines in Fig. 1A). All values in the continuous probability of occurrence map below this threshold were assigned a value of 0 (absence) while all values equal to or above this threshold were assigned a value of 1 (presence). All analyses were conducted on a global scale to maximize the data used to define this threshold, but resulting mosquito maps were cropped to just include North and South America and the Caribbean for subsequent analyses.

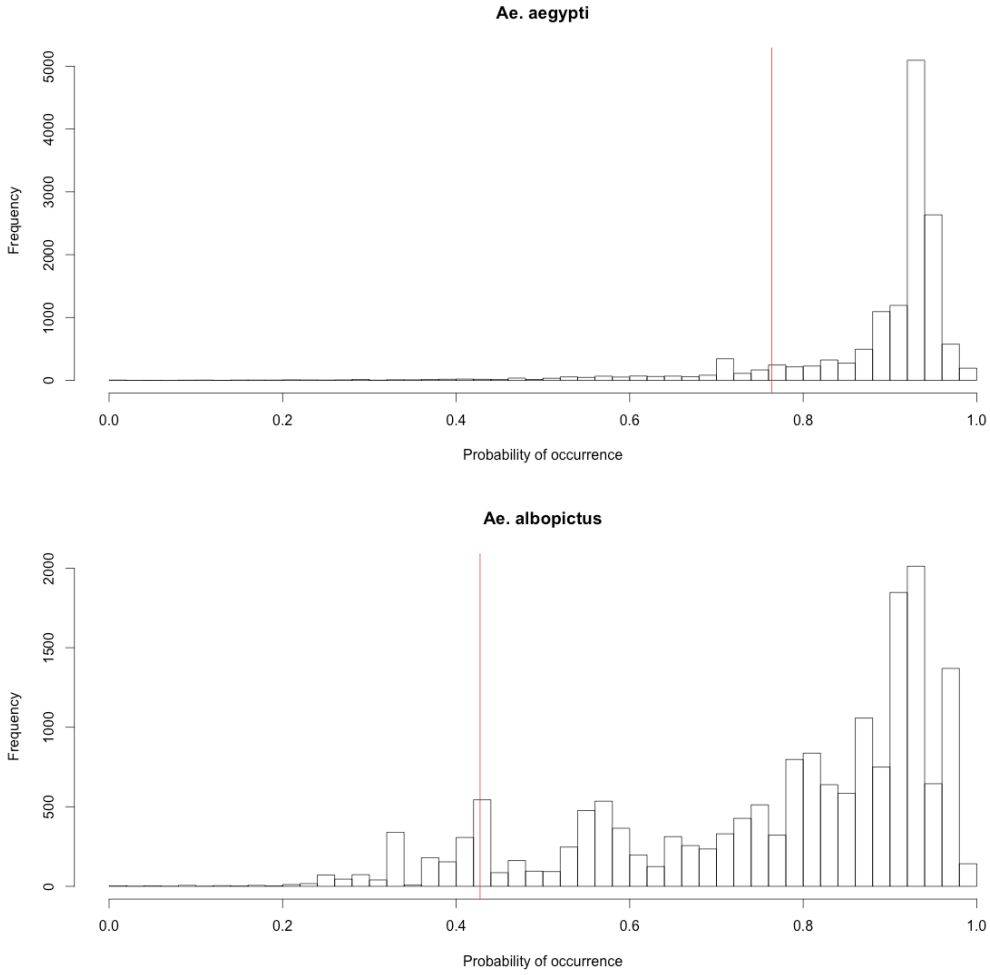


Figure A1. Histograms of the predicted probability of occurrence at locations where mosquitoes have previously been observed on a global scale in long-term average maps*. The red lines show threshold above which lies 90% of the data.*

1.1.2. The distribution of vector competence in the Americas

Once we had defined the geographic extent of *Ae. aegypti* and *Ae. albopictus,* the next step was to establish which regions are capable of sustaining ZIKV transmission throughout the year and thus present a significant risk of ZKIV exportation. To do this, we use a model of "temperature suitability" for dengue virus"- an index proportional to the temperature-driven components of vectorial capacity including

vector competence, vector survival, persistence and the effects of diurnal and seasonally fluctuating temperatures. While originally parameterized for dengue virus (DENV) the paucity of data on the extrinsic incubation period of ZIKV along with previous findings that arboviruses have similar EIPs [4] led us to consider that this is an appropriate model for ZIKV transmission potential. This model including parameterization, formulation and outputs is described in full in Brady et al. [5]. From this model output we use the persistence suitability index for each species, which calculates the number of days in a typical year on which mosquitoes would be capable of onward transmission given the presence of a viral source. Areas with a predicted persistence suitability of 355 days or greater were considered capable of sustaining year-round transmission, given the 10 days on which mosquitoes are not capable of onward transmission the virus would be able to persist in human populations [6].

Areas that were predicted to be both within the extent of either *Ae. aegypti* or *Ae. albopictus* and were capable of sustaining viral transmission for 355 days or greater in their respective mosquito species were assumed to present a risk of ZIKV export from the Americas in 2016. We subsequently refined this geographic range by excluding countries where transmission of another emerging arbovirus – chikungunya – was absent or limited in 2014-2015 (i.e. the United States, which had only 14 confirmed cases in Florida in 2014 was excluded).

1.2. Predicting the extent and timing of risk of introduction in Africa and Asia-Pacific

Once we had delimited areas at risk of ZIKV export within the Americas over the course of 2016 we then attempted to identify areas that may be capable of local

transmission in Africa and Asia-Pacific, should the virus be introduced via ZIKV-infected humans. To more fully understand how this risk may change over the rest of 2016 (and in subsequent years) we disaggregated our predictions by month as opposed to the annual summary maps used in 1.1.

1.2.1. The distribution of the mosquito vectors of ZIKV in Africa and Asia-Pacific

While *Aedes* populations persist year-round in the majority of locations within their range, their abundance is subject to significant intra-annual variation driven principally by rainfall, temperature among other factors [5, 7]. As such factors are particularly important in determining risk of a ZIKV introduction leading to a local outbreak we aimed to capture annual changes in relative abundance of each mosquito species.

To achieve this goal we used boosted regression tree models fitted to data from all time points (descried in Kraemer et al. 2015 [3]) and re-predicted the final maps substitution annual summary covariates for monthly specific covariates. Monthly covariates were available for *Aedes* oviposition persistence suitability [5], relative humidity (WorldClim v1.03) and precipitation (WorldClim v1.03) while covariates that showed little intra-annual change, such as urbanization were not varied from month-to-month. This resulted in 12 separate probability of occurrence maps for each mosquito species, which represented suitability of the midpoint of each month of a typical year.

These monthly maps were then processed to create an index of seasonal relative abundance that was compatible with annual average maps. This was done by

normalizing the predicted monthly values for each pixel to ensure their sum equaled the corresponding pixel value in the annual map:

$$x_{i,j} = z_{i,a} \frac{y_{i,j}}{\sum_{n=1}^{n=12} y_{i,n}}$$

where $x_{i,j}$ is the processed value of pixel i in processed map for month j, y is the value of the pixel before processing and $z_{i,a}$ is the value of pixel i in the separate annual average map from Kraemer et al. 2015 [3].

Each of these monthly maps was then converted to binary maps following a similar procedure as 1.1.1 to identify areas of the world that had sufficiently abundant mosquito populations for ZIKV transmission. As before, suitability values were extracted for locations where each species has been reported in the past. The maximum monthly value for each of these points was then plotted on a histogram and a threshold that incorporated 95% of the data chosen (Fig. A2). This approach attempted to define a threshold above which mosquito populations were abundant enough in a given month to be detected by routine trapping. It is then assumed that this level of abundance is sufficient for transmission of ZIKV. A more stringent threshold was used in comparison to the annual average maps (95% vs 90%) as in the latter, points could fall outside the range of the species due to both annual fluctuations and seasonal fluctuations. In the former, seasonal fluctuations were accounted for in the model and thus the 95% only excludes those points that are suspected to be due to inter-annual fluctuations in habitat suitability.

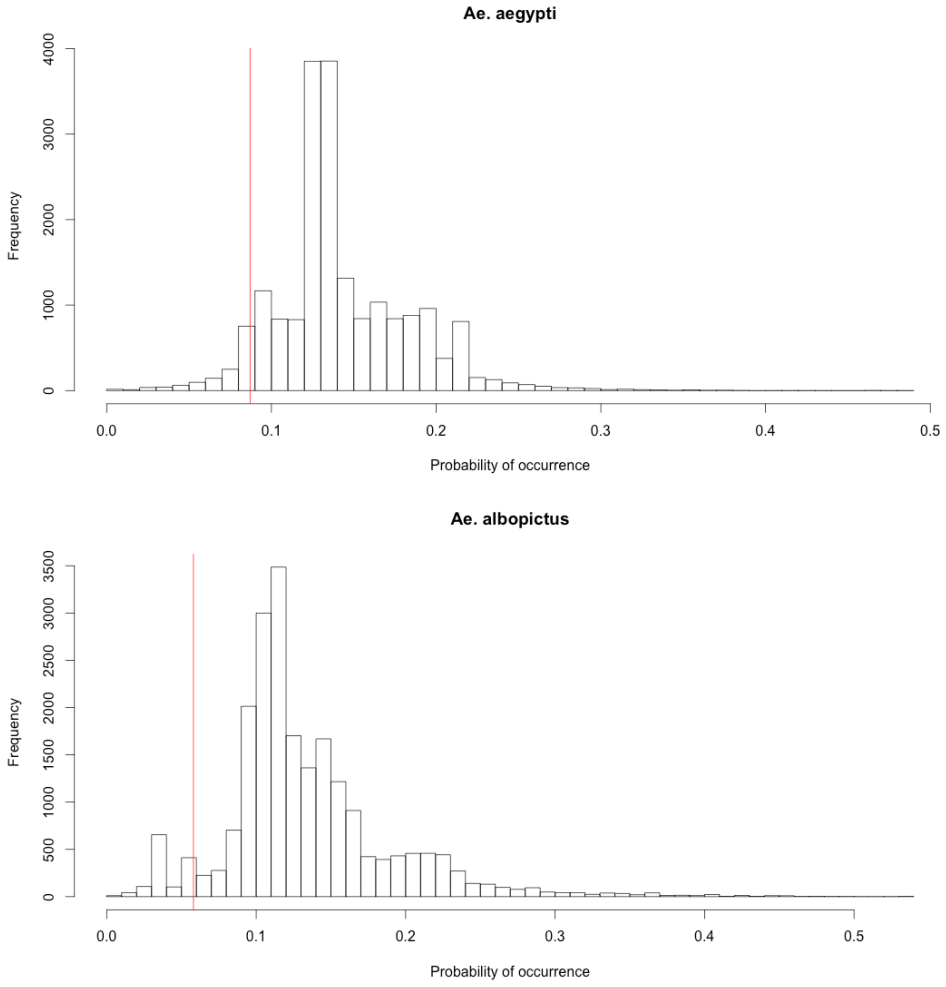


Figure A2. Histograms of the predicted probability of occurrence in the most suitable month at locations where mosquitoes have previously been observed on a global scale*. The red lines show threshold above which lies 95% of the data.*

1.2.2. The distribution of vector competence in Africa and Asia-Pacific

As in 1.1.2., the virus persistence model outputs from Brady et al. 2014 [5] were used to further refine areas capable of sustained virus transmission within areas where vector species are present and abundant. To do this, the monthly maps from this model output were used which show the number of days in each month on which

mosquitoes would be capable of onward transmission given an already present source of infection. If onward transmission was possible on at least 21 days an area was assumed at sufficient risk for an outbreak. Each monthly map was then threshold based on this 21 day threshold and each monthly vector competence map was then used to mask out areas where competent, abundant mosquito species are present, but where the temperature does not permit them to transmit ZIKV in that given month of the year.

1.3. Model validation from evidence in the Americas

ZIKV causes only relatively mild infections in humans and is thus it remains a challenge to directly infer the timing of the arrival of a pathogen using only epidemiological data [8]. Passenger flight data have been used successfully to estimate the risk of exportation from regions reporting ongoing ZIKV transmission using a combination of genetic and epidemiological evidence [9]. In addition, such models have linked the increase in passenger traffic to the timing of arrival of the virus. Previous studies have not used the level of receptivity of local transmission in the receiver country and thus may ignore an important element of the epidemiological significance of the spread of ZIKV.

1.4. ***Prediction uncertainty maps***

Below are global maps outlining the predicted distribution of *Ae. aegypti* and *Ae. albopictus*, and the predicted competence of each vector.

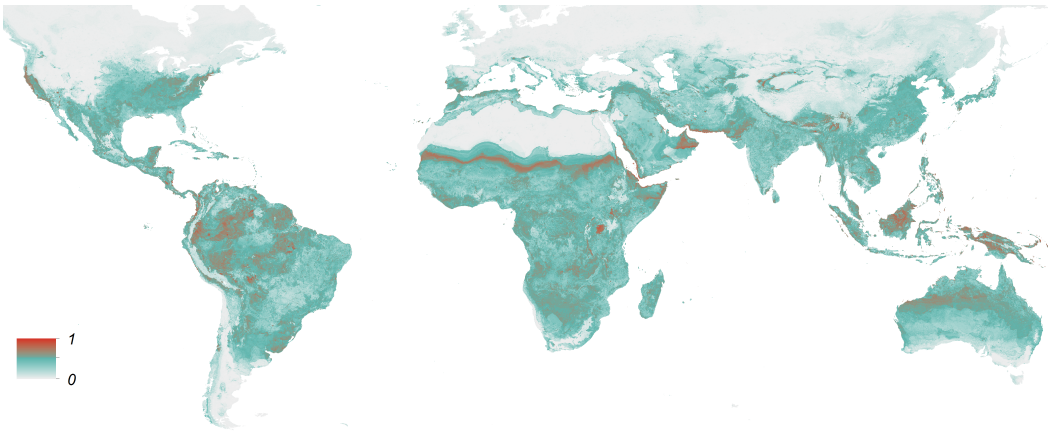


Figure A3. Predicted relative uncertainty from the *Ae. aegypti* distribution map. Full details in Kraemer et al. 2015 [3].

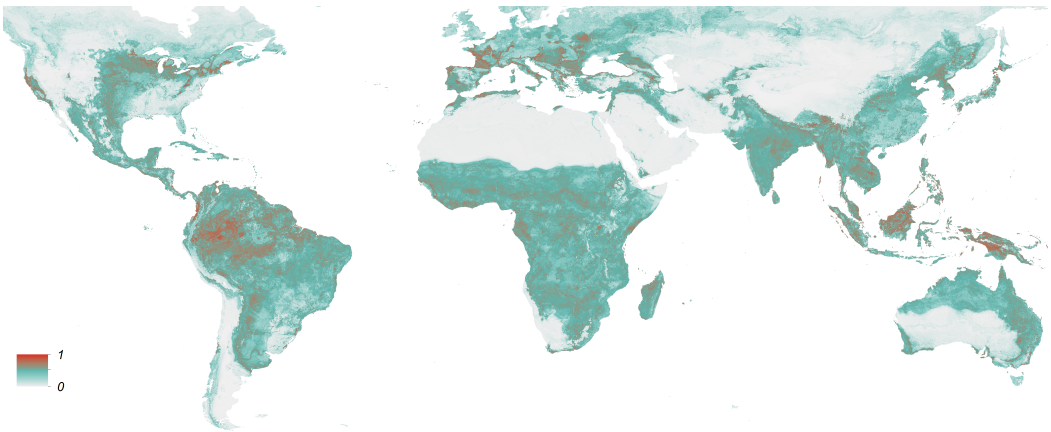


Figure A4. Predicted relative uncertainty from the *Ae. albopictus* distribution map. Full details in Kraemer et al. 2015 [3].

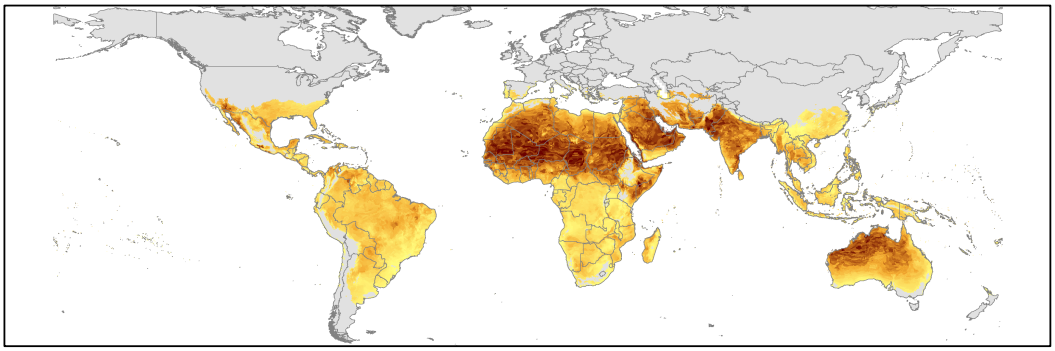


Figure A5. Predicted relative uncertainty from the *Ae. aegypti* vector competence model. Full details in Brady et al. 2014 [5].

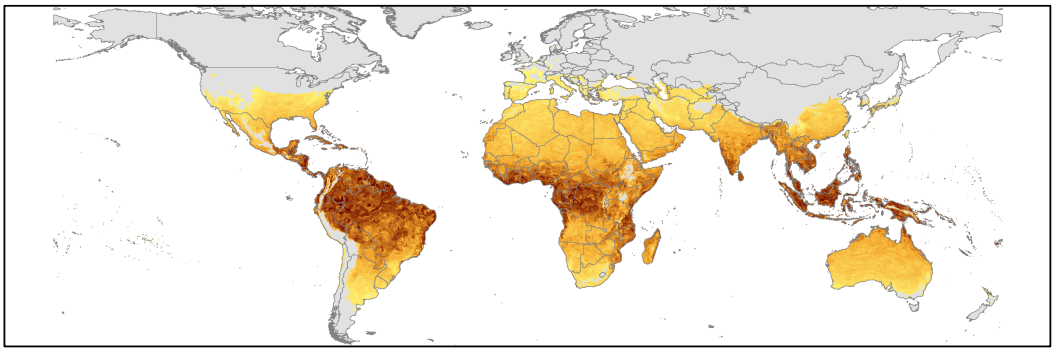
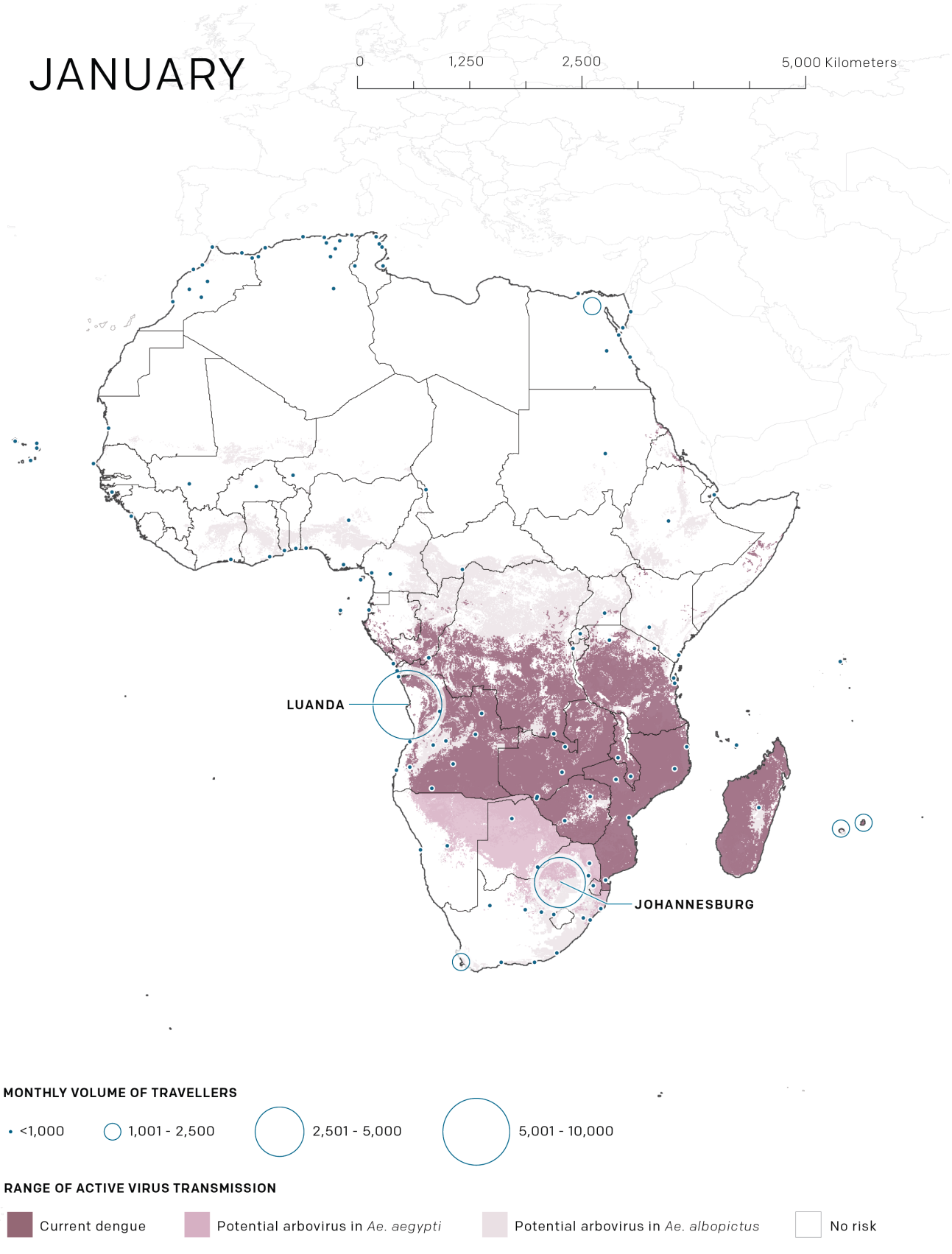


Figure A6. Predicted relative uncertainty from the *Ae. albopictus* vector competence model. Full details in Brady et al. 2014 [5].

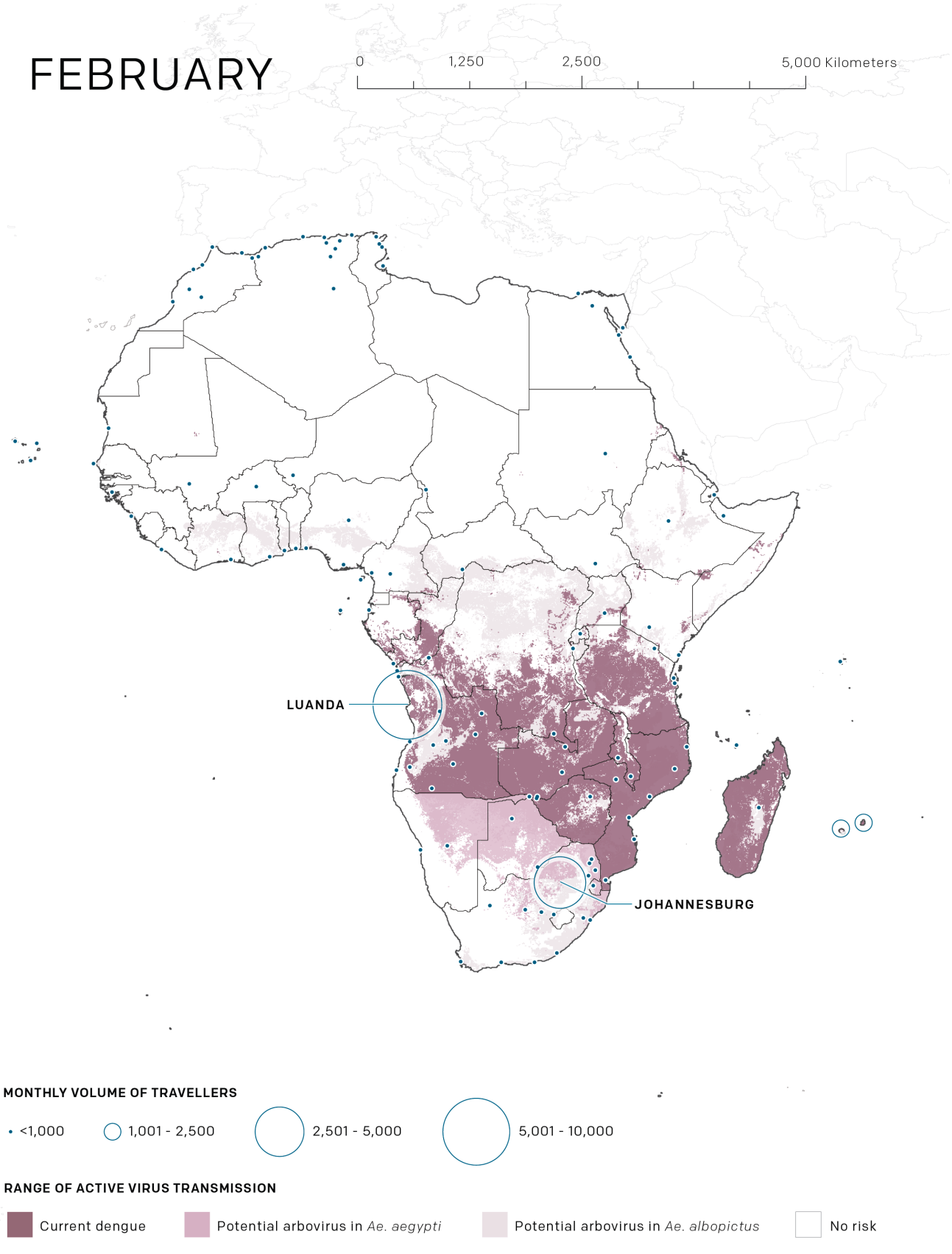
1.5. References

1. Chouin-Carneiro, T., et al., *Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus.* PLoS Negl Trop Dis, 2016. **10**(3): p. e0004543.
2. Kraemer, M.U., et al., *The global compendium of Aedes aegypti and Ae. albopictus occurrence.* Sci Data, 2015. **2**: p. 150035.
3. Kraemer, M.U., et al., *The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus.* Elife, 2015. **4**: p. e08347.
4. Lambrechts, L., et al., *Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7460-5.
5. Brady, O.J., et al., *Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission.* Parasit Vectors, 2014. **7**: p. 338.
6. Chan, M. and M.A. Johansson, *The incubation periods of Dengue viruses.* PLoS One, 2012. **7**(11): p. e50972.
7. Scott, T.W., et al., *Longitudinal studies of Aedes aegypti (Diptera: Culicidae) in Thailand and Puerto Rico: population dynamics.* J Med Entomol, 2000. **37**(1): p. 77-88.
8. Pacheco O, Beltrán M, Nelson CA, Valencia D, *et al.* Zika Virus Disease in Colombia - Preliminary Report. *N Engl J Med* 2016; [Epub ahead of print].
9. Faria NR, Azevedo R do S da S, Kraemer MUG, *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* 2016; **352**: 345–9.

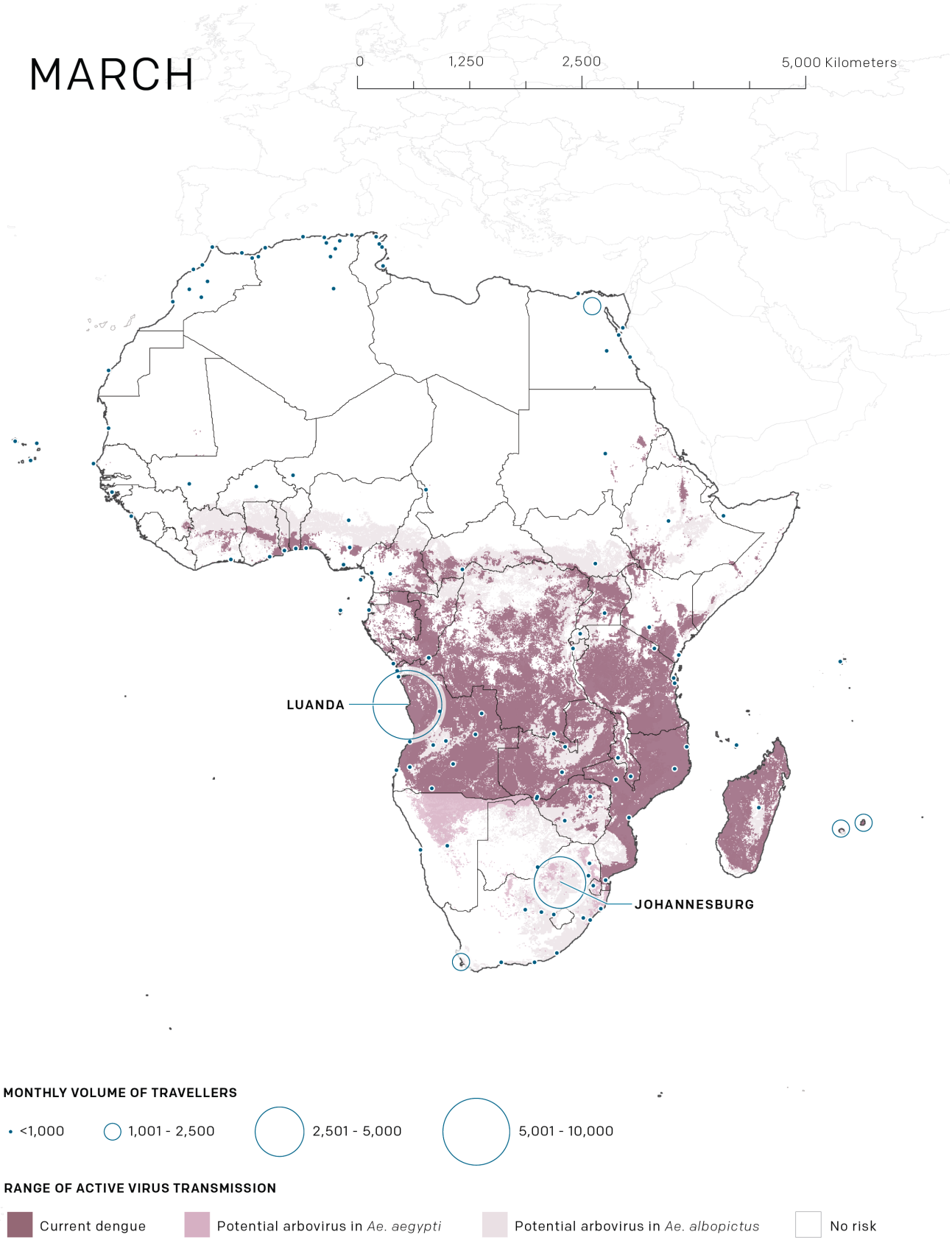
Appendix 2: Monthly African Model Outputs


JANUARY
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

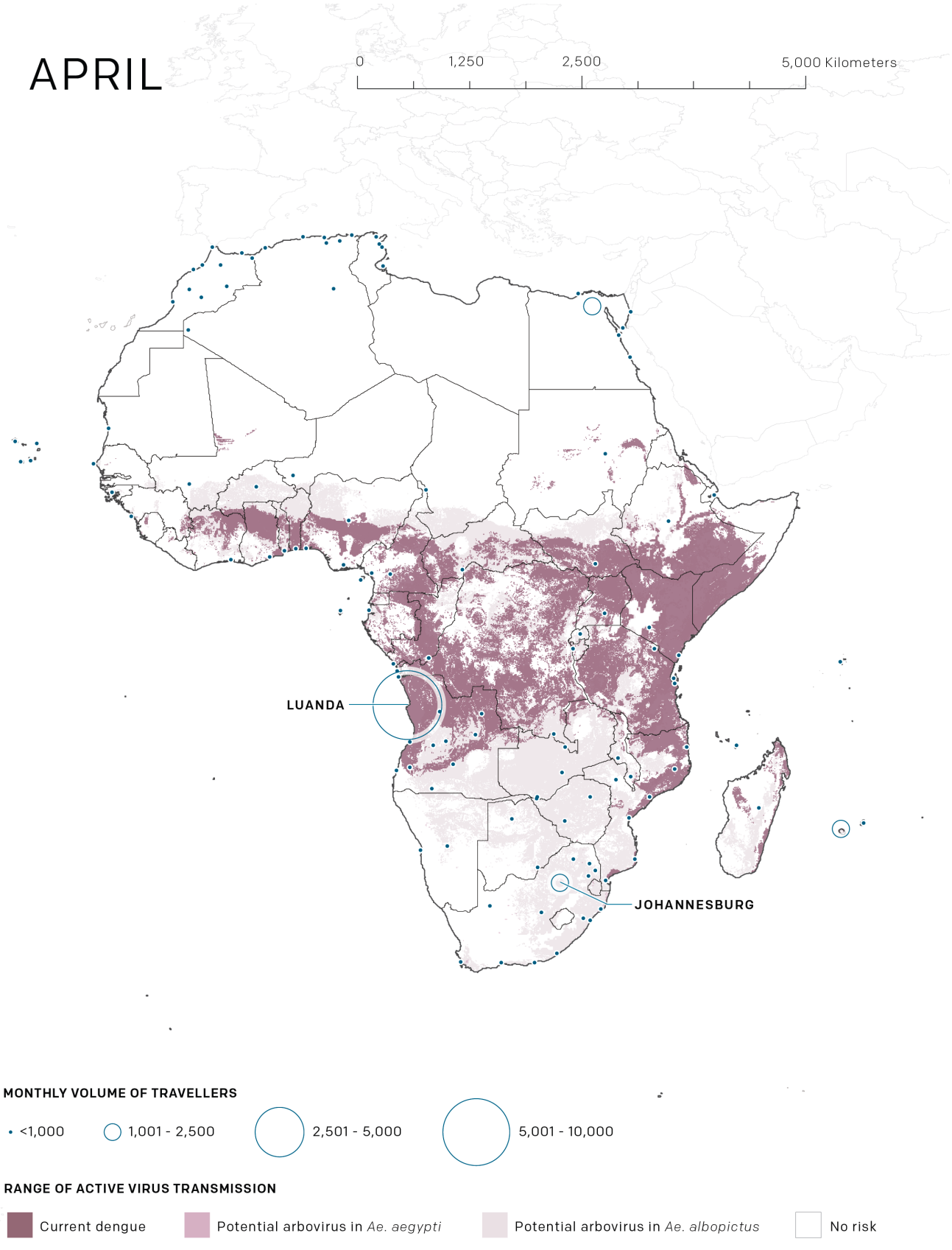
13


FEBRUARY
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

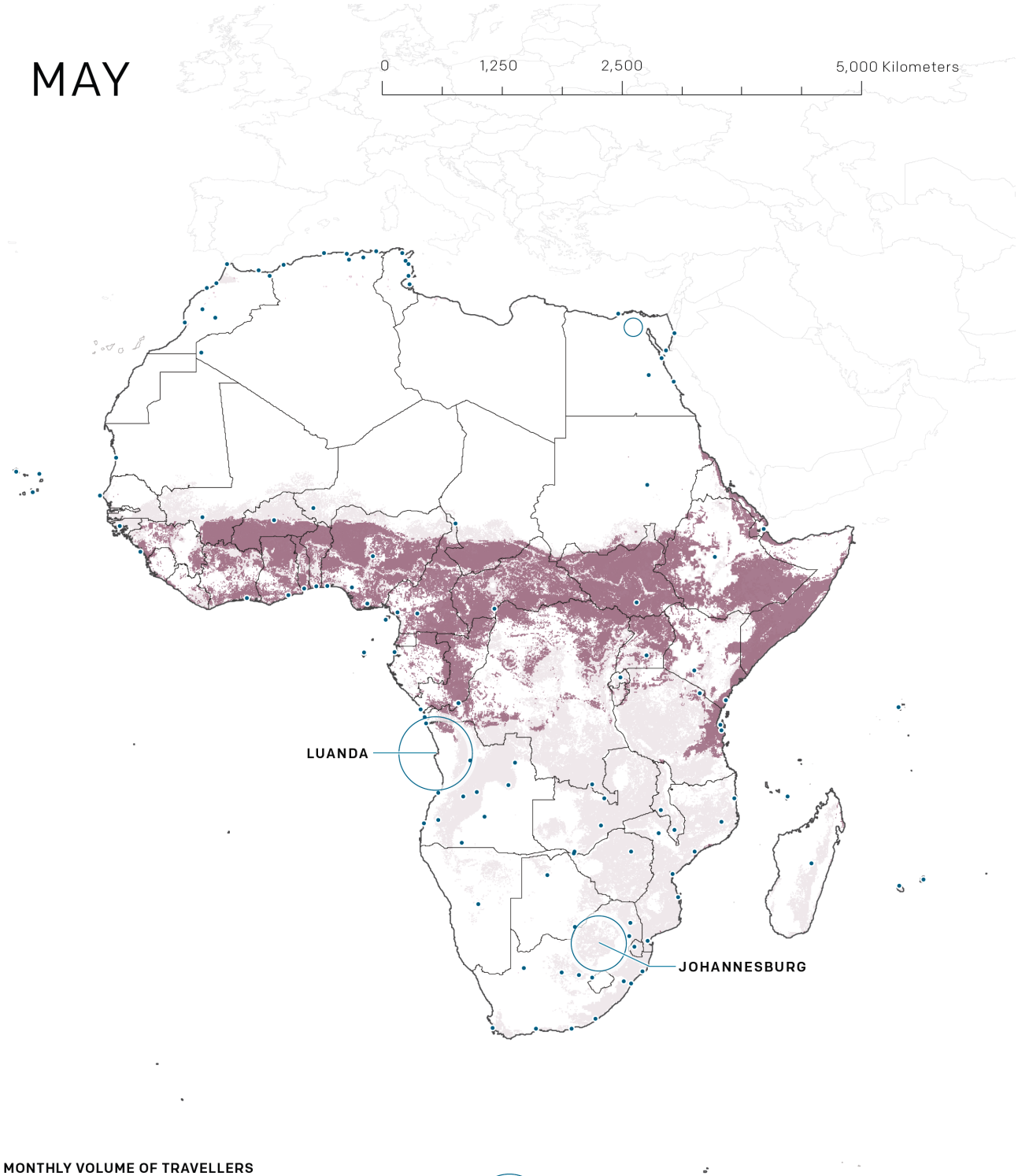
14


MARCH
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

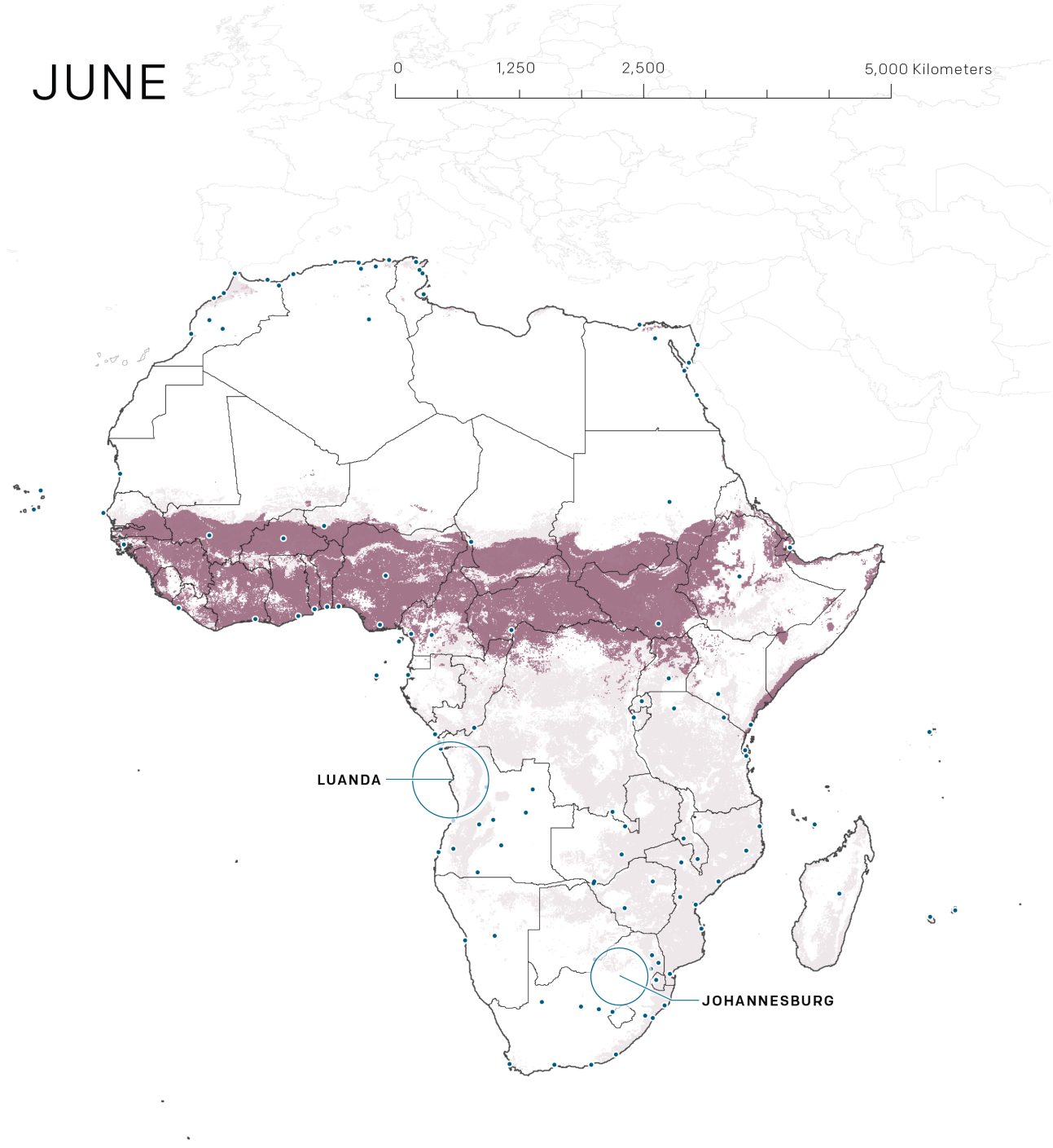
15

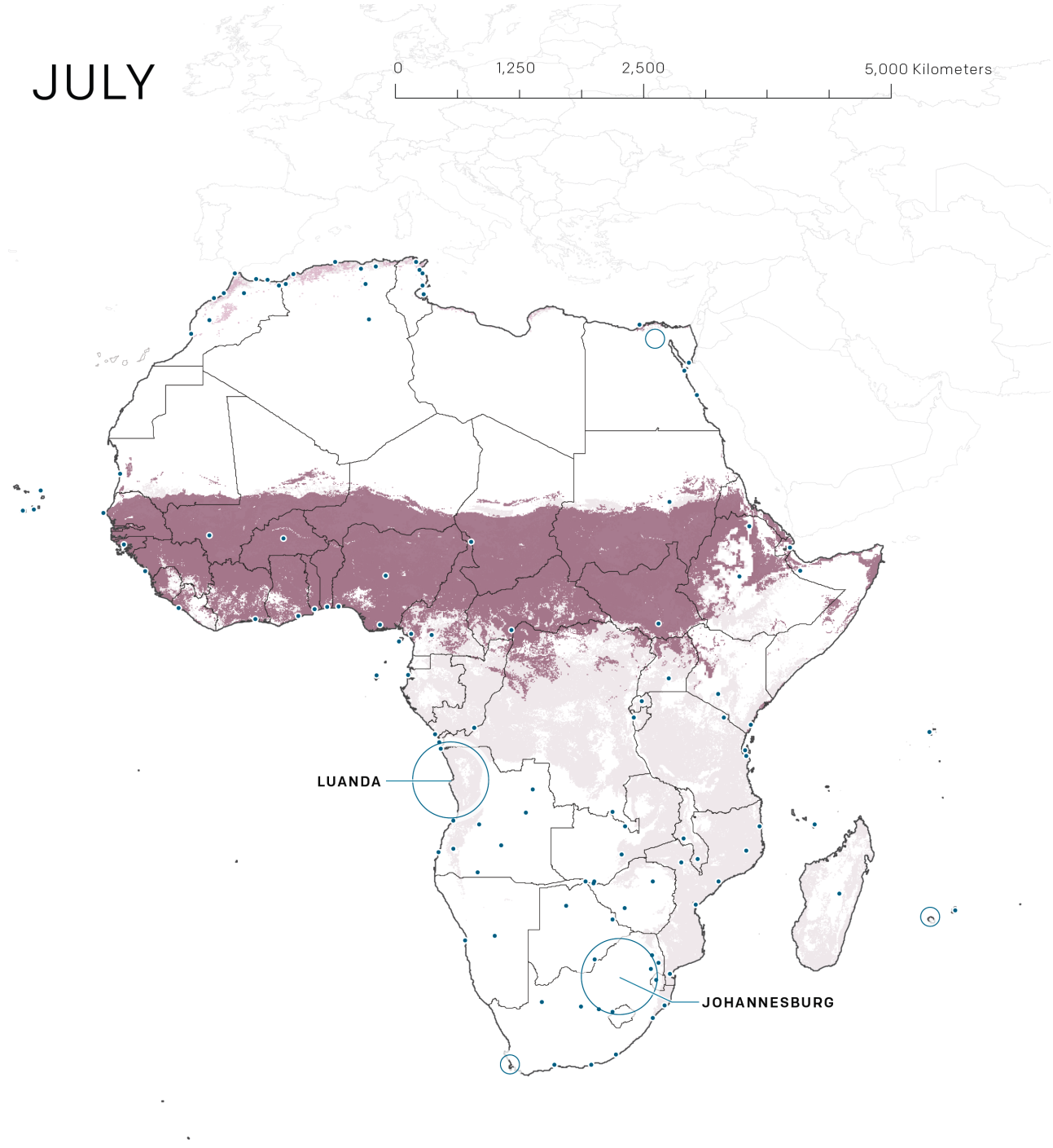

APRIL
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

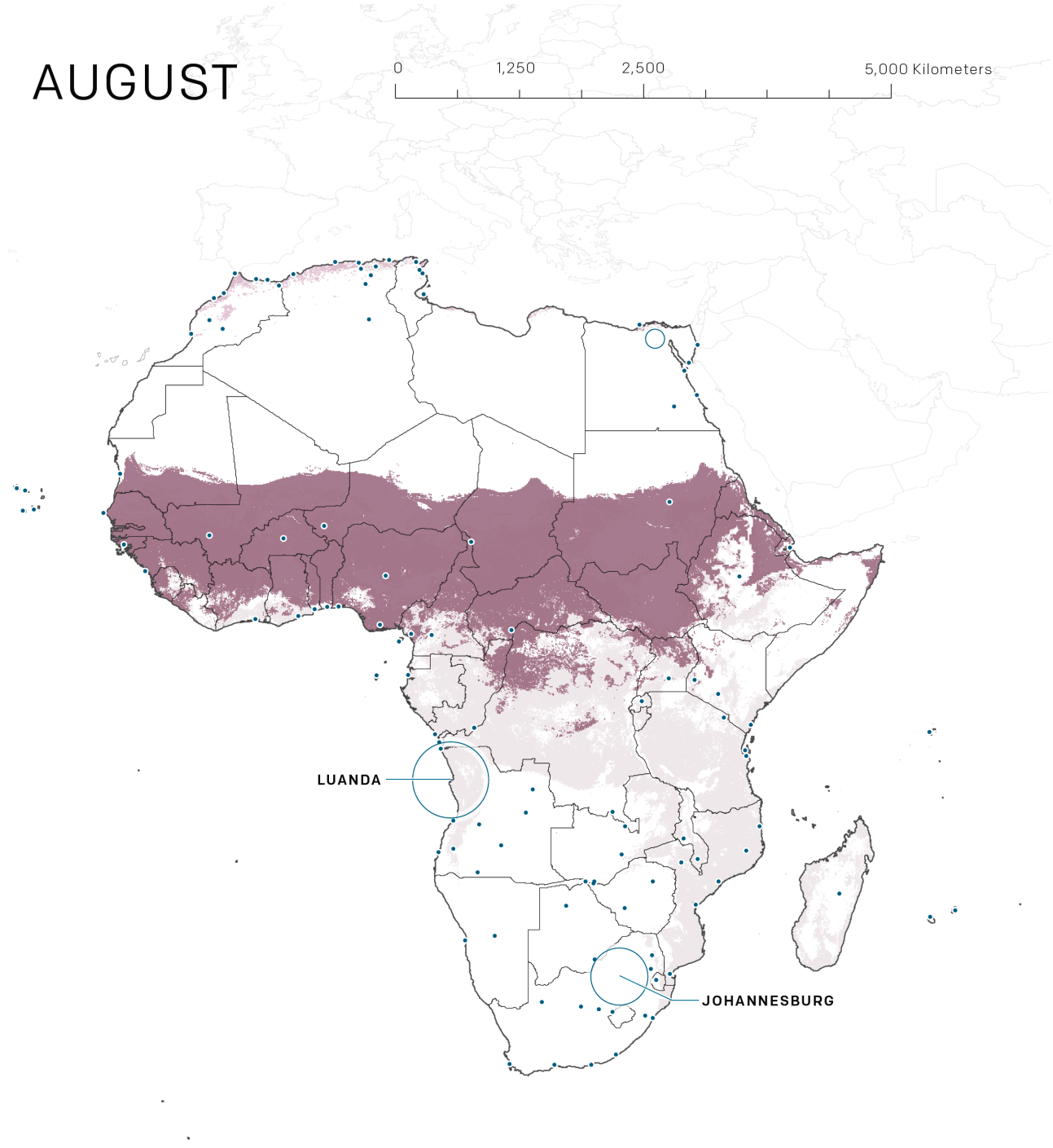
16


MAY
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

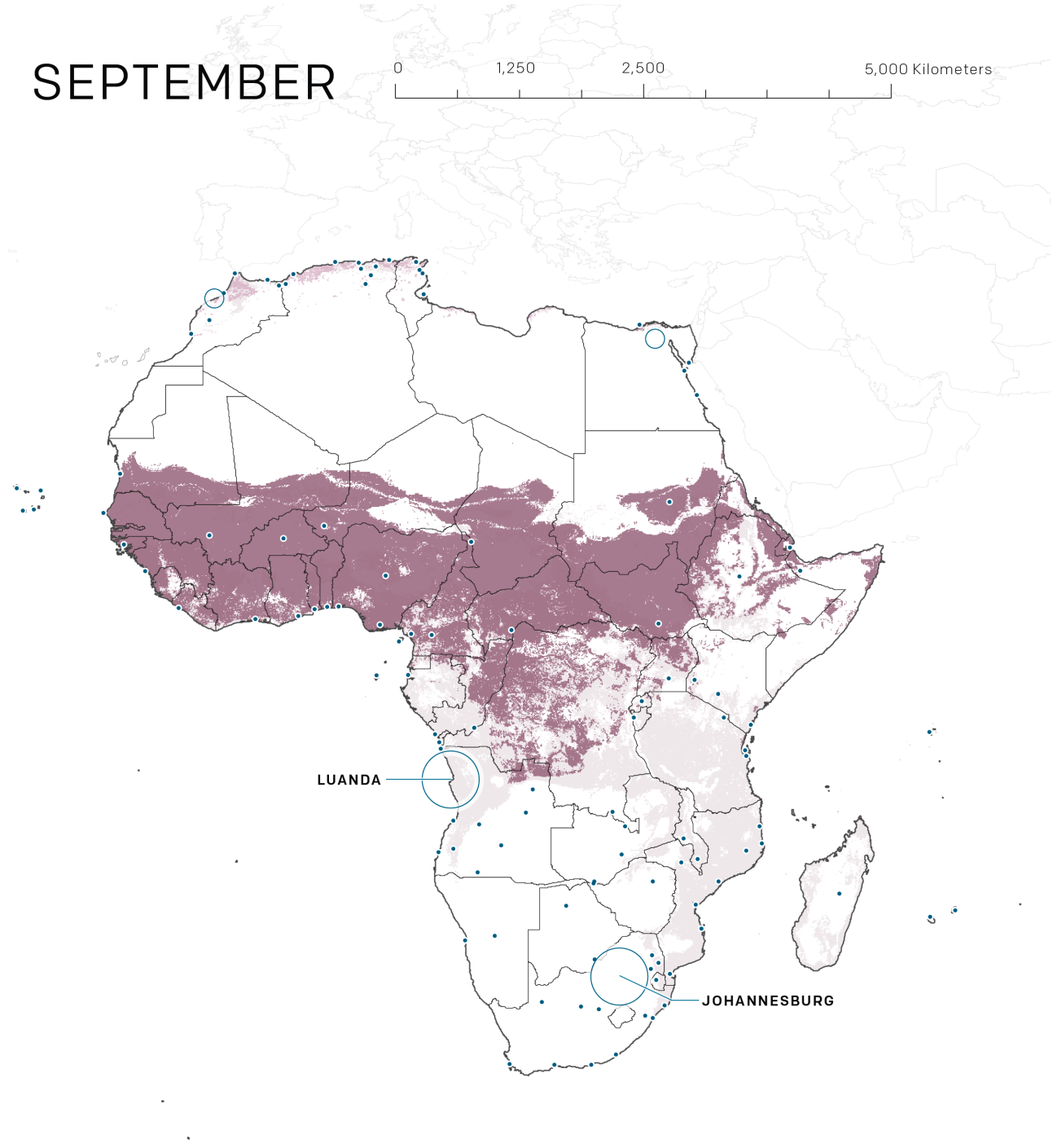
17


JUNE
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

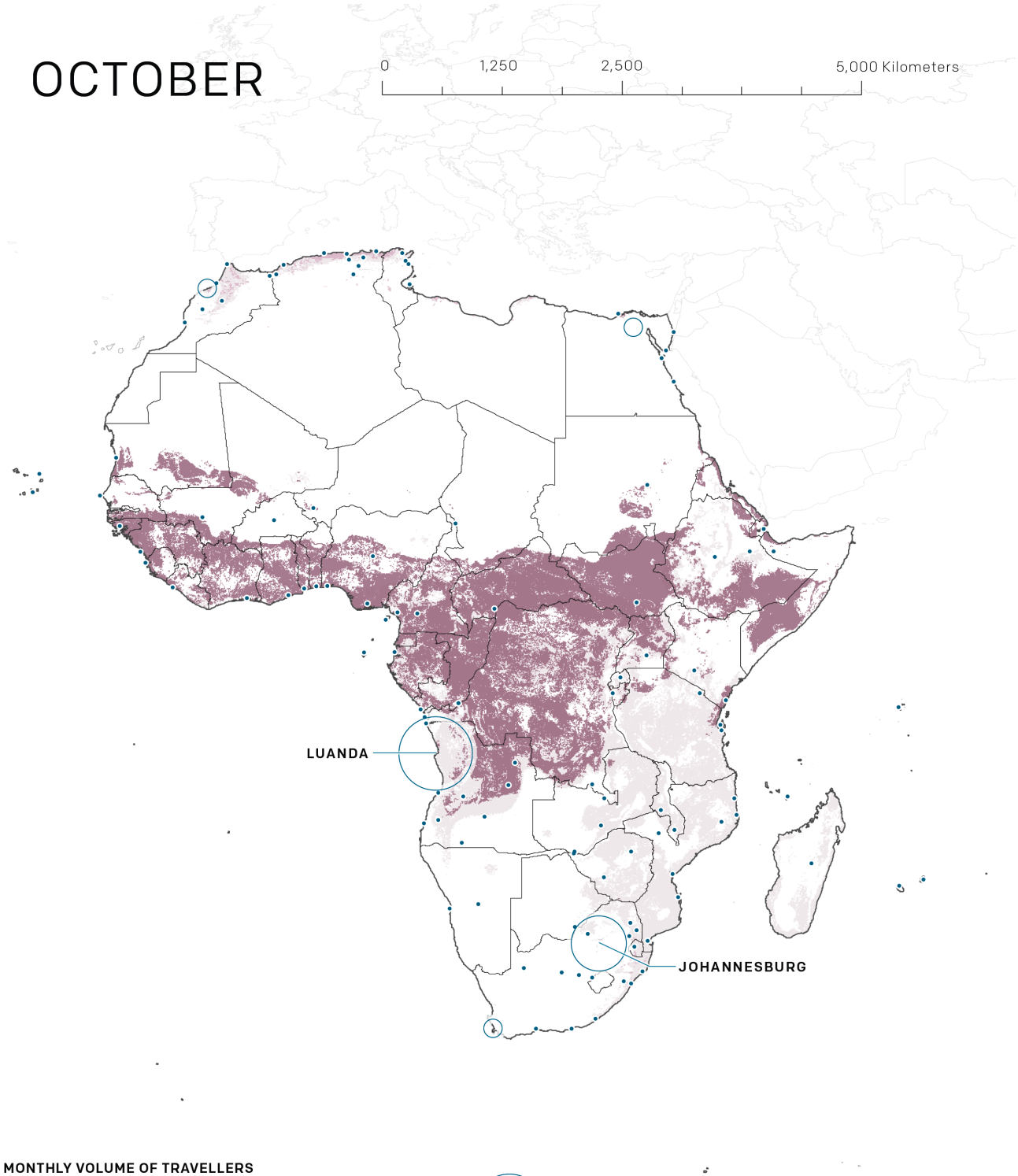

JULY
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


AUGUST
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

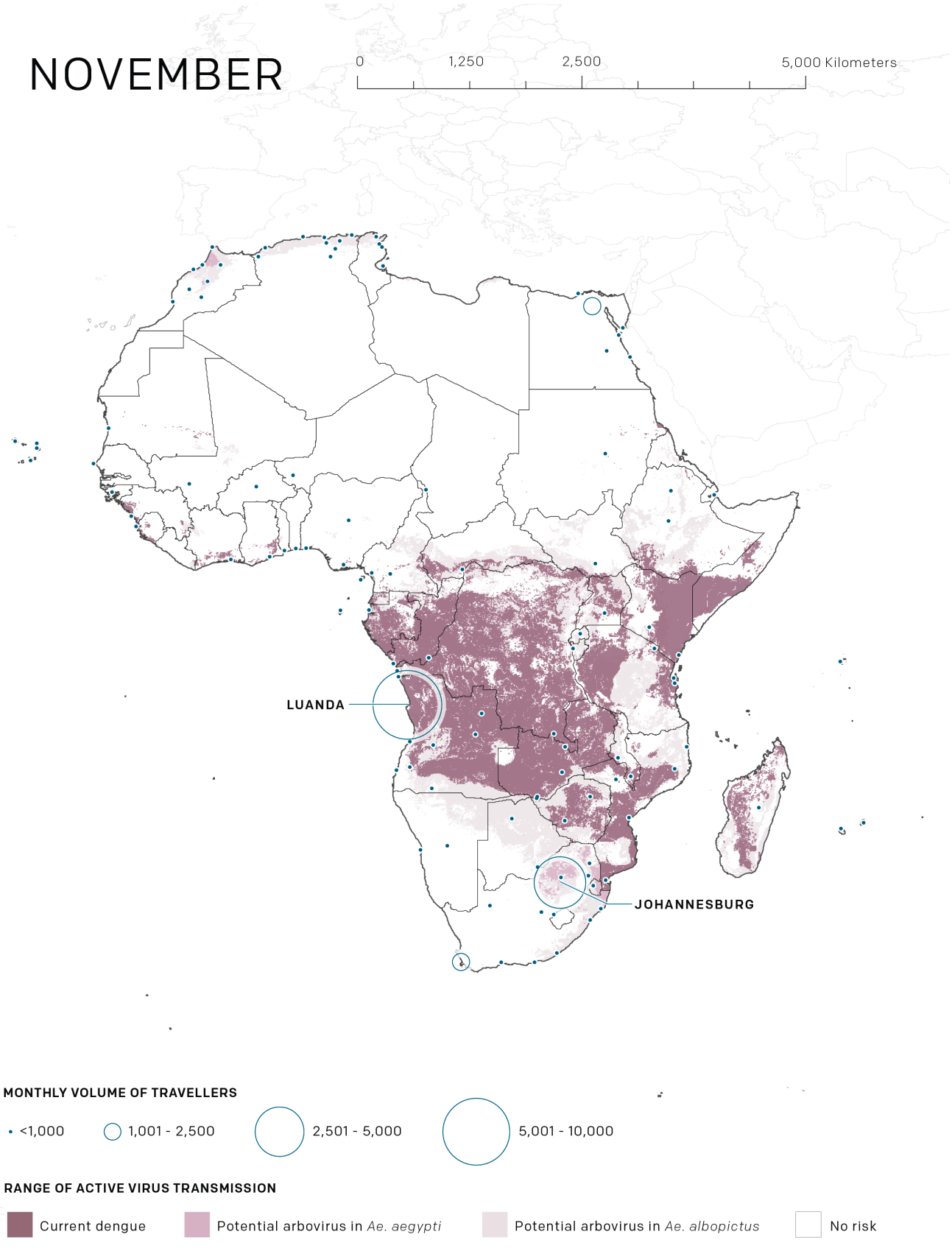
20


SEPTEMBER
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

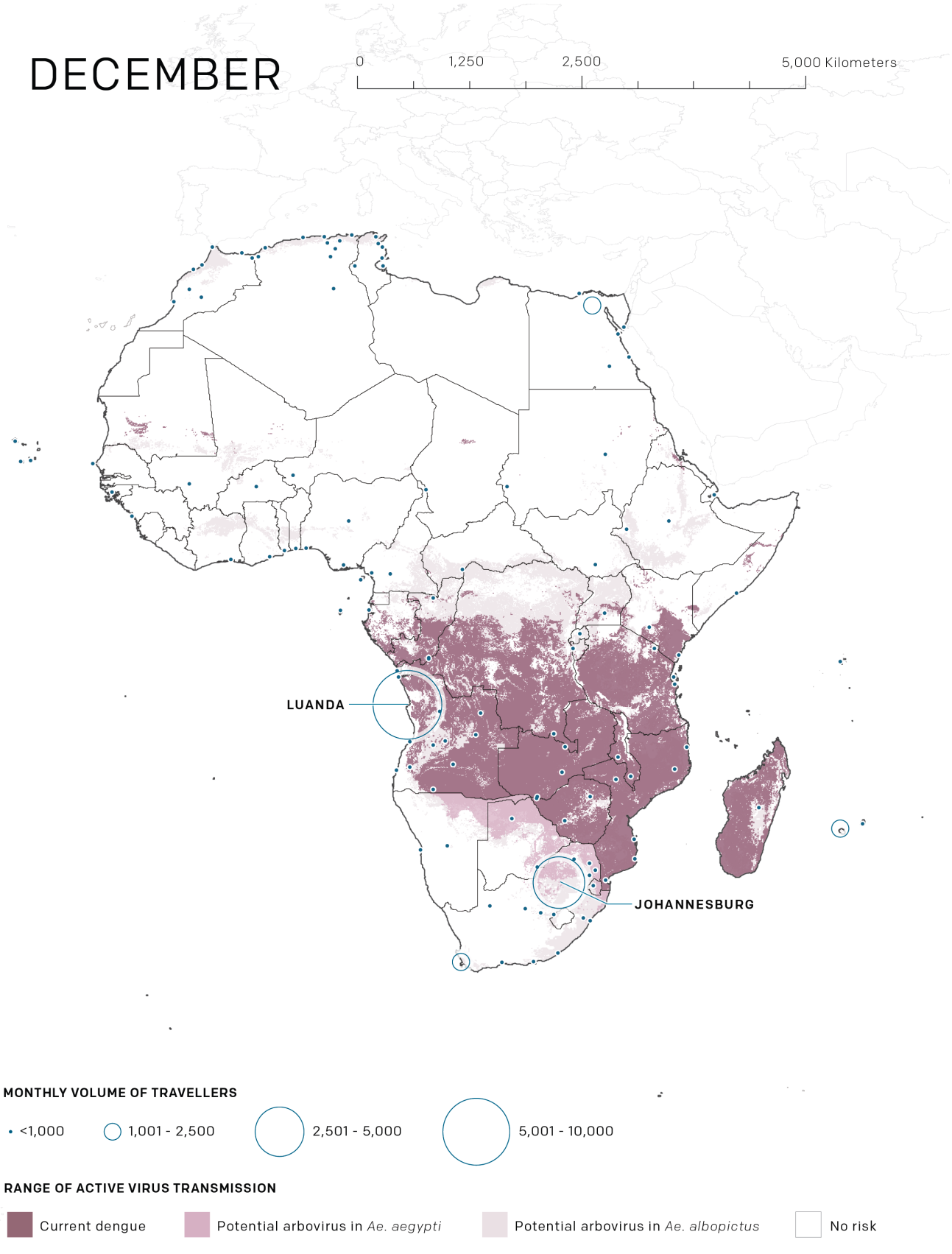
21


OCTOBER
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

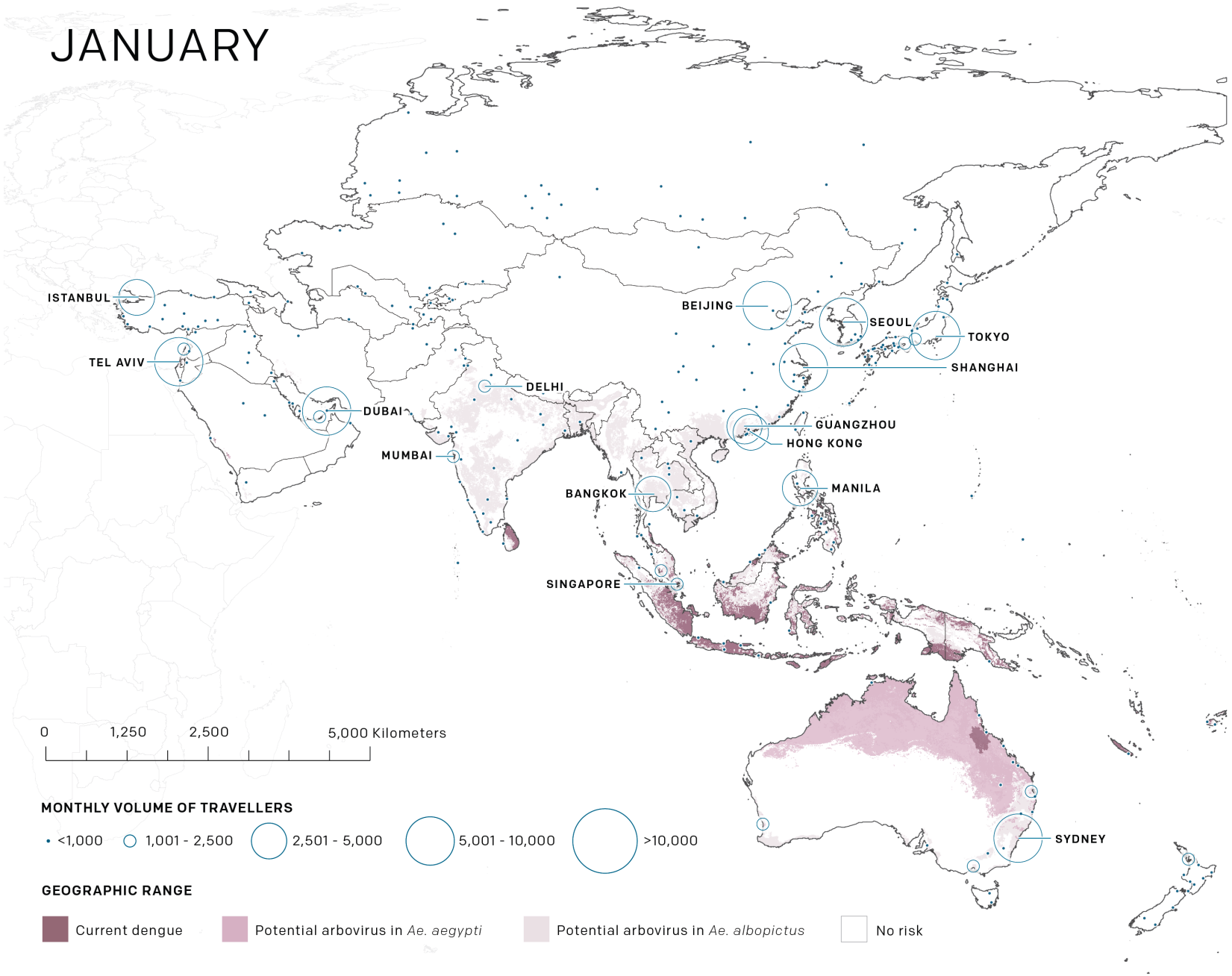
22

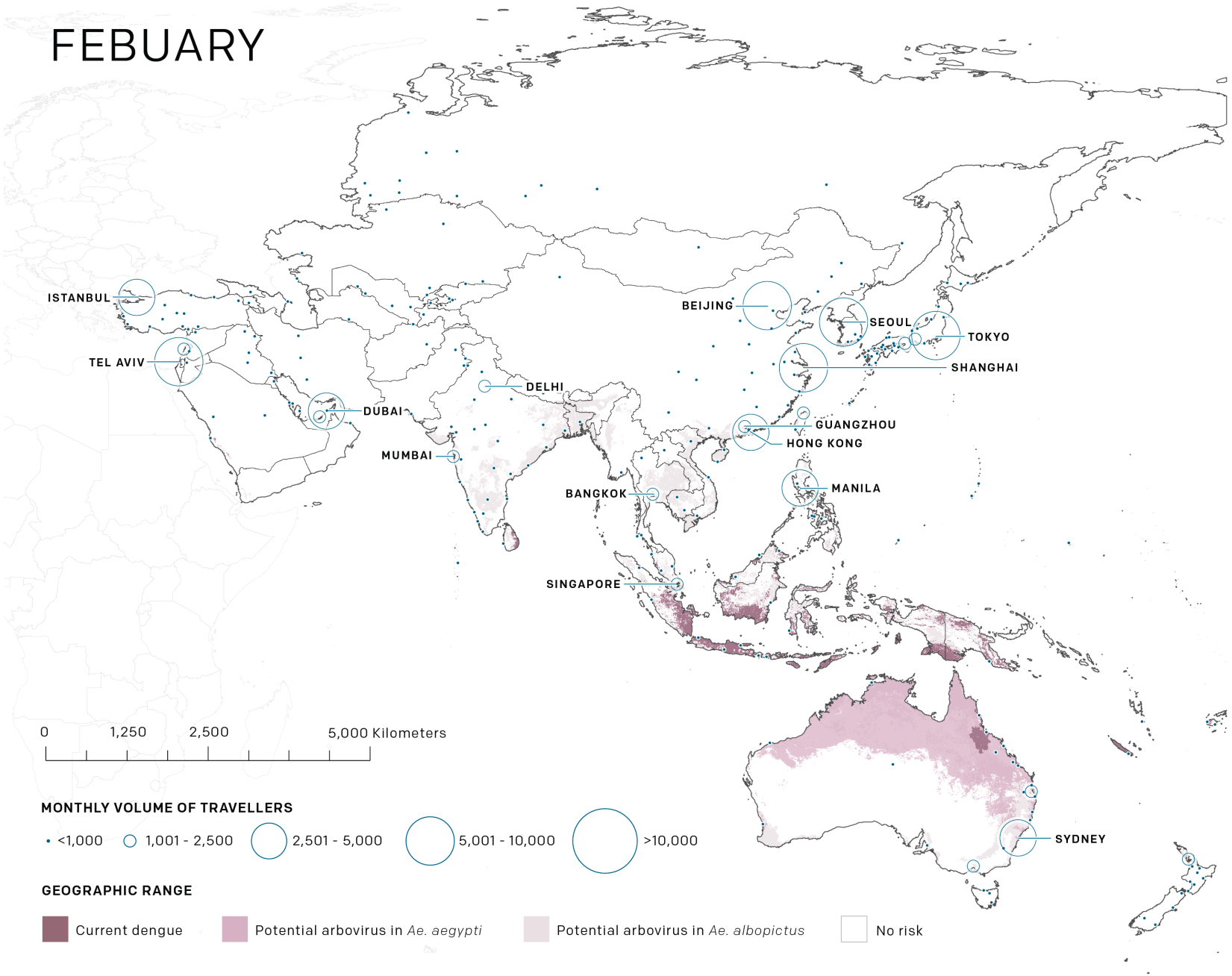

NOVEMBER
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

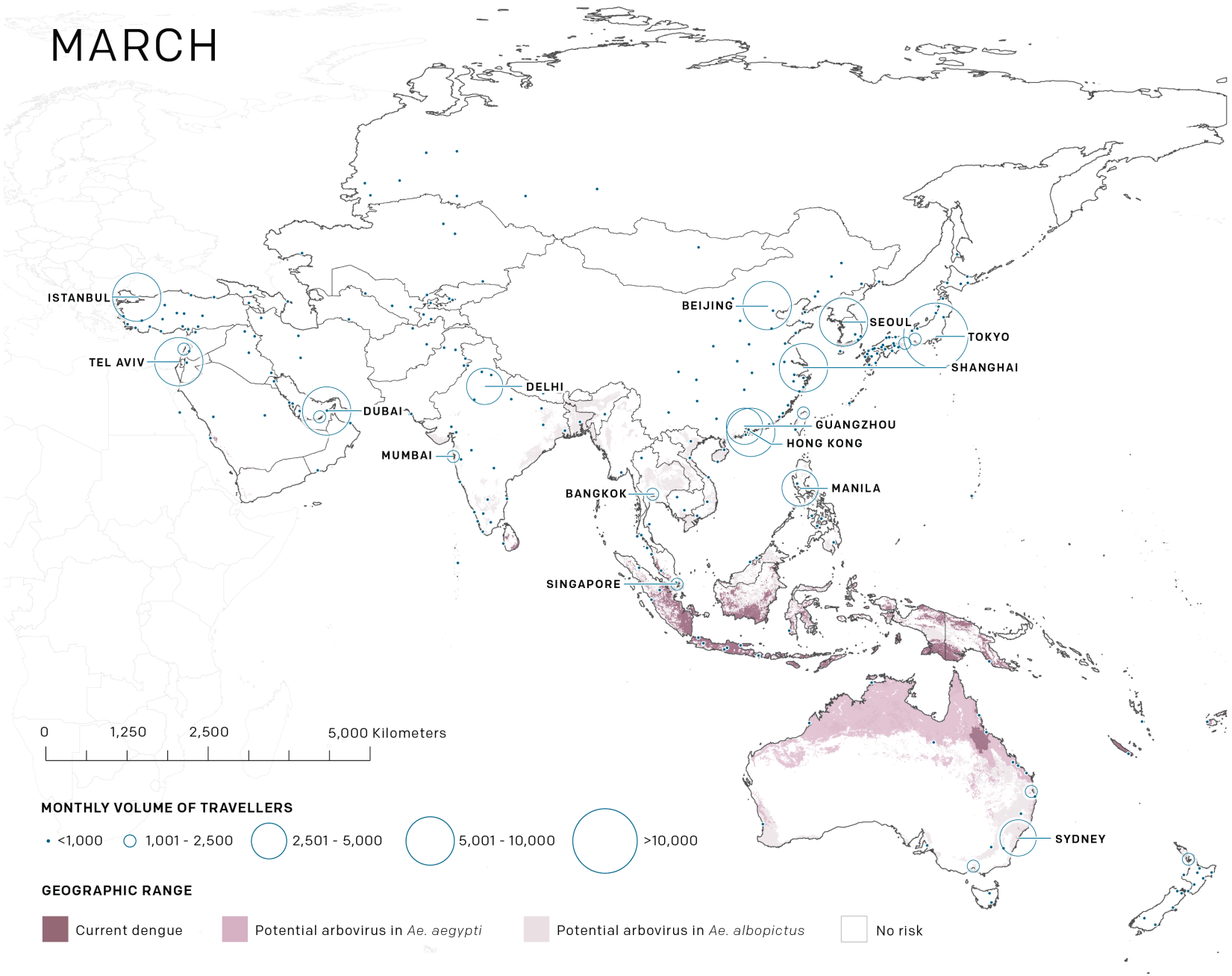
23

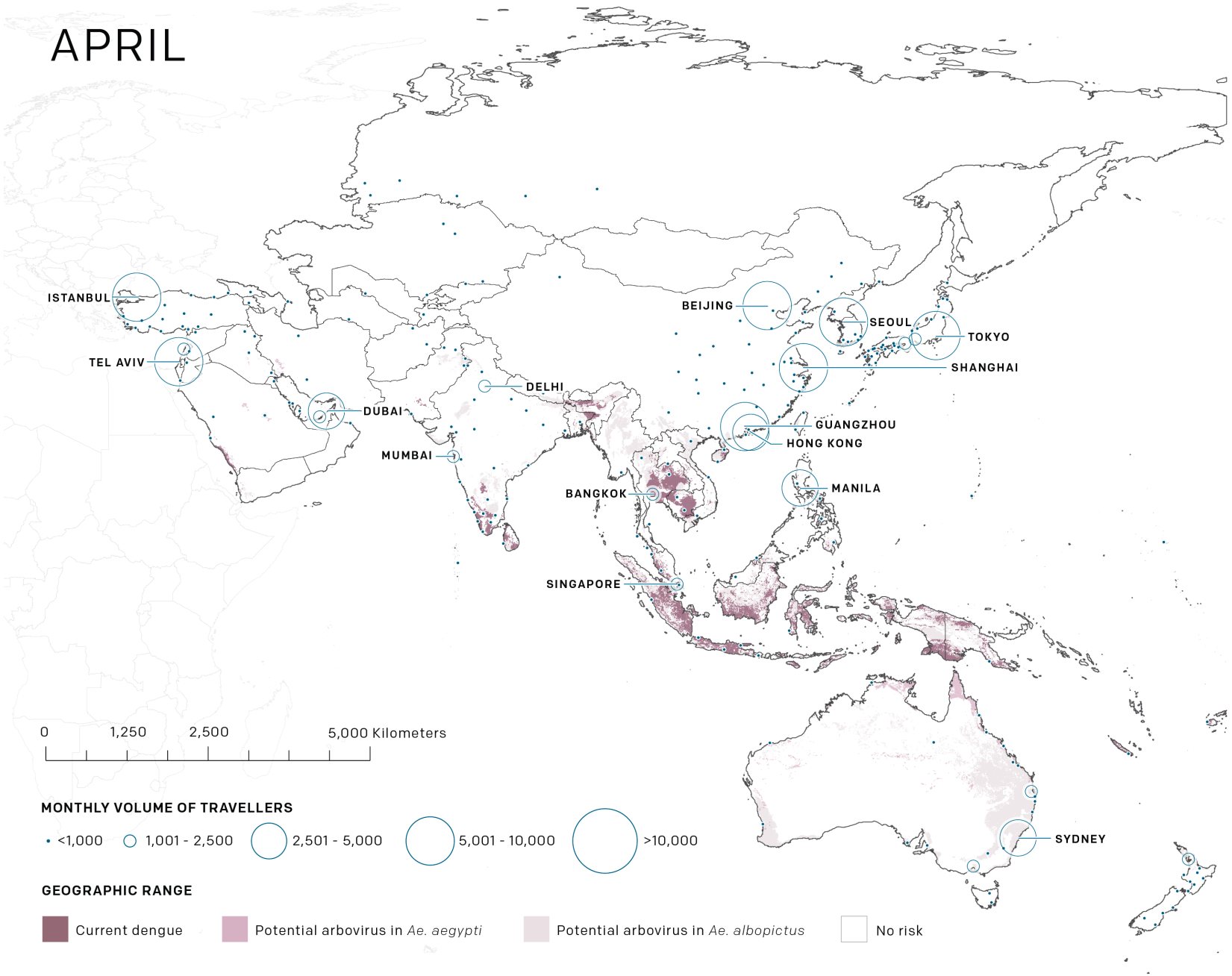

DECEMBER
0
1,250
2,500
5,000 Kilometers
LUANDA
JOHANNESBURG
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
RANGE OF ACTIVE VIRUS TRANSMISSION
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

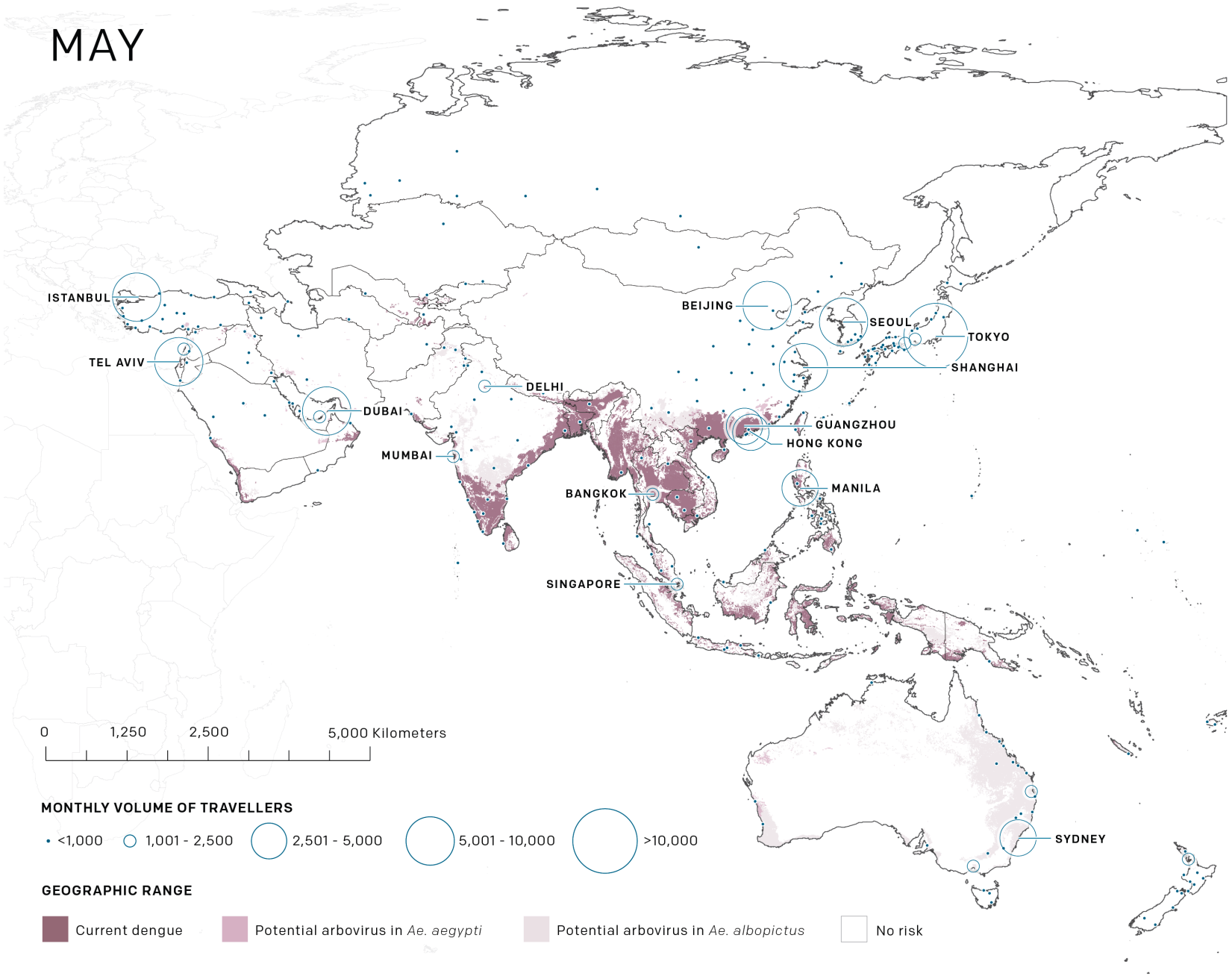
Appendix 3: Monthly Asia-Pacific Model Outputs

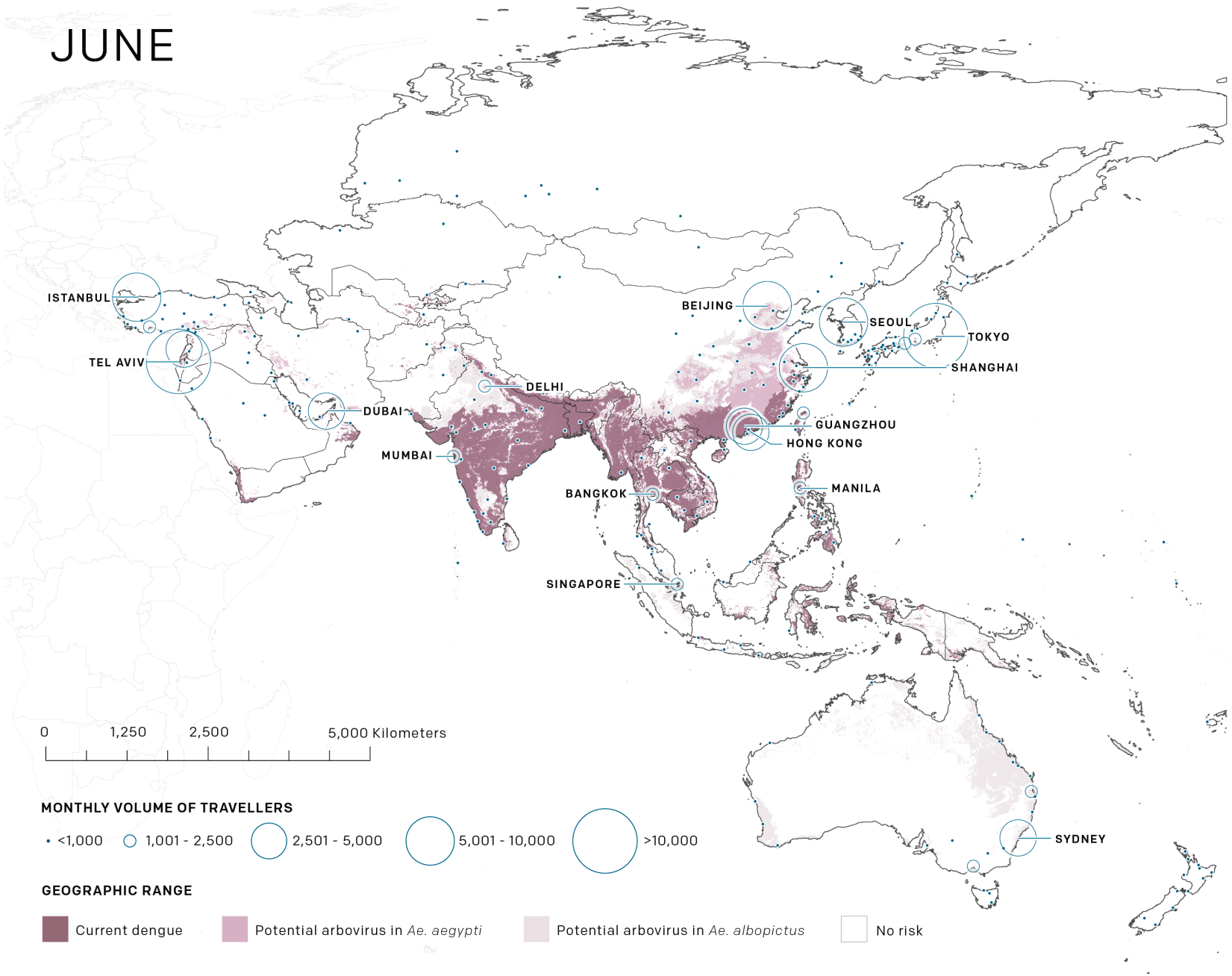

JANUARY
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

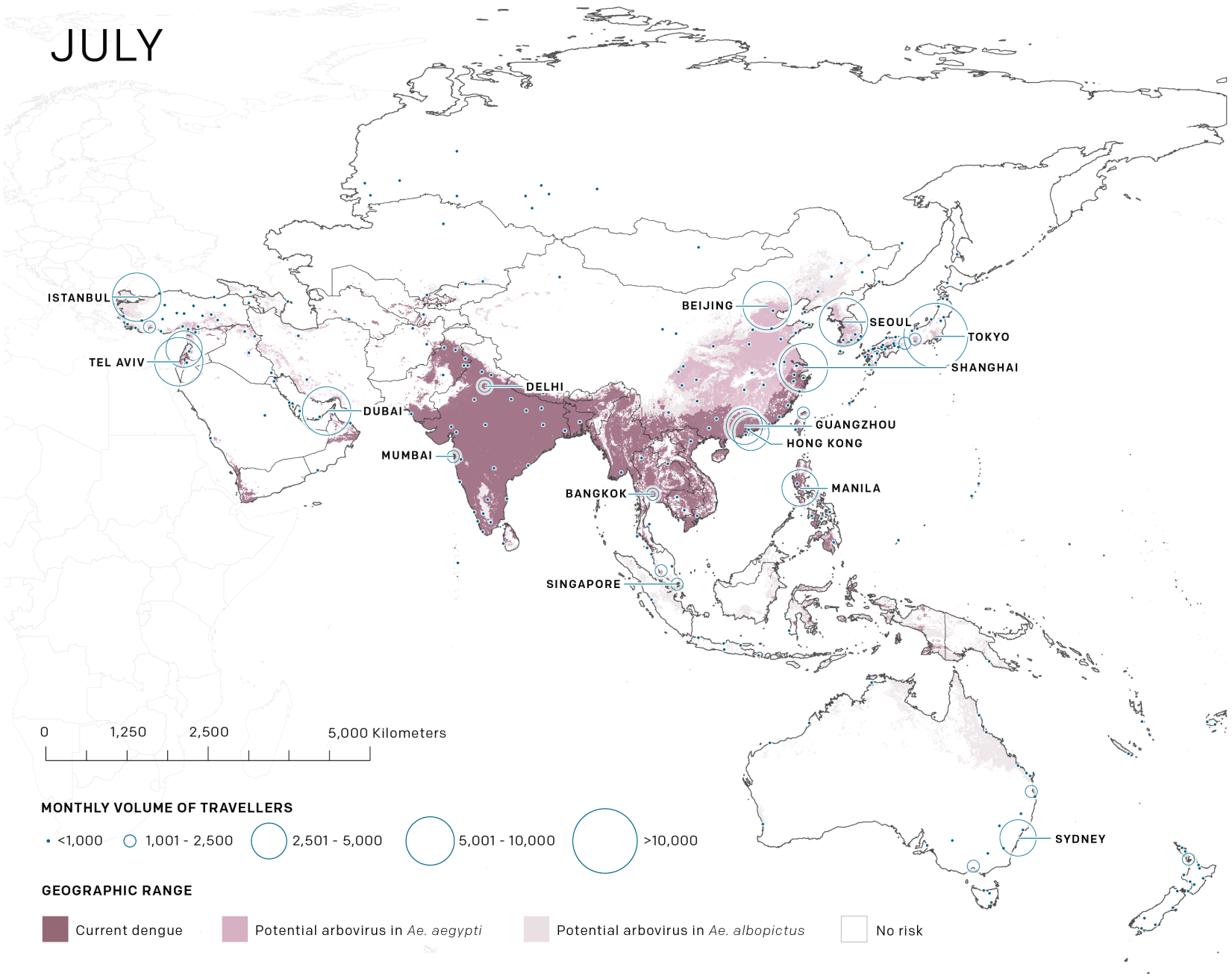

FEBUARY
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

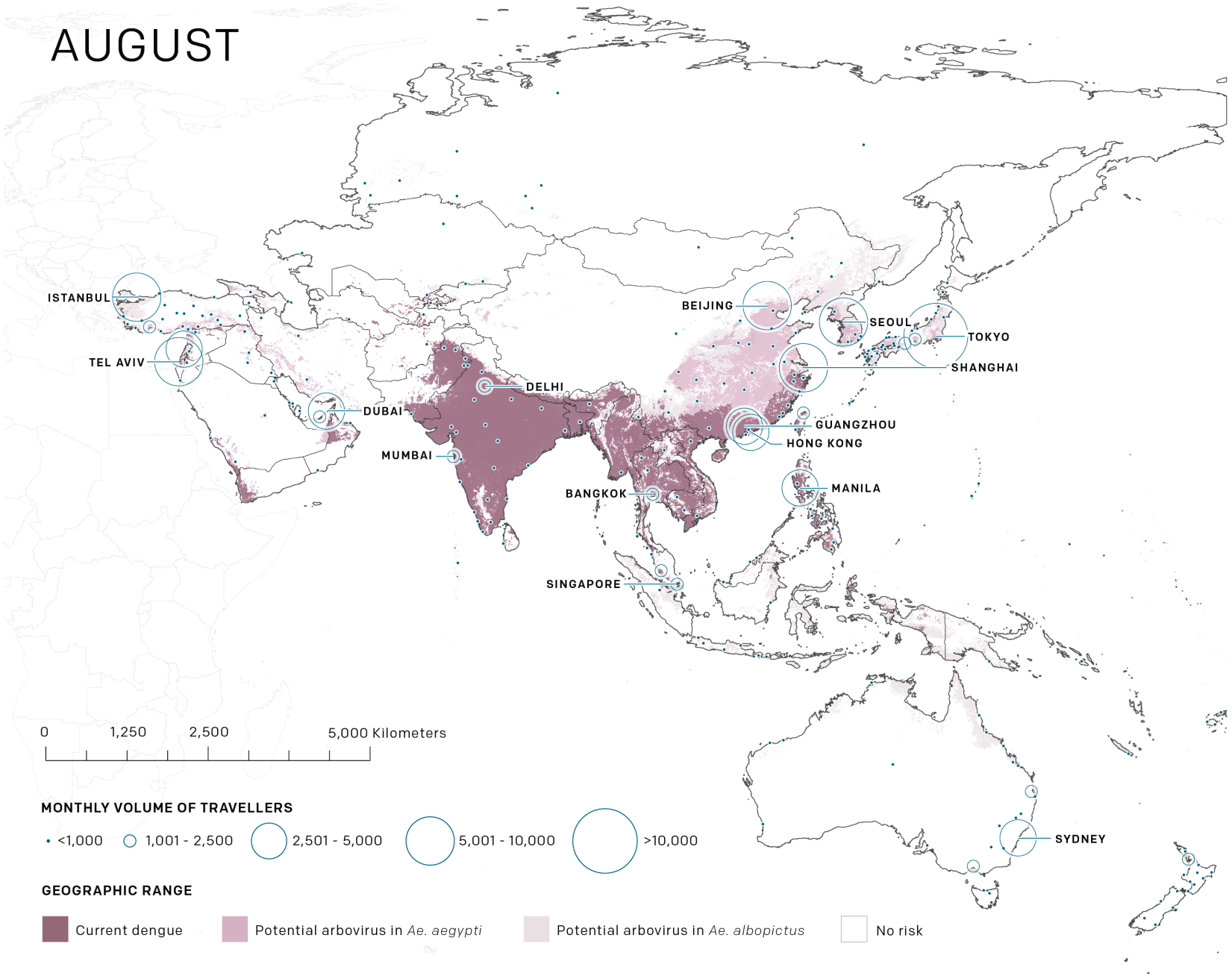

MARCH
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

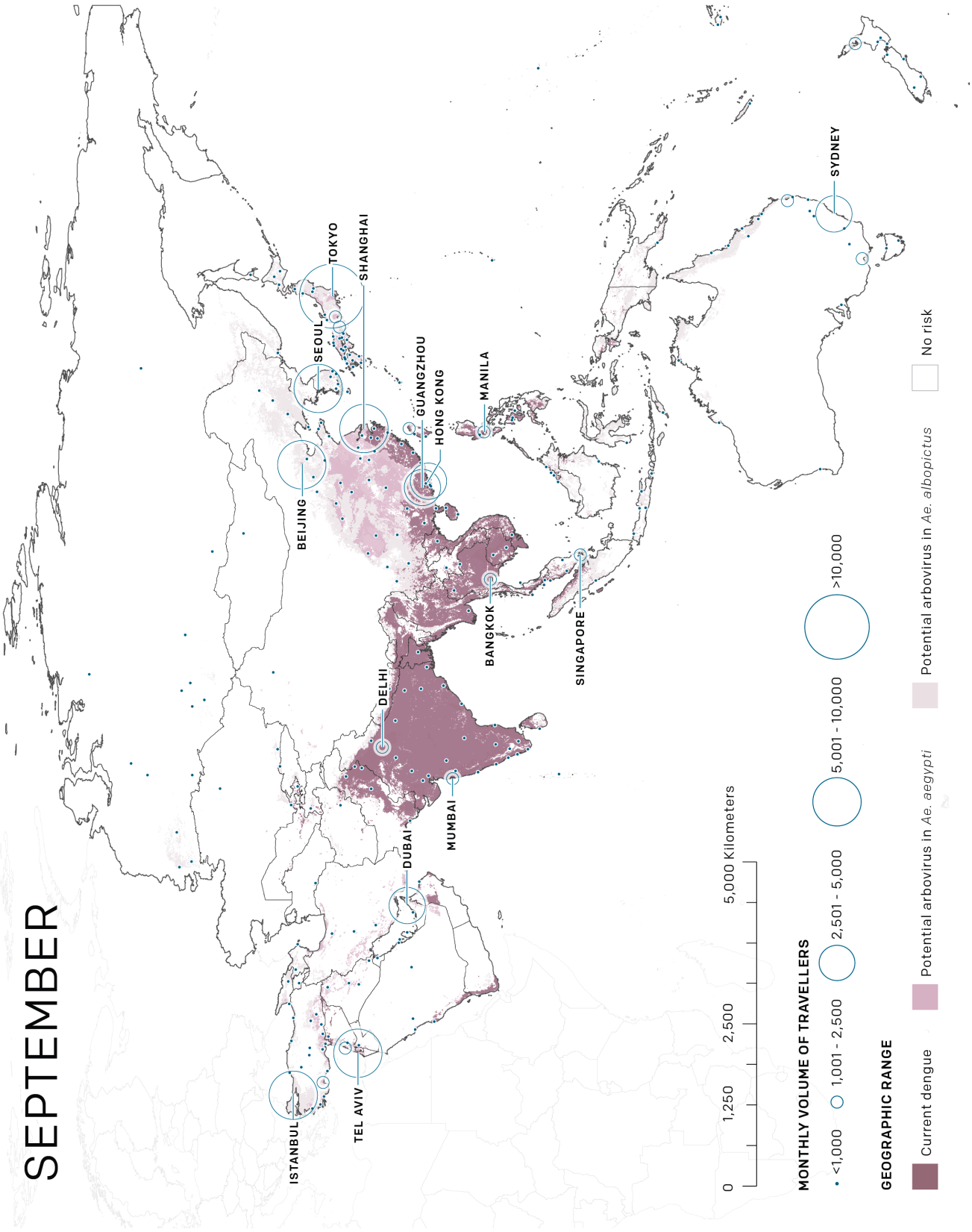

APRIL
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

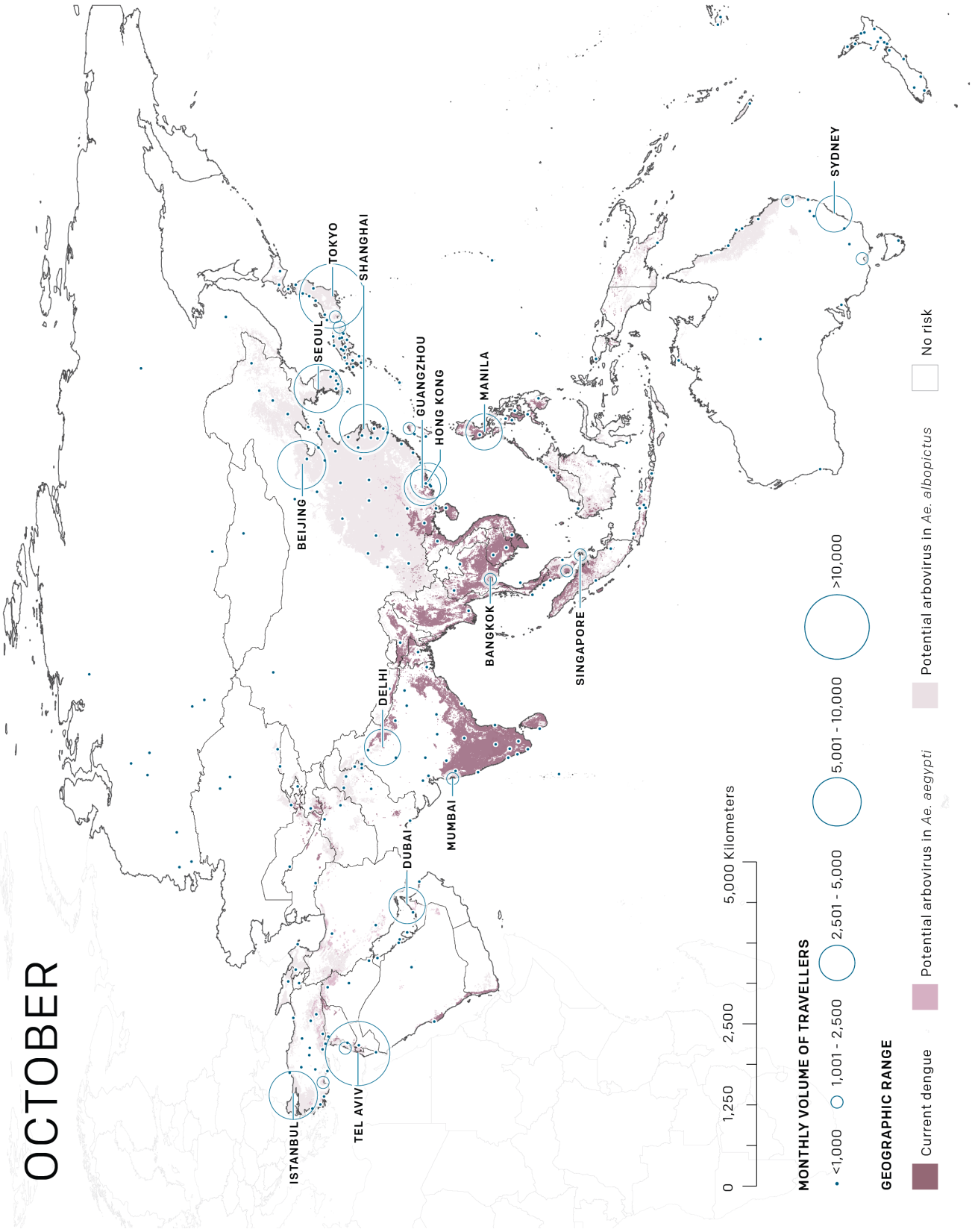

MAY
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

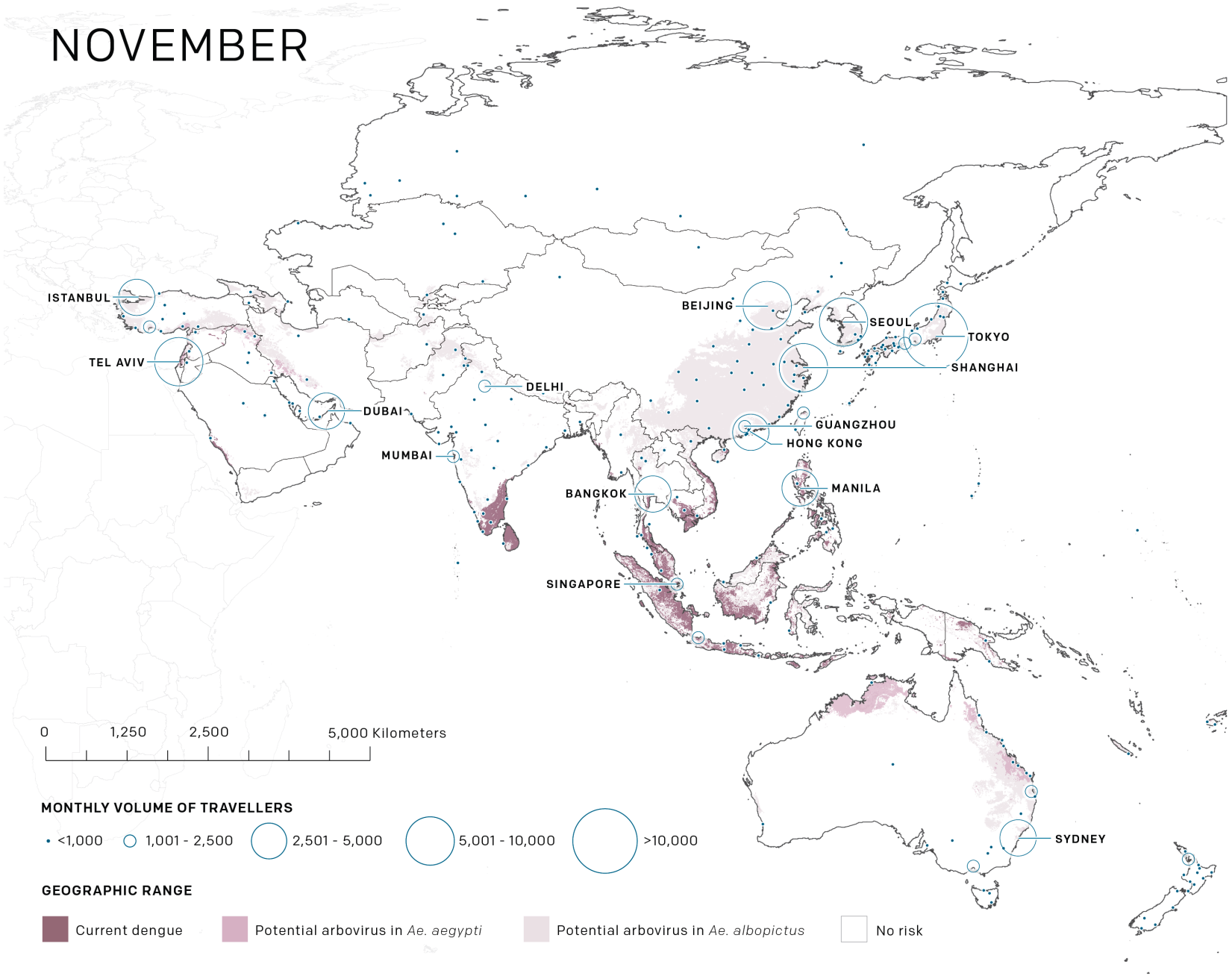

JUNE
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

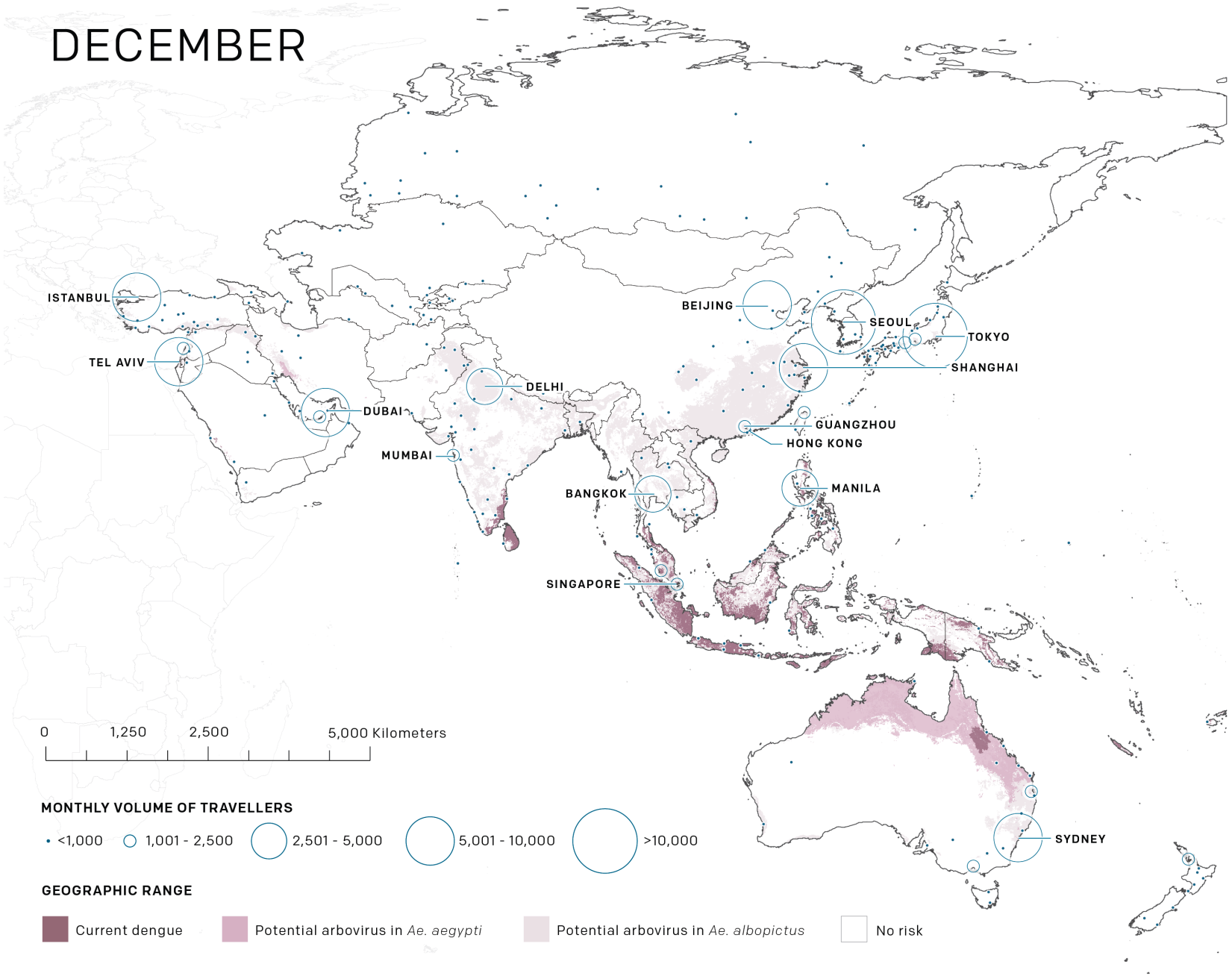

JULY
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


AUGUST
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BANGKOK
SINGAPORE
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


SEPTEMBER
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


OCTOBER
ISTANBUL
TEL AVIV
DUBAI
MUMBAI
DELHI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


NOVEMBER
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk


DECEMBER
ISTANBUL
TEL AVIV
DUBAI
DELHI
MUMBAI
BEIJING
SEOUL
TOKYO
SHANGHAI
GUANGZHOU
HONG KONG
MANILA
BANGKOK
SINGAPORE
SYDNEY
0
1,250
2,500
5,000 Kilometers
MONTHLY VOLUME OF TRAVELLERS
<1,000
1,001 - 2,500
2,501 - 5,000
5,001 - 10,000
>10,000
GEOGRAPHIC RANGE
Current dengue
Potential arbovirus in Ae. aegypti
Potential arbovirus in Ae. albopictus
No risk

A.3.2 Appendix to Chapter 3.2

Figure 1 - figure supplement 1: Maps of all covariates entered into the 300 BRT models: (a) probability of being urban, 2015; (b) enhanced vegetation index; (c) minimum relative humidity; (d) cumulative annual precipitation (mm); (e) temperature suitability for dengue *via Ae. aegypti*; (f) temperature suitability for dengue *via Ae. albopictus.*

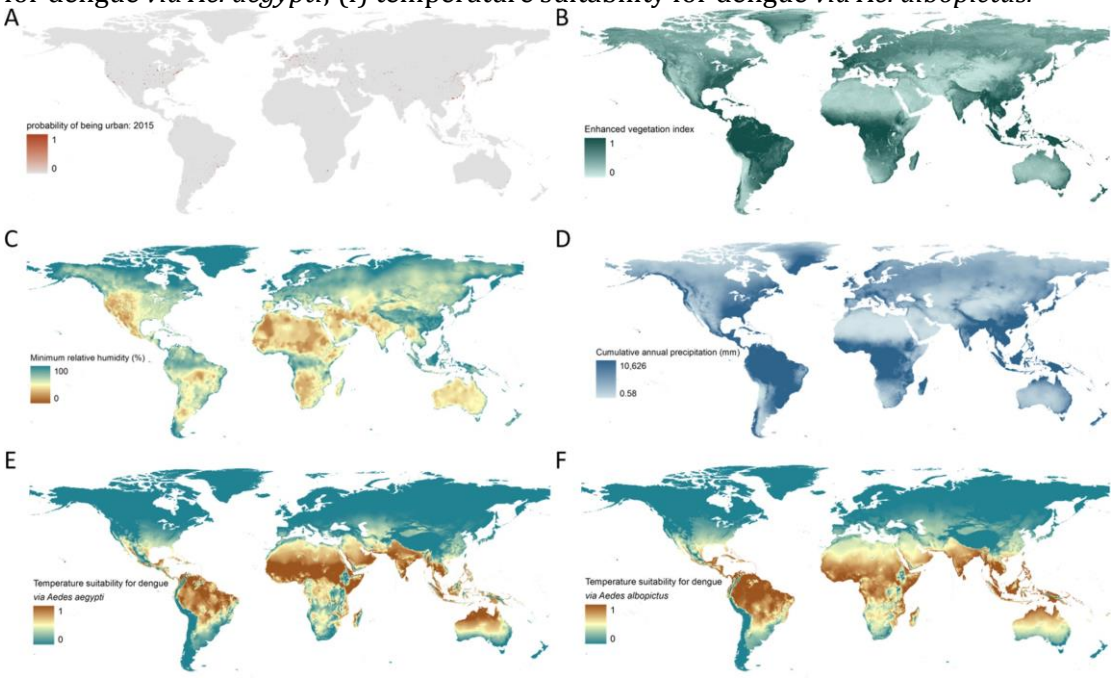


Figure 2 - figure supplement 1: Uncertainty around Zika suitability predictions displayed in main manuscript – Figure 2, ranging from less than 0.01 (very little uncertainty) to 0.94 (greatest uncertainty).

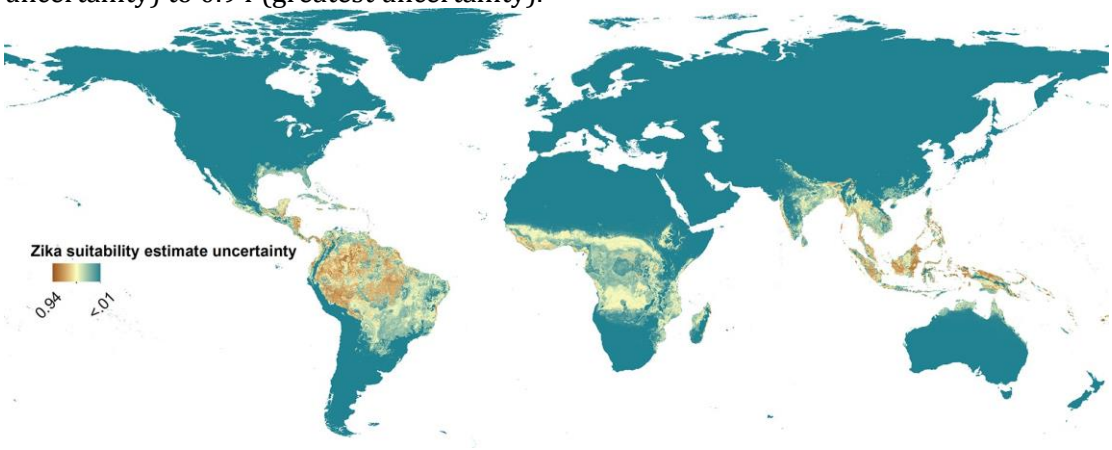


Figure 2 - figure supplement 2: Effect plots for each covariate entered into the ensemble of 300 BRT models: (a) minimum relative humidity; (b) cumulative annual precipitation (mm); (c) enhanced vegetation index; (d) probability of being urban (%); (e) temperature suitability for dengue *via Ae. aegypti*; (f) temperature suitability for dengue *via Ae. albopictus.*

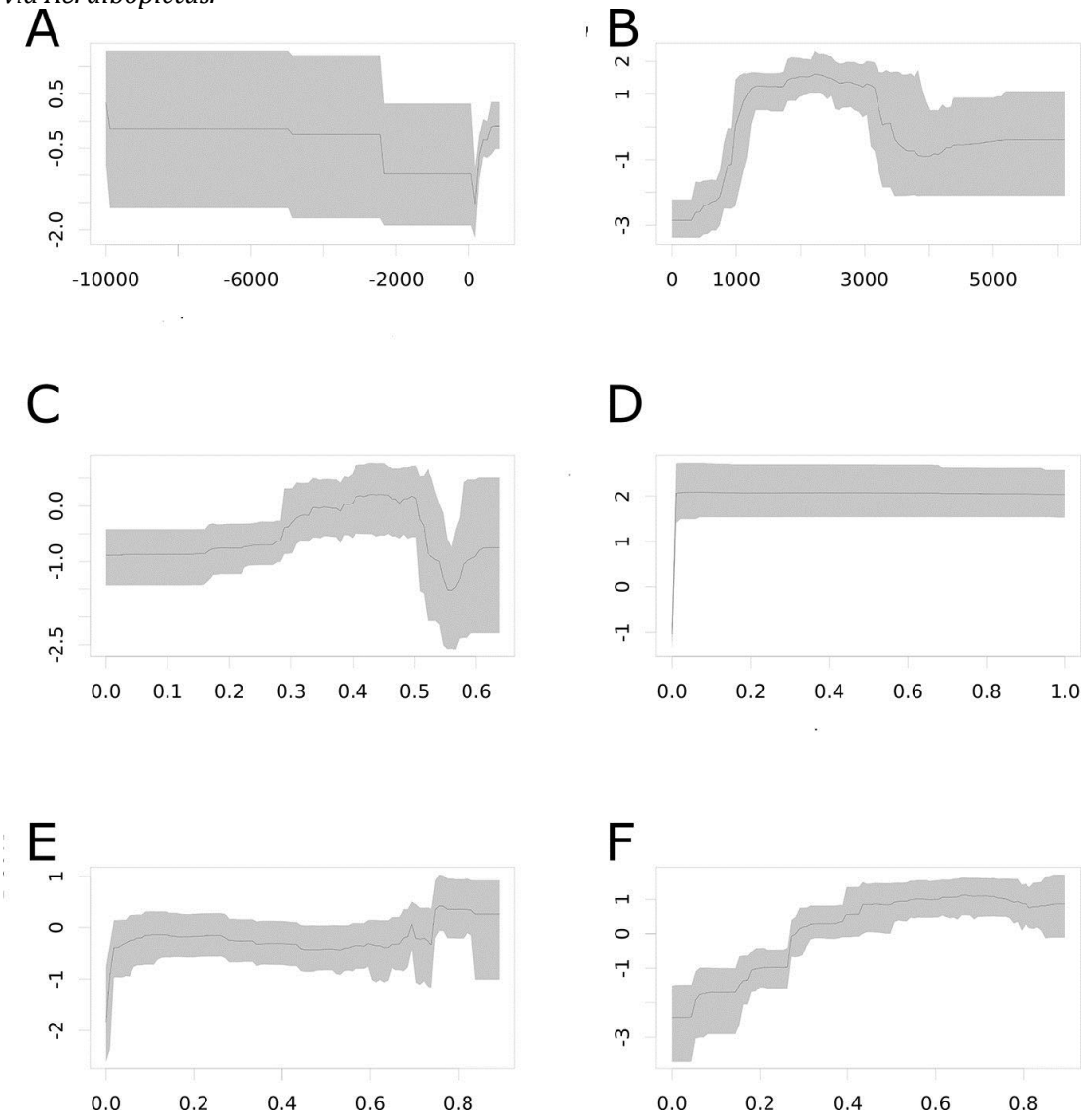


Figure 2 - figure supplement 3: Environmental suitability for Zika virus transmission to humans, not taking into account temperature suitability for dengue *via Aedes albopictus*. Covariate effects are as follows: cumulative annual precipitation (67.4%); temperature suitability for dengue *via Ae.* aegypti (16.9%); probability of being urban, 2015 (8.2%); enhanced vegetation index (5.1%); minimum relative humidity (2.4%).

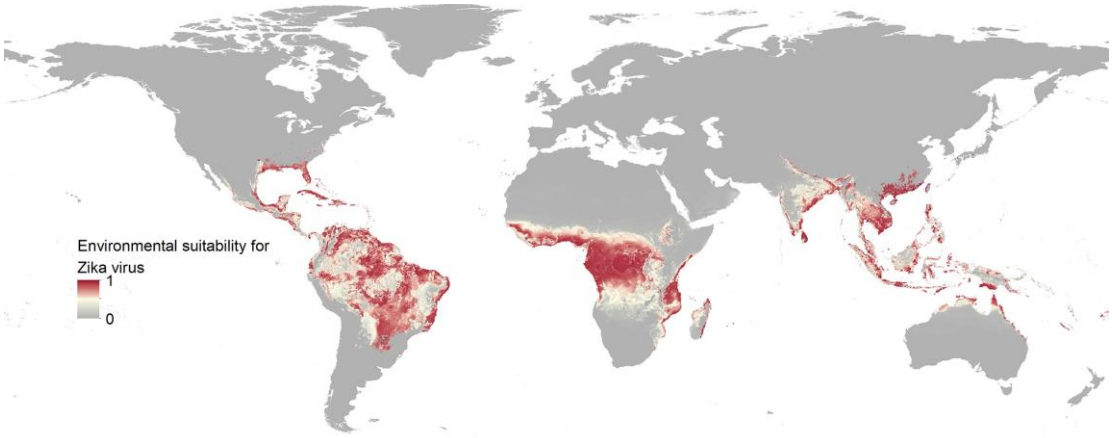
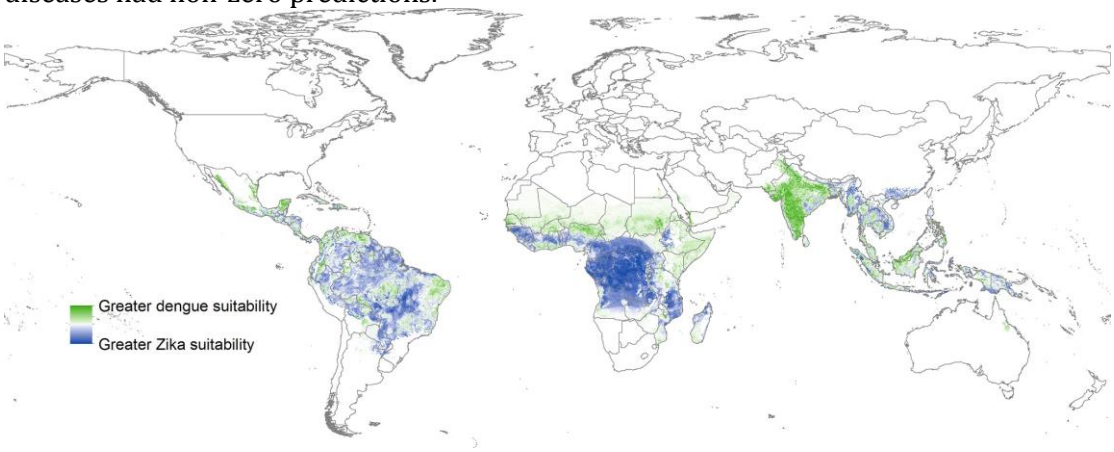


Figure 2 - figure supplement 4: Map showing areas predicted to have greater dengue suitability (from Bhatt *et al.* 2013, Nature) vs. those which are predicted to have greater Zika suitability in the current study. These values are restricted to areas where both diseases had non-zero predictions.



A.4. Appendix to Chapter 4

Table S1: Reported case numbers per year for each district of the study region, Punjab Province, Pakistan.

District	2011	2012	2013	2014	Total
Attock	7	1	0	18	26
Bagh	0	0	0	0	0
Bahawalnagar	65	0	0	0	65
Bahawalpur	65	1	40	3	109
Bhakkar	33	0	0	0	33
Bhimber	0	0	0	0	0
Chakwal	0	0	0	0	0
D. G. Khan	3	0	0	0	3
Faisalabad	981	58	125	0	1164
Gujranwala	61	0	0	0	61
Gujrat	0	24	0	0	24
Hafizabad	1	0	0	0	1
Jhang	6	0	0	0	6
Jhelum	75	0	0	33	108
Kasur	77	0	0	0	77
Khanewal	2	0	0	1	3
Khushab	37	0	0	0	37
Kotli	0	0	0	0	0
Lahore	18020	4013	11516	1799	35348
Leiah	27	1	0	0	28
Lodhran	0	0	0	0	0
Mandi Bahauddi	4	0	3	0	7
Mianwali	6	0	0	0	6
Mirpur	0	0	0	0	0
Multan	371	1	64	4	440
Muzaffarabad	0	0	0	0	0
Muzaffargarh	0	0	0	0	0
Narowal	26	2	0	0	28
Neelum	0	0	0	0	0
Okara	126	3	0	0	129
Pakpattan	179	5	0	0	184
Poonch	0	0	0	0	0
Rahim Yar Khan	0	0	0	0	0
Rajanpur	0	0	0	0	0
Rawalpindi	736	132	1845	2532	5245
Sahiwal	69	4	3	0	76
Sargodha	14	0	0	0	14
Sheikhupura	29	3	42	179	253
Sialkot	49	0	0	0	49
Sudhnoti	0	0	0	0	0
Toba Tek Singh	7	1	6	2	16
Vehari	106	264	228	18	616

Table S2: Environmental model: no variation in α.

Term	Estimate	Std. Error	t-value	p-value
Intercept	**2.51**	**0.679**	**3.696**	**0.00028**
α	**0.690**	**0.040**	**17.288**	**< 2e-16**
EVI (yearly average)	**-8.78**	**2.499**	**-3.392**	**0.00082**
Ae. albopictus	**2.28**	**0.544**	**4.196**	**3.93e-5**
Precipitation (yearly average)	**-0.021**	**0.0079**	**-2.590**	**0.0102**
Term	**edf**	**Ref. df**	**F**	**p-value**
"Seasonality"	**3.44**	**8.00**	**1.566**	**0.0034**
Temperature	**7.55**	**8.47**	**2.537**	**0.0102**
2-week lagged temperature	**5.47**	**6.67**	**2.300**	**0.0304**
EVI	**1.83**	**2.33**	**3.373**	**0.0299**

Table S3: Environmental model: α in Lahore differs from other districts.

Term	Estimate	Std. Error	t-value	p-value
Intercept	**2.548**	**0.676**	**3.769**	**0.0002**
α (Lahore)	**0.741**	**0.043**	**17.088**	**<2e-16**
α (Not Lahore)	**0.594**	**0.055**	**10.887**	**<2e-16**
EVI (yearly average)	**-7.636**	**2.527**	**-3.022**	**0.003**
Ae. albopictus	**1.159**	**0.679**	**1.708**	**0.089**
Precipitation (yearly average)	**-0.008**	**0.009**	**-0.870**	**0.385**
Term	**edf**	**Ref. df**	**F**	**p-value**
"Seasonality"	**3.238**	**8.000**	**1.278**	**0.0095**
Temperature	**7.502**	**8.435**	**3.195**	**0.0016**
2-week lagged temperature	**5.866**	**7.067**	**2.371**	**0.0231**
EVI	**2.179**	**2.784**	**3.690**	**0.0154**

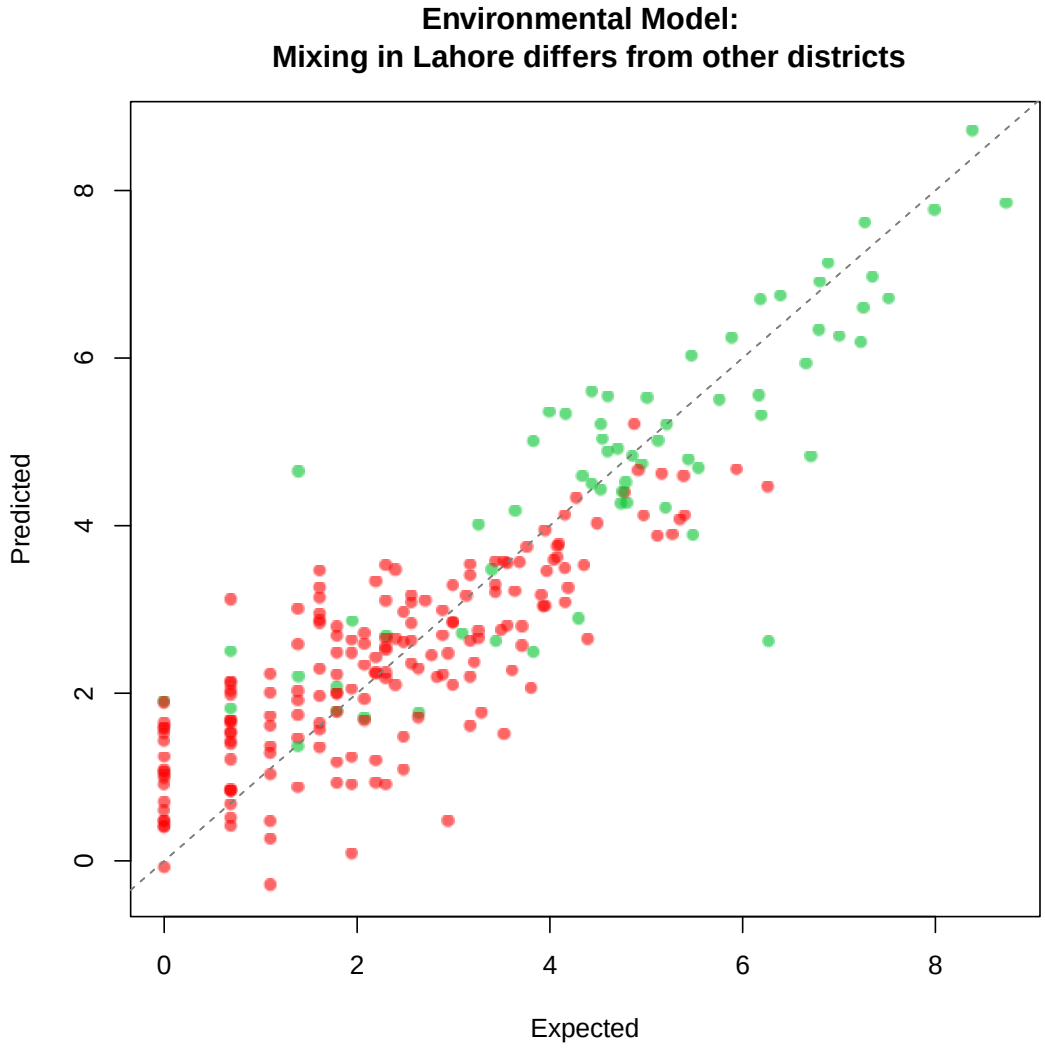
Table S4: Full model: no variation in α.

Term	Estimate	Std. Error	t-value	p-value
Intercept	**-34.83**	**9.803**	**-3.553**	**0.00047**
α	**0.578**	**0.0451**	**12.824**	**<2e-16**
EVI (yearly average)	**-60.12**	**15.04**	**-3.998**	**8.71e-5**
NDVI (yearly average)	**0.0383**	**0.0103**	**3.737**	**0.00024**
Ae. albopictus	**2.745**	**1.028**	**2.669**	**0.00826**
Weighted Access	**2.742e-5**	**6.464e-6**	**4.242**	**3.26e-5**
Urbanicity	**-2.468**	**0.9494**	**-2.600**	**0.00776**
Precipitation (yearly average)	**-0.0684**	**0.0185**	**-3.697**	**0.00028**
Term	**edf**	**Ref. df**	**F**	**p-value**
"Seasonality"	**4.480**	**8.000**	**4.173**	**3.97e-7**
Temperature	**7.607**	**8.504**	**4.396**	**4.14e-5**
2-week lagged temperature	**4.823**	**5.981**	**3.793**	**0.0013**

Table S5: Full model: α in Lahore differs from other districts.

Term	Estimate	Std. Error	t-value	p-value
Intercept	**-35.92**	**9.93**	**-3.618**	**0.00037**
α (Lahore)	**0.612**	**0.0643**	**8.510**	**<2e-16**
α (Not Lahore)	**0.555**	**0.0566**	**9.802**	**<2e-16**
EVI (yearly average)	**-61.59**	**15.19**	**-4.055**	**6.95e-5**
NDVI (yearly average)	**0.0395**	**0.010**	**3.8**	**0.00019**
Ae. albopictus	**2.81**	**1.033**	**2.721**	**0.0070**
Weighted Access	**2.42e-5**	**7.86e-6**	**3.072**	**0.0024**
Urbanicity	**-2.33**	**0.9697**	**-2.403**	**0.0171**
Precipitation (yearly average)	**-0.069**	**0.0186**	**-3.733**	**0.00024**
Term	**edf**	**Ref. df**	**F**	**p-value**
"Seasonality"	**4.350**	**8.000**	**3.817**	**1.31e-6**
Temperature	**7.620**	**8.511**	**4.335**	**4.97e-5**
2-week lagged temperature	**4.764**	**5.916**	**3.680**	**0.00178**

Figure S1: Predicted versus expected values for α in Lahore (green) and all other districts (red) for the environmental model.



Additional information about collection of epidemiological data.

Secondary data from hospital records were used from Punjab province in Pakistan. The data was initially collected by Punjab Health Department as part of the dengue prevention and eradication program. For ensuring accurate reporting from the health facilities, Punjab Health Department used the following three procedures: (i) Clinical case reporting, (ii) Lab case reporting, and (iii) Case management. All health facilities were liable to record the data and share them with Punjab Information Technology Board (PITB) within 24 hours. For clinical case reporting, as per Dengue Expert Advisory Group (DEAG), guidelines, dengue suspects, probable, and confirmed cases needed to be correctly entered on the PITB dashboard within 24 hours of admission. For lab case reporting, all private sector labs must send reports of positive dengue cases in the line list format to the respective Executive District Officer Health (EDOH) for online entry on the dashboard, again within 24 hours. For case management, the healthcare facilities are liable to mange dengue cases regularly in specified Dengue care units, OPDs, emergency units, or wards.

Video S1: Video showing the spatial and temporal dynamics of dengue in the study area across Pakistan.

A.5. Appendix to Chapter 5

Rationale for mobility approximation:

We acknowledge the possible misrepresentation of human movement models fitted from European cell phone data in regards to the local regional movement patterns in West Africa. To address this challenge we tested the ability of fitted human movement matrices from Europe to predict human migration from IPUMS. We confirm that fitted human movement models from short term human movement data can capture longer term relative migration even in the absence of local data (Figure S7) [1] and can be extended to other regions in outbreak situations. In addition we fitted a variety of movement models against the datasets to capture the variation in the movement patterns, so they do not directly translate into the patterns described in the actual flow matrices but are smoothed by information about the underlying population and distance structures in West Africa [2]. We also used a combination of adjacency matrices and movement matrices to try to smooth the effect of human mobility alone (Figure 1) and by using a variable selection method in the TSIR modelling framework they were only selected if an improvement of the predictive power (*i.e.,* cases next week) could be obtained [3].

Calculation of relative contribution of transmission:

Illustrated in a simple example with four districts and three covariates:

$$X_{i1} = \sum_{\substack{j=1 \\ j \neq i}}^{4} \rho_{ij} * \frac{x_j}{N_j} \quad (1)$$

$$X_{i2} = \sum_{\substack{j=1 \\ j \neq i}}^{4} \varphi_{ij} * \frac{x_j}{N_j} \quad (2)$$

$$X_{i3} = \sum_{\substack{j=1 \\ j \neq i}}^{4} \eta_{ij} * \frac{x_j}{N_j} \quad (3)$$

where $\rho, \varphi \; and \; \eta$ are connectivity matrices, and x_j and N_j are the number of cases and the total number of people respectively in district *j.* Equation (5) in the main text can be rewritten for the first district can as:

$$\log \beta_1 = S_1 * X_{11} + S_2 * X_{12} + S_3 * X_{13} \quad (4)$$

$$= S_1 \left(\sum_{j=2}^{4} \rho_{1j} * \frac{x_j}{N_j} \right) + S_2 \left(\sum_{j=2}^{4} \varphi_{1j} * \frac{x_j}{N_j} \right) + S_3 \left(\sum_{j=2}^{4} \eta_{1j} * \frac{x_j}{N_j} \right) (5)$$

$$= \sum_{j=2}^{4} \frac{x_j}{N_j} \left(S_1 \rho_{1j} + S_2 \varphi_{1j} + S_3 \eta_{1j} \right) (6)$$

Thus, district j contributes $\frac{x_j}{N_j}(S_1\rho_{1j} + S_2\varphi_{1j} + S_3\eta_{1j})$ to district 1's $\log \beta_1$. Finally, we can add up all the contributions that district *j* makes as:

$$C_j = \frac{x_j}{N_j}\sum_{\substack{i=1 \\ i\neq j}}^{4}(S_1\rho_{1j} + S_2\varphi_{1j} + S_2\eta_{1j}) \ (7)$$

For understanding the dynamics that evolved over the outbreak we computed the relative contribution (C_j) for the entire outbreak ($\overline{C_j}$), the first half ($\overline{C_j^I}$) of the outbreak (expanding phase) and the second half ($\overline{C_j^{II}}$):

$$\overline{C_j} = \frac{1}{83}\sum_{t=1}^{83} C_j\,(t) \ (8)$$

$$\overline{C_j^I} = \frac{1}{42}\sum_{t=1}^{42} C_j\,(t) \ (9)$$

$$\overline{C_j^{II}} = \frac{1}{41}\sum_{t=43}^{83} C_j\,(t) \ (10)$$

For each of the *j* = 63 district:

$$\overline{C} = \frac{1}{63}\sum_{j=1}^{63}\overline{C_j} \ (11)$$

$$\sigma(\overline{C}) = \sqrt{\frac{1}{63}\sum_{j=1}^{63}(\overline{C_j} - \overline{C})^2)} \ (12)$$

$$\overline{C^I} = \frac{1}{63}\sum_{j=1}^{63}\overline{C_j^I} \ (13)$$

$$\sigma\left(\overline{C^I}\right) = \sqrt{\frac{1}{63}\sum_{j=1}^{63}\left(\overline{C_j^I} - \overline{C^I}\right)^2}^{\ 2} \ (14)$$

$$\overline{C^{II}} = \frac{1}{63}\sum_{j=1}^{63}\overline{C_j^{II}} \ (15)$$

$$\sigma\left(\overline{C^{II}}\right) = \sqrt{\frac{1}{63}\sum_{j=1}^{63}\left(\overline{C_j^{II}} - \overline{C^{II}}\right)^2} \quad (16)$$

Finally, across all districts, standardized contributions for each district for the entire outbreak, the first half, the second half denoted as $c_j, c_j^I, and\ c_j^{II}$ respectively:

$$c_j = \frac{\overline{C_j} - \overline{C}}{\sigma(\overline{C})} \quad (17)$$

$$c_j^I = \frac{\overline{C_j^I} - \overline{C^I}}{\sigma(\overline{C^I})} \quad (18)$$

$$c_j^{II} = \frac{\overline{C_j^{II}} - \overline{C^{II}}}{\sigma(\overline{C^{II}})} \quad (19)$$

Country specific results for the second half of the outbreak:

Post week 42 transmission model by country: In Guinea, the base model (*i.e.*, covariate-free) fitted poorly (R^2-adjusted = 0.319) and the model whose structure was borrowed from the reduced model improved this fit (R^2-adjusted = 0.321). When starting with only the Guinea data and all the covariates and reducing that 'full' model, almost all covariates were eliminated leaving only the Gravity-France covariate (R^2-adjusted = 0.342). Using AIC, the base model beat the model whose structure was borrowed (647.13 and 655.60, respectively), and the Guinea-only reduced model was better but not strongly preferred (642.35).

In Liberia, the base model fitted poorly (R^2-adjusted = 0.416) and the model whose structure was borrowed from the full reduced model fitted significantly better and improved the percent of variation explained by the model (R^2-adjusted = 0.526). The Liberia-only reduced model further improved the fit (R^2-adjusted = 0.64). AIC showed clear separation from the three models with the Liberia-only reduced model being strongly preferred (AIC=173.37 versus 194.97 for the base model and 187.63 for the borrowed structural form model).

In Sierra Leone, the three models had comparable R^2-adjusted values (0.655 for the base model, 0.678 for the borrowed structure model, and 0.697 for the Sierra Leone-only reduced model). As with the Liberia models, AIC showed clear separation for the Sierra Leone-only reduced model being strongly preferred (AIC=585.57 versus 609.59 for the base model and 600.23 for the borrowed structural form model).

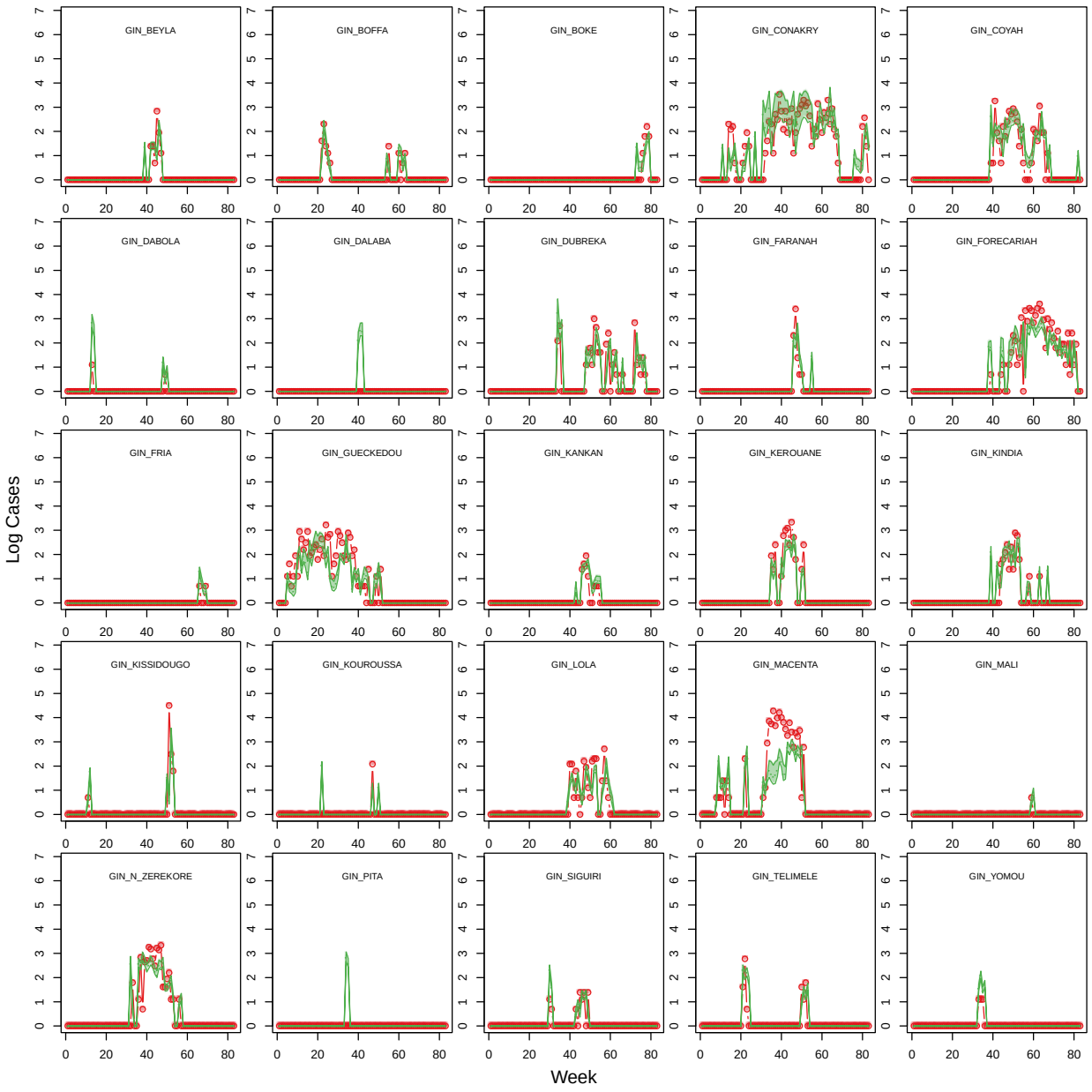


Figure S1: Results from the transmission model with a set of movement and adjacency matrices. Observed (red) versus predicted (green) case structure in Guinea.

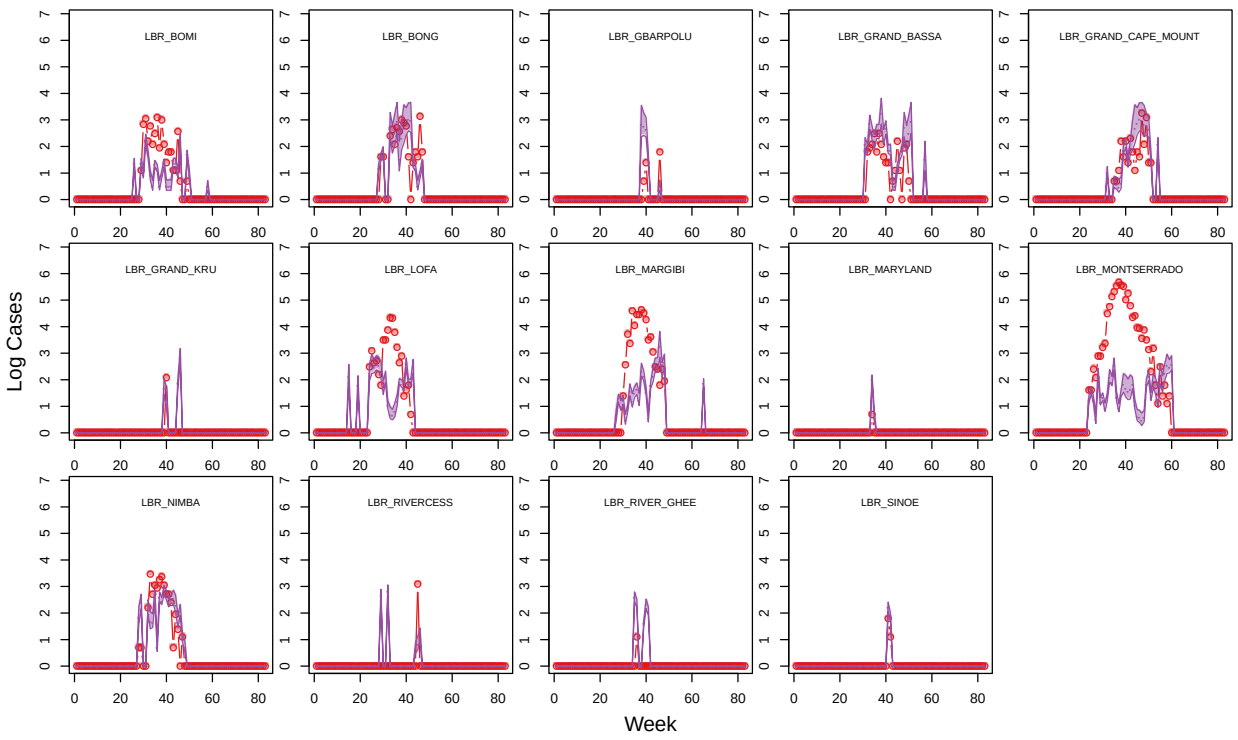


Figure S2: Results from the transmission model with a set of movement and adjacency matrices. Observed (red) versus predicted (purple) case structure in Liberia.

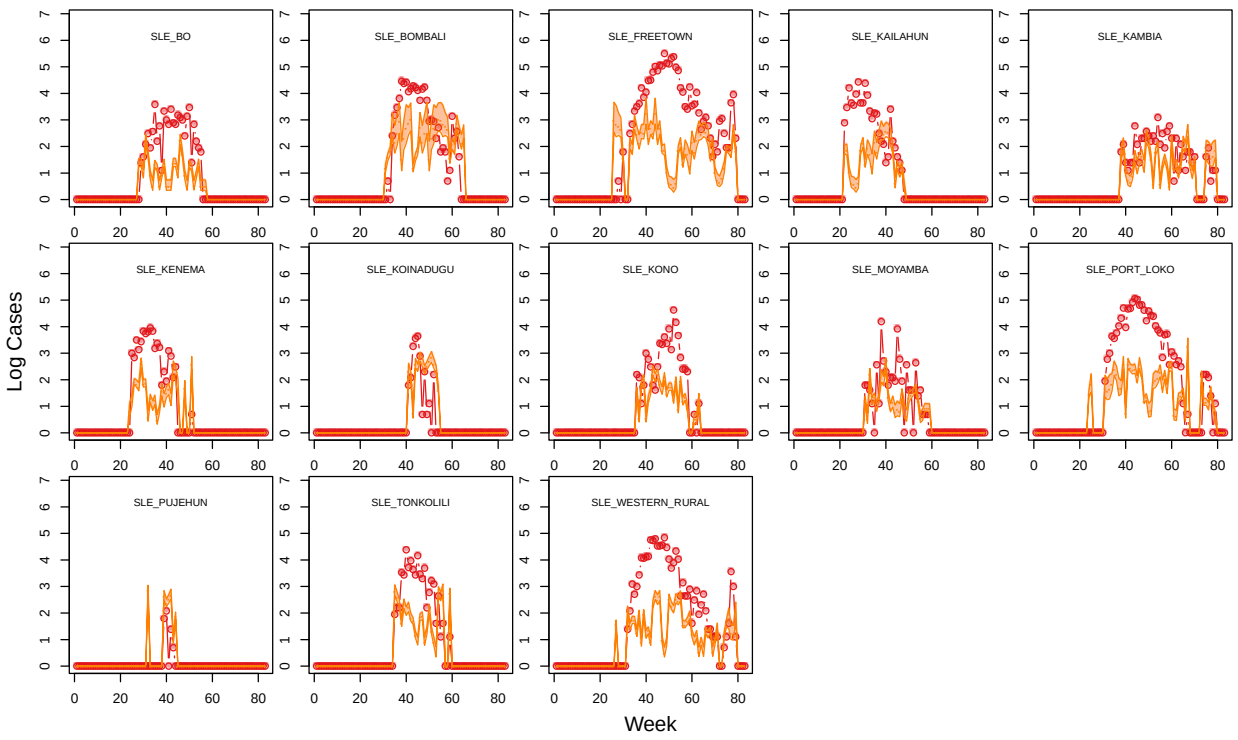


Figure S3: Results from the transmission model with a set of movement and adjacency matrices. Observed (red) versus predicted (orange) case structure in Sierra Leone.

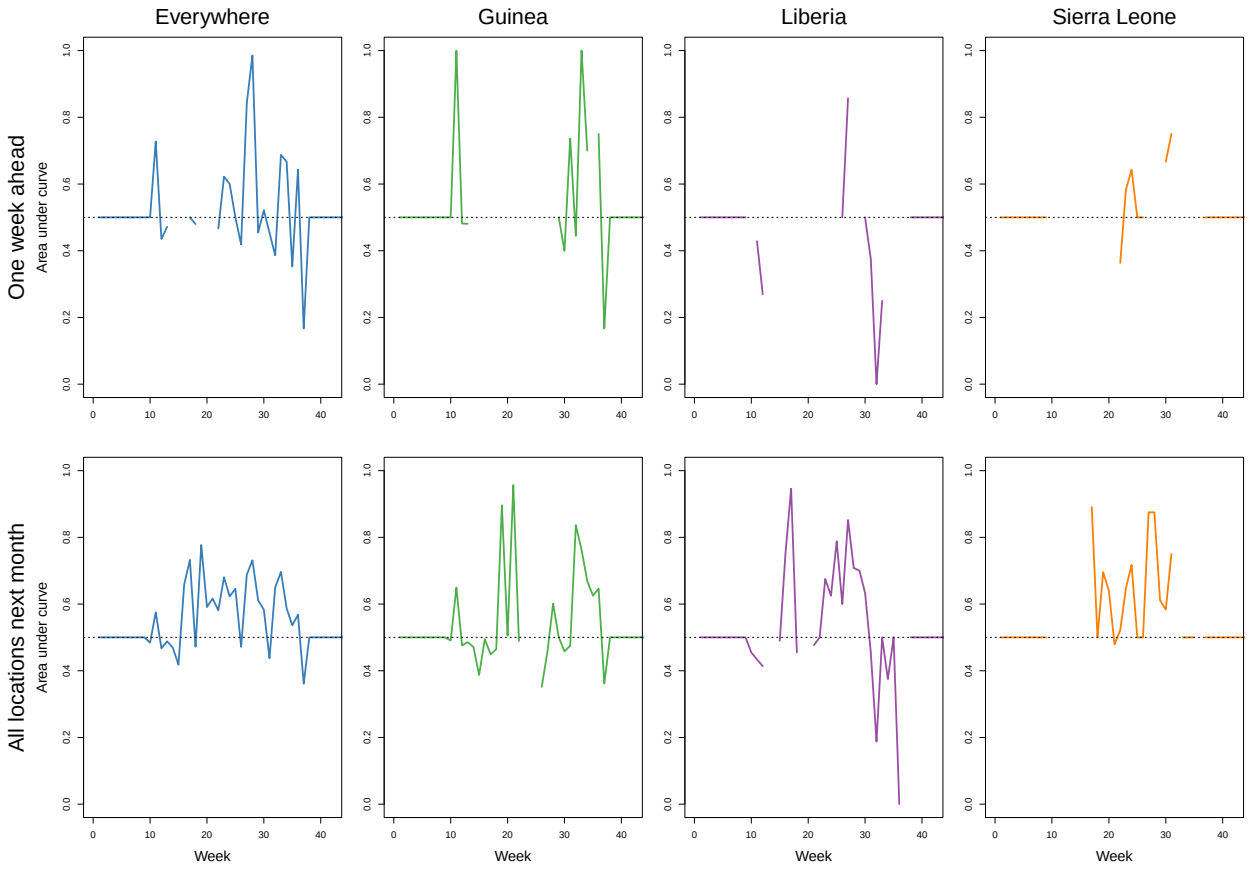


Figure S4: Results from the invasion model for one-week ahead predictions and next month predictions using AUC. 1 indicating that location of invasion (first recorded case) is predicted with high accuracy. No lines are shown for weeks/months where no invasion was observed.

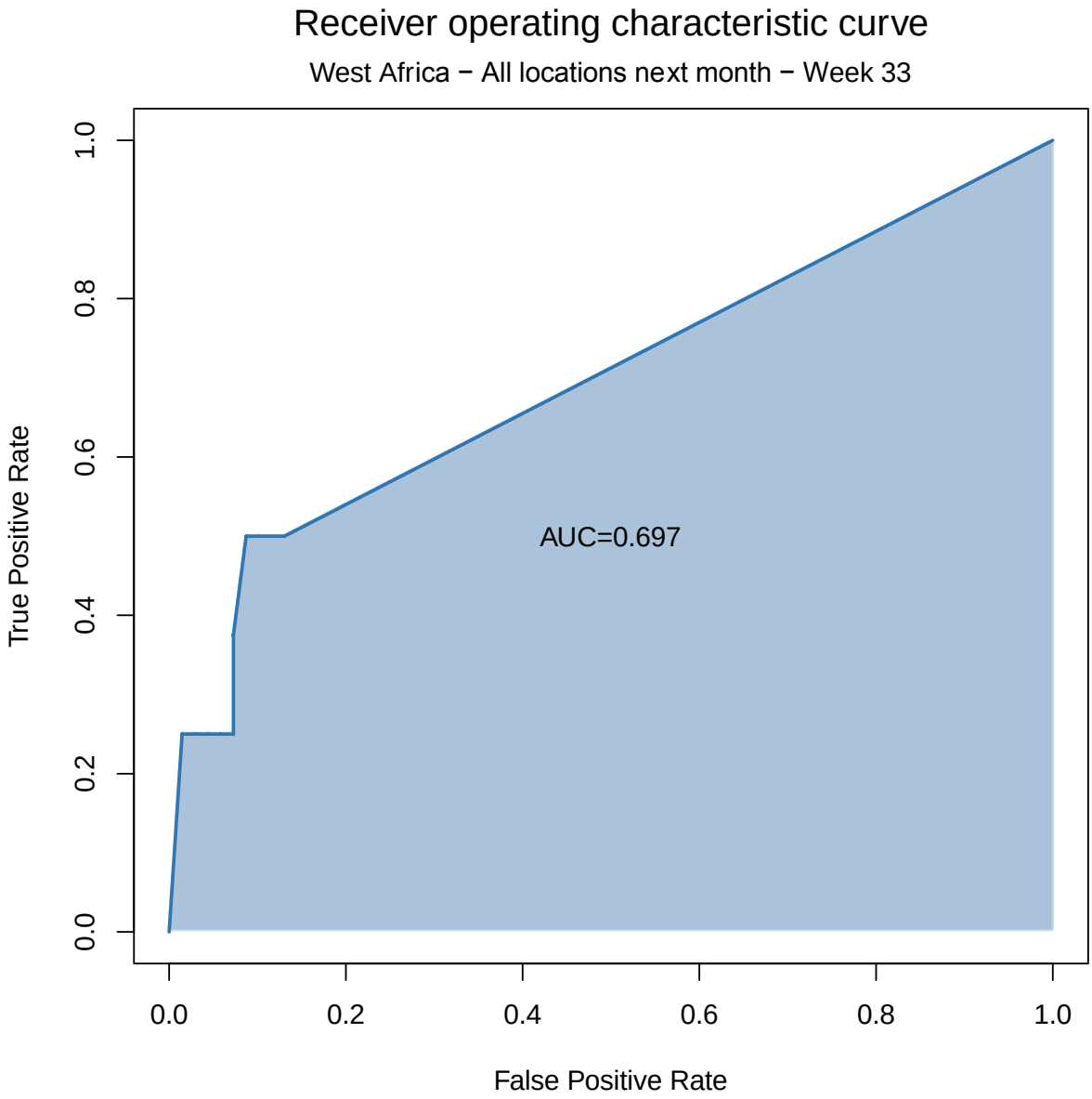


Figure S5: For evaluating the invasion model we used Receiver Operating Characteristic Curve to identify the relationship between the True and False Positive Rate shown here for Week 33 for all locations in the core three affected countries.

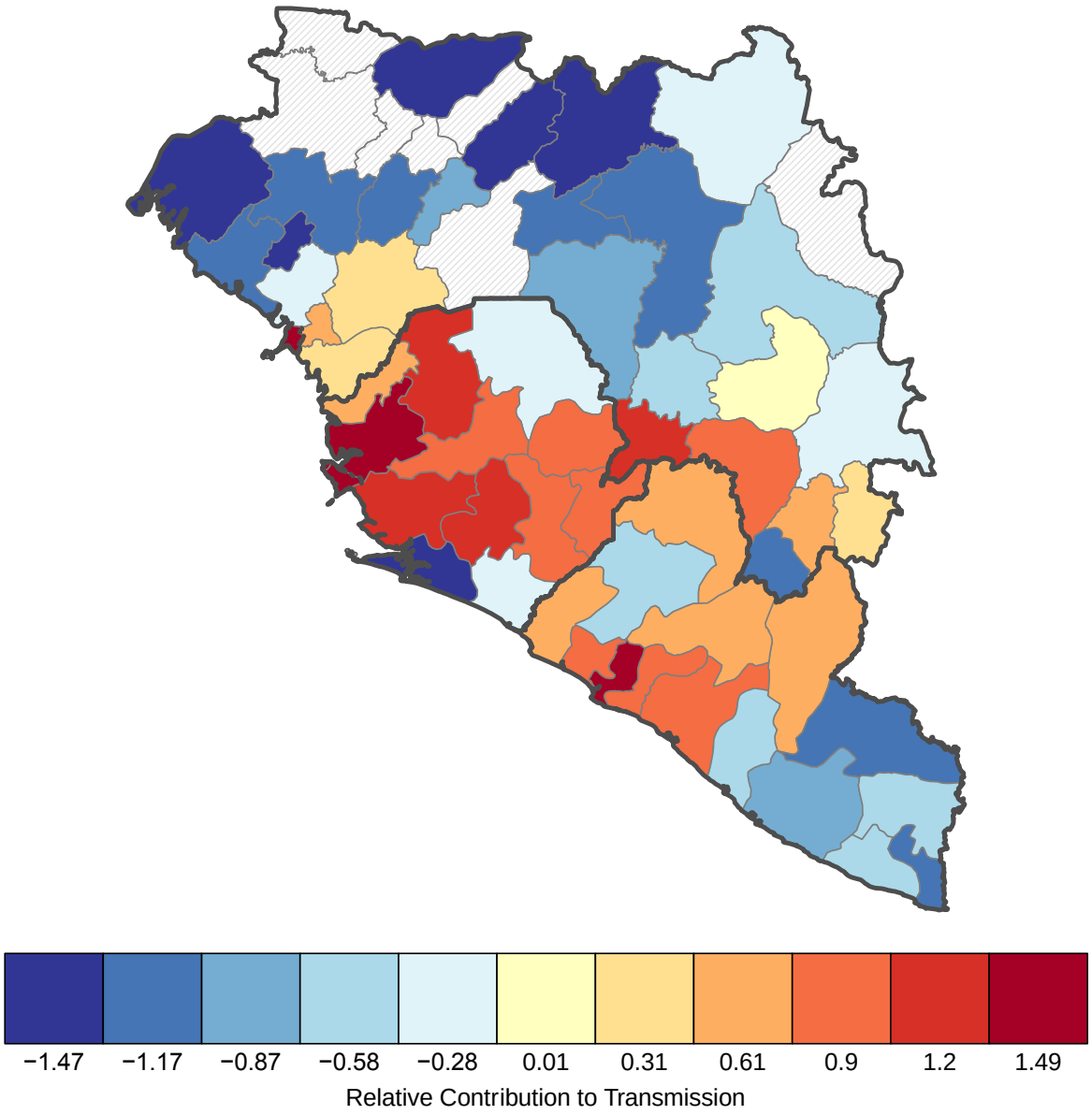


Figure S6: Average relative contribution to transmission for the whole course of the outbreak in West Africa. Districts coloured in red were net exporters of cases and districts in blue were net importers.

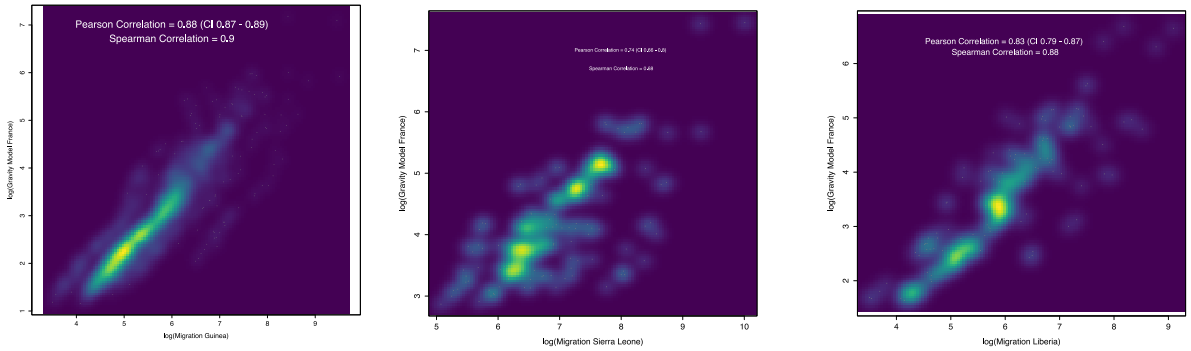


Figure S7: Correlation between predicted movements in Guinea, Sierra Leone, and Liberia from mobile phone data collected in France using the gravity model and migration data (IPUMS) from Guinea, Sierra Leone, and Liberia.

Table S1: Relative contribution of transmission by district (in alphabetical order) for the whole outbreak, the first half and the second half. NA indicates districts that did not have transmission during the timeframe considered.

	Entire Outbreak	First Half	Second Half
GIN_BEYLA	-0.25757661	-0.486448995	-0.183700341
GIN_BOFFA	-1.03945584	-1.189249145	-0.956257275
GIN_BOKE	-1.353444152	-1.375681607	-1.411479175
GIN_CONAKRY	1.641311933	1.39653973	1.425966966
GIN_COYAH	0.560285288	-0.173314192	1.263055449
GIN_DABOLA	-1.075431129	-1.295252387	-0.880320758
GIN_DALABA	-0.92382526	-0.504040539	NA
GIN_DINGUIRAYE	-1.615067751	-1.576689319	NA
GIN_DUBREKA	-0.215269208	-0.896749032	0.536578714
GIN_FARANAH	-0.811022466	-1.375888662	-0.137362512
GIN_FORECARIAH	0.369606388	-0.545110149	1.377927971
GIN_FRIA	-1.434363635	NA	-1.247062307
GIN_GAOUAL	NA	NA	NA
GIN_GUECKEDOU	1.159946598	1.687311383	-0.14449194
GIN_KANKAN	-0.548379421	-1.197638668	0.209664939
GIN_KEROUANE	0.112343313	0.038110146	-0.108833322
GIN_KINDIA	0.19052834	-0.447337141	0.809423227
GIN_KISSIDOUGO	-0.721516473	-0.899715652	-0.647784391
GIN_KOUBIA	NA	NA	NA
GIN_KOUNDARA	NA	NA	NA
GIN_KOUROUSSA	-1.060317047	-1.376006871	-0.722617929
GIN_LABE	NA	NA	NA
GIN_LELOUMA	NA	NA	NA
GIN_LOLA	0.204596775	-0.43781034	0.82804368
GIN_MACENTA	1.012430466	1.197488348	0.250358252
GIN_MALI	-1.504053462	NA	-1.410719357
GIN_MAMOU	NA	NA	NA
GIN_MANDIANA	NA	NA	NA
GIN_N_ZEREKORE	0.735791441	0.563427673	0.560262394
GIN_PITA	-1.041931375	-0.687313966	NA
GIN_SIGUIRI	-0.361111446	-0.78693779	0.027905719
GIN_TELIMELE	-1.17683781	-1.353508019	-1.030299586
GIN_TOUGUE	-1.596764989	NA	-1.628439728
GIN_YOMOU	-1.084342678	-0.753126517	NA
LBR_BOMI	0.909345946	1.155608001	0.071657774
LBR_BONG	0.500243504	0.922940805	-0.536957534
LBR_GBARPOLU	-0.538150594	-0.360996159	-1.032443599
LBR_GRAND_BASSA	0.865323108	0.924540583	0.317961084
LBR_GRAND_CAPE_MOUNT	0.706343586	0.568446539	0.483512844

LBR_GRAND_GEDEH	-1.23088528	-1.142037686	-1.477250464
LBR_GRAND_KRU	-0.51730223	-0.511936885	-0.755058527
LBR_LOFA	0.547616591	1.456644575	-1.233386733
LBR_MARGIBI	0.988099343	1.298724071	0.040015259
LBR_MARYLAND	-1.170332269	-0.886562513	NA
LBR_MONTSERRADO	1.474031816	1.44987366	0.952419752
LBR_NIMBA	0.591263983	0.984236211	-0.41596951
LBR_RIVER_GHEE	-0.562528132	-0.55620302	-0.794275411
LBR_RIVERCESS	-0.50371986	0.014211925	-1.519407346
LBR_SINOE	-0.739305147	-0.488333221	-1.312122594
SLE_BO	1.178153082	1.148106707	0.714266709
SLE_BOMBALI	1.26306605	0.953985393	1.207445928
SLE_BONTHE	-1.434617276	-1.296671837	NA
SLE_FREETOWN	1.624297288	1.091138754	1.848187721
SLE_KAILAHUN	1.019253758	1.461766535	-0.133562499
SLE_KAMBIA	0.602167755	-0.272134005	1.510959271
SLE_KENEMA	1.009134634	1.399783696	-0.063524404
SLE_KOINADUGU	-0.17574956	-0.764428939	0.429139139
SLE_KONO	0.755800062	0.30738997	0.994723529
SLE_MOYAMBA	1.07637182	0.876117731	0.886860714
SLE_PORT_LOKO	1.627483928	1.15529105	1.758586493
SLE_PUJEHUN	-0.266526044	0.198871456	-1.241843785
SLE_TONKOLILI	0.801044662	0.451582975	0.882760401
SLE_WESTERN_RURAL	1.433945687	0.934985338	1.637487096

Table S2: Reduced transmission model for data before week 42. Model was reduced using backwards selection and AIC (note this does not require only significant covariates to be retained).

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.827	0.089	9.251	$<2*10^{-16}$
x	0.666	0.037	17.820	$<2*10^{-16}$
Gravity - Portugal (GP)	-17.552	4.206	-4.173	$3.67*10^{-5}$
Gravity - Spain (GS)	22.337	5.326	4.194	$3.37*10^{-5}$
Radiation - France (RF)	0.312	0.095	3.281	0.001
'one away' weighted by GP	17.103	4.376	3.908	0.0001
'one away' weighted by GS	-21.791	5.446	-4.001	$7.48*10^{-5}$
'two away' weighted by RF	-0.289	0.115	-2.504	0.0127
'three away'	-0.366	0.203	-1.805	0.0719
'three away' weighted by GF	3.749	1.265	2.963	0.0032
'three away' weighted by GS	-3.809	1.209	-3.151	0.0017

Table S3: Reduced Guinea-specific transmission model for data before week 42. Model was reduced using backwards selection and AIC (note this does not require only significant covariates to be retained).

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.181	0.534	4.088	$7.67*10^{-5}$
x	0.424	0.044	5.489	$2.11*10^{-7}$
Gravity - France	-26.440	8.763	-3.017	0.0031
Gravity - Portugal (GP)	81.666	23.758	3.437	0.0008
Gravity - Spain (GS)	-41.732	11.922	-3.5	0.006
'one away' weighted by GP	-7.532	2.931	-2.57	0.113
'two away'	-3.598	1.073	3.353	0.001
'two away' weighted by GF	35.686	11.867	2.923	0.004
'two away' weighted by GP	-35.559	11.923	-2.982	0.003
'three away'	0.866	0.293	2.957	0.004
'three away' weighted by GP	-7.440	2.519	-2.954	0.004
'three away' weighted by GS	6.437	2.289	2.812	0.006

Table S4: Reduced Liberia-specific transmission model for data before week 42. Model was reduced using backwards selection and AIC (note this does not require only significant covariates to be retained).

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.026	0.717	2.825	0.0055
x	0.680	0.068	10.063	$<2*10^{-16}$
Gravity - France	-3.476	1.648	-2.110	0.0369
Gravity - Spain (GS)	-20.741	12.583	-1.650	0.1015
Radiation – France (RF)	49.173	27.478	1.790	0.0760
'one away'	0.262	0.141	1.864	0.0648
'one away' weighted by GS	23.142	11.183	2.069	0.0406
'one away' weighted by RF	-43.956	24.437	-1.799	0.0745
'two away' weighted by GP	-2.896	1.352	2.141	0.0343
'two away' weighted by GS	7.557	2.187	3.456	0.0008
'two away' weighted by RF	-12.415	6.553	-1.894	0.0605
'three away' weighted by GF	1.226	0.711	1.725	0.0871
'three away' weighted by RF	-3.924	2.618	-1.499	0.1366

Table S5: Reduced Sierra Leone-specific transmission model for data before week 42. Model was reduced using backwards selection and AIC (note this does not require only significant covariates to be retained).

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.244	0.157	7.937	$7.36*10^{-13}$
x	0.583	0.058	9.973	$<2*10^{-16}$
Gravity - France	20.876	4.555	4.583	$1.04*10^{-5}$
Gravity - Portugal (GP)	-62.270	14.143	-4.403	$2.16*10^{-5}$
Gravity - Spain (GS)	34.148	8.765	4.238	$4.16*10^{-5}$
'one away' weighted by GP	4.772	2.279	2.095	0.0381
'two away'	0.884	0.517	1.708	0.0899
'two away' weighted by GF	-7.152	3.324	-2.152	0.0332
'two away' weighted by GP	6.852	2.992	2.290	0.0236
'two away' weighted by RF	-0.308	0.171	-1.799	0.0743
'three away' weighted by GF	-6.692	2.271	-2.946	0.0038
'three away' weighted by GP	11.307	4.132	2.736	0.0071
'three away' weighted by GS	-4.488	2.376	-1.889	0.0611

Table S6: Pearson correlation between mobile phone data from France, Spain and Portugal and their correlation with IPUMS migration data.

	Guinea	Sierra Leone	Liberia
France Gravity	0.8799552*	0.7175581*	0.8822296*
France Radiation	0.8637236*	0.7372837*	0.8324313*
Spain Gravity	0.8978669*	0.7300599*	0.8996547*
Spain Radiation	0.8535703*	0.7360462*	0.8251793*
Portugal Gravity	0.8532444*	0.7062177*	0.8602595*
Portugal Radiation	0.870172*	0.7379413*	0.8372988*
			*p-value <0.001

References

1 Tizzoni M, Bajardi P, Decuyper A, *et al.* On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol* 2014; **10**: e1003716.

2 Wesolowski A, Buckee CO, Bengtsson L, Wetter E, Lu X, Tatem AJ. Commentary: Containing the Ebola outbreak – the potential and challenge of mobile network data. *PLOS Curr Outbreaks* 2014; **Sep 29**. DOI:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.

3 Hastie TJ, Tibshirani RJ. Generalized additive models, Vol. 43. CRC Press, 1990.

A.6. Appendix to Chapter 6

Calculation of the generation time distribution of yellow fever

Although it is still poorly characterised, we can try to use field and experimental data to reconstruct the generation time distribution of yellow fever.

Human incubation period (HI)

The incubation period is the time between infection and the time of symptom onset. For the human incubation period we used a truncated exponential distribution with a mean of 4 days and a maximum time of one week.[1]

Human to mosquito transmission (HM)

We assume that the duration of infectivity of human cases is exponentially distributed with a mean of 3 days for up to a maximum of 10 days.[2]

Mosquito infectiousness (MI)

The period of mosquito infectiousness depends on the lifespan of the mosquito and the extrinsic incubation period (the time between infection in the mosquito from blood feeding of an infectious human to it becoming infectious itself and able to transmit to a new host). The average lifespan of *Aedes aegypti* is 7 days with a maximum of 30 days.[3] The extrinsic incubation period for yellow fever has been estimated at 6·9 days.[2]

Generation time distribution

We derived the empirical distribution of the generation time by simulating values for HI, HM and MI. A human case contributed to the transmission process on each day they were infectious (so the number of mosquitoes infected by a case was proportional to the duration of infectivity of the case). The same was true for mosquitoes.

Figure S1 shows the empirical generation time distribution we obtained. The generation time is estimated to have a mean of 15·0 days and a standard deviation of 5·6 days.

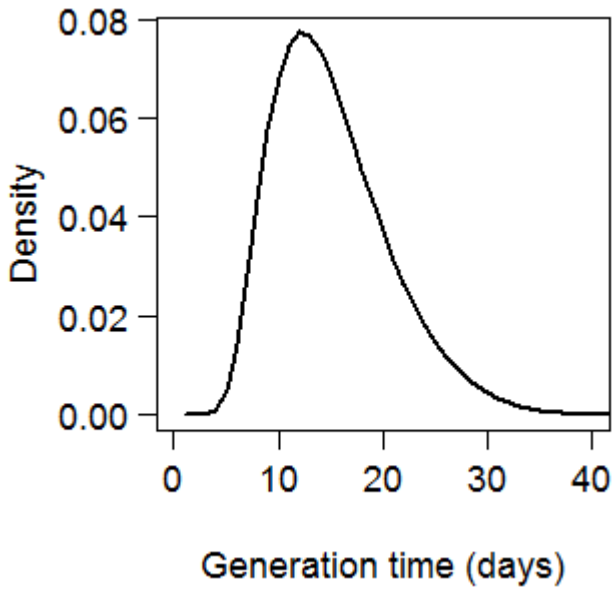


Figure S1: Empirical distribution of the generation time for Yellow Fever.

Estimation of the exponential growth rate, the doubling time and the reproduction number

Exponential growth rate

We fit a simple exponential growth rate model to the early stage of the epidemic:

$$I_w = I_0 exp(r_W . w)$$

where I_w is the number of cases on week w.

The time period for which exponential growth occurs is determined by plotting the log of the weekly number of cases (Figure S1) and selecting time period when this variable grows linearly. A simple linear model is then fitted on this time period to estimate r_W:

$$ln(I_w) = ln(I_0) + r_W . w$$

The daily exponential growth rate r is a simple function of the weekly exponential growth rate r_W:

$$r = r_W/7$$

Between week 1 and week 5 in 2016, we estimate that the weekly exponential growth rate r_W is 0·80 (95% CI: 0·71, 0·90) and the daily exponential growth rate r is 0·11 (95% CI: 0·10, 0·13).

Doubling time

The doubling time D can be derived from the exponential growth rate r with the following formula:

$$D = ln(2)/r$$

Reproduction number

Denote $g(.)$ the density of the generation time (i.e. time lag from the infection of a case to the infection of the persons they infect). The following formula can be used to derive the reproduction

number R from the exponential growth rate r if the generation time distribution $g(.)$ is assumed to be known.[4]

$$R = \frac{1}{\int_0^\infty exp(-r.t)g(t)dt}$$

Geographic spread model

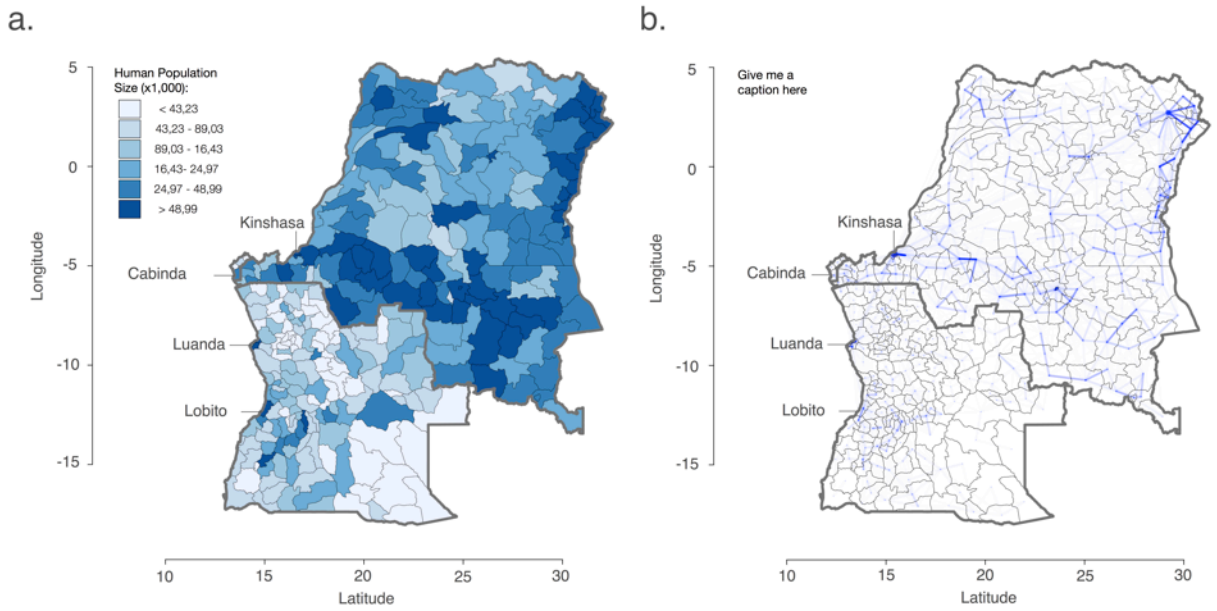


Figure S2: Population distribution in the study area (a) and relative human connectivity (b) between each district (Angola) and commune (DRC). The width of the blue arrows represent the strength of the connections.



Figure S3: Road network in Angola and southern Democratic Republic Congo (www.maps.google.com).

Table S1: List of variables used in geographic spread model.

Name	Reference
Great circle distance	rdist.earth function in the 'fields' package in R
One away adjacency	GADM shapefile (http://www.gadm.org)
Two away adjacency	GADM shapefile (http://www.gadm.org)
Three away adjacency	GADM shapefile (http://www.gadm.org)
Gravity model	Movement package in R, based on Zipf et al. 1946[5]
Radiation model	Movement package in R, Simini et al. 2012[6]
Uniform selection model	Simini et al. 2013[7]
Binary variable pre/post expansion phase	Before and after week 14
Aedes aegypti suitability	Kraemer et al. 2015[8]
Aedes aegypti suitability weighted by mobility	Kraemer et al. 2015[8], Simini et al. 2012[6]
Travel time distance	Uchida and Nelson 2008[9]

Supplementary Results

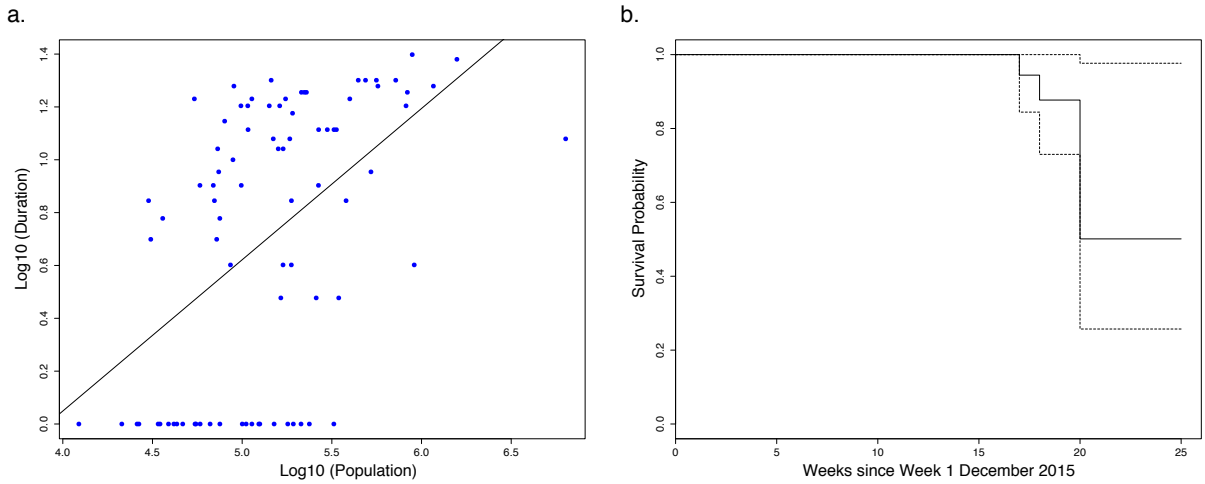


Figure S4: Results showing the relationship between population density and duration of transmission (a). Panel b) shows the results from the Cox model.

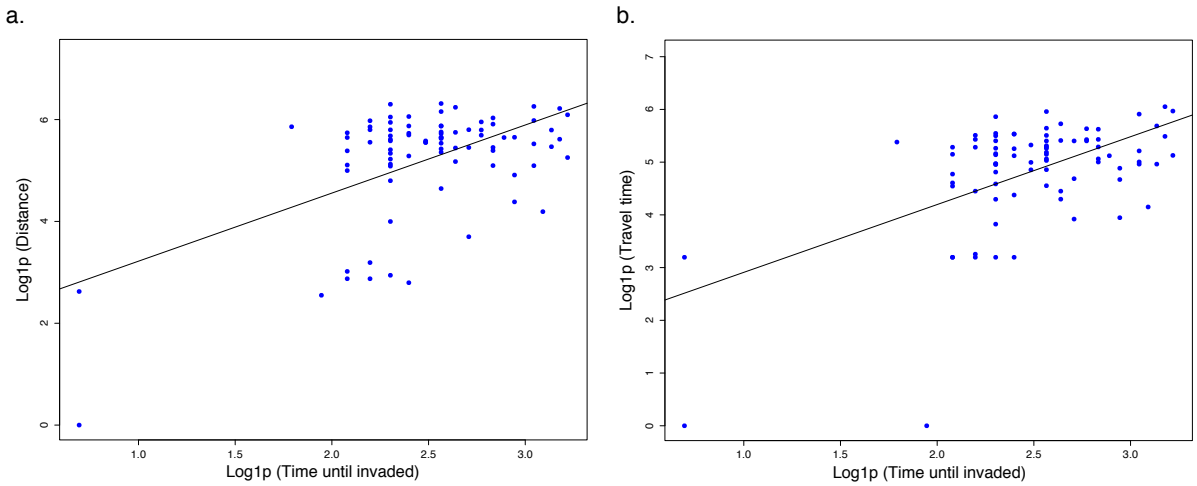


Figure S5: Relationship between distance (a), travel time (b) and time until a district was invaded.

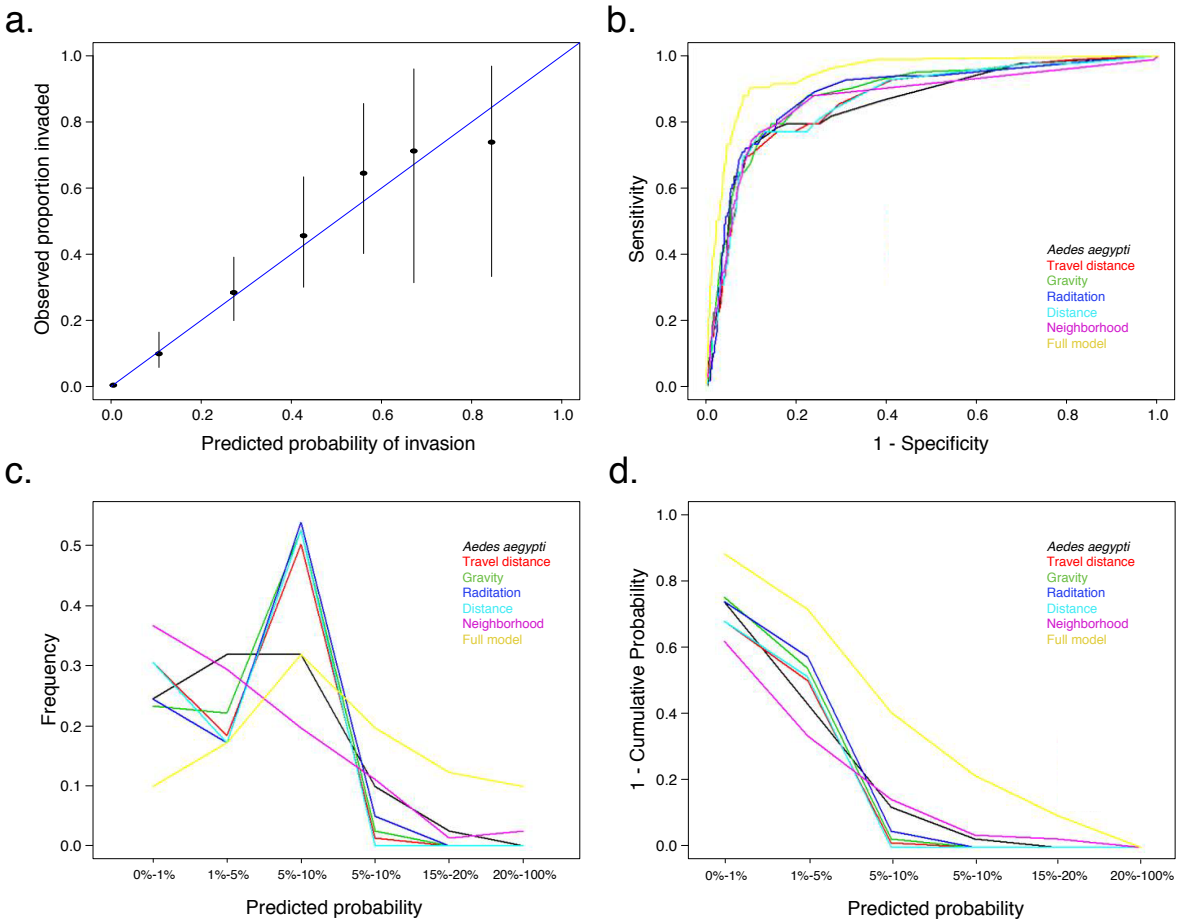


Figure S6: Model comparison between full model and univariate model. A) comparison between observed proportion of districts invaded and their predicted probabilities using the full model. Figures b shows the Area Under Curve (AUC) for the full model vs. the univariate models. Figures c and d show the frequency of predicted probabilities and cumulative probabilities for the full model compared to the univariate models.

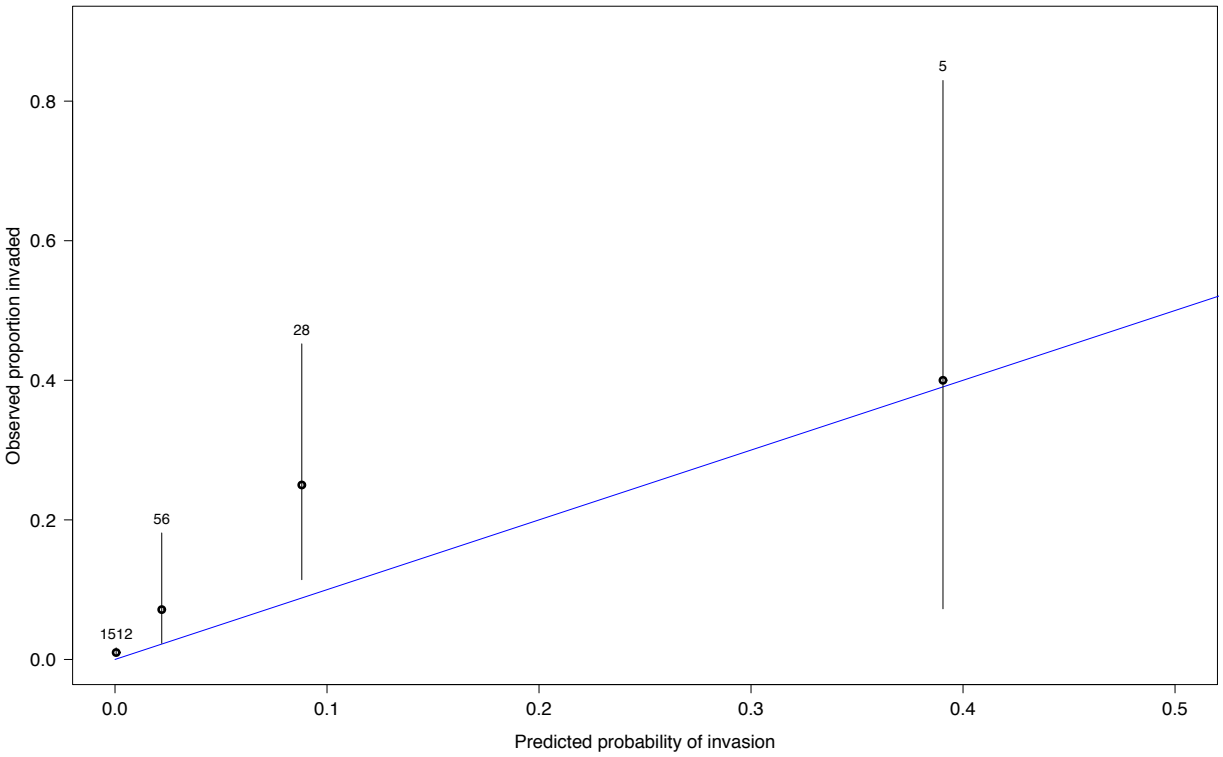


Figure S7: Model accuracy for the real-time predictions between weeks 13-18 during the outbreak with a model using data only until week 12. Blue line indicates perfect calibration.

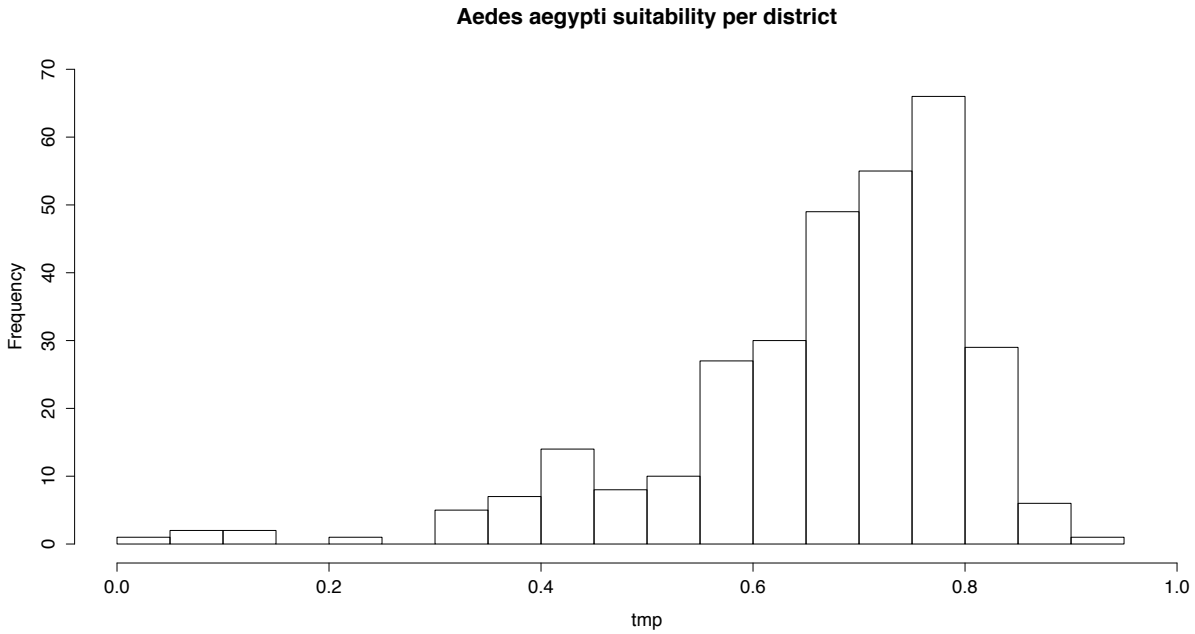


Figure S8: Histogram of district mean values of *Aedes aegypti* suitability for the study region. Estimates are taken from Kraemer et al. 2015.[8]

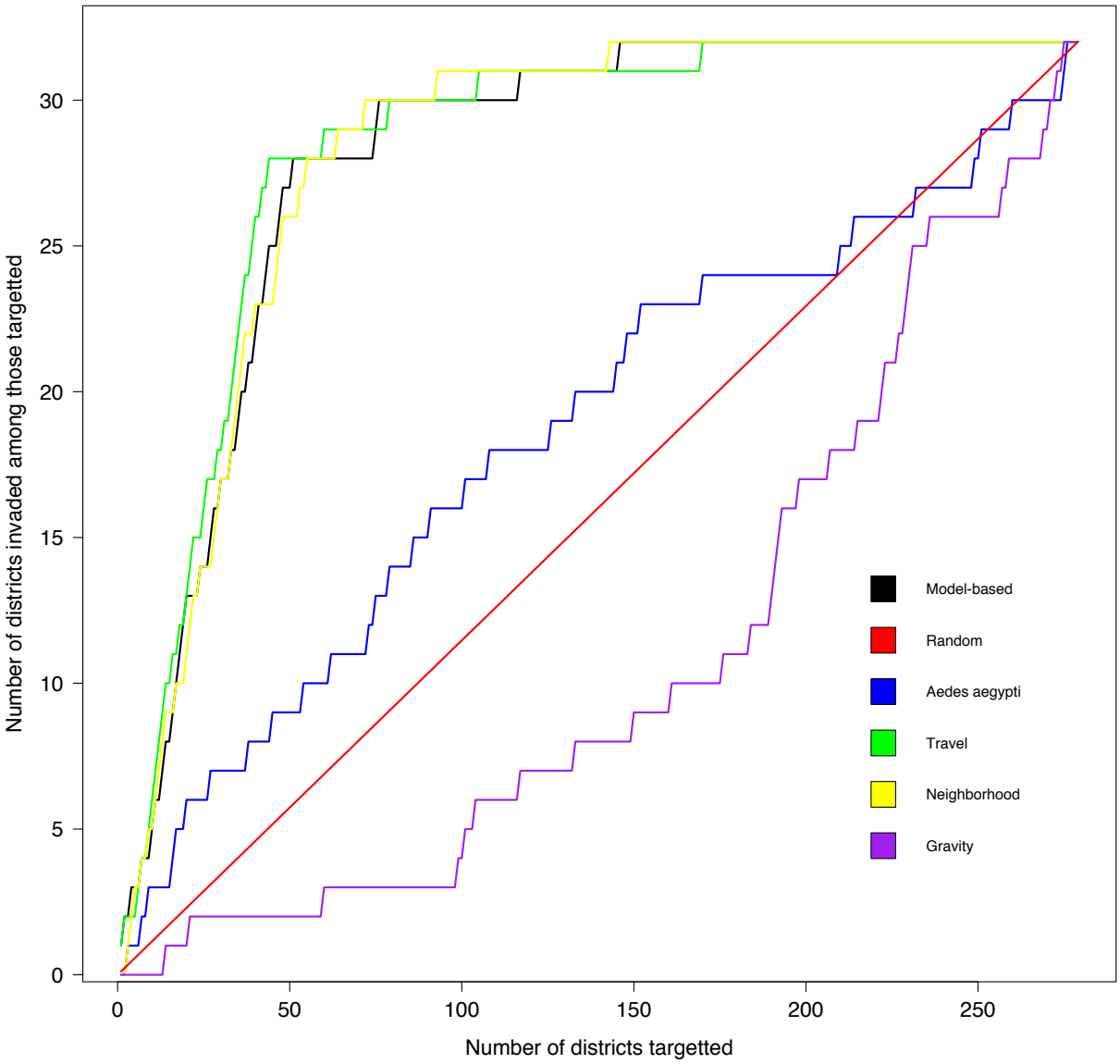


Figure S9: Model comparison of full model vs. univariate models for the expansion phase of the outbreak.

Table S2: Inclusion probability of variables in the model when using backward selection based on significance.

Name	Weeks variable was retained in the model (starting week 11 of outbreak)	Total number of weeks
Great circle distance	11, 18, 19, 21, 23, 25, 27 – 34	14
One away adjacency	14	1
Two away adjacency	15-34	20
Three away adjacency	13, 14, 16, 18	4
Gravity model	11-14, 16-18, 21-34	21
Radiation model	14, 15, 19-23, 25-34	17
Uniform selection model	11, 12, 21, 23-25, 27-34	14
Binary variable pre/post expansion phase	14-34	21
Aedes aegypti suitability	12, 13, 16-19, 21-24, 26-34	19
Aedes aegypti suitability weighted by mobility	13, 14, 16-18, 21-34	19
Travel time distance	12-17, 20	7

References

1. Johansson, M. A., Arana-Vizcarrondo, N., Biggerstaff, B. J. & Staples, J. E. Incubation periods of yellow fever virus. *Am. J. Trop. Med. Hyg.* **83,** 183–188 (2010).
2. Johansson, M. A. *et al.* Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *Am. J. Trop. Med. Hyg.* **86,** 349–58 (2012).
3. Bellan, S. E. The importance of age dependent mortality and the extrinsic incubation period in models of mosquito-borne disease transmission and control. *PLoS One* **5,** e10165 (2010).
4. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274,** 599–604 (2007).
5. Zipf, G. K. The P1 P2 / D hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* **11,** 677–86 (1946).
6. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484,** 96–100 (2012).
7. Simini, F., Maritan, A. & Néda, Z. Human mobility in a continuum approach. *PLoS One* **8,** e60069 (2013).
8. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *Elife* **4,** e08347 (2015).
9. Uchida, H. & Nelson, A. *Agglomeration index: towards a new measure of urban concentration*. (2008).

A.7. Additional Publications

Brady et al. Parasites & Vectors 2014, 7:338
http://www.parasitesandvectors.com/content/7/1/338

Parasites & Vectors

RESEARCH Open Access

Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission

Oliver J Brady[1*], Nick Golding[1], David M Pigott[1], Moritz U G Kraemer[1], Jane P Messina[1], Robert C Reiner Jr[2,3], Thomas W Scott[2,3], David L Smith[1,3,4], Peter W Gething[1] and Simon I Hay[1,3]

Abstract

Background: Dengue is a disease that has undergone significant expansion over the past hundred years. Understanding what factors limit the distribution of transmission can be used to predict current and future limits to further dengue expansion. While not the only factor, temperature plays an important role in defining these limits. Previous attempts to analyse the effect of temperature on the geographic distribution of dengue have not considered its dynamic intra-annual and diurnal change and its cumulative effects on mosquito and virus populations.

Methods: Here we expand an existing modelling framework with new temperature-based relationships to model an index proportional to the basic reproductive number of the dengue virus. This model framework is combined with high spatial and temporal resolution global temperature data to model the effects of temperature on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission.

Results: Our model predicted areas where temperature is not expected to permit transmission and/or *Aedes* persistence throughout the year. By reanalysing existing experimental data our analysis indicates that *Ae. albopictus*, often considered a minor vector of dengue, has comparable rates of virus dissemination to its primary vector, *Ae. aegypti*, and when the longer lifespan of *Ae. albopictus* is considered its competence for dengue virus transmission far exceeds that of *Ae. aegypti*.

Conclusions: These results can be used to analyse the effects of temperature and other contributing factors on the expansion of dengue or its *Aedes* vectors. Our finding that *Ae. albopictus* has a greater capacity for dengue transmission than *Ae. aegypti* is contrary to current explanations for the comparative rarity of dengue transmission in established *Ae. albopictus* populations. This suggests that the limited capacity of *Ae. albopictus* to transmit DENV is more dependent on its ecology than vector competence. The recommendations, which we explicitly outlined here, point to clear targets for entomological investigation.

Keywords: *Aedes*, *Aegypti*, *Albopictus*, Temperature, Distribution, Suitability, Competence, Disease model, Map

Background

Dengue is the most clinically important viral vector-borne disease with 96 million apparent infections per year among a population-at-risk of nearly four billion in 128 countries [1-3]. During the last century dengue virus (DENV) has rapidly expanded from its tropical origins to sub-tropical and temperate climates. Similarly, the geographic limits of transmission and endemic regions (with year round transmission) continue to expand [4-7]. While significant improvements have been made to our understanding of and our ability to map the current global distribution of dengue [1,2], we do not yet know which factors define the current and future geographic limits to the extent and endemicity of transmission of this still expanding viral infection. These limits are likely defined by a combination of different host (both mosquito and human) and virus factors in different environments. While the study of these factors in isolation is

* Correspondence: oliver.brady@zoo.ox.ac.uk
[1]Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom
Full list of author information is available at the end of the article

BioMed Central

© 2014 Brady et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

unlikely to define the full constraints of transmission, it nevertheless can identify areas where transmission cannot occur.

Of the many factors affecting dengue transmission, temperature has been most frequently investigated and rigorously quantified [8,9]. A variety of experiments have quantified how key vector and virus dynamics vary with temperature [10,11] and high spatial and temporal resolution meteorological data are available to apply these relationships on a global scale [12]. Using accumulated theoretical and experimental knowledge of the epidemiology of dengue, we use a dengue transmission model to identify where and when temperature is likely to be the most important factor limiting the rate of transmission. Among the limitations temperature can impose on DENV transmission, the most directly affected and most clearly understood mechanisms are its effects on adult female mosquito survivorship and DENV extrinsic incubation period (EIP). Transmission is only permitted when the longevity of the vector exceeds the EIP of the DENV. This is equivalent to the time between the mosquito imbibing an infected blood meal and becoming infectious (able to onwardly transmit the virus by bite). Similarly, a key limitation on vector persistence, the ability for mosquito population dynamics to sustain a population of adult mosquitoes, can be evaluated by determining the proportion of adult female mosquitoes that survive long enough to complete their first gonotrophic cycle (FGC), or the time between taking a blood meal and first oviposition [13]. Evaluating the effects of temperature on these two processes on a population level can determine when and where temperature causes breaks in transmission or vector population persistence. This can also give insights into the relative intensities of transmission and oviposition as if a greater proportion of mosquitoes survive the EIP and FGC they are thus more likely to deliver more infectious bites or oviposit a greater number of eggs respectively.

Evaluating the limits to global DENV transmission is complicated by differences in the global distributions of the DENV's primary vector, *Aedes aegypti*, and what is often considered a secondary vector, *Ae. albopictus* [14,15]. While *Ae. aegypti* appears to be confined to tropical and subtropical climates [8,16], *Ae. albopictus* eggs can undergo long periods of diapause that allow populations to persist in temperate environments despite winter temperatures that are unsuitable for adult survival. Survival of eggs for relatively long periods of time also allows them to be transported long distances using global travel and trade networks [17,18], resulting in multiple introductions outside their native range into tropical and some temperate environments, most well documented in Europe and North America [17]. The role of *Ae. albopictus* in DENV transmission remains unclear. In areas where it is the only DENV vector, outbreaks are rare and limited in incidence [15,19,20], whereas even recently introduced, isolated populations of *Ae. aegypti* have produced explosive and sustained outbreaks [21,22]. It has been suggested that this difference is primarily due to inter-species differences in the rate of DENV dissemination through the mosquito midgut to the salivary glands, leading to a longer EIP in *Ae. albopictus* [15]. The effect of this difference on transmission has not, however, been quantified taking into account the predicted longer lifespan of *Ae. albopictus* [10]. Alternative explanations involve differences in human blood feeding rates (*Ae. aegypti* has been known to deliver more frequent bites with a higher proportion of blood meals taken from humans than *Ae. albopictus* [23,24]) and possible differences in the proportion of mosquitoes per human resulting from their respective ecologies. *Ae. aegypti* typically inhabits more urban environments than *Ae. albopictus*, and even when both inhabit urban environments, *Ae. aegypti* typically lives in closer proximity to humans [25,26]. While both of these factors are also driven by temperature, they also have strong ecological influences that may in many cases be greater than the effect of temperature. We therefore chose to restrict our analysis to the effects of temperature on the duration of EIP and FGC.

Previous attempts to model the effects of temperature on *Ae. aegypti* and *Ae. albopictus* population persistence and ability to transmit DENV have either taken a statistical or mechanistic (biological or deterministic) modelling approach. Statistical or empirical approaches that attempt to find correlations between various measures of temperature and dengue or vector observations often find weak or only transient associations [27,28]. These approaches cannot be easily designed to include the complexities of the temperature response due to its highly non-linear effects and the dynamic time scale at which it acts [29]. Such limits are often simply defined by a constant mean or minimum temperature that fails to account for seasonal and daily temperature fluctuations throughout the year. Statistical approaches are further restricted in their ability to detect the limits of the distribution due to the inherent scarcity of reported data from environments where transmission or mosquito populations are marginal [30,31]. Conversely, previous mechanistic modelling approaches typically use temperature-based relationships defined by just one or two empirical studies and lack the geographic validation of the statistical approaches [9,32,33]. Furthermore, many studies using mechanistic models are not widely generalisable due to their reliance on locally-specific parameter estimates, particularly estimates of immature mosquito habitat and fluctuations in adult mosquito population sizes. This results in a high degree of uncertainty when trying to evaluate the effects of temperature on the persistence of vector

populations and their ability to transmit DENV and this uncertainty is rarely quantified in the final prediction.

In this study we attempt to elucidate the thermal limits of *Ae. aegypti* and *Ae. albopictus* persistence and DENV transmission at a high spatial and temporal resolution using an empirically-parameterised mechanistic model that explicitly considers the interplay between temperature-dependent EIP and adult vector survival. We use existing temperature-based relationships [10,11,34] and define new ones by applying complex modelling techniques to the increasing breadth of data from many different laboratory and field conditions, also including measures of uncertainty in our estimates. These relationships are then combined in a dynamic modelling framework that evaluates the cumulative effects of changing temperature regimes on an index proportional to vectorial capacity. This approach reveals different distinct limits to persistence and virus transmission for each species and goes on to indicate that temperature can also limit levels of endemicity of DENV transmission. In addition to these findings, the predictive maps produced from these models can be used as more biologically-appropriate covariates than typical temperature averages in statistical modelling of dengue and as an informative subcomponent to link vector and human models of DENV transmission.

Methods

Relating temperature to transmission potential

Here we adapt a modelling framework from Gething *et al.* [35,36] that evaluates the principal effects of temperature on the basic reproductive number (R_0) of a pathogen. The concept of R_0, the number of secondary cases that will arise from a single infectious human in a totally susceptible population, is a natural framework for evaluating relative changes in transmission potential for a variety of vector borne diseases [37-40]. The entomological, and thus strongly temperature dependent, components of R_0 can be summarised by the equation for vectorial capacity (V):

$$V = \frac{ma^2}{g} e^{-gn}, \quad (1)$$

where m is the ratio of adult female mosquitoes to humans, a is the human blood feeding rate, g is the instantaneous per-capita death rate of adult female mosquitoes and n is the length of EIP for the pathogen. All of the parameters in vectorial capacity are likely to be affected to some extent by temperature, either directly or indirectly through interaction with the environment. In our generalised global approach, specific details of the environment are unknown, so we therefore chose to confine our analyses to mosquito mortality (g) and pathogen extrinsic incubation period (n) due to the direct effects of temperature on these parameters and subsequent potential virus transmission and because of the abundance of information available. If the human population size is stable, m can be replaced by two entomological parameters describing mosquito population dynamics: $m = \lambda / g$, where λ is the emergence rate of new adult mosquitoes. Though mosquito mortality already appears twice in Eq. 1, this substitution reveals a third order effect of mortality on transmission through effects on the adult population size [41] and allows us to define an index of temperature suitability Z as a function of temperature, T, which is linearly proportional to vectorial capacity over small changes in temperature:

$$Z(T) = \frac{e^{-g(T)n(T)}}{g(T)^2} \propto V(T) = \lambda a^2 \left(\frac{e^{-g(T)n(T)}}{g(T)^2} \right). \quad (2)$$

$Z(T)$ therefore effectively measures the relative number of infectious mosquitoes supported in an environment with temperature T, given a constant emergence rate λ and human blood feeding rate a. As estimates for λ and a are not readily available across the wide range of different environments that exist globally, $Z(T)$ provides a suitable metric for evaluating the effects of temperature on vectorial capacity, and therefore R_0. While $Z(T)$ can only measure relative rather than absolute change in vectorial capacity, a value of $Z(T)$ approximately equal to zero indicates that an environment does not permit onward transmission and can thus be used to predict the geographical limits of transmission.

While $Z(T)$ can be used to estimate relative suitability for transmission as a function of temperature, a modified version of the model can be used to estimate suitability for the persistence of mosquito populations also as a function of temperature. Rather than surviving long enough to transmit the virus, we need to evaluate whether the mosquito can survive long enough to complete one gonotrophic cycle (c), the minimum necessary for reproduction and thus persistence of a permanent mosquito population. A simple substitution of c for n in equation 2 enables the calculation of temperature suitability for oviposition. This allows us to evaluate the limits temperature places on the persistence of adult *Ae. aegypti* and *Ae. albopictus* populations, assuming the environment can support the juvenile stages of each species:

$$Z_{ovi}(T) = \frac{e^{-g(T)c(T)}}{g(T)^2}. \quad (3)$$

Temperature-based relationships

Calculation of temperature suitability for transmission and vector persistence requires evidence-based models for mosquito mortality, EIP and the length of the FGC.

For the first two relationships we draw on two previously published studies investigating the effects of temperature on these parameters and incorporate both mean prediction and the uncertainty in these estimates. For FGC, we apply an existing model to newly available data from a literature review of entomological life cycle experiments.

Adult mosquito survivorship

Brady *et al.* [10] used generalised additive model regression to fit an empirical relationship between temperature and adult female survival in both *Ae. albopictus and Ae. aegypti* using over 350 laboratory and 50 field studies. An initial model of survival in a laboratory setting at a range of different temperatures was fitted, whilst accounting for potentially confounding factors. The laboratory survival model was then adapted to estimate survival under field conditions by calculating the additional hazard observed in the field as quantified by mark-release-recapture experiments. Laboratory and field models were derived separately for *Ae. aegypti* and *Ae. albopictus*, but in both mortality was modelled as a function of both temperature (T) and mosquito age (D):

$$logit\left(-\ln\left(g_{D,T}\right)\right) = f_s(D,T) + \gamma_s\,, \tag{4}$$

where f is a smooth term and y is the constant additional field mortality hazard for each mosquito species s.

DENV incubation period in Aedes aegypti

Chan and Johansson [11] used censored Bayesian time-to-event hazard models to parameterise a log-normal relationship between temperature and DENV EIP. This model was fitted to data from 140 mosquito transmission experiments where a cohort of mosquitoes was infected, incubated at constant temperature T, then transmission to human or primate hosts monitored at defined time intervals, accounting for inter-study differences in competence, to estimate the expected EIP duration (n).

$$n_T = e^{\beta X + \frac{1}{2\tau}} \tag{5}$$

$$\beta X = \beta_0 + \beta_T T \tag{6}$$

The vast majority of EIP experimental data was conducted using *Ae. aegypti* (96%), while only six experiments concerned EIP in *Ae. albopictus.* As these data were insufficient for detecting EIP differences between the two species, this model was fitted using only experiments that used *Ae. aegypti*, however, this yielded no difference in the fitted parameters than in the model that used data from both species. The final log-normal model was fitted with parameter estimates $\tau = 4.9$, $\beta_0 = 2.9$, and $\beta_T = -0.08$.

Inter-species differences in DENV incubation period

Few experiments have been conducted that have been designed to accurately measure DENV EIP in *Ae. albopictus.* To account for potential differences in EIPs for *Ae. albopictus*, we assessed data from vector competence assays in which cohorts of mosquitoes are experimentally infected and later assessed for disseminated DENV infection using immunofluorescent assays or similar techniques on cohorts extracted at one or more pre-defined time points.

To collate a comprehensive dataset with which to parameterise this model, databases of published literature (Pubmed and Google Scholar) were searched for the terms "aedes" AND "competence" and relevant data were extracted from the full text articles. Studies were included if DENV was ingested orally, DENV dissemination was measured in the head or salivary gland tissues and the mosquitoes had a laboratory colonisation history of five generations or fewer, as it has been found that long-term laboratory colonisation can increase vector competence [15]. The data extracted were: number infected at the beginning of the experiment, number sampled and number DENV positive at each time points and incubation temperature. We used this data to fit the following hierarchical model:

$$y_{t,T,\,i,j} \sim Bin\left(P_{t,T,\,i,j}\,,\; n_{t,\,T,\,i,j}\right)\,, \tag{7}$$

where y is the number of mosquitoes of species j in study i with DENV in their salivary glands at the assay time point t after incubation at temperature T, n is the number of mosquitoes infected at the beginning of the experiment and P is the probability of infection which is determined by:

$$P_{t,T,\,i,j} = \Phi\left(t_i \middle|\, \mu_{EIP_{T,i,j}},\; \sigma_{EIP}{}^2_{T,\,i,j}\right)\,,$$

in which Φ is the cumulative distribution function of the Gaussian distribution, t_i is the time in days between the initial infection and the assay, μ_{EIP} is the mean EIP and σ^2_{EIP} is the variance in EIP in each DENV assay experiment. We then used the assay data to calculate species-specific proportional differences (α_j) in the mean predicted EIP duration (from the EIP-temperature model, $\hat{\mu}_{EIP_T}$), taking into account inter-study variation due to different experimental designs, vector strains and virus types (ε_i):

$$\mu_{EIP_{T,i,j}} = \hat{\mu}_{EIP_T} + \alpha_j\,.\hat{\mu}_{EIP_T} + \varepsilon_i$$

$$\varepsilon_i \sim N\left(0,\; \sigma_{study}{}^2\right)$$

$$ln\left(\sigma_{EIP\,T,\,i,\,j}\right) \sim N\left(\mu_{assay},\; \sigma_{assay}{}^2\right)\,.$$

If viral dissemination and infectiousness are equivalent, we would expect $\alpha_{aegypti} = 0$, however due to various

factors associated with sensitivities to DENV detection, we may expect detection of DENV dissemination to occur after the mosquito becomes infectious ($\alpha_{aegypti} > 0$). Alternatively, transmission experiments may produce a higher rate of false-negatives than dissemination assays [42] ($\alpha_{aegypti} < 0$). We therefore calculated species-specific values for α_j and took the relative differences in the means to calculate the efficiency adjustment for EIP duration in *Ae. albopictus* (ϕ):

$$\phi = \frac{\hat{\mu}_{EIP_T} + \alpha_{aegypti}.\hat{\mu}_{EIP_T}}{\hat{\mu}_{EIP_T} + \alpha_{albopictus}.\hat{\mu}_{EIP_T}} = \frac{1 + \alpha_{aegypti}}{1 + \alpha_{albopictus}} .$$

The above model was fitted as a hierarchical Bayesian model. A vague zero-mean Gaussian prior was specified for α_j. Twenty-eight assay experiments estimated the infected population at multiple time points by taking samples from the overall population. In order to incorporate this information, a cumulative normal distribution function was fitted to these experiments by least squares regression to generate empirical (though weakly informative) priors for μ_{assay} and σ^2_{assay}. Finally, we specified weakly informative priors for σ^2_{study} based on previous models of DENV EIP [11]. See Additional file 1 for the specific priors used. The model was parameterised by MCMC using WinBUGS [43]. Three Markov chains were run for 10000 iterations each with the first 2000 iterations discarded. Convergence was assessed based on visualisations and the Gelman-Rubin statistic.

Length of first gonotrophic cycle

To calculate the relationship between temperature and length of the first gonotrophic cycle we fitted an enzyme kinetics model first suggested by Focks *et al.* [9] to an updated database. The database was updated by searching literature databases spanning 1960-2014 for the terms "aedes" and "oviposition" or "gonotrophic". Experiments that monitored time of emergence, temperature and time of first oviposition were included as long as the mosquitoes had access to a food source and oviposition substrate. Due to the limited detail with which the data were presented, we were only able to extract mean time to oviposition, thus limiting the scope of our modelling options. The four parameters of the following equation was then optimised by minimising mean-squared error using the "L-BFGS-B" method with no upper or lower bounds in R version 3.0.1 [44]:

$$c(T_k) = \frac{\rho(T_k/298)e^{[(H_A/1.987)((1/298)-(1/T_k))]}}{1 + e^{[(H_H/1.987)((1/T_{0.5})-(1/T_k))]}} ,$$

where T_k is the temperature in °K, and ρ, H_A, H_H and $T_{0.5}$ are parameters to fit. The model was run 500 times with different starting values to ensure global and not local minima were found. Uncertainty in these parameters was estimated using the inverse Hessian of the likelihood surface.

Calculating temperature suitability at high temporal and spatial resolution

Long term average (1950-2000) interpolated weather station temperature data was obtained from WorldClim in the form of global 5 km × 5 km global rasters containing monthly maximum and minimum temperatures [12]. These were aligned back to back into three repeated years then cubic spline interpolated using the stats package in R to give annual weekly maximum and minimum temperature values pixel-by-pixel. Finally, to simulate the effects of diurnal temperature changes, weekly maximum and minimum temperatures were used to fit a sinusoidal day and exponential night temperature profile for each day of the week [32,45-49]. This resulted in global rasters giving interpolated temperature values at two hour time steps throughout an average year. We limited our predictions of temperature suitability to areas with a maximum monthly temperature exceeding 13°C for *Ae. albopictus* and 14°C for *Ae. aegypti*. These thresholds represent the observed temperatures below which biting and movement behaviours are impaired [8,24,46,49]. This impairment affects survival, blood feeding and larval development in ways that are outside the scope of the mortality models included in this analysis. As such, we do not predict below these temperature limits.

Vectorial capacity, and thus $Z(T)$, is a measure of transmission potential indexed to the time point when a mosquito becomes infected after blood feeding on an infectious human. To calculate the number of infectious mosquito days that arise from this one event we therefore need to consider: i) mosquito mortality in the time preceding mosquito infection, ii) mosquito mortality and virus EIP proceeding infection but preceding mosquito infectiousness and iii) mosquito mortality after they become infectious. In an environment with constantly changing temperatures it is important to incorporate the sequence and magnitude of temperature changes over time and their resultant effect on $Z(T)$. In a temporal sense, therefore, $Z(T)_i$ measures the relative number of infectious mosquito days that will arise in the future following infection on day i.

To capture changes in $Z(T)_i$ over time, we used a simulation-based discrete time step approach at a high temporal resolution to estimate the integral of the $Z(T)$ time series, where T is a vector of different temperatures. This modelled adult mosquito population size in three compartments: i) uninfected, ii) infected but not yet infectious and iii) infectious, each made up of cohorts that emerge or are infected at different times. Temperature- and age-dependent death rates were applied to each compartment up until an imposed maximum lifetime of

22 days for *Ae. aegypti* and 65 days for *Ae. albopictus*, by which time best-case survival is below 1% in the mosquito survival model [10]. At each two hour time point a population of 100 adult mosquitoes (note that the choice of emerging population size is arbitrary as it has no effect on the relative measure $Z(T)$) was simulated to emerge into the uninfected compartment and all mosquitoes in all compartments were exposed to the mortality rate for that time point. Two days after emergence, when *Aedes* mosquitoes are first capable of blood feeding [50], all surviving mosquitoes in the uninfected compartment become infected. For infected mosquitoes, temperature dependent EIP completion proportion was tracked using equations 5 and 6 for each cohort in the infected compartment. Upon completion of the EIP, mosquitoes were transferred from the infected to infectious compartments then the number of infectious mosquito-days was summed up until the maximum lifetime.

By summing the number of infectious mosquito-days arising from cohorts of different ages all infected at time point i, the contribution of time step i to future transmission, or $Z(T)_i$, can be estimated. From the surveillance perspective, however, it would be useful to know the proportion of vectors infectious X_i at time step i given the environmental conditions that have occurred in the time leading up to that point. We therefore also calculate this number as the proportion of mosquitoes from all existing cohorts at time step i that are infectious. Each model was initiated with a burn in period exceeding the maximum lifetime of the vector, after which $Z(T)_i$ and $X(T)_i$ was calculated at two hour time steps from 1st January to 31st December. Equivalent time series for $Z_{ovi}(T)_i$ and $X_{ovi}(T)_i$ were also evaluated using the same approach with three compartments: i) newly emerged, ii) blood-fed but not ovipositing and iii) ovipositing.

Summarising annual time series

The four time series outputs from the temperature suitability simulation ($Z(T)$, $X(T)$, $Z_{ovi}(T)$ and $X_{ovi}(T)$) represent subtly different measures of the effect of temperature on dengue transmission and *Aedes* persistence. Over the course of an average year, if we take the sum of all the $Z(T)$ and $X(T)$ values for each time step $\sum Z(T) = \sum X(T)$, and $\sum Z_{ovi}(T) = \sum X_{ovi}(T)$, however, the timing of peaks and troughs will differ. $Z(T)_i$ and $Z_{ovi}(T)_i$ summarise the contribution of a single event at time i to future transmission and oviposition respectively. This is equivalent to the introduction of one infectious human or one emerging population of *Aedes* mosquitoes. By contrast $X(T)_i$ and $X_{ovi}(T)_i$ measure the contribution of all past events to suitability at time i, which is equivalent to multiple infectious events or emerging mosquito populations having occurred in the past that determine current suitability. Each of these measures has different implications depending on DENV endemicity and *Aedes* persistence status in the environments they summarise. The global distribution of dengue transmission varies by season, however, the global distribution of *Ae. aegypti* and *Ae. albopictus* shows little intra-annual variation mainly due to a limited flight range [8,51]. It follows that DENV transmission can exist in many stable states between sporadic transmission ($Z(T)$) and hyperendemicity ($X(T)$), however, *Aedes* populations are either absent with a risk of introduction ($Z_{ovi}(T)$), or persist year-round with varying degrees of suitability ($X_{ovi}(T)$). The measures for oviposition therefore reflect more absolute limits, while the measures of transmission document two ends of the spectrum of the many possible transmission intensities.

Vector temperature suitability

To determine the thermal limits of *Ae. aegypti* and *Ae. albopictus* persistence, we compared our measures of persistence suitability ($X_{ovi}(T)$) to known geographic occurrences of the vectors. Both *Ae. aegypti* and *Ae. albopictus* populations are robust to brief to medium term periods of environmental unsuitability due to transient behavioural and physiological adaptations at various life-cycle stages [52,53]. To determine the maximum unsuitable time period for each species the number of days in the year where $X_{ovi}(T)_i = 0$ was calculated for 4053 known geographical occurrences of *Ae. aegypti* and 1459 of *Ae. albopictus*. This database of known *Ae. aegypti* and *Ae. albopictus* occurrence points is composed of precise geographic locations where each species has been identified and reported in peer-reviewed literature sources. Full details of the database sources and extraction are available in Additional file 2. As the majority of these occurrence points only detect presence at one point in time, the database may contain identifications of temporary novel introductions as opposed to confirmed year round persistence of mosquito populations. As these records are likely to exist in the minority given their relative rarity, we chose a threshold of the number of suitable days ($X_{ovi}(T)_i > 0$) in an average year that incorporated 95% of these known occurrences. Separate thresholds were chosen for each species then the $X_{ovi}(T)$ index, normalised between 0 and 1, was displayed inside these limits, to show cumulative temperature suitability from established populations. In subsequent maps of temperature suitability for dengue transmission, the limits to vector establishment were used as a constraint, as significant dengue outbreaks can only be sustained in established vector populations.

Exploring spatial and temporal variations in suitability

To compare intra-annual variation in suitability across the globe two separate maps were created summarising the $Z(T)$ and $X(T)$ time series. The number of days in a

year where $Z(T)_i > 0$ or $X(T)_i > 0$ were calculated and mapped on a continuous scale.

Finally, to further distinguish areas that have year-round temperature suitability for oviposition or transmission we summed the total values of $Z(T)$ or $Z_{ovi}(T)$ for each time step and each vector species. These annual totals were then divided by the maximum pixel value for each species to give a relative index between 0 and 1 that can be compared across geographic areas, but not between species. To compare temperature suitability between species total values were normalised relative to the maximum value in either species.

Weekly average animations were also created to display intra-annual variation in temperature suitability. Animations were created for constrained $X_{ovi}(T)$, $Z(T)$ and $X(T)$ indices and are available in Additional files 3, 4, 5, 6, 7 and 8.

Quantifying uncertainty

To quantify uncertainty in our temperature suitability predictions, uncertainty in the individual temperature-based relationships parameterised from empirical data was propagated through the model using a Monte-Carlo method. Global pixels were divided into 10000 bins using k-means clustering based on their annual temperature profiles then, for each bin the pixel closest to the mean temperature regime of all pixels within the bin was selected as representative. The simulation model was re-run 100 times on the 10000 representative pixels but each time using a different set of parameters to define the temperature-based relationships. For each model run the parameters from one of the 200 bootstrap models used to fit the data [10] (selected at random) was used to form the temperature-vector survival relationship. For the temperature-EIP relationship for *Ae. aegypti* we took one sample at random from the posterior distribution of the parameters directly from the MCMC chain which gave us one set of model parameters. For the relative EIP differences between *Ae. aegypti* and *Ae. albopictus*, one sample from the posterior distribution of ϕ was selected at random from the MCMC chains. Parameter estimates for the temperature-time to oviposition relationship were obtained by randomly sampling from a multivariate normal distribution with mean vector given by the optimal parameters and covariance matrix given by the negative inverse Hessian matrix estimated during optimisation. The model for each species was run 100 times with these randomly selected parameters and uncertainty was quantified as the interquartile range of the predictions. The uncertainty analysis was performed at 20 km × 20 km resolution rather than the full 5 km × 5 km resolution of the final model due to computational constraints. Note that the model is entirely based on temperature and so is not scale-dependent.

Results

Temperature-based relationships

In this analysis we considered three separate temperature-dependent relationships for each of *Ae. aegypti* and *Ae. albopictus*. The relationship between temperature and adult female survival of the two species is available elsewhere [10,54,55]. Similarly, temperature-dependent EIP in *Ae. aegypti* has been modelled previously using data from natural transmission experiments [11], however the corresponding relationship for *Ae. albopictus* was calculated for the first time using assay data collected in this analysis. In total we found 29 references detailing 498 assay time-point measurements that met our inclusion criteria (373 for *Ae. aegypti*, 125 for *Ae. albopictus*). Of the 29 different studies, ten tested the competence of both *Ae. aegypti* and *Ae. albopictus*. Fitting the previously described Bayesian hierarchical model we found a significant difference in mean EIP as determined by transmission experiments and mean EIP as determined by assay experiments for *Ae. aegypti* ($\alpha_{aegypti} = 1.10$, standard deviation 0.97 – 1.23), suggesting that mosquitoes become infectious before the DENV is detectable in the head tissues through standard immunofluorescent assays. The corresponding difference between mean EIP for *Ae. aegypti*, as determined by transmission experiments, and mean EIP for *Ae. albopictus*, as determined by assay experiments, was slightly higher ($\alpha_{albopictus} = 1.17$, sd 1.03 – 1.30). Taking the difference between the two kinds of experiment into account, mean EIP for *Ae. albopictus* was found to only be 1.03 times longer than for *Ae. aegypti*. This equates to a very similar relative efficiency (ϕ) of 0.968 (0.95-0.98), which was used to rescale the temperature-EIP relationship, as determined by transmission experiments, to predict EIP for *Ae. albopictus*. We found the variance for the rate of EIP accumulation to be slightly higher for *Ae. aegypti*. This would mean infectious mosquitoes would be seen earlier, but complete infection of the cohort would be observed later, with the mean time to infection very similar to that of *Ae. albopictus* (Figure 1). The study effect in all analyses was found to be significant, indicating the high variability in measurement of mean EIP when using different experimental designs and different mosquito and viral strains.

The relationship between temperature and FGC length was defined by optimising an existing enzyme kinetics equation to new data. A total of 13 references contained 54 experiments that timed the mean length of the FGC (divided equally between *Ae. aegypti* and *Ae. albopictus*). The optimised enzyme kinetics equation for each species is shown in Figure 2. The model predicts broadly similar FGC length for the two species with *Ae. albopictus* having a slightly longer FGC length as well as higher uncertainty in its estimation.

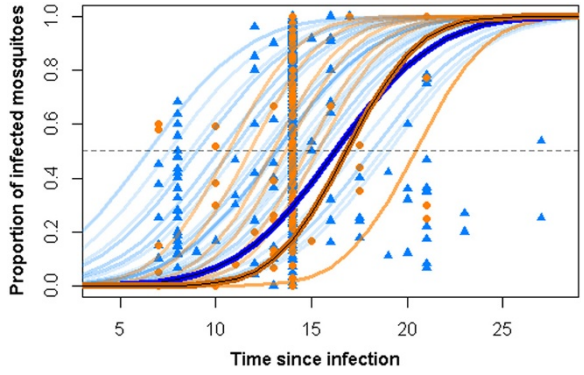


Figure 1 Rate of incubation in *Ae. aegypti* (blue) and *Ae. albopictus* (orange). The data from each individual assay are shown by blue triangles for *Ae. aegypti* (n = 373) and orange circles for *Ae. albopictus* (n = 125). The fitted lines for rate of EIP completion for each experiment are shown in faded blue lines for *Ae. aegypti* and faded orange lines for *Ae. alboipictus*. The mean fit for each species is shown in thick solid blue (*Ae. aegypti*) and orange/brown (*Ae. albopictus*) lines. The black dotted line indicates the point at which 50% of the population will have completed incubation, which is equivilent to the mean EIP.

Temperature suitability for oviposition

The thermal limits of persistence for each species were defined by comparing the predicted number of days in a year where $X_{ovi}(T)_i > 0$ at the occurrence points where each species has been reported. For *Ae. aegypti,* 95% of occurrences had 219 or more days in an average year suitable for oviposition. The equivalent value for persistence of *Ae. albopictus* was 365 days, suggesting that temperature does not limit the persistence of *Ae. Albopictus* through inability of the adults to complete oviposition. The limits defined by these thresholds were used to constrain the geographic extent of all the predictions in Figures 3 and 4.

Within these limits, temperature suitability for oviposition (relative values of $X_{ovi}(T)$) varies considerably, both geographically and between species. For *Ae. aegypti* oviposition, temperature suitability peaks in the warmest tropical environments in the northern Amazon, sub-Saharan Africa and South East Asia but drops by up to 100 fold as conditions become more temperate. The distribution of oviposition suitability for *Ae. albopictus* is similar, but the hottest regions are predicted to be less suitable due to the vector's reduced survival at higher temperatures relative to *Ae. aegypti*. The geographic extent of *Ae. albopictus* is also increased relative to that of *Ae. aegypti*, although often with low suitability, such as in the USA. The temperature suitability index also predicts high suitability in many arid areas where neither species has been observed, such as the Sahara desert and the interior of Australia. In these areas, factors other than temperature are clearly more likely to be the limiting factors defining the absolute extent of oviposition suitability.

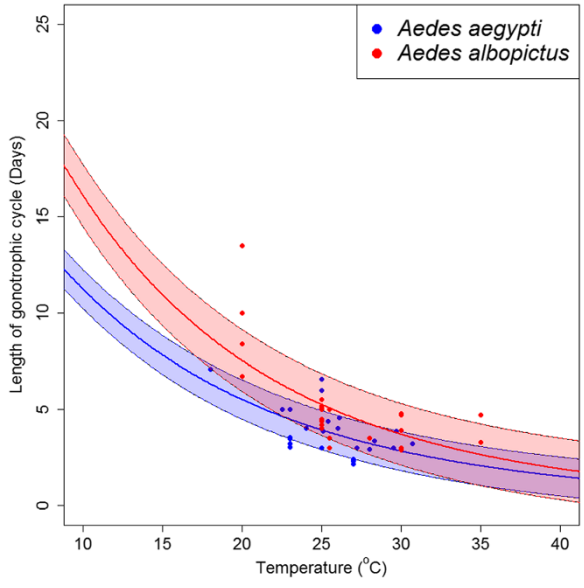


Figure 2 Relationship between temperature and length of the first gonotrophic cycle for *Ae. aegypti* and *Ae. albopictus*. The bold line shows the predicted mean first gonotrophic cycle length and the shaded areas show the prediction standard deviation for *Ae. aegypti* (blue) and *Ae. albopictus* (red).

Seasonal profile of temperature suitability

To summarise global intra-annual variation in temperature suitability two measures are presented in Figures 3B-C and 4B-C. The first summarises the number of days in an average year where $Z(T)_i > 0$. This summarises the number of days in a year where if an infectious human was introduced temperature would permit subsequent transmission, hereafter called introduction suitability, shown in Figures 3B and 4B. The latter of these indices summarises the number of days in which $X(T)_i > 0$, or if an infectious human population is omnipresent, the number of days in the year where temperature would permit an infectious mosquito population would be present, hereafter called persistence suitability, shown in Figures 3C and 4C. The former measure is more representative of a transmission free environment, while the latter is more representative of an endemic environment.

For both species, there are considerable differences between the introduction suitability and persistence suitability maps, indicating that temperature plays an important role in limiting not only the absolute geographic limits, but also in enabling different levels of endemicity. Surprisingly few environments were receptive to DENV introduction on every day of the year, which is consistent with the seasonal nature of reported dengue illness in many regions, such as southern Brazil, Mexico and China (Figures 3B-C and 4B-C). The map of introduction suitability does, however, highlight the prolonged periods of

Brady *et al. Parasites & Vectors* 2014, **7**:338
http://www.parasitesandvectors.com/content/7/1/338

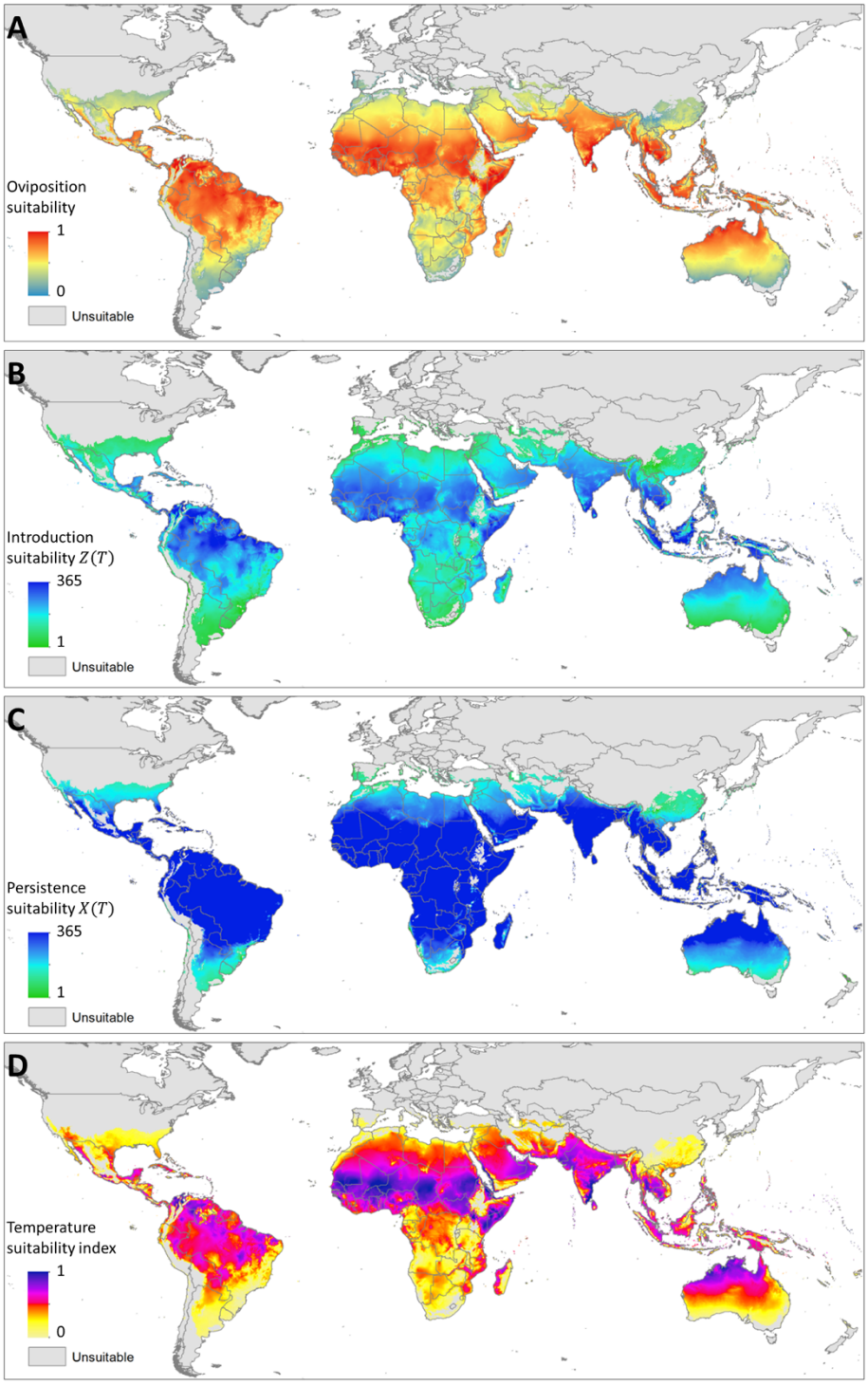


Figure 3 *Ae. aegypti* temperature suitability for persistence and DENV transmission. (A) The annualised summary of temperature suitability for oviposition ($X_{ovi}(T)$) on a normalised scale. **(B)** Introduction suitability; the number of days in a year where introduction of a DENV infected human would lead to ongoing transmission ($Z(T)_i > 0$). **(C)** Persistence suitability; the number of days in the year where onward DENV transmission could occur if a constant source of infectious humans were available ($X(T)_i > 0$). **(D)** The annualised summary of temperature suitability ($X(T)$) on a normalised scale. Predictions in all above maps are constrained to areas that permit oviposition ($X_{ovi}(T)_i > 0$) on 219 or more days in the year, as determined by comparison with known occurrences of *Ae.aegypti*.

Brady et al. Parasites & Vectors 2014, 7:338
http://www.parasitesandvectors.com/content/7/1/338

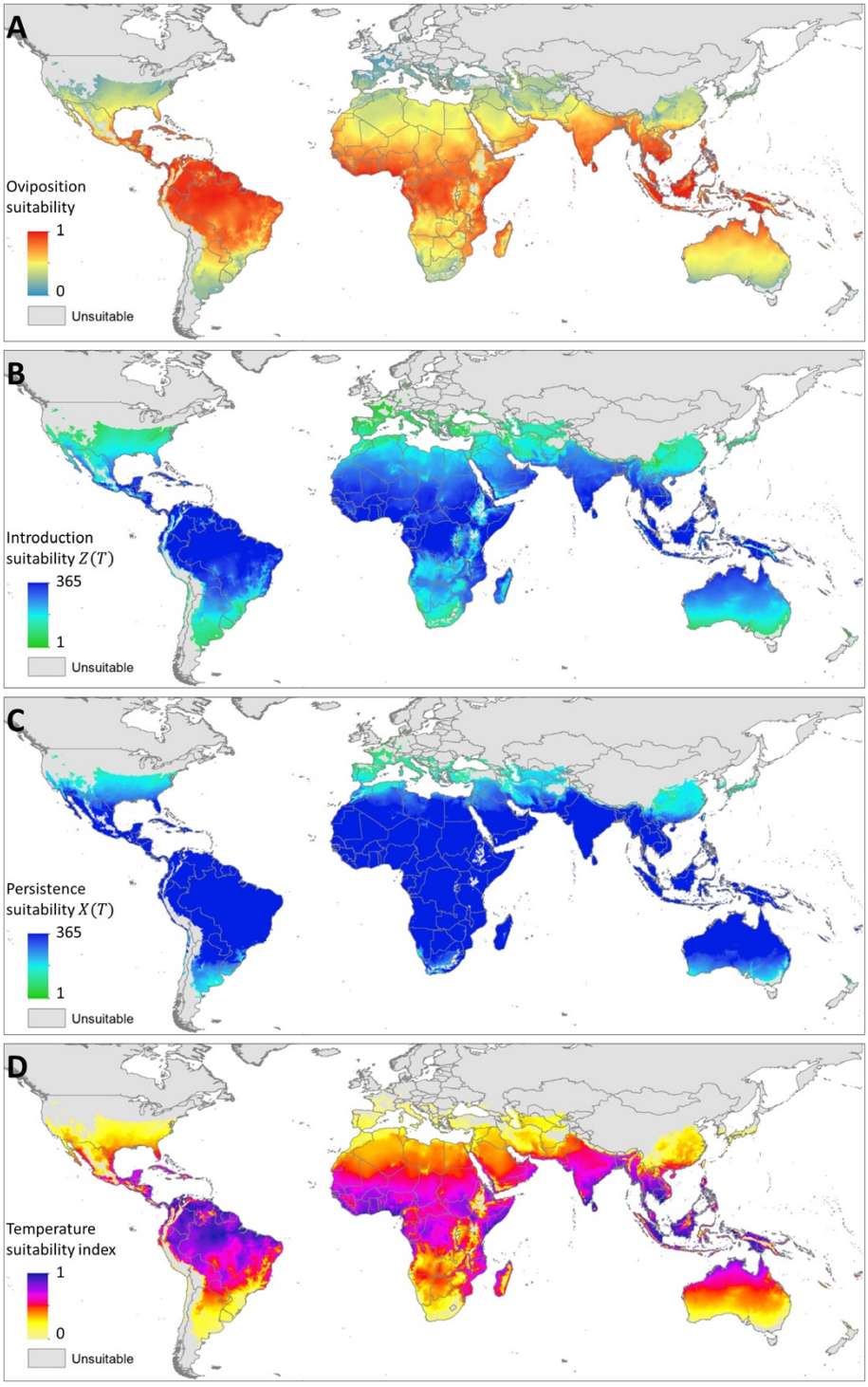


Figure 4 *Ae. albopictus* temperature suitability for persistence and DENV transmission. Panels correspond directly to those described for *Ae. aegypti* in Figure 3, but constraints are expanded to areas that permit oviposition for 365 days in the year.

suitability in areas where dengue is presently underreported, such as sub-Saharan Africa [2,22,56], or has yet to be reported, such as the eastern Arabian peninsula and Australia's Northern Territory. In such areas, if suitable mosquito breeding habitat and overlapping susceptible human populations were present, our results indicate that the potential for DENV introduction would be high (Figures 3D and 4D) and distributed throughout the year

(Figures 3B and 4B). By contrast, many temperate environments have a limited window of transmission, particularly in many areas of North America and Europe. The predicted distribution of persistence suitability is far less heterogeneous with large portions of the world where onward transmission would be enabled on every day of the year given sufficient reservoirs of infectious humans. At the fringes of the distribution, although the window of transmission is expected to be longer than that of introduction suitability, the transmission season is still predicted to be significantly restricted. This indicates that even if mosquitoes in these areas were frequently exposed to infectious humans through either autochthonous or imported cases, we expect temperature would prevent year-round transmission and thus prevent the area from becoming endemic. The limits temperature places on endemicity can be identified by areas that allow transmission for 365 days minus the incubation period in humans and the duration of human infectiousness (typically 3-10 days [11] and 2-5 days [57] respectively).

The maps of both introduction and persistence suitability for *Ae. aegypti* and *Ae. albopictus* show a similar distribution. This is consistent with DENV EIP being a main limiting factor to the seasonal extent of DENV transmission. Although, DENV EIP is slightly longer in *Ae. albopictus*, adult females are predicted to live longer leading to similar seasonal profiles. The only place in which the two predictions differ is at the fringes of transmission where the wider predicted distribution of *Ae. albopictus* could allow a very limited transmission season at higher latitudes.

Total suitability throughout the year

To summarise total suitability across the year $Z(T)$ values for each time step were summed then normalised relative to the maximum global pixel value. This index provides discrimination of the effects of suitability between environments that are suitable throughout the year and also highlights the contribution of environments that have shorter, but much more pronounced periods of suitability. Temperature suitability for *Ae. aegypti* peaks in the warmest regions of the world, particularly in the sub-Saharan region and southern India. By contrast, *Ae. albopictus* temperature suitability peaks in northern South America where the diurnal temperature fluctuation is less pronounced and thus more favourable to *Ae. albopictus* survival at extreme temperatures. For both species, temperature suitability in the most suitable regions (typically tropical regions) is 100-1000 times the value in the least suitable regions (typically temperate regions). Alternatively, given equivalent conditions temperate regions would need 100-1000 times the number of mosquitoes to sustain equivalent levels of transmission. Predicted weekly changes in the temperature and oviposition suitability indices throughout an average year are also shown in animations in Additional files 3, 4, 5, 6, 7 and 8.

Prediction uncertainty

Additional files 9 and 10 show the interquartile range in predictions for the temperature suitability index and oviposition suitability for both species. Uncertainty predictions of the temperature suitability index for *Ae. aegypti* (Additional file 9A) scales approximately linearly with the initial prediction (Figure 3D), while in the corresponding measure for *Ae. albopictus* (Additional file 10A) uncertainty values level off at high and low temperature suitability index extremes (Figure 4D). Uncertainty in predicting oviposition suitability for *Ae. aegypti* (Additional file 9B) is high in all but low suitability regions (Figure 3A), while uncertainty is highest in the mid oviposition suitability (Additional file 10B, Figure 4A) regions in the equivalent measure for *Ae. albopictus*. Average uncertainty in predicting the temperature suitability index for both species is similar, however uncertainty for oviposition suitability is lower for *Ae. albopictus* likely due to the longer lifespan reducing the overall effect of uncertainty in FGC length.

Relative role of *Ae. aegypti* and *Ae. albopictus*

Figure 5 shows the temperature suitability index displayed in Figures 3D and 4D, but with pixel values normalised to the maximum value of the temperature suitability index for *Ae. albopictus* and re-plotted on a log-scale. On average the temperature suitability index for *Ae. albopictus* is around 42 times higher than for *Ae. aegypti*, meaning that suitability is predicted to be far higher for *Ae. albopictus*. Even the most suitable regions for *Ae. aegypti* DENV transmission have equivalent temperature suitability values as relatively marginal areas for *Ae. albopictus* DENV transmission, for example temperature suitability for *Ae. aegypti* in the tropical regions of Brazil is equivalent to temperature suitability for *Ae. albopictus* in northern France (Figure 5). This difference is due to the much longer survival times of adult female *Ae. albopictus* which far offsets the minor differences in vector competence. The relative values of the temperature suitability for oviposition index also show the same degree of disparity between the two species with *Ae. albopictus* able to produce on average 33 times the population of parous females.

Discussion

In this study we used a range of biologically relevant temperature-based relationships in a dynamic model framework to evaluate the effects of temperature on DENV transmission on a global scale. For the first time we examined the global limits temperature places on DENV transmission taking into account seasonal and diurnal variations in temperature and their cumulative effects on

Brady et al. Parasites & Vectors 2014, 7:338
http://www.parasitesandvectors.com/content/7/1/338

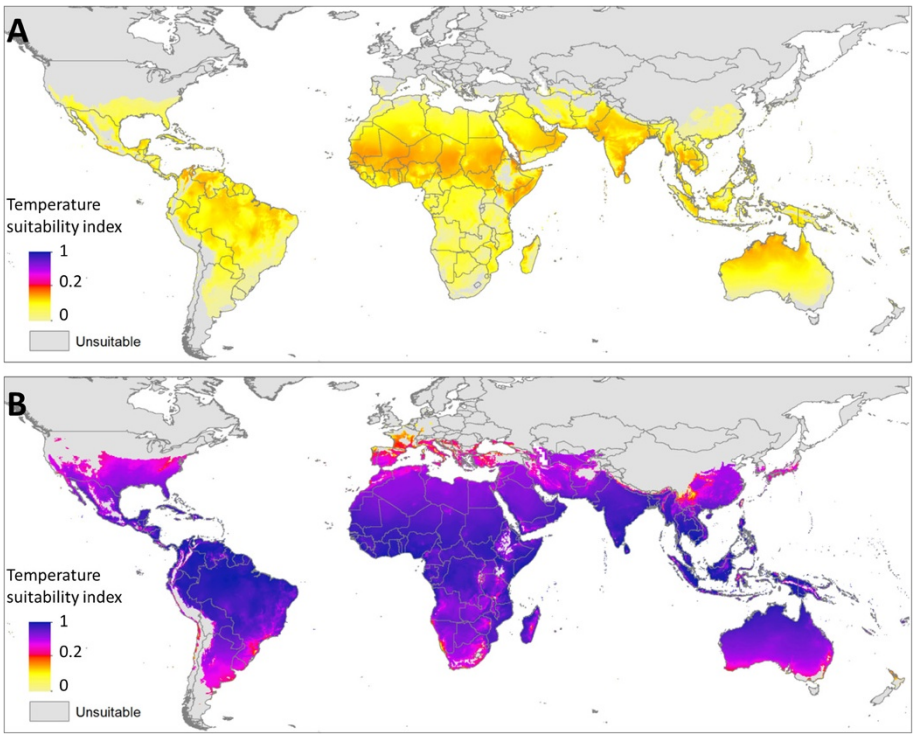


Figure 5 Comparative temperature suitability of *Ae. aegypti* (A) and *Ae. albopictus* (B). The annualised temperature suitability index ($X(T)$) normalised relative to the maximum value of both species and plotted on a logarithmic scale.

mosquito cohorts and present explicit high spatial and temporal resolution predictions. We also show that temperature has the potential to limit different levels of DENV endemicity and map the predicted temperature constraints in both transmission free environments and highly endemic environments and suggest that the true limits for other endemicity settings are likely to lie between the two. Finally, we conclude that when considering the currently available data from existing competence experiments, our analysis of temperature and its effects on DENV development and mosquito survival indicates that DENV transmission in *Ae. albopictus* is only marginally less efficient than in *Ae. aegypti* and when considering the increased longevity of the former, we predict that *Ae. albopictus* has the potential to be a more efficient DENV vector, even if this potential is not fully realised in its current niche.

The prediction that *Ae. albopictus* has a higher potential for DENV transmission than *Ae. aegypti* is at odds with the current understanding of the global epidemiology of DENV where sustained high incidence and explosive outbreaks are rarely observed in locations where *Ae. albopictus* is the sole vector. This was previously thought to be at least partly due to reduced competence of *Ae. albopictus* (reduced rates of virus dissemination from the midgut to the salivary glands) [15,58]. Our results indicate, however, that by incorporating the full breadth of data from previous experiments over a wide range of mosquito and virus strains and experimental designs, competence of *Ae. albopictus* was similar to that of *Ae. aegypti*. This small difference becomes even less meaningful when incorporated into a DENV transmission model due to the longer lifespan, and thus longer time to become and remain infectious, of *Ae. albopictus*. By assuming a normally distributed probability of EIP completion we assumed that given enough time, every infected mosquito would progress to becoming infectious and the inter-species difference is associated with the efficiency with which DENV can cross the mid-gut barrier and disseminate to the salivary glands. Although the existence of species-specific immune responses [59], protective gut microbiota [59,60], protective extreme temperatures [32] or *Wolbachia* [61] may block transmission, this would be challenging to demonstrate experimentally and full recovery from infection before onward transmission seems an unlikely outcome. Irrespective of whether this phenomenon occurs, the modelling approach effectively accounts for recovered mosquitoes in the data by predicting an EIP that exceeds their lifespan, therefore modelling the progress of DENV infection in the mosquito through EIP has the ability to incorporate a variety of species-specific aspects of competence.

Assuming EIP and adult survival are modelled correctly, species-specific differences in vectorial capacity can be accounted for by differences in i) the ratio of mosquitoes to humans, which seems unlikely to be driven by inter-species differences in mosquito population size given similar survey rates [62-64], but may be affected by shifts in habitat preferences that increase the degree of mosquito-human population overlap, ii) features of vector competence that are not accounted for in current study designs [65-67] or iii) differences in the human blood feeding rate. Differences in blood feeding behaviour and the degree of mosquito-human population overlap are perhaps the most plausible explanations, given that *Ae. albopictus* typically either inhabits more rural environments (or in urban environments often in much less close human association than the highly peridomestic *Ae. aegypti* [26,64]) and that *Ae. aegypti* has a well documented tendency to feed frequently on human hosts [68,69]. This has the effect of reducing the mosquito-human ratio, but also affects the source of potential blood meals, with *Ae. albopictus* known to feed opportunistically on a variety of non-human hosts [24]. It has been shown, however, that *Ae. albopictus* is capable of both inhabiting high density urban environments and will sometimes feed frequently and almost exclusively on human blood [23,70,71]. Although specifically designed comparative studies between *Ae. aegypti* and *Ae. albopictus* that control for the underlying mechanisms are needed to further clarify differences in their DENV vector competence [65-67], our results would suggest that the capacity of *Ae. albopictus* to transmit DENV is primarily driven by some aspect of its ecology that affects how closely *Ae. albopictus* habitats align with human population densities and how frequently they feed on human blood.

Further experiments need to be conducted on temporal and environmental variation in the ratio of mosquitoes to humans and human blood feeding habits of *Ae. albopictus* to explore these hypotheses further. Given the second order effect of human blood feeding rate on vectorial capacity, and given a conservative crude estimate that the mosquito-human proportion is halved for *Ae. albopictus,* we would expect an *Ae. albopictus* human blood feeding rate of less than $\frac{2}{\sqrt{42}} = 0.31$ times that of *Ae. aegypti*. This has been observed in some settings [23,72]. We do not mean to assert that *Ae. albopictus* is an unrecognised and underemphasised vector of DENV, but rather suggest that its involvement in transmission may be locally dependent and that further entomological investigation into human blood feeding rate and the overlap between mosquito and human populations needs to be undertaken to better understand its relative role in DENV transmission.

The results presented here apply generalised relationships to long-term average temperature data to predict the constraint temperature places on DENV transmission. While our predictions are likely to be representative of the majority of global environments most of the time, local adaptations and abnormal weather patterns may explain temporary deviations from predicted temperature constraints. It has been suggested that *Ae. albopictus* has adapted its immature stages to local environments to allow it to persist long inhospitable winters in Europe and the USA [73,74], however, when considering the temperature constraints solely on the adult forms, we found no difference between the temperature suitability for oviposition values attributed to *Ae. albopictus* occurrence points in its native Asian range and those in its newer invasive range. It has also been hypothesised that vector competence in *Ae. aegypti* varies geographically [75,76] and may even be influenced by co-evolution of local vector and virus populations [65,66,77]. Although we found the study effect to be highly significant in both EIP models, we were unable to distinguish the effects of experimental design from the effects of mosquito or viral strains. Previous work did not detect significant differences in EIP between DENV serotypes [11] and wide-scale differences in competence between different strains of *Ae. aegypti* are difficult to demonstrate considering the degree of variability in vector competence studies (Figure 1). Although there is some evidence of fine scale variation in competence in both mosquito [78-80] and virus [81,82] populations, more research is needed to determine if such structure exists at a wider scale considering human assisted rapid movement of vector and viruses may minimise opportunities for evolutionary interactions [83,84]. Until wider spatial scale structuring of mosquito-virus interaction can be demonstrated, a generalised prediction using vector and virus strains from a variety of different conditions is most likely to be representative of the majority of transmission environments. It is possible that behavioural adaptations to the micro-climates of the places where they rest or avoid extremes may allow adult mosquitoes to minimise the effects of harsh temperatures. While this may be possible in some local settings [52], protection is likely to be limited and environmental stochasticity is likely to minimise opportunities for permanent population establishment and subsequent DENV transmission. As such, we find no strong evidence that any local adaptations of the adult forms are directly temperature dependent and as a result, the predictions made here are likely to be indicative of the long-term environmental average.

The predictions of the limits temperature places on persistence and transmission broadly align with the known distribution of each species and the global distribution of DENV transmission (see Additional file 2, [1,2]). Discontinuities between the two most commonly occur in extremely arid environments, such as Australia and the Arabian Peninsula, where limited immature development

habitat (through either low precipitation or minimal open human water storage) prevents the establishment of permanent vector populations and where humidity may be an important additional limiting factor [85]. We also predict brief periods of low risk in some temperate environments that have yet to experience DENV transmission. This approach predicts the absolute limits temperature places on DENV transmission, however, in some low temperature suitability environments, the mosquito population size required to reach a certain threshold vectorial capacity may be biologically implausibly high, thus further limiting the constraints temperature places on global DENV transmission. Alternatively, DENV transmission may have yet to expand to these new areas to realise its full niche, or additional factors other than temperature, particularly humidity, may be providing additional limits to transmission. To fully delineate the temperature-permitted niche, the effects of temperature on aspects of immature mosquito population dynamics and other aspects of transmission would need to be predicted.

In this manuscript we predict the absolute limits temperature places on global expansion of DENV, but we do not evaluate what is driving its expansion. Further analyses need to be conducted to identify how temperature interacts with other environmental and genetic factors to limit mosquito vector and virus distributions. This kind of analysis could be done by studying areas that have undergone recent DENV expansion and test whether this coincides with peak periods of risk as determined by temperature suitability along with relevant intra-annual changes in other candidate covariates. This would bring about a greater quantitative understanding of global DENV expansion and allow preventative measures to be put in place in areas and times of peak or elevated risk.

By comparison with the global distribution of dengue we have revealed that while temperature provides key constraints on global DENV transmission it is not the only factor limiting DENV transmission on a global scale. Despite this, many attempts to model the current and future global distribution of dengue and its vectors have predominantly or exclusively relied on temperature as a predictor [86-91]. Through a detailed biological consideration of the way in which temperature acts, it is possible to better inform statistical approaches in two key ways: i) constraining predictions to eliminate areas where temperature does not permit transmission and ii) including this information as a more biologically relevant covariate than standard temperature measures, but always in addition to other covariates to explain the complex epidemiology of DENV transmission. Combining mechanistic and statistical approaches will allow us to utilise the wealth of theoretical and experimental knowledge of the epidemiology of DENV transmission in generalised analyses [1,40,92].

While in this manuscript the temperature suitability index has been summarised in a broadly spatial analysis, the same index could be used in models that analyse time series data of vectors or dengue cases using a similar method. Such biological consideration is essential if we are to extrapolate past or current conditions to make predictions about the future, such as in projections of the future distribution of dengue or in dengue early warning systems. Temperature suitability is just one example of this and consideration of further biological limitations, such as geographic spread of dengue's vectors, the role of other limiting factors such as humidity and the rate of dengue expansion need to be considered if projections of the future distribution of dengue are to be more reliable and useful.

The predictions from this analysis suggest that current day temperature regimes permit both vector establishment and DENV transmission beyond their current ranges. While factors other than temperature are likely to further limit the distribution of dengue and its vectors, if expansion is to occur, temperature suitability can elucidate the regions and suitable seasons most at risk. For *Ae. aegypti* and/or *Ae. albopictus* free areas, the introduction suitability ($Z_{ovi}(T)$) maps and annual profiles can be used to direct surveillance to the regions and times of the year where a single mosquito introduction would carry the highest risk of population persistence. For areas with year-round populations of *Ae. aegypti*, but without persistent DENV transmission, the DENV introduction ($Z(T)$) maps and annual profiles can be used to guide vector control to limit periods and areas of peak risk. For areas with persistent dengue transmission (endemic) the annual profiles of persistence suitability ($X(T)$) can be used to direct control towards the time of year that is most likely to break the on-going transmission cycle. Finally, in areas with established *Ae. albopictus* populations, their potential for DENV transmission can be evaluated through a combination of $Z(T)$ and $X(T)$ indices that elucidate areas and times of peak risk which can then guide further entomological investigation regarding human blood feeding rate and mosquito-human population overlap, thereby further improving estimates of DENV vectorial capacity of this species. This combination of high spatial and temporal resolutions of risk estimates coupled with guided local entomological investigation is likely to give a more complete understanding of future DENV expansion risk than a consideration of the risk imposed by extrapolated niche approaches alone [86,87,89-91].

Conclusions

Here we used a range of entomological and virological experimental data to parameterise temperature-based relationships in a dynamic transmission model to assess the limits that changing temperature regimes place on

Ae. aegypti and *Ae. albopictus* persistence as well as the global distribution of DENV transmission. The model predictions were used to map the absolute and seasonal temperature limits to vector persistence and DENV transmission on a global scale. These maps reveal that as well as constraining the absolute extent of transmission, temperature may also limit levels of DENV endemicity. A comparison of temperature suitability for DENV transmission by *Ae. aegypti* vs. *Ae. albopictus* revealed that given similar population sizes and human blood feeding rates, *Ae. albopictus* has a higher vectorial capacity. This finding contradicts existing explanations and suggests that ecological factors, such as blood feeding patterns and the degree of mosquito-human population overlap, may explain why areas with established *Ae. albopictus* populations do not experience sustained dengue outbreaks. The resulting models can inform statistical approaches to model DENV transmission and persistence of its vectors by replacing raw temperature variables to include known biological understanding of transmission. The risk of dengue expansion into novel areas and at novel times of the year can be estimated by using the seasonal profiles presented here to target entomological investigation of human biting rate and the degree of mosquito-human population overlap. As well as being informative for surveillance purposes, this further investigation will help identify and understand additional factors limiting DENV transmission and will inform and support new ways to limit the expansion of the geographic extent and range of endemicity of dengue.

Additional files

Additional file 1: Priors used for Bayesian Inter-species EIP comparison model.

Additional file 2: Assembly of the supplementary *Ae. aegypti* and *Ae. albopictus* database.

Additional file 3: Animation of *Ae. aegypti* temperature suitability for oviposition ($X_{ovi}(T)$) throughout an average year.

Additonal file 4: Animation of *Ae. albopictus* temperature suitability for oviposition ($X_{ovi}(T)$) throughout an average year.

Additional file 5: Animation of *Ae. aegypti* temperature suitability for introduction of dengue virus transmission ($Z(T)$) throughout an average year.

Additional file 6: Animation of *Ae. albopictus* temperature suitability for introduction of dengue virus transmission ($Z(T)$) throughout an average year.

Additional file 7: Animation of *Ae. aegypti* temperature suitability for persistence of dengue virus transmission ($X(T)$) throughout an average year.

Additional file 8: Animation of *Ae. albopictus* temperature suitability for persistence of dengue virus transmission ($X(T)$) throughout an average year.

Additional file 9: *Ae. aegypti* model prediction uncertainty for the temperature suitability index (A) and oviposition suitability (B). The output shows the interquartile range in predictions presented in maps 3A and 3D.

Additional file 10: *Ae. albopictus* model prediction uncertainty for the temperature suitability index (A) and oviposition suitability (B). The output shows the interquartile range in predictions presented in maps 4A and 4D.

Abbreviations
EIP: Extrinsic incubation period; DENV: Dengue virus; FGC: First gonotrophic cycle.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
OJB, DLS, PWG and SIH designed the experiment. OJB, wrote the manuscript and collected and analysed the data. MUK collected additional data. NG, DLS and MUK helped with data analysis. All authors helped with data and results interpretation and were involved in drafting, revising and approving the final version of the manuscript.

Acknowledgements
OJB is funded by a BBSRC studentship. PWG is a Medical Research Council (UK) Career Development Fellow (#K00669X) and receives support from the Bill and Melinda Gates Foundation (#OPP1068048). NG is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). SIH is funded by a Senior Research Fellowship from the Wellcome Trust (095066). MUGK is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. This study was partially funded by EU grant 21803 IDAMS (http://www.idams.eu) and is designated with IDAMS publication reference number IDAMS22. The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission. T.W.S. acknowledges funding from the Bill & Melinda Gates Foundation (OPP52250), the Innovative Vector Control Consortium, and the NIH (R01-AI069341, R01-AI091980, and R01-GM08322). RCR, TWS, DLS and SIH also acknowledges funding support from the Research and Policy in Infectious Disease Dynamics (RAPIDD) program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

Author details
[1]Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom. [2]Department of Entomology and Nematology, University of California, Davis, CA, USA. [3]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Received: 2 May 2014 Accepted: 27 June 2014
Published: 22 July 2014

References
1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI: **The global distribution and burden of dengue.** *Nature* 2013, **496**(7446):504.
2. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, Moyes CL, Farlow AW, Scott TW, Hay SI: **Refining the global spatial limits of dengue virus transmission by evidence-based consensus.** *PLoS Negl Trop Dis* 2012, **6**(8):e1760.
3. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI: **A global compendium of human dengue virus occurrence.** *Sci Data* 2014, **1**(140004). doi:10.1038/sdata.2014.1034.
4. Gubler DJ, Clark GG: **Dengue/dengue hemorrhagic fever: the emergence of a global health problem.** *Emerg Infect Dis* 1995, **1**(2):55–57.
5. Holmes EC, Twiddy SS: **The origin, emergence and evolutionary genetics of dengue virus.** *Infect Genet Evol* 2003, **3**(1):19–28.
6. Cuong HQ, Hien NT, Duong TN, Phong TV, Cam NN, Farrar J, Nam VS, Thai KT, Horby P: **Quantifying the emergence of dengue in Hanoi, Vietnam: 1998-2009.** *PLoS Negl Trop Dis* 2011, **5**(9):e1322.
7. Radke EG, Gregory CJ, Kintziger KW, Sauber-Schatz EK, Hunsperger EA, Gallagher GR, Barber JM, Biggerstaff BJ, Stanek DR, Tomashek KM, Blackmore CG: **Dengue outbreak in Key West, Florida, USA, 2009.** *Emerg Infect Dis* 2012, **18**(1):135–137.

8. Christophers R: *Aedes aegypti* (L.) the yellow fever mosquito: its life history, bionomics and structure. In 1st edition. Cambridge: Cambridge University Press; 1960.
9. Focks DA, Haile DG, Daniels E, Mount GA: Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *J Med Entomol* 1993, 30(6):1003–1017.
10. Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, Smith DL, Scott TW, Gething PW, Hay SI: Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasit Vectors* 2013, 6:351.
11. Chan M, Johansson MA: The incubation periods of dengue viruses. *PLoS One* 2012, 7(11):e50972.
12. WorldClim global climate data, Data for current conditions. http://www.worldclim.org/current.
13. Nasci R: The size of emerging and host-seeking *Aedes aegypti* and the relation of size to blood-feeding success in the field. *J Am Mosq Control Assoc* 1986, 2(1):61–62.
14. Gratz N: Critical review of the vector status of *Aedes albopictus*. *Med Vet Entomol* 2004, 18(3):215–227.
15. Lambrechts L, Scott TW, Gubler DJ: Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. *PLoS Negl Trop Dis* 2010, 4(5):e646.
16. Carbajo A, Curto S, Schweigmann N: Spatial distribution pattern of oviposition in the mosquito *Aedes aegypti* in relation to urbanization in Buenos Aires: southern fringe bionomics of an introduced vector. *Med Vet Entomol* 2006, 20(2):209–218.
17. Benedict MQ, Levine RS, Hawley WA, Lounibos LP: Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. *Vector Borne Zoonot Dis* 2007, 7(1):76–85.
18. Reiter P: *Aedes albopictus* and the world trade in used tires, 1988-1995: the shape of things to come? *J Am Mosq Control Assoc* 1998, 14(1):83–94.
19. La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, Lenglet A, Jourdain F, Leparc-Goffart I, Charlet F: First two autochthonous dengue virus infections in metropolitan France, September 2010. *Euro Surveill* 2010, 15(39):19676.
20. Effler PV, Pang L, Kitsutani P, Vorndam V, Nakata M, Ayers T, Elm J, Tom T, Reiter P, Rigau-Perez JG, Hayes JM, Mills K, Napier M, Clark GG, Gubler DJ, Hawaii Dengue Outbreak Investigation Team: Dengue fever, Hawaii, 2001-2002. *Emerg Infect Dis* 2005, 11(5):742–749.
21. ECDC: Rapid risk assessment on dengue cases in Madeira. In *European Centre for Disease Prevention and Control*. 2012. Available at: http://www.ecdc.europa.eu/en/press/news/_layouts/forms/News_DispForm.aspx?ID=568&List=8db7286c-fe2d-476c-9133-18ff4cb1b568.
22. Franco L, Di Caro A, Carletti F, Vapalahti O, Renaudat C, Zeller H, Tenorio A: Recent expansion of dengue virus serotype 3 in West Africa. *Euro Surveill* 2010, 15(7):578–583.
23. Valerio L, Marini F, Bongiorno G, Facchinelli L, Pombi M, Caputo B, Maroli M, della Torre A: Host-feeding patterns of *Aedes albopictus* (Diptera: Culicidae) in urban and rural contexts within Rome Province, Italy. *Vector Borne Zoonot Dis* 2010, 10(3):291–294.
24. Estrada-Franco JG, Craig GB: *Biology, Disease Relationships, and Control of Aedes albopictus*. Washington DC: Pan American Health Organization; 1995.
25. Ho B, Chan K, Chan Y: *Aedes aegypti* (L.) and *Aedes albopictus* (Skuse) in Singapore City: 3. Population fluctuations. *Bull World Health Organ* 1971, 44(5):635.
26. Tsuda Y, Suwonkerd W, Chawprom S, Prajakwong S, Takagi M: Different spatial distribution of *Aedes aegypti* and *Aedes albopictus* along an urban-rural gradient and the relating environmental factors examined in three villages in northern Thailand. *J Am Mosq Control Assoc* 2006, 22(2):222–228.
27. Johansson MA, Cummings DA, Glass GE: Multiyear climate variability and dengue—El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS Med* 2009, 6(11):e1000168.
28. Simoes TC, Codeço CT, Nobre AA, Eiras ÁE: Modeling the non-stationary climate dependent temporal dynamics of *Aedes aegypti*. *PLoS One* 2013, 8(8):e64773.
29. Campbell KM, Lin C, Iamsirithaworn S, Scott TW: The complex relationship between weather and dengue virus transmission in Thailand. *Am J Trop Med Hyg* 2013, 89(6):1066.
30. Nawrocki S, Hawley W: Estimation of the northern limits of distribution of *Aedes albopictus* in North America. *J Am Mosq Control Assoc* 1987, 3(2):314–317.
31. Kobayashi M, Nihei N, Kurihara T: Analysis of northern distribution of *Aedes albopictus* (Diptera: Culicidae) in Japan by geographical information system. *J Med Entomol* 2002, 39(1):4–11.
32. Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, Thomas MB, Scott TW: Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proc Natl Acad Sci U S A* 2011, 108(18):7460–7465.
33. Yang H, Macoris M, Galvani K, Andrighetti M, Wanderley D: Assessing the effects of temperature on the population of *Aedes aegypti*, the vector of dengue. *Epidemiol Infect* 2009, 137(8):1188.
34. Focks DA, Daniels E, Haile DG, Keesling JE: A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *Am J Trop Med Hyg* 1995, 53(5):489–506.
35. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI: Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit Vectors* 2011, 4:92.
36. Weiss DJ, Bhatt S, Mappin B, Van Boeckel TP, Smith DL, Hay SI, Gething PW: Air temperature suitability for *Plasmodium falciparum* malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malar J* 2014, 13(1):1–11.
37. Macdonald G: Epidemiological basis of malaria control. *Bull World Health Organ* 1956, 15(3–5):613.
38. Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, George DB, Le Menach A, Pulliam JR, Bisanzio D, Buckee C, Chiyaka C, Cummings DA, Garcia AJ, Gatton ML, Gething PW, Hartley DM, Johnston G, Klein EY, Michael E, Lindsay SW, Lloyd AL, Pigott DM, Reisen WK, Ruktanonchai N, Singh BK, Tatem AJ, Kitron U, Hay SI, Scott TW, *et al*: A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *J Royal Soc Interface* 2013, 10(81):20120921.
39. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE: Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathog* 2012, 8(4):e1002588.
40. Smith DL, Perkins TA, Reiner RC Jr, Barker CM, Niu T, Chaves LF, Ellis AM, George DB, Le Menach A, Pulliam JR, Bisanzio D, Buckee C, Chiyaka C, Cummings DA, Garcia AJ, Gatton ML, Gething PW, Hartley DM, Johnston G, Klein EY, Michael E, Lindsay SW, Lloyd AL, Pigott DM, Reisen WK, Ruktanonchai N, Singh B, Stoller J, Tatem AJ, Kitron U, *et al*: Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Trans R Soc Trop Med Hyg* 2014, 108(4):185–197.
41. Smith DL, McKenzie FE: Statics and dynamics of malaria infection in *Anopheles* mosquitoes. *Malar J* 2004, 3:13.
42. Lambrechts L, Fansiri T, Pongsiri A, Thaisomboonsuk B, Klungthong C, Richardson JH, Ponlawat A, Jarman RG, Scott TW: Dengue-1 virus clade replacement in Thailand associated with enhanced mosquito transmission. *J Virol* 2012, 86(3):1853–1861.
43. Lunn DJ, Thomas A, Best N, Spiegelhalter D: WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000, 10(4):325–337.
44. Byrd RH, Lu P, Nocedal J, Zhu C: A limited memory algorithm for bound constrained optimization. *Siam J Sci Comput* 1995, 16(5):1190–1208.
45. Parton WJ, Logan JA: A model for diurnal variation in soil and air temperature. *Agr Met* 1981, 23:205–216.
46. Carrington LB, Armijos MV, Lambrechts L, Barker CM, Scott TW: Effects of fluctuating daily temperatures at critical thermal extremes on *Aedes aegypti* life-history traits. *PLoS One* 2013, 8(3):e58824.
47. Carrington LB, Armijos MV, Lambrechts L, Scott TW: Fluctuations at a low mean temperature accelerate dengue virus transmission by *Aedes aegypti*. *PLoS Negl Trop Dis* 2013, 7(4):e2190.
48. Carrington LB, Seifert SN, Armijos MV, Lambrechts L, Scott TW: Reduction of *Aedes aegypti* vector competence for dengue virus under large temperature fluctuations. *Am J Trop Med Hyg* 2013, 88(4):689–697.
49. Carrington LB, Seifert SN, Willits NH, Lambrechts L, Scott TW: Large diurnal temperature fluctuations negatively influence *Aedes aegypti* (Diptera: Culicidae) life-history traits. *J Med Entomol* 2013, 50(1):43–51.
50. Costero A, Edman JD, Clark GG, Kittayapong P, Scott TW: Survival of starved *Aedes aegypti* (Diptera: Culicidae) in Puerto Rico and Thailand. *J Med Entomol* 1999, 36(3):272–276.
51. Guerra CA, Reiner RC, Perkins AT, Lindsay SW, Midega J, Brady OJ, Barker CM, Reisen WK, Harrington LC, Takken W, Kitron U, Lloyd AL, Hay SI, Scott TW, Smith DL: A global assembly of adult female mosquito mark-release-

recapture data to inform the control of mosquito-borne pathogens. *Parasit Vectors* 2014, **7**(1):276. doi:10.1186/1756-3305-7-276.

52. Canyon D, Hii J, Müller R: **Adaptation of *Aedes aegypti* (Diptera: Culicidae) oviposition behavior in response to humidity and diet.** *J Insect Physiol* 1999, **45**(10):959–964.
53. Hanson S, Craig G: **Cold acclimation, diapause, and geographic origin affect cold hardiness in eggs of *Aedes albopictus* (Diptera: Culicidae).** *J Med Entomol* 1994, **31**(2):192–201.
54. **Model of adult *Aedes albopictus* survival/mortality at different temperatures.** http://figshare.com/articles/Model_of_adult_Aedes_albopictus_survival_mortality_at_different_temperatures/865035.
55. **Model of adult *Aedes aegypti* survival/mortality at different temperatures.** http://figshare.com/articles/Model_of_adult_Aedes_aegypti_survival_mortality_at_different_temperatures/865034.
56. Vairo F, Nicastri E, Meschi S, Schepisi MS, Paglia MG, Bevilacqua N, Mangi S, Sciarrone MR, Chiappini R, Mohamed J, Racalbuto V, Di Caro A, Capobianchi MR, Ippolito G: **Seroprevalence of dengue infection: a cross-sectional survey in mainland Tanzania and on Pemba Island, Zanzibar.** *Int J Infect Dis* 2011, **16**(1):1–2.
57. Nguyen NM, Kien DTH, Tuan TV, Quyen NTH, Tran CN, Thi LV, Le Thi D, Nguyen HL, Farrar JJ, Holmes EC: **Host and viral features of human dengue cases shape the population of infected and infectious *Aedes aegypti* mosquitoes.** *Proc Natl Acad Sci U S A* 2013, **110**(22):9072–9077.
58. Chen WJ, Wei HL, Hsu EL, Chen ER: **Vector competence of *Aedes albopictus* and *Ae. aegypti* (Diptera: Culicidae) to dengue 1 virus on Taiwan: development of the virus in orally and parenterally infected mosquitoes.** *J Med Entomol* 1993, **30**(3):524–530.
59. Xi Z, Ramirez JL, Dimopoulos G: **The *Aedes aegypti* toll pathway controls dengue virus infection.** *PLoS Pathog* 2008, **4**(7):e1000098.
60. Ramirez JL, Souza-Neto J, Cosme RT, Rovira J, Ortiz A, Pascale JM, Dimopoulos G: **Reciprocal tripartite interactions between the *Aedes aegypti* midgut microbiota, innate immune system and dengue virus influences vector competence.** *PLoS Negl Trop Dis* 2012, **6**(3):e1561.
61. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M: **A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, chikungunya, and *Plasmodium*.** *Cell* 2009, **139**(7):1268–1278.
62. Mogi M, Khamboonruang C, Choochote W, Suwanpanti P: **Ovitrap surveys of dengue vector mosquitoes in Chiang Mai, northern Thailand: seasonal shifts in relative abundance of *Aedes albopictus* and *Ae. aegypti*.** *Med Vet Entomol* 1988, **2**(4):319–324.
63. Yap H: **Distribution of *Aedes aegypti* (Linnaeus) and *Aedes albopictus* (Skuse) in small towns and villages of Penang Island, Malaysia–an ovitrap survey.** *Southeast Asian J Trop Med Public Health* 1975, **6**(4):519–524.
64. Braks MA, Honório NA, Lourenço-De-Oliveira R, Juliano SA, Lounibos LP: **Convergent habitat segregation of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in southeastern Brazil and Florida.** *J Med Entomol* 2003, **40**(6):785–794.
65. Lambrechts L, Chevillon C, Albright RG, Thaisomboonsuk B, Richardson JH, Jarman RG, Scott TW: **Genetic specificity and potential for local adaptation between dengue viruses and mosquito vectors.** *BMC Evol Biol* 2009, **9**(1):160.
66. Lambrechts L, Quillery E, Noël V, Richardson JH, Jarman RG, Scott TW, Chevillon C: **Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene *Dicer-2*.** *Proc R Soc B* 2013, **280**(1751):20122437.
67. Fansiri T, Fontaine A, Diancourt L, Caro V, Thaisomboonsuk B, Richardson JH, Jarman RG, Ponlawat A, Lambrechts L: **Genetic mapping of specific interactions between *Aedes aegypti* mosquitoes and dengue viruses.** *PLoS Genet* 2013, **9**(8):e1003621.
68. Scott TW, Clark GG, Amerasinghe PH, Lorenz LH, Reiter P, Edman JD: **Detection of multiple blood feeding patterns in *Aedes aegypti* (Diptera: Culicidae) during a single gonotrophic cycle using a histological technique.** *J Med Entomol* 1993, **30:**94–99.
69. Scott TW, Amerasinghe PH, Morrison AC, Lorenz LH, Clark GG, Strickman D, Kittayapong P, Edman JD: **Longitudinal studies of *Aedes aegypti* (Diptera: *Culicidae*) in Thailand and Puerto Rico: blood feeding frequency.** *J Med Entomol* 2000, **37**(1):89–101.
70. Delatte H, Desvars A, Bouétard A, Bord S, Gimonneau G, Vourc'h G, Fontenille D: **Blood-feeding behavior of *Aedes albopictus*, a vector of Chikungunya on La Réunion.** *Vector Borne Zoonot Dis* 2010, **10**(3):249–258.
71. Ponlawat A, Harrington LC: **Blood feeding patterns of *Aedes aegypti* and *Aedes albopictus* in Thailand.** *J Med Entomol* 2005, **42**(5):844–849.
72. Richards SL, Ponnusamy L, Unnasch TR, Hassan HK, Apperson CS: **Host-feeding patterns of *Aedes albopictus* (Diptera: Culicidae) in relation to availability of human and domestic animals in suburban landscapes of central North Carolina.** *J Med Entomol* 2006, **43**(3):543.
73. Romi R, Severini F, Toma L: **Cold acclimation and overwintering of female *Aedes albopictus* in Roma.** *J Am Mosq Control Assoc* 2006, **22**(1):149–151.
74. Hawley W, Pumpuni C, Brady R, Craig G: **Overwintering survival of *Aedes albopictus* (Diptera: Culicidae) eggs in Indiana.** *J Med Entomol* 1989, **26**(2):122–129.
75. Bennett KE, Olson KE, Munoz Mde L, Fernandez-Salas I, Farfan-Ale JA, Higgs S, Black WC, Beaty BJ: **Variation in vector competence for dengue 2 virus among 24 collections of *Aedes aegypti* from Mexico and the United States.** *Am J Trop Med Hyg* 2002, **67**(1):85–92.
76. Boromisa RD, Rai KS, Grimstad PR: **Variation in the vector competence of geographic strains of *Aedes albopictus* for dengue 1 virus.** *J Am Mosq Control Assoc* 1987, **3**(3):378–386.
77. Failloux A, Vazeille-Falcoz M, Mousson L, Rodhain F: **Genetic control of vectorial competence in *Aedes* mosquitoes.** *Bull Soc Pathol Exot Filiales* 1998, **92**(4):266–273.
78. Bosio CF, Harrington LC, Jones JW, Sithiprasasna R, Norris DE, Scott TW: **Genetic structure of *Aedes aegypti* populations in Thailand using mitochondrial DNA.** *Am J Trop Med Hyg* 2005, **72**(4):434–442.
79. Brown JE, McBride CS, Johnson P, Ritchie S, Paupy C, Bossin H, Lutomiah J, Fernandez-Salas I, Ponlawat A, Cornel AJ: **Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases.** *Proc R Soc B* 2011, **278**(1717):2446–2454.
80. Lozano-Fuentes S, Fernandez-Salas I, de Lourdes MM, Garcia-Rejon J, Olson KE, Beaty BJ, Black WC IV: **The neovolcanic axis is a barrier to gene flow among *Aedes aegypti* populations in mexico that differ in vector competence for dengue 2 virus.** *PLoS Negl Trop Dis* 2009, **3**(6):e468.
81. Jarman RG, Holmes EC, Rodpradit P, Klungthong C, Gibbons RV, Nisalak A, Rothman AL, Libraty DH, Ennis FA, Mammen MP: **Microevolution of Dengue viruses circulating among primary school children in Kamphaeng Phet.** *Thailand J Virol* 2008, **82**(11):5494–5500.
82. Rabaa MA, Ty Hang VT, Wills B, Farrar J, Simmons CP, Holmes EC: **Phylogeography of recently emerged DENV-2 in southern Viet Nam.** *PLoS Negl Trop Dis* 2010, **4**(7):e766.
83. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, Bhatt S, Katzelnick L, Howes RE, Battle KE: **Global spread of dengue virus types: mapping the 70 year history.** *Trends Microbiol* 2014, **22**(3):138–146.
84. Tatem AJ, Hay SI, Rogers DJ: **Global traffic and disease vector dispersal.** *Proc Natl Acad Sci U S A* 2006, **103**(16):6242–6247.
85. Hales S, de Wet N, Maindonald J, Woodward A: **Potential effect of population and climate changes on global distribution of dengue fever: an empirical model.** *Lancet* 2002, **360**(9336):830–834.
86. Patz JA, Martens WJ, Focks DA, Jetten TH: **Dengue fever epidemic potential as projected by general circulation models of global climate change.** *Environ Health Perspect* 1998, **106**(3):147–153.
87. Jetten TH, Focks DA: **Potential changes in the distribution of dengue transmission under climate warming.** *Am J Trop Med Hyg* 1997, **57**(3):285–297.
88. Rogers DJ, Wilson AJ, Hay SI, Graham AJ: **The global distribution of yellow fever and dengue.** *Adv Parasitol* 2006, **62:**181–220.
89. European Centre for Disease Prevention and Control: *The climatic suitability for dengue transmission in continental Europe.* Stockholm: ECDC; 2012.
90. Fischer D, Thomas SM, Niemitz F, Reineking B, Beierkuhnlein C: **Projection of climatic suitability for *Aedes albopictus* Skuse (Culicidae) in Europe under climate change conditions.** *Global Planet Change* 2011, **78**(1):54–64.
91. Caminade C, Medlock JM, Ducheyne E, McIntyre KM, Leach S, Baylis M, Morse AP: **Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios.** *J Royal Soc Interface* 2012, **9**(75):2708–2717.
92. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, Tatem AJ, Hay SI: **A new world malaria map: *Plasmodium falciparum* endemicity in 2010.** *Malar J* 2011, **10:**378.

doi:10.1186/1756-3305-7-338
Cite this article as: Brady *et al.*: Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasites & Vectors* 2014 **7**:338.

ZIKA VIRUS OUTBREAK

Zika virus in the Americas: Early epidemiological and genetic findings

Nuno Rodrigues Faria,[1,2]* Raimunda do Socorro da Silva Azevedo,[3]* Moritz U. G. Kraemer,[2] Renato Souza,[4] Mariana Sequetin Cunha,[4] Sarah C. Hill,[2] Julien Thézé,[2] Michael B. Bonsall,[2] Thomas A. Bowden,[5] Ilona Rissanen,[5] Iray Maria Rocco,[4] Juliana Silva Nogueira,[4] Adriana Yurika Maeda,[4] Fernanda Giseli da Silva Vasami,[4] Fernando Luiz de Lima Macedo,[4] Akemi Suzuki,[4] Sueli Guerreiro Rodrigues,[3] Ana Cecilia Ribeiro Cruz,[3] Bruno Tardeli Nunes,[3] Daniele Barbosa de Almeida Medeiros,[3] Daniela Sueli Guerreiro Rodrigues,[3] Alice Louize Nunes Queiroz,[3] Eliana Vieira Pinto da Silva,[3] Daniele Freitas Henriques,[3] Elisabeth Salbe Travassos da Rosa,[3] Consuelo Silva de Oliveira,[3] Livia Caricio Martins,[3] Helena Baldez Vasconcelos,[3] Livia Medeiros Neves Casseb,[3] Darlene de Brito Simith,[3] Jane P. Messina,[2,6] Leandro Abade,[2] José Lourenço,[2] Luiz Carlos Junior Alcantara,[7] Maricélia Maia de Lima,[8] Marta Giovanetti,[7] Simon I. Hay,[9,5] Rodrigo Santos de Oliveira,[1] Poliana da Silva Lemos,[1] Layanna Freitas de Oliveira,[1] Clayton Pereira Silva de Lima,[1] Sandro Patroca da Silva,[1] Janaina Mota de Vasconcelos,[1] Luciano Franco,[1] Jedson Ferreira Cardoso,[1] João Lídio da Silva Gonçalves Vianez-Júnior,[1] Daiana Mir,[10] Gonzalo Bello,[10] Edson Delatorre,[10] Kamran Khan,[11,12] Marisa Creatore,[13] Giovanini Evelim Coelho,[14] Wanderson Kleber de Oliveira,[14] Robert Tesh,[15] Oliver G. Pybus,[2,6]†‡ Marcio R. T. Nunes,[1,15]†‡ Pedro F. C. Vasconcelos[3]†‡

Brazil has experienced an unprecedented epidemic of Zika virus (ZIKV), with ~30,000 cases reported to date. ZIKV was first detected in Brazil in May 2015, and cases of microcephaly potentially associated with ZIKV infection were identified in November 2015. We performed next-generation sequencing to generate seven Brazilian ZIKV genomes sampled from four self-limited cases, one blood donor, one fatal adult case, and one newborn with microcephaly and congenital malformations. Results of phylogenetic and molecular clock analyses show a single introduction of ZIKV into the Americas, which we estimated to have occurred between May and December 2013, more than 12 months before the detection of ZIKV in Brazil. The estimated date of origin coincides with an increase in air passengers to Brazil from ZIKV-endemic areas, as well as with reported outbreaks in the Pacific Islands. ZIKV genomes from Brazil are phylogenetically interspersed with those from other South American and Caribbean countries. Mapping mutations onto existing structural models revealed the context of viral amino acid changes present in the outbreak lineage; however, no shared amino acid changes were found among the three currently available virus genomes from microcephaly cases. Municipality-level incidence data indicate that reports of suspected microcephaly in Brazil best correlate with ZIKV incidence around week 17 of pregnancy, although this correlation does not demonstrate causation. Our genetic description and analysis of ZIKV isolates in Brazil provide a baseline for future studies of the evolution and molecular epidemiology of this emerging virus in the Americas.

Zika virus (ZIKV) is a single-stranded, positive-sense RNA virus with a 10.7-kb genome encoding a single polyprotein that is cleaved into three structural proteins (C, prM/M, E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (*1*). ZIKV is a member of the family *Flaviviridae*, genus *Flavivirus*, and is transmitted among humans by *Aedes* mosquito species such as *A. aegypti*, *A. albopictus*, and *A. africanus*. The virus was first isolated in 1947 from a sentinel rhesus monkey in the Zika forest in Uganda (*2*) and is classified by sequence analysis into two genotypes, African and Asian (*3*). In humans, ZIKV infection typically causes a mild and self-limiting illness known as Zika fever (*4*), which is accompanied by maculopapular rash, headache, conjunctivitis, and myalgia. In April 2007, a large epidemic of Asian genotype ZIKV was reported in Yap Island and Guam, Micronesia (*5*, *6*). Between 2013 and 2014, the ZIKV Asian genotype caused epidemics reported in several Pacific Islands, including French Polynesia (*7*), New Caledonia (*8*), the Cook Islands (*9*), Tahiti (*10*), and Easter Island (*11*).

By May 2015, ZIKV was reported in Brazil (*12*) and, subsequently, in several Central and South American countries and in the Caribbean. In Brazil, nearly 30,000 cases of ZIKV infection had been reported by 30 January 2016 (supplementary materials section 1.4). These occurrences point to an epidemic peak in mid-July 2015 (Fig. 1A), and most Brazilian ZIKV cases (93%) were reported in Bahia state (Fig. 1B). Surveillance of ZIKV in Brazil began after the country's first reported case and is conducted through the national Notifiable Diseases Information System (SINAN), which currently relies on passive case detection and reporting and therefore underestimates incidence (*13*). ZIKV is now widespread in Brazil, with autochthonous transmission and high incidence notified in 22 of 27 administrative states (*14*). ZIKV infection during pregnancy has been hypothesized to cause microcephaly and congenital abnormalities (*15*–*20*). The detection of ZIKV in fetal brain tissue (*17*, *20*) and amniotic fluid (*21*) supports the hypothesis that the virus is transmitted from mother to child (*22*); further, the virus infects neural progenitor cells in vitro (*23*). In Brazil, between November 2015 and 30 January 2016, 4783 suspected cases of microcephaly were reported electronically to the RESP database (www.resp.saude.gov.br; Ministry of Health, Brazil) (supplementary materials section 1.4) (Fig. 1C), although most suspected cases are still under investigation and a substantial proportion may represent misdiagnosis and overreporting (*24*). Using the 4 March 2016 World Health Organization guidelines for microcephaly diagnosis (*25*), we identified a total of 1118 suspected microcephaly cases suitable for analysis. The relation between total per-capita ZIKV incidence (Fig. 1B) and per-capita suspected microcephaly cases (Fig. 1C) in each state is weak and only significant under nonparametric correlation ($P < 0.01$) (fig. S1A); noise and uncertainty probably affect both variables. However, the relation is strengthened if suspected microcephaly cases are measured per pregnancy (fig. S1B). For municipalities with reported ZIKV incidence and cases of suspected microcephaly, we used a simple linear model to link microcephaly cases as a function of past ZIKV incidence (supplementary materials section 1.5). On average, suspected microcephaly cases are best predicted by ZIKV incidence during week 17 of pregnancy (95% confidence interval of mean = ±0.11 weeks) or week 14 for suspected

[1]Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA 67030-000, Brazil. [2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [3]Department of Arbovirology and Hemorrhagic Fevers, Evandro Chagas Institute, Ministry of Health, Ananindeua, Pará State, Brazil. [4]Instituto Adolfo Lutz, University of São Paulo, São Paulo, Brazil. [5]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [6]Metabiota, San Francisco, CA 94104, USA. [7]Oswaldo Cruz Foundation (FIOCRUZ), Salvador, Bahia, Brazil. [8]Centre of Post Graduation in Collective Health, Department of Health, Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil. [9]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA. [10]Laboratório de AIDS and Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil. [11]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada. [12]Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Ontario, Canada. [13]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [14]Brazilian Ministry of Health, Brasília, Brazil. [15]Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555, USA. *These authors contributed equally to this work. †These authors contributed equally to this work. **‡Correspondening author. E-mail: oliver.pybus@zoo.ox.ac.uk (O.G.P.); marcionunesbrasil@yahoo.com.br (M.R.T.N.); pedrovasconcelos@iec.pa.gov.br (P.F.C.V.)**

Downloaded from http://science.sciencemag.org/ on July 26, 2016

severe microcephaly cases (±0.08 weeks). These findings are in general agreement with individual reports of the timing of ZIKV symptoms in mothers of infants with microcephaly (*16*, *19*, *21*). We stress that these results quantify only the correlation between ZIKV and suspected microcephaly and do not demonstrate a causal link. Ongoing studies are aiming to establish whether ZIKV is a causal factor in microcephaly and other conditions (*15*–*17*, *23*, *26*).

We used phylogenetic, epidemiological, and mobility data to quantify ZIKV evolution and explore the introduction of the virus to the Americas. As part of ongoing surveillance by the Brazilian Ministry of Health, national laboratories, and other institutions, we used next-generation sequencing to generate seven complete ZIKV coding region sequences from samples collected during the outbreak. Our samples include one from a deceased newborn with microcephaly and congenital malformations collected in Ceará and one from a fatal adult case with lupus and rheumatoid disease from Maranhão state (Fig. 1B). None of the Brazilian patients reported overseas travel (information unavailable in one case), and one individual was a blood donor (supplementary materials section 2). A comparison of our genomes with other available Brazilian strains reveals that Brazilian ZIKV isolates differ at multiple nucleotide sites across the 10.3-kb coding region. The ZIKV genome recovered from isolate ZIKSP, from São Paulo, had 32 nucleotide changes compared with the microcephaly case (BeH823339) and 34 compared with the fatal case from Maranhão (BeH818305). Isolates BeH819966 (from Belém), BeH815744 (from Paraíba), and BeH18995 (from Belém) had a maximum of five nucleotide changes.

Maximum likelihood analysis of complete coding regions from our and other ZIKV genome sequences revealed that all viruses sampled in the Americas, including those from Brazil, form a robust monophyletic cluster (bootstrap score = 94%) within the Asian genotype (Fig. 2 and fig. S2) and share a common ancestor with the ZIKV strain that circulated in French Polynesia in November 2013 (Fig. 3). Previous analyses of outbreaks of related flaviviruses (*27*, *28*) suggest that, to be informative, molecular epidemiological studies of the current ZIKV epidemic should use full or near-complete coding region sequences.

We used a phylogenetic molecular clock approach to further explore the molecular epidemiology of ZIKV in the Americas. A strong correlation between genetic divergence and sampling time within the outbreak lineage (Fig. 2, inset) shows that our approach is appropriate, provided that whole genomes are used. The estimated time-scaled phylogeny (Fig. 3A) again contains a well-supported clade of American ZIKV strains [denoted B; posterior probability (*PP*) = 1.00] that share a common ancestor (denoted A) with the French Polynesian lineage (*PP* = 0.92). Within the American ZIKV lineage (clade B), Brazilian isolates are interspersed among isolates from elsewhere in the Americas. The mingling of ZIKV genomes from different countries reveals ZIKV movement within the Americas since its introduction to the continent. Two observations suggest that the common ancestor of the American ZIKV lineage existed in Brazil. First, Brazil was the first country in the Americas to detect ZIKV (*29*), and second, Brazilian strains are phylogenetically more diverse within clade B than those from elsewhere. However, these observations may reflect differences in surveillance intensity among countries, and more data are required before we can exclude the scenario that ZIKV was introduced to Brazil multiple times from other locations. Although two of three ZIKV-associated microcephaly isolates group together in the phylogeny, there is no reason to posit that this lineage is associated with increased disease severity.

Estimated rates of ZIKV molecular evolution are consistent among different evolutionary models and vary from 0.98×10^{-3} to 1.06×10^{-3} nucleotide substitutions per site per year (table S3). Although these rates are high compared with whole-genome rates for other flaviviruses (*28*), they are consistent with retrospective analyses of previous epidemics, which show that evolutionary rate estimates decline as the epidemic progresses (*30*, *31*). Hence, this result should not be interpreted as implying that ZIKV in the Americas is unusually mutable. We estimate that the date of the most recent common ancestor (TMRCA) of all Brazilian genomes (clade B) is Aug 2013 to Apr 2014 [95% Bayesian credible intervals (BCIs); point estimate = mid-December 2013] (Fig. 3B). The common ancestor of the French Polynesian and America lineages (clade A) was dated from December 2012 to September 2013 (BCIs; point estimate = late May 2013) (Fig. 3B). The posterior distribution for the age of clade B encompasses the recorded duration of the ZIKV outbreak in three of five island groups of French Polynesia (*4*) (Fig. 3C). Divergence date estimates are robust among different combinations of prior distributions, molecular clock models, and coalescent models (supplementary materials sections 4 and 5) and are more likely to shift into the past than toward the present as virus genomes accumulate through time (*30*).

To explore the possible routes of entry of ZIKV in Brazil, we collated airline flight data from all countries with reported ZIKV outbreaks between 2012 and the end of 2014. From late 2012, we find an increase in the number of travelers arriving in Brazil from these countries, rising from 3775 passengers per month in early 2013 to 5754 passengers per month a year later (Fig. 3C). This increase in visitors to Brazil from ZIKV-affected countries coincides with the period during which ZIKV is estimated to have entered the Americas (i.e., between TMRCAs of clades A and B) (Fig. 3B and supplementary materials section 5). If the ZIKV epidemic in Brazil did indeed arise from a single introduction, then the virus must have circulated in the country for at least 12 months before the first case was reported in May 2015. ZIKV clinical symptoms may be confused with those caused by

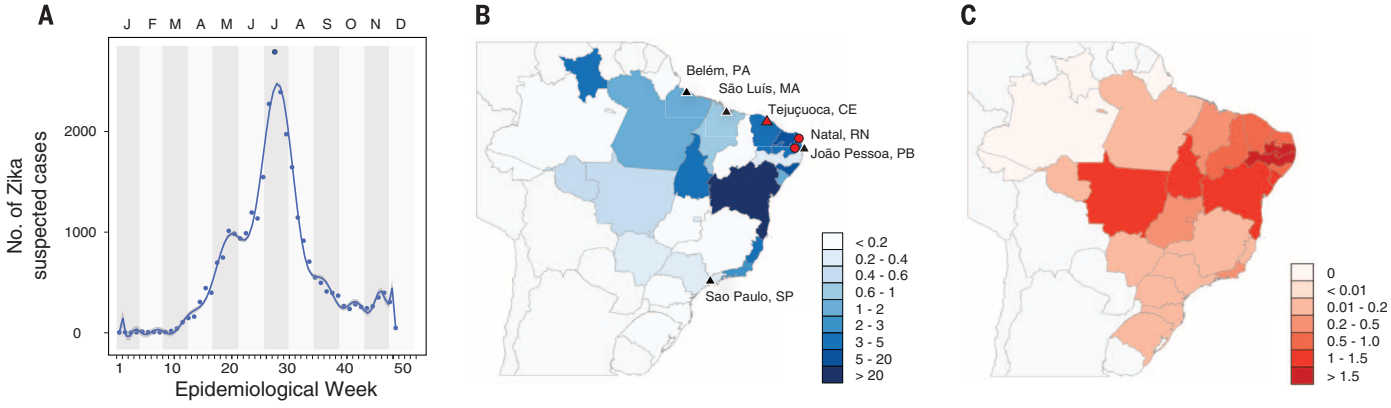


Fig. 1. Time series and cartography of reported Zika virus and microcephaly cases in Brazil. (**A**) Number of suspected cases of ZIKV per week in 5596 municipalities in Brazil. The epidemic peaked from 12 to 18 July 2015 (n = 2791 cases). Letters indicate months. (**B**) Total incidence of ZIKV cases per 100,000 people in each federal state. Triangles indicate sampling locations of the sequences reported here; circles indicate locations of other genomes from Brazil [municipality of Natal in Rio Grande do Norte state (*16*) and an unknown municipality in Paraiba state (*21*)]. Red symbols indicate ZIKV genomes isolated from microcephaly cases. Federal states are indicated by two-letter codes: PA, Pará; MA, Maranhão; CE, Ceará; RN, Rio Grande do Norte; PB, Paraíba; SP, São Paulo. Per-capita incidences in each state were calculated using high-resolution gridded human population–size data sets for Brazil (*45*). (**C**) Incidence of suspected microcephaly cases per 100,000 people in each federal state. Per-capita incidences for each state were calculated as described for (B).

Downloaded from http://science.sciencemag.org/ on July 26, 2016

dengue and chikungunya viruses, two endemic and epidemic viruses that cocirculate and share mosquito vectors with ZIKV in Brazil (*27*, *32*, *33*). Reliable differential diagnosis is possible only by using improved surveillance and laboratory diagnostics, which are now being implemented throughout the country.

There are two published hypotheses for how ZIKV came to be introduced into Brazil: during (i) the 2014 World Cup soccer tournament (12 June to 13 July 2014) (*29*) or (ii) the Va'a canoe event held in Rio de Janeiro between 12 and 17 August 2014 (*34*). Alternatively, introduction could have occurred during (iii) the 2013 Confederations Cup soccer tournament (15 to 30 June 2013). Notably, events (ii) and (iii) included competitors from French Polynesia. Our results suggest that the introduction of ZIKV to the Americas predated events (i) and (ii). Although the molecular clock dates are more consistent with the Confederations Cup tournament, that event ended before ZIKV cases were first reported in French Polynesia (*4*). Consequently, we believe that large-scale patterns in human mobility will provide more useful and testable hypotheses about viral introduction and emergence (*33*, *35*, *36*) than ad hoc hypotheses focused on specific events.

The ZIKV genome we obtained from a microcephaly case in Ceará, Brazil, contains eight amino acid changes not observed in any other complete genome in our data set. However, none of these mutations are shared with either of two recently published genomes from microcephaly cases (*16*, *21*). Thus, if a causal link between Asian lineage ZIKV and microcephaly is confirmed, it is possible that putative viral genetic determinants of disease will be found among the amino acid changes that occur on the ZIKV phylogeny branches ancestral to the French Polynesian and American ZIKV lineages (i.e., the two lineages associated with reports of microcephaly, Guillain-Barré syndrome, and congenital abnormalities) (*37*). Phylogenetic character mapping using parsimony reveals 11 amino acid changes on the four internal branches (Fig. 2, asterisks; and fig. S3) leading to these two lineages. We identified the structures of homologous proteins most closely related to the ZIKV proteins (supplementary materials section 7) and used them to map 7 of the 11 amino acid changes, in a structural context, to five proteins: the pr-peptide region of prM [changes Val^{123}→Ala^{123} (V123A) and S139N (S, Ser; N, Asn)], NS1 (A982V), the RNA helicase [NS3; N1902H and Y2086H (H, His; Y, Tyr)], the FtsJ-like methyl transferase domain [NS5; M2634V (M, Met)], and the thumb domain of RNA-directed RNA polymerase (NS5; M3392V) (fig. S7). None of these mutations are predicted to substantially affect the physicochemical properties of the protein environment, except possibly Y2086H (in the helicase; fig. S8), which may increase the hydrophilicity of the region. The remaining four amino acid changes could not be accurately mapped due to the absence of suitable related x-ray structures (supplementary materials section 7). Notably, none of the observed changes map to the E glycoprotein ectodomain, the primary target of humoral immune responses against flaviviruses (*38*, *39*). Factors other than viral genetic differences may be important for the proposed pathogenesis of ZIKV; hypothesized factors include coinfection with chikungunya virus (*40*), previous infection with dengue virus (*41*), or differences in human genetic predisposition to disease.

Besides vector-borne and mother-to-child transmission, Zika virus may also spread via sexual

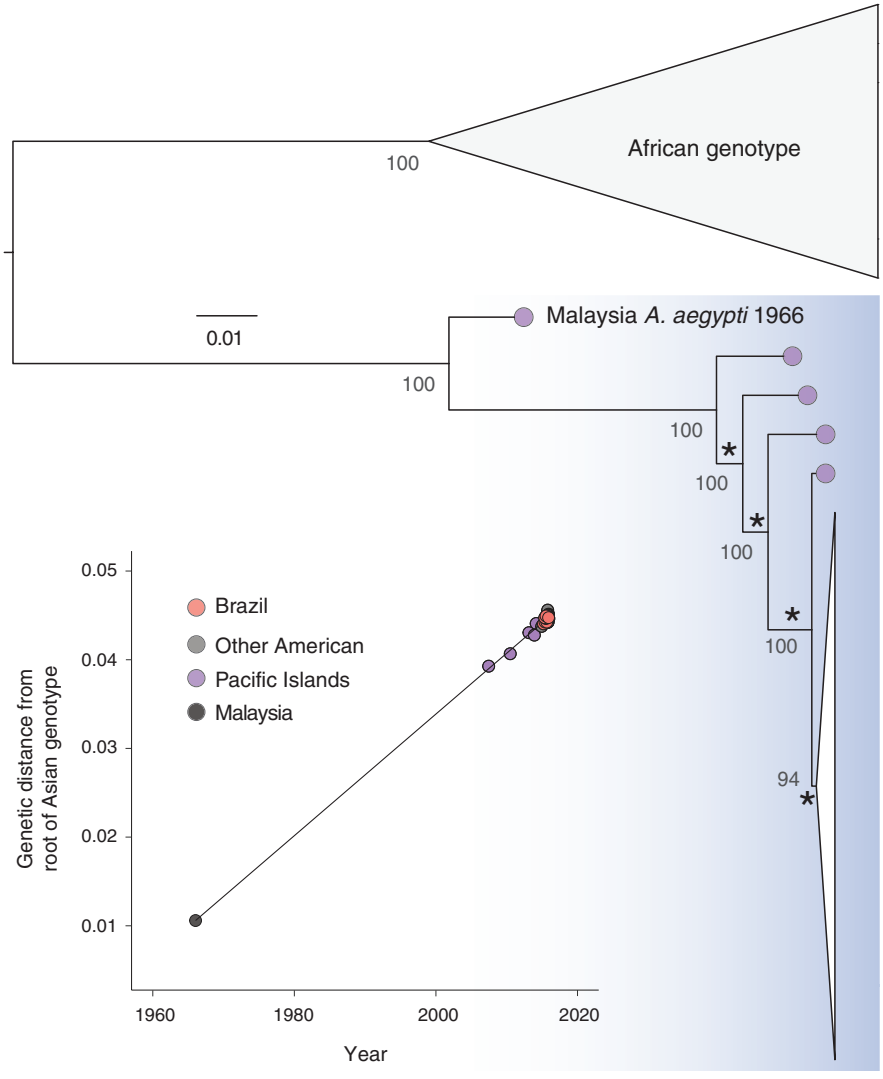


Fig. 2. Maximum likelihood phylogeny of ZIKV complete coding region sequences. Bootstrap scores are shown next to well-supported nodes, and the phylogeny was midpoint rooted. A fully annotated tree is provided in fig. S2. The American ZIKV outbreak clade is drawn as a narrow white triangle and is shown in detail in Fig. 3. Asterisks highlight the four internal branches that are ancestral to the American ZIKV lineage (see main text and fig. S3). There is a correlation between the sampling date of each sequence and the genetic distance of that sequence from the root of a maximum likelihood phylogeny of the Asian genotype (correlation coefficient R^2 = 0.997). A molecular clock phylogeny of these data is shown in Fig. 3. The Malaysian strain (HQ234499) sampled in 1966 is the oldest representative of the Asian genotype and falls on the regression line, indicating that it does not appear to be unusually divergent for its age. A similar analysis with the HQ234499 strain excluded is shown in fig. S5C.

Downloaded from http://science.sciencemag.org/ on July 26, 2016

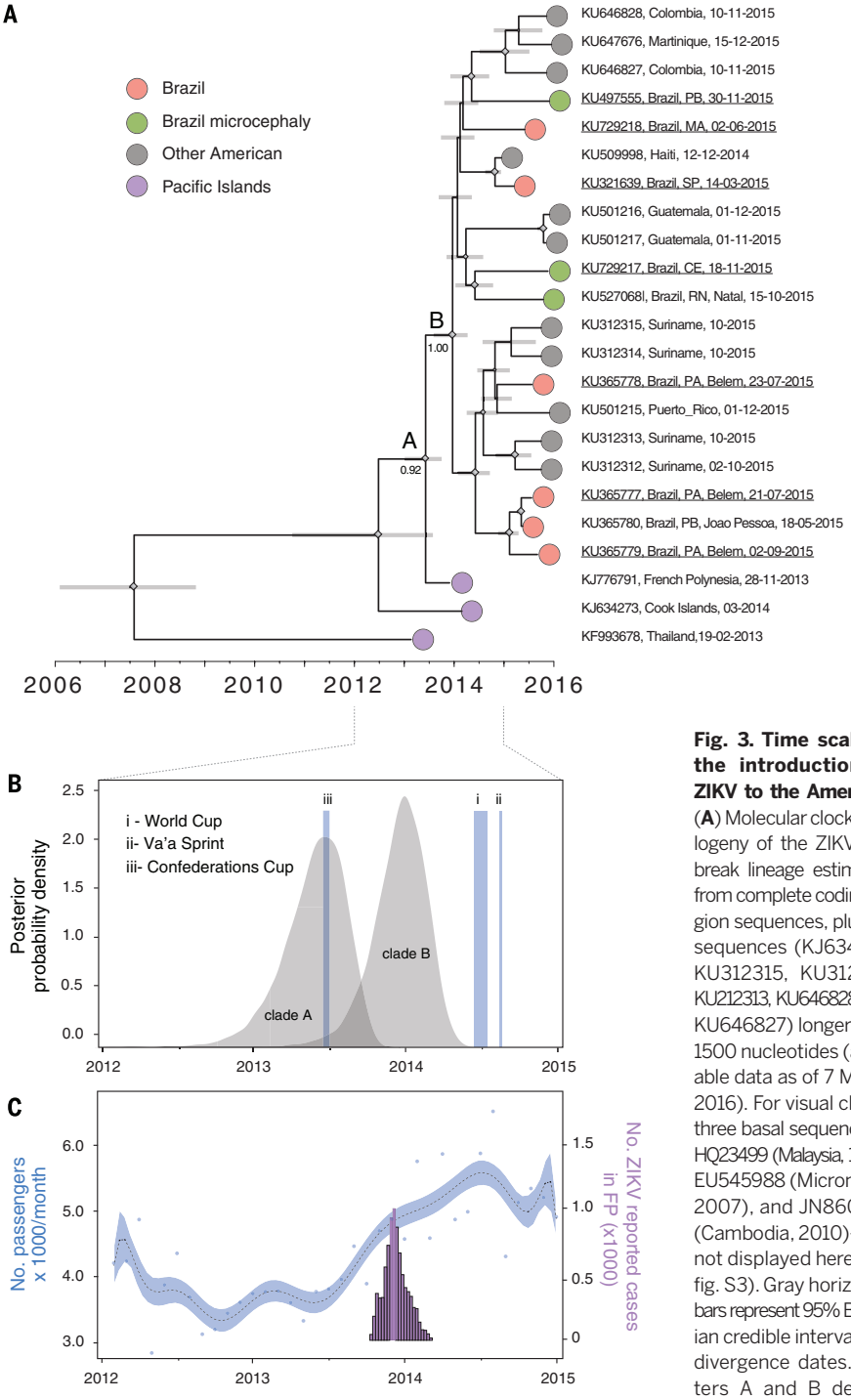


Fig. 3. Time scale of the introduction of ZIKV to the Americas. (**A**) Molecular clock phylogeny of the ZIKV outbreak lineage estimated from complete coding region sequences, plus six sequences (KJ634273, KU312315, KU312314, KU212313, KU646828, and KU646827) longer than 1500 nucleotides (available data as of 7 March 2016). For visual clarity, three basal sequences—HQ23499 (Malaysia, 1966), EU545988 (Micronesia, 2007), and JN860885 (Cambodia, 2010)—are not displayed here (see fig. S3). Gray horizontal bars represent 95% Bayesian credible intervals for divergence dates. Letters A and B denote clades discussed in the main text; numbers under the clade letters denote posterior probabilities. Diamond sizes represent, at each node, the posterior probability support of that node. Taxa are labeled with accession number, sampling location, and sampling date. Names of sequences generated in this study are underlined. (**B**) Posterior distributions of the estimated ages (TMRCAs) of clades A and B, estimated with BEAST software using the best-fitting evolutionary model (table S2). The time and duration of the three events (i to iii) discussed in the main text are shown. (**C**) Number of airline passengers from specific countries arriving in Brazil per month versus number of suspected cases of ZIKV in French Polynesia. The blue curve (left *y* axis) shows a polynomial fitting of the number of travelers (blue points) from countries with recorded ZIKV outbreaks between 2012 and 2015 (French Polynesia, Thailand, Indonesia, Malaysia, Cambodia, and New Caledonia) (supplementary materials section 6), aggregated across 20 Brazilian national airports. The purple bars represent weekly numbers of suspected ZIKV cases (right *y* axis) in French Polynesia (FP) from 30 October 2013 to 14 February 2014 (*4*).

contact (*42*, *43*) and blood transfusion (*44*). The evidence of ZIKV in blood donors raises the possibility of ZIKV transmission through transfusion and indicates that it may be prudent to consider the screening of blood donors.

REFERENCES AND NOTES

1. B. D. Lindenbach, C. M. Rice, *Adv. Virus Res.* **59**, 23–61 (2003).
2. G. W. Dick, S. F. Kitchen, A. J. Haddow, *Trans. R. Soc. Trop. Med. Hyg.* **46**, 509–520 (1952).
3. O. Faye *et al.*, *Virol. J.* **10**, 311 (2013).
4. S. Ioos *et al.*, *Med. Mal. Infect.* **44**, 302–307 (2014).
5. M. R. Duffy *et al.*, *N. Engl. J. Med.* **360**, 2536–2543 (2009).
6. A. D. Haddow *et al.*, *PLOS Negl. Trop. Dis.* **6**, e1477 (2012).
7. V. M. Cao-Lormeau *et al.*, *Emerg. Infect. Dis.* **20**, 1085–1086 (2014).
8. M. Dupont-Rouzeyrol *et al.*, *Emerg. Infect. Dis.* **21**, 381–382 (2015).
9. A. T. Pyke *et al.*, *PLOS Curr.* 10.1371/currents.outbreaks.4635a54dbffba2156fb2fd76dc49f65e (2014).
10. T. Wæhre, A. Maagard, D. Tappe, D. Cadar, J. Schmidt-Chanasit, *Emerg. Infect. Dis.* **20**, 1412–1414 (2014).
11. J. Tognarelli *et al.*, *Arch. Virol.* **161**, 665–668 (2016).
12. M. Hennessey, M. Fischer, J. E. Staples, *MMWR Morb. Mortal. Wkly. Rep.* **65**, 55–58 (2016).
13. M. M. O. Silva *et al.*, *Emerg. Infect. Dis.* **22**, 336–339 (2016).
14. Secretaria de Vigilância em Saúde, "Boletim epidemiológico: Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a semana epidemiológica 4, 2016" (in Portuguese) (Ministério da Saúde do Brasil, 2016); vol. 47, no. 7; http://portalsaude.saude.gov.br/images/pdf/2016/fevereiro/29/2016-005—Dengue-SE4-publica—-o.pdf.
15. European Centre for Disease Prevention and Control (ECDC), "Rapid risk assessment: Microcephaly in Brazil potentially linked to the Zika virus epidemic" (ECDC, 2015); http://ecdc.europa.eu/en/publications/Publications/zika-microcephaly-Brazil-rapid-risk-assessment-Nov-2015.pdf.
16. J. Mlakar *et al.*, *N. Engl. J. Med.* **374**, 951–958 (2016).
17. L. Schuler-Faccini *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **65**, 59–62 (2016).
18. C. V. Ventura *et al.*, *Arq. Bras. Oftalmol.* **79**, 1–3 (2016).
19. R. B. Martines *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **65**, 159–160 (2016).
20. Pan American Health Organization (PAHO), World Health Organization (WHO), "Epidemiological alert: Neurological syndrome, congenital malformations, and Zika virus infection. Implications for public health in the Americas" (PAHO/WHO, 2015).
21. G. Calvet *et al.*, *Lancet Infect. Dis.* 10.1016/S1473-3099(16)00095-5 (2016).
22. M. Besnard, S. Lastere, A. Teissier, V. Cao-Lormeau, D. Musso, *Euro Surveill.* **19**, 20751 (2014).
23. H. Tang *et al.*, *Cell Stem Cell* 10.1016/j.stem.2016.02.016 (2016).
24. C. G. Victora *et al.*, *Lancet* **387**, 621–624 (2016).
25. WHO, "Assessment of infants with microcephaly in the context of Zika virus" (WHO Interim Report, 2016); http://apps.who.int/iris/bitstream/10665/204475/1/WHO_ZIKV_MOC_16.3_eng.pdf.
26. C. V. Ventura, M. Maia, V. Bravo-Filho, A. L. Góis, R. Belfort Jr., *Lancet* **387**, 228 (2016).
27. M. R. Nunes *et al.*, *Emerg. Infect. Dis.* **18**, 1858–1864 (2012).
28. O. G. Pybus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15066–15071 (2012).
29. C. Zanluca *et al.*, *Mem. Inst. Oswaldo Cruz* **110**, 569–572 (2015).
30. A. G. Meyer, S. J. Spielman, T. Bedford, C. O. Wilke, *Virus Evol.* **1**, vev006 (2015).
31. D. J. Park *et al.*, *Cell* **161**, 1516–1526 (2015).
32. M. R. Nunes *et al.*, *BMC Med.* **13**, 102 (2015).
33. M. R. Nunes *et al.*, *PLOS Negl. Trop. Dis.* **8**, e2769 (2014).
34. D. Musso, *Emerg. Infect. Dis.* **21**, 1887 (2015).
35. P. Lemey *et al.*, *PLOS Pathog.* **10**, e1003932 (2014).
36. O. G. Pybus, A. J. Tatem, P. Lemey, *Proc. Biol. Sci.* **282**, 20142878 (2015).
37. V. M. Cao-Lormeau *et al.*, *Lancet* 10.1016/S0140-6736(16)00562-6 (2016).
38. K. A. Dowd, T. C. Pierson, *Virology* **411**, 306–315 (2011).
39. J. T. Roehrig, *Adv. Virus Res.* **59**, 141–175 (2003).
40. P. Gérardin *et al.*, *PLOS Negl. Trop. Dis.* **8**, e2996 (2014).
41. A. Fagbami, S. B. Halstead, N. Marchette, K. Larsen, *Am. J. Trop. Med. Hyg.* **38**, 205–207 (1988).
42. B. D. Foy *et al.*, *Emerg. Infect. Dis.* **17**, 880–882 (2011).
43. D. Musso *et al.*, *Emerg. Infect. Dis.* **21**, 359–361 (2015).

Downloaded from http://science.sciencemag.org/ on July 26, 2016

44. D. Musso *et al*., *Euro Surveill.* **19**, 20761 (2014).
45. A. Sorichetta *et al*., *Sci. Data* **2**, 150045 (2015).

ACKNOWLEDGMENTS

We thank X. de Lamballerie and J. Lednicky for permission to include their unpublished ZIKV genomes in our analysis. We thank the Death Verification Service (SVO); Central Laboratories of Public Health (LACEN); and health departments of the Ceará State and Maranhão State, Brazil, for collaboration. O.G.P. is supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant 614725-PATHPHYLODYN. J.L. is supported by the ERC under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant 268904-DIVERSITY. O.G.P. received consulting fees from Metabiota between 2015 and 2016. This study is made possible in part by the generous support of the American people through the United States Agency for International Development (USAID) Emerging Pandemic Threats Program - PREDICT. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the U.S. government. S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (095066) and grants from the Bill and Melinda Gates Foundation (OPP1119467, OPP1093011, OPP1106023, and OPP1132415). M.R.T.N. is funded as an associated researcher in public health by the Evandro Chagas Institute, Brazilian Ministry of Health, and as a researcher in scientific productivity by CNPq (Brazilian National Council for Scientific and Technological Development) grants 302032/2011-8 and 200024/2015-9 and is also supported in part by the National Institute of Science and Technology for Viral Hemorrhagic Fevers. R.T. is funded by grant R24 AT 120942 from the U.S. NIH. S.C.H. is supported by a Wellcome Trust grant (102427). T.A.B. and I.R. are supported by grants from the UK Medical Research Council (MR/L009528/1) and the Wellcome Trust (090532/Z/09/Z). P.F.C.V. is supported by CNPq–National Agency for Scientific and Technologic Development (grants 573739/2008-0, 301641/2010-2, and 457664/2013-4). All samples were obtained from persons visiting local clinics or hospitalized by the Brazilian Ministry of Health personnel as part of dengue, chikungunya, and Zika fever surveillance activities. In these cases, patient consent was oral and not recorded. The study was authorized by the Coordination of the National Program for Dengue, Chikungunya, and Zika Control coordinated by Brazil's Ministry of Health. The data are available at DRYAD (DOI: 10.5061/dryad.6kn23). The ZIKV genomes reported in this study are deposited in GenBank under accession numbers KU321639, KU365777 to KU365780, KU729217, and KU729218.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6283/345/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S5
References (*46*–*78*)

18 February 2016; accepted 16 March 2016
Published online 24 March 2016
10.1126/science.aaf5036

MALARIA DRUGS

Parasites resistant to the antimalarial atovaquone fail to transmit by mosquitoes

Christopher D. Goodman,[1]*† Josephine E. Siregar,[1,2,6]† Vanessa Mollard,[1] Joel Vega-Rodríguez,[3] Din Syafruddin,[2,4] Hiroyuki Matsuoka,[5] Motomichi Matsuzaki,[6] Tomoko Toyama,[1] Angelika Sturm,[1] Anton Cozijnsen,[1] Marcelo Jacobs-Lorena,[3] Kiyoshi Kita,[6,7] Sangkot Marzuki,[2]‡ Geoffrey I. McFadden[1]*‡

Drug resistance compromises control of malaria. Here, we show that resistance to a commonly used antimalarial medication, atovaquone, is apparently unable to spread. Atovaquone pressure selects parasites with mutations in cytochrome *b*, a respiratory protein with low but essential activity in the mammalian blood phase of the parasite life cycle. Resistance mutations rescue parasites from the drug but later prove lethal in the mosquito phase, where parasites require full respiration. Unable to respire efficiently, resistant parasites fail to complete mosquito development, arresting their life cycle. Because cytochrome *b* is encoded by the maternally inherited parasite mitochondrion, even outcrossing with wild-type strains cannot facilitate spread of resistance. Lack of transmission suggests that resistance will be unable to spread in the field, greatly enhancing the utility of atovaquone in malaria control.

Atovaquone, a component of the safe and effective antimalarial medication Malarone, kills both the blood and liver stages of malaria (*1*). The rollout of cheap generics should see increased atovaquone usage, and atovaquone derivatives are in development (*1*). Atovaquone is prone to resistance (*1*), and it has been assumed that this resistance will spread (*2*, *3*), as it has for other antimalarials (*4*, *5*). However, the target of atovaquone, cytochrome *b* (*cytB*) (*6*–*9*), has unique genetics (*10*–*12*) and experiences differential selection across the malaria parasite life cycle (*13*), which prompted us to investigate whether atovaquone resistance can spread via the mosquito vector.

We tested three atovaquone-resistant strains of the rodent malaria parasite *Plasmodium berghei*, each with different mutations in their mitochondrial DNA-encoded *cytB* gene (*14*, *15*), for transmissibility from mouse to mosquito and back to mouse (Table 1). *Anopheles stephensi* mosquitoes were fed on mice infected with either the parental *Pb*ANKA strain or one of the three atovaquone-resistant mutants, and sexual development of parasites in mosquitoes was assayed (Table 1). All three atovaquone-resistant parasite lines produced wild-type numbers of active male gametes (exflagellation) (Table 1). Parasites carrying the *Pb*M133I and *Pb*Y268C mutations in their *cytB* gene were able to self-fertilize, generate ookinetes, and successfully produce oocysts, but the oocysts produced had developmental defects (Fig. 1, A and B, and Table 1). Parasites with the *Pb*Y268N mutation were defective in the ability to self-fertilize and infect the mosquito host (Table 1) due to severely impaired female gamete activation (Fig. 1C). From 17 attempted mosquito infections, no parasite carrying an atovaquone-resistant *cytB* mutation was able to generate the sporozoite stages in the mosquito salivary glands or was able to infect a naïve mouse (Table 1). We conclude that the rodent malaria atovaquone-resistant *cytB* mutants tested—which represent a good cross section of the clinical atovaquone-resistant genotypes, including the common Y268 locus (*16*)—are unable to transmit from mouse to mouse via *A. stephensi* mosquitoes when self-fertilizing.

To determine whether outcrossing with atovaquone-sensitive parasites could help transmit the atovaquone resistance genes, we generated crosses of our atovaquone-resistant *P. berghei* lines with atovaquone-sensitive parasites. These experiments simulate what might happen if a mosquito bit an individual (or separate individuals) infected with both atovaquone-resistant and atovaquone-sensitive parasites, which can then mate in the mosquito gut. They allow us to assess whether the presence of wild-type copies of the *cytB* genes from one parent can complement a mutation in the other, as observed with deletions of electron transport components encoded in the nuclear genome (*17*). We first crossed *Pb*Y268C with an atovaquone-sensitive line (wild-type *cytB*) carrying a mutation (*15*) in the nucleus-encoded dihydrofolate reductase (*dhfr*) gene conferring pyrimethamine resistance (*Pbdhfr*S110N) by pooling blood from separate infected mice and then membrane-feeding mosquitoes. Sporozoites were produced, and all 14 naïve mice bitten by these mosquitoes (three trials) developed blood-stage infections. Genotyping of these progeny [passage

[1]School of BioSciences, University of Melbourne, Melbourne, VIC 3010, Australia. [2]Eijkman Institute for Molecular Biology, Jl Diponegoro no. 69, Jakarta, 10430, Indonesia. [3]Johns Hopkins University Bloomberg School of Public Health, Department of Molecular Microbiology and Immunology, Malaria Research Institute, Baltimore, MD 21205, USA. [4]Department of Parasitology, Faculty of Medicine, Hasanuddin University, Jalan Perintis Kemerdekaan Km10, Makassar 90245, Indonesia. [5]Division of Medical Zoology, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan. [6]Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [7]School of Tropical Medicine and Global Health, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan.
***Corresponding author. E-mail: gim@unimelb.edu.au (G.I.M.); deang@unimelb.edu.au (C.D.G.)** †These authors contributed equally to this work; order is alphabetical. ‡These authors contributed equally to this work; order is alphabetical.

Downloaded from http://science.sciencemag.org/ on July 26, 2016

Anticipating the international spread of Zika virus from Brazil

Published Online
January 14, 2016
http://dx.doi.org/10.1016/S0140-6736(16)00080-5

In May, 2015, locally acquired cases of Zika virus—an arbovirus found in Africa and Asia-Pacific and transmitted via *Aedes* mosquitoes—were confirmed in Brazil. The presence of *Aedes* mosquitoes across Latin America, coupled with suitable climatic conditions, have triggered a Zika virus epidemic in Brazil, currently estimated at 440 000–1 300 000 cases.[1] Viraemic travellers have now introduced Zika virus into at least 13 additional countries, where susceptible *Aedes* mosquitoes have become infected and perpetuated local transmission cycles. In Brazil, a precipitous surge in infants born with microcephaly and the detection of Zika virus RNA in the amniotic fluid of affected newborns has been reported.[1] We sought to identify high-risk international pathways for the dispersion of Zika virus and global geographies conducive to autochthonous transmission.

We created a global Zika virus spread model by adapting a seasonal model for dengue that integrates global ecological niche data for *Aedes aegypti* and *albopictus* and worldwide temperature profiles.[2,3] In Brazil, we identified airports within 50 km of areas conducive to year-round Zika virus transmission. We mapped the final destinations of international travellers departing from these airports from September, 2014, to August, 2015, using worldwide flight itinerary data from the International Air Transport Association. We used LandScan, a gridded global population dataset, to estimate numbers of people living in geographies at risk for autochthonous Zika virus transmission.

9·9 million travellers departed from the aforementioned Brazilian airports for international destinations, with 65% to the Americas (figure), 27% to Europe, and 5% to Asia. Traveller volumes were greatest to the USA (2 767 337), Argentina (1 314 694), Chile (614 687), Italy (419 955), Portugal (411 407), and France (404 525). China and Angola received the highest volume of travellers in Asia (84 332) and Africa (82 838), respectively. Argentina, Italy, and the USA have more than 60% of their populations residing in areas conducive to seasonal Zika

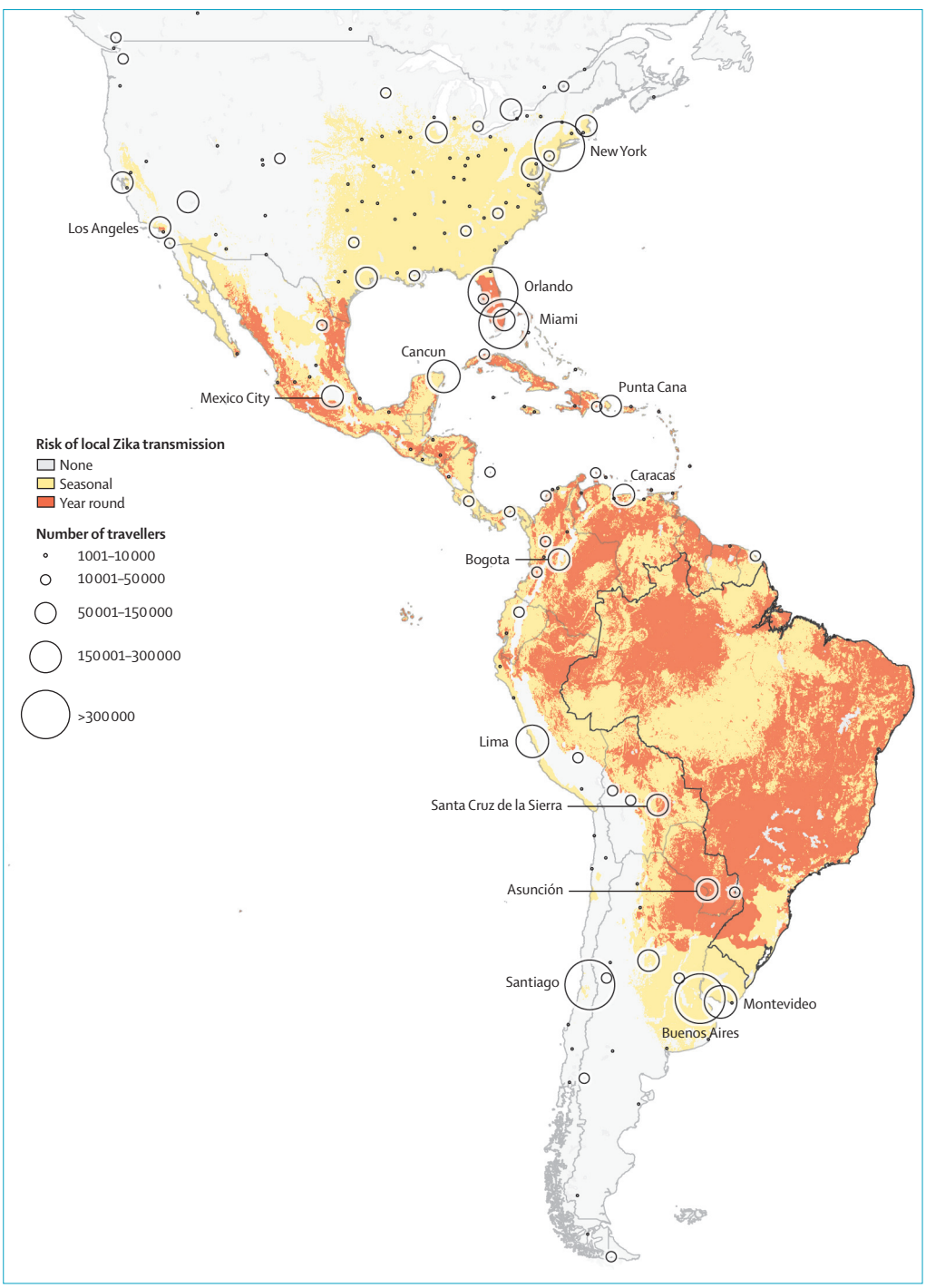


Figure: **Final destinations of travellers departing Brazil by potential for autochthonous Zika transmission**

virus transmission, whereas Mexico, Colombia, and the USA have an estimated 30·5, 23·2, and 22·7 million people, respectively, living in areas conducive to year-round transmission.

In parallel to the recent experience with chikungunya,[4] Zika virus has the potential to rapidly spread across Latin America and the Caribbean. With no vaccine or antiviral therapy available, possible interventions include: personal protection (ie, repellent use) and daytime avoidance of mosquito bites (especially pregnant women until more is known about the association between Zika virus infection and microcephaly); daytime avoidance of mosquito bites among Zika virus-infected individuals to disrupt human to mosquito to human transmission cycles (80% of infected individuals are asymptomatic and the remainder have clinical syndromes overlapping with dengue and chikungunya);[5] and community-level mosquito surveillance and control measures. The summer Olympic Games in Brazil in August, 2016, heighten the need for awareness of this emerging virus.

KK is the founder of BlueDot, a social benefit corporation that models global infectious disease threats. MIC, MG, and AW have received employment income from BlueDot. IIB has consulted to BlueDot. We acknowledge support from the Canadian Institutes of Health Research, National Institute of Health, R01 LM010812, the Wellcome Trust (#095066), the Bill & Melinda Gates Foundation (OPP1119467, OPP1106023, and OPP1093011), and the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

*Isaac I Bogoch, Oliver J Brady, Moritz U G Kraemer, Matthew German, Marisa I Creatore, Manisha A Kulkarni, John S Brownstein, Sumiko R Mekaru, Simon I Hay, Emily Groot, Alexander Watts, *Kamran Khan*
khank@smh.ca

Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Canada (IIB, KK); Divisions of Internal Medicine and Infectious Diseases, University Health Network, Toronto, Canada (IBB); "Dalla Lana School of Public Health, University of Toronto, Toronto, Canada (MIC); Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, UK (MUGK); Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada (MG, MIC, EG, AW, KK); School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada (MAK); Boston Children's Hospital, Harvard Medical School, Boston, MA, USA (JSB, SRM); Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK (SIH, OJB); Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA (SIH); and Fogarty International Center, National Institutes of Health, Bethesda, MD, USA (SIH)

1 European Centre for Disease Prevention and Control. Rapid risk assessment: Zika virus epidemic in the Americas: potential association with microcephaly and Guillain-Barré syndrome. 10 December 2015. Stockholm: ECDC, 2015.
2 Kraemer MUG, Sinka ME, Duda KA, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife* 2015; **4**: e08347.
3 Brady OJ, Golding N, Pigott DM, et al. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit Vectors* 2014; **7:** 338.
4 Weaver SC, Lecuit M. Chikungunya virus and the global spread of a mosquito-borne disease. *N Engl J Med* 2015; **372:** 1231–39.
5 Duffy MR, Chen T-H, Hancock WT, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 2009; **360:** 2536–43.

For the **COMPare project** see www.COMPare-Trials.org

Eltrombopag for chronic immune thrombocytopenia

The PETIT2 trial (Oct 24, p 1649)[1] reported outcomes that were different from those registered before trial commencement.[2] Of 22 prespecified secondary outcomes, 14 were reported in the Article, and eight were not reported anywhere in the paper. Additionally, the Article reported a new outcome ("concomitant drugs for immune thrombocytopenia") that was not prespecified, without flagging it as such. For clarity, we considered tables 2–4 in the Article to represent the main trial results, with secondary outcome measures split between them.

We declare no competing interests.

**Ioan Milosevic, Eirion Slade, Henry Drysdale, on behalf of the COMPare project team*
compare-team@ebmdatalab.net

Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Services, Oxford OX2 6GG, UK

1 Grainger JD, Locatelli F, Chotsampancharoen T, et al. Eltrombopag for children with chronic immune thrombocytopenia (PETIT2): a randomised, multicentre, placebo-controlled trial. *Lancet* 2015; **386:** 1649–58.
2 View of NCT01520909 on 2012_03_09. ClincalTrials.gov archive. https://clinicaltrials.gov/archive/NCT01520909/2012_03_09 (accessed Oct 29, 2015).

Authors' reply

We thank Ioan Milosevic and colleagues for their inquiry regarding the PETIT2 trial.[1] We agree that the CONSORT statement for reporting of randomised trials is an important guide for communicating trial results, and we used these guidelines when drafting our report. Because of word limits, we presented only the key findings that would be of particular interest to clinicians who treat children with immune thrombocytopenia. Additionally, results are publicly available on ClinicalTrials.gov (number NCT01520909) and the GlaxoSmithKline clinical study registry (number 115450).

We wish to clarify that the study endpoints in the protocol were not different from those reported in our Article.[1] All the primary and secondary endpoints in the Article were the same as those in the protocol, including "reduction or discontinuation of concomitant drugs for immune thrombocytopenia", which was a secondary endpoint. Vital signs and clinical laboratory values were part of the safety assessment. Clinical laboratory values related to liver function adverse events were reported in the text and were also provided in the Article[1] appendix (table S2). The results of the ophthalmic examination were reported in the text.

We did not report the five pharmacokinetic endpoints in our Article[1] but will include them in a separate publication, in which these data will be combined with similar data from the phase 2 PETIT study.

I have received research funding from Baxter and honoraria from Amgen, Baxter, Novartis, and GlaxoSmithKline.

nature microbiology

LETTERS

PUBLISHED: 25 JULY 2016 | ARTICLE NUMBER: 16126 | DOI: 10.1038/NMICROBIOL.2016.126

Model-based projections of Zika virus infections in childbearing women in the Americas

T. Alex Perkins[1]*, Amir S. Siraj[1], Corrine W. Ruktanonchai[2], Moritz U. G. Kraemer[3] and Andrew J. Tatem[2,4]

Zika virus is a mosquito-borne pathogen that is rapidly spreading across the Americas. Due to associations between Zika virus infection and a range of fetal maladies[1,2], the epidemic trajectory of this viral infection poses a significant concern for the nearly 15 million children born in the Americas each year. Ascertaining the portion of this population that is truly at risk is an important priority. One recent estimate[3] suggested that 5.42 million childbearing women live in areas of the Americas that are suitable for Zika occurrence. To improve on that estimate, which did not take into account the protective effects of herd immunity, we developed a new approach that combines classic results from epidemiological theory with seroprevalence data and highly spatially resolved data about drivers of transmission to make location-specific projections of epidemic attack rates. Our results suggest that 1.65 (1.45–2.06) million childbearing women and 93.4 (81.6–117.1) million people in total could become infected before the first wave of the epidemic concludes. Based on current estimates of rates of adverse fetal outcomes among infected women[2,4,5], these results suggest that tens of thousands of pregnancies could be negatively impacted by the first wave of the epidemic. These projections constitute a revised upper limit of populations at risk in the current Zika epidemic, and our approach offers a new way to make rapid assessments of the threat posed by emerging infectious diseases more generally.

On 1 February 2016, the World Health Organization (WHO) designated the ongoing Zika virus epidemic in the Americas as a Public Health Emergency of International Concern (PHEIC), defined as an 'extraordinary event' that 'potentially require[s] a coordinated international response'[6]. This declaration acknowledges the high potential for Zika to establish across the Americas given that its dominant vector, the *Aedes aegypti* mosquito, is endophilic and occupies an exceptionally broad geographical range[7]. Concern underlying this rare WHO declaration also stems from an association between Zika virus infection in pregnant women and a range of adverse fetal outcomes[2], most notably congenital microcephaly[1]. As of 30 June 2016, there were 1,674 confirmed cases of microcephaly associated with Zika virus infection in five countries[8], and there is widespread concern that these numbers could increase further as the virus continues to spread across the Americas[9].

A number of uncertainties surround the future of the Zika epidemic in the Americas, particularly questions about how many women may be at risk of having children with congenital microcephaly and other adverse outcomes associated with Zika virus infection[10]. Of women who become infected with Zika virus during a vulnerable stage of their pregnancy, evidence is emerging that 1–13% may go on to develop congenital microcephaly[2,4,5]. However, the number of women who become infected with Zika virus during that timeframe is difficult to ascertain. One recent study[3] estimated that 5.42 million births occurred in 2015 in regions of the Americas with 'suitability' for Zika 'occurrence'. Such estimates come with many caveats though, as they rely on a relatively limited number of reported cases and apply a method based on equilibrium assumptions to a situation involving active range expansion[11]. Most importantly, the estimate of 5.42 million births[3] reflects the total population within a demarcated area and does not take into account that large fractions of the populations in those areas may remain uninfected due to herd immunity generated over the course of the first wave of the epidemic[12,13].

To quantify the potential magnitude of the ongoing Zika epidemic in terms of people who realistically might become infected, we formulated and applied a method for projecting location-specific epidemic attack rates on highly spatially resolved human demographic projections[14]. The central concept behind our approach is that of the 'first-wave' epidemic. Zika and other mosquito-borne viruses have been known to exhibit explosive outbreaks, infecting as much as 75% of a population in a single year[15]. Classical epidemiological theory predicts that some proportion of a population will remain uninfected during an epidemic, because herd immunity eventually causes the epidemic to burn out[12]. A related prediction of this theory is that the proportion infected before epidemic burnout (that is, the epidemic attack rate) has a one-to-one relationship with the basic reproduction number, R_0 (ref. 13). The latter quantity has a well-known mechanistic formulation for mosquito-borne pathogens[16] that accommodates the effects of environmental drivers on transmission[17,18]. For example, the incubation periods of dengue viruses in the *Ae. aegypti* mosquitoes that transmit Zika virus have an empirically derived relationship with temperature[18], which can in turn be used to inform calculations of R_0. Together with similar relationships for other transmission parameters, it is possible to characterize R_0, a fundamental measure of transmission potential, as a function of local environmental conditions.

We leveraged these classic results from epidemiological theory to first perform highly spatially resolved calculations of R_0 and then to translate those calculations into location-specific projections of first-wave epidemic attack rates (Fig. 1). Because Zika-specific values of transmission parameters are largely unknown at present but may be well approximated by dengue-specific values[19], we used some parameter values for dengue virus in our R_0 calculations. We also calibrated our attack rate projections to match empirically estimated attack rates from 12 chikungunya epidemics and one Zika epidemic in naive populations (Supplementary Table 1). This step afforded us the flexibility to enhance the realism of the model with respect to firmly established but poorly quantified associations between

[1]Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, 100 Galvin Hall, Notre Dame, Indiana 46556, USA. [2]WorldPop Project, Department of Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK. [3]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [4]Flowminder Foundation, SE-11355 Stockholm, Sweden. *e-mail: taperkins@nd.edu

© 2016 Macmillan Publishers Limited. All rights reserved

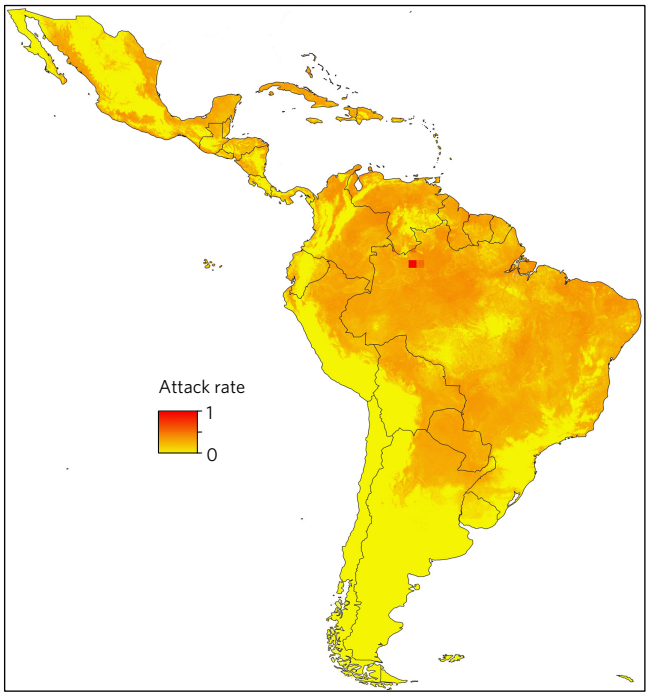


Figure 1 | Model-based projections of location-specific epidemic attack rates. Values shown in each 5 km × 5 km grid cell indicate the median of 1,000 values for that grid cell drawn from independent Monte Carlo samples of model parameters.

human–mosquito contact and economic prosperity[20]. In doing so, one departure from the classic relationship between R_0 and attack rate that we made was to rescale R_0 by an exponent $\alpha \in (0, 1]$ to allow for better correspondence with observed attack rates. Although there is no theoretical justification for this or any other particular scaling relationship, it is consistent with theoretical expectations[21] that attack rates should be lower in populations with equal R_0 values but more heterogeneous contact patterns, which are typical for transmission by *Ae. aegypti*[22]. To provide a point of reference for our model-based approach, we also fitted a statistical description of the 13 seroprevalence estimates as a function of the environmental drivers that we considered. For both approaches, we applied their respective location-specific attack rate projections to demographic projections on a 5 km × 5 km grid across Latin America and the Caribbean to obtain the expected numbers of infections in the overall population and among childbearing women in particular (Fig. 2a). All such calculations were performed for 1,000 Monte Carlo samples of model parameters.

In total, our median projection suggests that as many as 93.4 (range: 81.6–117.1) million people in Latin America and the Caribbean could become infected during the first wave of the epidemic (Table 1). To place this number into context, we refer to an estimate[23] that 53.8 (40.0–71.8) million dengue infections occurred in this region in 2010 alone. Our projections of nearly double this number for Zika are not surprising, given that there is extensive immunity to dengue but not Zika in this region and given that it would probably take longer than a year for the first wave of the epidemic to conclude in all locations within this region. At the country level, we project that Brazil will have the largest total number of infections by more than double that of any other country, due to a combination of its size and suitability for transmission. Island countries in the Caribbean are projected to experience the highest nationally averaged attack rates, with seven of the highest ten values projected for countries including Aruba, Haiti and Cuba. This projection is consistent with a frequent history of arbovirus

Table 1 | Country-level totals of median location-specific projections of Zika virus infections in the first-wave epidemic.

Country	Total (×10^6)	Births (×10^3)
Brazil	37.4 (30.6–49.2)	579 (473–755)
Mexico	14.9 (12.8–17.5)	263 (226–310)
Venezuela	7.4 (6.6–8.2)	144 (130–162)
Colombia	6.7 (6.0–8.1)	124 (110–153)
Cuba	3.7 (3.3–4.1)	33 (30–37)
Haiti	2.9 (2.5–4.5)	73 (63–121)
Argentina	2.7 (1.6–6.6)	44 (27–108)
Dom. Rep.	2.6 (2.4–2.8)	61 (56–68)
Others	15.6 (13.3–20.8)	336 (285–453)
Total	**93.4 (81.6–117.1)**	**1,650 (1,449–2,062)**

outbreaks on islands[24] and may be due to the uniformity of environmental conditions on the portions of islands where people tend to live. In more heterogeneous regions, the 5 km × 5 km spatial resolution of our maps allows for nuanced projections for areas of interest to local stakeholders (Fig. 2b,c). To facilitate the use of these local projections, we have made 5 km × 5 km minimum, median and maximum projections of attack rates, total infections and infections among childbearing women publically available online (http://github.com/TAlexPerkins/Zika_nmicrobiol_2016).

Among childbearing women, our median projection suggests that there could be as many as 1.65 (range: 1.45–2.06) million infections in Latin America and the Caribbean before the first wave of the epidemic concludes (Table 1). Assuming that birth rates are temporally constant, our projections are robust to uncertainty about the timing of local epidemics and the timeframe of the first wave of the epidemic, because they are based on cumulative proportions infected. These projections can also be used to postulate numbers at risk of microcephaly by multiplying them by the fraction of a year in which a pregnant woman is susceptible to developing microcephaly (for example, multiply by 1/4 in the case of first-trimester susceptibility). We also note that there were some discrepancies in our projections in terms of the rank order of countries experiencing the most infections among childbearing women versus the population as a whole. In particular, Cuba was fifth in terms of projected infections in the overall population but twelfth in terms of infections among childbearing women due to its low birth rate compared to other countries in the Americas[25]. Such discrepancies are also likely to exist subnationally[26], and their elucidation should be a priority for future work.

By accounting for uncertainty distributions for each of the key drivers of our model (Fig. 3a–e), we found that uncertainty distributions for infections across the region as a whole and by country were often multimodal (Fig. 3f–o) due to uncertainty in the shape of the relationship between mosquito–human contact and the local economic index that we considered (Fig. 3d). Summing our projections across Latin America and the Caribbean revealed variation that was modest, in the sense that none of our 1,000 Monte Carlo samples resulted in fewer than 81 million infections overall and 1.4 million among childbearing women (Fig. 3f,k). There are many reasons that even these numbers could be overestimates though. Our projections are conditioned on a local epidemic taking place in each 5 km × 5 km grid cell in the region, which is unlikely to happen given dispersal limitation, stochastic fadeout, geographic mismatches in seasonality and other factors. Therefore, it is most appropriate to interpret our projections as either a plausible worst-case scenario or an expectation of local epidemic size conditional on there being a local epidemic in the first place.

Although our approach was very much rooted in mechanistic models from epidemiological theory, two critical steps in our method involved fitting curves to describe theoretically motivated but heretofore unknown relationships: an association between

© 2016 Macmillan Publishers Limited. All rights reserved

NATURE MICROBIOLOGY DOI: 10.1038/NMICROBIOL.2016.126 LETTERS

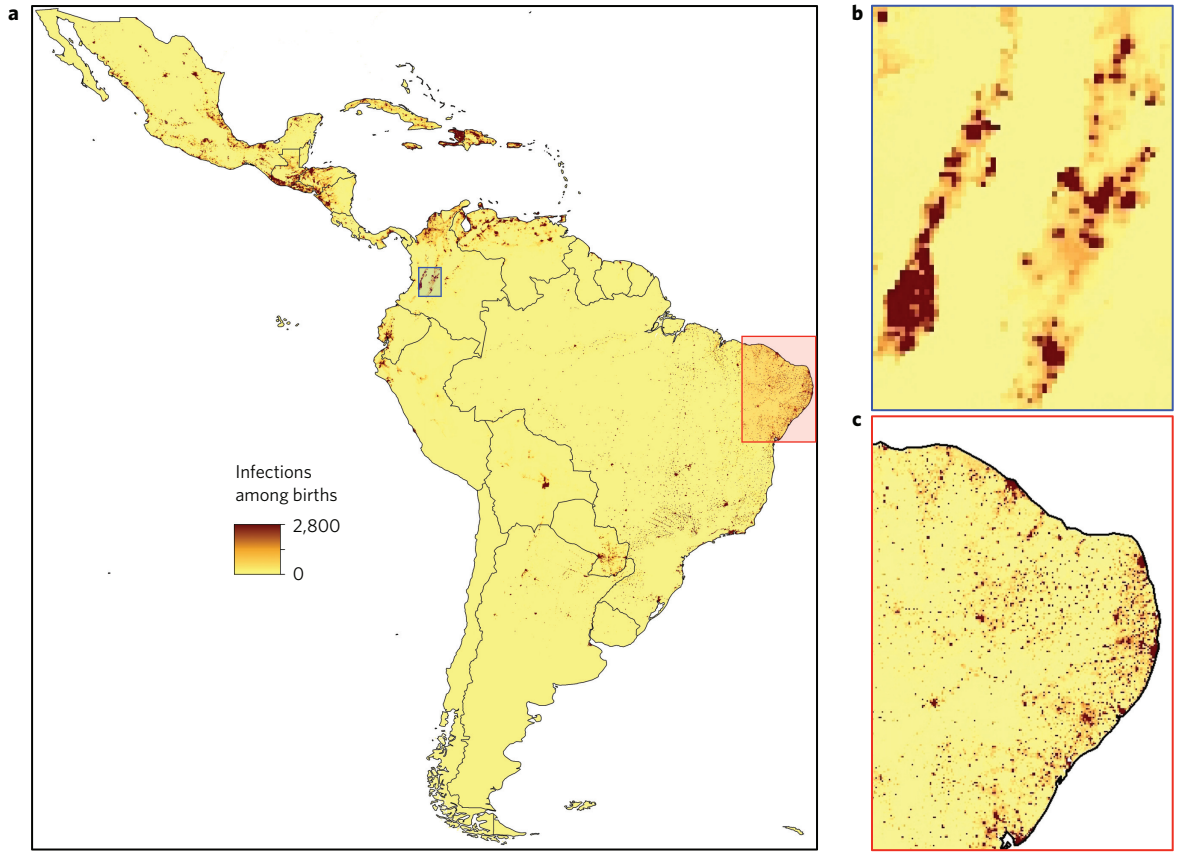


Figure 2 | Model-based projections of location-specific expected numbers of Zika virus infections among childbearing women. Values shown in each 5 km × 5 km represent projections of median epidemic attack rates from Fig. 1 multiplied by demographic projections[14]. **a**, Projections across Latin America and the Caribbean as a whole. **b**,**c**, More detailed projections for two areas: Cali, Colombia (**b**) and Recife, Brazil (**c**).

mosquito–human contact and economic prosperity (Fig. 3d) and a scaling relationship between R_0 and attack rates (Fig. 3e). Allowing these relationships to be informed by local seroprevalence estimates (Supplementary Table 1) left open the question of the extent to which our projections were informed by the mechanistic assumptions of the model versus statistical fits to the seroprevalence estimates that we used. On the one hand, an alternative statistical approach accounted for much more variation in seroprevalence estimates ($R^2 = 0.89$) than did the model-based approach ($R^2 = 0.32$). On the other hand, the statistical approach offered a dichotomous set of projections about numbers of infections outside the context of the data to which it was fitted: either everyone will become infected or very few people will (Fig. 4). Relationships between attack rates and predictor variables inferred by the statistical approach (Fig. 5d–i) were also implausible: a narrow temperature range in which attack rates increase sharply towards 100% (Fig. 5d–h) and a reversal of economic effects whereby wealthy populations experience higher attack rates than poor populations when mosquito occurrence probabilities are high (Fig. 5f,i). By contrast, the model-based approach yielded more moderate attack rate projections overall (Figs 3f versus 4a) in which temperature, economic prosperity and mosquito occurrence probability all had plausible relationships with attack rates (Fig. 5a–c).

In conclusion, our model-based approach offers a unique way to leverage a variety of spatially detailed data products[7,14,27,28] to make *a priori* projections of attack rates and infections that could be experienced in the first wave of the ongoing Zika epidemic. Projections such as these have an important role to play in the early stages of an epidemic, when planning for surveillance and outbreak response is actively under way both internationally and locally[9]. At the same time, it is important for consumers of this information to be aware of uncertainties in these and other projections, which often exceed the amount of uncertainty that can be identified *a priori*[29]. Similarly, following up on these projections in the aftermath of the epidemic—by comparing against projections made with alternative models and additional serological surveys[30]—will provide an exceptional opportunity to enhance capabilities to anticipate the severity of future epidemic threats.

Methods

Data sources and processing. *Human demography.* To estimate the annual numbers of pregnancies per 1 km × 1 km grid cell in 2015, methods developed by the WorldPop project (www.worldpop.org)[25,31] were adapted for the Americas region. High-resolution estimates of population counts per 100 m × 100 m grid cell for 2015 were recently constructed for Latin American, Asian and African countries[14,32]. With consistent subnational data on sex and age structures, as well as subnational age-specific fertility rate data across the Americas currently unavailable for fully replicating the approaches of Tatem and colleagues[31], national-level adjustments were made to construct pregnancy and birth counts. Data on estimated total numbers of births[33] and pregnancies[31] occurring annually in 2012 were assembled for all Latin American study countries, as well as births in 2015 (ref. 33). As no 2015 pregnancy estimates existed at the time of writing, the ratios of births to pregnancies for each country in the Americas were calculated using 2011 and 2012 estimates, and these were then applied to the 2015 birth numbers to obtain 2015 estimates of annual pregnancy numbers per country. This made the assumption that per-country births-to-pregnancies ratios remained the same in 2015 as they were in 2011 and 2012. The 100 m × 100 m gridded population totals were aggregated to 1 km × 1 km spatial resolution and the per-country totals were linearly adjusted to match the 2015 pregnancy estimates.

Temperature. We used interpolated meteorological station temperature data from the 1950–2000 period at 5 km × 5 km spatial resolution, processed to create climatological monthly averages that represent 'typical' conditions (www.worldclim.org)[27].

Ae. aegypti *occurrence probability.* To predict the likely distribution of *Ae. aegypti* mosquitoes, Kraemer *et al.*[7] generated high-resolution occurrence probability

© 2016 Macmillan Publishers Limited. All rights reserved

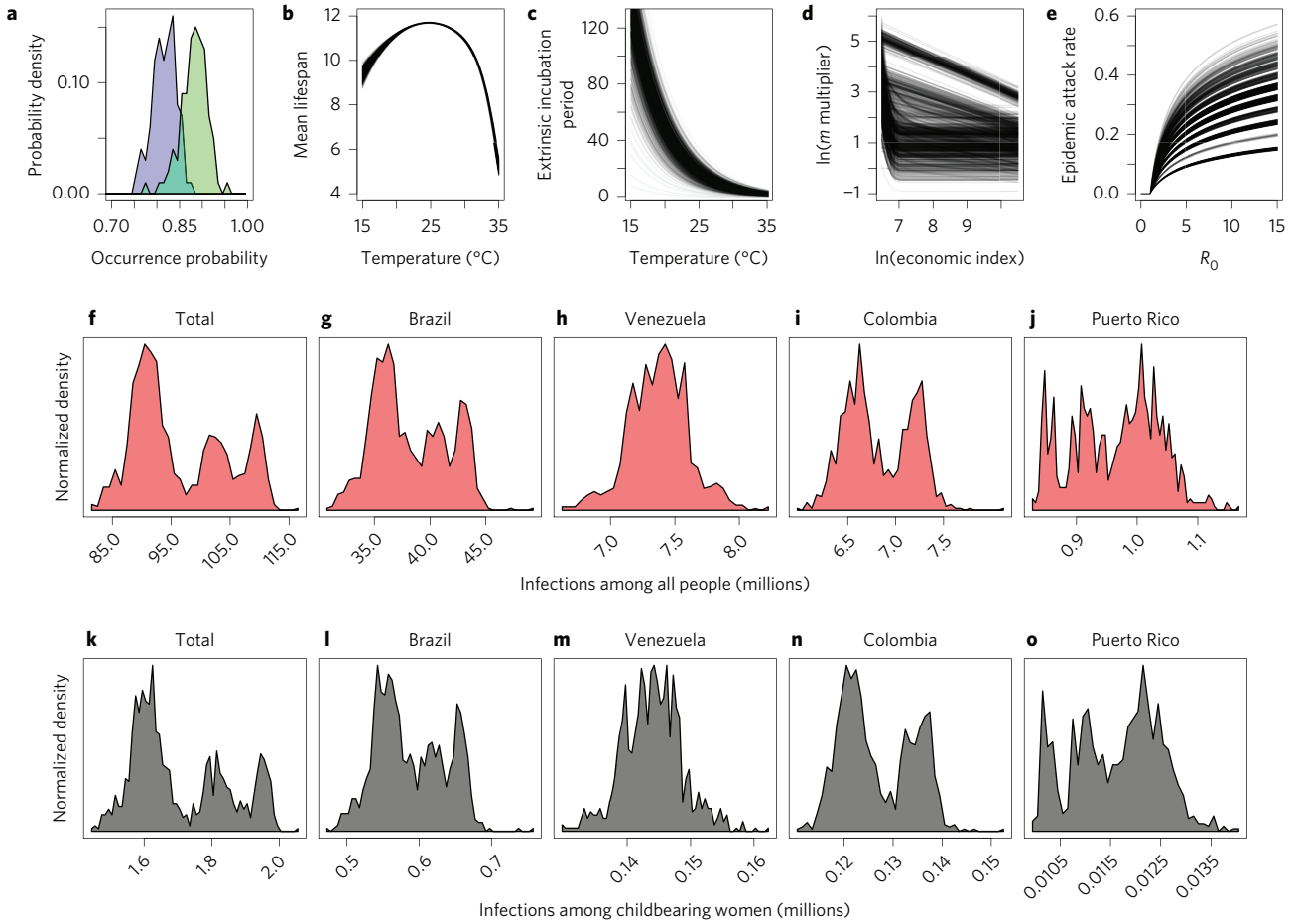


Figure 3 | Uncertainties in model-based projections of Zika virus infections. a–o, Monte Carlo samples from the uncertainty distributions surrounding each of the key drivers in the model (**a–e**) and uncertainty distributions for projected numbers of infections among all individuals (**f–j**) and among childbearing women (**k–o**) in different areas. All panels reflect the full range of uncertainty considered in 1,000 Monte Carlo samples. **a**, Posterior distributions of mosquito occurrence probabilities for two example 5 km × 5 km grid cells.

surfaces based on a species distribution modelling approach[11]. More specifically, a boosted regression tree model was applied using a comprehensive set of known occurrences (n = 19,930) of *Ae. aegypti* and a set of environmental predictors known to influence the distribution of the species[7]. Covariates included a temperature suitability index[17], contemporary mean and range maps of the Enhanced Vegetation Index and precipitation[34], and an urbanization index from the Global Rural Urban Mapping Project. We used a set of 100 spatial layers sampled from the posterior distribution estimated by Kraemer *et al.*[7]

Economic index. To account for socio-economic differences among populations residing in different regions, we used one-degree-resolution gridded estimates of purchasing power parity (PPP) in US$ from 2005 adjusted for inflation (G-Econ)[28]. When we encountered missing values, we imputed values in one of two ways. Grid cells in small island countries with data missing for the entire country were uniformly filled with population-adjusted PPP figures obtained from the US CIA World Factbook[35]. Missing values in continental grid cells were imputed with the mean of the surrounding eight grid cell values. Once we obtained a complete PPP grid layer at one-degree resolution, we resampled the layer to a resolution of 5 km × 5 km to match the resolution of gridded layers for human demography, temperature and *Ae. aegypti* occurrence probability.

Seroprevalence estimates. To calibrate our model, we identified published estimates of seroprevlance that were relevant to the context of our study (Supplementary Table 1). Specifically, we sought estimates of seroprevalence to either Zika or chikungunya viruses in populations that were presumably naive before an outbreak. Thus, we excluded some seroprevalence estimates that were obtained from endemic populations. We also excluded estimates from small islands—namely, Reunion and Grande Comore—for which it was clear that gridded temperature data were unrealistically low due to steep elevational gradients and other features of island geography. Although the focus of our analysis was on Latin America and the Caribbean, we were not able to exclude locations on the basis of location given that only 2 of 13 came from the focal region. Appropriately, however, a number of the seroprevalence estimates we obtained pertained specifically to pregnant women, although there did not appear to be differences in the seroprevalence of pregnant women and the population at large, at least in the context of a naive population following an outbreak[36].

Calculation of derived quantities. *Mosquito abundance.* Occurrence probabilities can be translated into proxies for abundance provided that an assumption is made about how abundance is distributed as a random variable[37]. Assuming that mosquito abundance is distributed as a Poisson random variable, the probability that there is at least one mosquito present in a given location is $1 - \exp(-\lambda)$, where λ is the expected abundance of mosquitoes. Inverting this relationship, we obtained an estimate $\lambda = -\ln(1 - \text{occurrence probability})$ of expected mosquito abundance under the Poisson model and used this as a proxy for mosquito abundance in our calculations.

Mosquito–human ratios. The estimates of mosquito occurrence probability that we used incorporated a number of environmental variables[7]. They did not account for factors that modulate contact between mosquitoes and humans, however. Due in part to economic differences, factors such as air conditioning and piped water can drastically limit mosquito–human contact and virus transmission, even when mosquitoes are abundant[20]. We accounted for the effect of economic differences between locations by multiplying our proxy for mosquito abundance λ by a multiplication factor, which we specified as a function of the aforementioned economic index. We specified the relationship between the economic index and the multiplication factor by using a shape constrained additive model (SCAM[38]). This allowed for flexibility in the shape of this relationship but constrained it such that the multiplication factor (and thus, presumed mosquito–human contact) could only have a monotonically decreasing relationship with increasing values of the economic index. The specific shape of this relationship was determined by fitting it to values of the multiplication factors that would be necessary for modelled attack rates to perfectly match published seroprevalence estimates.

© 2016 Macmillan Publishers Limited. All rights reserved

NATURE MICROBIOLOGY DOI: 10.1038/NMICROBIOL.2016.126 LETTERS

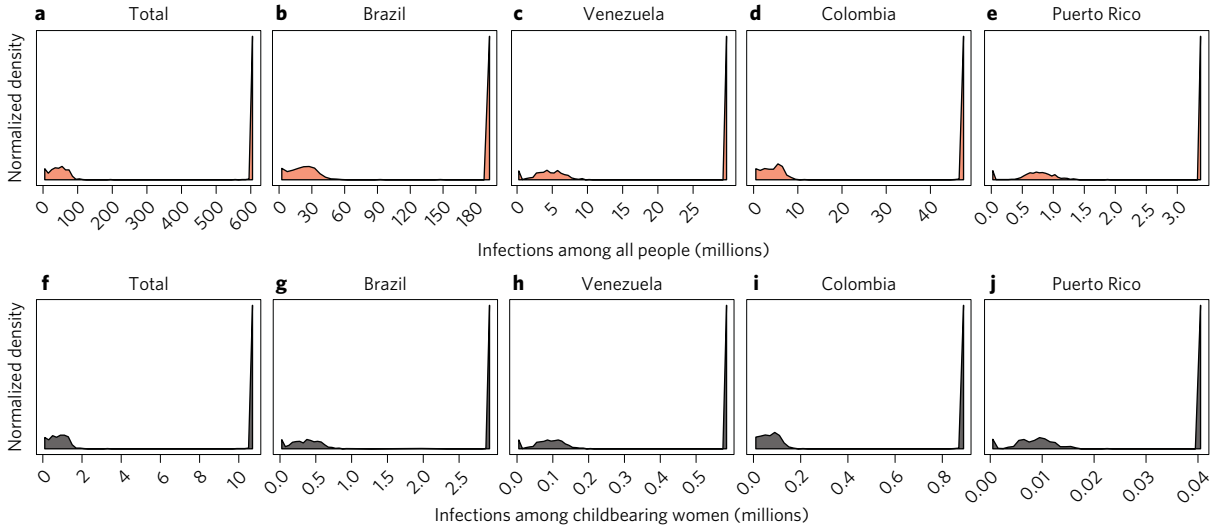


Figure 4 | Uncertainties in statistical projections of Zika virus infections. a–j, Projected numbers of infections among all individuals (**a–e**) and among childbearing women (**f–j**) for 1,000 Monte Carlo samples from the uncertainty distribution around parameters of the statistical model that was fitted to seroprevalence estimates from Supplementary Table 1.

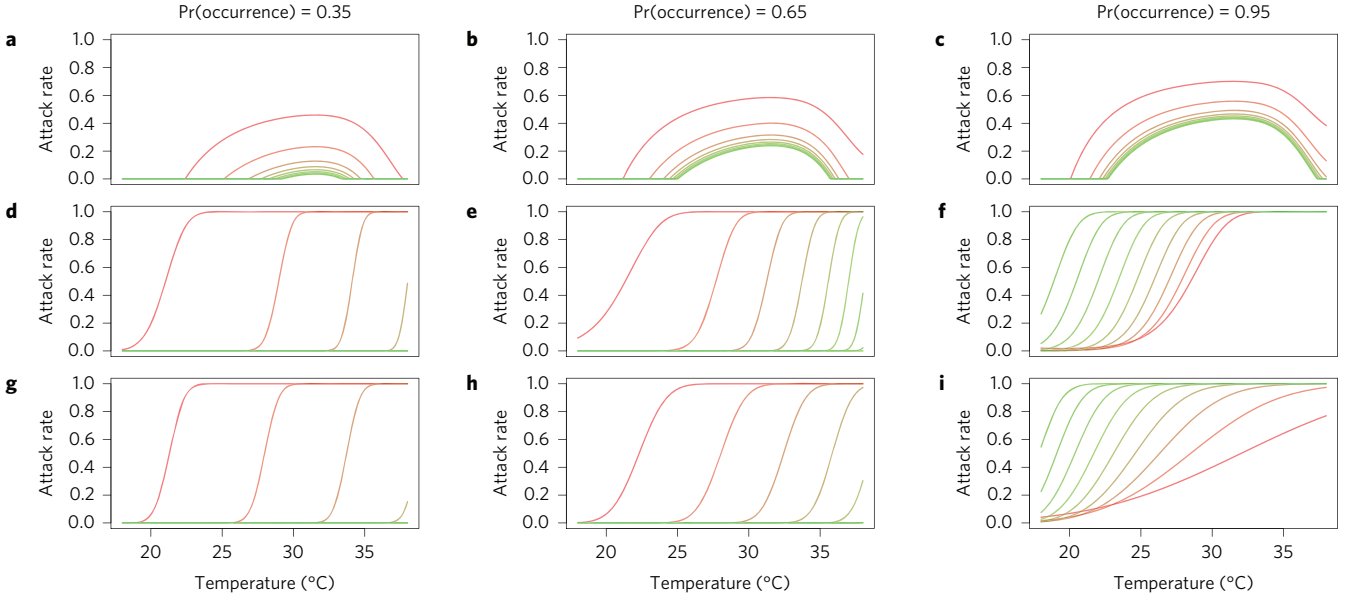


Figure 5 | Relationships between temperature (*x* axis), economic index (red = low, green = high), mosquito occurrence probabilities (columns) and projected epidemic attack rates (*y* axis). These relationships are shown for the model-based approach (**a–c**), the full statistical approach (**d–f**) and the version of the statistical approach chosen by stepwise model selection (**g–i**). Pr(occurrence) represents the local probability of occurrence of *Ae. aegypti*.

Basic reproduction number R_0. We calculated the basic reproduction number R_0 according to its classic Ross–Macdonald formulation and as a function of temperature T,

$$R_0(T) = \frac{mbca^2 e^{-\mu(T)n(T)}}{\mu(T)r} \quad (1)$$

with adult mosquito mortality μ and extrinsic incubation period n specified as functions of temperature. Because temperature values were available for each location on a monthly basis, we computed monthly values of R_0 for each location and then used the mean of the highest six monthly values of R_0 as a singular estimate of R_0 for each location. This approach was broadly consistent with the way in which a temperature suitability index was used to inform mosquito occurrence probabilities by Kraemer and co-authors[7].

For mosquito mortality, we used the temperature- and age-dependent model of Brady and colleagues[39], to which we added an additional force of extrinsic mortality (0.025 d^{-1}) to match an overall daily mortality value of 0.115 estimated in a mark–release–recapture experiment carried out under temperatures ranging from 20 to 34 °C (ref. 40). We then computed the mean of the age- and temperature-dependent lifespan distribution as a function of temperature to inform $\mu(T)$. For the relationship between temperature and mean duration of the extrinsic incubation period, we used the temperature-dependent exponential rate estimated by Chan and Johansson[18]. The ratio of mosquitoes to humans, m, was quantified using a combination of occurrence probabilities and the gross cell product economic index, as described in the previous two sections. Parameters that did not depend on temperature were set at the following values according to published estimates for *Ae. aegypti* and dengue virus: mosquito-to-human transmission probability, $b = 0.4$ (ref. 41); human-to-mosquito transmission probability times number of days of human infectiousness, $c/r = 3.5$ (ref. 42); mosquito biting rate, $a = 0.67$ (ref. 43). Although there is uncertainty around these parameter values, any such uncertainty was effectively subsumed by fitting m to seroprevalence data given that bca^2/r entered R_0 as a constant.

Attack rates under a model-based formulation. Under a susceptible–infected–recovered (SIR) transmission model, there is a one-to-one relationship between R_0 and final epidemic size, which is equivalent to the attack rate over the course of an epidemic[13]. Intuitively, the final epidemic size is reached once herd immunity is sufficient to limit contacts between infectious and susceptible individuals to the

© 2016 Macmillan Publishers Limited. All rights reserved

extent necessary to reduce the pathogen's force of infection to zero. There is no explicit solution for final epidemic size as a function of model parameters, but it can be calculated numerically by obtaining an implicit solution of $S_\infty = e^{-R_0(1-S_\infty)}$ for S_∞, which is the proportion remaining susceptible after the epidemic has burned out[13]. Under the assumptions of the SIR model, the attack rate over the course of an epidemic is $AR = 1 - S_\infty$.

To apply this theoretical insight to Zika or other mosquito-borne pathogens, several limiting assumptions of the SIR model must first be reconciled. One such assumption is that individuals become infectious immediately upon becoming infected and remain infectious for an exponentially distributed period of time[44]; mosquito-borne pathogens such as Zika virus are instead characterized by a distinct lag between human and mosquito infection[45]. Despite this discrepancy between assumptions of the SIR model and the reality of many pathogen systems, mathematical analyses[46] have shown that final epidemic size is insensitive to details about the shape of the distribution that characterizes the time period between successive cases (that is, the generation interval).

Another limiting assumption of the SIR model is that of homogeneous encounters between people and mosquitoes[44], which are understood to be extensive for mosquito-borne diseases[22]. Mathematical analyses[21] in this case show that a seemingly infinite complexity of relationships between R_0 and final epidemic size are possible in a heterogeneous system. As a general rule, however, final epidemic size in a system with contact heterogeneity and proportional mixing is expected to be strictly less than the final epidemic size in an otherwise equivalent system with homogeneous contacts[21]. How the ratio of these final epidemic sizes scales as a function of R_0 depends entirely on the details of a given system and would therefore be extremely difficult to generalize without copious data on mosquito–human contact and further investigation, which is beyond the scope of our study.

To capture the potentially very strong effects of heterogeneity in reducing final epidemic size in populations subject to Zika epidemics, we scaled the final epidemic size by substituting R_0^α for R_0 in the SIR-based final epidemic size formula given some constant $\alpha \in (0, 1]$. Although there is no theoretical justification for this or any other choice of how to scale R_0 and AR in the presence of contact heterogeneity, the choice we made has the following desirable properties: (1) it implies that $AR \to 1$ as $R_0 \to \infty$; (2) it leads to the function $AR(R_0)$ having a more gradual slope and thereby allows for intermediate attack rates to be more common than they would be otherwise; (3) it preserves the property that $AR = 0$ for $R_0 < 1$. At the same time, this and possible alternative formulations are limited by a general lack of understanding about the relationship between R_0 and AR in heterogeneous systems, relationships that may furthermore be heterogeneous themselves across different areas[47].

To estimate α, we performed the following procedure for candidate values of α between 0.01 and 1 in increments of 0.01: (1) calculate R_0 according to equation (1) and assume $m = \lambda$ for each of the 13 sites from which seroprevalence estimates were derived; (2) use those R_0 values to calculate AR values for each of those sites based on the classic SIR formulation; (3) calculate what multiplication factor of R_0 would be necessary for AR to match the empirical seroprevalence estimate; (4) fit a SCAM model of the economic index to the multiplication factors; and (5) use the fitted SCAM values to recalculate R_0 and then AR for each site. Next, we calculated the sum of squares between the final predicted AR values associated with each α and the empirical seroprevalence estimates and we then selected the value of α that minimized the sum of squares. Supplementary Fig. 1 illustrates this process given mean estimates of λ from *Ae. aegypti* occurrence probabilities, $\mu(T)$ and $n(T)$.

Attack rates under a statistical formulation. As an alternative to our model-based characterization of attack rates, we also considered a purely statistical approach that modelled probit-transformed seroprevalence observations as functions of averaged monthly temperatures, *Ae. aegypti* occurrence probabilities and the economic index. We considered all combinations of linear, quadratic and pairwise interaction terms of these variables, comparing them on the basis of the Akaike Information Criterion using the lm and step functions in R (ref. 48). Although additional functional forms would have been of interest, this suite of models was as complex as the limited set of 13 seroprevalence observations would support.

Quantifying uncertainty around attack rate projections. To quantify uncertainty associated with our projections, we generated 1,000 Monte Carlo samples from the uncertainty distributions of each model parameter as described in each of the references[7,17,18] in which those parameters were originally described. For $\mu(T)$ and $n(T)$, we took random draws of their parameters consistent with published descriptions of uncertainty in the parameters of those functions from their original sources[17,18]. For *Ae. aegypti* occurrence probabilities, we drew randomly with replacement from 100 sample layers from the posterior distribution[7]. For the relationship involving the economic index and the R_0 scaling factor α, we used best-fit SCAM models and α values corresponding to each set of random draws of the parameters of $\mu(T)$, $n(T)$ and λ from the *Ae. aegypti* layers. For each of the 1,000 Monte Carlo samples of the statistical model, we performed resampling with replacement among the 13 seroprevalence values, performed the same model fitting and model selection procedure described in the previous section, and took a multivariate normal random sample of the parameter values of the best-fit model based on the model's best-fit parameters and variance–covariance matrix.

Projecting attack rates and numbers of infections. To obtain estimates of the numbers of infections in total and among childbearing women for the model-based and statistical approaches, we multiplied their respective attack rate projections applied to 5 km × 5 km grids across Latin America and the Caribbean by human demographic layers for total population and births in 2015. For both the model-based and statistical approaches, we performed these calculations and summed at the country level once for each of the 1,000 Monte Carlo samples that we produced. High-resolution spatial projections of attack rates and numbers of infected childbearing women under the model-based approach are presented in Supplementary Figs 2–10. Most projections based on the statistical approach resulted in attack rates of 100% in nearly all locations throughout Latin America and the Caribbean.

Code availability. Code in the R language for reproducing all analyses is available at http://github.com/TAlexPerkins/Zika_nmicrobiol_2016.

Received 17 February 2016; accepted 28 June 2016; published 25 July 2016

References

1. Mlakar, J. *et al.* Zika virus associated with microcephaly. *New Engl. J. Med.* **374,** 951–958 (2016).
2. Brasil, P. *et al.* Zika virus infection in pregnant women in Rio de Janeiro—preliminary report. *New Engl. J. Med.* http://dx.doi.org/10.1056/NEJMoa1602412 (2016).
3. Messina, J. P. *et al.* Mapping global environmental suitability for Zika virus. *eLife* **5,** 15272 (2016).
4. Cauchemez, S. *et al.* Association between Zika virus infection and microcephaly in French Polynesia, 2013–2015: a retrospective study. *Lancet* **387,** 2125–2132 (2016).
5. Johansson, M. A. *et al.* Risk estimates for microcephaly related to Zika virus infection—from French Polynesia to Bahia, Brazil. Preprint at http://dx.doi.org/10.1101/051060 (2016).
6. *WHO Statement on the First Meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika Virus and Observed Increase in Neurological Disorders and Neonatal Malformations* (WHO, 2016); http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/
7. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus. eLife* **4,** e08347 (2015).
8. *Zika—Epidemiological Update* (Pan American Health Organization, 2016).
9. Lucey, D. R. & Gostin, L. O. The emerging Zika pandemic: enhancing preparedness. *J. Am. Med. Assoc.* **315,** 865–866 (2016).
10. The next steps on Zika. *Nature* **530,** 5 (2016).
11. Elith, J. & Leathwick, J. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. System.* **40,** 677–697 (2009).
12. Kermack, W. & McKendrick, A. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* **115,** 700–721 (1927).
13. Keeling, M. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton Univ. Press, 2007).
14. Sorichetta, A. *et al.* High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Sci. Data* **2,** 150045 (2015).
15. Sergon, K. *et al.* Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. *Am. J. Trop. Med. Hyg.* **78,** 333–337 (2008).
16. Smith, D. L. *et al.* Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLOS Pathogens* **8,** e1002588 (2012).
17. Brady, O. J. *et al.* Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasites Vector.* **7,** 338 (2014).
18. Chan, M. & Johansson, M. A. The incubation periods of dengue viruses. *PLoS ONE* **7,** e50972 (2012).
19. Funk, S. *et al.* Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. Preprint at http://dx.doi.org/10.1101/043265 (2016).
20. Reiter, P. *et al.* Texas lifestyle limits transmission of dengue virus. *Emerg. Infect. Dis.* **9,** 86–89 (2003).
21. Andreasen, V. The final size of an epidemic and its relation to the basic reproduction number. *Bull. Math. Biol.* **73,** 2305–2321 (2011).
22. Liebman, K. A. *et al.* Determinants of heterogeneous blood feeding patterns by *Aedes aegypti* in Iquitos, Peru. *PLoS Negl. Trop. Dis.* **8,** e2702 (2014).
23. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
24. Feldstein, L., Brownstein, J., Brady, O. J., Hay, S. I. & Johansson, M. A. Dengue on islands: a Bayesian approach to understanding the global ecology of dengue viruses. *Trans. R. Soc. Trop. Med. Hyg.* **109,** 303–312 (2015).
25. *The State of the World's Midwifery 2014* (UNFPA, 2014).
26. Tatem, A. J. *et al.* Millenium development health metrics: where do Africa's children and women of childbearing age live? *Popul. Health Metrics* **11,** 11 (2013).

© 2016 Macmillan Publishers Limited. All rights reserved

27. Hijmans, R. *et al.* Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25,** 1965–1978 (2005).
28. Nordhaus, W. Geography and macroeconomics: new data and new findings. *Proc. Natl Acad. Sci. USA* **103,** 3510–3517 (2006).
29. Taylor, B. P. *et al.* Stochasticity and the limits to confidence when estimating R_0 of Ebola and other emerging infectious diseases. Preprint at http://arxiv.org/abs/1601.06829 (2016).
30. Metcalf, C. J. E. *et al.* Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet* http://dx.doi.org/10.1016/S0140-6736(16)30164-7 (2016).
31. Tatem, A. J. *et al.* Mapping for maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *Int. J. Health Geogr.* **13,** 2 (2013).
32. Stevens, F. R. *et al.* Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* **10,** e0107042 (2015).
33. *World Population Prospects: The 2015 Revision* (UN, 2015).
34. Weiss, D. *et al.* An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogr. Remote Sens.* **98,** 106–118 (2014).
35. *The World Factbook* (US Central Intelligence Agency, 2016); https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html
36. Gerardin, P. *et al.* Estimating chikungunya prevalence in Reunion Island outbreak by serosurveys: two methods for two critical times of the epidemic. *BMC Infect. Dis.* **8,** 99 (2008).
37. Wright, D. Correlations between incidence and abundance are expected by chance. *J. Biogeogr.* **18,** 463–466 (1991).
38. Pya, N. & Wood, S. Shape constrained additive models. *Statist. Comput.* **25,** 543–559 (2015).
39. Brady, O. J. *et al.* Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasites Vector.* **6,** 351–362 (2013).
40. Muir, L. & Kay, B. *Aedes aegypti* survival and dispersal estimated by mark–release–recapture in northern Australia. *Am. J. Trop. Med. Hyg.* **58,** 277–282 (1998).
41. Nishiura, H. & Halstead, S. Natural history of dengue virus (DENV)-1 and DENV-4 infections: reanalysis of classic studies. *J. Infect. Dis.* **195,** 1007–1013 (2007).
42. Scott, T. W. *et al.* Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. *J. Med. Entomol.* **37,** 89–101 (2000).
43. Reiner, R. C. *et al.* A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *J. R. Soc. Interface* **10,** 20120921 (2013).
44. Lessler, J. *et al.* Times to key events in the course of Zika infection and their implications for surveillance: a systematic review and pooled analysis. Preprint at http://dx.doi.org/10.1101/041913 (2016).
45. Ma, J. & Earn, D. J. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* **68,** 679–702 (2006).
46. Kraemer, M. U. G. *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12,** 20150468 (2015).
47. R Core Team *R: A Language for Statistical Computing (R Foundation for Statistical Computing*, 2014); http://www.R-project.org/
48. Duffy, M. *et al.* Zika virus outbreak on Yap Island, Federated States of Micronesia. *New Engl. J. Med.* **360,** 2536–2543 (2009).
49. Schwarz, N. *et al.* Seroprevalence of antibodies against chikungunya, dengue, and Rift Valley fever viruses after febrile illness outbreak, Madagascar. *Emerg. Infect. Dis.* **18,** 1780–1786 (2012).
50. Sissoko, D. *et al.* Outbreak of chikungunya fever in Mayotte, Comoros archipelago, 2005–2006. *Trans. R. Soc. Trop. Med. Hyg.* **102,** 780–786 (2008).
51. Dwibedi, B. *et al.* Rapid spread of chikungunya virus infection in Orissa, India. *Indian J. Med. Res.* **133,** 316–321 (2011).
52. Gay, N. *et al.* Seroprevalence and asymptomatic rates of Asian lineage chikungunya virus infection on Saint Martin, Caribbean. *Am. J. Trop. Med. Hyg.* **94,** 393–396 (2015).
53. Moro, M. *et al.* Chikungunya virus in north-eastern Italy: a seroprevalence study. *Am. J. Trop. Med. Hyg.* **82,** 508–511 (2010).
54. Balmaseda, A. *et al.* Clinical attack rate of chikungunya in a cohort of Nicaraguan children. *Am. J. Trop. Med. Hyg.* **94,** 397–399 (2016).

Acknowledgements

The authors thank three anonymous reviewers, as well as J. Ashander, C.M. Barker, M.A. Johansson, R.C. Reiner, S.T. Stoddard, J.C. Miller and members of the Perkins Lab for discussions. The authors thank O.J. Brady for sharing code for calculating mosquito mortality as a function of temperature. T.A.P., A.S.S. and A.J.T. are supported by funding from the National Science Foundation (DEB 1641130). C.W.R. is supported by funding through the University of Southampton's Economic and Social Research Council's Doctoral Training Centre. M.U.G.K. receives funding from the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS; European Commission 7th Framework Programme, 21893). A.J.T. is supported by funding from NIH/NIAID (U19AI089674), the BMGF (OPP1106427, 1032350), NORAD and a Wellcome Trust Sustaining Health Grant (106866/Z/15/Z). A.J.T. and C.W.R. acknowledge the support of the WorldPop (www.worldpop.org) and Flowminder Foundation (www.flowminder.org) teams in demographic data set production, and T.A.P. and A.S.S. acknowledge support from the Notre Dame Center for Research Computing.

Author contributions

T.A.P. conceived the research, designed the analysis and wrote the first draft of the manuscript. A.S.S. assembled data, performed calculations and contributed to writing. C.W.R. assembled data, produced map visuals and contributed to writing. M.U.G.K. assembled data and contributed to writing. A.J.T. contributed to the analysis, map visuals and writing.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A.P.

Competing interests

The authors declare no competing financial interests.

© 2016 Macmillan Publishers Limited. All rights reserved

EBioMedicine xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Assessing Seasonal Risks for the Introduction and Mosquito-borne Spread of Zika Virus in Europe

Joacim Rocklöv [a,*], Mikkel Brandon Quam [a,1], Bertrand Sudre [b,1], Matthew German [c,1], Moritz U.G. Kraemer [d], Oliver Brady [e], Isaac I. Bogoch [f,g], Jing Liu-Helmersson [a], Annelies Wilder-Smith [a,h], Jan C. Semenza [b], Mark Ong [i], Kaja Kaasik Aaslav [b], Kamran Khan [c,f]

[a] Department of Public Health and Clinical Medicine, Epidemiology and Global Health, Umeå University, Sweden
[b] European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
[c] Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada
[d] Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford; Oxford, UK
[e] Wellcome Trust Centre for Human Genetics, University of Oxford; Oxford, UK
[f] Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, ON, Canada
[g] Divisions of Internal Medicine and Infectious Diseases, University Health Network, Toronto, Canada
[h] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[i] Geomatics Program, University of Waterloo, Waterloo, ON, Canada

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 26 May 2016
Accepted 6 June 2016
Available online xxxx

Keywords:
ZIKV
Zika
Air travel
Globalization, mosquito
Climate
Aedes

ABSTRACT

The explosive Zika virus epidemic in the Americas is amplifying spread of this emerging pathogen into previously unaffected regions of the world, including Europe (Gulland, 2016), where local populations are immunologically naïve. As summertime approaches in the northern hemisphere, *Aedes* mosquitoes in Europe may find suitable climatic conditions to acquire and subsequently transmit Zika virus from viremic travellers to local populations. While *Aedes albopictus* has proven to be a vector for the transmission of dengue and chikungunya viruses in Europe (Delisle et al., 2015; ECDC, n.d.) there is growing experimental and ecological evidence to suggest that it may also be competent for Zika virus(Chouin-Carneiro et al., 2016; Grard et al., 2014; Li et al., 2012; Wong et al., 2013). Here we analyze and overlay the monthly flows of airline travellers arriving into European cities from Zika affected areas across the Americas, the predicted monthly estimates of the basic reproduction number of Zika virus in areas where *Aedes* mosquito populations reside in Europe (*Aedes aegypti* in Madeira, Portugal and *Ae. albopictus* in continental Europe), and human populations living within areas where mosquito-borne transmission of Zika virus may be possible. We highlight specific geographic areas and timing of risk for Zika virus introduction and possible spread within Europe to inform the efficient use of human disease surveillance, vector surveillance and control, and public education resources.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

On May 17th, 2015, the Program for Monitoring Emerging Diseases (ProMED-mail) published a report confirming locally acquired cases of Zika virus (ZIKV) in several northeastern Brazilian states, marking the first time this virus is known to have spread within the Americas (Promed, 2015). Eight months later, on February 1st, 2016 the World Health Organization declared the ZIKV epidemic in the Americas a Public Health Emergency of International Concern, in part due to an emerging association with congenital birth anomalies such as microcephaly (Calvet et al., 2016; Mlakar et al., 2016; Rodrigues, 2016) and Guillain-Barré syndrome (Cao-Lormeau et al., 2016). After the virus' introduction into Brazil, the epidemic has swiftly spread across Latin America and the Caribbean (Faria et al., 2016; Petersen et al., 2016a). Potential reasons for this rapid spread include the presence of immunologically naïve populations and an abundance of *Aedes* mosquitoes (Kraemer et al., 2015) within a conducive environment.

As the epidemic expands in scale and geographic range, a growing number of travellers are exporting ZIKV to other regions of the world, including Europe, where *Aedes* vectors are known to be present (Maria et al., 2016; Zammarchi et al., 2015; http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx, n.d.). In Europe, *Aedes aegypti* is known to exist on the island of Madeira, Portugal (http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/

* Corresponding author.
E-mail address: joacim.rocklov@umu.se (J. Rocklöv).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ebiom.2016.06.009
2352-3964/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Pages/VBORNET_maps.aspx, n.d.) and in parts of Georgia and southwestern Russia, whereas *Aedes albopictus* is established along much of the Mediterranean coast (http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx, n.d.). While virus importation events could trigger epidemics in distant geographies where competent *Aedes* mosquito vectors exist, this risk has to date, been mitigated by winter temperatures in the northern hemisphere.

Given the growing experimental and ecological evidence to suggest that *Ae. albopictus* may be a competent vector for ZIKV (Chouin-Carneiro et al., 2016; Grard et al., 2014; Li et al., 2012; Wong et al., 2013), health officials must plan for the possibility of locally acquired ZIKV infections in parts of Europe. The imminent arrival of summer in the northern hemisphere, when *Aedes* mosquito populations will peak and viral replication within these vectors will be most efficient, could lead to autochthonous transmission, not unlike the recent localized and transient European epidemics of dengue and chikungunya (Angelini et al., 2007; Wilder-Smith et al., 2014).

To assist public health decision-making, we (i) modeled the risks of ZIKV importation into Europe via airline travellers departing areas in the Americas where ZIKV activity has been confirmed or where suitable conditions exist for its transmission year round (Bogoch et al., 2016), (ii) used a temperature driven vectorial capacity model to quantify the potential for European *Aedes* mosquitoes to support autochthonous transmission of ZIKV, assuming that *Ae. albopictus* is a competent vector, and (iii) quantified the size of populations living in European areas where mosquito-borne transmission of ZIKV may be possible at the height of summer.

2. Materials & Methods

2.1. Overview

We developed a mathematical model that outputs basic reproduction numbers (R_0) for ZIKV transmission with biological anchoring in *Aedes* mosquito vectorial capacity and validation against surveillance data from the current ZIKV outbreak in Latin America and the Caribbean. We applied this vectorial capacity model to estimate R_0 potential for ZIKV in Europe this spring to autumn, while superimposing data on airline travellers arriving from areas in the Americas where ZIKV is active, as well as populations living in areas of Europe where mosquito-borne transmission of ZIKV is possible. Finally we discuss model assumptions and limitations in our approach.

2.2. Model Development

R_0 is used to characterize the epidemic potential of a pathogen. It represents the expected number of new infections generated by one infectious individual within a fully susceptible population. In the context of a mosquito-borne illness, R_0 is a function of vectorial capacity (VC), and the period of viremia in humans (T_h) (Anderson and R, 1991), given mathematically by: $R_0 = VC \times T_h$. Transmission increases when R_0 exceeds 1 (i.e. potential for an epidemic), and diminishes when R_0 is less than 1. VC in turn is a function of vector competence (inherent ability of the vector to transmit a particular pathogen), vector lifespan, and the extrinsic incubation period (Lambrechts et al., 2011). Since *Aedes* mosquitoes are ectotherms, VC is highly dependent upon mean

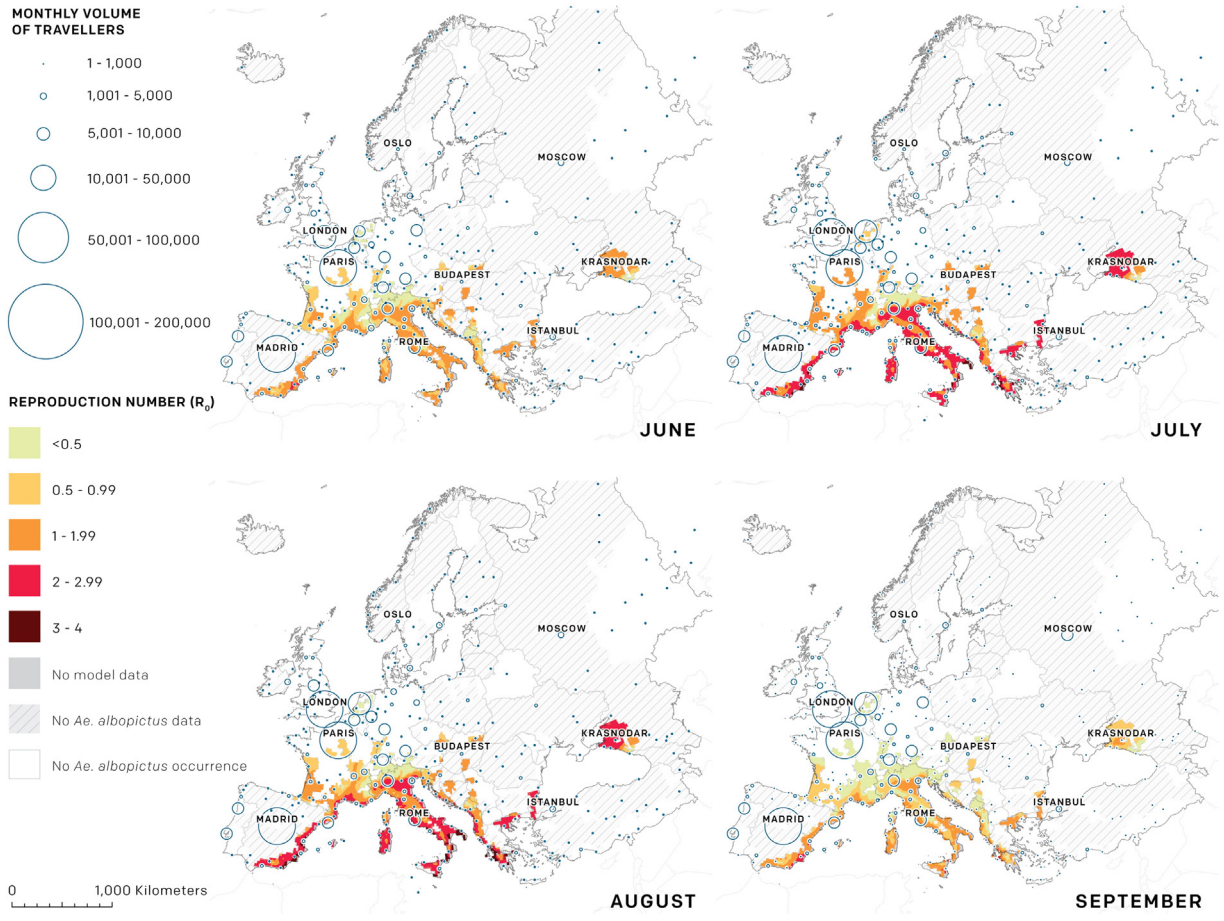


Fig. 1. Monthly stratified maps (June–Sept) of the potential basic reproduction number (R_0) of Zika virus in Europe via *Aedes albopictus* overlaid with monthly estimates of airline travellers arriving from areas with potential for year-round Zika transmission in the Americas.

Table 1
Observed and predicted R_0 for Zika for *Ae. aegypti* (Ae) and *Ae. albopictus* (Aa).

Location	Period	Number of cases	Mean Observed R_0	Mean R_0 predicted		
				Model 1: *Ae*, *Aa*	*Model 2*: *Ae*, *Aa*	*Model 3*: *Aa*, *Ae*
Guadeloupe, French overseas region	29 Jan. to 10 March, 2016	717[b]	2.1	3.7, 3.2	1.0, 0.80	2.0, 1.6
Martinique, French overseas region	21 Dec. to 21 Jan., 2015–2016	102[a]	7.2	4.4, 3.9	1.3, 0.96	2.6, 1.9
Dominican Republic, whole country	30 Jan. to 27 Feb., 2016	573[b]	1.8	3.3, 2.9	1.1, 0.75	2.2, 1.5
Mexico, 1 administrative region	1 Jan. to 13 March, 2016	129[a]	2.1	3.4, 3.0	0.94, 0.79	1.9, 1.6
Guatemala, 8 administrative regions	16 Nov. to 15 Feb., 2015–2016	386[c]	5.0	4.1, 3.6	1.2, 0.96	2.4, 1.93
Panama, 1 administrative region	1 Jan. to 13 March, 2016	49[a]	1.9	2.6, 2.2	0.73, 0.61	1.5, 1.2
Puerto Rico, whole country	26 Nov. to 24 Feb., 2015–2016	198[a]	3.7	4.2, 3.7	1.2, 0.96	2.4, 1.93
El Salvador, 11 administrative regions	1 Jan. to 13 March, 2016	5618[b]	4.8	5.5, 4.8	1.8, 1.2	3.6, 2.4
Colombia, 15 administrative regions	1 Jan. to 13 March, 2016	1593[a]	3.2	3.0, 2.7	1.1, 0.68	2.2, 1.4
Ecuador, 2 administrative regions	1 Jan. to 13 March, 2016	96[b]	7.0	4.4, 3.8	1.2, 0.96	2.4, 1.93
		Total	Mean			
		9461	3.9	3.9, 3.4	1.2, 0.87	2.4, 1.74

[a] Confirmed cases
[b] Suspected cases
[c] Confirmed plus suspected cases

temperature and diurnal temperature variation (Lambrechts et al., 2011; Brady et al., 2014; Liu-Helmersson et al., 2014; Liu-Helmersson et al., 2016). It is given mathematically by(Liu-Helmersson et al., 2016):

$$VC = \frac{ma^2 b_h b_m e^{-\mu_m n}}{\mu_m}$$

The six vector parameters in the above equation include: 1) the average vector biting rate a, 2) the probability of vector-to-human transmission per bite b_h, 3) the probability of human-to-vector infection per bite, b_m, 4) the duration of the extrinsic incubation period (i.e. this represents the duration between acquisition of a pathogen by a vector and the ability for that vector to then transmit the same pathogen to a susceptible host), n, 5) vector mortality rate μ_m, and 6) the female vector-to-human population ratio, m.

Dengue virus (DENV) for example, is also transmitted by *Ae. aegypti* and *Ae. albopictus* mosquitoes, for which the aforementioned parameters and their relationship to temperature has been described (Lambrechts et al., 2011; Brady et al., 2014; Liu-Helmersson et al., 2014; Liu-Helmersson et al., 2016). This has relevance for ZIKV transmission by *Ae. albopictus* and *Ae. aegypti*, as the mortality rate, μ_m, and biting rates, a, for *Aedes* mosquitoes are independent of whether they carry ZIKV or DENV. Further, the competence of *Aedes* mosquito species to transmit ZIKV, as described by n, b_h and b_m, is driven by the infection, dissemination, and transmission rates of the virus in an infected vector, measuring the rate of virus presence in the mosquito midgut after a blood meal, and the time and efficiency of its replication to the mosquito's salivary glands. Recent data suggests that *Aedes* mosquitoes may have the same vectorial competence for ZIKV dissemination and transmission rate as DENV at 29 °C (Chouin-Carneiro et al., 2016; Li et al., 2012; Wong et al., 2013; Lambrechts et al., 2011; Brady et al., 2014; Liu-Helmersson et al., 2014; Liu-Helmersson et al., 2016). However, geographical differences in vector competence may exist, as *Aedes* mosquitoes from Brazil, the U.S. and Martinique have been reported as being less potent in their ability to replicate ZIKV, while *Aedes* mosquitoes from French Guiana and Guadeloupe appear competent and efficient (Chouin-Carneiro et al., 2016). The extrinsic incubation period of ZIKV (n) is not yet determined, but recent studies estimate it to be 7 days (Chouin-Carneiro et al., 2016; Li et al., 2012; Wong et al., 2013), while n for DENV is estimated to be 8–9 days at 28–29 °C (Liu-Helmersson et al., 2016).

The duration of viremia for ZIKV (T_h) is also currently under investigation however recent evidence suggests that the virus may be detectable in blood for 1 to 10 days (Musso et al., 2015a; Lessler et al., 2016). Similarly, DENV is detected in the blood for 3 to 12 days, with most individuals being viremic for about 5 days (Liu-Helmersson et al., 2016). The serial time window between pairs of infected human cases, encapsulating part of the period of viremia and the extrinsic and intrinsic incubation periods, has been estimated to be 15 days for ZIKV and 17 days for DENV(Majumder et al., 2016) indicating similarities between ZIKV and DENV in real-world settings. For both *Ae. aegypti* and *Ae. albopictus* mosquitoes, the female vector-to-human population ratio (m), is temperature dependent and presumed to be similar across species. In the absence of valid vector abundance data, as an indicator of vector populations, we used a method described in previous studies relating population density to survival (Lambrechts et al., 2011; Brady et al., 2014; Liu-Helmersson et al., 2014; Liu-Helmersson et al., 2016).

We developed three vectorial capacity models for *Aedes* mosquitoes to transmit ZIKV and selected the model that best fit observed transmission dynamics in the Americas. Each model was developed by adapting existing temperature-driven models for DENV transmission via *Ae. aegypti* and *Ae. albopictus* (Liu-Helmersson et al., 2016) while incorporating current data on ZIKV parameters. Our best fitting model was then adapted to Europe using European climatic data to estimate R_0

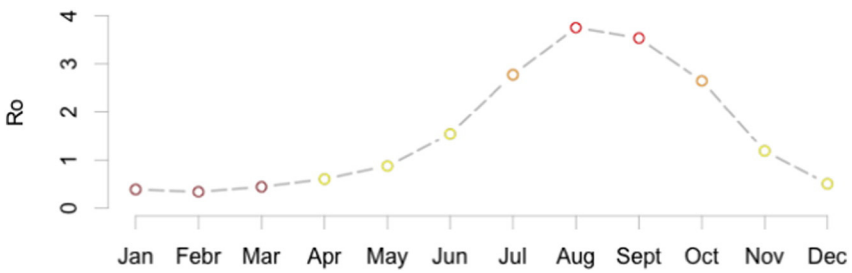


Fig. 2. Potential basic reproduction numbers (R_0) for Zika virus in Madeira, Portugal via *Aedes aegypti*.

potential this spring to autumn. Below we synthesize the findings described above into three alternative models describing the vectorial capacity and basic reproduction number of ZIKV:

- *Model 1*: characterized by high vector competence compared with dengue (Liu-Helmersson et al., 2016) in assuming a viremic period of 5 days.
- *Model 2*: similar to *model 1* with alternations in b_m being reduced to 76.7% and b_h reduced to 21.4% of the *model 1* values for *Ae. aegypti* (Chouin-Carneiro et al., 2016). For *Ae. albopictus*, *model 2* is similar to *model 1* however b_m are reduced to 50.0% of the *model 1* value (Chouin-Carneiro et al., 2016).
- *Model 3*: same configuration as *model 2* but with a viremic period of 10 days (Lessler et al., 2016).

The model 1–3 parameters and their relationship to temperature for both vectors are described in the supplementary information, Figs. 7–12.

2.3. Model Validation

We collected data on the number, location and time of confirmed and suspected ZIKV infections across Latin America and the Caribbean regions up until the end of March 2016 using data from national surveillance systems (see Table 1). Disease data were reported for individual countries or sub-national administrative regions. We also collected temperature measurements from terrestrial meteorological stations and satellite sensors reading land surface temperatures (see Table 3; extended data).

We estimated R_0 values for the initial phases of the epidemic in the Americas assuming a Poisson distribution using weekly count data (*R package version 1.2–5, 2014; Wallinga, 2007*). The serial time interval range of ZIKV infection was taken to be 10 to 23 days, with a mean of 16 days (Majumder et al., 2016). We then used a lognormal serial time distribution function with a mean of 16 days and standard deviation of 3 days covering the interval. Estimates of observed R_0 values were made at the sub-national administrative region level and aggregated to national averages as described in Table 1. We selected smaller countries and regions to better fulfill the assumption of homogeneous mixing in the estimation of R_0.

We tested and validated our three temperature-driven vectorial capacity models by comparing R_0 values generated for countries in Latin America and the Caribbean to R_0 estimates derived from the same region using ZIKV surveillance data (Table 1). We used temperature data from one month prior to and two months into the epidemic (Table 3; extended data). R_0 values were computed assuming that *Ae. aegypti* or *Ae. albopictus* alone were the sole vector for ZIKV transmission.

2.4. Estimating R_0 in Europe

Since *Model 1* best fit observed data on ZIKV transmission in the Americas, we used it to estimate potential for mosquito-borne transmission, outputted as R_0 values, across Europe in 2016. We did this by integrating daily temperature observations (mean, minimum, maximum) from the E-OBS 12.0 dataset at locations gridded at 0.25 × 0.25° (approximately 25 × 25 km at the equator) latitude and longitude (Haylock et al., 2008) for the period between January 1st 2006 to December 31st 2015. Daily R_0 estimates required interpolating diurnal temperature values from daily temperature observations. We estimated R_0 values for Europe for geographic areas where *Ae. aegypti or Ae. albopictus* mosquito populations have been reported by the European Centre for Disease Prevention and Control in 2016 (http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx, n.d.). Maps were presented for each month from May to October after averaging daily R_0 observations. Model outputs were classified into R_0 zones for ease of interpretation (<0.5, 0.5–0.99, 1–1.99, 2–2.99, and 3–4).

2.5. Airline Travellers into Europe from the Americas

To estimate the potential for travellers infected with ZIKV arriving into European cities from the Americas, we analyzed worldwide airline ticket sales data from the International Air Transport Association (IATA). This dataset includes anonymized, full-route, flight itinerary data on an estimated 90% of all passenger trips on commercial flights worldwide, while the remainder is modeled using airline market intelligence. Each flight itinerary includes information on passengers' initial point of embarkation, final destination, and where applicable, connecting flights.

To select commercial airports in the Americas where departing travellers might be infected with ZIKV, we assumed that ZIKV would be active in all countries in the Americas except Canada and Chile by the summer of 2016. Subsequently, we narrowed this list of countries to those that have experienced *significant* transmission of chikungunya virus since the onset of the 2013 outbreak in the Americas (Musso et al., 2015b). In this step, we excluded the United States as it has only experienced limited autochthonous transmission of chikungunya virus in Florida. Conversely, we included Argentina because it is currently experiencing a DENV epidemic (Gil et al., 2016). Within each country in our remaining list, we identified subnational areas with potential for year-round ZIKV transmission by either *Ae. aegypti* or *Ae. albopictus* based upon an ecological suitability analysis (Bogoch et al., 2016). We then created a 50 km buffer zone around these areas to accommodate for the potential movements of individuals travelling by land from areas of ZIKV activity to neighboring commercial airports, where travellers could embark upon international trips. After selecting our final list of commercial airports in the Americas, we quantified and mapped the monthly final European destinations of all travellers departing these airports between May and October 2015.

2.6. Populations at Risk for Locally Acquired Infection

To estimate the size of populations living in areas at risk for mosquito-borne ZIKV transmission, we extracted population data from LandScan 2014, a satellite-based dataset of ambient population density worldwide in 1-km^2 grids (Dobson et al., 2000). To link our vectorial capacity model data with LandScan population density data, our R_0 map (at 25-km^2 resolution) was resampled to 1-km^2 pixels using a nearest-neighbor sampling algorithm. Population estimates were then extracted for each R_0 zone for each country across Europe in the month of August, when vectorial capacity was at its peak.

3. Results

3.1. Travellers Arriving to Europe from the Americas

In continental Europe, the leading final destinations of airline travellers departing areas in the Americas with known ZIKV activity or suitable conditions for year-round autochthonous transmission, from May to October 2015 were i) Paris (~120,000 to 200,000 travellers peaking in July and August), ii) London (~100,000 to 130,000 travellers peaking in August), iii) Madrid (~75,000 to 125,000 travellers peaking in July), iv) Amsterdam (~50,000 to 70,000 travellers peaking in August), v) Frankfurt (~40,000 travellers with no clear peak), vi) Milan (~25,000 to 40,000 travellers peaking in August), vii) Lisbon (~20,000 to 40,000 travellers peaking in July), viii) Barcelona (~25,000 to 35,000 travellers peaking in September and October), and ix) Rome (~20,000 to 35,000 travellers peaking in August; see Fig. 1). Outside of continental Europe, Madeira, Portugal (where *Ae. aegypti* is established) receives ~500 to 2500 monthly travellers with peak flows in July.

3.2. Basic Reproductive Number for Zika virus in Europe

Based on temperature inputs, our vectorial capacity model predicted a basic reproduction number (R_0) of 3.9 (assuming transmission by *Ae. aegypti*), and 3.4 (assuming transmission by *Ae. albopictus*) for the current ZIKV epidemic in the Americas. We compared these R_0 estimates with those derived using ZIKV surveillance data in the Americas and found agreement between our predictions and observations as summarized in Table 1. Adapting our vectorial capacity model to Europe using empirical temperature data from 2006–2015, we estimated R_0 values for *Ae. albopictus* across southern Europe from France and Spain in the west to southwestern Russia in the east (see Fig. 1; higher resolution monthly maps are supplied in the supplementary information, Figs. 1–6). Most of the potential for autochthonous transmission was centered on Italy, southeastern France including the island of Corsica, the southern and eastern coasts of Spain, and western regions of the Balkans from Slovenia and Croatia in the north, to Albania and Greece in the south. While predicted R_0 values begin to exceed 1 (i.e. indicating epidemic potential) across areas of southern Europe in June, they increase in July (R_0 values of 2–3) and peak in August (R_0 values of 3–4) before falling again in September (R_0 values of 1–2) and October (R_0 falls below 1). For the sub-tropical island of Madeira in Portugal, R_0 values for the vector *Ae. aegypti*, were estimated to exceed 2 from July through October (see Fig. 2).

3.3. Populations at Risk in Europe

Of the total population residing within the geographic range of our analysis (i.e. 779 million people), we find that in the month of August (i.e. when temperatures and vectorial capacity in Europe are peaking), approximately 47% of people (i.e. 366 million) reside in areas with no known occurrences of *Ae. albopictus*, 35% (i.e. 272 million) in areas where data on *Aedes* mosquito occurrences are absent, 11% (i.e. 83 million) in areas where R_0 estimates from our model exceed 1, 4% (i.e. 31 million) in areas where R_0 estimates are less than 1 (i.e. low risk for sporadic transmission), and 3% (i.e. 27 million) in areas where our model lacked data to estimate R_0.

Countries with a large proportion of their population residing in areas where our R_0 estimates exceeded 1 in August included: Albania (83%), Croatia (44%), France (20%), Greece (25%), Italy (78%), Montenegro (39%), Slovenia (28%), and Spain (19%). Population sizes in these areas were largest for Italy (45 million people), France (12 million), and Spain (8 million). For Greece and Spain, more than half of their populations resided in areas with either no *Aedes* mosquito surveillance data or no data outputs from our model. The geographic extents of *Ae. albopictus* mosquitoes overlaid with our model's R_0 outputs are shown in Fig. 1, while actual values of populations affected are found in Table 2 (extended data).

4. Discussion

Our analysis indicates that the peak flow of travellers departing areas in the Americas where they might be exposed to ZIKV (and who have final destinations in Europe) coincides with the peak in vectorial capacity for ZIKV transmission in Europe (i.e. in July and August). These intersections in risk are most apparent within or adjacent to several major cities such as Barcelona, Milan and Rome.

Since *Ae. albopictus* mosquitoes might prove to be competent vectors for ZIKV (Chouin-Carneiro et al., 2016; Grard et al., 2014; Li et al., 2012; Wong et al., 2013), the public, healthcare providers and public health officials across Europe could use these findings to identify regions at greatest risk for the importation of ZIKV, and its potential transmission within ecologically suitable areas. Although the volume of travellers arriving from the Americas to Madeira, Portugal is substantially lower compared to other major cities in continental Europe, the known occurrence of *Ae. aegypti*, a longer season with high vectorial capacity, the explosive epidemic of dengue fever in 2012 (Wilder-Smith et al., 2014), and the recent Zika epidemic in nearby Cape Verde (Attar, 2016), collectively highlight the potential for autochthonous transmission of ZIKV on this sub-tropical island.

Our analysis highlights necessary, but not always sufficient conditions for autochthonous transmission of ZIKV. While the introduction of ZIKV into Europe, the presence of competent mosquito vectors, and suitable climatic conditions are all prerequisites for local mosquito-borne transmission, a multitude of other factors, including but not limited to, population density, housing conditions, and socioeconomic factors, could influence the likelihood of observing ZIKV epidemics, as seen with other arbovirus infections such as dengue (Clark, 2008; Reiter et al., 2003).

Our model is founded on a number of assumptions, most notably that continental European strains of *Ae. albopictus* possess competence for the transmission of ZIKV. While there is growing evidence to suggest that *Ae. albopictus* can become infected with ZIKV under experimental conditions (Chouin-Carneiro et al., 2016; Li et al., 2012), empirical data on its role as a vector in nature exist (Grard et al., 2014), but are limited. Recent evidence from the Americas also suggests that *Ae. aegypti* and *Ae. albopictus* may be less competent vectors than anticipated (Chouin-Carneiro et al., 2016), and that other factors such as population density, immunologically naïve populations, additional modes of transmission (e.g. sexual (Oster et al., 2016; Foy et al., 2011)), and possibly even other mosquito species might play a role in this epidemic (Ayres, 2016). A key strength of our study is that our model's R_0 outputs were comparable to estimates we derived using ZIKV surveillance data from the ongoing epidemic in the Americas. However, we assumed that our model, validated against ZIKV data from the Americas where *Ae. aegypti* is thought to be the primary driver for transmission, would be transferrable to a European setting where *Ae. albopictus* is the dominant vector.

Cities across Europe are already experiencing an increase in ZIKV importations via travellers from the Americas (Maria et al., 2016; Zammarchi et al., 2015). Two recent studies assessed the risk of transmission of ZIKV in Europe reaching contrasting conclusions (Guzzetta et al., 2016; WHO, 2016). One of the studies finds minimal risk based on an assumed low vectorial capacity of European *Ae. albopictus* (Guzzetta et al., 2016), while the other study, aligning climatic suitability of vectors and coarse patterns in air traffic, estimate and identify European countries at high risk for Zika transmission (WHO, 2016). Our study increases output resolution through our transmission model and passenger-level air travel data, while drawing from epidemiological data from the current outbreak in the Americas. We conclude, that while there remains uncertainty about the capacity for *Ae. albopictus* to transmit ZIKV in nature, European health officials must consider the possibility of mosquito-borne transmission within Europe during warmer periods of the year. With the imminent arrival of summer, our findings could help guide the efficient use of finite resources for human disease surveillance, vector surveillance and control, and public education following guidelines from the European Centre for Disease Prevention and Control (European Centre for Disease Prevention and Control, 2016), for vector and non-vector borne modes of ZIKV transmission. Furthermore, our travel analysis can help target efforts to educate the public and healthcare providers alike about the potential for Zika virus transmission via travellers returning from Zika affected areas in the Americas to their sexual partners living in Europe (Oster et al., 2016; Petersen et al., 2016b).

4.1. Assumptions and Limitations

The overall range and average values of the basic reproduction number for ZIKV estimated in our study for Latin America and the Caribbean is supported by several recent studies indicating similarities to dengue (Funk et al., 2016; Kucharski et al., 2016) at least for transmission via *Ae. aegypti*. Data on the competence of *Ae. albopictus* as a vector to

transmit ZIKV in natural settings are limited. While recent experimental studies have indicated that North American *Ae. albopictus* strains may not be highly competent (Chouin-Carneiro et al., 2016), others support a higher vector competence (Li et al., 2012; Wong et al., 2013), including one study where *Ae. albopictus* appears to have been the principal vector in an urban ZIKV outbreak in Gabon (Grard et al., 2014). In this analysis, we assumed that *Ae. albopictus* would be capable of transmitting ZIKV in continental Europe. New experimental results indicate European *Ae. albopictus* mosquitoes may exhibit lower competence to transmit ZIKV (Di Luca et al., 2016; Jupille et al., 2016) contrasting to African (Grard et al., 2014) and Asian (Wong et al., 2013) varieties. However, more evidence is needed to fully understand vector competence in Europe.

We also assumed that DENV and ZIKV have similar temperature dependencies in relation to vector competence. While the similarities between our model's R_0 outputs for ZIKV and R_0 estimates derived from the current ZIKV epidemic in the Americas supports this assumption, we recognize that other factors, related to human behavior, socioeconomic, and environmental factors may create differential vector-to-human contacts rates between Europe and Latin America. These differences could contribute to different transmission dynamics not accounted for in our predictive models. For example, during a recent DENV outbreak along the Texas-Mexico border, non-climatic factors significantly influenced transmission dynamics (Clark, 2008; Reiter et al., 2003). The observed transmission of chikungunya and dengue in Europe, however, support that European vectors and environment are capable of supporting local transmission (Angelini et al., 2007; Wilder-Smith et al., 2014). We further note that vector populations, as measured by ovitraps, indicate high abundance in Europe (Bellini et al., 1996), and further, that human population densities are higher in some European areas compared with areas in Latin America with persistent transmission (*http://www.worldpop.org.uk/*). In relation to European vector populations it is important to note that the vector surveillance is limited in some European regions, with no reported surveillance data. Thus, our estimates of the geographic range of European areas potential suitable for ZIKV transmission are conservative (Kraemer et al., 2015).

Due to a lack of detailed parameterization data, we did not consider the effects of wind, humidity and precipitation in our model. The impact of precipitation, interacting with human behavior (e.g. storing of water and irrigation), appear unrelated to fluctuations in vector abundance in Thailand (Lambrechts et al., 2011). Tran *et al.* modeled *Ae. albopictus* abundance in Europe and demonstrated that cycles were highly seasonal and related to vector lifespan (Tran et al., 2013). While we were unable to include vector abundance data in our models, we created a proxy for abundance derived from a temperature-dependent mosquito survival function (Brady et al., 2014; Liu-Helmersson et al., 2016). We also note that populations of *Aedes* mosquitoes in the Netherlands have not established outside greenhouses, suggesting that they are tropical strains not adapted to temperate climates.

Although DENV and chikungunya viruses may offer certain insights into the unknowns of ZIKV spread and control (Musso et al., 2015b), differences between ZIKV, DENV and chikungunya vector-virus interactions and transmission routes cannot be ignored (Christofferson, 2016), In addition to mosquito-borne transmission, other modes of spread including via sexual contact (Oster et al., 2016) were not accounted for in this analysis. Current data suggest that ZIKV can be found in semen, and hence could potentially be transmitted for months post-infection (Mansuy et al., 2016).

The 2015 IATA data used in this analysis is the most current available to quantify the flow of travellers from the Americas to Europe across this spring to autumn. Although the ZIKV epidemic may decrease traveller flows to and from affected countries, we assumed that the seasonal patterns and proportions of travellers arriving to various European cities would be preserved relative to 2015. Another consideration not accounted for in our analysis is the potential impact of the Summer Olympic Games in Rio in August on travel patterns between Brazil and Europe. However, the local transmission of Zika during the Olympics might decrease due to seasonal transmission changes (Funk et al., 2016; Kucharski et al., 2016).

Our weather driven models use averaged monthly temperature values over the most recent 10-year period (i.e. 2006–2015). However, 2016 weather patterns may vary under the influence of El Nino this year, which in turn, might increase or decrease vectorial capacity.

Finally, stochasticity plays a key role in the appearance of any epidemic. In this analysis, we present conditions that are necessary, but alone may be insufficient for the emergence of an epidemic. Hence our analysis assesses the potential for ZIKV transmission within Europe, not the probability of ZIKV epidemics. The purpose of this analysis is to highlight geographies and times where epidemic potential is greatest as a way to help guide public health resources to prepare for, and if required, mobilize an appropriate response to mosquito-borne ZIKV transmission in Europe later this year, in accordance with guidelines developed by the European Centre for Disease Prevention and Control (European Centre for Disease Prevention and Control, 2016).

Disclosure Statements

Kamran Khan is the founder of BlueDot, a social benefit corporation that develops public health tools and analytics to help prepare for and respond to global epidemics. Matthew German has been employed by BlueDot and Isaac I. Bogoch has consulted to BlueDot.

Author Contributions

Joacim Rocklöv and Kamran Khan conceived the research idea, made the first analyses and drafted the text. Mikkel Quam helped develop the Zika mathematical models. Bertrand Sudre helped gather the Zika surveillance data and update the manuscript based on literature reviews on Zika epidemiology and entomology. Matthew German conducted the analyses of commercial air travel flows. Moritz Kraemer and Oliver Brady helped to develop the method for the Zika transmission zones in Latin America. Mark Ong produced estimates of populations living in areas at potential risk for autochthonous transmission of Zika virus across Europe. All authors made significant contributions to the editing and critical assessment of the manuscript.

The following are the supplementary data related to this article.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2016.06.009.

Acknowledgements

This study is part of the DengueTools project funded by the European Union Seventh Framework Program FP7/2007–2013 under grant agreement no. 282589. We acknowledge the support from the epidemic intelligence department at the European Centre for Disease Prevention and Control (ECDC).

References

Mosquito maps (ECDC-EFSA, 2016 – VECTORNET) available at: http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx <http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx> (

Anderson, R.M., May, R., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press.

Angelini, R., et al., 2007. Chikungunya in north-eastern Italy: a summing up of the outbreak. Euro Surveill. 12 (E071122 071122).

Attar, N., 2016. ZIKA virus circulates in new regions. Nat. Rev. Microbiol. 14, 62. http://dx.doi.org/10.1038/nrmicro.2015.28.

Ayres, C.F., 2016. Identification of Zika virus vectors and implications for control. Lancet Infect. Dis. 16, 278–279. http://dx.doi.org/10.1016/S1473-3099(16)00073-6.

Bellini, R., Carrieri, M., Burgio, G., Bacchi, M., 1996. Efficacy of different ovitraps and binomial sampling in *Aedes albopictus* surveillance activity. J. Am. Mosq. Control Assoc. 12, 632–636.

Bogoch, I.I., et al., 2016. Anticipating the international spread of Zika virus from Brazil. Lancet 387, 335–336. http://dx.doi.org/10.1016/S0140-6736(16)00080-5.

Brady, O.J., et al., 2014. Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. Parasit Vectors 7, 338. http://dx.doi.org/10.1186/1756–3305–7-338.

Calvet, G., et al., 2016. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet Infect. Dis. http://dx.doi.org/10.1016/S1473-3099(16)00095-5.

Cao-Lormeau, V.M., et al., 2016. Guillain-Barre Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet http://dx.doi.org/10.1016/S0140-6736(16)00562-6.

Chouin-Carneiro, T., V.-R., A., Vazeille, M., Yebakima, A., Girod, R., Goindin, R., Dupont-Rouzeyrol, M., Lourenço-de-Oliveira, R., Failloux, A.B., 2016. Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika virus. PLoS Negl. Trop. Dis. 10, e0004543. http://dx.doi.org/10.1371/journal.pntd.0004543.

Christofferson, R.C., 2016. Zika virus emergence and expansion: lessons learned from dengue and chikungunya may not provide all the answers. Am.J.Trop. Med. Hyg. http://dx.doi.org/10.4269/ajtmh.15-0866.

Clark, G.G., 2008. Dengue and dengue hemorrhagic fever in northern Mexico and south Texas: do they really respect the border? Am.J.Trop. Med. Hyg. 78, 361–362.

Delisle, E., et al., 2015. Chikungunya outbreak in Montpellier, France, September to October 2014. Euro Surveill. 20.

Di Luca, M., et al., 2016. Experimental studies of susceptibility of Italian *Aedes albopictus* to Zika virus. Euro Surveill. 21. http://dx.doi.org/10.2807/1560-7917.ES.2016.21.18.30223.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote. Sens. 66, 9.

ECDC. *Aedes albopictus*, <http://ecdc.europa.eu/en/healthtopics/vectors/mosquitoes/Pages/aedes-albopictus.aspx> (

European Centre for Disease Prevention and Control, 2016. Zika virus disease epidemic: preparedness planning guide for diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. ECDC, Stockholm (2016).

Faria, N.R., et al., 2016. Zika virus in the Americas: early epidemiological and genetic findings. Science http://dx.doi.org/10.1126/science.aaf5036.

Foy, B.D., et al., 2011. Probable non-vector-borne transmission of Zika virus, Colorado, USA. Emerg. Infect. Dis. 17, 880–882. http://dx.doi.org/10.3201/eid1705.101939.

Funk, S., K. A. J., Camacho, A., Eggo, R.M., Yakob, L., Edmunds, W.J., 2016. Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. bioRxiv http://dx.doi.org/10.1101/043265.

Gil, J.F., et al., 2016. Spatial spread of dengue in a non-endemic tropical city in northern Argentina. Acta Trop. 158, 24–31. http://dx.doi.org/10.1016/j.actatropica.2016.02.003.

Grard, G., Caron, M., Mombo, I.M., Nkoghe, D., Ondo, S.M., Jiolle, D., Fonenille, D., Paupy, C., Leroy, E.M., 2014. Zika virus in Gabon (central Africa) — 2007: a new threat from *Aedes albopictus*? PLoS Negl. Trop. Dis. 8.

Gulland, A., 2016. WHO warns European countries to be on alert for Zika. BMJ 352, i753. http://dx.doi.org/10.1136/bmj.i753.

Guzzetta, G., et al., 2016. Assessing the potential risk of Zika virus epidemics in temperate areas with established *Aedes albopictus* populations. Euro Surveill. 21. http://dx.doi.org/10.2807/1560–7917.ES.2016.21.15.30199.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res. Atmos. 113. http://dx.doi.org/10.1029/2008JD10201.

Jupille, H., Seixas, G., Mousson, L., Sousa, C., Failloux, A.B., 2016. Zika virus, a new threat for Europe? bioRxiv http://dx.doi.org/10.1101/048454.

Kraemer, M.U., et al., 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. Elife 4, e08347. http://dx.doi.org/10.7554/eLife.08347.

Kucharski, A.J., et al., 2016. Transmission dynamics of zika virus in island populations: a modelling analysis of the 2013–14 French Polynesia outbreak. PLoS Negl. Trop. Dis. 10, e0004726. http://dx.doi.org/10.1371/journal.pntd.0004726.

Lambrechts, L., et al., 2011. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. Proc. Natl. Acad. Sci. U. S. A. 108, 7460–7465. http://dx.doi.org/10.1073/pnas.1101377108.

Lessler, J.T., O.C.T., Carcelen, A.C., Konikoff, J.M., Williamson, J., Bi, Q., Kurcika, L.M., Cummings, D.A.T., Reich, N., Chaisson, L.H., 2016. Times to key events in the course of Zika infection and their implications: a systematic review and pooled analysis. Bull. World Health Organ. http://dx.doi.org/10.2471/BLT.16.174540.

Li, M.I., Wong, P.S., Ng, L.C., Tan, C.H., 2012. Oral susceptibility of Singapore *Aedes* (Stegomyia) *aegypti* (Linnaeus) to Zika virus. PLoS Negl. Trop. Dis. 6, e1792. http://dx.doi.org/10.1371/journal.pntd.0001792.

Liu-Helmersson, J., Quam, M., Wilder-Smith, A., Stenlund, H., Ebi, K., Massad, E., Rocklöv, J., 2016. Climate change and aedes vectors: 21st century projections for dengue transmission in Europe. eBioMedicine.

Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A., Rocklov, J., 2014. Vectorial capacity of *Aedes aegypti*: effects of temperature and implications for global dengue epidemic potential. PLoS One 9, e89783. http://dx.doi.org/10.1371/journal.pone.0089783.

Majumder, M.S., Cohn, E., Fish, D., Brownstein, J.S., 2016. Estimating a feasible serial interval range for Zika fever [submitted]. Bull. World Health Organ. http://dx.doi.org/10.2471/BLT.16.171009.

Mansuy, J.M., et al., 2016. Zika virus: high infectious viral load in semen, a new sexually transmitted pathogen? Lancet Infect. Dis. 16, 405. http://dx.doi.org/10.1016/S1473-3099(16)00138-9.

Maria, A.T., et al., 2016. Zika virus infections in three travellers returning from South America and the Caribbean respectively, to Montpellier, France, December 2015 to January 2016. Euro Surveill. 21. http://dx.doi.org/10.2807/1560-7917.ES.2016.21.6.30131.

Mlakar, J., et al., 2016. Zika virus associated with microcephaly. N. Engl. J. Med. 374, 951–958. http://dx.doi.org/10.1056/NEJMoa1600651.

Musso, D., Cao-Lormeau, V.M., Gubler, D.J., 2015b. Zika virus: following the path of dengue and chikungunya? Lancet 386, 243–244. http://dx.doi.org/10.1016/S0140-6736(15)61273-9.

Musso, D., et al., 2015a. Detection of Zika virus in saliva. J. Clin. Virol. 68, 53–55. http://dx.doi.org/10.1016/j.jcv.2015.04.021.

Oster, A.M., et al., 2016. Update: interim guidance for prevention of sexual transmission of Zika virus — United States, 2016. MMWR Morb. Mortal. Wkly Rep. 65, 323–325. http://dx.doi.org/10.15585/mmwr.mm6512e3.

Petersen, E.E., et al., 2016b. Update: interim guidance for health care providers caring for women of reproductive age with possible zika virus exposure — United States, 2016. MMWR Morb. Mortal. Wkly Rep. 65, 315–322. http://dx.doi.org/10.15585/mmwr.mm6512e2.

Petersen, L.R., Jamieson, D.J., Powers, A.M., Honein, M.A., 2016a. Zika virus. N. Engl. J. Med. http://dx.doi.org/10.1056/NEJMra1602113.

Promed, 2015. ZIKA virus - BRAZIL: confirmed. http://www.promedmail.org/post/3370768.

R_0: estimation of R_0 and real-time reproduction number from epidemics. R Package Version 1.2–5 (R).

Reiter, P., et al., 2003. Texas lifestyle limits transmission of dengue virus. Emerg. Infect. Dis. 9, 86–89. http://dx.doi.org/10.3201/eid0901.020220.

Rodrigues, L.C., 2016. Microcephaly and zika virus infection. Lancet http://dx.doi.org/10.1016/S0140-6736(16)00742-X.

Tran, A., et al., 2013. A rainfall- and temperature-driven abundance model for *Aedes albopictus* populations. Int. J. Environ. Res. Public Health 10, 1698–1719. http://dx.doi.org/10.3390/ijerph10051698.

Wallinga, J.L.M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B Biol. Sci. 274.

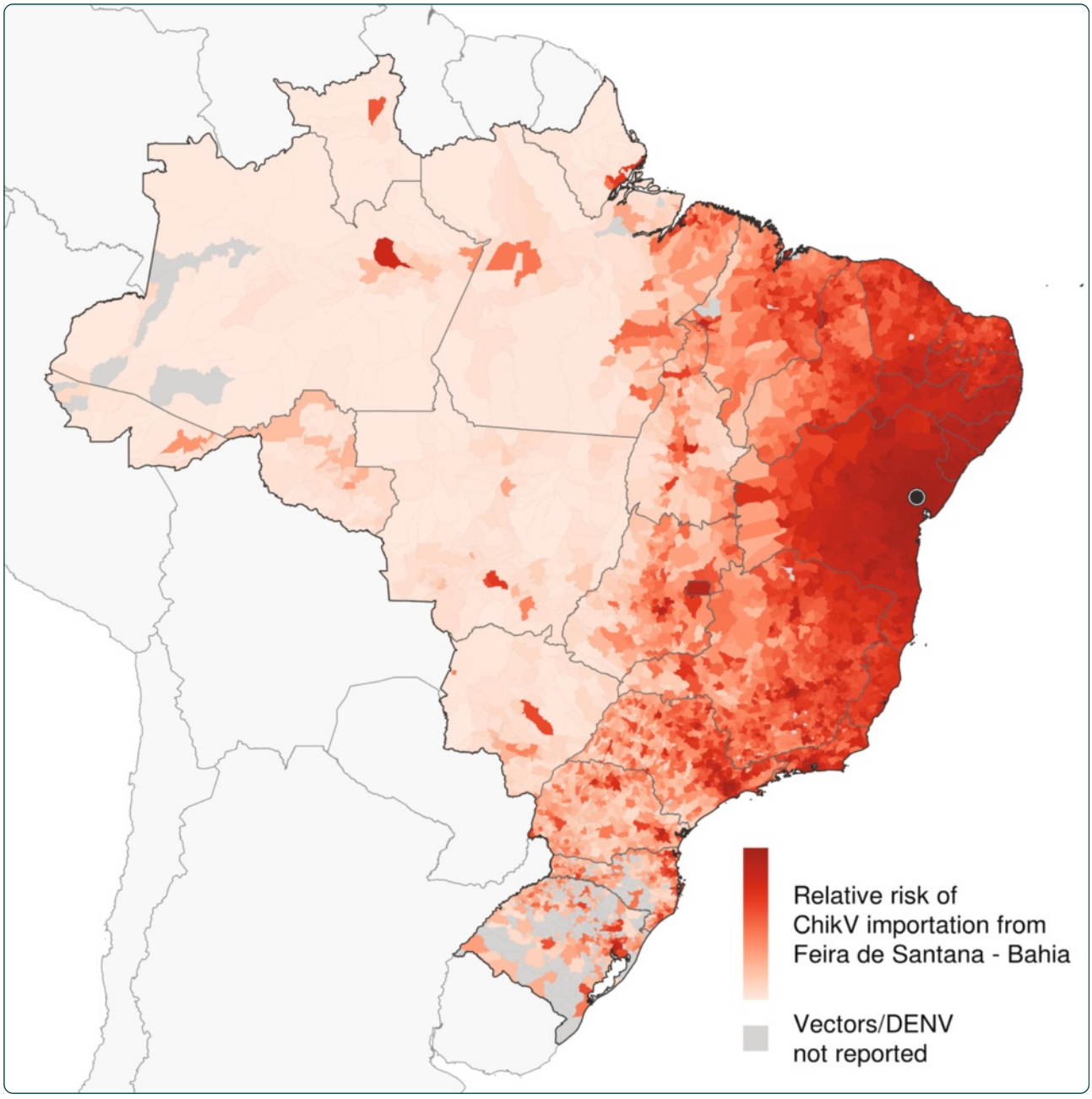
WHO, 2016. Zika virus technical report. Interim risk assessment for WHO European region. http://www.euro.who.int/en/health-topics/emergencies/zika-virus/technical-reports-and-guidelines-on-zika-virus/zika-virus-technical-report.-interim-risk-assessment-for-who-european-region.

Wilder-Smith, A., et al., 2014. The 2012 dengue outbreak in Madeira: exploring the origins. Euro Surveill. 19, 20718.

Wong, P.S., Li, M.Z., Chong, C.S., Ng, L.C., Tan, C.H., 2013. Aedes (Stegomyia) albopictus (Skuse): a potential vector of Zika virus in Singapore. PLoS Negl. Trop. Dis. 7, e2348. http://dx.doi.org/10.1371/journal.pntd.0002348.

Zammarchi, L., et al., 2015. Zika virus infections imported to Italy: clinical, immunological and virological findings, and public health implications. J. Clin. Virol. 63, 32–35. http://dx.doi.org/10.1016/j.jcv.2014.12.005.

BMC Medicine



Emergence and potential for spread of Chikungunya virus in Brazil

Nunes *et al.*

BioMed Central

Nunes *et al. BMC Medicine* (2015) 13:102
DOI 10.1186/s12916-015-0348-x

Nunes et al. BMC Medicine (2015) 13:102
DOI 10.1186/s12916-015-0348-x

BMC Medicine

RESEARCH ARTICLE Open Access

Emergence and potential for spread of Chikungunya virus in Brazil

Marcio Roberto Teixeira Nunes[1*†], Nuno Rodrigues Faria[2†], Janaina Mota de Vasconcelos[1†], Nick Golding[3], Moritz UG Kraemer[3], Layanna Freitas de Oliveira[1], Raimunda do Socorro da Silva Azevedo[4], Daisy Elaine Andrade da Silva[1], Eliana Vieira Pinto da Silva[4], Sandro Patroca da Silva[1], Valéria Lima Carvalho[4], Giovanini Evelim Coelho[5], Ana Cecília Ribeiro Cruz[4], Sueli Guerreiro Rodrigues[4], Joao Lídio da Silva Gonçalves Vianez Jr[1], Bruno Tardelli Diniz Nunes[4], Jedson Ferreira Cardoso[1], Robert B Tesh[6], Simon I Hay[3,7], Oliver G Pybus[2†] and Pedro Fernando da Costa Vasconcelos[4,8†]

Abstract

Background: In December 2013, an outbreak of Chikungunya virus (CHIKV) caused by the Asian genotype was notified in the Caribbean. The outbreak has since spread to 38 regions in the Americas. By September 2014, the first autochthonous CHIKV infections were confirmed in Oiapoque, North Brazil, and in Feira de Santana, Northeast Brazil.

Methods: We compiled epidemiological and clinical data on suspected CHIKV cases in Brazil and polymerase-chain-reaction-based diagnostic was conducted on 68 serum samples from patients with symptom onset between April and September 2014. Two imported and four autochthonous cases were selected for virus propagation, RNA isolation, full-length genome sequencing, and phylogenetic analysis. We then followed CDC/PAHO guidelines to estimate the risk of establishment of CHIKV in Brazilian municipalities.

Results: We detected 41 CHIKV importations and 27 autochthonous cases in Brazil. Epidemiological and phylogenetic analyses indicated local transmission of the Asian CHIKV genotype in Oiapoque. Unexpectedly, we also discovered that the ECSA genotype is circulating in Feira de Santana. The presumed index case of the ECSA genotype was an individual who had recently returned from Angola and developed symptoms in Feira de Santana. We estimate that, if CHIKV becomes established in Brazil, transmission could occur in 94% of municipalities in the country and provide maps of the risk of importation of each strain of CHIKV in Brazil.

Conclusions: The etiological strains associated with the early-phase CHIKV outbreaks in Brazil belong to the Asian and ECSA genotypes. Continued surveillance and vector mitigation strategies are needed to reduce the future public health impact of CHIKV in the Americas.

Keywords: Chikungunya virus, National surveillance, Public health, Spatial prediction, Statistical methods

Background

The Chikungunya virus (CHIKV) is a re-emerging mosquito-borne alphavirus posing a significant public health problem in tropical and subtropical regions [1]. CHIKV infection is usually characterized by an acute onset of fever, rash, and arthralgias, and is often accompanied by headache, joint swelling, and conjunctivitis. Unlike dengue virus (DENV), CHIKV infection is associated with recurrent polyarthralgias and high rates of symptomatic infections. CHIKV is typically transmitted between humans by the anthropophilic vectors *Aedes aegypti* and *Ae. albopictus* [2].

Four CHIKV genotypes have been identified since its discovery in 1952 [3]. The East-Central-South-African (ECSA) and West African genotypes are endemic and cause epidemics in sub-Saharan Africa, whereas the Asian genotype circulates in *Ae. aegypti*-human urban transmission cycles in Southeast Asia. The Indian Ocean lineage (IOL) caused explosive epidemics in Indian

* Correspondence: marcionunesbrasil@yahoo.com.br
†Equal contributors
[1]Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA 67030-000, Brazil
Full list of author information is available at the end of the article

BioMed Central

© 2015 Nunes et al.; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Ocean islands and Asia between 2005 and 2011. Several IOL strains harbour mutations that augment infectivity and transmissibility in *Ae. albopictus* [4]. Imported cases caused limited outbreaks in temperate regions where only *Ae. albopictus* is present [5]. On December 5, 2013, Asian genotype CHIKV infections were reported in the Caribbean island of Saint Martin [6]. For the first time in recorded history, CHIKV had established a mosquito-human cycle in the Americas. As of October 31, 2014, >1,222,000 suspected CHIKV cases have been reported in the Americas and autochthonous infections have been confirmed in 50 territories in the region [7].

The first autochthonous cases of CHIKV in Brazil were confirmed in Oiapoque, Amapa State, on September 13, 2014. Seven days later, autochthonous cases were also confirmed in Feira de Santana, Bahia state. By October 18, 2014, 682 confirmed autochthonous cases had been notified to the Brazilian Ministry of Health [8]. Yet, the source of the ongoing outbreaks in these municipalities and the potential for virus establishment in the country remain unknown. We performed serological and virological analysis on 68 samples collected from June to September 2014 and full-length viral genome sequencing on six isolates. CHIKV and DENV have similar transmission cycles [9], and DENV is hyperendemic in Brazil [10]. Herein, we describe the epidemiological and genetic characteristics of CHIKV emergence in Brazil and provide a prediction of the risk of CHIKV importation and establishment in each Brazilian municipality for the coming year, using data on human mobility, vector distribution, and retrospective DENV incidence.

Methods

Patients and diagnostic assays

Serum was sampled from 68 patients exhibiting symptoms consistent with CHIKV infection and detected through a passive surveillance national system. Demographic data included patient age, sex, municipality, and federal state where the case was reported, date of onset of symptoms, date of sample collection, history of travel and residence, plus CHIKV and DENV serology. Data were provided by the Evandro Chagas Institute and the National Program for Dengue Control, Department of Surveillance in Health, Brazil. Samples were sent to the Evandro Chagas Institute for confirmation of CHIKV infection using IgM ELISA, real-time PCR, and C6/36 cell culture. Samples were obtained from persons visiting local clinics by Ministry of Health personnel as part of dengue surveillance activities. In these situations, patient consent is oral and not recorded. The study was authorized by the Coordination of the National Program for Dengue Control run by Brazil's Ministry of Health.

Virus propagation and RNA isolation

CHIKV isolates were recovered from sera of six acutely infected patients (P25, P34, P36, P37, P38, P39; Additional file 1: Text S1 and Table S1) in low-passage monolayer cultures of *Ae. albopictus* C6/36 cells, harvested 6 to 10 days post infection or after evident cytopathic effect. Culture suspensions were centrifuged and treated with 50% polyethylene glycol 8000 and 23% NaCl for virus precipitation. After centrifugation, virus pellets were eluted in RNase-free water and used for RNA isolation. Samples were subjected to a TRIzol separation step and the aqueous phase was further processed using Qiamp Viral RNA Minikit (Qiagen) according to the manufacturer's instructions.

Viral genome sequencing

Viral genomes were recovered using the Ion Torrent PGM [11]. Briefly, the cDNA synthesis reaction was performed using an in-house protocol with random primers. Samples were fragmented by enzymatic digestion and libraries built by the automated AB Library Builder System (Applied Biosystems). To normalize the number of molecules required for emulsion PCR, a quantitation step was performed using the automated Ion OneTouch 2 platform. The emulsion PCR reaction result comprising barcoded-pooled samples was loaded on a 318v2 chip and the sequencing reaction was performed on the PGM device. Raw reads were assembled using Mira v4.0 [12] (Additional file 1: Text S1). The six CHIKV genomes generated here have GenBank accession numbers: KP164567–KP164572.

Phylogenetic analysis

To investigate the origins of CHIKV in Brazil, 155 publicly available genomes (>10,000 nt, by June 2014) with known dates and locations of sampling were collected from GenBank [13] and added to the new Brazilian genomes. Sequences were aligned using MAFFT [14]. Viral phylogenies based on full-length nucleotide sequences were estimated using maximum likelihood implemented in PhyML [15] and using BEAST [16], a program for Bayesian analysis of molecular sequences using Markov Chain Monte Carlo (MCMC). The GTR + 4Γ + I substitution model was used for phylogenetic reconstruction, as suggested by jModelTest [17]. After excluding 85 IOL strains, a total of 76 CHIKV genomes representing all four viral genotypes were used. Specifically, the final full-length dataset of 76 full-length genomes included 11 West African, 12 ECSA, 17 IOL, and 30 Asian epidemic strains, plus 6 new Brazilian strains (Accession Numbers in Additional file 1: Text S1). A maximum likelihood analysis was also performed using an E1 envelope alignment including all 554 available sequences for which the date and country of collection were available (Additional file 1: Figure S2). For the full-length dataset, a relaxed lognormal molecular clock model [18] and the non-

parametric skygrid coalescent model [19] were employed. Using a strict molecular clock model and the skyline coalescent model [20] produced similar results. Six runs using the full-length dataset were computed for 50 million MCMC steps and convergence inspected using Tracer v.1.6 [16]. A maximum clade credibility phylogeny was obtained from the combined posterior distributions of the full-length dataset runs (excluding 10% burn-in) using TreeAnnotator [16]. Similar analyses were also computed separately for each genotype. For each estimated parameter, we report the mean and the corresponding uncertainty as 95% Bayesian credible intervals (BCI).

Predicting importation and establishment risk

Given the poor dispersal ability of *Ae. aegypti* and *Ae. albopictus* [21], introduction of CHIKV in Brazil will largely be driven by human movements. We therefore predicted the risk of CHIKV importation into each Brazilian municipality using a radiation parametric model of human mobility with selection, which has shown strong predictive power at sub-national scales [22]. Empirical data concerning short-term human movements (as opposed to long-term relocations) were unavailable for Brazil or surrounding countries, hence model parameters were estimated using a previously published directed commuting network for France, obtained from 5,695,974 anonymized mobile phone records comprising directional movements among 329 districts [23]. Population sizes in each of these districts and in Brazilian municipalities were extracted from the Gridded Population of the World, v3 dataset [24] using administrative boundary data from the GADM database [25]. Parameters were estimated by maximum likelihood using a standard numerical optimisation routine [26] in R software v3.1.1 [27] and applied to predict population-level movements between 5,494 municipalities in Brazil (R code available on request). For each municipality, we included the presence of *Ae. aegypti* and *Ae. albopictus* recorded from annual larval surveys conducted between 2007 and 2014 [28], or of reported DENV transmission between 2001 and 2013 [29] to determine the risk of CHIKV transmission. As DENV is transmitted by the same vectors as CHIKV, and is subject to similar dynamics, this is likely to represent a rational, although preliminary, estimate of establishment risk [2]. As an indication of the likely intensity of CHIKV transmission if the virus does become established, we also estimated an average DENV incidence rate for each municipality from the same dataset.

Results

Characterization of CHIKV cases in Brazil

Table 1 summarizes the demographic characteristics of the 68 patients with confirmed CHIKV infection whose date of symptom onset was between 8 April and 15 September, 2014 (further details in Additional file 1: Table S1). Travel history was known for 67 (99%) patients. Forty-one (60%) reported travelling or residing abroad (defined as imported cases), 21 (31%) were Brazilian military and missionary personnel that had returned from Haiti, 32% travelled to Caribbean islands, and 6 (15%) to other South American countries. The remaining 27 cases were defined as autochthonous.

The number, timeline, and geographic location of cases are shown in Figure 1. Local transmission was restricted to Amapa state (48%; with 70% of these being reported in the Oiapoque municipality) and Bahia state (52%). There was a rise in the number of confirmed autochthonous cases during the last week of August 2014, but no significant difference (unpaired *t*-test two sided P value = 0.72) in imported cases during the month of the World Cup event.

Asian and ECSA genotypes in Brazil

The time-scaled phylogeny in Figure 2 includes the six Brazilian CHIKV genomes reported here (2 imported, 4 autochthonous). Both imported cases (P25, P34) plus the Oiapoque autochthonous case (P37) belonged to the Asian genotype of CHIKV. Patient P25 resides in Recife and travelled in June to the Dominican Republic. Patient P34 travelled from Guadeloupe to Belem, Para state, in August. Patient P37 resides in Oiapoque, Amapa state, did not travel abroad, and exhibited symptoms at the end of August 2014. The three Asian genotype isolates clustered closely with available genomes from the Caribbean outbreak (from Saint Martin and British Virgin Islands). The CHIKV genome from patient P37 differed from the Saint Martin isolate by 7 mutations.

We discovered and confirmed a separate CHIKV introduction to Brazil in Feira de Santana, which comprises ECSA genotype infections, a CHIKV genotype previously undetected in the Americas (Figure 2). ECSA genotype genomes were obtained from three patients (P36, P38, P39) that had not travelled abroad. Specifically, the three genome sequences were near-identical, form a phylogenetic cluster with maximum support (Figure 2), and were proximate in time and space. This supports local transmission derived from a single viral importation. The Brazilian ECSA lineage is most closely related to a CHIKV genome isolated in 1962 in Angola (Figure 2, Additional file 1: Figure S1). Similar conclusions were obtained with a larger E1 envelope dataset (Additional file 1: Figure S2). The ECSA lineage index case arrived in Feira de Santana from Angola in June 2014 and three contacts of the index case reported CHIKV-like symptoms. None of the six genomes sequenced here contain the A226V (E1 protein) or the L210Q (E2 protein) mutations that increase CHIKV transmissibility and persistence in *Ae. albopictus* [30].

Table 1 Demographic characteristics of CHIKV confirmed cases in Brazil 2014

	Number of cases (%)
Age (years)	
0–19	8 (11.8%)
20–39	39 (57.4%)
40–59	15 (22.1%)
60–70	5 (7.4%)
Unknown	1 (1.5%)
Sex	
Female	30 (60%)
Male	20 (40%)
Unknown	18 (27%)
State of notification	
Amapa	16 (23.5%)
Bahia	14 (20.6%)
Ceara	5 (7.4%)
Brasilia	2 (2.9%)
Goiania	2 (2.9%)
Maranhao	1 (1.5%)
Para	1 (1.5%)
Paraná	2 (2.9%)
Rio de Janeiro	3 (4.4%)
Rio Grande do Sul	2 (2.9%)
Roraima	5 (7.4%)
Sao Paulo	15 (22.1%)
Country of origin	
Brazil	65 (95.6%)
Outside Brazil	3 (4.4%)
Transmission cases	
Imported	41 (60.3%)
Autochthonous	27 (39.7%)
Place of infection (Autochthonous)	
Amapa state	13 (48.2%)
Bahia state	14 (51.8%)
Place of infection (Imported)	
Haiti[a]	21 (51.2%)
French Guiana	1 (2.4%)
British Guiana	2 (4.9%)
Dominican Republic[b]	10 (24.4%)
Guadalupe[c]	2 (4.9%)
Venezuela	3 (7.3%)
Unknown[d]	2 (4.9%)

Data was collated between April and September 2014 from the Evandro Chagas Institute and reports from the National Program for Dengue Control. Patient's epidemiological and clinical details can be found in Additional file 1: Table S1.
[a]Includes two travellers from Haiti; [b]Includes one case that also travelled to Haiti; [c]Includes one traveller from Guadalupe; [d]Includes one patient that was infected in an unspecified Caribbean island.

The genomic evolutionary rate of the Asian genotype was estimated at 4.71 (95% BCI: 3.84–5.65) nucleotide substitutions per year, higher than that of the ECSA genotype (2.31, 95% BCI: 1.89–2.71), as previously noted [31]. We estimate that the common ancestor of CHIKV in the Caribbean existed around 23 August, 2012 (95% BCI: 3 October 2011 to 2 June 2013). We cannot estimate the date of introduction of the Asian genotype to Oiapoque because only one non-imported genome (P37) was available. However, we can date the common ancestor of the Brazilian ESCA genomes, estimated to be around March 6, 2014 (95% BCI: 29 July 2013, 15 August 2014) (Figure 2). Notably, these uncertainty intervals include the arrival date (June 2014) of the index case to Feira de Santana.

Estimating risks of importation and establishment in Brazil

Modelling of human mobility predicted different geographic patterns of CHIKV importation risk from Oiapoque and Feira de Santana (Figure 3). A total of 5,172 of the 5,494 municipalities in Brazil have reported the presence of either *Ae. aegypti* or *Ae. albopictus*, or DENV transmission. These municipalities are home to the vast majority (98.6%) of Brazil's population. Whilst only a subset of people in these regions are likely to be at risk of CHIKV infection, the public health impact of the virus is likely to be significant should it become established. We estimate that 2,093 (38.1%) municipalities in Brazil are in the top 50th percentile of relative risk of importation from both Oiapoque and Feira de Santana.

Of the 35 municipalities in the top 95th percentile of importation risk from both Oiapoque and Feira de Santana, 57% are located in the Northeast and 31% in the Southeast region, which includes Rio de Janeiro and Sao Paulo (Additional file 1: Table S2). This suggests a chance of two CHIKV genotypes becoming established in the same geographical areas. Retrospective analysis of DENV transmission indicates that CHIKV transmission is expected to occur in year-round transmission cycles, with transmission peaks starting in November/December and peaking between January to April, maintained mostly

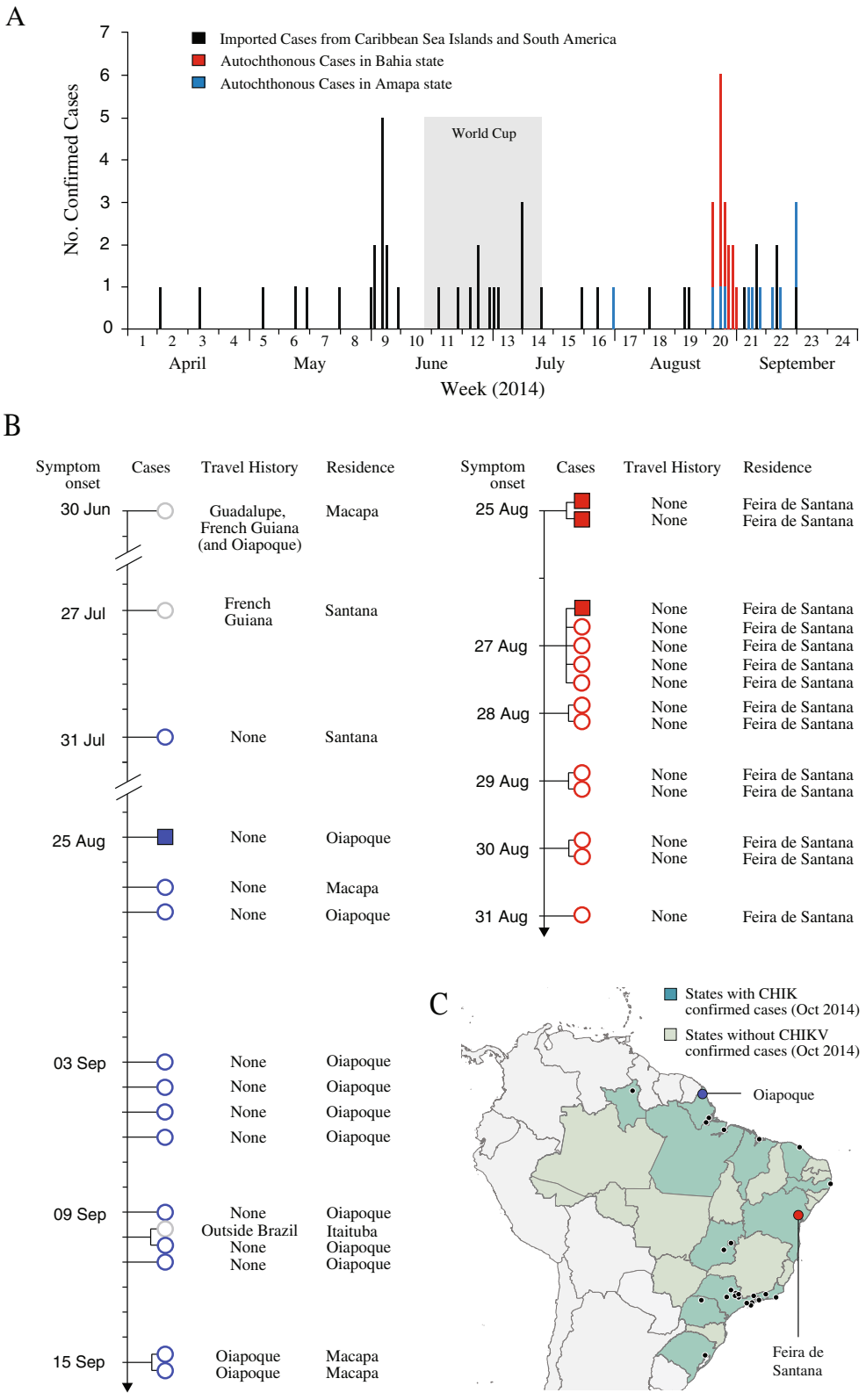


Figure 1 (See legend on next page.)

(See figure on previous page.)

Figure 1 Timeline of laboratory confirmed CHIKV cases in Brazil 2014. Shown are imported (black) and autochthonous (red and blue) CHIKV cases in Brazil between April and September 2014 **(A)**. As of September, autochthonous cases had been reported in Feira de Santana, a municipality in Bahia state, north-eastern Brazil (red) and in Amapa state, north Brazil, bordering French Guiana. Also shown is the timeline of confirmed CHIKV in Amapa and Bahia state **(B)**, along with patients' travel history and municipality of residence. Grey circles indicate imported cases, while blue and red circles indicate localized transmission. Filled squares indicate patient samples whose viruses were sequenced. **(C)** Map of confirmed cases and federal states with confirmed cases.

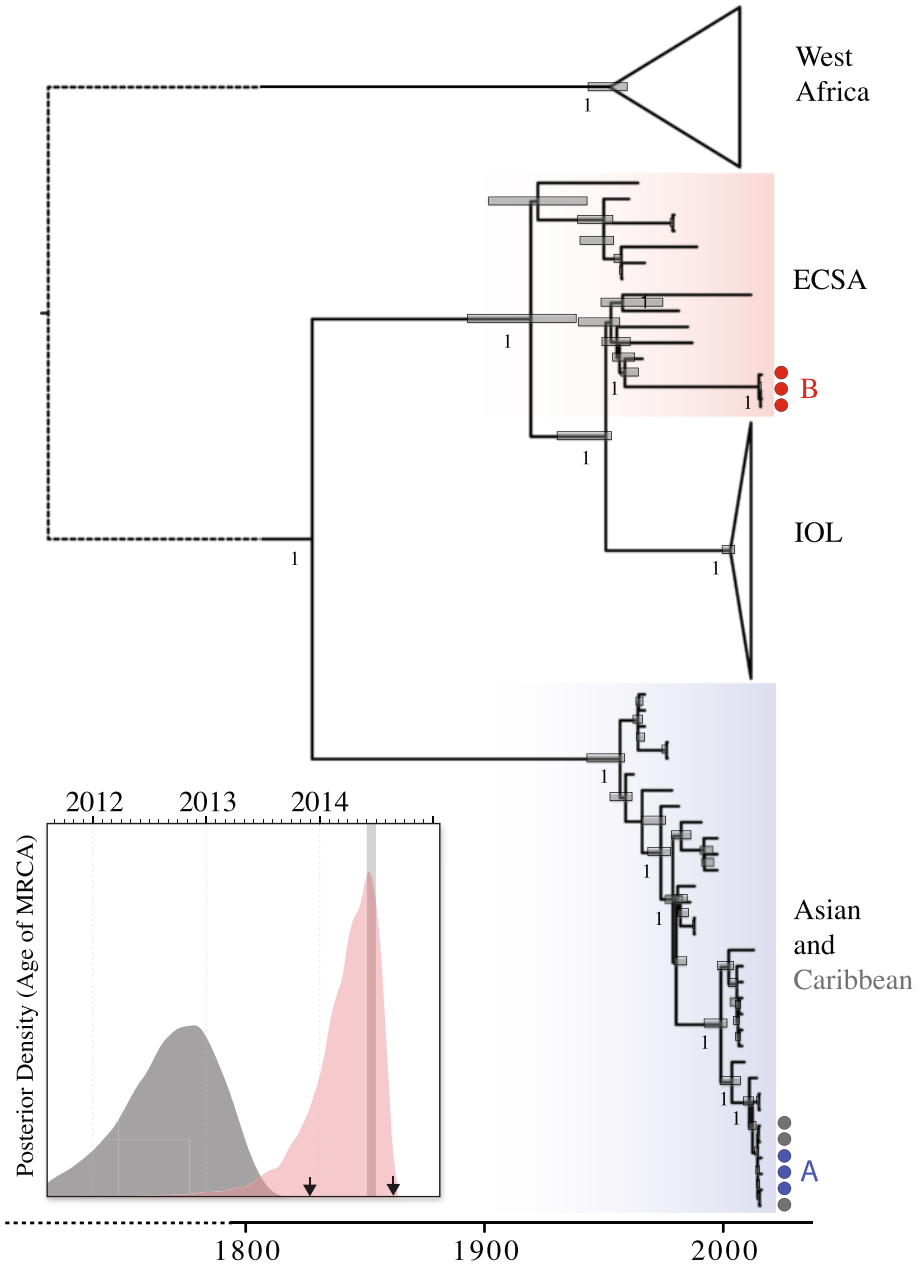


Figure 2 Dated phylogenetic reconstruction of CHIKV including Brazilian genomes. The genotypes involved in autochthonous Brazilian transmission in Oiapoque and Feira de Santana are highlighted in blue **(A)** and red **(B)**, respectively. On the right, isolates sampled in Feira de Santana and Oiapoque are depicted as red and blue circles, whilst isolates from the Caribbean outbreak are shown in grey circles. The inset shows the posterior density of the age of the most recent common ancestor (MRCA) of the Caribbean outbreak (**A**, grey) and the Feira de Santana cluster (**B**, red). The arrows highlight the dates of the first notifications of CHIKV cases in the Caribbean and in Feira de Santana, respectively. The index case of the ECSA genotype in Brazil arrived in Feira de Santana in June (grey bar). A ML tree with tip information can be found in Additional file 1: Figure S1.

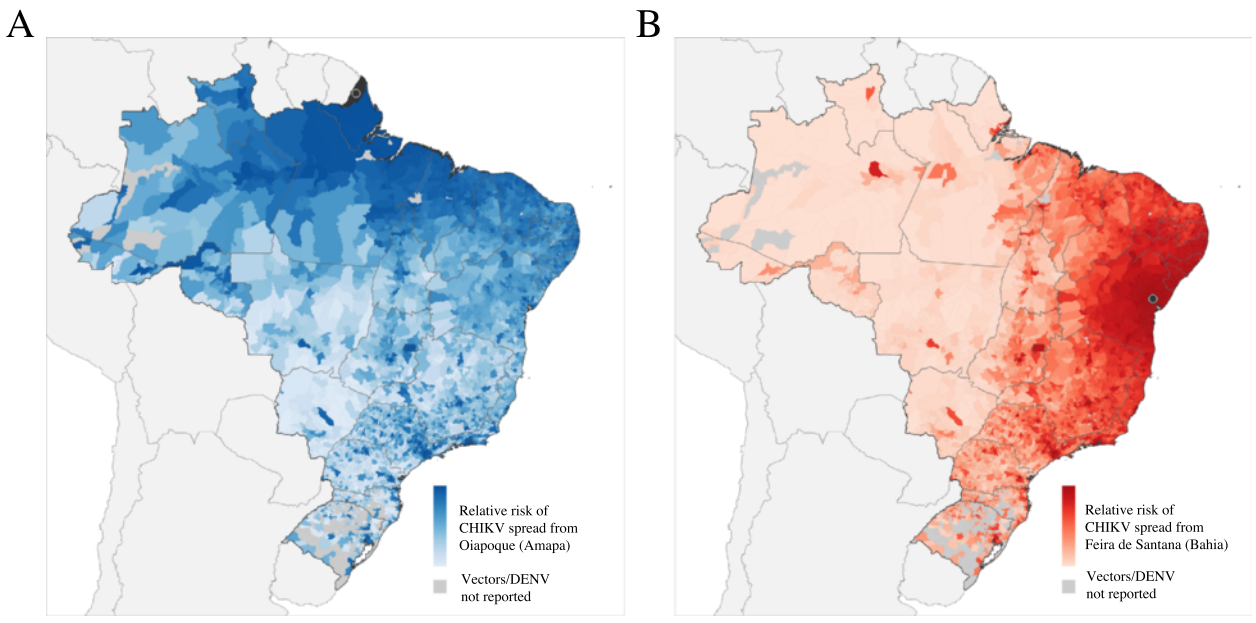


Figure 3 Risk maps for CHIKV spread from Oiapoque and Feira de Santana, Brazil. Shown is the relative risk of CHIKV spread from Oiapoque **(A)** and Feira de Santana **(B)**. Black circles denote the locations of these municipalities.

in less populated areas located in the north and central areas of Brazil (Figure 4).

Discussion

This study reports the emergence of both Asian and ECSA CHIKV genotypes in Brazil. This is the first time that an outbreak of the ECSA genotype has been reported in the Americas. Genetic and epidemiological data suggest that the ECSA genotype was introduced from Angola to Feira de Santana in June 2014, where epidemiological investigation is ongoing. CHIKV appears to be endemic in Angola [32], where thousands of Brazilians, mostly from Northeast and Southeast areas [33], work in the petroleum and mining industries. Genetic data suggests multiple introductions of the Asian genotype to Brazil from ongoing epidemics in the Caribbean and South America. Surveillance data suggests autochthonous transmission in Oiapoque began by early September 2014 [8], and we posit that the strain circulating there was imported from French Guiana, a country bordering Oiapoque that has reported a steady increase in autochthonous cases since January 2014 [8].

The clinical picture of patients with confirmed CHIKV infection included fever, arthralgia, rash, myalgia, and headache, symptoms similar to DENV infection, which is endemic in Brazil. The fact that we observe two different strains is sufficient to demonstrate that the epidemics in Oiapoque and Feira de Santana resulted from separate introductions and this result is robust to the number of sequences obtained. Our genetic estimate of the date of introduction of the ECSA genotype to Feira de Santana has an upper bound of 15 August, 2014, suggesting that the CHIKV surveillance network in Brazil was able to detect early cases of this introduction. Moreover, the confirmation of ~7 (3 to 13) importations of CHIKV to Brazil per month implies that Brazil is at risk of additional importations from endemic regions and that CHIKV may be exported from Brazil to other locations.

It is possible that the sustained transmission described here in results from the introduction of CHIKV strains into a location with appropriate vector abundance and during a period with appropriate climate conditions [34], particularly during monsoon rainfall period and high temperatures [35]. Some additional observations deserve further investigation. Prediction of future transmission of CHIKV within Brazil indicates that the geographic ranges of the current Asian and ECSA lineages will likely overlap and that nearly 99% of the population in Brazil may be at risk for CHIKV infection. There is no population immunity to CHIKV in Brazil, so incidence is expected to increase [36]. Training health care workers for differential diagnosis of DENV and CHIKV, increased availability of diagnostic tests for patient sera and mosquitoes, as well as educational programs and a national active surveillance system are needed to control CHIKV spread [3]. A small percentage of those living in rural northern and western areas of the country have antibodies to Mayaro virus (MAYV), a related alphavirus associated with Haemagogus mosquitoes [34]. Cross-reactivity between the antigenically-related MAYV and CHIKV [37] has been reported [38] but it is unknown if prior exposure to MAYV provides protection against

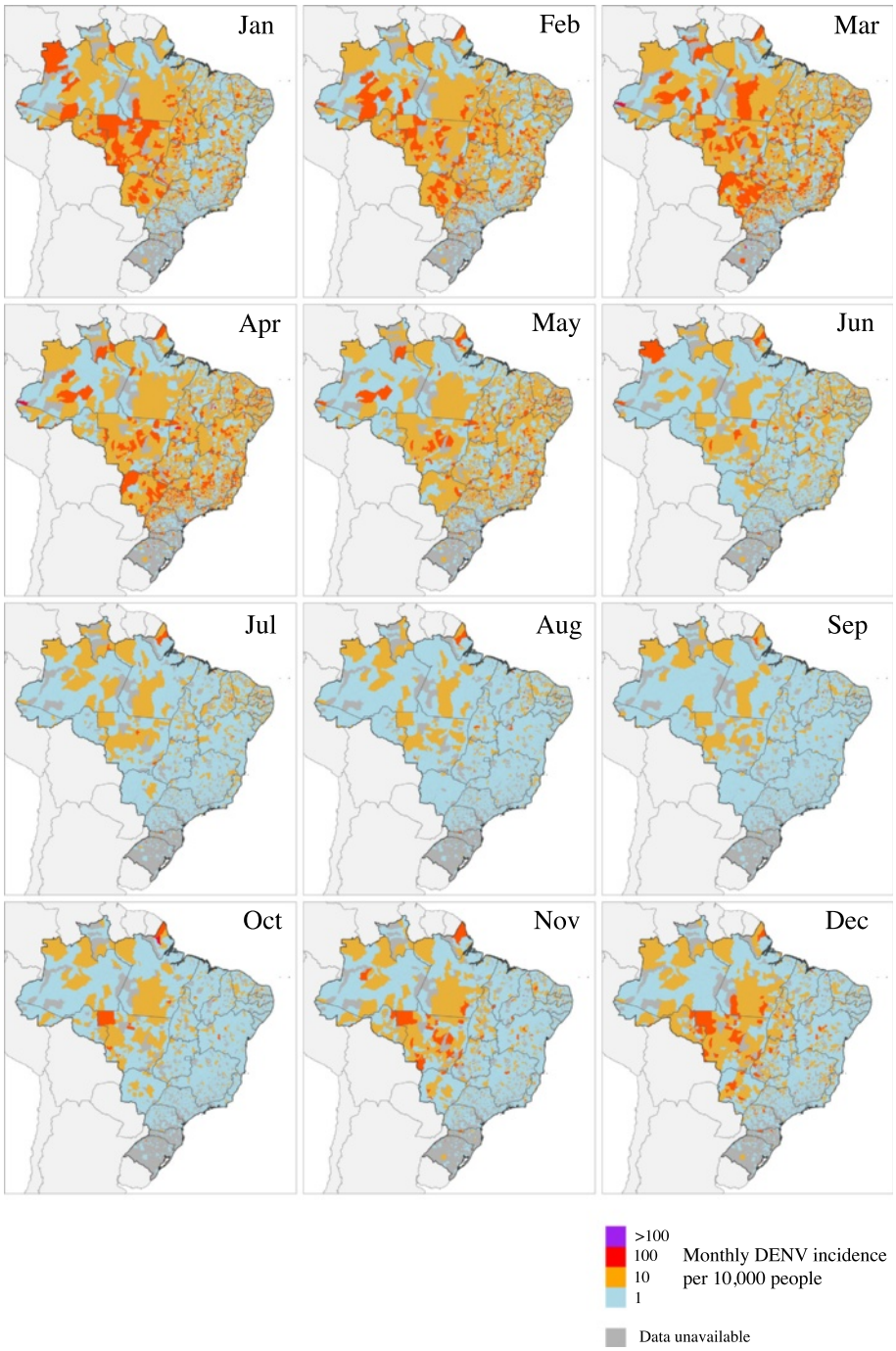


Figure 4 Retrospective analysis of dengue virus transmission to identify municipalities where CHIKV is expected to circulate. Spatial dengue virus monthly incidence (averaged between 2001 and 2013) is shown in Brazilian municipalities.

CHIKV. As CHIKV genotypes are more similar to each other than to MAYV it seems probable that immunity against one CHIKV genotype will provide at least partial protection against another [39].

We follow international guidance and evaluate CHIKV importation risk using data on past DENV transmission [2]. The similarity of DENV and CHIKV transmission cycles [9] suggests that CHIKV may become endemic in several Brazilian municipalities, with transmission peaks occurring between January and April (Figure 4). Risk of CHIKV transmission is driven by a range of factors, including the presence and local abundance of suitable vector species, environmental variables such as ambient temperature, and socioeconomic factors. Whilst few of

these factors are well understood or quantified for Brazil, many are likely to be shared with DENV, and it is known that the CHIKV-competent *Ae. aegypti* and *Ae. albopictus* vectors [40] are widespread in Brazil. More specifically, *Ae. aegypti* is more dispersed in Brazil with higher incidence in northern, north-eastern, central-eastern regions and less frequent in southern Brazil, due to a cooler climate [41]. In contrast, *Ae. albopictus* has high incidence in subtropical areas, more specifically in southern areas of the country [28]. Although our index of introduction risk (Figure 3) indicates the municipalities in Brazil at greatest risk for CHIKV transmission, and can be used to prioritise disease surveillance, we cannot estimate the absolute number of importations expected in each location over a given time frame because the relevant data on human mobility within Brazil are currently unavailable. This could be obtained from, for example, anonymised cell phone call records [42] or through collating existing microcensus data.

The ECSA outbreak in Feira de Santana persists and is disseminating to other regions in Brazil – and potentially to locations outside the country. Therefore, continued genetic surveillance will be necessary to investigate whether the ECSA genotype will acquire the A226V mutation in the viral E1 protein that can increase viral transmission in *Ae. albopictus* and which was associated with explosive CHIKV outbreaks in the Indian Ocean nations and Indian subcontinent [4]. Other scenarios are also possible: the Asian CHIKV genotype may become dominant in tropical regions of Brazil, where *Ae. aegypti* is well established, and the ECSA genotype in subtropical and more temperate regions, where *Ae. albopictus* is more abundant. The ECSA genotype has acquired mutations that increase transmissibility and persistence in *Ae. albopictus* on multiple independent occasions [3], whereas the Asian genotype appears less able to accrue such adaptations. In the long term, both genotypes could potentially disappear from the region if levels of human population immunity increase sufficiently. Alternatively, CHIKV could establish an enzootic cycle in the region with sporadic human epidemics. This pattern has been seen in Africa [31] and potentially in Southeast Asia [43], where there is some evidence of a sylvatic CHIKV transmission cycle involving non-human primates and forest-dwelling mosquitoes, similar to that observed for sylvatic yellow fever virus, which was introduced into tropical Americas from Africa with the slave trade and is still endemic in the Amazon and Orinoco River basins [44]. DENV serotypes also exhibit sylvatic cycles in some areas of Southeast Asia and Africa, but this has not been yet observed in the Americas.

Conclusions

In summary, this is the first report of CHIKV emergence in Brazil. Remarkably, we detected for the first time, ongoing transmission of the ECSA CHIKV genotype in the Americas. With CHIKV transmission established in highly connected regions in the north and northeast of Brazil and with the peak suitability season fast approaching, now is the time for preventative action. Effective, targeted, and sustained viral and case surveillance and vector control measures have the potential to avert invasion of CHIKV and avoid the overwhelming of one of the world's largest healthcare systems.

Additional file

Additional file 1: Methodological details for sequencing, genetic and spatial data analysis, and supplementary tables and figures.

Abbreviations

BCI: Bayesian credible intervals; CHIKV: Chikungunya virus; DENV: Dengue virus; ECSA: East-Central-South-African; IOL: Indian Ocean lineage; MAYV: Mayaro virus; MCMC: Markov Chain Monte Carlo.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MRTN and NRF designed the research study, performed the evolutionary analyses, analysed data, and wrote the manuscript. JMV collected data, performed genome sequencing, and edited the manuscript. NG and MUGK conceived the statistical analysis, evaluated results, and edited the manuscript. LFO, DEAS, and SPS performed genome sequencing. RSSA contributed with clinical data. EVPS and VLC carried out the cell culture analysis. SPS and JFC participated in sequence alignment. GEC collected epidemiological data. ACRC performed the molecular biology diagnostic tests. SGR performed serological tests. JLSGVJr participated in sequence alignment and edited the manuscript. BTDN performed the qPCR analyses. RBT and SIH analysed data and revised the manuscript. OGP and PFCV helped design the study and interpret results, contributed epidemiological information, and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to the Brazilian Dengue and Chikungunya network composed by 17 national laboratories for their assistance on preliminary screening diagnostics. We thank Oliver Brady for useful discussions and Thessika Hialla Araujo, Luiz Carlos Jr. Alcantara, and Daniele Barbosa Medeiros for providing epidemiological information. MRTN is supported by the Brazilian National Council of Scientific and Technological Development (CNPq) grant no. 302032/2011-8. PFCV is supported by CNPq grant no. 573.739/2008-0 and 240/2008/FAPESPA/INCT. JMV is a PhD student supported by CNPq grant no. 142492/2011-6. NG is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). MUGK is funded by the German Academic Exchange Service (DAAD) graduate scholarship. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) and receives support from the Li Ka Shing Foundation and also acknowledges funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center (FIC), National Institutes of Health (NIH). This research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 614725.

Author details

[1]Center for Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA 67030-000, Brazil. [2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [3]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [4]Department of Arbovirology and Hemorrhagic Fevers, Evandro Chagas Institute, Ministry of Health, Ananindeua, PA 67030-000, Brazil.

[5]National Dengue Control Program, Brazilian Ministry of Health, Brasilia, DF 70058-900, Brazil. [6]Department of Pathology, University of Texas Medical Branch, Galveston, Texas TX 77555-0609, USA. [7]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [8]Department of Pathology, Para State University, Belem, PA 66087-670, Brazil.

Received: 5 January 2015 Accepted: 13 April 2015
Published online: 30 April 2015

References

1. Pialoux G, Gauzere BA, Jaureguiberry S, Strobel M. Chikungunya, an epidemic arbovirosis. Lancet Infect Dis. 2007;7:319–27.
2. Pan American Health Organization. Preparedness and response for Chikungunya virus: introduction in the Americas. Washington, DC: PAHO/CDC; 2011.
3. Weaver SC. Arrival of Chikungunya virus in the new world: prospects for spread and impact on public health. PLoS Negl Trop Dis. 2014;8, e2921.
4. Tsetsarkin KA, Chen R, Yun R, Rossi SL, Plante KS, Guerbois M, et al. Multi-peaked adaptive landscape for Chikungunya virus evolution predicts continued fitness optimization in Aedes albopictus mosquitoes. Nat Commun. 2014;5:4084.
5. Soumahoro MK, Fontenille D, Turbelin C, Pelat C, Boyd A, Flahault A, et al. Imported Chikungunya virus infection. Emerg Infect Dis. 2010;16:162–3.
6. Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. Chikungunya in the Americas. Lancet. 2014;383:514.
7. Pan American Health Organization. Chikungunya. Washington, DC: PAHO. http://www.paho.org/hq/index.php?option=com_topics&view=article&id=343&Itemid=40931.
8. Brazilian Health Portal. Boletim Epidemiológico. http://portalsaude.saude.gov.br/. 2015.
9. Roth A, Mercier A, Lepers C, Hoy D, Duituturaga S, Benyon E, et al. Concurrent outbreaks of dengue, Chikungunya and Zika virus infections – an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012–2014. Euro Surveill. 2014;19:1.
10. Villabona-Arenas CJ, de Oliveira JL, Capra Cde S, Balarini K, Loureiro M, Fonseca CR, et al. Detection of four dengue serotypes suggests rise in hyperendemicity in urban centers of Brazil. PLoS Negl Trop Dis. 2014;8, e2620.
11. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348–52.
12. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T, et al. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004;14:1147–59.
13. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2014;42:D32–7.
14. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005;33:511–8.
15. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73.
17. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9:772.
18. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4, e88.
19. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol. 2013;30:713–24.
20. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185–92.
21. Honorio NA, Silva Wda C, Leite PJ, Goncalves JM, Lounibos LP, Lourenco-de-Oliveira R. Dispersal of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. Mem Inst Oswaldo Cruz. 2003;98:191–8.
22. Simini F, Maritan A, Neda Z. Human mobility in a continuum approach. PLoS One. 2013;8, e60069.
23. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, et al. Dynamic population mapping using mobile phone data. PNAS. 2014;111:15888–93.
24. Socioeconomic Data and Applications Center (sedac). Gridded population of the world, version 3 (GPWv3) and the Global Rural–urban Mapping Project (GRUM); 2010. http://sedac.ciesin.columbia.edu/data/collection/gpw-v3.
25. Global Administrative Areas. http://www.gadm.org.
26. Nelder JA, Mead R. A simplex method for function minimization. Comput J. 1965;7:308–13.
27. Team RC. R: a language and environment for computing. Vienna: R Foundation for Statistical Computing; 2014.
28. Carvalho RG, Lourenco-de-Oliveira R, Braga IA. Updating the geographical distribution and frequency of Aedes albopictus in Brazil with remarks regarding its range in the Americas. Mem Inst Oswaldo Cruz. 2014;109:787–96.
29. Saúde MD: Dengue: Notificações registradas no Sistema de Informação de Agravos de Notificação – SINAN, Brazil. http://dtr2004.saude.gov.br/sinanweb.
30. Tsetsarkin KA, Weaver SC. Sequential adaptive mutations enhance efficient vector switching by Chikungunya virus and its epidemic emergence. PLoS Pathog. 2011;7, e1002412.
31. Volk SM, Chen R, Tsetsarkin KA, Adams AP, Garcia TI, Sall AA, et al. Genome-scale phylogenetic analyses of chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. J Virol. 2010;84:6497–504.
32. Filipe AF, Pinto MR. Arbovirus studies in Luanda, Angola. 2. Virological and serological studies during an outbreak of dengue-like disease caused by the Chikungunya virus. Bull World Health Organ. 1973;49:37–40.
33. IBGE. Instituto Brasileiro de Geografia e Estatistica (SIDRA) Available at http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=3173&z=t&o=3. Accessed 07 Nov 2014.
34. Vasconcelos PFC, Travassos da Rosa APA, Degallier N, Travassos da Rosa JFS, Pinheiro FR. Clinical and ecoepidemiological situation of human arboviruses in Brazilian Amazonia. J Braz Assoc Adv Sci. 1992;43:117–24.
35. Codeco CT, Lima AW, Araujo SC, Lima JB, Maciel-de-Freitas R, Honorio NA, et al. Surveillance of Aedes aegypti: Comparison of House Index with Four Alternative Traps. PLoS Negl Trop Dis. 2015;9, e0003475.
36. Staples JE, Fischer M. Chikungunya virus in the Americas–what a vectorborne pathogen can do. N Engl J Med. 2014;371:887–9.
37. Powers AM, Brault AC, Shirako Y, et al. Evolutionary relationships and systematics of the alphaviruses. J Virol. 2001;75:10118–31.
38. Hassing RJ, Leparc-Goffart I, Tolou H, van Doornum G, van Genderen PJ. Cross-reactivity of antibodies to viruses belonging to the Semliki forest sero-complex. Euro surveillance 2010; 15.
39. Sam IC, Chan YF, Chan SY, Loong SK, Chin HK, Hooi PS, et al. Chikungunya virus of Asian and Central/East African genotypes in Malaysia. J Clin Virol. 2009;46:180–3.
40. Vega-Rúa A, Zouache K, Girod R, Failloux A-B, Lourenço-de-Oliveira R. High level of vector competence of Aedes aegypti and Aedes albopictus from ten American countries as a crucial factor in the spread of Chikungunya virus. J Virol. 2014;88:6294–306.
41. Coelho GE. Challenges in the control of Aedes aegypti. Rev Inst Med Trop Sao Paulo. 2012;54:S13–14.
42. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science. 2012;338:267–70.
43. Apandi Y, Nazni WA, Noor Azleen ZA, Vythilingam I, Noorazian MY, Azahari AH, et al. The first isolation of Chikungunya virus from non-human primates in Malaysia. J Gen Molec Virol. 2009;1:35–9.
44. Taylor RM, FdC J, et al. An epidemiological study of jungle yellow fever in an endemic area in Brazil; epidemiology of human infections; investigation of vertebrate hosts and arthropod vectors. AmJTrop Med Hyg. 1946;26:69.

VECTOR-BORNE DISEASES

The many projected futures of dengue

Jane P. Messina[1], Oliver J. Brady[1], David M. Pigott[1], Nick Golding[1], Moritz U. G. Kraemer[1], Thomas W. Scott[2,3], G. R. William Wint[4], David L. Smith[1,3] and Simon I. Hay[1,3]

Abstract | Dengue is a vector-borne disease that causes a substantial public health burden within its expanding range. Several modelling studies have attempted to predict the future global distribution of dengue. However, the resulting projections are difficult to compare and are sometimes contradictory because the models differ in their approach, in the quality of the disease data that they use and in the choice of variables that drive disease distribution. In this Review, we compare the main approaches that have been used to model the future global distribution of dengue and propose a set of minimum criteria for future projections that, by analogy, are applicable to other vector-borne diseases.

Hyper-endemicity
A term used to indicate that a disease is constantly present in a location, with high incidence and/or prevalence rates and affecting all age groups. For dengue, co-circulation of all four serotypes in a location can be an indicator of hyper-endemic transmission.

[1]*Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.*
[2]*Department of Entomology and Nematology, University of California Davis, Davis, California 95616, USA.*
[3]*Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.*
[4]*Environmental Research Group Oxford, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.*
Correspondence to J.P.M.
e-mail: jane.messina@zoo.ox.ac.uk
doi:10.1038/nrmicro3430
Published online 2 March 2015

Several pathogens with an important impact on human health are transmitted between hosts via vectors, particularly insects and ticks, and the distribution of these vector-borne diseases is determined by a complex interplay between environmental and social factors. Several studies have evaluated the potential effects of climate change on the future spread of vector-borne diseases[1–3], including mosquito-borne diseases such as malaria[4–7], dengue[8–10], West Nile virus[11,12] and Japanese encephalitis[13], as well as tick-borne diseases such as Lyme disease[14,15] and tick-borne encephalitis[16–19]. These studies respond to growing concern over the impact of climate change on human health[20–25] and make use of the fact that many biological aspects of vector-borne disease transmission have quantifiable processes that depend on specific meteorological factors[26]. For example, warmer temperatures are associated with increased vector reproduction, more frequent biting behaviour and changes in vector survival, whereas the incubation period of pathogens inside the vectors can decrease as temperature rises[27–29]. Changes in humidity and precipitation may also affect the availability of aquatic habitats required for the juvenile stages of vectors such as mosquito larvae[2]. Therefore, the effect of climate change on these factors has become a widespread public health concern, and it is clearly fundamental to include meteorological variables when modelling the present or future distribution of a vector-borne disease.

However, the distributions of vector-borne diseases are not only linked to climatic conditions; the spread, establishment and persistence of these diseases also depend greatly on characteristics of the natural (that is, vegetation and geological features) and human-built environments. For example, urbanization trends and human movement patterns are thought to have been key drivers of the geographical expansion and increased endemicity of dengue over the past 70 years[30,31]. The expansion of urban environments and the increased frequency of international travel are likely to be equal, if not greater, contributors to change in the future landscape of dengue[9,32]. It is projected that by 2030, more than half of the world's population will reside in urban parts of the tropics owing to both population growth in existing urban areas and migration away from rural areas[33]. Given that the primary dengue vector — the *Aedes aegypti* mosquito (BOX 1) — tends to be found in peridomestic urban environments, these trends are likely to have a significant impact on the future distribution of this disease[34]. Similarly, because adult *Ae. aegypti* and the secondary dengue vector *Aedes albopictus*[35–37] fly relatively short distances throughout their lifetime[38,39], the rising frequency of international human travel is predicted to be the main driver for the geographical spread and increased co-circulation of the four dengue virus types (DENV1–DENV4) across many areas of the world (BOX 1). Importantly, an increase in the number of dengue virus types co-circulating in a particular region is a key indicator of progression towards disease hyper-endemicity and increased frequency of severe disease[40].

Dengue occurs in more than 125 countries and causes approximately 100 million symptomatic infections per year[41] among the 3.97 billion people that, as of 2010, are estimated to live in areas at risk from the disease[42] (BOX 1). It has been hypothesized that dengue virus transmission will both intensify in already endemic areas[43,44] and spread to currently low-risk parts of Asia,

© 2015 Macmillan Publishers Limited. All rights reserved

Box 1 | **The basics of dengue fever**

The global burden of dengue

Dengue is a vector-borne disease that is ubiquitous throughout the tropics and subtropics, with approximately 4 billion of the world's population at risk and nearly 400 million infections per year (25% of which are symptomatic). In most cases, the disease manifests itself as flu-like symptoms, such as fever and headache, although more severe cases result in haemorrhagic fever. There are currently no effective antiviral agents to treat dengue infection, and there is no licensed vaccine despite the existence of some promising candidates.

Dengue viruses

There are four main phylogenetically and antigenically distinct dengue viruses (DENV1–DENV4); although infection with one type confers long-term homologous immunity, it does not provide heterotypic immunity. Furthermore, secondary infection with a heterologous dengue virus type may increase the probability of severe disease.

Dengue vectors

Dengue is transmitted between humans by mosquitoes of the genus *Aedes*, which take blood meals from humans during the day and lay their eggs in small, open water-holding containers in domestic environments. Most *Aedes* species, including the primary dengue vector *Aedes aegypti*, require warm temperatures for persistence, and dengue virus incubation inside these mosquitoes also requires a minimum temperature. For these reasons, *Ae. aegypti* is an efficient urban vector of dengue virus in tropical and subtropical areas. By contrast, *Aedes albopictus* has adapted to more temperate climates in Europe and North America, but various aspects of its ecology and the temperatures required for dengue virus incubation mean that it does not vector sustained dengue outbreaks.

Dengue transmission and environmental change

As many aspects of dengue transmission have quantifiable processes that depend on temperature, the effect of climate change on future dengue transmission has become a widespread public health concern. Changes in humidity and precipitation may affect adult mosquito behaviour and the availability of aquatic habitats required for *Aedes* larval development. The persistence of dengue also depends on characteristics and changes in the environment and human population. For example, urbanization trends and human movement patterns, including the increased frequency of international travel, are likely to be equal, if not greater, contributors to change in the future landscape of dengue as climate change.

Europe and North America[44–46]. However, the relative importance of human and climate change-related factors in facilitating this spread remains a topic of debate[47–49]. Although projected trends in many of these factors (such as rising temperatures and increased urbanization) might suggest future increases in dengue incidence, wide-scale coordinated intervention efforts have been shown to reduce *Ae. aegypti* populations in the Americas in the past[50,51] and may be effective at lowering dengue incidence in some regions in the future. Furthermore, it has also been postulated that changes in the socio-economic status in some areas could prevent the establishment of dengue via behavioural changes such as the increased use of air conditioning, which may increase the amount of time humans spend indoors and therefore reduce the frequency of contact between mosquitoes and humans[52–55]. For example, the widespread access to such technologies in the United States has been indicated as the cause of lower dengue transmission in southern United States border towns compared with their Mexican neighbours[56]. Similarly, it has been suggested that changes in individual human behaviours and effective vector control have led to the decline and even elimination of dengue transmission in certain parts of the United States and Australia[57].

Unfortunately, it is difficult for models to accurately reconstruct the historical range of dengue transmission because of disparities in data quality and quantity. At best, changes in the number of reported occurrences of the disease can be shown over time, but these must be considered within the context of variable reporting practices and accuracy of dengue diagnostic methods throughout the history of the disease[40]. Attempting to define direct statistical associations between changes in reported dengue occurrence and environmental or socio-economic changes over long periods of time would be ill advised. However, for dengue[41] and other vector-borne diseases such as malaria[58,59], schistosomiasis[60], Lassa fever[61] and leishmaniasis[62], their current global distribution can be modelled based on recent, smaller-scale studies addressing the associations between disease occurrence and different variables that drive disease distribution (known as covariates). Incorporating projected changes in these underlying drivers of disease occurrence can then be used to project the future disease distribution; for dengue, several studies have done so already[63–67]. However, because different models have been used, the resulting projections are difficult to compare and are sometimes contradictory.

In this Review, we compare and contrast the different studies that have projected the future global distribution of dengue and evaluate their strengths and limitations based on the different covariates used in the models and on considerations of methodological rigour, the quality of disease data and how the different models incorporate bias and uncertainty. Building on the lessons learned in this evaluation, we suggest a framework for making future projections of the global distribution of dengue and, by analogy, other vector-borne diseases.

Dengue distribution models and projections

Projections of the future global distribution of dengue have been generated using either a biologically based mechanistic modelling approach or an empirically based statistical approach[68,69]. Mechanistic modelling is appropriate when specific biological processes related to virus transmission in relation to the vector and/or human host have been well described and parameterized through controlled experiments. In this approach, relationships are defined based on theoretical knowledge or from field and/or laboratory experiments and can be independently projected as long as their parameters are also projectable into the future. In the specific case of dengue, much attention has been devoted to the impact of meteorological variables on *Ae. aegypti* (and, to a lesser degree, *Ae. albopictus*) survival rates and their competence as dengue virus vectors[70–75].

By contrast, statistical modelling defines an empirical relationship between locations of known virus occurrence and a set of hypothesized underlying covariates, and then extrapolates these relationships to a future time point using projections of the covariates. For dengue, although mechanistic modelling approaches may better capture the temperature-dependent physiological aspects of dengue virus transmission and therefore be applicable in a broad range of transmission settings,

Covariates
Variables such as temperature or rainfall that may be used to predict disease occurrence. These may be direct or indirect in terms of the hypothesized relationship with the outcome.

© 2015 Macmillan Publishers Limited. All rights reserved

statistical approaches can better encompass the locally and temporally variable ecological aspects of transmission. For example, water storage practices in low-income urban and peri-urban housing areas, along with sufficient amounts of precipitation, could increase the abundance of aquatic habitats required for *Ae. aegypti* larval development[2], thereby increasing the incidence of dengue in these areas. Such relationships between environmental conditions and observed disease occurrence can be best accounted for by using statistical approaches applied to data on human infection rates in specific locations.

Mechanistic studies. The three mechanistic models that have been developed to project global dengue distribution used an output metric named epidemic potential[63,64] (also known as potential transmission intensity[65]). These terms represent a critical vector density threshold required to maintain virus transmission. This threshold was derived from the vectorial capacity equation, which measures the number of potentially infectious bites delivered by a vector population in 1 day[76,77]. In all three studies, the epidemic potential was modelled exclusively as a function of temperature along with certain fixed biological parameters derived from previous laboratory and field findings. Historical temperature averages either from global meteorological stations or from climate model data were used, and biological parameters included factors such as the probabilities of dengue being transferred between humans and mosquitoes during mosquito biting, and the survival rate and biting frequency of the mosquitoes. Two of the studies additionally applied their model to predict the expected maximum number of cases during peak dengue season for a small number of cities worldwide that have well-documented dengue virus transmission information[63,65]. These studies also validated their models by comparing these predictions with the reported case numbers in the corresponding cities[63,65]. Although the observed and expected cases followed similar trends in these cities, the amount of information provided is limited and the variance is not formally quantified. Finally, these two studies recomputed the epidemic potential for the same locations under specific warming scenarios (2 °C or 4 °C), or according to general circulation models of climate change for the year 2050, to provide projections of the relative change in the global distribution of dengue given rises in temperature (FIG. 1).

Interestingly, the three mechanistic models highlighted several locations as having baseline transmission potential that had not yet experienced any dengue outbreaks — for example, in southern California, where breeding populations of *Ae. aegypti* were only reported for the first time in 2013 (REF. 78) (FIG. 1). Projecting marked increases (as high as tenfold) in transmission potential from these biased baseline estimates is therefore potentially misleading. Similarly, one of these studies[65] also projected that new dengue transmission would occur in the Sahara desert under the climate warming scenarios; in reality, this region is unlikely to have enough human hosts or larval development sites for transmission to occur.

Statistical studies. The two statistical studies that have been applied to dengue used historical national-level (and subnational-level, where available) reports on dengue occurrence to determine contemporary dengue presence or absence at a global level. These studies then drew associations between dengue presence or absence and a set of hypothesized underlying covariates. Finally, the studies extrapolated their models to future years by using projections of the different covariates to map projected global dengue risk. However, the studies differed in the historical data on dengue occurrence that was used, in the particular covariates tested and in the specific statistical methodologies that were applied.

The earlier of these studies[67] included dengue reports from 1975 to 1996 and used logistic regression to model several meteorological variables related to temperature, precipitation and humidity, as well as the interactions between the variables as modelling covariates. The authors ultimately selected a model with a single dependent parameter (annual mean vapour pressure) based on parsimonious model selection criteria. They then extrapolated this model to future years using projections of changes in annual mean vapour pressure to map projected global dengue risk (FIG. 2). The later study[66] included dengue reports from 1975 to 2009 and used generalized additive models to include national-level gross domestic product (GDP) per capita in addition to meteorological variables in baseline models and projections, using a probability cut-off to then map dengue presence or absence in a location (FIG. 2).

There are some notable distinctions between the two studies that highlight the extent-limiting effect of economic information in the statistical modelling procedure. For example, the first model[67] projected very high dengue risk in the south-eastern United States and northern Australia, as well as parts of Brazil, Paraguay and China (FIG. 2). However, these areas are expected to experience a contraction of dengue when GDP is also considered, as predicted by the second model[66] (FIG. 2). These observations indicate that although climate change may lead to more suitable weather conditions for dengue transmission in some parts of the world, socio-economic factors could markedly limit the extent to which transmission will actually occur.

Lessons learned from previous projections

Although all of the studies conducted to date project an increase in the overall global extent of dengue transmission, the results do not agree with regard to the specific geographical areas where expansion or intensification is likely to occur. As projecting the global distribution of dengue involves the inclusion of multiple covariates, the comparison of new projections would be simplified if the international organizations involved with projecting changes in population, climate and economic factors agreed on a standard set of future dates for projections, as well as a minimum spatial resolution. Greater convergence in these fundamental aspects would also be advantageous when evaluating the likely future distributions of a number of other diseases of public health concern, including other vector-borne and climate-dependent infectious

Epidemic potential
The reciprocal of the critical density threshold (that is, the average number of adult female mosquitoes per person required for one infectious human case of dengue to give rise to a new one in a susceptible human population). A greater epidemic potential in a location indicates that the climate conditions in that location are such that fewer vectors are needed to effectively spread a vector-borne disease such as dengue.

General circulation models
Mathematical models of the general circulation of the atmosphere or ocean, which constitute an important component of most global climate models. These models can be applied at a variety of temporal scales and used to project climate conditions up to 100 years in the future.

Logistic regression
A probabilistic statistical model that is used to predict a binary response (for example, presence versus absence of a disease) based on a linear combination of hypothesized predictor variables or covariates.

Annual mean vapour pressure
In meteorology, this refers to the partial pressure of water vapour in the atmosphere as measured (or estimated) and averaged over 1 year. It has been used as a measure of humidity in dengue modelling.

Parsimonious model selection
When the goodness of fit of a statistical model is weighted against its complexity when choosing a final model. More-parsimonious models are less complex, with fewer covariates chosen for inclusion. This makes them generally more interpretable and more straightforward to extrapolate into different environments.

© 2015 Macmillan Publishers Limited. All rights reserved

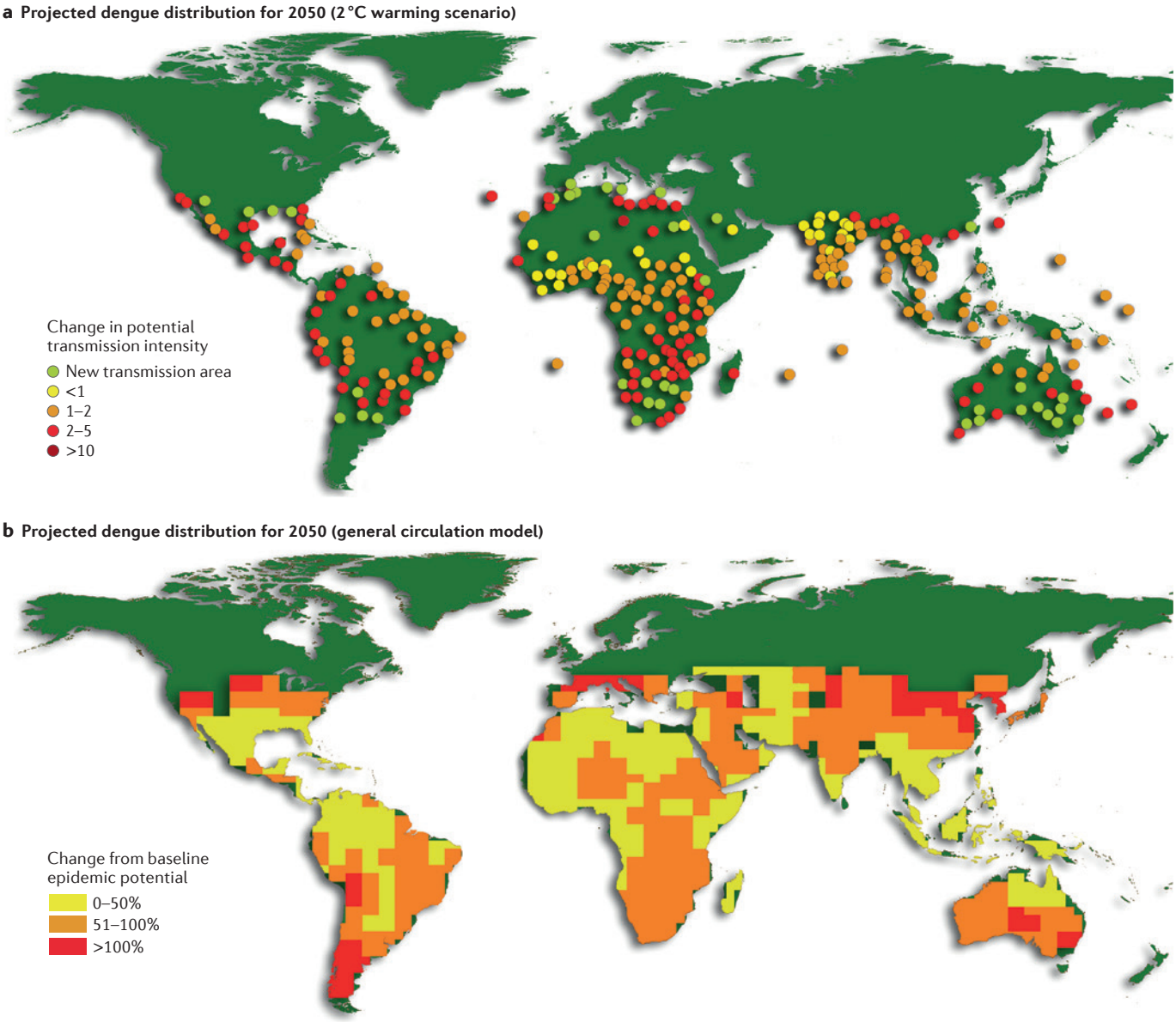


Figure 1 | **Mechanistic models to project the global distribution of dengue.** **a** | Projection of the potential distribution of dengue transmission in 2050 under a 2 °C warming scenario is shown. Each point represents a meteorological station, with colours representing the expected magnitude of change in potential transmission intensity from baseline (1931–1960) temperature data[65]. **b** | Projection of the potential distribution of dengue transmission in 2050 based on a general circulation model (GFLD89 GCM) is shown. Each 250 km × 250 km cell is coloured according to the expected change in epidemic potential from baseline (1931–1980) temperature data[63]. Note that the terms potential transmission intensity and epidemic potential both refer to the same measurement. Part **a** reproduced with permission from REF. 65, American Society of Tropical Medicine and Hygiene. Part **b** reproduced from REF. 63.

Generalized additive models
Statistical models relating a response variable to a set of covariates, modelling the response as the sum of nonlinear relationships with different covariates. The response variable is assigned a specific distribution and a link function used to relate it to the sum of covariate relationships, as in a generalized linear model.

diseases. So far, this type of standardization of modelling elements has not been performed for any disease. Below, we aim to identify the elements that can be developed further for dengue and other vector-borne diseases, including the general modelling approach, the quality and quantity of contemporary disease information, the mitigation of bias and the comprehensiveness of modelling covariates.

Modelling approaches. It is important to consider that if precise and globally representative data were available on all of the key processes influencing transmission, a mechanistic modelling approach may be the most suitable method for mapping an infectious disease. Unfortunately, few studies have identified physiological relationships between dengue transmission and factors other than temperature, and even these existing studies are highly specific to local conditions; this explains why the mechanistic models for dengue expansion[63–65] were carried out purely as a function of temperature.

Numerous other factors are important drivers of dengue virus transmission across multiple scales, but they cannot realistically be linked to dengue without

© 2015 Macmillan Publishers Limited. All rights reserved

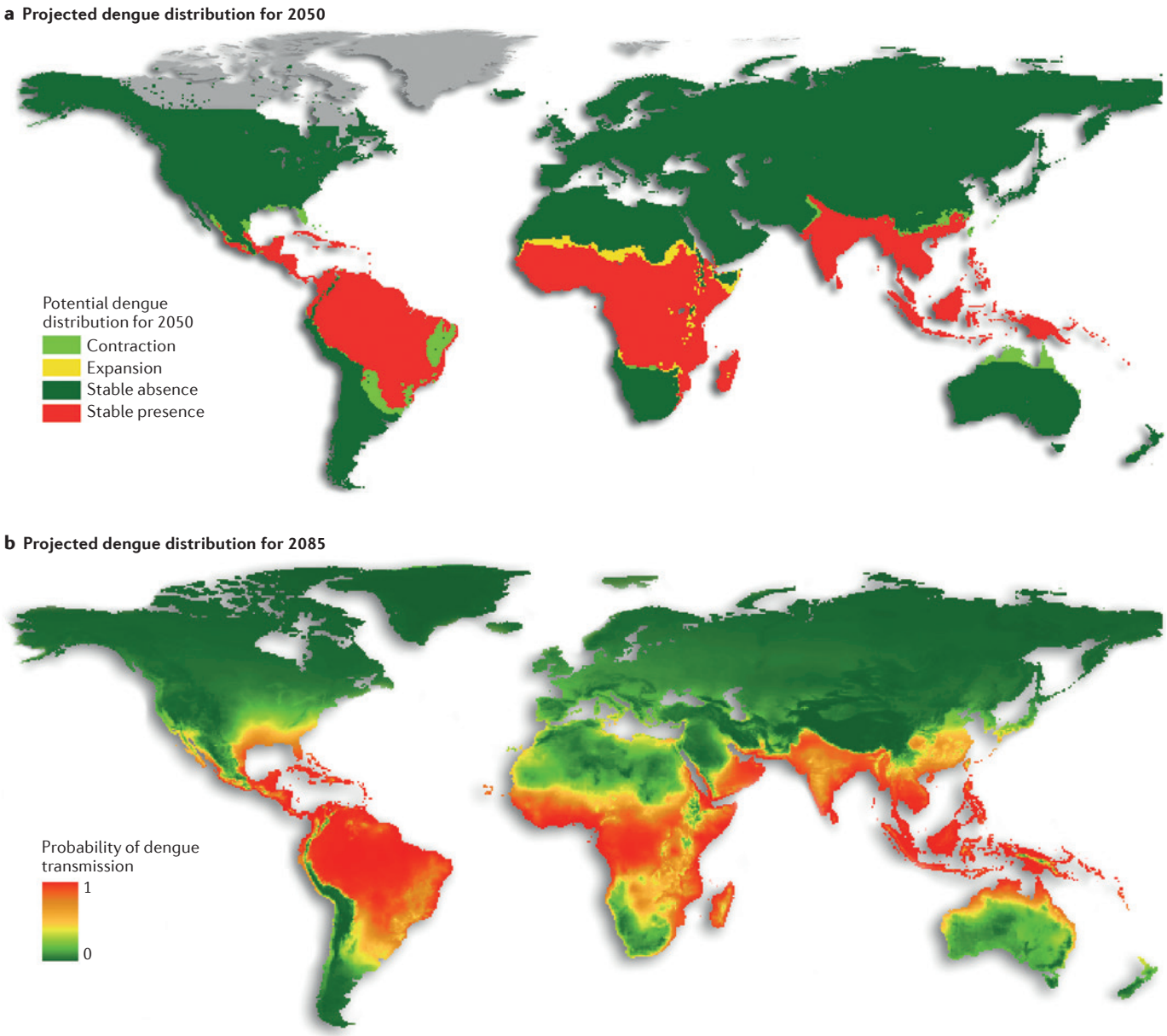


Figure 2 | **Statistical models to project the global distribution of dengue.** **a** | Projection of the potential distribution of dengue transmission in 2050 is shown, which accounts for the combined effect of climate change and gross domestic product (GDP) per capita, and is based on all five climate change projection models. In red areas, the models indicate that dengue transmission occurs in the present and is expected to persist through to 2050. Dark green areas currently do not have dengue transmission, nor are they expected to in the future in these models. The yellow and light green areas represent places where dengue can be expected to expand or contract from its current distribution, respectively. **b** | Projection of the potential distribution of dengue transmission in 2085 based on climate projections for 2080–2100 from a general circulation model (CCGCMA2) of vapour pressure (humidity) is shown. Green areas on the map represent places where the model projects very little or no dengue transmission, whereas in red areas, the model projects that dengue is almost certain to occur. Part **a** reproduced with permission from REF. 66, with kind permission from Springer Science and Business Media. Part **b** reprinted from *Lancet*, **360**, Hales, S., de Wet, N., Maindonald, J. & Woodward, A., Potential effect of population and climate changes on global distribution of dengue fever: an empirical model, 830–834, Copyright (2002), with permission from Elsevier.

Gross domestic product (GDP) per capita
The value of all final goods and services produced within a given year in a country, divided by the average or mid-year population in the country for that year. This metric is often used as an indicator of the overall standard of living in a country, but it does not convey variation in this standard of living across populations or locations within the country.

empirical observations and statistical inference. Several studies have explored these empirical relationships at local or national levels, implicating not only other meteorological drivers of vector abundance in dengue virus transmission (such as precipitation[79–87] and humidity[28,73,85,88–91]), but also physical environmental factors (including vegetation[29,92–94] and moisture availability[95–98]), human environmental factors (such as urbanization[99] and population density[100–104]) and societal factors (such as socio-economic status[52,54,105,106–109] and

© 2015 Macmillan Publishers Limited. All rights reserved

household water storage practices[110–114]). Importantly, not all of these variables can be measured in a globally systematic manner or projected with a high degree of confidence, but their potential ability to add strength to the resulting projections, both individually and through their interactions, might outweigh the uncertainties they carry with them.

Considering the above, the best compromise could be to carry out a statistical model based on a comprehensive and high-quality disease occurrence database (see below) that also includes biologically relevant variables derived from the more well-defined mechanistic relationships. For example, temperature suitability metrics have been computed for both malaria and dengue modelling[41,71,115] by using a mathematical model of the temperature-driven components of vectorial capacity. This type of metric can be used as just one of many covariates in a statistical model and offers an improvement over conventional climate averages. Future measures for temperature suitability can also be calculated by replacing current temperature data with projected temperature estimates.

Quality and quantity of contemporary disease data. Clearly, if a model cannot predict the present accurately, future projections based on it are inherently questionable. The ideal database representing the contemporary global distribution of dengue would comprise spatially explicit information from population-representative disease prevalence surveys distributed across the entire range of the disease and would incorporate information about both asymptomatic and symptomatic infections. This would enable geostatistical mapping of disease intensity along with uncertainty estimates (for example, using Bayesian inference methods, such as has been conducted for malaria[58,59]). However, for dengue and for many neglected tropical diseases, this type of information is rare. Instead, data giving the locations of reported cases provide the best evidence for the geographical distribution of the disease. Often, these data are geographically imprecise or sparse, with no information about disease prevalence in the population or the magnitude of the outbreak. Therefore, environmental modelling approaches are the most appropriate technique for predictive mapping[116] until a greater number of dengue incidence and prevalence studies have been carried out.

The two statistical studies reviewed here[66,67] used a data set based on national- or subnational-level information on the presence or absence of dengue that was transformed to match the resolution of finer-scale covariate data sets. Neither study defined a mechanism for distributing the national-level data at the finer scale at which they carried out their analysis. This can introduce bias into the modelling process by disregarding spatial variations in dengue occurrence within countries and assuming dengue presence to be uniform across large areas[117]. As a result, inaccurate relationships may have been defined between disease presence and the underlying covariates in the statistical models. When compounded with other sources of modelling bias, this becomes particularly problematic and hinders meaningful interpretation of the resulting projections.

Considering these limitations and in the absence of population prevalence information, an exhaustive and high-resolution database of dengue occurrence locations would be the most preferable alternative. Importantly, an extensive dengue occurrence database for the years 1960 to 2012 (REF. 118) — compiled from published literature, case reports and informal online sources — is now freely available. Even though this type of database still lacks robust information on where the disease is absent, as lack of reporting does not mean true absence, the greater geographical precision of the disease occurrence records can improve the accuracy of correlations found with the set of underlying covariates. This database has been used in past work to model the contemporary global distribution of dengue[41] and can be updated with new information as necessary.

As the reporting rate and geographical precision of occurrence data for dengue and many other diseases vary greatly by country, a combination of three approaches can be used to ameliorate reporting bias. First, one can create a background measure of certainty on disease status, termed evidence consensus in previous work[41,42], which is quantified by taking into account the reliability and amount of evidence for disease presence in a region from a variety of sources. This can be used in pseudo-data generation procedures common to species distribution models[119–121] and, more recently, dengue models[1], as well as to limit the geographical extents within which baseline predictions are made. Second, cross-validation or systematic removal and testing of a different subset of the data can determine prediction ability and model fit[67,68,122]. The results of many models containing alternative sets of pseudo-data points can also be averaged to generate uncertainty estimates alongside risk predictions. These uncertainty estimates can be propagated through to projections and combined with the uncertainty associated with projections of underlying covariates. In doing so, it is possible to provide more accurate projections of the future distribution of a disease that include appropriate caveats.

Comprehensive set of modelling covariates. Overall, the existing dengue distribution projections that resulted from mechanistic models based only on temperature-dependent biological relationships demonstrate a greater number of incongruities with actual dengue occurrence and an overall weaker ability to communicate current and future dengue risk. These models are better described as characterizing temperature suitability for dengue transmission, rather than actual dengue transmission. If the resulting maps are to depict more plausible distributions for dengue (both present and future), it is essential to incorporate a range of human and meteorological variables into the modelling process. This means considering not only the biology of the virus and its vector, but also other important meteorological and non-meteorological drivers of transmission between the vector and the host. For example, the most recent statistical study[66] was the first to consider economic information when projecting the future global distribution of dengue. However, other important factors, such as patterns of human movement

Uncertainty estimates
A measure of how uncertain each prediction from a model is. This acts as a quantitative estimate of how well the model is able to make predictions. When predictions are made across a spatial grid, uncertainty estimates can be made for each cell. This enables a map of uncertainty to be produced, showing places where the model performs relatively better or worse.

© 2015 Macmillan Publishers Limited. All rights reserved

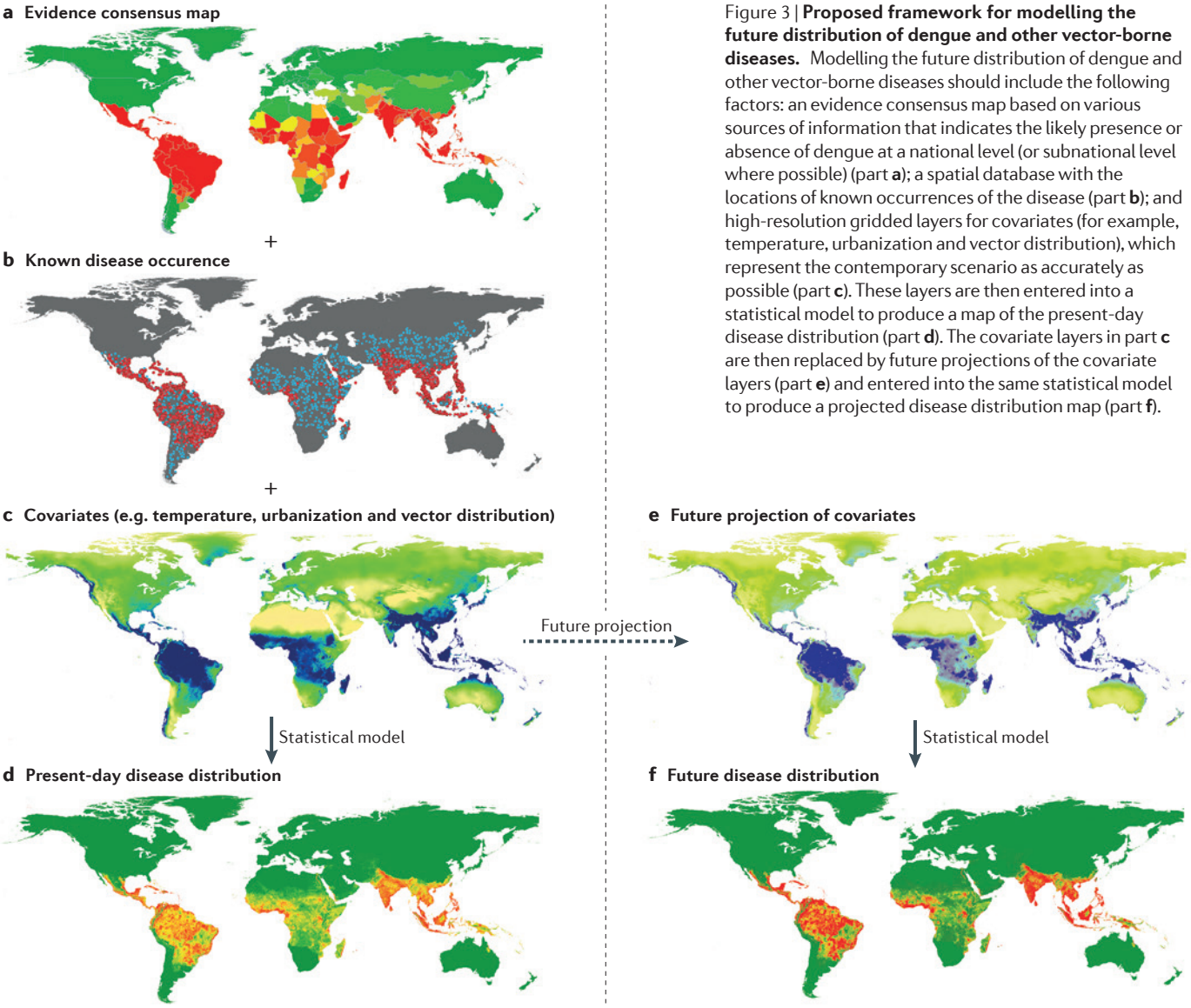


Figure 3 | **Proposed framework for modelling the future distribution of dengue and other vector-borne diseases.** Modelling the future distribution of dengue and other vector-borne diseases should include the following factors: an evidence consensus map based on various sources of information that indicates the likely presence or absence of dengue at a national level (or subnational level where possible) (part **a**); a spatial database with the locations of known occurrences of the disease (part **b**); and high-resolution gridded layers for covariates (for example, temperature, urbanization and vector distribution), which represent the contemporary scenario as accurately as possible (part **c**). These layers are then entered into a statistical model to produce a map of the present-day disease distribution (part **d**). The covariate layers in part **c** are then replaced by future projections of the covariate layers (part **e**) and entered into the same statistical model to produce a projected disease distribution map (part **f**).

or changes in urban environments, have not been considered in any global dengue distribution projections; these may be important correlates of dengue transmission alongside economic information. For example, with optimal weather conditions, densely populated urban and peri-urban slums with reduced water supply and inadequate waste management infrastructure are more likely to contain development sites for *Ae. aegypti* mosquitoes in close proximity to an abundance of human hosts.

Without an established competent vector, there can be no dengue transmission, and baseline vector distributions must be accurately characterized to determine true transmission potential. Even though the behaviour and ecology of *Ae. aegypti* have been studied in various settings, population measures for *Aedes* species have not been standardized. Moreover, detailed global distributions of dengue vectors have yet to be formally defined, and the relationship between *Aedes* species abundance and risk of human dengue virus infection has not been delineated. Standardized methods for modelling these distributions would therefore be useful, enabling maps of *Aedes* species to be included as modelling covariates for mapping dengue distribution. So far, temperature has been included as an indirect indicator of the distribution and virus transmission potential of *Aedes* species, but high spatial-resolution estimates of *Aedes* species distributions would improve the plausibility of both contemporary and projected maps by being incorporated either as covariates or as layers of information that are capable of limiting the extent of baseline risk predictions.

Furthermore, it is important to note that future environmental suitability for dengue vectors does not necessarily equate to their certain future presence. Non-environmental factors such as human movement and shipping practices[123] will also have an important role in determining whether the vectors actually spread to areas suitable for them to survive and be competent vectors. Credible maps of the future distribution of *Aedes* species will therefore require not only projections of

© 2015 Macmillan Publishers Limited. All rights reserved

environmental suitability for the vector but also probabilistic models of their geographical spread. Although the studies we have reviewed here based their projections on the biology of *Ae. aegypti*, the expansion of this species into cooler regions is not as likely as for *Ae. albopictus* owing to the intolerance of *Ae. aegypti* eggs to temperate winters[124]. As *Ae. albopictus* has a longer lifespan and greater tolerance of cooler temperatures, this vector may become increasingly important for dengue virus transmission in temperate areas[35,72,125]. Therefore, separate distribution models for *Ae. aegypti* and *Ae. albopictus* would also be useful in making projections for the future distribution of dengue.

The need for scientific consensus

Several efforts have been made to project the future distribution of dengue, but a clear consensus is lacking regarding the parts of the world where transmission is expected to expand or intensify, or where it might contract or diminish. Although locally and temporally specific environmental variations will undoubtedly have a role in determining future changes in the spatial distribution of dengue, global-scale phenomena such as climate change, economic development and demographic shifts will be the key underlying drivers of its expansion into new geographical regions and intensification within areas that are already endemic. Although not every potential advancement or shift in these realms can be foreseen, global-scale dengue distribution projections are a valuable exercise, enabling policymakers to plan future public health interventions. Despite the inherent difficulty in making future projections, quantitative estimates of uncertainty in these predictions have yet to be provided alongside global dengue distribution projections. However, it is possible to provide useful projections (which incorporate appropriate caveats) of future change in the geographical distribution of dengue and other vector-borne diseases (FIG. 3).

Public health policymakers, vaccine developers and vector control agencies require a range of robust estimates for the future global distribution of dengue. Although modelling the distribution of dengue should be encouraged across a range of projection dates, future scenarios, and spatial scales and resolutions, for global estimates to be comparable across individual efforts, certain guidelines should be recommended by agencies involved in global public health policy such as the WHO, the US Centers for Disease Control and the European Centre for Disease Control. These guidelines must be based on the real needs of international and national health policymakers, and we suggest that these guidelines take into consideration several key factors.

Our first suggestion is to harmonize projection dates. It may be most practical to project to the years 2020, 2050 and 2080, as these 30-year intervals are commonly used for making climate projections based on the International Panel on Climate Change (IPCC) Special Report on Emissions Scenarios[126]. The IPCC report presents four main scenarios based on different combinations of potential emissions and economic changes; within these scenarios, there are a multitude of general circulation models, which result from different methods for modelling specific processes and feedbacks. Therefore, our second suggestion is to summarize meteorological variables across all general circulation models within each of the main four scenarios, allowing averages and credible intervals for these variables to be incorporated into overall dengue distribution projections and accounted for in uncertainty estimates. Finally, because gridded geographical data on population density and the extents of urban areas are currently available at a 1 km × 1 km resolution (for example, data sets from the Socioeconomic Data and Applications Center), this could serve as a minimum spatial resolution for dengue distribution projections, unless they are being made for specific policy purposes that only require results at a coarser scale. Should subsequent studies show that the spatial scale of dengue virus transmission is finer than 1 km × 1 km and finer-scale data sets for all covariates of interest become available, it would be possible to increase this minimum spatial resolution.

Projecting the geographical distribution of dengue and other vector-borne diseases should also require a set of minimum criteria that include: comprehensive data representing the current distribution of the disease; a model proven to accurately predict this distribution; projections for factors that are important in this model when extrapolating it to the future; and quantification of uncertainty introduced by assumptions inherent to these criteria. In this manner, the evidence consensus, disease occurrence and pseudo-data, and contemporary and projected covariate layers can be used to project disease distribution at some future time point (FIG. 3).

Although the focus of this Review has been on future projections of dengue distribution, the modelling requirements we have proposed can be applied to other vector-borne diseases, the distribution of which is also dependent on both the natural and human-built environments. Collective recognition of certain minimum criteria, along with adherence to specific guidelines regarding projection dates and spatial resolution, would facilitate the scientific consensus on the probable geographical spread of certain vector-borne diseases, as well as the rate at which this spread may occur and the size of the populations that will be at risk. Owing to the substantial public health threat already posed by dengue and many other vector-borne diseases, it is crucial to convey this scientific consensus to public health policymakers, whose efforts will play an important part in shaping the future landscapes of vector-borne diseases, interventions and their associated burden on human populations.

1. McMichael, A. J., Woodruff, R. E. & Hales, S. Climate change and human health: present and future risks. *Lancet* **367**, 859–869 (2006).
2. Rogers, D. J. & Randolph, S. E. Climate change and vector-borne diseases. *Adv. Parasitol.* **62**, 345–381 (2006).
3. Kovats, R. S., Campbell-Lendrum, D. H., McMichael, A. J., Woodward, A. & Cox, J. S. Early effects of climate change: do they include changes in vector-borne disease? *Phil. Trans. R. Soc. Lond. B* **356**, 1057–1068 (2001).
4. Tanser, F. C., Sharp, B. & le Sueur, D. Potential effect of climate change on malaria transmission in Africa. *Lancet* **362**, 1792–1798 (2003).
5. Patz, J. A. & Olson, S. H. Malaria risk and temperature: influences from global climate change and local land use practices. *Proc. Natl Acad. Sci. USA* **103**, 5635–5636 (2006).
6. van Lieshout, M., Kovats, R. S., Livermore, M. T. J. & Martens, P. Climate change and malaria: analysis of the SRES climate and socio-economic scenarios. *Glob. Environ. Change* **14**, 87–99 (2004).
7. Gething, P. W. *et al.* Climate change and the global malaria recession. *Nature* **465**, 342–345 (2010).

© 2015 Macmillan Publishers Limited. All rights reserved

8. Barclay, E. Is climate change affecting dengue in the Americas? *Lancet* **371**, 973–974 (2008).
9. Russell, R. C. *et al.* Dengue and climate change in Australia: predictions for the future should incorporate knowledge from the past. *Med. J. Aust.* **190**, 265–268 (2009).
10. Banu, S., Hu, W., Hurst, C. & Tong, S. Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors. *Trop. Med. Int. Health* **16**, 598–607 (2011).
11. Epstein, P. R. West Nile virus and the climate. *J. Urban Health* **78**, 367–371 (2001).
12. Paz, S. The West Nile Virus outbreak in Israel (2000) from a new perspective: the regional impact of climate change. *Int. J. Environ. Health Res.* **16**, 1–13 (2006).
13. Hsu, S. M., Yen, A. M. & Chen, T. H. The impact of climate on Japanese encephalitis. *Epidemiol. Infect.* **136**, 980–987 (2008).
14. Ogden, N. H. *et al.* Estimated effects of projected climate change on the basic reproductive number of the Lyme disease vector *Ixodes scapularis*. *Environ. Health Perspect.* **122**, 631–638 (2014).
15. Subak, S. Effects of climate on variability in Lyme disease incidence in the northeastern United States. *Am. J. Epidemiol.* **157**, 531–538 (2003).
16. Lindgren, E. & Gustafson, R. Tick-borne encephalitis in Sweden and climate change. *Lancet* **358**, 16–18 (2001).
17. Sumilo, D. *et al.* Climate change cannot explain the upsurge of tick-borne encephalitis in the Baltics. *PLoS ONE* **2**, e500 (2007).
18. Zeman, P. & Benes, C. A tick-borne encephalitis ceiling in Central Europe has moved upwards during the last 30 years: possible impact of global warming? *Int. J. Med. Microbiol.* **293**, 48–54 (2004).
19. Danielova, V. *et al.* Tick-borne encephalitis virus expansion to higher altitudes correlated with climate warming. *Int. J. Med. Microbiol.* **298**, 68–72 (2008).
20. Campbell-Lendrum, D., Bertollini, R., Neira, M., Ebi, K. & McMichael, A. Health and climate change: a roadmap for applied research. *Lancet* **373**, 1663–1665 (2009).
21. Colwell, R. R. *et al.* Climate change and human health. *Science* **279**, 968–969 (1998).
22. Frumkin, H. & McMichael, A. J. Climate change and public health: thinking, communicating, acting. *Am. J. Prev. Med.* **35**, 403–410 (2008).
23. Frumkin, H., McMichael, A. J. & Hess, J. J. Climate change and the health of the public. *Am. J. Prev. Med.* **35**, 401–402 (2008).
24. Haines, A. & McMichael, A. J. Climate change and health: implications for research, monitoring, and policy. *BMJ* **315**, 870–874 (1997).
25. Hales, S., Weinstein, P. & Woodward, A. Public health impacts of global climate change. *Rev. Environ. Health* **12**, 191–199 (1997).
26. World Health Organization. *Atlas of Health and Climate* (WHO, 2012).
27. Tabachnick, W. J. Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *J. Exp. Biol.* **213**, 946–954 (2010).
28. Mattar, S., Morales, V., Cassab, A. & Rodriguez-Morales, A. J. Effect of climate variables on dengue incidence in a tropical Caribbean municipality of Colombia, Cerete, 2003–2008. *Int. J. Infect. Dis.* **17**, e358–e359 (2013).
29. Sarfraz, M. S. *et al.* Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. *BMC Publ. Health* **12**, 853 (2012).
30. Gubler, D. J. The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? *Comp. Immunol. Microbiol. Infect. Dis.* **27**, 319–330 (2004).
31. Phillips, M. L. Dengue reborn: widespread resurgence of a resilient vector. *Environ. Health Perspect.* **116**, A382–A388 (2008).
32. Gubler, D. J. The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. *Ann. Acad. Med. Singapore* **27**, 227–234 (1998).
33. Seto, K. C., Guneralp, B. & Hutyra, L. R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl Acad. Sci. USA* **109**, 16083–16088 (2012).
34. Jansen, C. C. & Beebe, N. W. The dengue vector *Aedes aegypti:* what comes next. *Microbes Infect.* **12**, 272–279 (2010).
35. Rezza, G. *Aedes albopictus* and the reemergence of dengue. *BMC Publ. Health* **12**, 72 (2012).
36. Richards, S. L., Anderson, S. L. & Alto, B. W. Vector competence of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) for dengue virus in the Florida Keys. *J. Med. Entomol.* **49**, 942–946 (2012).
37. Lambrechts, L., Scott, T. W. & Gubler, D. J. Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. *PLoS Negl. Trop. Dis.* **4**, e646 (2010).
38. Wilson, M. E. in *Water and Sanitation Related Diseases and the Environment: Challenges, Interventions, and Preventive Measures* (ed. Selendy, J. M. H.) Ch. 31 (Wiley-Blackwell, 2011).
39. Tatem, A. J., Rogers, D. J. & Hay, S. I. Global transport networks and infectious disease spread. *Adv. Parasitol.* **62**, 293–343 (2006).
40. Messina, J. P. *et al.* Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol.* **22**, 138–146 (2014).
41. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
42. Brady, O. J. *et al.* Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* **6**, e1760 (2012).
43. Gubler, D. J. Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Trop. Med. Health* **39**, 3–11 (2011).
44. Murray, N. E., Quam, M. B. & Wilder-Smith, A. Epidemiology of dengue: past, present and future prospects. *Clin. Epidemiol.* **5**, 299–309 (2013).
45. Senior, K. Vector-borne diseases threaten Europe. *Lancet Infect. Dis.* **8**, 531–532 (2008).
46. Monath, T. P. Dengue: the risk to developed and developing countries. *Proc. Natl Acad. Sci. USA* **91**, 2395–2400 (1994).
47. Reiter, P. Climate change and mosquito-borne disease. *Environ. Health Perspect.* **109** (Suppl. 1), 141–161 (2001).
48. Erickson, R. A. *et al.* Potential impacts of climate change on the ecology of dengue and its mosquito vector the Asian tiger mosquito (*Aedes albopictus*). *Environ. Res. Lett.* **7**, 034003 (2012).
49. Erickson, R. A., Presley, S. M., Allen, L. J. S., Long, K. R. & Cox, S. B. A dengue model with a dynamic *Aedes albopictus* vector population. *Ecol. Model.* **221**, 2899–2908 (2010).
50. Ocampo, C. B., Mina, N. J., Carabali, M., Alexander, N. & Osorio, L. Reduction in dengue cases observed during mass control of *Aedes* in street catch basins in an endemic urban area in Colombia. *Acta Trop.* **132**, 15–22 (2014).
51. Soper, F. L. The 1964 status of *Aedes aegypti* eradication and yellow fever in the Americas. *Am. J. Trop. Med. Hyg.* **14**, 887–891 (1965).
52. Khormi, H. M. & Kumar, L. Assessing the risk for dengue fever based on socioeconomic and environmental variables in a geographical information system environment. *Geospat. Health* **6**, 171–176 (2012).
53. Khormi, H. M. & Kumar, L. Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing based case study. *Sci. Total Environ.* **409**, 4713–4719 (2011).
54. Mondini, A. & Chiaravalloti-Neto, F. Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a Brazilian city. *Sci. Total Environ.* **393**, 241–248 (2008).
55. Reiter, P. *et al.* Texas lifestyle limits transmission of dengue virus. *Emerg. Infect. Dis.* **9**, 86–89 (2003).
56. Ramos, M. M. *et al.* Epidemic dengue and dengue hemorrhagic fever at the Texas–Mexico border: results of a household-based seroepidemiologic survey, December 2005. *Am. J. Trop. Med. Hyg.* **78**, 364–369 (2008).
57. Williams, C. R., Bader, C. A., Kearney, M. R., Ritchie, S. A. & Russell, R. C. The extinction of dengue through natural vulnerability of its vectors. *PLoS Negl. Trop. Dis.* **4**, e922 (2010).
58. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6**, e1814 (2012).
59. Gething, P. W. *et al.* A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10**, 378 (2011).
60. Brooker, S., Kabatereine, N. B., Gyapong, J. O., Stothard, J. R. & Utzinger, J. Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology* **136**, 1707–1718 (2009).
61. Fichet-Calvet, E. & Rogers, D. J. Risk maps of Lassa fever in West Africa. *PLoS Negl. Trop. Dis.* **3**, e388 (2009).
62. Pigott, D. M. *et al.* Global distribution maps of the leishmaniases. *eLife* **3**, e02851 (2014).
63. Patz, J. A., Martens, W. J., Focks, D. A. & Jetten, T. H. Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environ. Health Perspect.* **106**, 147–153 (1998).
This study describes a mechanistic model that uses temperature projections from specific global climate models from the IPCC to project the global distribution of dengue in 2050.
64. Martens, W. J. M., Jetten, T. H. & Focks, D. A. Sensitivity of malaria, schistosomiasis and dengue to global warming. *Clim. Change* **35**, 145–156 (1997).
This study describes a mechanistic model that compares the results of projections for three vector-borne diseases based on climate change scenarios.
65. Jetten, T. H. & Focks, D. A. Potential changes in the distribution of dengue transmission under climate warming. *Am. J. Trop. Med. Hyg.* **57**, 285–297 (1997).
This study describes a mechanistic model that applies projected scenarios of 2 °C and 4 °C rises in temperature to present-day scenarios.
66. Astrom, C. *et al.* Potential distribution of dengue fever under scenarios of climate change and economic development. *Ecohealth* **9**, 448–454 (2012).
This study describes a statistical model that uses present-day and projected estimates for annual mean vapour pressure and GDP estimates to map baseline and future dengue transmission.
67. Hales, S., de Wet, N., Maindonald, J. & Woodward, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *Lancet* **360**, 830–834 (2002).
This study describes a statistical model that uses present-day and projected estimates for annual mean vapour pressure to map baseline and future dengue transmission.
68. Rogers, D. J. & Sedda, L. Statistical models for spatially explicit biological data. *Parasitology* **139**, 1852–1869 (2012).
69. Dormann, C. F. *et al.* Correlation and process in species distribution models: bridging a dichotomy. *J. Biogeogr.* **39**, 2119–2131 (2012).
70. Chan, M. & Johansson, M. A. The incubation periods of dengue viruses. *PLoS ONE* **7**, e50972 (2012).
71. Brady, O. J. *et al.* Modelling adult *Aedes aegypti* and *Aedes albopictus* survival at different temperatures in laboratory and field settings. *Parasit. Vectors* **6**, 351 (2013).
72. Brady, O. J. *et al.* Global temperature constraints on *Aedes aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasit. Vectors* **7**, 338 (2014).
73. Waldock, J. *et al.* The role of environmental variables on *Aedes albopictus* biology and chikungunya epidemiology. *Pathog. Glob. Health* **107**, 224–241 (2013).
74. Focks, D. A., Haile, D. G., Daniels, E. & Mount, G. A. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *J. Med. Entomol.* **30**, 1003–1017 (1993).
75. Kearney, M., Porter, W. P., Williams, C., Ritchie, S. & Hoffmann, A. A. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Funct. Ecol.* **23**, 528–538 (2009).
76. Smith, D. L. *et al.* Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathog.* **8**, e1002588 (2012).
77. Smith, D. L. *et al.* Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Trans. R. Soc. Trop. Med. Hyg.* **108**, 185–197 (2014).
78. Gloria-Soria, A., Brown, J. E., Kramer, V., Hardstone Yoshimizu, M. & Powell, J. R. Origin of the dengue fever mosquito, *Aedes aegypti*, in California. *PLoS Negl. Trop. Dis.* **8**, e3029 (2014).
79. Nagao, Y. & Koelle, K. Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. *Proc. Natl Acad. Sci. USA* **105**, 2238–2243 (2008).
80. Chowell, G., Cazelles, B., Broutin, H. & Munayco, C. V. The influence of geographic and climate factors on the timing of dengue epidemics in Peru, 1994–2008. *BMC Infect. Dis.* **11**, 164 (2011).
81. Jury, M. R. Climate influence on dengue epidemics in Puerto Rico. *Int. J. Environ. Health Res.* **18**, 323–334 (2008).
82. Johansson, M. A., Dominici, F. & Glass, G. E. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl. Trop. Dis.* **3**, e382 (2009).

© 2015 Macmillan Publishers Limited. All rights reserved

83. Hii, Y. L. *et al.* Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Glob Health Action* http://dx.doi.org/10.3402/gha.v2i0.2036 (2009).
84. Chen, M. J. *et al.* Effects of extreme precipitation to the distribution of infectious diseases in Taiwan, 1994–2008. *PLoS ONE* **7**, e34651 (2012).
85. Sankari, T., Hoti, S. L., Singh, T. B. & Shanmugavel, J. Outbreak of dengue virus serotype 2 (DENV-2) of Cambodian origin in Manipur, India — association with meteorological factors. *Indian J. Med. Res.* **136**, 649–655 (2012).
86. Descloux, E. *et al.* Climate-based models for understanding and forecasting dengue epidemics. *PLoS Negl. Trop. Dis.* **6**, e1470 (2012).
87. Chandy, S., Ramanathan, K., Manoharan, A., Mathai, D. & Baruah, K. Assessing effect of climate on the incidence of dengue in Tamil Nadu. *Indian J. Med. Microbiol.* **31**, 283–286 (2013).
88. Yasuoka, J. & Levins, R. Ecology of vector mosquitoes in Sri Lanka: suggestions for future mosquito control in rice ecosystems. *Southeast Asian J. Trop. Med. Publ. Health* **38**, 646–657 (2007).
89. Lu, L. *et al.* Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Publ. Health* **9**, 395 (2009).
90. Yang, T. C. *et al.* Epidemiology and vector efficiency during a dengue fever outbreak in Cixi, Zhejiang Province, China. *J. Vector Ecol.* **34**, 148–154 (2009).
91. Gharbi, M. *et al.* Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infect. Dis.* **11**, 166 (2011).
92. Kamgang, B. *et al.* Geographic and ecological distribution of the dengue and chikungunya virus vectors *Aedes aegypti* and *Aedes albopictus* in three major Cameroonian towns. *Med. Vet. Entomol.* **24**, 132–141 (2010).
93. Troyo, A., Fuller, D. O., Calderon-Arguedas, O., Solano, M. E. & Beier, J. C. Urban structure and dengue fever in Puntarenas, Costa Rica. *Singap. J. Trop. Geogr.* **30**, 265–282 (2009).
94. Smith, J., Amador, M. & Barrera, R. Seasonal and habitat effects on dengue and West Nile virus vectors in San Juan, Puerto Rico. *J. Am. Mosq. Control Assoc.* **25**, 38–46 (2009).
95. Estallo, E. L. *et al.* Effectiveness of normalized difference water index in modelling *Aedes aegypti* house index. *Int. J. Remote Sens.* **33**, 4254–4265 (2012).
96. Tourre, Y. M., Jarlan, L., Lacaux, J. P., Rotela, C. H. & Lafaye, M. Spatio-temporal variability of NDVI-precipitation over southernmost South America: possible linkages between climate signals and epidemics. *Environ. Res. Lett.* **3**, 044008 (2008).
97. Schreiber, K. V. An investigation of relationships between climate and dengue using a water budgeting technique. *Int. J. Biometeorol.* **45**, 81–89 (2001).
98. Cheng, S. Q., Kalkstein, L. S., Focks, D. A. & Nnaji, A. New procedures to estimate water temperatures and water depths for application in climate-dengue modeling. *J. Med. Entomol.* **35**, 646–652 (1998).
99. Rios-Velasquez, C. M. *et al.* Distribution of dengue vectors in neighborhoods with different urbanization types of Manaus, state of Amazonas, Brazil. *Mem. Inst. Oswaldo Cruz* **102**, 617–623 (2007).
100. Padmanabha, H., Durham, D., Correa, F., Diuk-Wasser, M. & Galvani, A. The interactive roles of *Aedes aegypti* super-production and human density in dengue transmission. *PLoS Negl. Trop. Dis.* **6**, e1799 (2012).
101. Hsueh, Y. H., Lee, J. & Beltz, L. Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008. *Appl. Geogr.* **34**, 587–594 (2012).
102. Hassan, H., Shohaimi, S. & Hashim, N. R. Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospat. Health* **7**, 21–25 (2012).
103. Schmidt, W. P. *et al.* Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. *PLoS Med.* **8**, e1001082 (2011).
104. Lin, C. H. & Wen, T. H. Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *Int. J. Environ. Res. Publ. Health* **8**, 2798–2815 (2011).
105. Teixeira, T. R. D. & Medronho, R. D. Socio-demographic factors and the dengue fever epidemic in 2002 in the state of Rio de Janeiro, Brazil. *Cad. Saude Publ.* **24**, 2160–2170 (in Portuguese) (2008).
106. Kienberger, S., Hagenlocher, M., Delmelle, E. & Casas, I. A. WebGIS tool for visualizing and exploring socioeconomic vulnerability to dengue fever in Cali, Colombia. *Geospat. Health* **8**, 313–316 (2013).
107. Hagenlocher, M., Delmelle, E., Casas, I. & Kienberger, S. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical versus expert-based modeling. *Int. J. Health Geogr.* **12**, 36 (2013).
108. Syed, M. *et al.* Knowledge, attitudes and practices regarding dengue fever among adults of high and low socioeconomic groups. *J. Pak. Med. Assoc.* **60**, 243–247 (2010).
109. Almeida, M. C. D., Caiaffa, W. T., Assuncao, R. M. & Proietti, F. A. Spatial vulnerability to dengue in a Brazilian urban area during a 7-year surveillance. *J. Urban Health* **84**, 334–345 (2007).
110. Banerjee, S., Aditya, G. & Saha, G. K. Household disposables as breeding habitats of dengue vectors: linking wastes and public health. *Waste Manag.* **33**, 233–239 (2013).
111. Caprara, A. *et al.* Irregular water supply, household usage and dengue: a bio-social study in the Brazilian northeast. *Cad. Saude Publ.* **25**, S125–S136 (2009).
112. Tran, H. P. *et al.* Householder perspectives and preferences on water storage and use, with reference to dengue, in the Mekong Delta, southern Vietnam. *Int. Health* **2**, 136–142 (2010).
113. Padmanabha, H., Soto, E., Mosquera, M., Lord, C. C. & Lounibos, L. P. Ecological links between water storage behaviors and *Aedes aegypti* production: implications for dengue vector control in variable climates. *Ecohealth* **7**, 78–90 (2010).
114. Kusumawathie, P. H., Yapabandarab, A. M., Jayasooriya, G. A. & Walisinghe, C. Effectiveness of net covers on water storage tanks for the control of dengue vectors in Sri Lanka. *J. Vector Borne Dis.* **46**, 160–163 (2009).
115. Gething, P. W. *et al.* Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasit. Vectors* **4**, 92 (2011).
116. Hay, S. I. *et al.* Global mapping of infectious disease. *Phil. Trans. R. Soc. B* **368**, 20120250 (2013).
117. Gotway, C. A. & Young, L. J. Combining incompatible spatial data. *J. Am. Stat. Assoc.* **97**, 632–648 (2002).
118. Messina, J. P. *et al.* A global compendium of human dengue virus occurrence. *Sci. Data* **1**, 140004 (2014).
119. Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* **3**, 327–338 (2012).
120. Phillips, S. J. *et al.* Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).
121. Pearce, J. L. & Boyce, M. S. Modelling distribution and abundance with presence-only data. *J. Appl. Ecol.* **43**, 405–412 (2006).
122. Hijmans, R. J. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93**, 679–688 (2012).
123. Seebens, H., Gastner, M. T. & Blasius, B. The risk of marine bioinvasion caused by global shipping. *Ecol. Lett.* **16**, 782–790 (2013).
124. Gould, E. A. & Higgs, S. Impact of climate change and other factors on emerging arbovirus diseases. *Trans. R. Soc. Trop. Med. Hyg.* **103**, 109–121 (2009).
125. Weaver, S. C. & Reisen, W. K. Present and future arboviral threats. *Antiviral Res.* **85**, 328–345 (2010).
126. Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).

Acknowledgments
J.P.M., G.R.W.W., T.W.S. and S.I.H. receive funding from, and O.J.B. acknowledges the support of, the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS; European Commission 7th Framework Programme (21803)). O.J.B. is funded by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship. D.M.P. is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford, UK. N.G. is funded by a grant from the Bill & Melinda Gates Foundation (OPP1053338). M.U.G.K. is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. T.W.S. acknowledges funding from the Bill & Melinda Gates Foundation (OPP52250), the Innovative Vector Control Consortium and the US National Institutes of Health (NIH; R01-AI069341, R01-AI091980, R01-GM08322 and P01-AI098670). S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (095066). T.W.S., D.L.S. and S.I.H. also acknowledge funding support from the Research and Policy for Infectious Disease Dynamics (RAPIDD) Program of the Science and Technology Directorate, the US Department of Homeland Security and the Fogarty International Center, NIH. Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All authors contributed to the writing of the paper.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
Socioeconomic Data and Applications Center:
http://sedac.ciesin.columbia.edu/data/sets/browse

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2015 Macmillan Publishers Limited. All rights reserved

eLIFE elifesciences.org

RESEARCH ARTICLE

Mapping the zoonotic niche of Ebola virus disease in Africa

David M Pigott[1†], Nick Golding[1†], Adrian Mylne[1], Zhi Huang[1], Andrew J Henry[1], Daniel J Weiss[1], Oliver J Brady[1], Moritz UG Kraemer[1], David L Smith[1,2], Catherine L Moyes[1], Samir Bhatt[1], Peter W Gething[1], Peter W Horby[3], Isaac I Bogoch[4,5], John S Brownstein[6,7], Sumiko R Mekaru[8], Andrew J Tatem[9,10,13], Kamran Khan[4,11], Simon I Hay[1,12]*

[1]Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom; [2]Sanaria Institute for Global Health and Tropical Medicine, Rockville, United States; [3]Epidemic Diseases Research Group, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom; [4]Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Canada; [5]Divisions of Internal Medicine and Infectious Diseases, University Health Network, Toronto, Toronto, Canada; [6]Department of Pediatrics, Harvard Medical School, Boston, United States; [7]Children's Hospital Informatics Program, Boston Children's Hospital, Boston, United States; [8]Children's Hospital Informatics Program, Boston Children's Hospital, Boston, United States; [9]Department of Geography and Environment, University of Southampton, Southampton, United Kingdom; [10]Fogarty International Center, National Institutes of Health, Bethesda, United States; [11]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada; [12]Fogarty International Center, National Institutes of Health, Bethseda, United States; [13]Flowminder Foundation, Stockholm, Sweden

***For correspondence:** simon.hay@zoo.ox.ac.uk

†These authors contributed equally to this work

Competing interests: See page 20

Funding: See page 21

Received: 18 August 2014
Accepted: 31 August 2014
Published: 08 September 2014

Reviewing editor: Prabhat Jha, University of Toronto, Canada

Copyright Pigott et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract Ebola virus disease (EVD) is a complex zoonosis that is highly virulent in humans. The largest recorded outbreak of EVD is ongoing in West Africa, outside of its previously reported and predicted niche. We assembled location data on all recorded zoonotic transmission to humans and Ebola virus infection in bats and primates (1976–2014). Using species distribution models, these occurrence data were paired with environmental covariates to predict a zoonotic transmission niche covering 22 countries across Central and West Africa. Vegetation, elevation, temperature, evapotranspiration, and suspected reservoir bat distributions define this relationship. At-risk areas are inhabited by 22 million people; however, the rarity of human outbreaks emphasises the very low probability of transmission to humans. Increasing population sizes and international connectivity by air since the first detection of EVD in 1976 suggest that the dynamics of human-to-human secondary transmission in contemporary outbreaks will be very different to those of the past.
DOI: 10.7554/eLife.04395.001

Introduction

Ebola viruses have for the last forty years been responsible for a number of outbreaks of Ebola virus disease (EVD) in humans (***Pattyn et al., 1977***), with high case fatality rates typically around 60–70%, but potentially reaching as high as 90% (***Feldmann and Geisbert, 2011***). The most recent outbreak began in Guinea in December 2013 (***Baize et al., 2014***; ***Bausch and Schwarz, 2014***) and has subsequently spread to Liberia, Sierra Leone and Nigeria (***ECDC, 2014***). The unprecedented size and scale of this ongoing outbreak has the potential to destabilise already fragile economies and healthcare

eLife digest Since the first outbreaks of Ebola virus disease in 1976, there have been numerous other outbreaks in humans across Africa with fatality rates ranging from 50% to 90%. Humans can become infected with the Ebola virus after direct contact with blood or bodily fluids from an infected person or animal. The virus also infects and kills other primates—such as chimpanzees or gorillas—though Old World fruit bats are suspected to be the most likely carriers of the virus in the wild.

The largest recorded outbreak of Ebola virus disease is ongoing in West Africa: more people have been infected in this current outbreak than in all previous outbreaks combined. The current outbreak is also the first to occur in West Africa—which is outside the previously known range of the Ebola virus.

Pigott et al. have now updated predictions about where in Africa wild animals may harbour the virus and where the transmission of the virus from these animals to humans is possible. As such, the map identifies the regions that are most at risk of a future Ebola outbreak. The data behind these new maps include the locations of all recorded primary cases of Ebola in human populations—the 'index' cases—many of which have been linked to animal sources. The data also include the locations of recorded cases of Ebola virus infections in wild bats and primates from the last forty years. The maps, which were modelled using more flexible methods than previous predictions, also include new information—collected using satellites—about environmental factors and new predictions of the range of wild fruit bats.

Pigott et al. report that the transmission of Ebola virus from animals to humans is possible in 22 countries across Central and West Africa—and that 22 million people live in the areas at risk. However, outbreaks in human populations are rare and the likelihood of a human getting the disease from an infected animal still remains very low. The updated map does not include data about how infections spread from one person to another, so the next challenge is to use existing data on human-to-human transmission to better understand the likely size and extent of current and future outbreaks. As more people live in, and travel to and from, the at-risk regions than ever before, Pigott et al. note that new outbreaks of Ebola virus disease are likely to be very different to those of the past.

DOI: 10.7554/eLife.04395.002

systems (***Fauci, 2014***), and fears of international spread of a Category A Priority Pathogen (***NIH, 2014***) have made this a massive focus for international public health (***Chan, 2014***). This has led to the current outbreak being declared a Public Health Emergency of International Concern on the 8 August 2014 (***Briand et al., 2014***; ***Gostin et al., 2014***; ***WHO, 2014d***).

The *Filoviridae*, of which *Ebolavirus* is a constituent genus, belong to the order *Mononegavirales*. Two other genera complete the family: *Marburgvirus*, itself responsible for a number of outbreaks of haemorrhagic fever across Africa (***Gear et al., 1975***; ***Conrad et al., 1978***; ***Smith et al., 1982***; ***Towner et al., 2006***) and *Cuevavirus*, recently isolated from bats in northern Spain (***Negredo et al., 2011***). Five species of *Ebolavirus* have been isolated to date (***Kuhn et al., 2010***; ***King et al., 2012***); the earliest recognised outbreaks of EVD were reported in Zaire (now the Democratic Republic of the Congo [DRC]) and Sudan in 1976 (***International Commission, 1978***; ***WHO International Study Team, 1978***). The causative viruses were isolated (***Pattyn et al., 1977***) and later identified to be distinct species, *Zaire ebolavirus* (EBOV) and *Sudan ebolavirus* (SUDV). A third species of *Ebolavirus*, *Reston ebolavirus*, was isolated from Cynomologus monkeys imported from the Philippines to a facility in the United States, where they experienced severe haemorrhaging (***Jahrling et al., 1990***). Whilst serological evidence of infection with this species has been reported in individuals in the Philippines (***Miranda et al., 1991***), no pathogenicity has been reported beyond primates and porcids (***Barrette et al., 2009***; ***Feldmann and Geisbert, 2011***). In 1994 a fourth species, *Tai Forest ebolavirus* was isolated from a veterinarian who had autopsied a chimpanzee in Côte d'Ivoire (***Le Guenno et al., 1995***), though the virus has not been detected subsequently. The final species, *Bundibugyo ebolavirus*, was responsible for an outbreak of EVD in Uganda in 2007 (***Towner et al., 2008***), as well as a more recent outbreak in the DRC (***WHO, 2012b***).

Initial analysis suggested that the viruses isolated from the current outbreak, originating in Guinea, formed a separate clade within the five *Ebolavirus* species (***Baize et al., 2014***). Subsequent re-analysis

of the same sequences however, indicated that these isolates instead nest within the *Zaire ebolavirus* lineage (**Dudas and Rambaut, 2014**), and diverged from Central Africa strains approximately ten years ago (**Gire et al., 2014**).

Which reservoir species are responsible for maintaining Ebola transmission between outbreaks is not well understood (**Peterson et al., 2004b**), but over the last decade significant progress has been made in narrowing down the list of likely hosts (**Peterson et al., 2007**) (**Figure 1**). Primates have long been known to harbour filoviral infections, with the first Marburg strains identified in African green monkeys in 1967 (**Siegert et al., 1967**; **Beer et al., 1999**). Significant mortality has also been reported in wild primate populations across Africa, most notably in gorilla (*Gorilla gorilla*) and chimpanzee (*Pan troglodytes*) populations (**Formenty et al., 1999**; **Rouquet et al., 2005**; **Bermejo et al., 2006**). The high case fatality rates recorded in the great apes combined with their declining populations and limited geographical range, indicate they are likely dead-end hosts for the virus and not reservoir species (**Groseth et al., 2007**). A large survey of small mammals in and around Gabon identified three species of bats which were infected with Ebola viruses—*Hypsignathus monstrosus*, *Epomops franqueti* and *Myonycetris torquata* (**Leroy et al., 2005**). Subsequent serological surveys (**Pourrut et al., 2009**; **Hayman et al., 2010**) and evidence linking the potential source of human outbreaks to bats (**Leroy et al., 2009**) lend support to the hypothesis of a bat reservoir. This, coupled with repeated detection of *Marburgvirus* in the fruit bat *Rousettus aegypticus* (**Towner et al., 2009**) and the only isolations of *Cuevavirus* also from bats (specifically *Llovia virus* [**Negredo et al., 2011**]), all support the suspicion that Chiroptera play an important role in the natural life-cycle of the filoviruses.

Humans represent a dead-end host for the virus, with only stuttering chains of transmission reported between humans in the majority of previous outbreaks (**Chowell et al., 2004**; **Legrand et al., 2007**) and no indication that humans can reintroduce the virus back into reservoir species (**Karesh et al., 2012**). The incubation period in humans ranges from two days to three weeks, after which a variety of clinical symptoms arise, affecting multiple organs of the body. At the peak of illness, haemorrhaging shock and widespread tissue damage can occur and can eventually lead to death within 6–16 days (**Feldmann and Geisbert, 2011**). Human-to-human transmission is mainly through direct unprotected contact with infected individuals and cadavers, with infectious particles detected in a number of different body fluids (**Feldmann and Geisbert, 2011**). The typical outbreak profile is defined by an index individual that has recently come into contact with the blood of another mammal through either hunting or the butchering of animal carcasses (**Pourrut et al., 2005**). Whilst it has been difficult to identify the zoonotic source for the index cases of some outbreaks, a recurring theme of hunting and handling bushmeat is suspected (**Table 1**; **Boumandouki et al., 2005**; **Nkoghe et al., 2005**, **2011**; **Leroy et al., 2009**). For some outbreaks, including the most recent, the initial source of zoonotic transmission has not been identified. In subsequent human-to-human transmission, the highest risk activities are those that bring humans into close contact with infected individuals. These include medical settings where insufficient infection control precautions have been taken, as well as home care and funeral preparations carried out by families or close friends (**Baron et al., 1983**; **Georges et al., 1999**; **Boumandouki et al., 2005**). As the conditions required for transmission are culturally and contextually dependent, opportunities for sustained transmission are highly heterogeneously distributed. Typically, chains of infection do not exceed three or four sequential transmission events, although occasionally (and particularly in

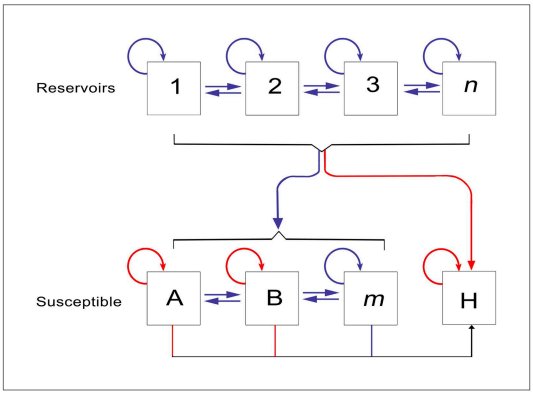


Figure 1. The epidemiology of Ebola virus transmission in Africa. Of the suspected reservoir species, 1, 2 and 3 represent the three bat species from which Ebola virus has been isolated (*Hypsignathus monstrosus*, *Myonycteris torquata* and *Epomops franqueti*) and *n* represents unknown reservoirs of the disease yet to be discovered. Of the susceptible species, A represents *Pan troglodytes*, B *Gorilla gorilla* and *m* represents other organisms susceptible to the disease, such as duikers. H represents humans. Blue arrows indicate unknown transmission cycles or infection routes and red arrow routes have been confirmed or are suspected. Adapted from **Groseth et al. (2007)**.

DOI: 10.7554/eLife.04395.003

Table 1. Locations of outbreaks of Ebola virus disease in humans

Outbreak	Countries	Date range	Location	Species	Reference
1	South Sudan	Jun–Nov 1976	Nzara	SUDV	(***WHO International Study Team, 1978***)
2	DRC	Sep–Oct 1976	Yambuku	EBOV	(***International Commission, 1978***)
3	DRC	Jun 1977	Bonduni	EBOV	(***Heymann et al., 1980***)
4	South Sudan	Jul–Oct 1979	Nzara	SUDV	(***Baron et al., 1983***)
5	Côte d'Ivoire	Nov 1994	Tai Forest	TAFV	(***Le Guenno et al., 1995***; ***Formenty et al., 1999***)
6	Gabon	Nov 1994–Feb 1995	Mekouka and Andock mining camps	EBOV	(***Amblard et al., 1997***; ***Georges et al., 1999***; ***Milleliri et al., 2004***)
7	DRC	Jan–Jul 1995	Mwembe Forest	EBOV	(***Muyembe and Kipasa, 1995***; ***Khan et al., 1999***)
8	Gabon	Jan–Mar 1996	Mayibout 2	EBOV	(***Georges et al., 1999***; ***Milleliri et al., 2004***)
9	Gabon	Jul 1996–Jan 1997	Booue	EBOV	(***Georges et al., 1999***; ***Milleliri et al., 2004***)
10	Uganda	Oct 2000–Feb 2001	Rwot-Obillo	SUDV	(***WHO, 2001***; ***Okware et al., 2002***; ***Lamunu et al., 2004***)
11	Gabon & ROC	Oct 2001–Mar 2002	Memdemba Entsiami, Abolo and Ambomi	EBOV	(WHO, 2003; ***Milleliri et al., 2004***; ***Nkoghe et al., 2005***; ***Pourrut et al., 2005***)
			Ekata		
			Oloba		
			Etakangaye		
			Grand Etoumbi		
12	ROC	Dec 2002–Apr 2003	Yembelangoye	EBOV	(WHO, 2003; ***Pourrut et al., 2005***)
			Nearby hunting camp		
			Mvoula		
13	ROC	Oct–Dec 2003	Mbandza	EBOV	(***Boumandouki et al., 2005***)
14	South Sudan	Apr–Jun 2004	Forests bordering Yambio	SUDV	(***WHO, 2005***; ***Onyango et al., 2007***)
15	ROC	Apr–May 2005	Odzala National Park	EBOV	(***Nkoghe et al., 2011***)
16	DRC	May–Nov 2007	Mombo Mounene 2 market	EBOV	(***Leroy et al., 2009***)
17	Uganda	Aug–Dec 2007	Kabango	BDBV	(***Towner et al., 2008***; ***MacNeil et al., 2010***; ***Wamala et al., 2010***)
18	DRC	Nov 2008–Feb 2009	Luebo	EBOV	(***Grard et al., 2011***)
19	Uganda	May 2011	Nakisamata	SUDV	(***Shoemaker et al., 2012***)
20	DRC	July–Nov 2012	Isiro	BDBV	(***CDC, 2014***; ***WHO, 2012b***)
21	Uganda	July–Oct 2012	Nyanswiga	SUDV	(***CDC, 2014***; ***WHO, 2012a***)
22	Uganda	Nov 2012–Jan 2013	Luwero District	SUDV	(***WHO, 2012c***; ***CDC, 2014***)
23	Guinea	Dec 2013 -	Meliandou	EBOV	(***Baize et al., 2014***; ***Bausch and Schwarz, 2014***)

DRC = Democratic Republic of the Congo, ROC = Republic of Congo.

DOI: 10.7554/eLife.04395.004

the early stages of infection) a single individual may be responsible for directly infecting a large number of others (**Brady et al., 2014**). In the outbreak in Gabon in 1996, a single person was responsible for infecting ten other individuals (**Milleliri et al., 2004**) whilst in the 1995 outbreak in the DRC, thirty five cases resulted from one individual (**Khan et al., 1999**). Secondary transmission can be restricted by effective case detection and isolation measures (**Shoemaker et al., 2012**; **WHO, 2014c**). Where this cannot be achieved, either due to a lack of infrastructure, poor understanding of the disease, or distrust of medical practices, secondary cases can continue to occur (**Khan et al., 1999**; **Larkin, 2003**; **Hewlett et al., 2005**). As the number of infections grows, the ability of healthcare systems to control the further spread diminishes and the risk of a large outbreak increases.

The recent outbreak in Guinea and surrounding countries indicate that the previous paradigm for Ebola outbreaks is shifting (**Briand et al., 2014**; **Chan, 2014**). The last 40 years of EVD outbreaks were accompanied by considerable changes in demographic patterns throughout Africa. There has been a large increase in population size coupled with increasing urbanisation (**Cohen, 2004**; **Seto et al., 2012**; **Linard et al., 2013**). African populations have also become better connected internally and internationally (**Linard et al., 2012**; **Huang and Tatem, 2013**). Only recently have we begun to understand the dynamic nature of these travel patterns (**Garcia et al., 2014**; **Gonzalez et al., 2008**; **Simini et al., 2012**; **Wesolowski et al., 2013**, **2012**) which have been clearly demonstrated to influence disease transmission over different temporal and spatial scales (**Hufnagel et al., 2004**; **Yang et al., 2008**; **Stoddard et al., 2009**; **Talbi et al., 2010**; **Brockmann and Helbing, 2013**; **Pindolia et al., 2014**). Changes in land use and penetration into previously remote areas of rainforest bring humans into contact with potential new reservoirs (**Daszak, 2000**), while changes in human mobility and connectivity will likely have profound impacts on the dispersion of Ebola cases during outbreaks. These conditions are thought to have a major role in setting the stage for the current outbreak.

This paper aims to define the areas suitable for zoonotic transmission of *Ebolavirus* (i.e., those routes defined in **Figure 1** excluding human-to-human transmission) through species distribution modelling techniques. The fundamental niche of a species can be conceptualised as the confluence of environmental conditions that support its presence in a particular location (**Franklin, 2009**). Species distribution models quantitatively describe this niche based on known occurrence records of the organism and their associated environmental conditions, enabling predictions of the likely geographic distribution of the species in other regions (**Elith and Leathwick, 2009**). The era of satellites and geographical information systems has made high resolution global data on environmental conditions increasingly available (**Hay et al., 2006**; **Weiss et al., 2014b**). Species distribution modelling using flexible machine learning approaches have been successfully applied to map the global distributions of disease vectors (**Sinka et al., 2012**) and pathogens such as dengue (**Bhatt et al., 2013**), influenza (**Gilbert et al., 2014**) and leishmaniasis (**Pigott et al., 2014**).

Previous studies applied the GARP (Genetic Algorithm for Rule-set Production) species distribution modelling approach (**Stockwell and Peters, 1999**) to the locations of 12 Ebola outbreaks in humans between 1976 and 2002 to map the likely distribution of Ebola viruses (**Peterson et al., 2004a**) and as a mechanism to identify potential reservoir hosts (**Peterson et al., 2004b**; **Peterson et al., 2007**). Here we update and improve the maps of the zoonotic transmission niche of EVD by: (i) incorporating more recent outbreak data from outside the formerly predicted niche of EVD; (ii) integrating for the first time data on outbreaks in primates and the occurrence of the virus in the suspected Old World fruit bat (OWFB) reservoirs; (iii) using new satellite-derived information on bespoke environmental covariates from Africa, including new distribution maps of the OWFB; and (iv) using new increasingly flexible niche mapping techniques in the modelling framework. To elucidate the relevance of these maps for transmission, we have also calculated the population at risk of primary spillover outbreaks from the zoonotic niche of EVD in Africa, and we investigated the changing nature of the populations within this niche.

Results

Reported EVD outbreaks

In total, 23 outbreaks of Ebola virus were identified in humans across Africa, consisting of a hypothesised 30 independent primary infection events (**Table 1**; **Figure 2**). These outbreaks span the last 40 years from the first outbreaks in 1976 to the five outbreaks that have occurred since 2010 (**Table 1**). The locations of the index cases span from West Africa, with the most westerly outbreak ongoing in Guinea, to Gabon, the Republic of Congo (ROC), the DRC, South Sudan and Uganda. Before December

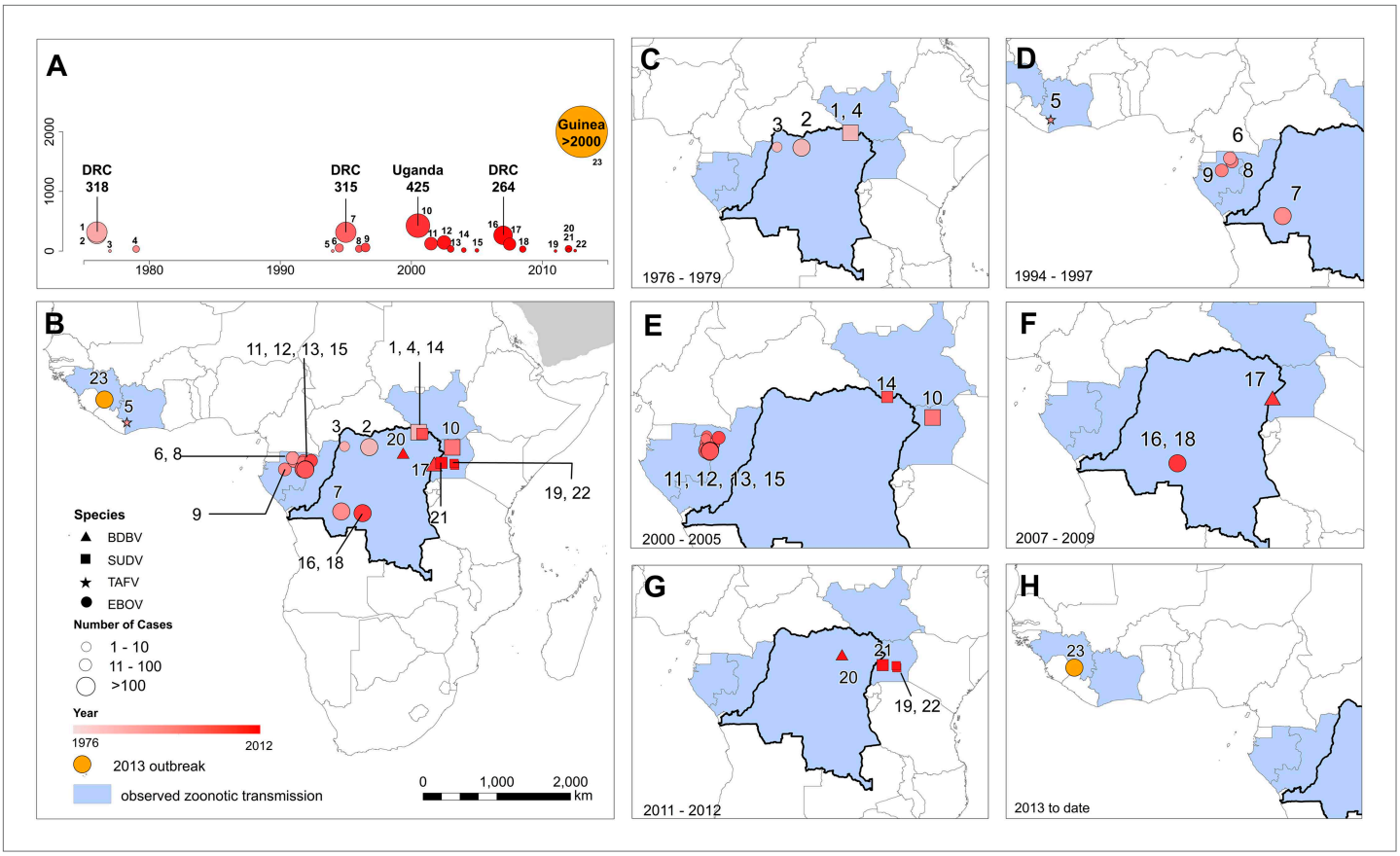


Figure 2. The locations of Ebola virus disease outbreaks in humans in Africa. (**A**) Illustrates the 23 reported outbreaks of Ebola virus disease through time, with the area of each circle and its position along the y-axis representing the number of cases. The onset year is represented by the colour as per (**B**). (**B**) Shows a map of the index cases for each of these outbreaks. (**C**–**H**) Show these outbreaks over a series of time periods. Numbers refer to outbreaks as listed in ***Table 1***. In (**B**–**H**) the species of Ebola virus responsible for the outbreak is illustrated by the symbol shape, the number of resulting cases and onset date by symbol colour. The most recent outbreak (#23) is indicated in orange. Countries in which zoonotic transmission to humans has been reported or is assumed to have occurred are coloured in blue. In each map the Democratic Republic of Congo is outlined for reference.
DOI: 10.7554/eLife.04395.005

2013, a total of 2322 cases had occurred from *Ebolavirus* infections, a number already overtaken by the likely underreported current case count of the ongoing outbreak >2250 (***WHO, 2014a***) (***Figure 2A***). Of the four viruses circulating in Africa, *Zaire ebolavirus* has been responsible for the most outbreaks (13), followed by *Sudan ebolavirus* (7) and *Bundigbuyo ebolavirus* with just two outbreaks in 2007/8 and 2012. Tai Forest has caused one confirmed infection in humans, from which the patient recovered (***Le Guenno et al., 1995***; ***Formenty et al., 1999***). Although outbreaks have been reported since 1976, there was an absence of reported outbreaks in humans for 15 years between 1979 and 1994 (although antibodies in humans were identified over the period [***Kuhn, 2008***]) and the frequency of outbreaks has increased substantially post 2000 (***Figure 2A***).

Reported Ebola virus infections in animals

A total of 51 surveyed locations reporting infections in animals were identified in the literature since the discovery of the disease (***Table 2***; ***Figure 3***). These comprised 17 infections in gorillas (*Gorilla gorilla*), nine infections in chimpanzees (*Pan troglodytes*), 18 in OWFB and 2 in duikers (*Cephalophus* spp.). A large proportion of the great ape cases originated from the ROC/Gabon border, coinciding with the main known distributions of both chimpanzees and gorillas (***Petter and Desbordes, 2013***) and representing a period of well-documented great ape Ebola outbreaks in and around the Lossi Animal Sanctuary (***Rouquet et al., 2005***; ***Bermejo et al., 2006***; ***Walsh et al., 2009***). All animal isolations of Ebola viruses have come from countries that have also reported index cases of human outbreaks, with the exception of several seropositive bats from a survey in southern Ghana.

Table 2. Locations of reported infections with Ebola virus in animals

Site	Country	Date range	Location	Species	Diagnosis	Reference
1	Côte d'Ivoire	Oct–Nov 1994	Tai Forest	Chimpanzee	Serology	(**Formenty et al., 1999**)
2	Gabon	Jan 1996	Mayiboth 2	Chimpanzee	PCR	(**Lahm et al., 2007**)
3	Gabon	Jul 1996	Near Booue	Chimpanzee	Serology	(**Georges-Courbot et al., 1997**)
4	Gabon	Sept 1996	Lope National Park	Chimpanzee	PCR	(**Lahm et al., 2007**)
5	Gabon & ROC	Aug 2001	Mendemba/Lossi Animal Sanctuary	Chimpanzee	PCR	(**Lahm et al., 2007**)
6	Gabon & ROC	Aug 2001	Mendemba/Lossi Animal Sanctuary	Gorilla	PCR	(**Lahm et al., 2007**)
7	Gabon & ROC	Aug 2001	Mendemba/Lossi Animal Sanctuary	*Cephalophus dorsalis*	PCR	(**Lahm et al., 2007**)
8	Gabon	Nov 2001	Zadie	Gorilla	PCR	(**Rouquet et al., 2005**)
9	Gabon	Nov 2001	Ekata	Gorilla	PCR	(**Wittmann et al., 2007**)
10	Gabon	Dec 2001	Medemba and neighbouring villages	Chimpanzee and Gorilla	PCR	(**Leroy et al., 2002**)
11	Gabon	Feb 2002	Zadie	Gorilla	PCR	(**Rouquet et al., 2005**)
12	Gabon	Feb 2002	Ekata	Various bat species	Serology	(**Leroy et al., 2005**)
13	Gabon	Mar 2002	Zadie	Gorilla	PCR	(**Rouquet et al., 2005**)
14	Gabon	Mar 2002	Grand Etoumbi	Gorilla	PCR	(**Wittmann et al., 2007**)
15	Gabon	Apr 2002	Ekata	Gorilla	PCR	(**Wittmann et al., 2007**)
16	ROC	May 2002	Oloba	Chimpanzee	PCR	(**Lahm et al., 2007**)
17	ROC	Dec 2002	Lossi Animal Sanctuary	Gorilla	PCR	(**Rouquet et al., 2005**)
18	ROC	Dec 2002	Lossi Animal Sanctuary	Gorilla	PCR	(**Rouquet et al., 2005**)
19	ROC	Dec 2002	Lossi Animal Sanctuary	Chimpanzee	Serology	(**Rouquet et al., 2005**)
20	ROC	Dec 2002	Lossi Animal Sanctuary	Gorilla	PCR	(**Rouquet et al., 2005**)
21	ROC	Dec 2002	Lossi Animal Sanctuary	Gorilla	PCR	(**Rouquet et al., 2005**)
22	ROC	Dec 2002	Lossi Animal Sanctuary	*Cephalophus* spp.	PCR	(**Rouquet et al., 2005**)
23	Gabon	Feb 2003	Mbomo	Various bat species	PCR	(**Leroy et al., 2005**)
24	ROC	Feb 2003	Lossi Animal Sanctuary	Gorilla	Serology	(**Rouquet et al., 2005**)
25	Gabon	Feb 2003	Lossi Animal Sanctuary	Chimpanzee	PCR	(**Wittmann et al., 2007**)
26	Gabon	Jun 2003	Mbomo	Various bat species	PCR and serology	(**Leroy et al., 2005**)
27	ROC	Jun 2003	Near Mbomo and Ozala National Park	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)
28	ROC	Jun 2003	Near Mbomo and Ozala National Park	*Hypsignathus monstrosus*	Serology	(**Pourrut et al., 2009**)
29	ROC	Jun 2003	Near Mbomo and Ozala National Park	*Myonycteris torquata*	Serology	(**Pourrut et al., 2009**)

Table 2. Continued on next page

Table 2. Continued

Site	Country	Date range	Location	Species	Diagnosis	Reference
30	ROC	Jun 2003	Mbanza	Gorilla	PCR	(**Rouquet et al., 2005**)
31	ROC	Jan–Jun 2004	Lokoué	Gorilla	Reported	(**Caillaud et al., 2006**)
32	ROC	May 2004	Lokoué	Gorilla	PCR	(**Wittmann et al., 2007**)
33	Gabon	Feb 2005	Near Franceville	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)
34	Gabon	Feb 2005	Near Franceville	*Myonycteris torquata*	Serology	(**Pourrut et al., 2009**)
35	Gabon	Apr 2005	Near Lambarene	*Epomops franqueti* and *Hypsignathus monstrosus*	Serology	(**Pourrut et al., 2007**)
36	ROC	May 2005	Near Mbomo and Ozala National Park	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)
37	ROC	May 2005	Near Mbomo and Ozala National Park	*Hypsignathus monstrosus*	Serology	(**Pourrut et al., 2009**)
38	ROC	May 2005	Near Mbomo and Ozala National Park	*Myonycteris torquata*	Serology	(**Pourrut et al., 2009**)
39	ROC	Jun 2005	Odzala National Park	Gorilla	PCR	(**Wittmann et al., 2007**)
40	Gabon	Feb 2006	Near Tchibanga	Various bat species	Serology	(**Pourrut et al., 2009**)
41	ROC	May 2006	Near Mbomo and Ozala National Park	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)
42	ROC	May 2006	Near Mbomo and Ozala National Park	*Hypsignathus monstrosus*	Serology	(**Pourrut et al., 2009**)
43	ROC	May 2006	Near Mbomo and Ozala National Park	*Myonycteris torquata*	Serology	(**Pourrut et al., 2009**)
44	Gabon	Oct 2006	Near Franceville	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)
45	Ghana	May 2007	Sagyimase	*Epomops franqueti*	Serology	(**Hayman et al., 2012**)
46	Ghana	May 2007	Sagyimase	*Hypsignathus monstrosus*	Serology	(**Hayman et al., 2012**)
47	Ghana	May 2007	Adoagyir	*Epomophorus gambianus*	Serology	(**Hayman et al., 2012**)
48	Ghana	May 2007	Adoagyir	*Epomops franqueti*	Serology	(**Hayman et al., 2012**)
49	Ghana	Jun 2007	Oyibi	*Epomophorus gambianus*	Serology	(**Hayman et al., 2012**)
50	Ghana	Jan 2008	Accra	*Eidolon helvum*	Serology	(**Hayman et al., 2010**)
51	Gabon	Mar 2008	Near Franceville	*Epomops franqueti*	Serology	(**Pourrut et al., 2009**)

ROC = Republic of Congo.

DOI: 10.7554/eLife.04395.006

Predicted distribution of suspected reservoir species of bats

Three species of bats, *Hypsignathus monstrosus*, *Myonycteris torquata* and *Epomops franqueti*, were identified as the most likely candidates to be reservoir species for Ebola viruses due to high

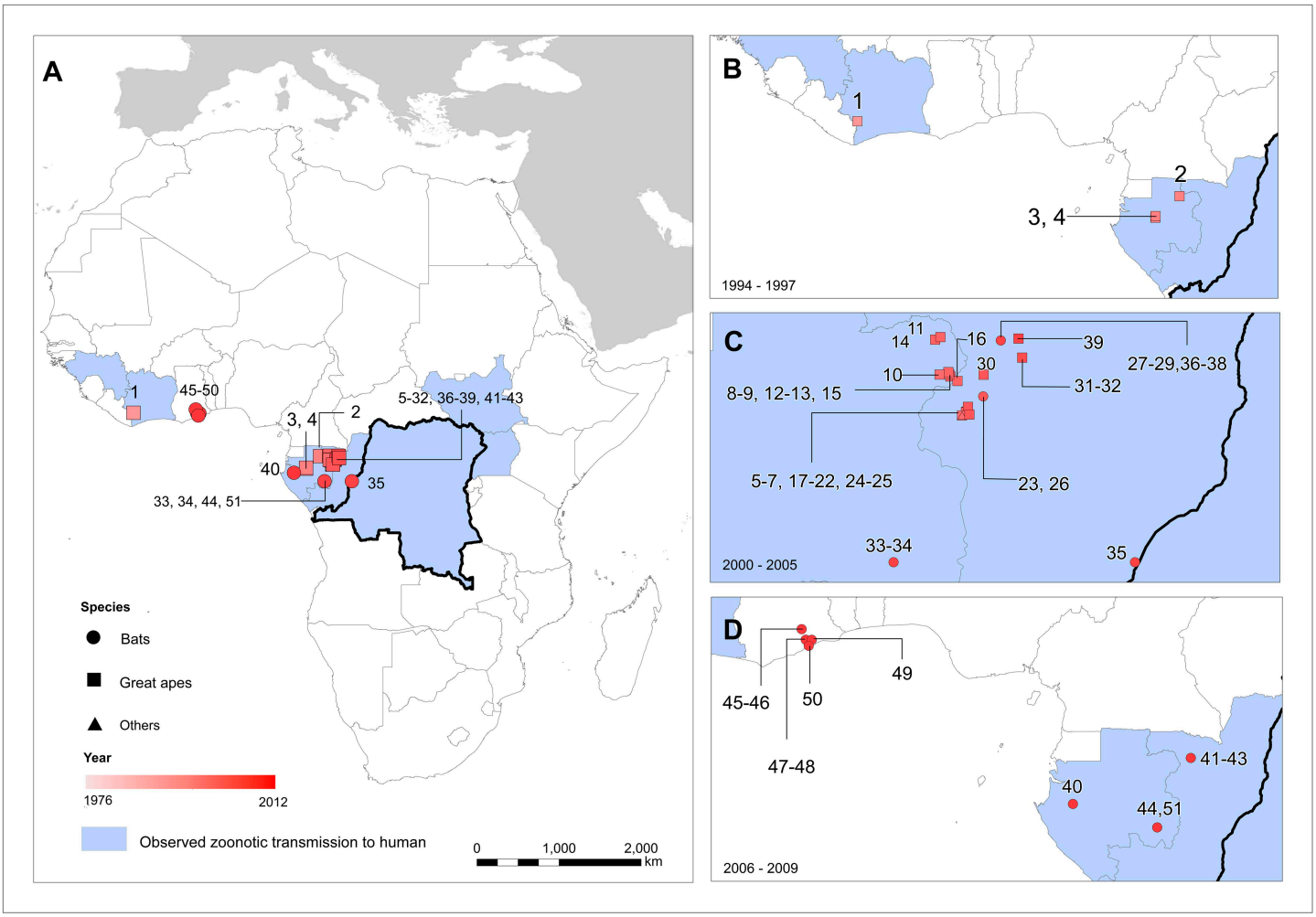


Figure 3. The locations of reported Ebola virus infection in animals in Africa. (**A**) Shows the locations of reported Ebola virus infection in animals. (**B**–**D**) Show these records in animals over three different time periods. Numbers refer to records as listed in ***Table 2***. In all panels, the species in which infection was detected is given by symbol shape and the year recorded by symbol colour. Blue countries represent locations where zoonotic transmission to humans has been reported or is assumed to have occurred. In each map the Democratic Republic of Congo is outlined for reference.

DOI: 10.7554/eLife.04395.007

seroprevalence and the isolation of RNA closely related to *Zaire ebolavirus* (***Leroy et al., 2005***; ***Olival and Hayman, 2014***). In total, 239 locations were identified from the Global Biodiversity Information Facility (GBIF) (***GBIF, 2014***): 67 for *H. monstrosus* (***Figure 4A***), 52 for *M. torquata* (***Figure 4B***) and 120 for *E. franqueti* (***Figure 4C***). Distribution models for all three species demonstrated predictive skill (indicated by an area under the curve (AUC) greater than 0.5) as follows: *H. monstrosus* AUC 0.63 ± 0.04; *M. torquata* AUC = 0.59 ± 0.04; *E. franqueti* AUC = 0.58 ± 0.03, n = 50 submodels for all three species. In addition, each species was broadly predicted within its considered expert opinion range (***Figure 4A–C***) (***Schipper et al., 2008***). The marginal effect plots (not shown) were strongly influenced by land surface temperature (LST) and vegetation (as measured by the enhanced vegetation index [EVI]). The predicted combined distribution of these species (***Figure 4D***), covers West and Central Africa, specifically the moist forests of the northeastern, western and central Congo basin, and Guinea, as well as the Congolian coastal forest ecoregions (***WWF, 2014***).

Predicted environmental suitability for zoonotic transmission of Ebola

The predicted environmental niche for zoonotic transmission of EVD is shown in ***Figure 5***. All countries with observed index cases of EVD (n = 7, hereafter Set 1) have areas of the highest environmental suitability (see list in ***Table 1***). In addition, areas of high environmental suitability for zoonotic

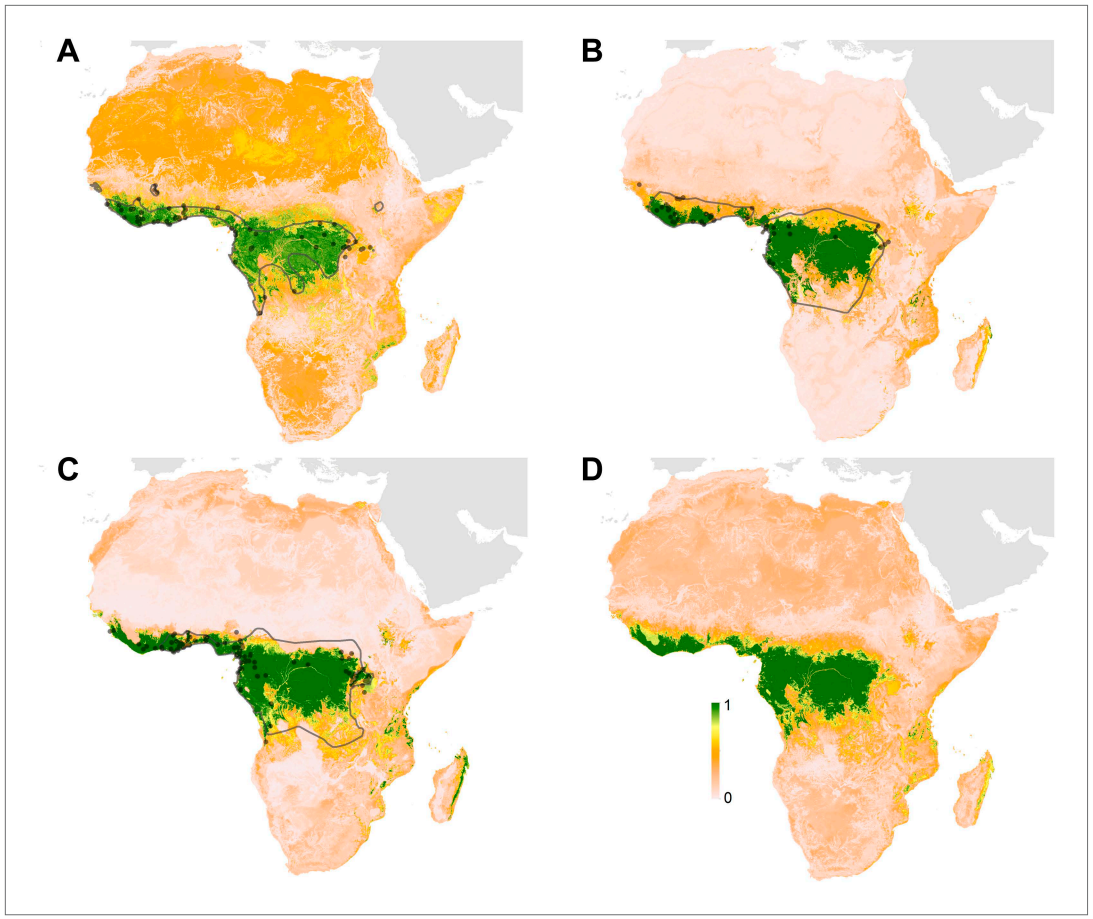


Figure 4. Predicted geographical distribution of the three species of Megachiroptera suspected to reservoir Ebola virus. (**A**) Shows the distribution of the hammer-headed bat (*Hypsignathus monstrosus*), (**B**) The little collared fruit bat (*Myonycteris torquata*) and (**C**) Franquet's epauletted fruit bat (*Epomops franqueti*). In each map, the locations of reported observations of each species, extracted and curated from the Global Biodiversity Information Facility (***GBIF, 2014***) and used to train each model are given as grey points (*H. monstrosus*, n = 67; *E. franqueti*, n = 120 and *M. torquata*, n = 52). Expert opinion maps of the known range of each species, generated by the IUCN (***Schipper et al., 2008***), are outlined in grey. The colour legend represents a scale of the relative probability that the species occurs in that location from 0 (white, low) to 1 (green, high). Area under the curve statistics, calculated under a stringent ten-fold cross validation procedure, are 0.63 ± 0.04, 0.59 ± 0.04 and 0.58 ± 0.03 for *H. monstrosus*, *M. torquata* and *E. franqueti* respectively. (**D**) Is a composite distribution map giving the mean, relative probability of occurrence from (**A**–**C**).

DOI: 10.7554/eLife.04395.008

transmission are predicted in a further 15 countries where, to date, index cases of the four African species of *Ebolavirus* have not been recorded. These are Nigeria, Cameroon, Central African Republic (CAR), Ghana, Liberia, Sierra Leone, Angola, Tanzania, Togo, Ethiopia, Mozambique, Burundi, Equatorial Guinea, Madagascar and Malawi (hereafter Set 2).

The AUC for the Ebola model was relatively high (AUC = 0.85 ± 0.04, n = 500 submodels) indicating that the model could strongly distinguish regions of environmental suitability for EVD. Enhanced vegetation index had the greatest impact on the distribution (relative contribution [RC] of 65.3%) followed by elevation (RC = 11.7%), night-time land surface temperature (LST) (RC = 7.7%), potential evapotranspiration (PET) (RC = 5.7%) and combined bat distribution (RC = 3.8%). Marginal effect plots are presented in ***Figure 5—figure supplement 2***.

In total, 22.2 million people are predicted to live in areas suitable for zoonotic transmission of Ebola. The vast majority, 21.7 million (approximately 97%), live in rural areas, as opposed to urban or peri-urban areas (***CIESIN/IFPRI/WB/CIAT, 2007***; ***WorldPop, 2014***). Of these, 15.2 million are in Set 1 and 7 million are in Set 2. In terms of ranked populations at risk, DRC, Guinea and Uganda are highest

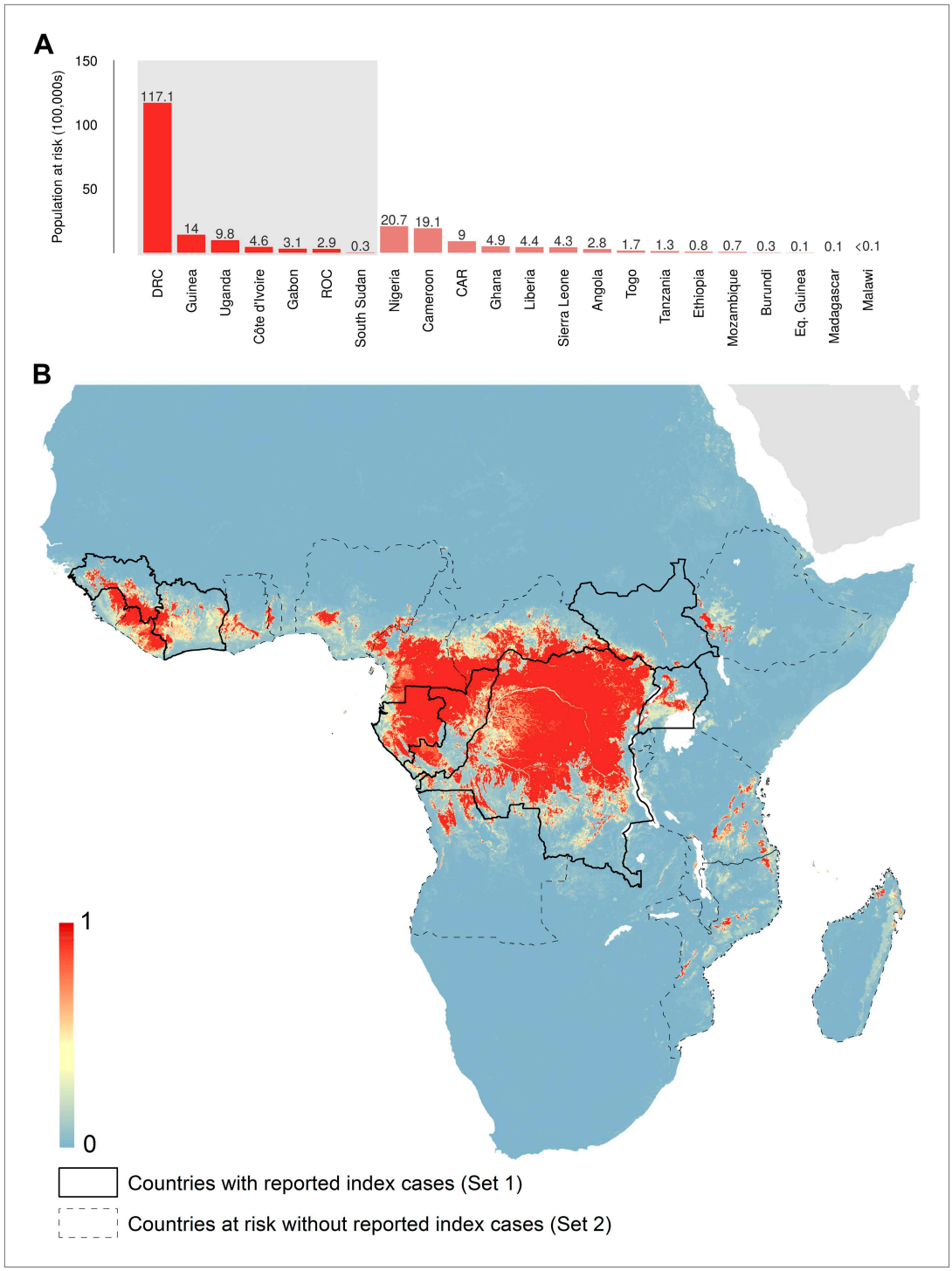


Figure 5. Predicted geographical distribution of the zoonotic niche for Ebola virus. (**A**) Shows the total populations living in areas of risk of zoonotic transmission for each at-risk country. The grey rectangle highlights countries in which index cases of Ebola virus disease have been reported (Set 1); the remainder are countries in which risk of zoonotic transmission is predicted, but in which index cases of Ebola have not been reported (Set 2). These countries are ranked by population at risk within each set. The population at risk Figure in 100,000 s is given above each bar. (**B**) Shows the predicted distribution of zoonotic Ebola virus. The scale reflects the relative probability that zoonotic transmission of Ebola virus could occur at these locations; areas closer to 1 (red) are more likely to harbour zoonotic transmission than those closer to 0 (blue). Countries with borders outlined are those which are predicted to contain at-risk areas for zoonotic transmission based on a thresholding approach (see 'Materials and methods').

Figure 5. Continued on next page

Figure 5. Continued

The area under the curve statistic, calculated under a stringent 10-fold cross-validation procedure is 0.85 ± 0.04. Solid lines represent Set 1 whilst dashed lines delimit Set 2. Areas covered by major lakes have been masked white.
DOI: 10.7554/eLife.04395.009

The following figure supplements are available for figure 5:

Figure supplement 1. Covariates used in predicting zoonotic transmission niche of Ebola.
DOI: 10.7554/eLife.04395.010

Figure supplement 2. Marginal effect plots for each covariate used in the Ebola virus distribution model.
DOI: 10.7554/eLife.04395.011

Figure supplement 3. Comparison of predictions for zoonotic niche of Ebola virus excluding the Guinea outbreak.
DOI: 10.7554/eLife.04395.012

in Set 1 and Nigeria, Cameroon and CAR are top in Set 2. For a full listing of these populations living in areas of risk, see the stacked bar plot in ***Figure 5A***.

National level demographic and mobility changes

Over the 40 year period since discovery of EVD, the total population living in those countries predicted to be within the zoonotic niche has nearly tripled (from 230 million to 639 million) and the proportion of the population in these countries living in an urban (rather than rural) setting has changed from 25.5% to 59.2% (***Figure 6***).

Data on the connectivity of human populations over this period were not available. We can infer however, intuitively, empirically and theoretically (***Zipf, 1946***; ***Simini et al., 2012***) that rates of population movement within a country will scale directly in proportion to population growth.

International connectivity by airline traffic

Records of passenger seat capacity are available since 2000 and show substantive increases over the period in Set 1 (from 2.96 to 4.77 million, a fractional change of 1.61) and Set 2 (from 5.6 to 15.6 million, a change of 2.8) (***Figure 7A***). More specific data on passenger volumes show almost universally similar increases since 2005 with Set 1 nations changing from 2 million to 2.5 million, a fractional change of 1.22 and Set 2 changing from 5 million to 7.9 million, a change of 1.57 (***Figure 7B***).

Global analysis of airline passenger volumes demonstrates that international connectivity has increased amongst all global regions and national income strata (***Figure 8***). Total passenger volumes

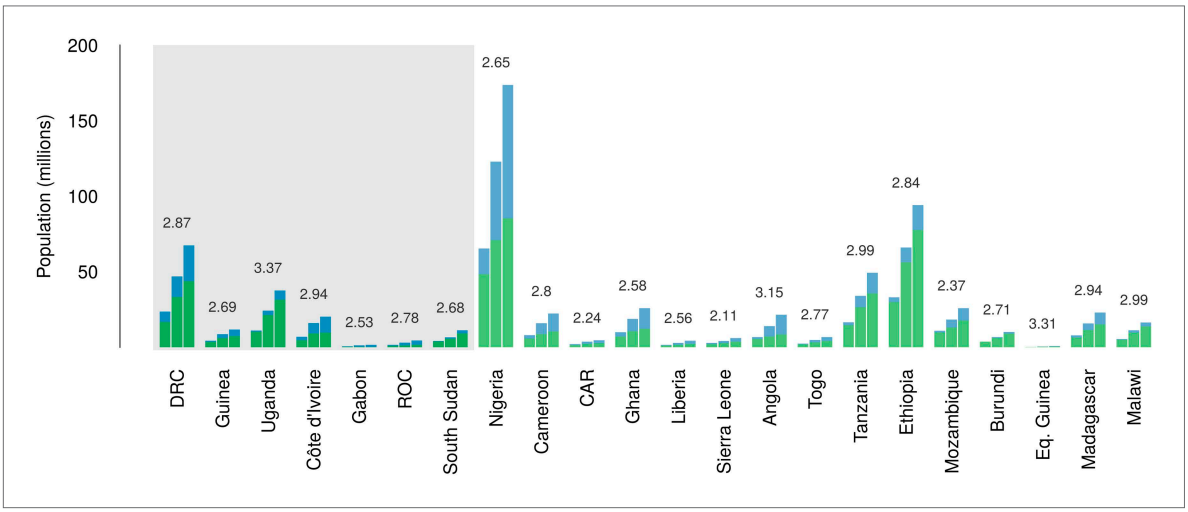


Figure 6. Changes in national population for countries predicted to contain areas at-risk of zoonotic Ebola virus transmission. For each country the population (in millions) is presented for three time periods (1976, 2000 and 2014) as three bars. Each stacked bar gives the rural (green) and urban (blue) populations of the country. The grey rectangle highlights countries in which index cases of Ebola virus diseases have been reported (Set 1); the remainder are countries in which risk of zoonotic transmission is predicted, but where index cases have not been reported (Set 2). The fractional change in population between 1976 and 2014 is given above each set of bars.
DOI: 10.7554/eLife.04395.013

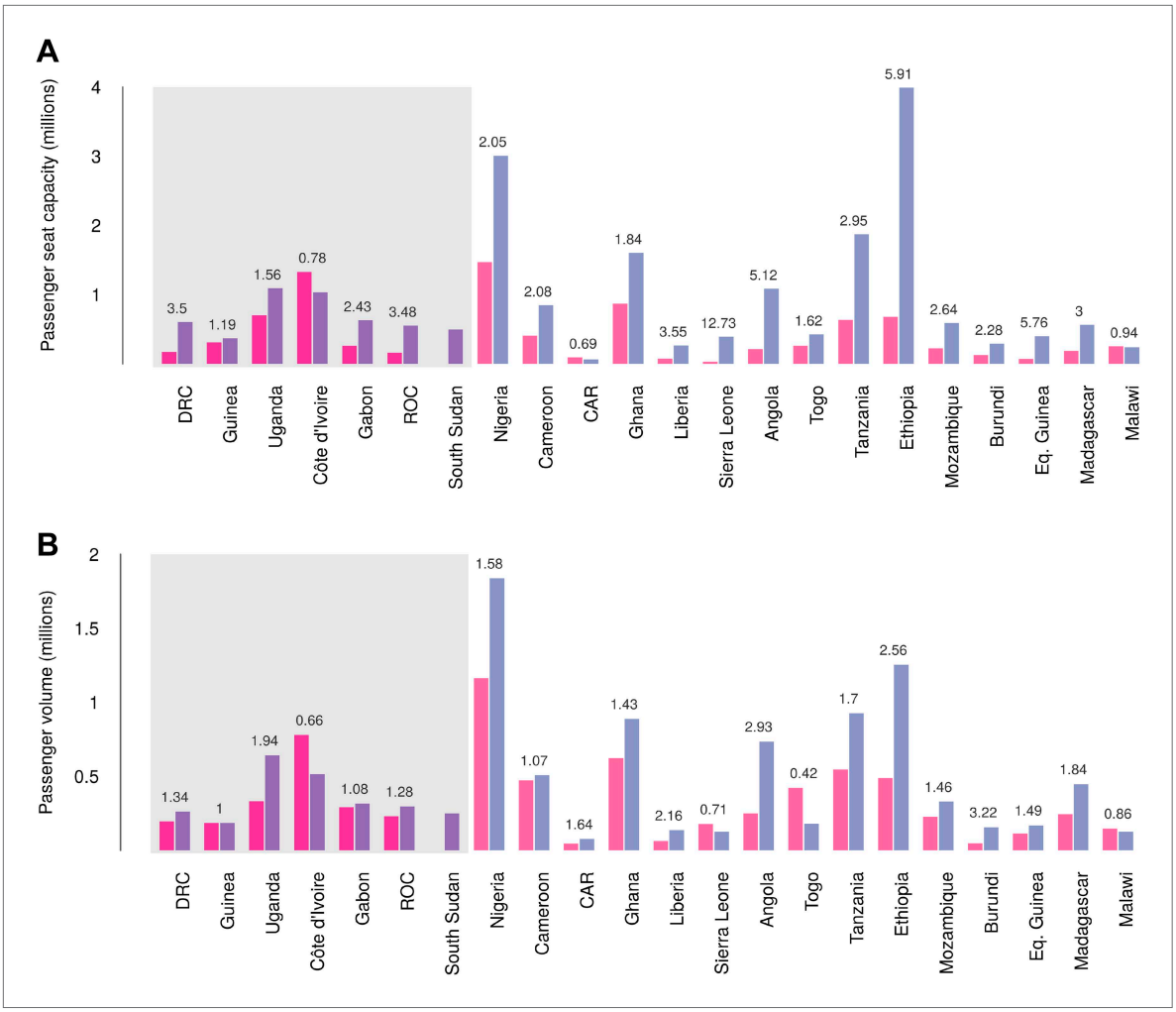


Figure 7. Changes in international flight capacity and traveller volumes for countries predicted to contain areas at-risk of zoonotic Ebola virus transmission. The grey rectangle highlights countries in which index cases of EVD have been reported (Set 1). The remainder are countries in which risk of zoonotic transmission is predicted, but where index cases have not been reported (Set 2). (**A**) Shows changes in annual outbound international seat capacity (between 2000 in red and 2013 in blue). (**B**) Depicts changes in annual outbound international passenger volume by country (between 2005 in red and 2012 in blue). For each country, the fractional change in volume is given above each set of bars. Note that only one bar is presented for South Sudan as data for this region prior to formation of the country in 2011 were unavailable.

DOI: 10.7554/eLife.04395.014

have increased by a third from 9.5 to over 14 million during the eight year window (2005–2012) where records are available. The largest increases have occurred in WHO regions (***WHO, 2014b***) outside of the sub-Saharan African region (AFRO) (***Figure 8A,B***). In 2012, almost half of the final destinations of those travelling from these at-risk countries were to other AFRO nations (47%). Other frequent destinations were in Europe (EURO; 27%) and the Eastern Mediterranean (EMRO; 13%). Similarly, analysis of passenger volumes by World Bank national income groupings (***WHO, 2014b***) (***Figure 8C,D***) show that in 2012 40% of all passenger final destinations were to low or low-middle income countries.

Discussion

Summary of the main findings

We have re-evaluated the zoonotic niche for EVD in Africa. In doing so we have (i) used all existing outbreaks to assemble an inventory of index cases (n = 30); (ii) added to this all confirmed records of Ebola virus in animals (n = 51); (iii) assembled more accurate and contemporary environmental covariates including new maps of the distribution of confirmed OWFB reservoirs of the disease; and (iv) used

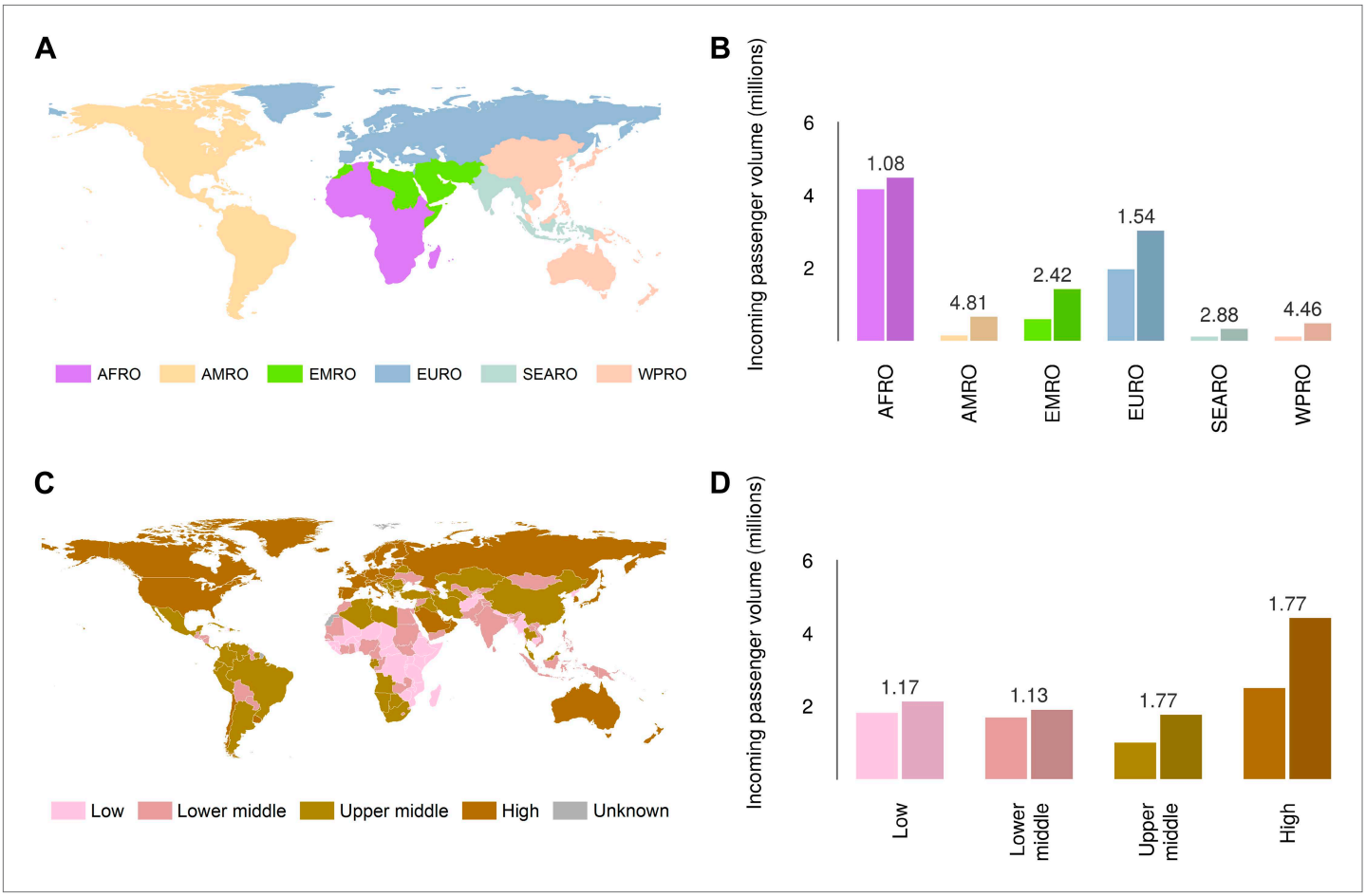


Figure 8. Numbers of airline passengers arriving from at-risk countries to other countries stratified by major geographic regions and national income groups. (**A**) Shows the locations of WHO regions (AFRO–African Region; AMRO–Region of the Americas; EMRO–Eastern Mediterranean Region; EURO–European Region; SEARO–South-East Asian Region; WPRO–Western Pacific Region). (**B**) Displays the numbers of passengers arriving in each of these regions from countries predicted to contain areas at risk of zoonotic Ebola virus transmission (Sets 1 and 2) in 2005 and 2012. (**C**) Shows the income tiers of all countries as defined by the World Bank. (**D**) Displays the total numbers of passengers arriving in countries in each of these income strata from at-risk countries in 2005 and 2012. The number above each pair of bars indicates the fractional change in these numbers of incoming passengers between 2005 and 2012.

DOI: 10.7554/eLife.04395.015

the latest niche modelling techniques to predict the geographic distribution of potential zoonotic transmission of the disease. Using these predictions we have estimated the populations at risk of EVD both in countries which have confirmed index cases (Set 1, n = 7) and those for which we predict strong environmental suitability for outbreaks (Set 2, n = 15). In all countries at risk we show that since the discovery of EVD in 1976, urban and rural populations have increased and have become more interconnected both within and across national borders. During the last 40 years the increasing size and connectivity of these populations may have facilitated the subsequent spread of EVD outbreaks. These factors underline a change in the way in which EVD interacts with human populations.

Interpreting the zoonotic niche

The remote and isolated nature of Ebola zoonotic transmission events, paired with the relatively poor diagnostics and understanding of the disease transmission routes in early outbreaks, mean that under-reporting of previous outbreaks is probable. An increasing understanding and description of a broader range of symptoms used in case definitions of EVD (***Leroy et al., 2000***; ***Feldmann and Geisbert, 2011***) also increase the possibility that past outbreaks may have been misattributed to different diseases (***Tignor et al., 1993***). This poor detectability of EVD also clearly limits capacity to accurately

identify the locations and transmission routes of index cases (**Heymann et al., 1980**; **Baize et al., 2014**). We must assume, as has been done previously (**Peterson et al., 2004a**; **Jones et al., 2008**), that the first reported cases are representative of the true location of the index cases. Where possible we have represented this geographic uncertainty by attributing the index case to a wide-area polygon which then incorporated this uncertainty into the mapping process (see 'Materials and methods').

The relationship between the EVD niche and the environmental covariates (**Figure 5—figure supplement 2**), particularly the high relative contribution of the vegetation index, underscore that there are clear environmental limits to transmission of the virus from animals to humans, and that ecoregions dominated by rainforest are the primary home of such zoonotic cycles. Our analysis has shown that the zoonotic niche of the pathogen is more widespread than previously predicted or appreciated (**Peterson et al., 2004a**), most notably in West Africa.

This analysis used information from all human outbreaks and animal infections to delineate the likely zoonotic niche of the disease. Further analysis, excluding the existing outbreak focussed in Guinea from the dataset used to train the model (**Figure 5—figure supplement 3**), still resulted in prediction of high suitability in this region, with the presumed index village located within 5 km of an at-risk pixel. This implies that the eco-epidemiological situation in Guinea is very similar to that in past outbreaks, mirroring phylogenetic similarity in the causative viruses (**Dudas and Rambaut, 2014**; **Gire et al., 2014**). The ecological similarity between the past and current outbreaks also lends support to the notion that the scale of this outbreak is more heavily influenced by patterns of human-to-human transmission than any expansion of the zoonotic niche.

Interpreting population at risk

It is important to appreciate that this zoonotic niche map delineates areas in which populations are at-risk of zoonotic transmission of EVD (**Figure 5B**). It does not predict the likelihood of EVD spillover, the likelihood of an outbreak establishing, or its subsequent rate of spread within a population. Increasing human encroachment and certain cultural practices sometimes linked with poverty, such as bushmeat hunting, result in increasing exposure of humans to animals which may harbour diseases including Ebola (**Daszak, 2000**; **Wolfe et al., 2005**, **2007**). Increasing human population may accelerate the degree of risk through these processes but spatially refined information on these factors is not available comprehensively. It is hoped that as the understanding of the risk factors for zoonotic transmission of *Ebolavirus* to humans increases, it will be possible to incorporate this information into future risk mapping assessments.

Previous considerations of the geographic distribution of EVD have used human outbreaks alone. We have updated this work to include the last decade of outbreaks, as well as disaggregated outbreaks where evidence suggests multiple independent zoonotic transmission events overlap in space and time. Furthermore, our modelling process accommodates uncertainty in geopositioning of these index cases by utilising both point and polygon data. In addition, we include occurrence of infection in wildlife, important to the wider scale of zoonotic transmission (**Figure 1**), which in total has increased the dataset used in the model to 81 occurrences. The rareness of EVD outbreaks and the prevalence of detectable Ebola virus in reservoir species suggests that there will always be a limited set of observation data when compared to mapping of more prevalent zoonoses (**Pigott et al., 2014**). The results demonstrate predictive skill using a stringent validation procedure, however, indicating strong model performance even with this relatively limited observation dataset.

A broad zoonotic niche is predicted across 22 countries in Central and West Africa. Whilst several of these countries have reported index cases of EVD, others have not, although serological evidence in some regions points to possible underreporting of small-scale outbreaks (**Kuhn, 2008**). With improved ecological understanding, particularly with improvements to our knowledge of specific reservoir species and their distributions, it may be possible to delineate areas not at risk due to the absence of these species.

Despite relatively a large population living in areas of risk and the widespread practice of bushmeat hunting in these predicted areas (**Wolfe et al., 2005**; **Mfunda and Røskaft, 2010**; **Brashares et al., 2011**; **Kamins et al., 2011**), *Ebolavirus* is rare both in suspected animal reservoirs (**Leroy et al., 2005**; **Olival and Hayman, 2014**) and in terms of human outbreaks (**Table 1**). There is some indication however, that the frequency of Ebola outbreaks has increased since 2000, as shown in **Figure 2A**. We have shown that the human population living within this niche is larger, more mobile and better internationally connected than when the pathogen was first observed. As a result, when spillover events do

occur, the likelihood of continued spread amongst the human population is greater, particularly in areas with poor healthcare infrastructure (**Briand et al., 2014**; **Fauci, 2014**).

Whilst rare in comparison to other high burden diseases prevalent in this region, such as malaria (**Gething et al., 2011**; **Murray et al., 2012**), Ebola outbreaks can have a considerable economic and political impact, and the subsequent destabilisation of basic health care provisioning in affected regions increases the toll of unrecorded morbidity and mortality of more common infectious diseases (**Murray et al., 2014**; **Wang et al., 2014**), throughout and after the epidemic period. The number of concurrent infections during the present outbreak represents a significant strain on healthcare systems that are already poorly provisioned (**Briand et al., 2014**; **Chan, 2014**; **Fauci, 2014**) and many other Set 1 and Set 2 countries rank amongst the lowest per capita healthcare spenders. These considerations should be paramount when international organizations debate the financing requirements for the improvement of healthcare needed in the region and the urgency with which new therapeutics and vaccines can be brought into production (**Brady et al., 2014**; **Goodman, 2014**).

Together, these considerations necessitate prioritisation of efforts to reinforce and improve existing surveillance and control, and encourage the development of therapeutics and vaccines. The national population at risk estimates presented here would be a strong rationale for improving, prioritising and stratifying surveillance for EVD outbreaks and diagnostic capacity in these countries. We believe it would be prudent to test OWFB species in Set 2 countries for Ebola virus (**Hayman et al., 2012**), particularly during the bat breeding season to maximise chances of isolation in order to clarify the outbreak risk in these countries. In all Set 1 and Set 2 countries, raising awareness about the risk presented by reservoir bats and incidental primate hosts and the modes of transmission of this disease could be of value. Finally, increasing our capacity to rapidly map ever changing biological threats is also a core need (**Hay et al., 2013b**).

Interpreting International connectivity

The increasing connectedness of the Africa region means that EVD is now a problem of international concern. While most EVD secondary transmission occurs locally and is likely transported via ground transit (**Francesconi et al., 2003**), the potential for international spread of infection is possible, as demonstrated by the importation of the disease from Liberia to Nigeria, culminating in further secondary transmission in Lagos (**WHO, 2014a**). The aetiology of EVD infection and disease progression means that an international outbreak propagated by air travel remains unlikely, particularly in high-income countries better able to handle EVD cases (**Fauci, 2014**). Nevertheless, a non-negligible threat remains, particularly in the low and middle income destinations and the rapid increase in global connectivity of these at-risk regions indicates that international airports could see more imported cases (**Chan, 2014**).

Future work

We have focussed on reanalysing the zoonotic niche for EVD transmission and the characterisation of the populations at risk to improve the landscape in which future risk and impact of EVD outbreaks can be discussed. During the current emergency much of the work will concentrate on routes of secondary transmission in the human population—conceptually the red arrow of the H box in **Figure 1**. An important task is to stratify the risk of EVD spread both within and between countries and identify the most likely pathways of spread for characterisation and surveillance. Our next priority therefore is to investigate aspects of secondary human-to-human transmission by documenting the rate of geographic spread of EVD during the past and ongoing epidemics to help understand changes in these patterns in the historical record. Simulating these movements in a real landscape of population movement patterns, inferred from population movements assessed by mobile phones and other data (**Garcia et al., 2014**), as well as parametric movement models (**Simini et al., 2013**) is a logical next step, and can be used in future targeting of interventions and potential new treatments for both the current and future outbreaks (**Brady et al., 2014**; **Goodman, 2014**).

As previously discussed, whilst there is the risk of human travel during the latent phase of infection, and therefore potential for international spread, the high pathogenicity during infectiousness (immobilising infected persons) and the likely rapid and effective isolation measures implemented in regions with strong health care systems, limit the pandemic potential of EVD. Nevertheless, improvement of international containment plans and informed discussions of potential risks to airline carriers and populations of other regions will be supported by knowledge of local, regional and international population flows. Assessing these flows by air traffic volumes is an ongoing priority.

There are several other zoonotic viral haemorrhagic fevers (for example *Marburgvirus*, Lassa fever, hantaviral infections and arenaviruses) that are of similar public health and biosecurity concern (**Bannister, 2010**), due to their high virulence and mortality and their potential to cause outbreaks and spread internationally. Despite this their geographical distributions are poorly understood (**Hay et al., 2013a**). Many of the methods applied here can be adapted to these diseases and improve our geographical understanding of the risk presented by these pathogens.

We are in the midst of a public health emergency that will likely last for many more months (**Chan, 2014**) and which has brought EVD to global attention. We emphasise that the maps of zoonotic transmission presented here do not enable assessment of secondary transmission rates in human populations, but they do act as an evidenced-based indicator of locations with potential for future zoonotic transmission and thus outbreaks. Interestingly, early reports of another independent zoonotic outbreak in the DRC (**MSF, 2014**) are in predicted at-risk areas. An improved understanding of the spatial extent of the zoonotic niche can only help future efforts in biosurveillance.

Materials and methods

Methodological overview

A boosted regression tree (BRT) modelling framework was used to generate predictive risk maps of the zoonotic Ebola virus niche in Africa. This methodology combines regression trees, where trees are built according to optimal decision rules based on how binary decisions best accommodate a given dataset (**De'ath, 2007**; **Elith et al., 2008**), and boosting, which selects the tree that minimises the loss function. In doing so, a parameter space is defined which captures the greatest amount of variation present in the dataset. In order to train the model, four component datasets were compiled: (i) a comprehensive dataset of the reported locations of Ebola virus transmission from a zoonotic reservoir to a human; (ii) a dataset of the locations of Ebola virus infections in suspected reservoir and (non-human) susceptible host species (iii) a suite of ecologically relevant environmental covariates for Africa, including predicted distribution maps of suspected reservoir bat species and (iv) background (or pseudo-absence) records representing locations where zoonotic Ebola virus has not been reported. This study was limited to the African continent since no natural outbreaks of EVD have occurred outside the continent (**CDC, 2014**). Only *Reston ebolavirus* has a distribution reported outside of Africa, focussed in the Philippines, but has never been reported as pathogenic in humans; as a result this species was not included in the analysis.

Identifying index cases and reconstructing zoonotic transmission events in space and time

Tables detailing proven outbreaks of Ebola virus, initially sourced from the scientific literature (**Kuhn, 2008**) and from health reporting organisations (**CDC, 2014**), were used to coordinate searches of the formal scientific literature using Web of Science and PubMed for each specific outbreak. Relevant papers were abstracted and where possible outbreak-specific epidemiological surveys were sourced. The citations in these references were obtained in order to reconstruct the outbreak in detail and to identify the most probable index case. Index cases were defined as any human infection resulting from interaction with non-human sources of the disease. Some of these cases arose from presumed interactions with zoonotic reservoirs or hosts, such as primates and other mammals during hunting trips (**Boumandouki et al., 2005**; **Nkoghe et al., 2005**, **2011**; WHO, 2003) or butchering of bats (**Leroy et al., 2009**). Any cases arising from existing human infections are considered as secondary infections rather than index cases. Similar to methodology employed elsewhere (**Messina et al., 2014**), the site, or supposed site, of this zoonotic transmission event was geopositioned using Google Earth. For locations where precise geographic information (e.g., geographic coordinates of a hunting camp) was provided by the authors, these were used. Where precise geographic information could not be accurately geopositioned, a geographic area (termed a polygon) was defined covering the reported region. In several cases only the first reported patient could be identified, with the source of infection unknown. With these outbreaks the location of the first patient was geopositioned under the assumption that an initial zoonotic spillover event occurred in the vicinity of this location. In two outbreaks multiple independent zoonotic transmission events were identified (WHO, 2003; **Nkoghe et al., 2005**; **Pourrut et al., 2005**), and each unique event was geopositioned and included in the model as separate entities. ***Table 1*** catalogues the outbreaks used in this study.

Assembling a database of reported infections in animals

A literature search was conducted in Web of Science using the search term "Ebola" that returned 8635 citations. The abstracts were examined and for those that contained possible data on animal Ebola infection, the full text was obtained. The sampling site or location of the animal in the study was identified and geopositioned using Google Maps. These locations were recorded either as precise locations or as polygons, as with human index cases. Records for which local transmission of Ebola virus was deemed unlikely (e.g., seropositive primates tested in containment facilities several years after their capture) were excluded from the study. The non-human Ebola virus occurrence data collected are detailed in ***Table 2***, including the diagnostic methods used.

GenBank isolates

The open access sequence database GenBank (***NCBI, 2014***) was searched using MESH Umbrella search terms for Ebola virus, returning 181 results. These were then manually cross-referenced with the existing human and animal Ebola information, collected above, and 30 duplicates were removed. For the remaining isolates, original references and GenBank information fields were examined, but as there was insufficient information to establish precise location of isolation and/or whether the isolate represented an index case for any of these data sources, they were excluded from subsequent analyses.

Covariates assembled and used in the analyses

A suite of ecologically relevant gridded environmental covariates for Africa was compiled, each having a nominal resolution of 5 km × 5 km. The environmental covariates used in this analysis were: elevation (from the shuttle radar topography mission [***ORNL DAAC, 2000***]); the mean value, and a measure of spatial variation (range, described below) between 2000 and 2012 of Enhanced Vegetation Index (EVI), daytime Land Surface Temperature (LST) and night-time LST; and mean potential evapotranspiration from 1950–2000 (***Trabucco and Zomer, 2009***) (***Figure 5—figure supplement 1***).

The EVI and LST datasets were derived from satellite imagery collected by NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) remote sensing platform (***Tatem et al., 2004***). EVI is a metric designed to characterise vegetation density and vigour based on the ratio of absorbed photosynthetically active radiation to near infrared radiation (***Huete et al., 2002***). LST is a modelled product derived from emissivity as measured by the MODIS thermal sensor (***Wan and Li, 1997***), which is correlated, though not synonymous with air temperature, and effective for differentiating landscapes based on a combination of thermal energy and properties of surface types. The MODIS datasets utilized in this research (EVI was derived from the MCD43B4 product and the MOD11A2 LST product was used directly) were acquired as composite datasets created using imagery collected over multiple days, a procedure that results in products with 8-day temporal resolutions. Despite compositing, the EVI and LST datasets contained gaps due to persistent cloud cover found in forested equatorial regions, and these gaps were filled using a previously described approach (***Weiss et al., 2014a***). The EVI and LST datasets were then aggregated from their native 1 km × 1 km spatial resolution to a final 5 km × 5 km resolution, calculating both the mean and the range of the values of the subpixels making up each larger pixel. These spatial summaries therefore characterise both the mean temperature in each location as well as the degree of spatial heterogeneity within that pixel. This is of interest as humans and susceptible species are more likely to come into contact in transitional areas (e.g., boundary areas between areas of highly suitable susceptible species habitat and areas heavily utilised by humans). The final covariate production step consisted of summarising temporally across the 13-year data archive to produce synoptic datasets devoid of annual or seasonal anomalies (***Weiss et al., 2014a***).

Implicated bat reservoir distributions

Over recent years, significant research has been undertaken in investigating the role bats have to play in the transmission cycle of Ebola viruses (***Olival and Hayman, 2014***) and evidence of asymptomatic infection in fruit bats has been documented to varying extents (***Leroy et al., 2005***; ***Pourrut et al., 2007***, ***2009***; ***Hayman et al., 2010***; ***Hayman et al., 2012***). In order to incorporate this potential driver of Ebola virus transmission into the model we developed predicted distribution maps for three species of fruit bat implicated as primary reservoirs of the disease: *Hypsignathus monstrosus*, *Epomops franqueti* and *Myonycteris torquata*. The evidence was strongest for these three species having a reservoir role

as Ebola virus RNA (all nested within the *Zaire ebolavirus* phylogeny [**Leroy et al., 2005**]) has been detected in all three. Whilst a handful of other bat species have been found to be seropositive, no further viral isolations have been recorded (**Olival and Hayman, 2014**).

Whilst expert opinion range maps for these species exist (**Schipper et al., 2008**), there is some disagreement with independently-sourced occurrence data (all archived in the Global Biodiversity Information Facility). As a result, a predictive modelling approach was used to create a continuous surface of habitat suitability for these species which we then included as a predictor in the model. Occurrence data for all Megachiroptera in Africa was extracted from GBIF (**GBIF, 2014**) using the R packages dismo (**Hijmans et al., 2014**) and taxize (**Chamberlain et al., 2014**). To remove apparently erroneous records in the GBIF archive all occurrence records more than 100 km from the species known ranges, as determined by expert-opinion range maps (**Schipper et al., 2008**), were excluded, as were duplicate records and those located in the ocean. This resulted in a total dataset of 1341 unique occurrence records.

The occurrence database was then used to train separate boosted regression tree species distribution models (**Elith et al., 2008**) to predict the likely distribution of each of these suspected reservoir species. For each model, occurrence records for the target species (*H. monstrosus*, n = 67; *E. franqueti*, n = 120; and *M. torquata*, n = 52) were considered presence records and occurrence records of all other species were used as background records. This procedure is designed to account for the potentially biasing effect of spatial variation in recording of Megachiroptera occurrences (**Phillips et al., 2009**).

For each species we ran fifty submodels each trained to a randomly selected bootstrap of this dataset, subject to the constraint that each bootstrap contained a minimum of 10 occurrence and 10 background records. Each submodel was fitted using the gbm.step subroutine (**Elith et al., 2008**) in the dismo R package. In each submodel the background records were down weighted so that the weighted sum of presence records equalled the weighted sum of background records (**Barbet-Massin et al., 2012**) in order to maximise the discrimination capacity of the model. We generated a prediction map from each of these submodels and calculated both the mean prediction and 95% confidence interval around the prediction for each 5 km × 5 km pixel for each species.

Model accuracy was assessed by calculating the mean area under the curve (AUC) statistic for each submodel under a stringent 10-fold cross validation for each submodel and obtaining the mean and standard deviation across all 50 submodels. Under this procedure the dataset was split into ten subsets, each containing approximately equal numbers of presence and background points. The ability of a model trained on each subset to predict the distribution of the other 90% of records was assessed by AUC and the mean value taken. As so few presence records were used to train each fold model (i.e., around 5 presence records for *M. torquata* up to 12 for *E. franqueti*), this represents a very stringent test of the model's predictive capacity. Additionally, to prevent inflation of the accuracy statistics due to spatial sorting bias, these statistics were estimated using a pairwise distance sampling procedure (**Hijmans, 2012**). Consequently, the AUC statistics presented here are lower than would be returned by standard procedures but gives a more realistic quantification of the model's ability to extrapolate predictions to new regions (**Wenger and Olden, 2012**). We also generated marginal effect plots with associated uncertainty intervals and relative contribution statistics (how often each covariate was selected during the model fitting process) as quantification of the sensitivity of the model to the different covariates. These allow us to make inferences about the ecological relationship between each species and its environment as well as to identify where this relationship is most uncertain.

To generate a single surface describing the distribution of the bat reservoir species to be used as a covariate in the subsequent Ebola modelling, the three mean prediction distribution maps were merged by taking the average habitat suitability for each of the three bat species at each pixel.

Ebola distribution modelling

The Ebola virus occurrence dataset was supplemented with a background record dataset generated by randomly sampling 10,000 locations across Africa, biased towards more populous areas as a proxy for reporting bias (**Phillips et al., 2009**). We fitted 500 submodels to bootstraps of this dataset. To account for uncertainty in the geographic location of those occurrences reported as polygons, for each submodel one point was randomly selected from each of these occurrence polygons. This Monte

Carlo procedure enabled the model to efficiently integrate over the environmental uncertainty associated with imprecise geographic data. A bootstrap sample was then taken from each of these datasets and used to train the BRT model using the same procedure and weighting of background records as for the bat distribution models. Similarly, we generated a prediction map from each of these models and calculated both the mean prediction and corresponding 95% confidence intervals for each pixel and analysed prediction accuracy using the same stringent cross validation and sensitivity analysis procedure as for the bat distribution models (detailed above).

The predicted distribution map produced by this approach represents the environmental suitability of each pixel for zoonotic Ebola virus transmission. This may be interpreted as a relative probability of presence in the sense that more suitable pixels are more likely to contain zoonotic transmission than less suitable pixels, though not an absolute probability that transmission occurs in a given pixel. Similarly, the presence of zoonotic transmission increases the risk of transmission to a human, though this is also contingent on how humans interact with these zoonotic pools, through hunting or other activities.

Population living in areas of environmental suitability for zoonotic transmission

In order to identify areas which are likely to be at risk of transmission of *Ebolavirus* from zoonotic reservoir hosts to humans, the continuous map of the predicted environmental suitability for zoonotic transmission (shown in ***Figure 5***) was converted into a binary map classifying pixels as either at risk or not at risk. A pixel was assumed to be at risk if its predicted environmental suitability for zoonotic Ebola virus transmission was greater than 0.673, the lowest suitability value predicted at the locations of known transmission to humans (point records of human index cases). Countries containing at least one at-risk pixel are shown in ***Figure 5B***—those countries that previously report an index case were defined as Set 1; countries with at least one at-risk pixel with no previous index cases of EVD were categorised as Set 2. The number of people living in at-risk areas in each of these countries was calculated by summing the estimated population of at-risk pixels using population density maps from the AfriPop project (***Linard et al., 2012***; ***WorldPop, 2014***) and the proportion of those living in urban, periurban and rural areas was evaluated using the Global Rural Urban Mapping Project (***CIESIN/IFPRI/WB/CIAT, 2007***).

The R code used for all of the analysis has been made available on an open source basis (https://github.com/SEEG-Oxford/ebola_zoonotic).

National level demographic and mobility data

For three separate years (1976, 2000 and 2014), total national populations were retrieved and the proportion of rural to urban populations noted from World Bank statistics (***World Bank, 2014***). To describe global air travel patterns from Set 1 and Set 2 countries, flight schedules data from the Official Airline Guide, reflecting an estimated 95% of all commercial flights worldwide, were analysed between 2000 and 2013 to calculate the annual volume of seats on direct flights that depart from each predicted country and which have an international destination. Complementing these seat capacity data, worldwide data on anonymised, individual passenger flight itineraries from the International Air Transport Association (2012) (***IATA, 2014***) were analysed between 2005 and 2012 to calculate the annual volume of international passenger departures out of each Set 1 and Set 2 country. The IATA dataset represents an estimated 93% of the world's commercial air traffic at the passenger level and includes points of departure and arrival and final destination information for travellers as well as their connecting flights.

Acknowledgements

We thank Katherine Battle, Maria Devine and Kirsten Duda for proof-reading and Jane Messina for creating ***Figure 1***. We also thank Andrew Rambaut for his comments on the final draft.

Additional information

Competing interests

SIH: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
University Of Oxford	Sir Richard Southwood Graduate Scholarship	David M Pigott
Bill and Melinda Gates Foundation	OPP1053338	Nick Golding
Medical Research Council	K00669X	Peter W Gething
Biotechnology and Biological Sciences Research Council	Studentship	Oliver J Brady
German Academic Exchange Service	Graduate Scholarship	Moritz UG Kraemer
U.S. National Library of Medicine	R01LM010812	John S Brownstein, Sumiko R Mekaru
7th Framework Programme for Research and Technological Development (EU FP7)	602525	Peter W Horby
Canadian Institutes of Health Research		Isaac I Bogoch, Kamran Khan
Wellcome Trust	095066	Adrian Mylne, Simon I Hay
RAPIDD program of the Science & Technology Directorate		Andrew J Tatem, Simon I Hay
Fogarty International Center		Andrew J Tatem, Simon I Hay
Bill and Melinda Gates Foundation	OPP1106023	Zhi Huang, Andrew J Henry, Catherine L Moyes
Bill and Melinda Gates Foundation	OPP1068048	Daniel J Weiss, Samir Bhatt
Bill and Melinda Gates Foundation	OPP1093011	Catherine L Moyes, John S Brownstein, Sumiko R Mekaru, Simon I Hay
National Institute of Allergy and Infectious Diseases	U19AI089674	Andrew J Tatem
Bill and Melinda Gates Foundation	OPP1106427	Andrew J Tatem
Bill and Melinda Gates Foundation	OPP1032350	Andrew J Tatem

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

DMP, Advised and assembled the outbreak and animal infection data, Wrote the article; NG, Extracted the bat data, Conducted all of the analysis, Drafting or revising the article; AM, Assembled and geo-positioned the outbreak and animal infection data, Drafting or revising the article; ZH, Geo-positioned the outbreak and animal infection data, Provided all maps, Drafting or revising the article; AJH, Analysed international flight data, Helped assemble maps, Drafting or revising the article; IIB, Assisted with international transportation analysis, Edited the final draft of the article; SRM, Assisted in geopositioning, Drafting or revising the article; AJT, Provided information on urban change and migration data, Drafting or revising the article; KK, Provided data, Conducted all analyses on international air traffic patterns, Drafting or revising the article; DJW, Assembled the covariate layers, Drafting or revising the article; SB, Extracted Ebola information from GenBank, Drafting or revising the article; OJB, Screened GenBank data, Provided ***Figure 2A***, Drafting or revising the article; MUGK, DLS, CLM, Analysis and interpretation of data, Drafting or revising the article; PWG, Assembled the covariate layers, Drafting or revising the article; PWH, JSB, Advised on international public health context, Edited the final draft of the article; SIH, Conceived the work and analysis, Wrote content and edited the article at all stages of development, Acts as guarantor of the paper

Author ORCIDs

David M Pigott, http://orcid.org/0000-0002-6731-4034
Nick Golding, http://orcid.org/0000-0001-8916-5570
David L Smith, http://orcid.org/0000-0003-4367-3849
Simon I Hay, http://orcid.org/0000-0002-0611-7272

References

Amblard J, Obiang P, Edzang S, Prehaud C, Bouloy M, Guenno BL. 1997. Identification of the Ebola virus in Gabon in 1994. *Lancet* **349**:181–182. doi: 10.1016/S0140-6736(05)60984-1.

Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Soropogui B, Sow MS, Keïta S, De Clerck H, Tiffany A, Dominguez G, Loua M, Traoré A, Kolié M, Malano ER, Heleze E, Bocquin A, Mély S, Raoul H, Caro V, Cadar D, Gabriel M, Pahlmann M, Tappe D, Schmidt-Chanasit J, Impouma B, Diallo AK, Formenty P, Van Herp M, Günther S. 2014. Emergence of Zaire Ebola virus disease in Guinea - preliminary report. *The New England Journal of Medicine* doi: 10.1056/NEJMoa1404505.

Bannister B. 2010. Viral haemorrhagic fevers imported into non-endemic countries: risk assessment and management. *British Medical Bulletin* **95**:193–225. doi: 10.1093/Bmb/Ldq022.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**:327–338. doi: 10.1111/j.2041-210X.2011.00172.x.

Baron RC, McCormick JB, Zubeir OA. 1983. Ebola virus disease in southern Sudan: hospital dissemination and intrafamilial spread. *Bulletin of the World Health Organization* **61**:997–1003.

Barrette RW, Metwally SA, Rowland JM, Xu LZ, Zaki SR, Nichol ST, Rollin PE, Towner JS, Shieh WJ, Batten B, Sealy TK, Carrillo C, Moran KE, Bracht AJ, Mayr GA, Sirios-Cruz M, Catbagan DP, Lautner EA, Ksiazek TG, White WR, McIntosh MT. 2009. Discovery of swine as a host for the Reston ebolavirus. *Science* **325**:204–206. doi: 10.1126/science.1172705.

Bausch DG, Schwarz L. 2014. Outbreak of Ebola virus disease in Guinea: where ecology meets economy. *PLOS Neglected Tropical Diseases* **8**:e3056. doi: 10.1371/journal.pntd.0003056.

Beer B, Kurth R, Bukreyev A. 1999. Characteristics of Filoviridae: Marburg and Ebola viruses. *Die Naturwissenschaften* **86**:8–17. doi: 10.1007/s001140050562.

Bermejo M, Rodríguez-Teijeiro JD, Illera G, Barroso A, Vilà C, Walsh PD. 2006. Ebola outbreak killed 5000 gorillas. *Science* **314**:1564. doi: 10.1126/science.1133105.

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. 2013. The global distribution and burden of dengue. *Nature* **496**:504–507. doi: 10.1038/Nature12060.

Boumandouki P, Formenty P, Epelboin A, Campbell P, Atsangandoko C, Allarangar Y, Leroy EM, Kone ML, Molamou A, Dinga-Longa O, Salemo A, Kounkou RY, Mombouli V, Ibara JR, Gaturuku P, Nkunku S, Lucht A, Feldmann H. 2005. [Clinical management of patients and deceased during the Ebola outbreak from October to December 2003 in Republic of Congo]. *Bulletin de la Société de Pathologie Exotique* **98**:218–223.

Brady OJ, Hay SI, Horby P. 2014. Estimating vaccine and drug requirements for Ebola outbreaks. *Nature* **512**:7514. doi: 10.1038/512233a.

Brashares JS, Golden CD, Weinbaum KZ, Barrett CB, Okello GV. 2011. Economic and geographic drivers of wildlife consumption in rural Africa. *Proceedings of the National Academy of Sciences of USA* **108**:13931–13936. doi: 10.1073/pnas.1011526108.

Briand S, Bertherat E, Cox P, Formenty P, Kieny MP, Myhre JK, Roth C, Shindo N, Dye C. 2014. The international Ebola emergency. *The New England Journal of Medicine* doi: 10.1056/NEJMp1409858.

Brockmann D, Helbing D. 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**:1337–1342. doi: 10.1126/science.1245200.

Caillaud D, Levréro F, Cristescu R, Gatti S, Dewas M, Douadi M, Gautier-Hion A, Raymond M, Ménard N. 2006. Gorilla susceptibility to Ebola virus: the cost of sociality. *Current Biology* **16**:R489–R491. doi: 10.1016/j.cub.2006.06.017.

CDC. 2014. Chronology of Ebola hemorrhagic fever outbreaks. Accessed: August 2014. Available: http://www.cdc.gov/vhf/ebola/resources/outbreak-table.html.

Chamberlain S, Szocs E, Boettiger C, Ram K, Bartomeus I, Baumgartner J. 2014. taxize: taxonomic information from around the web. *R package*. Accessed: August 2014. Available: https://github.com/ropensci/taxize.

Chan M. 2014. Ebola virus disease in West Africa - no early end to the outbreak. *The New England Journal of Medicine* online. doi: 10.156/NEJMp1409859.

Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. 2004. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology* **229**:119–126. doi: 10.1016/j.jtbi.2004.03.006.

CIESIN/IFPRI/WB/CIAT. 2007. Global Rural Urban Mapping Project (GRUMP): gridded population of the world, version 3. Accessed: December 2013. Available: http://sedac.ciesin.columbia.edu/gpw.

Cohen B. 2004. Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World Development* **32**:23–51. doi: 10.1016/j.worlddev.2003.04.008.

Conrad JL, Isaacson M, Smith EB, Wulff H, Crees M, Geldenhuys P, Johnston J. 1978. Epidemiologic investigation of Marburg virus disease, Southern Africa, 1975. *The American Journal of Tropical Medicine and Hygiene* **27**:1210–1215.

Daszak P. 2000. Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science* **287**:443–449. doi: 10.1126/science.287.5452.443.

De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**:243–251. doi: 10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.

Dudas G, Rambaut A. 2014. Phylogenetic analysis of Guinea 2014 EBOV ebolavirus outbreak. *PLOS Currents* **6**:ecurrents.outbreaks.84eefe85ce43ec89dc80bf0670f0677b0678b0417d. doi: 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

ECDC. 2014. *Outbreak of Ebola virus disease in West Africa*. Stockholm: ECDC. p. 17.

Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**:677–697. doi: 10.1146/annurev.ecolsys.110308.120159.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *The Journal of Animal Ecology* **77**:802–813. doi: 10.1111/j.1365-2656.2008.01390.x.

Fauci AS. 2014. Ebola - underscoring the global disparities in health care resources. *The New England Journal of Medicine* doi: 10.1056/NEJMp1409494.

Feldmann H, Geisbert TW. 2011. Ebola haemorrhagic fever. *Lancet* **377**:849–862. doi: 10.1016/S0140-6736(10)60667-8.

Formenty P, Boesch C, Wyers M, Steiner C, Donati F, Dind F, Walker F, Le Guenno B. 1999. Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire. *The Journal of Infectious Diseases* **179**(suppl 1): S120–S126. doi: 10.1086/514296.

Francesconi P, Yoti Z, Declich S, Onek PA, Fabiani M, Olango J, Andraghetti R, Rollin PE, Opira C, Greco D, Salmaso S. 2003. Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. *Emerging Infectious Diseases* **9**:1430–1437. doi: 10.3201/eid0911.030339.

Franklin J. 2009. *Mapping species distributions*. Cambridge: Cambridge University Press. p. 320.

Garcia AJ, Pindolia DK, Lopiano KK, Tatem AJ. 2014. Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* in press. doi: 10.1093/migration/mnu036.

GBIF. 2014. Global biodiversity information facility. Accessed: August 2014. Available: http://www.gbif.org/.

Gear JS, Cassel GA, Gear AJ, Trappler B, Clausen L, Meyers AM, Kew MC, Bothwell TH, Sher R, Miller GB, Schneider J, Koornhof HJ, Gomperts ED, Isaäcson M, Gear JH. 1975. Outbreake of Marburg virus disease in Johannesburg. *British Medical Journal* **4**:489–493. doi: 10.1136/bmj.4.5995.489.

Georges AJ, Leroy EM, Renaut AA, Benissan CT, Nabias RJ, Ngoc MT, Obiang PI, Lepage JP, Bertherat EJ, Bénoni DD, Wickings EJ, Amblard JP, Lansoud-Soukate JM, Milleliri JM, Baize S, Georges-Courbot MC. 1999. Ebola hemorrhagic fever outbreaks in Gabon, 1994-1997: epidemiologic and health control issues. *The Journal of Infectious Diseases* **179**:S65–S75. doi: 10.1086/514290.

Georges-Courbot MC, Sanchez A, Lu CY, Baize S, Leroy E, Lansout-Soukate J, Tévi-Bénissan C, Georges AJ, Trappier SG, Zaki SR, Swanepoel R, Leman PA, Rollin PE, Peters CJ, Nichol ST, Ksiazek TG. 1997. Isolation and phylogenetic characterization of Ebola viruses causing different outbreaks in Gabon. *Emerging Infectious Diseases* **3**:59–62. doi: 10.3201/eid0301.970107.

Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, Tatem AJ, Hay SI. 2011. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* **10**:378. doi: 10.1186/1475-2875-10-378.

Gilbert M, Golding N, Zhou H, Wint GR, Robinson TP, Tatem AJ, Lai S, Zhou S, Jiang H, Guo D, Huang Z, Messina JP, Xiao X, Linard C, Van Boeckel TP, Martin V, Bhatt S, Gething PW, Farrar JJ, Hay SI, Yu H. 2014. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nature Communications* **5**:4116. doi: 10.1038/ncomms5116.

Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JLB, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheiffelin JS, Lander ES, Happi C, Gevao SM, Gnirke A, Rambaut A, Garry RF, Khan SH, Sabeti PC. 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**:1369–1372. doi: 10.1126/science.1259657.

Gonzalez MC, Hidalgo CA, Barabasi AL. 2008. Understanding individual human mobility patterns. *Nature* **453**:779–782. doi: 10.1038/Nature06958.

Goodman JL. 2014. Studying "secret serums" - toward safe, effective Ebola treatments. *The New England Journal of Medicine*.

Gostin LO, Lucey D, Phelan A. 2014. The ebola epidemic: a global health emergency. *JAMA* E1–E2. doi: 10.1001/jama.2014.11176.

Grard G, Biek R, Tamfum JJM, Fair J, Wolfe N, Formenty P, Paweska J, Leroy E. 2011. Emergence of divergent Zaire Ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *The Journal of Infectious Diseases* **204**:S776–S784. doi: 10.1093/infdis/jir364.

Groseth A, Feldmann H, Strong JE. 2007. The ecology of Ebola virus. *Trends in Microbiology* **15**:408–416. doi: 10.1016/j.tim.2007.08.001.

Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, Brownstein JS, Collier N, Myers MF, George DB, Gething PW. 2013a. Global mapping of infectious disease. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **368**:20120250. doi: 10.1098/rstb.2012.0250.

Hay SI, George DB, Moyes CL, Brownstein JS. 2013b. Big data opportunities for global infectious disease surveillance. *PLOS Medicine* **10**:e1001413. doi: 10.1371/journal.pmed.1001413.

Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. 2006. Global environmental data for mapping infectious disease distribution. *Advances in Parasitology* **62**:37–77. doi: 10.1016/S0065-308x(05)62002-7.

Hayman DT, Emmerich P, Yu M, Wang LF, Suu-Ire R, Fooks AR, Cunningham AA, Wood JL. 2010. Long-term survival of an urban fruit bat seropositive for Ebola and Lagos Bat viruses. *PLOS ONE* **5**:e11978. doi: 10.1371/journal.pone.0011978.

Hayman DT, Yu M, Crameri G, Wang LF, Suu-Ire R, Wood JL, Cunningham AA. 2012. Ebola virus antibodies in fruit bats, Ghana, West Africa. *Emerging Infectious Diseases* **18**:1207–1209. doi: 10.3201/eid1807.111654.

Hewlett BS, Epelboin A, Hewlett BL, Formenty P. 2005. Medical anthropology and Ebola in Congo: cultural models and humanistic care. *Bulletin de la Société de Pathologie Exotique* **98**:230–236.

Heymann DL, Weisfeld JS, Webb PA, Johnson KM, Cairns T, Berquist H. 1980. Ebola hemorrhagic fever: Tandala, Zaire, 1977-1978. *The Journal of Infectious Diseases* **142**:372–376. doi: 10.1093/infdis/142.3.372.

Hijmans RJ. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93**:679–688. doi: 10.1890/11-0826.1.

Hijmans RJ, Phillips S, Leathwick J, Elith J. 2014. dismo. R package. Accessed: August 2014. Available: http://cran.r-project.org/web/packages/dismo/dismo.pdf.

Huang ZJ, Tatem AJ. 2013. Global malaria connectivity through air travel. *Malaria Journal* **12**:269. doi: 10.1186/1475-2875-12-269.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **83**:195–213. doi: 10.1016/S0034-4257(02)00096-2.

Hufnagel L, Brockmann D, Geisel T. 2004. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of USA* **101**:15124–15129. doi: 10.1073/pnas.0308344101.

IATA. 2014. International Air Transport Association. Accessed: August 2014. Available: http://www.iata.org/Pages/default.aspx.

International Commission. 1978. Ebola haemorrhagic fever in Zaire, 1976. *Bulletin of the World Health Organization* **56**:271–293.

Jahrling PB, Geisbert TW, Dalgard DW, Johnson ED, Ksiazek TG, Hall WC, Peters CJ. 1990. Preliminary report: isolation of Ebola virus from monkeys imported to USA. *Lancet* **335**:502–505. doi: 10.1016/0140-6736(90)90737-P.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. *Nature* **451**:990–994. doi: 10.1038/Nature06536.

Kamins AO, Restif O, Ntiamoa-Baidu Y, Suu-Ire R, Hayman DT, Cunningham AA, Wood JL, Rowcliffe JM. 2011. Uncovering the fruit bat bushmeat commodity chain and the true extent of fruit bat hunting in Ghana, West Africa. *Biological Conservation* **144**:3000–3008. doi: 10.1016/j.biocon.2011.09.003.

Karesh WB, Dobson A, Lloyd-Smith JO, Lubroth J, Dixon MA, Bennett M, Aldrich S, Harrington T, Formenty P, Loh EH, Machalaba CC, Thomas MJ, Heymann DL. 2012. Ecology of zoonoses: natural and unnatural histories. *Lancet* **380**:1936–1945. doi: 10.1016/S0140-6736(12)61678-X.

Khan AS, Tshioko FK, Heymann DL, Le Guenno B, Nabeth P, Kerstiëns B, Fleerackers Y, Kilmarx PH, Rodier GR, Nkuku O, Rollin PE, Sanchez A, Zaki SR, Swanepoel R, Tomori O, Nichol ST, Peters CJ, Muyembe-Tamfum JJ, Ksiazek TG. 1999. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. *The Journal of Infectious Diseases* **179**:S76–S86. doi: 10.1086/514306.

King AMQ, Adams MJ, Carstens EB, Lefkowitz E. 2012. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. London: Elsevier. p. 1338.

Kuhn JH. 2008. Filoviruses: a compendium of 40 years of epidemiological, clinical, and laboratory studies. Vienna: Springer-Verlag. p. 413.

Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, Lipkin WI, Negredo AI, Netesov SV, Nichol ST, Palacios G, Peters CJ, Tenorio A, Volchkov VE, Jahrling PB. 2010. Proposal for a revised taxonomy of the family *Filoviridae*: classification, names of taxa and viruses, and virus abbreviations. *Archives of Virology* **155**:2083–2103. doi: 10.1007/s00705-010-0814-x.

Lahm SA, Kombila M, Swanepoel R, Barnes RF. 2007. Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994-2003. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**:64–78. doi: 10.1016/j.trstmh.2006.07.002.

Lamunu M, Lutwama JJ, Kamugisha J, Opio A, Nambooze J, Ndayimirije N, Okware S. 2004. Containing a haemorrhagic fever epidemic: the Ebola experience in Uganda (October 2000-January 2001). *International Journal of Infectious Diseases* **8**:27–37. doi: 10.1016/j.ijid.2003.04.001.

Larkin M. 2003. Ebola outbreak in the news. *The Lancet Infectious Diseases* **3**:255. doi: 10.1016/S1473-3099(03)00584-X.

Le Guenno B, Formenty P, Wyers M, Gounon P, Walker F, Boesch C. 1995. Isolation and partial characterization of a new strain of Ebola virus. *Lancet* **345**:1271–1274. doi: 10.1016/S0140-6736(95)90925-7.

Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. 2007. Understanding the dynamics of Ebola epidemics. *Epidemiology and Infection* **135**:610–621. doi: 10.1017/S0950268806007217.

Leroy EM, Baize S, Volchkov VE, Fisher-Hoch SP, Georges-Courbot MC, Lansoud-Soukate J, Capron M, Debré P, McCormick JB, Georges AJ. 2000. Human asymptomatic Ebola infection and strong inflammatory response. *Lancet* **355**:2210–2215. doi: 10.1016/S0140-6736(00)02405-3.

Leroy EM, Epelboin A, Mondonge V, Pourrut X, Gonzalez JP, Muyembe-Tamfum JJ, Formenty P. 2009. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne and Zoonotic Diseases* **9**:723–728. doi: 10.1089/vbz.2008.0167.

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT, Gonzalez JP, Swanepoel R. 2005. Fruit bats as reservoirs of Ebola virus. *Nature* **438**:575–576. doi: 10.1038/438575a.

Leroy EM, Souquière S, Rouquet P, Drevet D. 2002. Re-emergence of ebola haemorrhagic fever in Gabon. *Lancet* **359**:712–712. doi: 10.1016/S0140-6736(02)07796-6.

Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLOS ONE* **7**:e31743. doi: 10.1371/journal.pone.0031743.

Linard C, Tatem AJ, Gilbert M. 2013. Modelling spatial patterns of urban growth in Africa. *Applied Geography* **44**:23–32. doi: 10.1016/j.apgeog.2013.07.009.

MacNeil A, Farnon EC, Wamala J, Okware S, Cannon DL, Reed Z, Towner JS, Tappero JW, Lutwama J, Downing R, Nichol ST, Ksiazek TG, Rollin PE. 2010. Proportion of deaths and clinical features in Bundibugyo Ebola virus infection, Uganda. *Emerging Infectious Diseases* **16**:1969–1972. doi: 10.3201/eid1612.100627.

Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. 2014. A global compendium of human dengue virus occurrence. *Scientific Data* **1**:140004. doi: 10.1038/sdata.2014.4.

Mfunda IM, Røskaft E. 2010. Bushmeat hunting in Serengeti, Tanzania: an important economic activity to local people. *International Journal of Biodiversity and Conservation* **2**:263–272.

Milleliri JM, Tevi-Benissan C, Baize S, Leroy E, Georges-Courbot MC. 2004. [Epidemics of Ebola haemorrhagic fever in Gabon (1994-2002). Epidemiologic aspects and considerations on control measures]. *Bulletin de la Société de Pathologie Exotique* **97**:199–205.

Miranda ME, White ME, Dayrit MM, Hayes CG, Ksiazek TG, Burans JP. 1991. Seroepidemiological study of filovirus related to Ebola in the Philippines. *Lancet* **337**:425–426. doi: 10.1016/0140-6736(91)91199-5.

MSF. 2014. Ebola epidemic confirmed in Democratic Republic of Congo: MSF sends specialists and material to the epicentre. Accessed: August 2014. Available: http://www.msf.org/article/ebola-epidemic-confirmed-democratic-republic-congo-msf-sends-specialists-and-material.

Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, Dansereau EA, Graetz N, Barber RM, Brown JC, Wang H, Duber HC, Naghavi M, Dicker D, Dandona L, Salomon JA, Heuton KR, Foreman K, Phillips DE, Fleming TD, Flaxman AD, Phillips BK, Johnson EK, Coggeshall MS, Abd-Allah F, Abera SF, Abraham JP, Abubakar I, Abu-Raddad LJ, Abu-Rmeileh NM, Achoki T, Adeyemo AO, Adou AK, Adsuar JC, Agardh EE, Akena D, Al Kahbouri MJ, Alasfoor D, Albittar MI, Alcalá-Cerra G, Alegretti MA, Alemu ZA, Alfonso-Cristancho R, Alhabib S, Ali R, Alla F, Allen PJ, Alsharif U, Alvarez E, Alvis-Guzman N, Amankwaa AA, Amare AT, Amini H, Ammar W, Anderson BO, Antonio CA, Anwari P, Arnlöv J, Arsenijevic VS, Artaman A, Asghar RJ, Assadi R, Atkins LS, Badawi A, Balakrishnan K, Banerjee A, Basu S, Beardsley J, Bekele T, Bell ML, Bernabe E, Beyene TJ, Bhala N, Bhalla A, Bhutta ZA, Abdulhak AB, Binagwaho A, Blore JD, Bose D, Brainin M, Breitborde N, Castañeda-Orjuela CA, Catalá-López F, Chadha VK, Chang JC, Chiang PP, Chuang TW, Colomar M, Cooper LT, Cooper C, Courville KJ, Cowie BC, Criqui MH, Dandona R, Dayama A, De Leo D, Degenhardt L, Del Pozo-Cruz B, Deribe K, Des Jarlais DC, Dessalegn M, Dharmaratne SD, Dilmen U, Ding EL, Driscoll TR, Durrani AM, Ellenbogen RG, Ermakov SP, Esteghamati A, Faraon EJ, Farzadfar F, Fereshtehnejad SM, Fijabi DO, Forouzanfar MH, Fra Paleo U, Gaffikin L, Gamkrelidze A, Gankpé FG, Geleijnse JM, Gessner BD, Gibney KB, Ginawi IA, Glaser EL, Gona P, Goto A, Gouda HN, Gugnani HC, Gupta R, Gupta R, Hafezi-Nejad N, Hamadeh RR, Hammami M, Hankey GJ, Harb HL, Haro JM, Havmoeller R, Hay SI, Hedayati MT, Pi IB, Hoek HW, Hornberger JC, Hosgood HD, Hotez PJ, Hoy DG, Huang JJ, Iburg KM, Idrisov BT, Innos K, Jacobsen KH, Jeemon P, Jensen PN, Jha V, Jiang G, Jonas JB, Juel K, Kan H, Kankindi I, Karam NE, Karch A, Karema CK, Kaul A, Kawakami N, Kazi DS, Kemp AH, Kengne AP, Keren A, Kereselidze M, Khader YS, Khalifa SE, Khan EA, Khang YH, Khonelidze I, Kinfu Y, Kinge JM, Knibbs L, Kokubo Y, Kosen S, Defo BK, Kulkarni VS, Kulkarni C, Kumar K, Kumar RB, Kumar GA, Kwan GF, Lai T, Balaji AL, Lam H, Lan Q, Lansingh VC, Larson HJ, Larsson A, Lee JT, Leigh J, Leinsalu M, Leung R, Li Y, Li Y, De Lima GM, Lin HH, Lipshultz SE, Liu S, Liu Y, Lloyd BK, Lotufo PA, Machado VM, Maclachlan JH, Magis-Rodriguez C, Majdan M, Mapoma CC, Marcenes W, Marzan MB, Masci JR, Mashal MT, Mason-Jones AJ, Mayosi BM, Mazorodze TT, Mckay AC, Meaney PA, Mehndiratta MM, Mejia-Rodriguez F, Melaku YA, Memish ZA, Mendoza W, Miller TR, Mills EJ, Mohammad KA, Mokdad AH, Mola GL, Monasta L, Montico M, Moore AR, Mori R, Moturi WN, Mukaigawara M, Murthy KS, Naheed A, Naidoo KS, Naldi L, Nangia V, Narayan KM, Nash D, Nejjari C, Nelson RG, Neupane SP, Newton CR, Ng M, Nisar MI, Nolte S, Norheim OF, Nowaseb V, Nyakarahuka L, Oh IH, Ohkubo T, Olusanya BO, Omer SB, Opio JN, Orisakwe OE, Pandian JD, Papachristou C, Caicedo AJ, Patten SB, Paul VK, Pavlin BI, Pearce N, Pereira DM, Pervaiz A, Pesudovs K, Petzold M, Pourmalek F, Qato D, Quezada AD, Quistberg DA, Rafay A, Rahimi K, Rahimi-Movaghar V, Rahman SU, Raju M, Rana SM, Razavi H, Reilly RQ, Remuzzi G, Richardus JH, Ronfani L, Roy N, Sabin N, Saeedi MY, Sahraian MA, Samonte GM, Sawhney M, Schneider IJ, Schwebel DC, Seedat S, Sepanlou SG, Servan-Mori EE, Sheikhbahaei S, Shibuya K, Shin HH, Shiue I, Shivakoti R, Sigfusdottir ID, Silberberg DH, Silva AP, Simard EP, Singh JA, Skirbekk V, Sliwa K, Soneji S, Soshnikov SS, Sreeramareddy CT, Stathopoulou VK, Stroumpoulis K, Swaminathan S, Sykes BL, Tabb KM, Talongwa RT, Tenkorang EY, Terkawi AS, Thomson AJ, Thorne-Lyman AL, Towbin JA, Traebert J, Tran BX, Dimbuene ZT, Tsilimbaris M, Uchendu US, Ukwaja KN, Vallely AJ, Vasankari TJ, Venketasubramanian N, Violante FS, Vlassov VV, Waller S, Wallin MT, Wang L, Wang SX, Wang Y, Weichenthal S, Weiderpass E, Weintraub RG, Westerman R, White RA, Wilkinson JD, Williams TN, Woldeyohannes SM, Wong JQ, Xu G, Yang YC, Yano Y, Yip P, Yonemoto N, Yoon SJ, Younis M, Yu C, Jin KY, El Sayed Zaki M, Zhao Y, Zheng Y, Zhou M, Zhu J, Zou XN, Lopez AD, Vos T. 2014. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. doi: 10.1016/S0140-6736(14)60844-8.

Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fèvre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gonzalez-Medina D, Gosselin R, Grainger R, Grant B, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laden F, Lalloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Levinson D, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mock C, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Nevitt MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA III, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leòn FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivakoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ, Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiebe N, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD, AlMazroa MA, Memish ZA. 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**: 2197–2223. doi: 10.1016/S0140-6736(12)61689-4.

Muyembe T, Kipasa M. 1995. Ebola haemorrhagic fever in Kikwit, Zaire. International Scientific and Technical Committee and WHO Collaborating Centre for Haemorrhagic Fevers. *Lancet* **345**:1448. doi: 10.1016/S0140-6736(95)92640-2.

NCBI. 2014. GenBank - ebola. Accessed: August 2014. Available: http://www.ncbi.nlm.nih.gov/genbank/.

Negredo A, Palacios G, Vazquez-Moron S, Gonzalez F, Dopazo H, Molero F, Juste J, Quetglas J, Savji N, de la Cruz Martínez M, Herrera JE, Pizarro M, Hutchison SK, Echevarría JE, Lipkin WI, Tenorio A. 2011. Discovery of an ebolavirus-like filovirus in Europe. *PLOS Pathogens* **7**:e1002304. doi: 10.1371/journal.ppat.1002304.

NIH. 2014. NIAID emerging infectious diseases/pathogens. Available: Accessed: August 2014.

Nkoghe D, Formenty P, Leroy EM, Nnegue S, Edou SY, Ba JI, Allarangar Y, Cabore J, Bachy C, Andraghetti R, de Benoist AC, Galanis E, Rose A, Bausch D, Reynolds M, Rollin P, Choueibou C, Shongo R, Gergonne B, Koné LM, Yada A, Roth C, Mve MT. 2005. [Multiple Ebola virus haemorrhagic fever outbreaks in Gabon, from October 2001 to April 2002]. *Bulletin de la Société de Pathologie Exotique* **98**:224–229.

Nkoghe D, Kone ML, Yada A, Leroy E. 2011. A limited outbreak of Ebola haemorrhagic fever in Etoumbi, Republic of Congo, 2005. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **105**:466–472. doi: 10.1016/j.trstmh.2011.04.011.

Okware SI, Omaswa FG, Zaramba S, Opio A, Lutwama JJ, Kamugisha J, Rwaguma EB, Kagwa P, Lamunu M. 2002. An outbreak of Ebola in Uganda. *Tropical Medicine & International Health* **7**:1068–1075. doi: 10.1046/j.1365-3156.2002.00944.x.

Olival KJ, Hayman DT. 2014. Filoviruses in bats: current knowledge and future directions. *Viruses* **6**:1759–1788. doi: 10.3390/V6041759.

Onyango CO, Opoka ML, Ksiazek TG, Formenty P, Ahmed A, Tukei PM, Sang RC, Ofula VO, Konongoi SL, Coldren RL, Grein T, Legros D, Bell M, De Cock KM, Bellini WJ, Towner JS, Nichol ST, Rollin PE. 2007. Laboratory diagnosis of Ebola hemorrhagic fever during an outbreak in Yambio, Sudan, 2004. *The Journal of Infectious Diseases* **196**:S193–S198. doi: 10.1086/520609.

ORNL DAAC. 2000. Shuttle radar topography mission near-global digital elevation models. Accessed: August 2014. Available: http://webmap.ornl.gov/wcsdown/wcsdown.jsp?dg_id=10008_1.

Pattyn S, Jacob W, Vandergroen G, Piot P, Courteille G. 1977. Isolation of Marburg-like virus from a case of hemorrhagic fever in Zaire. *Lancet* **1**:573–574. doi: 10.1016/S0140-6736(77)92002-5.

Peterson AT, Bauer JT, Mills JN. 2004a. Ecologic and geographic distribution of filovirus disease. *Emerging Infectious Diseases* **10**:40–47. doi: 10.3201/eid1001.030125.

Peterson AT, Carroll DS, Mills JN, Johnson KM. 2004b. Potential mammalian filovirus reservoirs. *Emerging Infectious Diseases* **10**:2073–2081. doi: 10.3201/eid1012.040346.

Peterson AT, Papes M, Carroll DS, Leirs H, Johnson KM. 2007. Mammal taxa constituting potential coevolved reservoirs of filoviruses. *Journal of Mammalogy* **88**:1544–1554. doi: 10.1644/06-Mamm-a-280r1.1.

Petter JJ, Desbordes F. 2013. *Primates of the world*. Princeton: Princeton University Press. p. 186.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**:181–197. doi: 10.1890/07-2153.1.

Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld C, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI. 2014. Global distribution maps of the leishmaniases. *eLife* e02851. doi: 10.7554/eLife.02851.

Pindolia DK, Garcia AJ, Huang ZJ, Fik T, Smith DL, Tatem AJ. 2014. Quantifying cross-border movements and migrations for guiding the strategic planning of malaria control and elimination. *Malaria Journal* **13**:169. doi: 10.1186/1475-2875-13-169.

Pourrut X, Delicat A, Rollin PE, Ksiazek TG, Gonzalez JP, Leroy EM. 2007. Spatial and temporal patterns of Zaire ebolavirus antibody prevalence in the possible reservoir bat species. *The Journal of Infectious Diseases* **196**:S176–S183. doi: 10.1086/520541.

Pourrut X, Kumulungui B, Wittmann T, Moussavou G, Délicat A, Yaba P, Nkoghe D, Gonzalez JP, Leroy EM. 2005. The natural history of Ebola virus in Africa. *Microbes and Infection* **7**:1005–1014. doi: 10.1016/j.micinf.2005.04.006.

Pourrut X, Souris M, Towner JS, Rollin PE, Nichol ST, Gonzalez JP, Leroy E. 2009. Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in Rousettus aegyptiacus. *BMC Infectious Diseases* **9**:e159. doi: 10.1186/1471-2334-9-159.

Rouquet P, Froment JM, Bermejo M, Kilbourn A, Karesh W, Reed P, Kumulungui B, Yaba P, Délicat A, Rollin PE, Leroy EM. 2005. Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001-2003. *Emerging Infectious Diseases* **11**:283–290. doi: 10.3201/eid1102.040533.

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, Lamoreux J, Rodrigues AS, Stuart SN, Temple HJ, Baillie J, Boitani L, Lacher TE Jnr, Mittermeier RA, Smith AT, Absolon D, Aguiar JM, Amori G, Bakkour N, Baldi R, Berridge RJ, Bielby J, Black PA, Blanc JJ, Brooks TM, Burton JA, Butynski TM, Catullo G, Chapman R, Cokeliss Z, Collen B, Conroy J, Cooke JG, da Fonseca GA, Derocher AE, Dublin HT, Duckworth JW, Emmons L, Emslie RH, Festa-Bianchet M, Foster M, Foster S, Garshelis DL, Gates C, Gimenez-Dixon M, Gonzalez S, Gonzalez-Maya JF, Good TC, Hammerson G, Hammond PS, Happold D, Happold M, Hare J, Harris RB, Hawkins CE, Haywood M, Heaney LR, Hedges S, Helgen KM, Hilton-Taylor C, Hussain SA, Ishii N, Jefferson TA, Jenkins RK, Johnston CH, Keith M, Kingdon J, Knox DH, Kovacs KM, Langhammer P, Leus K, Lewison R, Lichtenstein G, Lowry LF, Macavoy Z, Mace GM, Mallon DP, Masi M, McKnight MW, Medellín RA, Medici P, Mills G, Moehlman PD, Molur S, Mora A, Nowell K, Oates JF, Olech W, Oliver WR, Oprea M, Patterson BD, Perrin WF, Polidoro BA, Pollock C, Powel A, Protas Y, Racey P, Ragle J, Ramani P, Rathbun G, Reeves RR, Reilly SB, Reynolds JE III, Rondinini C, Rosell-Ambal RG, Rulli M, Rylands AB, Savini S, Schank CJ, Sechrest W, Self-Sullivan C, Shoemaker A, Sillero-Zubiri C, De Silva N, Smith DE, Srinivasulu C, Stephenson PJ, van Strien N, Talukdar BK, Taylor BL, Timmins R, Tirira DG, Tognelli MF, Tsytsulina K, Veiga LM, Vié JC, Williamson EA, Wyatt SA, Xie Y, Young BE. 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**:225–230. doi: 10.1126/science.1165115.

Seto KC, Guneralp B, Hutyra LR. 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences of USA* **109**:16083–16088. doi: 10.1073/pnas.1211658109.

Shoemaker T, MacNeil A, Balinandi S, Campbell S, Wamala JF, McMullan LK, Downing R, Lutwama J, Mbidde E, Ströher U, Rollin PE, Nichol ST. 2012. Reemerging Sudan ebola virus disease in Uganda, 2011. *Emerging Infectious Diseases* **18**:1480–1483. doi: 10.3201/eid1809.111536.

Siegert R, Shu H-L, Slenczka W, Peters D, Mueller G. 1967. Zur Ätiologie einer unbekannten, von Affen ausgegangenen menschlichen Infektionskrankheit. *Deutsche Medizinische Wochenschrift* **92**:2341–2343. doi: 10.1055/s-0028-1106144.

Simini F, Gonzalez MC, Maritan A, Barabasi AL. 2012. A universal model for mobility and migration patterns. *Nature* **484**:96–100. doi: 10.1038/nature10856.

Simini F, Maritan A, Neda Z. 2013. Human mobility in a continuum approach. *PLOS ONE* **8**:e60069. doi: 10.1371/journal.pone.0060069.

Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Burkot TR, Harbach RE, Hay SI. 2012. A global map of dominant malaria vectors. *Parasites & Vectors* **5**:69. doi: 10.1186/1756-3305-5-69.

Smith DH, Isaacson M, Johnson KM, Bagshawe A, Johnson BK, Swanapoel R, Killey M, Siongok T, Keruga WK. 1982. Marburg virus disease in Kenya. *Lancet* **1**:816–820. doi: 10.1016/S0140-6736(82)91871-2.

Stockwell D, Peters D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**:143–158. doi: 10.1080/136588199241391.

Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Soldan VP, Kochel TJ, Kitron U, Elder JP, Scott TW. 2009. The role of human movement in the transmission of vector-borne pathogens. *PLOS Neglected Tropical Diseases* **3**:e481. doi: 10.1371/journal.pntd.0000481.

Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Nourlil J, Faouzi A, Echevarría JE, Vazquez Morón S, Rambaut A, Campiz N, Tatem AJ, Holmes EC, Bourhy H. 2010. Phylodynamics and human-mediated dispersal of a zoonotic virus. *PLOS Pathogens* **6**:e1001166. doi: 10.1371/journal.ppat.1001166.

Tatem AJ, Goetz SJ, Hay SI. 2004. Terra and aqua: new data for epidemiology and public health. *International Journal of Applied Earth Observation and Geoinformation* **6**:33–46. doi: 10.1016/j.jag.2004.07.001.

Tignor GH, Casals J, Shope RE. 1993. The Yellow Fever epidemic in Ethiopia, 1961-1962-retrospective serological evidence for concomitant Ebola or Ebola-like virus infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **87**:162–162. doi: 10.1016/0035-9203(93)90471-2.

Towner JS, Amman BR, Sealy TK, Carroll SAR, Comer JA, Kemp A, Swanepoel R, Paddock CD, Balinandi S, Khristova ML, Formenty PB, Albarino CG, Miller DM, Reed ZD, Kayiwa JT, Mills JN, Cannon DL, Greer PW, Byaruhanga E, Farnon EC, Atimnedi P, Okware S, Katongole-Mbidde E, Downing R, Tappero JW, Zaki SR, Ksiazek TG, Nichol ST, Rollin PE. 2009. Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. *PLOS Pathogens* **5**:e1000536. doi: 10.1371/journal.ppat.1000536.

Towner JS, Khristova ML, Sealy TK, Vincent MJ, Erickson BR, Bawiec DA, Hartman AL, Comer JA, Zaki SR, Ströher U, Gomes da Silva F, del Castillo F, Rollin PE, Ksiazek TG, Nichol ST. 2006. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *Journal of Virology* **80**: 6497–6516. doi: 10.1128/JVI.00069-06.

Towner JS, Sealy TK, Khristova ML, Albarino CG, Conlan S, Reeder SA, Quan PL, Lipkin WI, Downing R, Tappero JW, Okware S, Lutwama J, Bakamutumaho B, Kayiwa J, Comer JA, Rollin PE, Ksiazek TG, Nichol ST. 2008. Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLOS Pathogens* **4**:e1000212. doi: 10.1371/journal.ppat.1000212.

Trabucco A, Zomer RJ. 2009. Global Aridity Index (Global-Aridity) and global potential Evapo-Transpiration (Global-PET) Geospatial database. Accessed: August 2014. Available: http://www.csi.cgiar.org.

Walsh PD, Bermejo M, Rodriguez-Teijeiro JD. 2009. Disease avoidance and the evolution of primate social connectivity: Ebola, bats, gorillas, and chimpanzees. In: Huffman MA, Chapman CA, editors. *Primate parasite ecology: the dynamics and study of host-parasite relationships*. Cambridge: Cambridge University Press. p. 183–197.

Wamala JF, Lukwago L, Malimbo M, Nguku P, Yoti Z, Musenero M, Amone J, Mbabazi W, Nanyunja M, Zaramba S, Opio A, Lutwama JJ, Talisuna AO, Okware SI. 2010. Ebola hemorrhagic fever associated with novel virus strain, Uganda, 2007-2008. *Emerging Infectious Diseases* **16**:1087–1092. doi: 10.3201/eid1607.091525.

Wan ZM, Li ZL. 1997. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote Sensing* **35**:980–996. doi: 10.1109/36.602541.

Wang H, Liddell CA, Coates MM, Mooney MD, Levitz CE, Schumacher AE, Apfel H, Iannarone M, Phillips B, Lofgren KT, Sandar L, Dorrington RE, Rakovac I, Jacobs TA, Liang X, Zhou M, Zhu J, Yang G, Wang Y, Liu S, Li Y, Ozgoren AA, Abera SF, Abubakar I, Achoki T, Adelekan A, Ademi Z, Alemu ZA, Allen PJ, Almazroa MA, Alvarez E, Amankwaa AA, Amare AT, Ammar W, Anwari P, Cunningham SA, Asad MM, Assadi R, Banerjee A, Basu S, Bedi N, Bekele T, Bell ML, Bhutta Z, Blore J, Basara BB, Boufous S, Breitborde N, Bruce NG, Bui LN, Carapetis JR, Cárdenas R, Carpenter DO, Caso V, Castro RE, Catalá-Lopéz F, Cavlin A, Che X, Chiang PP, Chowdhury R, Christophi CA, Chuang TW, Cirillo M, da Costa Leite I, Courville KJ, Dandona L, Dandona R, Davis A, Dayama A, Deribe K, Dharmaratne SD, Dherani MK, Dilmen U, Ding EL, Edmond KM, Ermakov SP, Farzadfar F, Fereshtehnejad SM, Fijabi DO, Foigt N, Forouzanfar MH, Garcia AC, Geleijnse JM, Gessner BD, Goginashvili K, Gona P, Goto A, Gouda HN, Green MA, Greenwell KF, Gugnani HC, Gupta R, Hamadeh RR, Hammami M, Harb HL, Hay S, Hedayati MT, Hosgood HD, Hoy DG, Idrisov BT, Islami F, Ismayilova S, Jha V, Jiang G, Jonas JB, Juel K, Kabagambe EK, Kazi DS, Kengne AP, Kereselidze M, Khader YS, Khalifa SE, Khang YH, Kim D, Kinfu Y, Kinge JM, Kokubo Y, Kosen S, Defo BK, Kumar GA, Kumar K, Kumar RB, Lai T, Lan Q, Larsson A, Lee JT, Leinsalu M, Lim SS, Lipshultz SE, Logroscino G, Lotufo PA, Lunevicius R, Lyons RA, Ma S, Mahdi AA, Marzan MB, Mashal MT, Mazorodze TT, McGrath JJ, Memish ZA, Mendoza W, Mensah GA, Meretoja A, Miller TR, Mills EJ, Mohammad KA, Mokdad AH, Monasta L, Montico M, Moore AR, Moschandreas J, Msemburi WT, Mueller UO, Muszynska MM, Naghavi M, Naidoo KS, Narayan KV, Nejjari C, Ng M, de Dieu Ngirabega J, Nieuwenhuijsen MJ, Nyakarahuka L, Ohkubo T, Omer SB, Caicedo AJ, Wyk VP, Pope D, Prabhakaran D, Rahman SU, Rana SM, Reilly RQ, Rojas-Rueda D, Ronfani L, Rushton L, Saeedi MY, Salomon J, Sampson U, Santos IS, Sawhney M, Schmidt JC, Shakh-Nazarova M, She J, Sheikhbahaei S, Shibuya K, Shin HH, Shishani K, Shiue I, Sigfusdottir ID, Singh JA, Skirbekk V, Sliwa K, Soshnikov SS, Sposato LA, Stathopoulou VK, Stroumpoulis K, Tabb KM, Talongwa RT, Teixeira CM, Terkawi AS, Thomson AJ, Thorne-Lyman AL, Toyoshima H, Dimbuene ZT, Uwaliraye P, Uzun SB, Vasankari TJ, Vasconcelos AM, Vlassov VV, Vollset SE, Vos T, Waller S, Wan X, Weichenthal S, Weiderpass E, Weintraub RG, Westerman R, Wilkinson JD, Williams HC, Yang YC, Yentur GK, Yip P, Yonemoto N, Younis M, Yu C, Jin KY, El Sayed Zaki M, Zhu S, Lopez AD, Murray CJ. 2014. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*:doi: 10.1016/S0140-6736(14)60497-9.

Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. 2014a. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing* accepted manuscript.

Weiss DJ, Bhatt S, Mappin B, Van Boeckel TP, Smith DL, Hay SI, Gething PW. 2014b. Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malaria Journal* **13**:171. doi: 10.1186/1475-2875-13-171.

Wenger SJ, Olden JD. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* **3**:260–267. doi: 10.1111/j.2041-210X.2011.00170.x.

Wesolowski A, Buckee CO, Pindolia DK, Eagle N, Smith DL, Garcia AJ, Tatem AJ. 2013. The use of census migration data to approximate human movement patterns across temporal scales. *PLOS ONE* **8**:e52971. doi: 10.1371/journal.pone.0052971.

Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. 2012. Quantifying the impact of human mobility on malaria. *Science* **338**:267–270. doi: 10.1126/science.1223467.

WHO. 2001. Outbreak of Ebola haemorrhagic fever, Uganda, August 2000-January 2001. *Weekly Epidemiological Record* **76**:41–46.

WHO. 2003a. Outbreak(s) of Ebola haemorrhagic fever, Congo and Gabon, October 2001-July 2002. *Weekly Epidemiological Record* **78**:223–228.

WHO. 2003b. Outbreak(s) of Ebola haemorrhagic fever in the Republic of the Congo, January-April 2003. *Weekly Epidemiological Record* **78**:285–289.

WHO. 2005. Outbreak of Ebola haemorrhagic fever in Yambio, south Sudan, April - June 2004. *Weekly Epidemiological Record* **80**:370–375.

WHO. 2012a. Ebola in Uganda - 29 July 2012. Accessed: August 2014. Available: http://www.who.int/csr/don/2012_07_29/en/.

WHO. 2012b. Ebola outbreak in democratic Republic of Congo - 17 August 2012. Accessed: August 2014. Available: http://www.who.int/csr/don/2012_08_18/en/.

WHO. 2012c. Ebola in Uganda - 17 November 2012. Accessed: August 2014. Available: http://www.who.int/csr/don/2012_11_17/en/.

WHO. 2014a. Ebola virus disease update - west Africa. Accessed: August 2014. Available: http://www.who.int/csr/don/2014_08_15_ebola/en/.

WHO. 2014b. Health statistics and information systems - definition of region groupings. Accessed: August 2014. Available: http://www.who.int/healthinfo/global_burden_disease/definition_regions/en/.

WHO. 2014c. *Interim manual - Ebola and Marburg virus disease epidemics: preparedness, alert, control, and evaluation*. Geneva: World Health Organization. P. 123.

WHO. 2014d. WHO statement on the meeting of the International Health Regulations Emergency Committee regarding the 2014 Ebola outbreak in West Africa. Accessed: August 2014. Available: http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/.

WHO International Study Team. 1978. Ebola haemorrhagic fever in Sudan, 1976. *Bulletin of the World Health Organization* **56**:247–270.

Wittmann TJ, Biek R, Hassanin A, Rouquet P, Reed P, Yaba P, Pourrut X, Real LA, Gonzalez JP, Leroy EM. 2007. Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. *Proceedings of the National Academy of Sciences of USA* **104**:17123–17127. doi: 10.1073/pnas.0704076104.

Wolfe ND, Daszak P, Kilpatrick AM, Burke DS. 2005. Bushmeat Hunting, deforestation, and prediction of zoonotic disease emergence. *Emerging Infectious Diseases* **11**:1822–1827. doi: 10.3201/eid1112.040789.

Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. *Nature* **447**:279–283. doi: 10.1038/Nature05775.

World Bank. 2014. World Bank open dataset. Accessed: July 2014. Available: http://data.worldbank.org/.

WorldPop. 2014. WorldPop project. Accessed: August 2014. Available: http://www.worldpop.org.uk/.

WWF. 2014. List of terrestrial ecoregions. Accessed: August 2014. Available: http://wwf.panda.org/about_our_earth/ecoregions/ecoregion_list/.

Yang Y, Atkinson P, Ettema D. 2008. Individual space-time activity-based modelling of infectious disease transmission within a city. *Journal of the Royal Society, Interface* **5**:759–772. doi: 10.1098/rsif.2007.1218.

Zipf GK. 1946. The P_1 P_2/D hypothesis: on the intercity movement of persons. *American Sociological Review* **11**:677–686. doi: 10.2307/2087063.

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Viral infection

» Ecological epidemiology

» Ebola virus

A comprehensive database of the geographic spread of past human Ebola outbreaks

Adrian Mylne[1,*], Oliver J. Brady[1,*], Zhi Huang[1], David M. Pigott[1], Nick Golding[1], Moritz U.G. Kraemer[1] & Simon I. Hay[1,2]

Received: 15 September 2014
Accepted: 06 October 2014
Published: 23 October 2014

Ebola is a zoonotic filovirus that has the potential to cause outbreaks of variable magnitude in human populations. This database collates our existing knowledge of all known human outbreaks of Ebola for the first time by extracting details of their suspected zoonotic origin and subsequent human-to-human spread from a range of published and non-published sources. In total, 22 unique Ebola outbreaks were identified, composed of 117 unique geographic transmission clusters. Details of the index case and geographic spread of secondary and imported cases were recorded as well as summaries of patient numbers and case fatality rates. A brief text summary describing suspected routes and means of spread for each outbreak was also included. While we cannot yet include the ongoing Guinea and DRC outbreaks until they are over, these data and compiled maps can be used to gain an improved understanding of the initial spread of past Ebola outbreaks and help evaluate surveillance and control guidelines for limiting the spread of future epidemics.

Design Type(s)	observation design • epidemiological study • data integration
Measurement Type(s)	Viral Epidemiology
Technology Type(s)	data collection method
Factor Type(s)	emergence and spread locations
Sample Characteristic(s)	Ebolavirus • South Sudan • Sudan • Congo, the Democratic Republic of the • Cote d'Ivoire • Gabon • Uganda • Congo • anthropogenic habitat

[1]Spatial Ecology & Epidemiology Group, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK. [2]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA. *Joint first authors. Correspondence and requests for materials should be addressed to S.I.H. (email: simon.hay@zoo.ox.ac.uk)

Background & Summary

The genus *Ebolavirus* belongs to the family *Filoviridae* and contains five known species to date that vary in their distribution, reservoir hosts and their pathogenicity to humans. With the exception of *Reston ebolavirus* which has only shown pathogenicity among primates and porcids, all of these have shown some capacity to spill over from their natural reservoirs and cause human cases[1,2]. While only a single human case of *Tai Forest ebolavirus* has been documented[3], the remaining three species (*Zaire ebolavirus*, *Sudan ebolavirus* and *Bundibugyo ebolavirus*) are all known to permit human to human transmission after the initial suspected zoonotic transfer resulting in outbreaks of different sizes, geographic extents and case fatality rates (Fig. 1)[4,5].

Which reservoir species are responsible for maintaining Ebolavirus transmission between outbreaks is not well understood[6], but several candidate species have been identified. In Gabon three species of bats (*Hypsignathus monstrosus*, *Epomops franqueti* and *Myonycteris torquata*) were found to be infected with Ebola virus[7] and some human outbreaks have been directly linked to bat exposure[8]. While it is increasingly clear that gorillas (*Gorilla gorilla*) and chimpanzees (*Pan troglodytes*) are dead end hosts for the virus, infection in populations of these species is frequently found and they still present a risk of animal to human transmission[3,9–14]. For many outbreaks it has been difficult to definitively identify the source of human Ebola index cases, but activities that bring humans into close contact with the blood of mammals through activities such as hunting and the bushmeat trade are common to many of the index cases[8,14–16].

Following the initial suspected zoonotic transfer, secondary transmission can result from close contact between infectious individuals or corpses and other humans, usually through exposure to infectious bodily fluids[2]. Due to the close degree of contact required for secondary transmission, certain specific community activities are commonly associated with hotspots of secondary transmission such as family home care, traditional burial practices that involve washing the corpse, or healthcare settings where sufficient protective measures are not in place[15,17,18]. These focal transmission events combined with an incubation period of 5–9 days mean transmission can often be observed in waves of cases[2,18] in both space and time.

Spread of cases over longer distances is often associated with treatment seeking that draws people from rural villages that typify the index case locations to big urban centres with central medical facilities (Supplementary Figures 1–22). While this mostly involves domestic land travel[19], some instances of international importation by air travel have been documented[18]. Following travel of an infectious individual, either secondary clusters of Ebola cases will occur, or transmission will be interrupted by control methods such as quarantine and patient contact tracing[20,21]. Due to the variable rate of progression of symptoms of Ebola virus disease (EVD) (onset of Ebola Haemorrhagic fever can range from 2–21 days[2]), case fatality rates can vary significantly depending on a number of factors associated

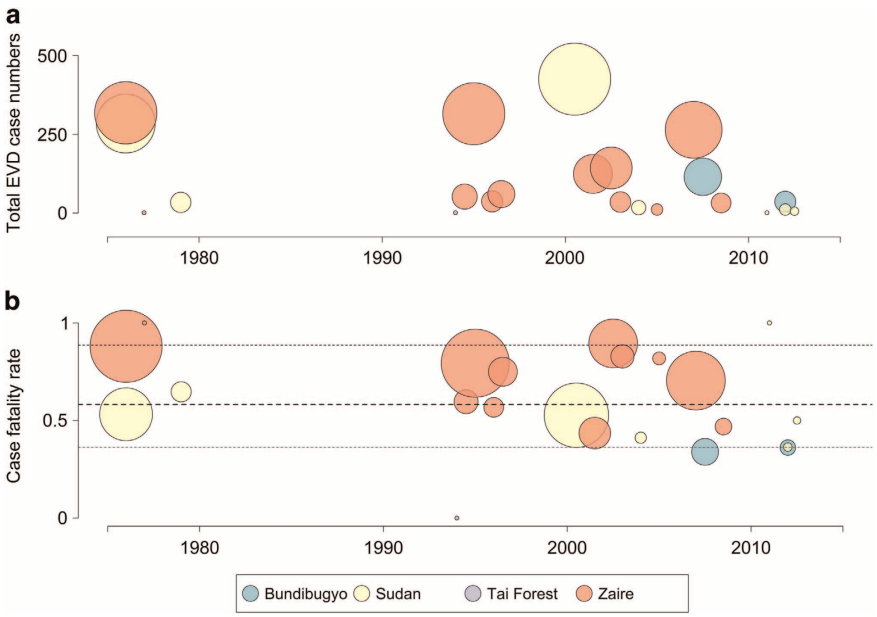


Figure 1. The size (**a**) and case fatality rate (**b**) of the 22 previous Ebola outbreaks (suspected and confirmed cases). Circle area is proportional to the total number of reported cases (**a**) or deaths (**b**) for each outbreak. Circle colour represents different species of Ebolavirus. Black dotted lines in (**b**) show the median and upper and lower 75% quantiles of outbreak case fatality rate.

with Ebola virus pathogenesis and the quality and timing of symptomatic care. The complex interaction between surveillance, control, treatment seeking, patient-contact rates and their combined effects on the dynamics of transmission dictate the spread, magnitude and case fatality rate of an Ebola outbreak.

This database collates existing knowledge on the geographic spread of past Ebola outbreaks in a standardised format that allows the dynamics of different outbreaks to be compared. Procedures for data abstraction are outlined and each outbreak is summarised with a map and brief text description. These data will be useful for conducting spatial analyses of Ebola outbreak spread. We include every outbreak preceding the atypical 2013 Guinea epidemic which has spread further and faster than any previous epidemic. Once the current Guinea and Democratic Republic of the Congo (DRC) outbreaks are over, this database will be updated to include the same standardised data fields for these contemporary outbreaks. Periodic updates to include any additional Ebola outbreaks will also ensure this resource has on-going relevance in Ebola spread analyses. In particular, a comparison between the Guinea 2013 outbreak and historical outbreaks will allow an evaluation of surveillance and control guidelines in terms of their appropriateness for mitigating the spread of future Ebola outbreaks of variable magnitude. In the meantime, it is hoped that these data will support research into EVD epidemiology which can be brought to bear on the current outbreak.

Methods

Data collection

Tables listing proven outbreaks of Ebola virus, sourced from the scientific literature[5] and from health reporting organisations[22], were used to coordinate initial searches of the formal scientific literature using Web of Science and PubMed for each specific outbreak. Relevant papers were abstracted and, where possible, outbreak-specific epidemiological surveys were sourced. The citations in these references were obtained in order to reconstruct the outbreak in detail and extract a range of epidemiological data relating to geographic spread, case and fatality numbers.

In this analysis we excluded ongoing outbreaks meaning that the current Guinea and DRC spread data are not yet included. When this outbreak is over and data has been assimilated we will update this database to incorporate the Guinea and DRC outbreaks using the same procedures described below, which will allow a comparison with historical outbreaks of Ebola.

Definitions of index, secondary and imported cases

Index cases were defined as any human infection resulting from interaction with non-human sources of the disease. Among the sources examined, index cases were identified based on reported interaction with suspected zoonotic reservoirs or hosts, such as non-human mammals during hunting trips[8,14–16,23]. In cases where a mode of suspected zoonotic transmission could not be established, the first reported case was assumed to be the index case. Any cases arising from existing human infections are considered as secondary infections. Cases reported after the index cases were assumed to be secondary cases unless they were accompanied by specific details of likely exposure to a zoonotic reservoir or non-human host. If a case was reported in an area where no further cases occurred and no continued transmission was documented, these were termed imported cases.

Procedure for geo-positioning

For each index, secondary or imported case cluster that could be linked to a unique geographic location we performed a range of procedures similar to methods employed elsewhere[24,25] to assign geographic coordinates. For index and secondary cases these locations were representative of the site of suspected zoonotic transfer or human-to-human transmission respectively. Index cases were geopositioned as the location where exposure to the suspected zoonotic reservoir was likely to be highest. For example hunters who butchered carcases on hunting expeditions were geopositioned as a polygon covering the area of the hunting trip not a point at the hunters' homes. In contrast, if an individual purchased bushmeat from a local market for preparation and consumption at home, the home of the individual was georeferenced as the index case, as it was considered the location of highest exposure to the suspected zoonotic reservoir. For imported cases these locations were representative of all locations that infected patients travelled to, but did not cause onward transmission. For the purpose of this analysis we excluded international imported cases that did not cause autochthonous secondary transmission as, in most cases, they represented diagnosed foreign workers or expatriates who were evacuated for specialised treatment[3,10]. In such circumstances appropriate preventative measures are employed meaning that risk of onward transmission is negligible.

Each occurrence was reported to the highest degree of spatial resolution available based upon the information provided, as long as they could be categorised into one of: index, secondary or imported cases. This ranged from point locations (indicative of a precise location, such as a village), to areas, termed polygon locations, which correspond approximately to administrative regions or custom digitised areas based on site descriptors within the primary articles. Administrative regions were defined as classified by the Food and Agriculture Organization's Global Administrative Unit Layers (GAUL) coding[26]. These classify national boundaries as admin0 units, states or provinces as admin1 units and districts as admin2 units. By classifying Ebola occurrences as polygons we were able to represent the geographic uncertainty around the exact location of Ebola transmission which could have occurred anywhere within the defined

region. For towns, the coordinates of the centre were recorded, unless a specific part of the town (or an explicit latitude and longitude) was described. Coordinates for point locations were extracted using Google Earth (version 7.1.2). If the area concerned could not be assigned to a finer resolution than 5 km × 5 km, it was entered as polygon rather than point data. If specified regions could not be linked to an admin2 or admin1 unit, custom polygons were digitised using site descriptions in the text articles. For imprecise descriptors e.g., '15 km from the town' with no direction specified or 'cases occurred on a north-south road between village X and Y', circular polygons were digitised based on radius distances given or extreme points that defined the geographic limits of transmission. These circular polygons could be trimmed if their area included admin1 or admin2 administrative regions which reported no EVD patients. Some articles referred to 'healthcare districts' that did not correspond to admin1 or admin2 units, but were definable based on maps presented in the primary literature that were digitised or were available on the map sharing website IKI (www.ikimap.com)[27]. For index cases that referred to suspected zoonotic transfer in specific forests or game reserves, polygons were drawn based on the specified park or forest geographic boundary as shown in Google Earth. Two exceptional cases were present. In the first, for outbreak 13, the index case transmission site description merely mentioned a case being reported as 'near the town of Mbandza'. In this instance a circular polygon was defined with radius of half the distance to the next specified location of transmission (7.5 km)[15]. In the second, for outbreak 9, two locations described in the primary literature could not be located, but were described as 'near to the town of Booue'[18]. As a result the same procedure was undertaken and a radius of 30 km was defined around the village of Booue. All digitising was performed using ArcGIS 10.1[28].

Key outbreak metrics recorded

The total number of cases for each outbreak was obtained from the most recent primary source (Table 1 (available online only)). Cases included both clinically suspected and laboratory confirmed cases at the point of care, or diagnosed retrospectively. The number of people who died with a suspected or confirmed diagnosis of EVD was also recorded. These data were spatially disaggregated as much as possible from the information given in the text to give measures of spatial variation in case fatality rate within an outbreak.

Outbreak start and end time were defined by reports of the first (index) and last cases respectively. For each secondary and imported case that occurred in a novel location, the source of importation was recorded where reported. When possible this was reconciled with the date of the first secondary or imported case in each cluster to determine geographic spread order. We only included confirmed sources of origin, as determined by detailed patient histories, not suspected sources and no differentiation of spread order to two new clusters from the same source was made unless both new clusters documented the date of their first secondary case. If two secondary case clusters had a confirmed source but no differentiation of order we assigned them the same spread order. Similarly if a secondary case cluster could have come from two specified sources both were included as sources of spread. Due to a lack of epidemiological investigation at the time, or through information being lost in the reporting chain, the spread order of many secondary and imported case clusters was partially complete or missing altogether (Table 2).

For each outbreak the above details were combined with additional information from the original articles on method and timing of spread to construct the text descriptions accompanying the spread maps in Supplementary Figures 1–22 (selected examples in Figs 2,3,4). Spread is defined as movement of an infected individual from the location of infection to a transmission-free area. For example an individual who was likely infected from the suspected zoonotic reservoir on a hunting expedition in the forest then caused secondary infections back in their local village, would qualify as having spread the virus from the forest to the village. By contrast if the first reported case occurred in a village where subsequent secondary cases then occurred, no spread would have been recorded as index and secondary transmission occur in the same location.

	Spread order (%)	Source (%)	Onset timing (%)	End timing (%)	Case data (%)
Complete	51	69	32	6	27
Partial	44	2	41	0	21
Unknown	5	29	27	94	52

Table 2. Completeness of epidemiological details among the 117 Ebola virus transmission occurrences. Spread order is considered complete if the order of each occurrence in their respective outbreaks is known, partial if their order could not be distinguished from every occurrence in each outbreak and unknown if their order was unknown in the outbreak. Source was considered complete if spread could be linked to one source occurrence, partial if spread is known to have come from one of a number of source occurrences and unknown if source was unidentifiable (index cases excluded). Both onset and end timing were considered complete if day, date and year was known, partial if just month and year was known. Case data was considered complete if the number of cases and deaths was known, partial if either cases or deaths was known.

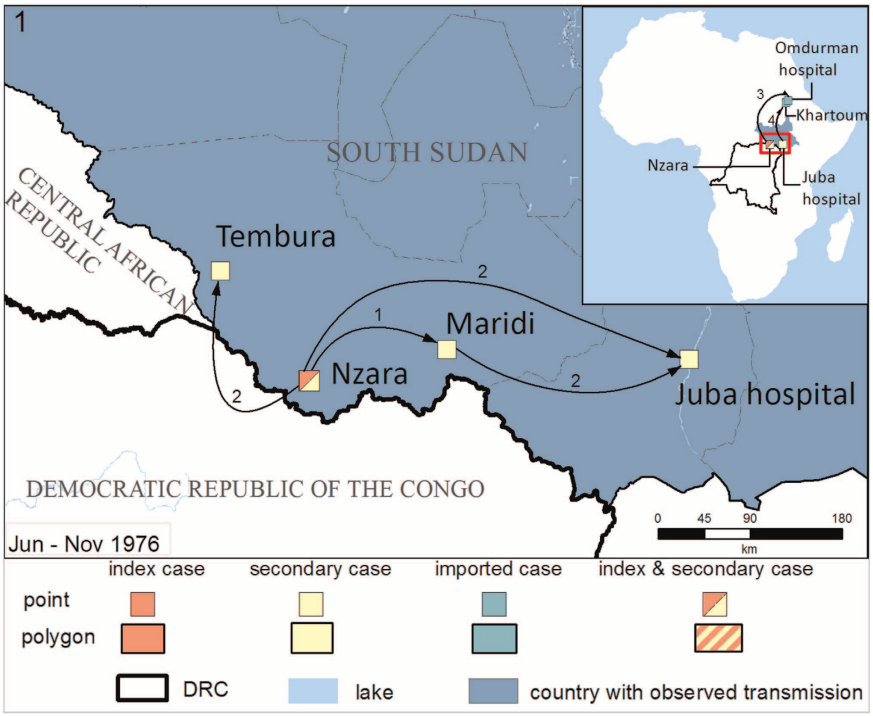


Figure 2. Map of the South Sudan (1976) outbreak. The first reported cases of Sudan Ebola virus were in three workers at a cotton factory in Nzara, in close proximity to three game reserves. The method of acquisition was unknown. The first secondary cases arose in Nzara infecting a total of 67 people who were primarily family members of the factory workers. Further secondary transmission clusters emerged in Maridi following spread from Nzara due to seeking treatment, after which further cases occurred in Juba due to patients who were referred. Additional cases from Maridi were also referred directly to Juba making the source of infection in Juba difficult to identify. Secondary transmission also emerged in Tembura due to a patient seeking family care, although the source of this infection is unknown. Imported cases from Juba to Khartoum and from Nzara to Omdurman were also reported following a patient seeking treatment and a referral for diagnosis respectively (see inset). The principal mode of transmission in this outbreak was initially familial, although in Maridi secondary transmission arose through nosocomial transmission. Seeking of treatment was the principal cause of geographic spread. The first index case became ill on the 27 June 1976 before the first secondary cases in July and subsequent secondary transmission clusters from August to October. Cases peaked in September (138 cases, 65 deaths). The final case was reported on 25 November 1976. Imported cases in Omdurman and Khartoum were reported in August and September, respectively. Overall, 284 cases were reported with 151 deaths giving a CFR of 53.2%. This figure varied in different locations: Nzara (67,31,46%), Maridi (213,116,55%), Tembura (3,3,100%), Juba (1,1,100%). Arrows indicate order of spread. Where spread order is known, numbers are indicative of the order of spread. Arrows sharing the same number indicate that it was not possible to distinguish which spread happened first.

Selected examples are presented in Figs. 2,3,4. In the first, initial focal infection jumped from village to village, as well as longer distance dispersal primarily through treatment seeking of infected patients (Fig. 2). In the second (Fig. 3) more limited outbreak, cases occurred in two main clusters, with one containing working camps surrounding the site of suspected zoonotic transfer and the other villages and a local hospital in the vicinity of Makokou. In the final example outbreak (Fig. 4) secondary transmission spread radially out from the village of Balimba following treatment seeking of a faith healer. Similar to outbreak 1, infected patients travelled to local and national healthcare centres for treatment, in some cases causing secondary transmission at great distances from the index case.

In the text descriptions, methods, sources and order of spread were described in a standardised manner, where information allowed. A brief summary of the evolution of the outbreak over time and the breakdown of total number of cases, deaths and case fatality rate over the course of the entire outbreak is also given.

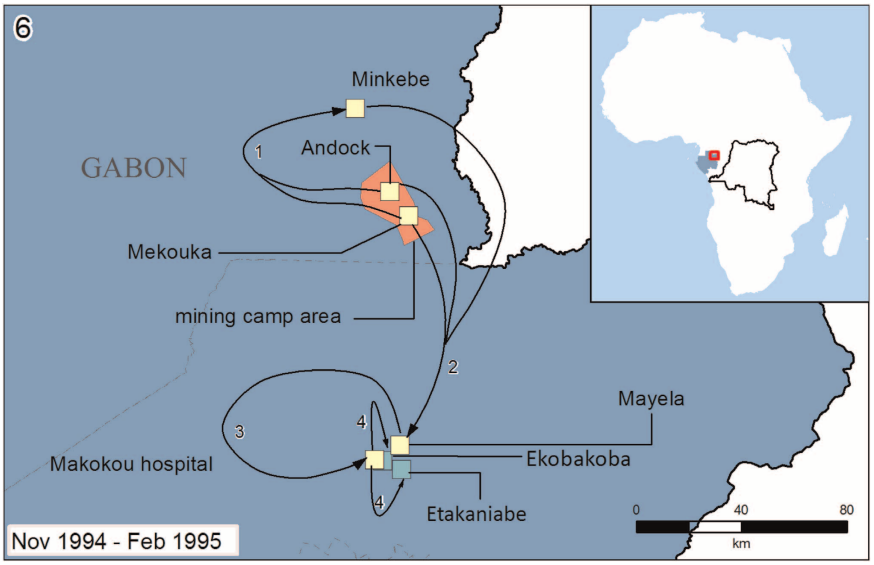


Figure 3. Map of the Gabon (1994) outbreak. The first reported cases of Zaire Ebola virus were in miners from the Mekouka and Andock encampments, suspected to have contracted the infection in the surrounding area. The method of acquisition was unknown. The first secondary cases arose within these two encampments and then spread to the Minkebe camp. Further secondary transmission clusters emerged in Mayela then Makokou general hospital after 32 patients from the forest encampments sought treatment. Cases were also reported in Ekataniabe and Ekobakoba who had recent travel histories to Makokou general hospital. The principal modes of transmission were among workers at first, followed by nosocomial in Makokou general hospital and familial in Mayela (connected by a single traditional healer). The initial case was reported on 13 November 1994 before secondary transmission clusters occurred from the end of January to February 1995. Cases and deaths peaked in December (26 cases, 14 deaths (53.8% CFR)). The final case was reported on 9 February 1995 in Ekobakoba. Overall, 49 cases were reported with 30 deaths, giving a CFR of 61.2%. For map key, see Fig. 2.

A wider-scale map showing the locations of all the outbreaks since 1976 is also available in Supplementary Fig. 23 and a table detailing which type of data were obtained from which source for each outbreak is given in Table 1 (available online only)[3,8,10,13–18,20,29–46].

Data Records

The data from this analysis are summarised in two types of data format (Data Citation 1). First a data table details unique geographic locations of Ebola occurrence, including information on type of transmission, location, spread, timing and case number. These geographic locations are grouped into individual outbreaks ($n = 22$) and summary statistics on timings and case and death numbers are given for each outbreak. Second, geographic information files are provided that match the information presented in the data table to explicit geographic areas. These are available in a variety of formats that can be read by geographic information system (GIS) applications. Information from these two file types were used to make the outbreak summary maps and text in Supplementary Figures 1–22 (selected examples in Figs. 2,3,4).

Data table of unique Ebola virus transmission locations

The table includes the following fields, detailed below. The value 'NA' was entered if information was unknown, unreported or indeterminable. The term 'occurrence' refers to Ebola transmission locations that are either unique in the geographic location or in their type of transmission (index, secondary or imported). Each row in the table represents a unique Ebola occurrence. A group of occurrences make up contained 'outbreaks' and fields with the prefix 'OB' summarise various metrics related to the entire outbreak that each occurrence belongs to.

UNIQ_ID: A unique identification number for each occurrence at which index, secondary or imported cases of EVD have occurred at unique geographic locations ($n = 117$).

NAME: Text description of the point or polygon that defines the location of the occurrence.

COUNTRY: The country where the majority of cases occurred in each outbreak.

VIRUS: The Ebola virus species of each outbreak.

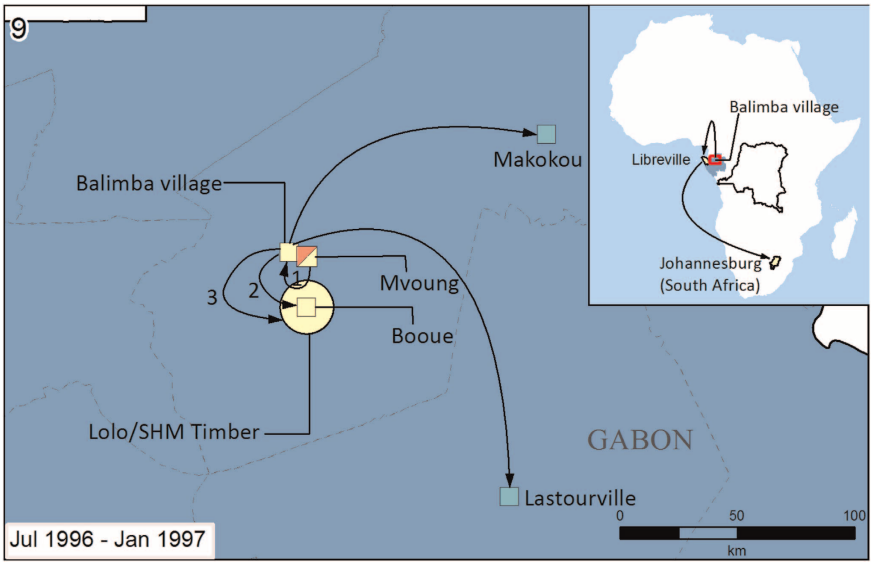


Figure 4. Map of the Gabon (1996b) outbreak. The index case of Zaire Ebola virus likely came from one of three infected hunters in a logging camp near Mvoung. The timing of infection makes it difficult to distinguish index cases from secondary cases during the early stages of this outbreak, but it is likely that the first secondary cases emerged amongst the hunters who then sought treatment from a traditional healer in Balimba. After falling ill, the traditional healer from Balimba sought treatment in Booue, where the disease then radially spread through the communities in the surrounding areas. A further secondary transmission cluster emerged in Libreville (see inset) after patients from Balimba sought treatment there. In Libreville one doctor became infected and flew to Johannesburg, South Africa for treatment before receiving a diagnosis of Ebola. Limited further nosocomial transmission (1 case) occurred upon his arrival in Johannesburg. Imported cases in Makokou General Hospital and Lastourville were also reported after patients from Balimba sought treatment. No clear principal mode of transmission was observed for the early stages of the outbreak, but in Libreville secondary transmission mainly arose through nosocomial transmission. The index case was reported on the 13 July 1976 before the first secondary cases in September and subsequent secondary transmission clusters from September to January. Cases peaked in September and deaths peaked in October. The final case was reported on 18 January 1997. Overall 60 cases were reported with 45 deaths, giving a CFR of 75%. For map key, see Fig. 2.

CASE_TYPE: The type of transmission represented by the Ebola occurrence. Can be either 'index', 'secondary' or 'import'.

DATA_TYPE: Whether the occurrence represents a point or larger polygon location.

LAT: The latitude of the centre point of the point or polygon of the occurrence.

LONG: The longitude of the centre point of the point or polygon of the occurrence.

LOC_NTS: Additional notes describing the site location of the occurrence.

SPR_ORDER: The order of spread between occurrences over the course of the outbreak, as determined by the date of onset of the first case in a given occurrence. Index cases are represented with the value '1'. Two or more occurrences share the same spread order if it is unknown which of the two areas Ebola virus transmission spread to first.

SOURCE_1: The unique identification number of the occurrence where the first EVD patient came from.

SOURCE_2: The unique identification number of the occurrence where the first EVD patient came from. An occurrence may have more than one source if infected patients came from more than one source but it is unknown which triggered secondary transmission.

SOURCE_3: The unique identification number of the occurrence where the first EVD patient came from. An occurrence may have more than one source if infected patients came from more than one source but it is unknown which triggered secondary transmission.

STR_DAY: Day of first reported case in the occurrence.

STR_MNTH: Month of first reported case in the occurrence.

STR_YEAR: Year of first reported case in the occurrence.

END_DAY: Day of last reported case in the occurrence.

END_MNTH: Month of last reported case in the occurrence.

END_YEAR: Year of last reported case in the occurrence.

REP_CASE: The total number of cases (suspected or confirmed) reported over the course of the outbreak, but only within the occurrence.

REP_DEATH: The total number of deaths (suspected or confirmed) reported over the course of the outbreak, but only within the occurrence.

OB_ID: A unique identification number for each outbreak ($n = 22$).

OB_STR_DAY: Day of first reported case of the outbreak.

OB_STR_MNTH: Month of first reported case of the outbreak.

OB_STR_YEAR: Year of first reported case of the outbreak.

OB_END_DAY: Day of last reported case of the outbreak.

OB_END_MNTH: Month of last reported case of the outbreak.

OB_END_YEAR: Year of last reported case of the outbreak.

OB_CASE: The total number of cases (suspected or confirmed) reported over the course of the outbreak in all areas.

OB_DEATH: The total number of deaths (suspected or confirmed) reported over the course of the outbreak in all areas.

Geographic information files

The geographic information files include index, secondary and imported cases linked to geographic locations. All fields match those in the data table. Unknown, unreported or indeterminable data are represented by the character combination 'NA'.

Technical Validation

For each outbreak, all relevant text articles were used to confirm or reach a consensus on the likely result. Where differing results were found (e.g., total case count), figures from primary research articles took preference over review, summary articles or periodical epidemiological reports.

The extracted geographic and epidemiological summary information was cross-checked by at least two different researchers to ensure accuracy.

All point and polygon locations were checked in ArcGIS 10.1[28] against national and subnational boundaries to ensure their location matched the text descriptions. A gridded raster file from the Global Lakes and Wetlands Database[47] giving the locations of rivers and lakes was also included to check points and polygons fell on land rather than water. Any points that fell in water were moved to the nearest land pixel.

Usage Notes

The data presented here can be used in combination with spatial and temporal meteorological and socioeconomic information to generate hypotheses about the factors that may be important in the emergence of index cases and the spread of secondary and imported cases of EVD.

Pigott *et al.* combined data on Ebola index cases with a suite of environmental information in a species distribution model to map the zoonotic niche of Ebola transmission across Africa[48]. Matching more varied and finer scale local information at the sites of the index cases presented here could help develop our understanding of the complex process of Ebola virus emergence and the risk posed by certain human activities and land use patterns.

Using the secondary and imported case data, it would be possible to model and investigate causes of spread of human Ebola outbreaks. Understanding spread of the pathogen in past outbreaks may aid control of the current outbreak. A comparison between the spread of historical outbreaks and the spread of the current ongoing outbreak once it is over will be useful for informing Ebola surveillance and control guidelines to minimise the size and burden of these sporadic zoonoses.

Finally, this dataset can be used to investigate how the rate, extent and the environment in which Ebola outbreaks spread relates to important outbreak measures such as the total number of cases and case fatality rate. Understanding what distinguishes brief, geographically limited and low mortality Ebola outbreaks from those that impose a much higher public health burden will be important for informing future surveillance, control and treatment efforts[49]. This dataset provides the most comprehensive collection of standardised data on Ebola outbreak spread currently available and will be an important resource for these uses. These records will also be updated periodically with data from ongoing and future Ebola outbreaks to ensure the ongoing relevance of this database.

References

1. Barrette, R. W. *et al.* Discovery of swine as a host for the Reston ebolavirus. *Science* **325,** 204–206 (2009).
2. Feldmann, H. & Geisbert, T. W. Ebola haemorrhagic fever. *The Lancet* **377,** 849–862 (2011).
3. Le Guenno, B. *et al.* Isolation and partial characterisation of a new strain of Ebola virus. *The Lancet* **345,** 1271–1274 (1995).
4. King, A. M., Adams, M. J., Lefkowitz, E. J. & Carstens, E. B. *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses* Vol. 9. (Elsevier, 2012).
5. Kuhn, J. & Calisher, C. H. *Filoviruses: A Compendium of 40 Years of Epidemiological, Clinical, and Laboratory Studies* 1st edn, Vol. 20 (Springer, 2008).
6. Peterson, A. T., Carroll, D. S., Mills, J. N. & Johnson, K. M. Potential mammalian filovirus reservoirs. *Emerg. Infect. Dis.* **10,** 2073–2081 (2004).
7. Leroy, E. M. *et al.* Fruit bats as reservoirs of Ebola virus. *Nature* **438,** 575–576 (2005).

8. Leroy, E. M. *et al.* Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonot. Dis.* **9,** 723–728 (2009).
9. Bermejo, M. *et al.* Ebola outbreak killed 5000 gorillas. *Science* **314,** 1564–1564 (2006).
10. Formenty, P. *et al.* Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire. *J. Infect. Dis.* **179,** S120–S126 (1999).
11. Rouquet, P. *et al.* Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001–2003. *Emerg. Infect. Dis.* **11,** 283–290 (2005).
12. Groseth, A., Feldmann, H. & Strong, J. E. The ecology of Ebola virus. *Trends Microbiol.* **15,** 408–416 (2007).
13. Milleliri, J., Tévi-Benissan, C., Baize, S., Leroy, E. & Georges-Courbot, M. Les épidémies de fièvre hémorragique due au virus Ebola au Gabon (1994-2002). *Bull. Soc. Pathol. Exot. Filiales* **97,** 199–205 (2004).
14. Nkoghe Mba, D. *et al.* Plusieurs épidémies de fièvre hémorragique due au virus Ebola au Gabon, d'octobre 2001 à avril 2002. *Bull. Soc. Pathol. Exot. Filiales* **98,** 224–229 (2005).
15. Boumandouki, P. *et al.* Prise en charge des malades et des défunts lors de l'épidémie de fièvre hémorragique due au virus Ebola d'octobre à décembre 2003 au Congo. *Bull. Soc. Pathol. Exot. Filiales* **98,** 218–223 (2005).
16. Nkoghe, D., Kone, M. L., Yada, A. & Leroy, E. A limited outbreak of Ebola haemorrhagic fever in Etoumbi, Republic of Congo, 2005. *Trans. R. Soc. Trop. Med. Hyg.* **105,** 466–472 (2011).
17. Baron, R. C., McCormick, J. B. & Zubeir, O. A. Ebola virus disease in southern Sudan - hospital dissemination and intrafamilial spread. *Bull. World Health Organ.* **61,** 997–1003 (1983).
18. Georges, A.-J. *et al.* Ebola hemorrhagic fever outbreaks in Gabon, 1994–1997: epidemiologic and health control issues. *J. Infect. Dis.* **179,** S65–S75 (1999).
19. Francesconi, P. *et al.* Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. *Emerg. Infect. Dis.* **9,** 1430 (2003).
20. Shoemaker, T. *et al.* Reemerging Sudan ebola virus disease in Uganda, 2011. *Emerg. Infect. Dis* **18,** 1480 (2012).
21. WHO. *Interim Manual - Ebola and Marburg Virus Disease Epidemics: Preparedness, Alert, Control, and Evaluation*, http://www.who.int/csr/disease/ebola/manual_EVD/en/ (2014).
22. Centers for Disease Control. *Known Cases and Outbreaks of Ebola Hemorrhagic Fever, in Chronological Order*, http://www.cdc.gov/vhf/ebola/resources/outbreak-table.html (2014).
23. WHO. Outbreak(s) of Ebola haemorrhagic fever, Congo and Gabon, October 2001–July 2002. *Wkly. Epidemiol. Rec.* **78,** 24 (2003).
24. Messina, J. P. *et al.* A global compendium of human dengue virus occurrence. *Sci. Data* **1,** 140004 (2014).
25. Pigott, D. M. *et al.* Global database of leishmaniasis occurrence locations, 1960–2012. *Sci. Data* **1,** 140036 (2014).
26. Food and Agriculture Organization of the United Nations. *The Global Administrative Unit Layers (GAUL): Technical Aspects*. Food and Agriculture Organization of the United Nations, EC-FAO Food Security Programme (ESTG), (2008).
27. IKI maps, http://www.ikimap.com/.
28. ArcGIS Desktop (Environmental Systems Research Institute, Redlands, CA, 2012).
29. WHO/International Study Team. Ebola haemorrhagic fever in Sudan, 1976. *Bull. World Health Organ.* **56,** 247–270 (1978).
30. International Commission. Ebola haemorrhagic fever in Zaire, 1976. *Bull. World Health Organ.* **56,** 271–293 (1978).
31. Heymann, D. *et al.* Ebola hemorrhagic fever: Tandala, Zaire, 1977–1978. *J. Infect. Dis.* **142,** 372–376 (1980).
32. Lamunu, M. *et al.* Containing a haemorrhagic fever epidemic: the Ebola experience in Uganda (October 2000–January 2001). *Int. J. Infect. Dis.* **8,** 27–37 (2004).
33. World Health Organization. Outbreak of Ebola haemorrhagic fever, Uganda, August 2000–January 2001. *Wkly. Epidemiol. Rec.* **76,** 41–48 (2001).
34. Okware, S. *et al.* An outbreak of Ebola in Uganda. *Trop. Med. Int. Health* **7,** 1068–1075 (2002).
35. World Health Organization. Outbreak(s) of Ebola haemorrhagic fever in the Republic of the Congo, January-April 2003. *Wkly. Epidemiol. Rec.* **78,** 285–289 (2003).
36. World Health Organization. Outbreak of Ebola haemorrhagic fever in Yambio, south Sudan, April-June 2004. *Wkly. Epidemiol. Rec.* **80,** 370–375 (2005).
37. Onyango, C. O. *et al.* Laboratory diagnosis of Ebola hemorrhagic fever during an outbreak in Yambio, Sudan, 2004. *J. Infect. Dis.* **196,** S193–S198 (2007).
38. Grard, G. *et al.* Emergence of divergent Zaire ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *J. Infect. Dis.* **204,** S776–S784 (2011).
39. Wamala, J. F. *et al.* Ebola hemorrhagic fever associated with novel virus strain, Uganda, 2007–2008. *Emerg. Infect. Dis.* **16,** 1087–1092 (2010).
40. MacNeil, A. *et al.* Proportion of deaths and clinical features in Bundibugyo Ebola virus infection, Uganda. *Emerg. Infect. Dis.* **16,** 1969 (2010).
41. Towner, J. S. *et al.* Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* **4,** e1000212 (2008).
42. World Health Organisation. *End of Ebola Outbreak in the Democratic Republic of the Congo*, http://www.who.int/csr/don/2009_02_17/en/ (2009).
43. Albarino, C. *et al.* Genomic analysis of filoviruses associated with four viral hemorrhagic fever outbreaks in Uganda and the Democratic Republic of the Congo in 2012. *Virology* **442,** 97–100 (2013).
44. World Health Organization. *DR Congo: Ebola (Situation as of 01 October 2012)*, http://www.afro.who.int/pt/grupos-organicos-e-programas/ddc/alerta-e-resposta-epidemias-e-pandemias/outbreak-news/3698-dr-congo-ebola-situation-as-of-01-october-2012.html (2012).
45. World Health Organisation. *Ebola in Uganda*, http://www.who.int/csr/don/2012_07_29/en/ (2012).
46. World Health Organisation. *Uganda: Ebola (situation as of 27 August 2012)*, http://www.afro.who.int/en/clusters-a-programmes/dpc/epidemic-a-pandemic-alert-and-response/outbreak-news/3674-uganda--ebola-situation-as-of-27-august-2012-.html (2012).
47. WWF. Global Lakes and Wetlands Database, http://www.worldwildlife.org/pages/global-lakes-and-wetlands-database (2014).
48. Pigott, D. M. *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3,** e04395 (2014).
49. Brady, O., Hay, S. & Horby, P. Scale up supply of experimental Ebola drugs. *Nature* **512,** 233 (2014).

Data Citation

1. Mylne, A. *et al. Figshare* http://dx.doi.org/10.6084/m9.figshare.1168886 (2014).

Acknowledgements

O.J.B. is funded by a BBSRC studentship. S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (#095066) which also supports A.M. and a grant from the Bill & Melinda Gates Foundation (#OPP1093011). S.I.H. would also like to acknowledge funding support from the RAPIDD

program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. Z.H. is funded by the Bill & Melinda Gates Foundation (#OPP1106023). D.M.P. is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford. N.G. is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). M.U.G.K. is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. Funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

O.J.B. and A.M. wrote the manuscript with editing and approval from all authors. O.J.B., A.M. and D.M.P. compiled the data records. Z.H. digitized the geographic data and produced the outbreak maps. N.G., D.M.P. and M.U.G.K. helped with data cross-checking. S.I.H. conceived database design.

Additional information

Table 1 is only available in the online version of this paper.

Supplementary information accompanies this paper at http://www.nature.com/sdata

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mylne, A. *et al.* A comprehensive database of the geographic spread of past human Ebola outbreaks. *Sci. Data* 1:140042 doi: 10.1038/sdata.2014.42 (2014).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0

Metadata associated with this Data Descriptor is available at http://www.nature.com/sdata/ and is released under the CC0 waiver to maximize reuse.

eLIFE elifesciences.org

RESEARCH ADVANCE

Updates to the zoonotic niche map of Ebola virus disease in Africa

David M Pigott[1,2]*, Anoushka I Millear[1], Lucas Earl[1], Chloe Morozoff[1]*, Barbara A Han[3], Freya M Shearer[2], Daniel J Weiss[4,5], Oliver J Brady[2], Moritz UG Kraemer[4,5], Catherine L Moyes[2], Samir Bhatt[4,5], Peter W Gething[4,5], Nick Golding[2,6], Simon I Hay[1,2]

[1]Institute for Health Metrics and Evaluation, University of Washington, Seattle, United States; [2]Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Oxford, United Kingdom; [3]Cary Institute of Ecosystem Studies, New York, United States; [4]Spatial Ecology and Epidemiology Group, University of Oxford, Oxford, United Kingdom; [5]Department of Zoology, University of Oxford, Oxford, United Kingdom; [6]Department of BioSciences, University of Melbourne, Parkville, Australia

Abstract As the outbreak of Ebola virus disease (EVD) in West Africa is now contained, attention is turning from control to future outbreak prediction and prevention. Building on a previously published zoonotic niche map (***Pigott et al., 2014***), this study incorporates new human and animal occurrence data and expands upon the way in which potential bat EVD reservoir species are incorporated. This update demonstrates the potential for incorporating and updating data used to generate the predicted suitability map. A new data portal for sharing such maps is discussed. This output represents the most up-to-date estimate of the extent of EVD zoonotic risk in Africa. These maps can assist in strengthening surveillance and response capacity to contain viral haemorrhagic fevers.

DOI: 10.7554/eLife.16412.001

***For correspondence:** pigottdm@uw.edu (DMP); chloemor@uw.edu (CM)

Competing interest: See page 10

Funding: See page 10

Received: 30 March 2016
Accepted: 20 June 2016
Published: 14 July 2016

Reviewing editor: Mark Jit, London School of Hygiene & Tropical Medicine, and Public Health England, United Kingdom

© Copyright Pigott et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Since the index case in 2013, the West African Ebola epidemic has killed more than 11,000 people (***World Heath Organization, 2016***) and exposed national and international inadequacies in pandemic preparedness and response (***Moon et al., 2015***). In 2014 a zoonotic niche map for Ebola virus disease (EVD) was produced (***Pigott et al., 2014***) in part to assess the expected geographical extent of spillover risk. This research was then expanded to explore how changes in demography and international connectivity may have facilitated the establishment and rapid subsequent spread of the epidemic (***Bogoch et al., 2015***). The West African outbreak of EVD has again highlighted key information gaps that exist with respect to the broader epidemiology of Ebola virus, particularly concerning viral persistence in reservoirs (***Funk and Piot, 2014***; ***Mari Saez et al., 2015***; ***Leendertz, 2016***), and prompted a variety of questions concerning the role bats play in transmission (***Leendertz et al., 2016***). Identifying reservoirs of zoonotic disease is a complex process (***Viana et al., 2014***; ***Haydon et al., 2002***) and whilst considerable sampling effort has been undertaken over the years (***Kuhn, 2008***; ***Leirs et al., 1999***), isolation of Ebolavirus from living animals has been rare (***Leroy et al., 2005***). The original eLife study only incorporated the three bat species found to be RNA-positive (***Leroy et al., 2005***). Whilst this remains currently the best evidence for an animal reservoir species, it is important to consider that other sampling efforts may by chance represent false negatives, particularly if infection is rare.

Consequently, to contribute to these broader discussions, the original paper (**Pigott et al., 2014**) was updated with new occurrence data and expanded to consider a wider range of potential bat reservoir species. Bats remained the priority mammalian order given the previous viral isolation and the repeated anecdotal implications in previous outbreaks (**Leroy et al., 2009**; **Mari Saez et al., 2015**). Since there are a large number of bat species found in Africa, we defined three groupings, based upon the strength of evidence supporting their potential Ebola reservoir status. As a result, not only were the original three RNA-positive bats included (**Leroy et al., 2005**), but also those species with serological evidence of EVD infection (**Olival and Hayman, 2014**) and those identified through trait-based machine learning approaches as being similar to species already reporting filoviral infection (**Han et al., 2016**).

Results

Six additional records of EVD were incorporated into the disease occurrence database: one human outbreak in the Democratic Republic of Congo (**Maganga et al., 2014**); two reports of infections in animals in Zambia (**Ogawa et al., 2015**); and three animal infections in Central African Republic (**Morvan et al., 1999**) (**Figure 1**). Of these new occurrences, two in southern Central African Republic are found in areas predicted to be at-risk by the previous model (**Pigott et al., 2014**), with the index case from the Democratic Republic of Congo located in close proximity (<10 km) to at-risk areas. The occurrences in Zambia and northern Central African Republic lie, respectively, to the south and north of previously predicted at-risk regions.

Figure 2 depicts the three new consolidated bat distributions. The revised distribution of the Group 1 bats (*i.e.* those found to have been Ebolavirus RNA positive) is broadly consistent with that published in the original paper except that the peripheries of Central Africa are now predicted to be environmentally suitable for these bats, as well as some parts of East Africa, particularly Tanzania, Mozambique and Madagascar. The Group 2 and Group 3 bat species are predicted to be distributed across much of Africa stretching from West to East Sub-Saharan Africa, as well as much of the coastline of the continent.

The revised niche map, incorporating the updated bat covariates and disease occurrence database, is presented in **Figure 3**. The map shows the predicted areas of environmental suitability for zoonotic Ebola virus transmission to be consistent with previous attempts, but the relative environmental suitability within this distribution differs from the previous estimates. **Figure 3—figure supplement 1** demonstrates that Cameroon, Gabon, Republic of Congo and mainland Equatorial Guinea are now predicted to be more environmentally suitable than in the previous analysis. The regions of Central Africa (particularly Gabon and the Republic of Congo) identified as being most environmentally suitable for zoonotic EVD transmission in the previous analysis remain so in this analysis. The revised number of predicted at-risk countries, determined by thresholding the map by a probability that captures 95% of the occurrence dataset, is 23 (**Table 1**).

The similar AUC values (0.85 ± 0.04 compared to 0.8236 ± 0.080) between the previous and current iterations suggest that the updated model fits the new occurrence dataset as well as the previous model fitted the older dataset. Mean enhanced vegetation index (EVI) remains the highest relative predictor covariate for zoonotic EVD transmission while the relative importance of Group 1 bat distributions moved from being the fifth most important to the second. Mean night-time land surface temperature (LST), elevation and mean daytime LST complete the top five predictors (**Table 2**).

When separate bat layers were used in the model, as opposed to the consolidated covariates, the predictions were geographically similar (**Figure 3—figure supplement 2**) however, four bat species were identified as explaining more of the variation than the rest; *Hypsignathus monstrosus, Epomops franqueti* (from both of which Ebolavirus RNA has been isolated), *Otomops martiensseni* and *Epomophorus labiatus* (both from Group 3). The explanatory power of this model (evaluated using AUC) was comparable to the model results described above (AUC = 0.819 ± 0.080).

Discussion

This research advance integrates new data as well as a more thorough consideration of the bat species that act as a reservoir for the virus in order to update our modelled estimate of the zoonotic

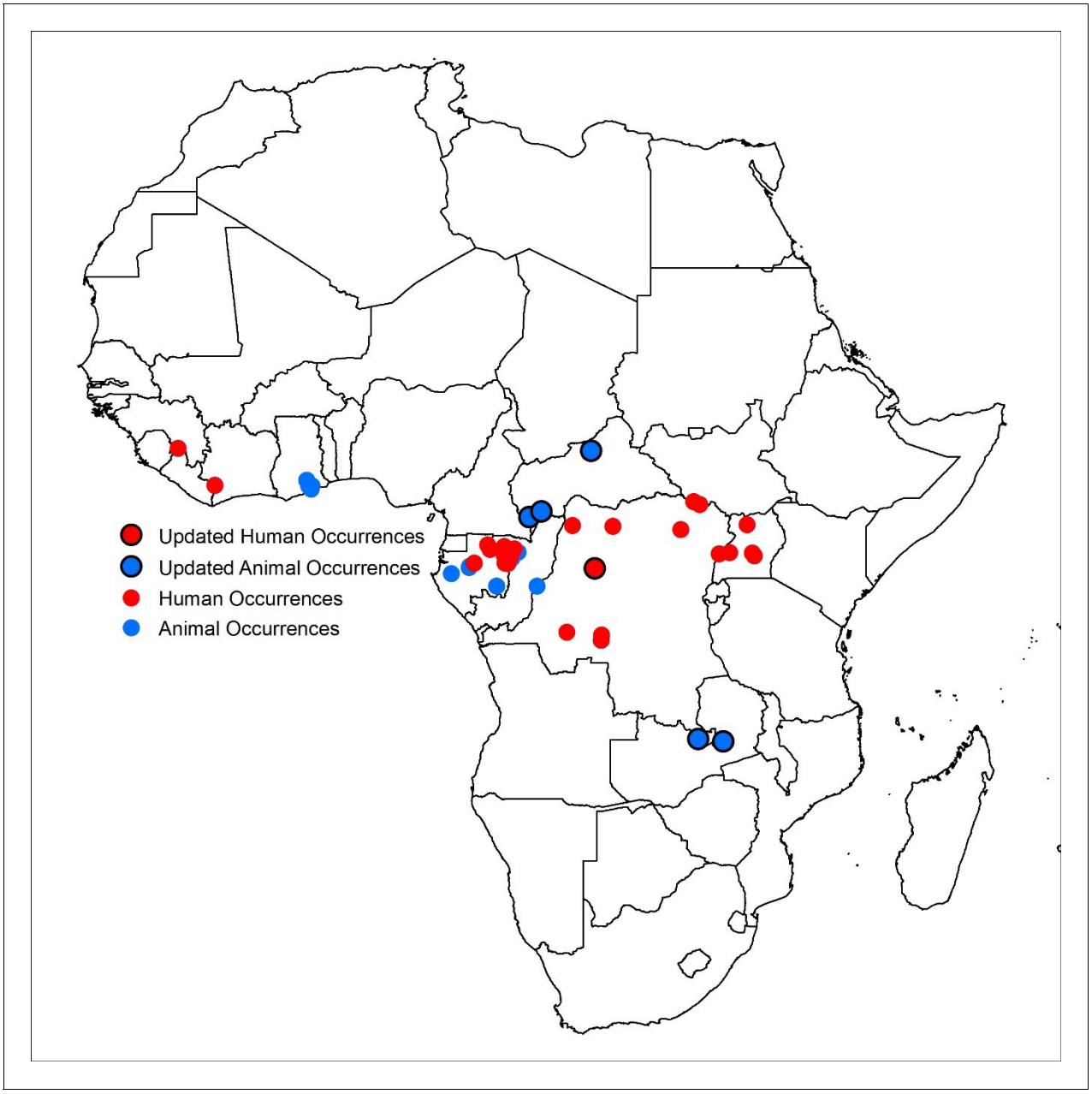


Figure 1. Updated Ebola virus disease occurrence database. Human index cases are represented by red circles, animal occurrences in blue. New occurrence information is indicated by the black circle. The coordinates of polygon centroids are displayed for occurrences defined by an area greater than 5 km x 5 km.

DOI: 10.7554/eLife.16412.002

niche of EVD. The area estimated to be at potential risk of zoonotic EVD transmission has now expanded to include Kenya and the influence of additional bat species demonstrates that continued focus should be placed on rigorously identifying reservoir species and the role they play in sustaining viral transmission (***Leendertz, 2016***). The fact that *O. martiensseni* and *E. labiatus* contribute explanatory power to the model, in comparison with their distributions on the eastern and southern periphery of reported cases of EVD (***Figure 2—figure supplement 3***) suggests that different regions of the continent may support transmission cycles with differing reservoir species. This, coupled with the potential for each of the pathogenic species of *Ebolavirus* potentially having differing distributions (***Peterson et al., 2004***), cannot currently be explored more rigorously due to insufficient data.

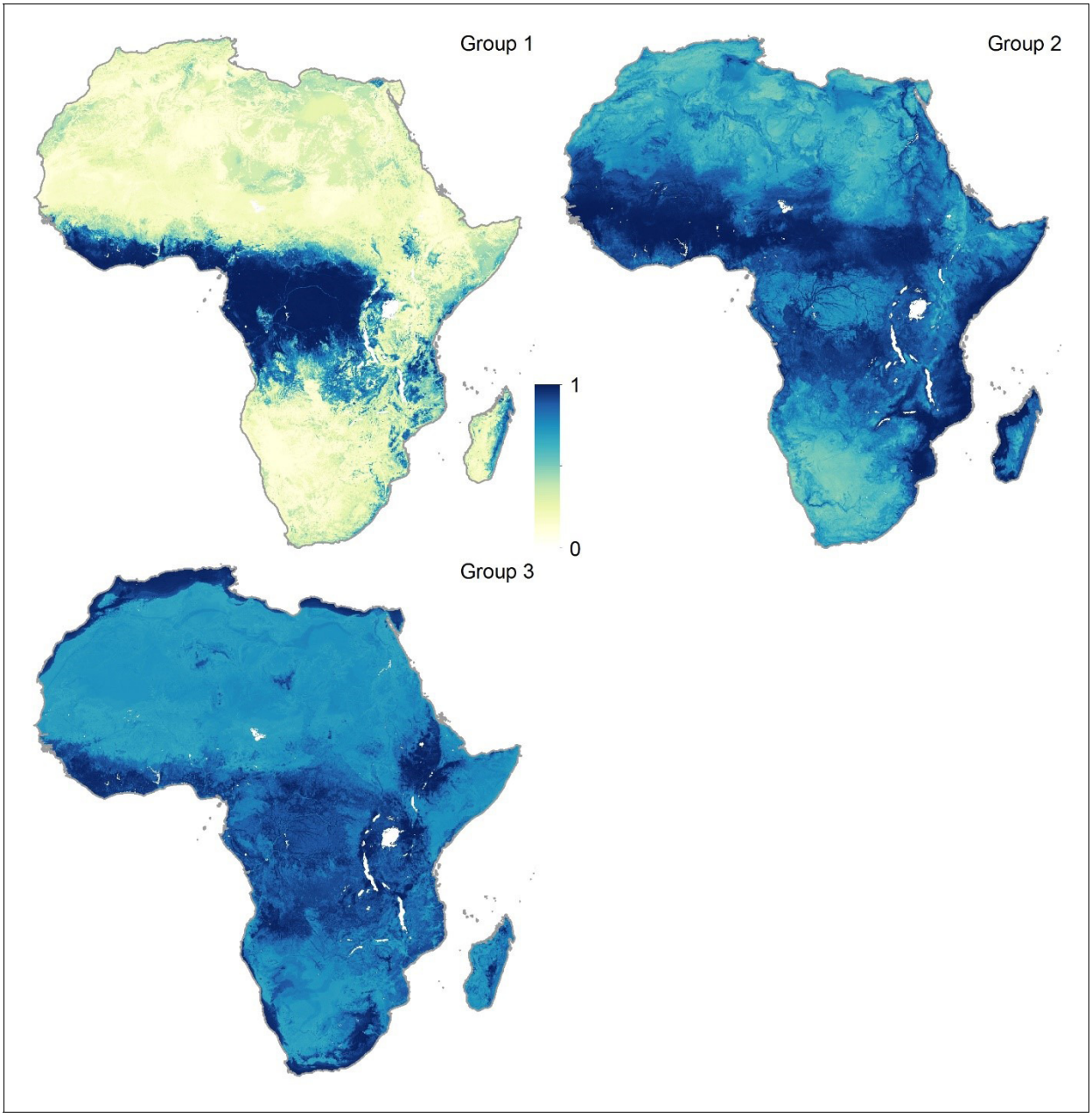


Figure 2. Combined suitability surfaces for each of the potential reservoir bat groupings. For each layer the species specific suitability maps were combined to produce a surface approximating the probability that any bat species in that group may be present. Regions in blue (1) are most environmentally similar to locations reporting bat records. Areas in yellow (0) are the least environmentally similar. The top left panel depicts Group 1, top right Group 2 and bottom left Group 3 bats.

DOI: 10.7554/eLife.16412.003

The following figure supplements are available for figure 2:

Figure supplement 1. Group 1 bat distributions.

DOI: 10.7554/eLife.16412.004

Figure supplement 2. Group 2 bat distributions.

DOI: 10.7554/eLife.16412.005

Figure supplement 3. Group 3 bat distributions.

DOI: 10.7554/eLife.16412.006

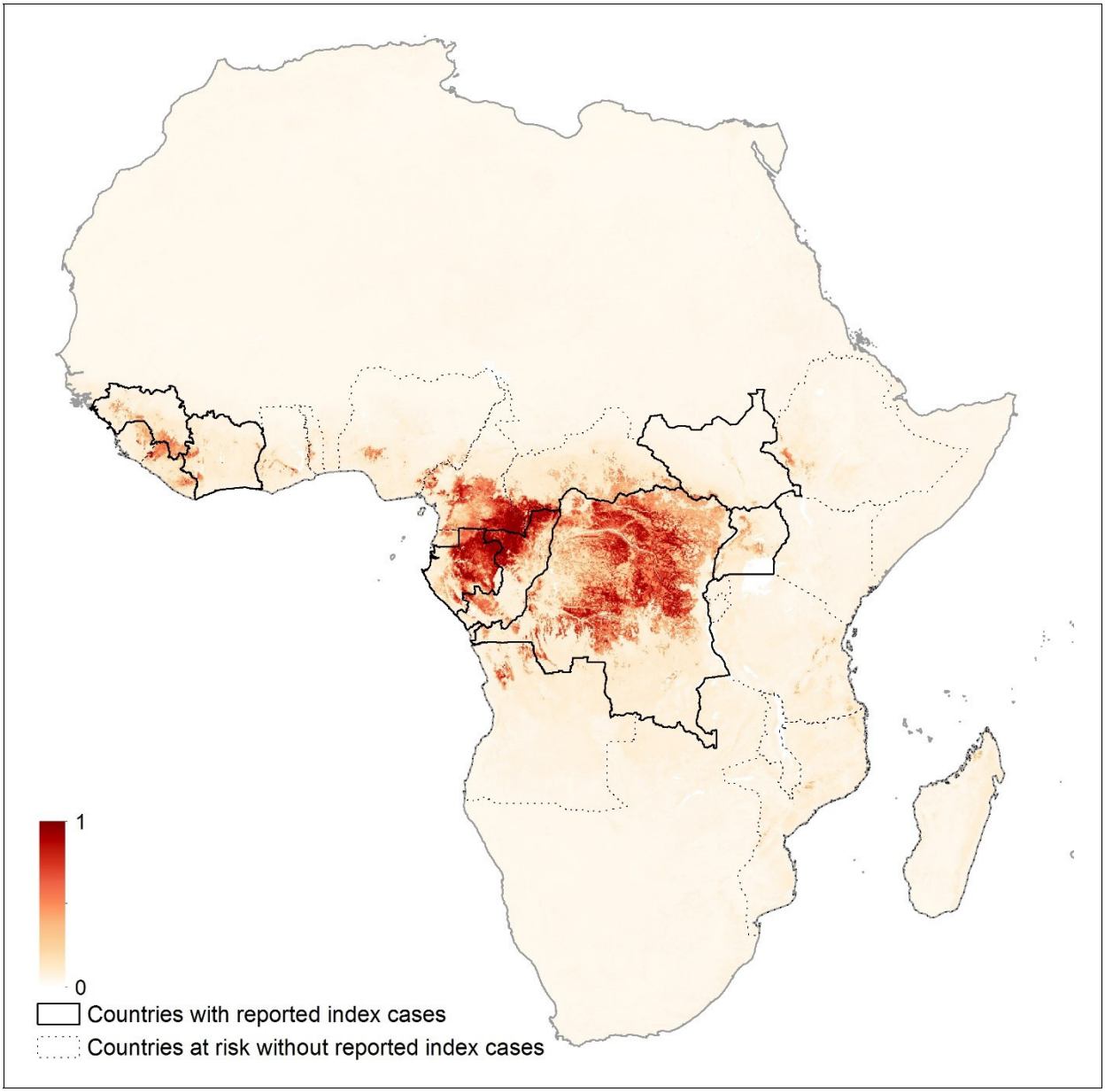


Figure 3. Updated map showing areas most environmentally suitable for the zoonotic transmission of Ebola virus. Areas closer to dark red (1) are most environmentally similar to locations reporting Ebola virus occurrences; areas in light yellow (0) are least similar. Countries with borders outlined are those which are predicted to contain at-risk areas for zoonotic transmission based on a thresholding approach. Output displayed generated from model using the three consolidated bat covariates.

DOI: 10.7554/eLife.16412.007

The following figure supplements are available for figure 3:

Figure supplement 1. Absolute differences between previous and revised maps.

DOI: 10.7554/eLife.16412.008

Figure supplement 2. Zoonotic niche map based upon inclusion of individual bat covariate layers.

DOI: 10.7554/eLife.16412.009

As with the original publication, it must be reiterated that environmental suitability does not inevitably lead to spillover events. Currently absolute human population residing in at-risk pixels is used as a proxy for spillover likelihood, however, a variety of factors will influence the outbreak risk within each location (***Plowright et al., 2015***) and only by including covariates that consider human behaviour (***Woldehanna and Zimicki, 2015***), patterns of susceptibility in other animals (***Walsh et al.,***

Table 1. National populations at risk.

	Country	Population-at-risk (in 100,000s)
Countries previously reporting index cases	Democratic Republic of the Congo	170.18
	Uganda	21.58
	Guinea	17.61
	Côte d'Ivoire	4.08
	Gabon	3.65
	South Sudan	1.80
	Republic of Congo	1.07
Countries with no reported index cases	Nigeria	29.13
	Cameroon	22.90
	Central African Republic	7.62
	Liberia	5.88
	Ghana	4.04
	Sierra Leone	3.94
	Angola	3.25
	Togo	1.78
	Ethiopia	1.75
	Equatorial Guinea	1.22
	Tanzania	1.18
	Burundi	1.07
	Mozambique	0.55
	Madagascar	<0.1
	Kenya	<0.1
	Malawi	<0.1

DOI: 10.7554/eLife.16412.010

2007; ***2009***), impacts of land use change (***Rosales-Chilama et al., 2015***) and within-host viral dynamics (***Amman et al., 2012***; ***Hayman, 2015***) can an approximation of spillover risk be defined.

These updates demonstrate the ease and speed with which new data and covariate considerations can be incorporated within existing empirical models (***Kraemer et al., 2016***). As the wider discussion on EVD turns to focus on strategies to prevent or contain future spillover events as well as developing long-term in-country containment capacities (***Currie et al., 2016***), it is hoped that maps

Table 2. Comparison of previous and revised niche models.

	Revised niche map (with summary bat layers)	Revised niche map (with individual bat layers)	Previous eLife niche map (*Pigott et al., 2014*)
AUC	0.8236 ± 0.08	0.8195 ± 0.08	0.85 ± 0.04
Occurrences	n = 57 (animals), n = 31 (humans)	n = 57 (animals), n = 31 (humans)	n = 51 (animals), n = 30 (humans)
Ranked relative contributions	EVI mean (0.55)	EVI mean (0.46)	EVI mean (0.65)
	Group 1 bat distribution (0.18)	*Hypsignathus monstrosus* (0.15)	Elevation (0.12)
	LST mean (night) (0.08)	*Epomops franqueti* (0.08)	LST mean (night) (0.08)
	Elevation (0.06)	*Otomops martiensseni* (0.06)	PET mean (0.06)
	LST mean (day) (0.04)	*Epomophorus labiatus* (0.04)	Bat distribution (0.04)

DOI: 10.7554/eLife.16412.011

such as these convey the heterogeneities in spillover risk that exist within Africa. To better enable researchers and policymakers to consider EVD preparedness and necessary contingencies, a new online tool has been developed which allows users to interrogate the revised maps in more detail, in areas of specific interest (http://vizhub.healthdata.org/ebola). As part of this tool, the zoonotic niche output and Group 1 bat layers are now available, along with filters for identifying at-risk countries and locations of previous index cases from outbreaks.

Geographic datasets such as these provide context to broader discussions as our aspirations transition from controlling outbreaks to mitigating the risk of future spillover events prioritised by their potential for more widespread epidemics. Such data are particularly important for determining where best to investigate the frequency of potentially transmissible contacts between reservoir and susceptible species and humans. Previous niche maps served as an important impetus in the search for potential reservoirs (**Peterson et al., 2004**) and these iterations can continue to inform such work. As researchers and policy makers seek to resolve outstanding questions about EVD epidemiology, it is hoped that the continued updating and dissemination of this information can contribute to this discussion.

Materials and methods

Updating the occurrence database

Since the previous publication, an outbreak of EVD occurred in humans in the Democratic Republic of the Congo (**Maganga et al., 2014**). The outbreak is thought to have originated in Inkanamongo, a village near Boende, Équateur province and resulted in 66 probable and confirmed cases and 49 probable and confirmed deaths (**Rosello et al., 2015**). A polygon of radius 10 km centered on the town of Boende was included to capture the location of the index case, increasing the database of assumed independent animal-to-human spillover events to 31 as part of 24 distinct reported outbreaks (**Mylne et al., 2014**).

In addition, a re-analysis of the literature available on infections in animal species was completed on 7th October 2015. Due to the poor differential capacity of immunological tests to discriminate Ebola virus from other viruses we retained the following inclusion criteria for the database; for susceptible species mortality events linked to Ebolavirus by any diagnostic methodology *or* PCR-positive diagnosis of Ebolavirus were included. Inclusion criteria for potential bat reservoir species were either PCR-positivity or serological evidence suggesting Ebolavirus infection. Serological studies were included without fatal outcomes (unlike with susceptible species) due to the hypothesised asymptomatic nature of infection in the reservoir hosts. As a result of these inclusion criteria, studies with serological detection of Ebolavirus in healthy non-Chiropteran species were excluded, such as surveys in dogs (**Allela et al., 2005**).

In total six new records of EVD occurrence in animals were identified and included within the database to increase the total to 57. These records were obtained from two research articles. The first of these assessed Ebolavirus load in a variety of mammal species and identified PCR-positivity in a number of small mammals across three sites in Central African Republic (**Morvan et al., 1999**). In total, four separate occurrences, consisting of three different species, were identified as being PCR positive: a member of the *Praomys* complex, Peter's mouse (*Mus setulosus*) and the greater forest shrew (*Sylvisorex ollula*). The second study investigated serological responses in straw-coloured fruit bats (*Eidolon helvum*) caught in two districts in Zambia (**Ogawa et al., 2015**). Specific latitudes and longitudes of the study sites were supplied for the Central African Republic study and were used to generate point occurrences. For the Zambian study it was necessary to use administrative data representing the two districts where the bats were caught (Serenje and Ndola districts).

Expanding potential bat reservoir species

Potential bat reservoir species were stratified into three groupings based upon the strength of evidence suggesting their reservoir status (**Table 3**). Group 1 contained the three species of bat from which Ebolavirus RNA has been detected and therefore have the strongest evidence to support potential reservoir status (**Leroy et al., 2005**). Group 2 species are those that, using a variety of serological tests, have been reported to be Ebolavirus seropositive, suggesting potential reservoir status. A previous review (**Olival and Hayman, 2014**), identified nine species as seropositive for

Table 3. Final bats included in analysis classified by evidence grouping.

Grouping	Bat	Occurrences
Group 1	Franquet's epauletted fruit bat (*Epomops franqueti*)	442
	Hammerheaded fruit bat (*Hypsignathus monstrosus*)	254
	Little collared fruit bat (*Myonycteris torquata*)	107
Group 2	Angolan free-tailed bat (*Tadarida condylura*, formerly *Mops condylurus*)	179
	Egyptian fruit bat (*Rousettus aegyptiacus*)	177
	Gambian epauletted fruit bat (*Epomophorus gambianus*)	166
	Peter's dwarf epauletted fruit bat (*Micropteropus pusillus*)	208
	Straw-coloured fruit bat (*Eidolon helvum*)	282
Group 3	Buettikofer's epauletted fruit bat (*Epomops buettikoferi*)	50
	Common bent-wing bat (*Miniopterus schreibersii*)	31
	Eloquent horseshoe bat (*Rhinolophus eloquens*)	61
	Ethiopian epauletted fruit bat (*Epomophorus labiatus*)	187
	Giant leaf-nosed bat (*Hipposideros gigas*)	21
	Greater long-fingered bat (*Miniopterus inflatus*)	56
	Large-eared free-tailed bat (*Otomops martiensseni*)	33

DOI: 10.7554/eLife.16412.012

Ebolavirus. This was reduced to five species after the removal of the three species already categorised in Group 1 and Leschenault's Rousette, *Rousettus leschenaultii*, which is not found in Africa.

Finally, Group 3 species were identified *via* generalized boosted regression analysis, which discriminates the bats reported to be filovirus-positive by learning trait patterns that distinguish them from all other bat species (***Han et al., 2016***). Generalized boosted regression (***Elith et al., 2008***) was applied to traits describing all bat species, including life history, physiological, ecological, morphological and demographic variables collected from numerous published sources. In addition to traits, the filovirus status of each bat species was assigned as a binary score (0 – not currently known to be positive for any filoviruses; 1 – published evidence). This analysis produces a rank list of all bat species according to their probability of being a filovirus carrier on the basis of their trait similarities with known filovirus-positive bat species. Bats found in the 90th percentile of likely filovirus carriers were initially considered, and then filtered to include only those which have home ranges in Africa (***Schipper et al., 2008***). As per the original publication, occurrence records were extracted from the Global Biodiversity Information Facility (GBIF). Species for which there were fewer than 20 unique GBIF records in Africa were dropped from the analysis due to data paucity. ***Table 3*** reports the bat species and corresponding numbers of occurrences included in the analysis.

For Group 1 species, occurrence records were supplemented by searching PubMed and Web of Knowledge for additional reports. A literature review was completed on the 8th September 2015 using the following sets of keywords:

- 'Hypsignathus monstrosus' or 'hammer-headed bat' or 'hammer headed bat' or 'hammer-headed bat' or 'big-lipped bat' or 'big lipped bat' or 'Hypsignathus labrosus' or 'Hypsignathus macrocephalus'
- 'Myonycteris torquata' or 'little collared fruit bat' or 'Myonycteris collaris' or 'Myonycteris leptodon' or 'Myonycteris wroughtoni'
- 'Epomops franqueti' or "Franquet's epauletted fruit bat" or 'Epomops comptus' or 'Epomops strepitans'

A total of 34 articles were identified for inclusion, from which 564 additional occurrences were sourced.

All bat species were modelled separately using boosted regression trees (**Elith et al., 2008**) utilising the same modelling procedure as outlined in the original article except that 100, rather than 50, bootstrap models were fitted. This resulted in 15 individual environmental suitability maps for bat species (see **Figure 2—figure supplements 1**, **2** and **3**), as well as three consolidated bat layers combining the environmental suitability maps for the bats within each of the three groupings (**Figure 2**).

Revising the predicted zoonotic niche map

A species distribution model, specifically a boosted regression trees approach (**Elith et al., 2008**), was implemented. The model generates regression trees based upon binary splits of linked covariates, which are iteratively improved upon by boosting. The regression trees are capable of characterising complex environmental interactions and correlations since each tree is built from a hierarchy of multiple nodes, each based upon different successive binary splits of the covariates. The model extracts environmental information for each reported occurrence of Ebolavirus to define an optimal relationship between presence of the disease and environmental factors. Predictive performance is improved by including a comparison background dataset that acts as a hypothesised environmental negative control (**Phillips et al., 2009**). As per the previous analysis, this dataset was generated by randomly sampling across Africa biased towards areas of high population density. By including human population density in this way, some potential sampling biases present in human index case reporting can be mitigated as cases are more likely to be reported in more populous areas. The boosted regression trees were re-run using the same parameters and covariates (elevation, mean evapotranspiration rate, and mean and range measures of enhanced vegetation index, daytime land surface temperature (LST), and night-time LST) as the previous publication except for the inclusion of the new occurrence data outlined above and the new bat layers. Two model iterations were run: one with the three consolidated bat layers (i.e. Groups 1, 2 and 3) and the other with all the bat species layers considered separately.

Estimating populations at risk

The continuous suitability surface was converted into a binary at-risk versus not-at-risk surface by determining a threshold value that included 95% of the estimated suitability values of pixels with reported human index cases (**Pigott et al., 2015**). For sites represented by a specific latitude and longitude the suitability score was taken from the corresponding pixel; for polygon estimates covering a number of cells, the mean suitability was taken across all pixels covered by the polygon.

Whilst not included directly as a covariate in the modelling process, human population layers were assessed in at-risk locations as a potential proxy for spillover frequency. The populations living within the gridded cells thought to be at-risk of potential Ebolavirus transmission from zoonotic sources were calculated using an updated contemporary gridded estimate of population (**WorldPop Project, 2015**).

Acknowledgements

This work was funded by a grant from The Paul G. Allen Family Foundation (#11878) to SIH which also provided funding for DMP, AIM, LE and CM. BAH acknowledges support from the National Institute of General Medical Sciences of the National Institutes of Health (#U01GM110744). FMS is funded by the Rhodes Trust. OJB is funded by the Bill & Melinda Gates Foundation (#OPP1119467). MUGK is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship. CLM and NG acknowledge support from the Bill & Melinda Gates Foundation (#OPP1053338).

PWG is a Career Development Fellow (#K00669X) jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and receives support from the Bill & Melinda Gates Foundation (#OPP1068048, #OPP1106023) which also supports DJW and SB. SIH is also funded by a Senior Research Fellowship from the Wellcome Trust (#095066) and grants from the Bill & Melinda Gates Foundation (#OPP1093011, #OPP1132415). DMP and BAH also acknowledge the support for a collaborative meeting from the RAPIDD program of the Science & Technology Directorate. SIH also acknowledges a grant from the Research for Health in Humanitarian Crises (R2HC) Programme, managed by ELHRA (#13468) and funded equally by the Wellcome Trust and DFID, which also supported NG and MUGK.

Additional information

Competing interests

SIH: Reviewing editor, *eLife*. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Paul G. Allen Family Foundation	11878	David M Pigott Anoushka I Millear Lucas Earl Chloe Morozoff Simon I Hay
Science and Technology Directorate	RAPIDD program	David M Pigott Barbara A Han
National Institute of General Medical Sciences	U01GM110744	Barbara A Han
Rhodes Scholarships		Freya M Shearer
Bill and Melinda Gates Foundation	OPP1068048	Daniel J Weiss Samir Bhatt Peter W Gething
Bill and Melinda Gates Foundation	OPP1106023	Daniel J Weiss Samir Bhatt Peter W Gething
Bill and Melinda Gates Foundation	OPP1119467	Oliver J Brady
German Academic Exchange Service London	Graduate Scholarship	Moritz UG Kraemer
Research for Health in Humanitarian Crises (R2HC) Programme	ELHRA (#13468)	Moritz UG Kraemer Nick Golding Simon I Hay
Wellcome Trust		Moritz UG Kraemer Nick Golding
Department for International Development		Moritz UG Kraemer Nick Golding
Bill and Melinda Gates Foundation	OPP1053338	Catherine L Moyes Nick Golding
Medical Research Council	MRC/DFID Concordat agreement	Peter W Gething
Department for International Development	MRC/DFID Concordat agreement	Peter W Gething
Wellcome Trust	Senior Research Fellowship #095066	Simon I Hay
Bill and Melinda Gates Foundation	OPP1093011	Simon I Hay

Bill and Melinda Gates Foundation	OPP1132415	Simon I Hay

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

DMP, SIH, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; AIM, BAH, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; LE, CM, FMS, OJB, MUGK, Analysis and interpretation of data, Drafting or revising the article; DJW, Acquisition of data, Drafting or revising the article; CLM, NG, Conception and design, Analysis and interpretation of data, Drafting or revising the article; SB, PWG, Conception and design, Drafting or revising the article

Author ORCIDs

Chloe Morozoff, http://orcid.org/0000-0002-3254-5553
Moritz UG Kraemer, http://orcid.org/0000-0001-8838-7147
Catherine L Moyes, http://orcid.org/0000-0002-8028-4079
Simon I Hay, http://orcid.org/0000-0002-0611-7272

Additional files

Major datasets

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
David M Pigott, Anoushka I Millear, Lucas Earl, Chloe Morozoff, Barbara A Han, Freya M Shearer, Daniel J Weiss, Oliver J Brady, Moritz UG Kraemer, Catherine L Moyes, Samir Bhatt, Peter W Gething, Nick Golding, Simon I Hay	2016	eLife Advances - Updates to the zoonotic niche map of Ebola virus disease in Africa	http://dx.doi.org/10.6084/m9.figshare.c.3267965.v1	Publicly available at figshare (https://figshare.com/)

References

Allela L, Boury O, Pouillot R, Délicat A, Yaba P, Kumulungui B, Rouquet P, Gonzalez JP, Leroy EM. 2005. Ebola virus antibody prevalence in dogs and human risk. *Emerging Infectious Diseases* **11**:385–390. doi: 10.3201/eid1103.040981

Amman BR, Carroll SA, Reed ZD, Sealy TK, Balinandi S, Swanepoel R, Kemp A, Erickson BR, Comer JA, Campbell S, Cannon DL, Khristova ML, Atimnedi P, Paddock CD, Crockett RJ, Flietstra TD, Warfield KL, Unfer R, Katongole-Mbidde E, Downing R, et al. 2012. Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLoS Pathogens* **8**: e1002877. doi: 10.1371/journal.ppat.1002877

Bogoch II, Creatore MI, Cetron MS, Brownstein JS, Pesik N, Miniota J, Tam T, Hu W, Nicolucci A, Ahmed S, Yoon JW, Berry I, Hay SI, Anema A, Tatem AJ, MacFadden D, German M, Khan K. 2015. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *The Lancet* **385**:29–35. doi: 10.1016/S0140-6736(14)61828-6

Currie J, Grenfell B, Farrar J. 2016. Infectious diseases. Beyond Ebola. *Science* **351**:815–816. doi: 10.1126/science.aad8521

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**:802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Funk S, Piot P. 2014. Mapping Ebola in wild animals for better disease control. *eLife* **3**:e04565. doi: 10.7554/eLife.04565

Han BA, Schmidt JP, Alexander LW, Bowden SE, Hayman DTS, Drake JM. 2016. Undiscovered bat hosts of filoviruses. *PLoS Neglected Tropical Diseases* **10**:e0004815. doi: 10.1371/journal.pntd.0004815

Haydon DT, Cleaveland S, Taylor LH, Laurenson MK. 2002. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerging Infectious Diseases* **8**:1468–1473. doi: 10.3201/eid0812.010317

Hayman DT. 2015. Biannual birth pulses allow filoviruses to persist in bat populations. *Proceedings of the Royal Society B* **282**:20142591. doi: 10.1098/rspb.2014.2591

Kraemer MU, Hay SI, Pigott DM, Smith DL, Wint GR, Golding N. 2016. Progress and challenges in infectious disease cartography. *Trends in Parasitology* **32**:19–29. doi: 10.1016/j.pt.2015.09.006

Kuhn JH. 2008. Filoviruses. A compendium of 40 years of epidemiological, clinical, and laboratory studies. *Archives of Virology. Supplementum* **20**:13–360.

Leendertz S. 2016. Testing new hypotheses regarding ebolavirus reservoirs. *Viruses* **8**:30. doi: 10.3390/v8020030

Leendertz SA, Gogarten JF, Düx A, Calvignac-Spencer S, Leendertz FH. 2016. Assessing the evidence supporting fruit bats as the primary reservoirs for Ebola viruses. *EcoHealth* **13**:18–25. doi: 10.1007/s10393-015-1053-0

Leirs H, Mills JN, Krebs JW, Childs JE, Akaibe D, Woollen N, Ludwig G, Peters CJ, Ksiazek TG. 1999. Search for the Ebola virus reservoir in Kikwit, Democratic Republic of the Congo: reflections on a vertebrate collection. *Journal of Infectious Diseases* **179 Suppl 1**:S155–163. doi: 10.1086/514299

Leroy EM, Epelboin A, Mondonge V, Pourrut X, Gonzalez JP, Muyembe-Tamfum JJ, Formenty P. 2009. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases* **9**:723–728. doi: 10.1089/vbz.2008.0167

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT, Gonzalez JP, Swanepoel R. 2005. Fruit bats as reservoirs of Ebola virus. *Nature* **438**:575–576. doi: 10.1038/438575a

Maganga GD, Kapetshi J, Berthet N, Kebela Ilunga B, Kabange F, Mbala Kingebeni P, Mondonge V, Muyembe JJ, Bertherat E, Briand S, Cabore J, Epelboin A, Formenty P, Kobinger G, González-Angulo L, Labouba I, Manuguerra JC, Okwo-Bele JM, Dye C, Leroy EM. 2014. Ebola virus disease in the Democratic Republic of Congo. *New England Journal of Medicine* **371**:2083–2091. doi: 10.1056/NEJMoa1411099

Mari Saez A, Weiss S, Nowak K, Lapeyre V, Zimmermann F, Dux A, Kuhl HS, Kaba M, Regnaut S, Merkel K, Sachse A, Thiesen U, Villanyi L, Boesch C, Dabrowski PW, Radonic A, Nitsche A, Leendertz SAJ, Petterson S, Becker S, et al. 2015. Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine* **7**:17–23. doi: 10.15252/emmm.201404792

Moon S, Sridhar D, Pate MA, Jha AK, Clinton C, Delaunay S, Edwin V, Fallah M, Fidler DP, Garrett L, Goosby E, Gostin LO, Heymann DL, Lee K, Leung GM, Morrison JS, Saavedra J, Tanner M, Leigh JA, Hawkins B, et al. 2015. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the harvard-LSHTM independent panel on the global response to Ebola. *The Lancet* **386**:2204–2221. doi: 10.1016/S0140-6736(15)00946-0

Morvan JM, Deubel V, Gounon P, Nakouné E, Barrière P, Murri S, Perpète O, Selekon B, Coudrier D, Gautier-Hion A, Colyn M, Volehkov V. 1999. Identification of Ebola virus sequences present as RNA or DNA in organs of terrestrial small mammals of the Central African Republic. *Microbes and Infection* **1**:1193–1201. doi: 10.1016/S1286-4579(99)00242-7

Mylne A, Brady OJ, Huang Z, Pigott DM, Golding N, Kraemer MUG, Hay SI. 2014. A comprehensive database of the geographic spread of past human Ebola outbreaks. *Scientific Data* **1**:140042. doi: 10.1038/sdata.2014.42

Ogawa H, Miyamoto H, Nakayama E, Yoshida R, Nakamura I, Sawa H, Ishii A, Thomas Y, Nakagawa E, Matsuno K, Kajihara M, Maruyama J, Nao N, Muramatsu M, Kuroda M, Simulundu E, Changula K, Hang'ombe B, Namangala B, Nambota A, et al. 2015. Seroepidemiological prevalence of multiple species of filoviruses in fruit bats (*Eidolon helvum*) migrating in Africa. *Journal of Infectious Diseases* **212 Suppl 2**:S101–108. doi: 10.1093/infdis/jiv063

Olival KJ, Hayman DT. 2014. Filoviruses in bats: current knowledge and future directions. *Viruses* **6**:1759–1788. doi: 10.3390/v6041759

Peterson AT, Bauer JT, Mills JN. 2004. Ecologic and geographic distribution of filovirus disease. *Emerging Infectious Diseases* **10**:40–47. doi: 10.3201/eid1001.030125

Peterson AT, Carroll DS, Mills JN, Johnson KM. 2004. Potential mammalian filovirus reservoirs. *Emerging Infectious Diseases* **10**:2073–2081. doi: 10.3201/eid1012.040346

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**:181–197. doi: 10.1890/07-2153.1

Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, Brady OJ Kraemer MUG, Smith DL, Moyes CL, Bhatt S, Gething PW, Horby PW, Bogoch II, Brownstein JS, Mekaru SR, Tatem AJ, Khan K, Hay SI. , 2014. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**:e04395. doi: 10.7554/eLife.04395

Pigott DM, Golding N, Mylne A, Huang Z, Weiss DJ, Brady OJ, Kraemer MU, Hay SI. 2015. Mapping the zoonotic niche of Marburg virus disease in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **109**:366–378. doi: 10.1093/trstmh/trv024

Plowright RK, Eby P, Hudson PJ, Smith IL, Westcott D, Bryden WL, Middleton D, Reid PA, McFarlane RA, Martin G, Tabor GM, Skerratt LF, Anderson DL, Crameri G, Quammen D, Jordan D, Freeman P, Wang LF, Epstein JH, Marsh GA, et al. 2015. Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B* **282**:20142124. doi: 10.1098/rspb.2014.2124

Rosales-Chilama M, Gongora RE, Valderrama L, Jojoa J, Alexander N, Rubiano LC, Cossio A, Adams ER, Saravia NG, Gomez MA. 2015. Parasitological Confirmation and Analysis of Leishmania Diversity in Asymptomatic and

Subclinical Infection following Resolution of Cutaneous Leishmaniasis. *PLoS Neglected Tropical Diseases* **9**: e0003652. doi: 10.1371/journal.pntd.0004273

Rosello A, Mossoko M, Flasche S, Van Hoek AJ, Mbala P, Camacho A, Funk S, Kucharski A, Ilunga BK, Edmunds WJ, Piot P, Baguelin M, Tamfum JJ. 2015. Ebola virus disease in the Democratic Republic of the Congo, 1976-2014. *eLife* **4**:e09015. doi: 10.7554/eLife.09015

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, Lamoreux J, Rodrigues AS, Stuart SN, Temple HJ, Baillie J, Boitani L, Lacher TE, Mittermeier RA, Smith AT, Absolon D, Aguiar JM, Amori G, Bakkour N, Baldi R, et al. 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**:225–230. doi: 10.1126/science.1165115

Toledo J, George L, Martinez E, Lazaro A, Han WW, Coelho GE, Runge Ranzinger S, Horstick O, Han BA. 2016. Relevance of Non-communicable Comorbidities for the Development of the Severe Forms of Dengue: A Systematic Literature Review. *PLoS Neglected Tropical Diseases* **10**:e0004284. doi: 10.1371/journal.pntd.0004284

Viana M, Mancy R, Biek R, Cleaveland S, Cross PC, Lloyd-Smith JO, Haydon DT. 2014. Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution* **29**:270–279. doi: 10.1016/j.tree.2014.03.002

Walsh PD, Bermejo M, Rodriguez-Teijeiro JD. 2009. Disease avoidance and the evolution of primate social connectivity. In: *Primate Parasite Ecology: The Dynamics and Study of Host-Parasite Relationships*. 183–197.

Walsh PD, Breuer T, Sanz C, Morgan D, Doran-Sheehy D. 2007. Potential for Ebola transmission between gorilla and chimpanzee social groups. *American Naturalist* **169**:684–689. doi: 10.1086/513494

Woldehanna S, Zimicki S. 2015. An expanded one health model: integrating social science and one health to inform study of the human-animal interface. *Social Science & Medicine* **129**:87–95. doi: 10.1016/j.socscimed.2014.10.059

World Health Organization. 2016. Ebola situation reports. Available at http://www.who.int/csr/disease/ebola/en/. Accessed March 14, 2016.

WorldPop Project. 2015. Gridded population surface of Africa. Available at http://www.worldpop.org.uk/. Accessed November 6, 2015.

Transactions of the Royal Society of Tropical Medicine and Hygiene Advance Access published March 27, 2015



Trans R Soc Trop Med Hyg
doi:10.1093/trstmh/trv024

ORIGINAL ARTICLE

Mapping the zoonotic niche of Marburg virus disease in Africa

David M. Pigott[a,*], Nick Golding[a], Adrian Mylne[a], Zhi Huang[a], Daniel J. Weiss[a], Oliver J. Brady[a], Moritz U. G. Kraemer[a] and Simon I. Hay[a,b]

[a]*Spatial Ecology & Epidemiology Group, Department of Zoology, University of Oxford, Oxford, UK;*
[b]*Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA*

*Corresponding author: Tel: +44 1865 271137; E-mail: david.pigott@zoo.ox.ac.uk

Received 19 December 2014; revised 20 February 2015; accepted 23 February 2015

Background: Marburg virus disease (MVD) describes a viral haemorrhagic fever responsible for a number of outbreaks across eastern and southern Africa. It is a zoonotic disease, with the Egyptian rousette (*Rousettus aegyptiacus*) identified as a reservoir host. Infection is suspected to result from contact between this reservoir and human populations, with occasional secondary human-to-human transmission.

Methods: Index cases of previous human outbreaks were identified and reports of infection in animals recorded. These data were modelled within a species distribution modelling framework in order to generate a probabilistic surface of zoonotic transmission potential of MVD across sub-Saharan Africa.

Results: Areas suitable for zoonotic transmission of MVD are predicted in 27 countries inhabited by 105 million people. Regions are suggested for exploratory surveys to better characterise the geographical distribution of the disease, as well as for directing efforts to communicate the risk of practices enhancing zoonotic contact.

Conclusions: These maps can inform future contingency and preparedness strategies for MVD control, especially where secondary transmission is a risk. Coupling this risk map with patient travel histories could be used to guide the differential diagnosis of highly transmissible pathogens, enabling more rapid response to outbreaks of haemorrhagic fever.

Keywords: Boosted regression trees, Filovirus, Marburg virus disease, *Rousettus aegyptiacus*, Species distribution models, Viral haemorrhagic fever

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

Introduction

In 1967, outbreaks of a previously undescribed disease in workers of three laboratories in West Germany and Yugoslavia were reported, characterised by high fever, haemorrhaging and organ failure.[1] A novel virus, named Marburg virus (MARV), the first described in the *Filoviridae* family, was subsequently identified as the causative pathogen.[2] In 1975, the first recognised case of the disease outside of a laboratory occurred in Rhodesia (now Zimbabwe), with one case in 1980 due to MARV and another in 1987 due to Ravn virus (RAVV), another marburgvirus.[3] Not until 1998, when a series of fatal haemorrhagic cases were identified in the vicinity of Durba, Democratic Republic of the Congo (DRC), was a large-scale outbreak reported. A total of 154 cases were reported, with the source of infection traced back to bat colonies in local gold mines.[4] While a large number of cases were reported between 1998 and 2000, it was found that multiple introductions of the virus from the same zoonotic pool were responsible for the continued outbreak rather than only human-to-human transmission, as more commonly reported with Ebola virus disease (EVD).[4–6] In 2004 however, a large outbreak in Uige province, Angola, occurred where, unlike in Durba, continued cases were driven by subsequent human-to-human transmission rather than repeated introductions from a natural source.[7] More recent outbreaks have been smaller in comparison (Figure 1).[8–12]

The wider epidemiology of Marburg virus disease (MVD) remains relatively unknown (Figure 2). While non-human primates are susceptible to the disease, there have been no reported transmission events from primates to humans outside of a laboratory setting. Furthermore, no significant epizootics have been reported among non-human primates, unlike the closely related ebolaviruses.[13,14] Past outbreaks have strongly implicated bats as the origin of initial index cases in humans. Serological and molecular surveys conducted in caves and

© *The Author 2015. Published by Oxford University Press on behalf of Royal Society of Tropical Medicine and Hygiene.*
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

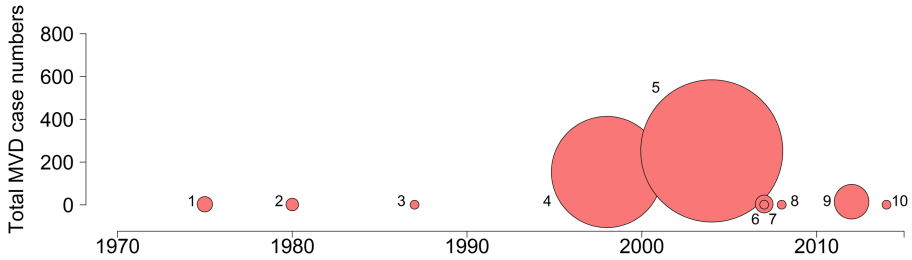


Figure 1. Case numbers in previous Marburg virus disease outbreaks. The size of each circle is proportional to the number of cases of the disease in a given outbreak. Outbreaks are labelled as per Table 1. This figure is available in black and white in print and in colour at Transactions online.

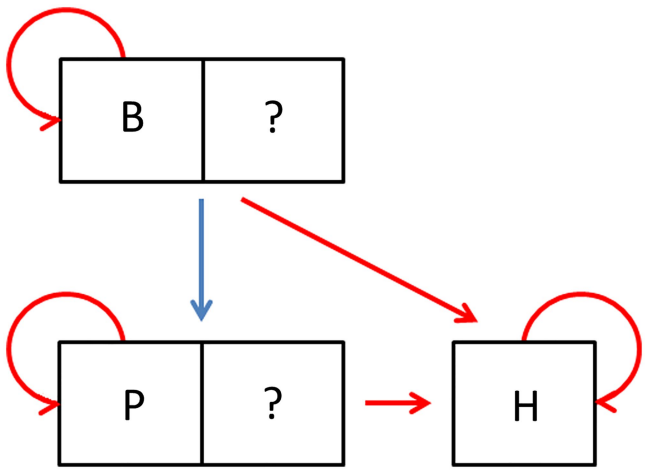


Figure 2. The epidemiology of marburgvirus transmission in Africa. B represents suspected bat reservoirs (including Egyptian rousettes). Susceptible animals include non-human primates, such as the monkeys responsible for the 1967 outbreaks (P). H represents humans. Question marks indicate potential animals of other species. All routes have been confirmed or are suspected to occur apart from transmission between bats and primates, which remains unknown. Adapted from Laminger and Prinz and Groseth et al.[15,16] This figure is available in black and white in print and in colour at Transactions online.

mines visited by the infected individuals, have identified the virus in the Egyptian rousette (*Rousettus aegyptiacus*).[17,18] Colonies of bats have also been reported in the vicinity of the supposed index case in other outbreaks.[19,20]

In order to better understand the nature of MVD risk, this study attempts to define those areas where zoonotic transmission of MVD may occur in order to identify people at potential risk of zoonotic spillover. Such a methodology has previously been employed with EVD to identify 22 western and equatorial African nations where ebolavirus transmission may occur.[21] Ecological niche modelling of MVD has previously been undertaken and this work seeks to update these efforts by including more recent outbreaks, records of infection in animals, improved environmental covariate layers and recent advances in modelling techniques.[21–25] The need for such information is critical, not only to assist in differential diagnosis of fevers across Africa, but also to increase awareness of the potential risk of more widespread outbreaks that could arise from a delay in the response to initial cases.[26]

Materials and methods

Methodological overview

A species distribution model, specifically an ensemble boosted regression trees (BRT) framework, was used to model the zoonotic niche of MVD. This model optimally builds ensembles of trees based upon binary decisions used to classify suitable environmental covariates in reference to a database of known occurrence locations.[27,28] Areas which are environmentally similar to those with reported zoonotic transmission of MVD are predicted to be at higher levels of risk. To perform this analysis, we obtained four key information components: 1. a database of cases where MVD has been transmitted from animals to humans; 2. reported infections of MARV and RAVV in animals; 3. a collection of spatially gridded environmental variables that are likely to be correlates of disease presence; and 4. background (pseudo-absence) data indicating locations where MVD has not been reported. The model was restricted to the African continent since there have been no reported natural outbreaks, in humans or animals, outside this region.

Identifying human and animal infections with marburgviruses

Outbreaks of MVD in humans were identified from review articles and by sourcing original references.[29] Where possible, index cases (individuals infected by animal reservoir species) were located and the supposed location of animal to human transfer of MARV and RAVV was geopositioned using Google Earth. When an accurate site location could not be determined, a geographic area (termed a polygon) was defined covering the reported region, identified using the source articles (e.g., a specified landmark, or an area referenced in relation to another directly identifiable site); otherwise a precise, site-specific latitude and longitude was recorded. For larger settlements, the centroid of the site was recorded. In some instances, only the first reported patient could be identified, with little information on the initial route of infection. In these instances we assumed that the index case occurred where the zoonotic transmission event took place. For some outbreaks there was sufficient evidence to suggest multiple independent zoonotic transmission events. For these outbreaks, each individual transmission event was separately positioned.

To obtain a comprehensive database of MARV infections in animals, a literature search was conducted in Web of Science using the search term 'Marburg reservoir OR Marburg monkey OR Marburg bat OR Marburg primate'. This procedure returned 1544

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

unique citations. Abstracts for these citations were processed and where they indicated that the article might contain spatial information on Marburg infections in animals the full articles were obtained. Once identified, the references of these articles (as well as more general review articles discussing MVD reservoirs) were followed up in case relevant articles were omitted from the initial literature search. Locations of infected animals were geopositioned using the same methodology as for human index cases.

Covariates used in the analyses

A suite of ecologically relevant gridded environmental covariates for Africa was compiled, each having a nominal resolution of 5 km × 5 km. A number of environmental covariates thought to potentially influence MVD distribution were selected for inclusion in this analysis, namely range and mean values of enhanced vegetation index (EVI) and land surface temperature (LST) (day and night) derived from satellite data and parsed through gap-filling algorithms, as well as elevation and potential evapotranspiration (PET).[21,25,30,31] Many of these have been considered in previous investigations.[24] In addition, distance to the nearest Karst formation was included as a covariate.[32] Karst landscapes typically form when soluble rocks dissolve and can create expansive cave networks and as such were used in the model as a proxy for the subterranean roosting habitat of the supposed disease reservoir, the Egyptian rousette.[33,34] Previous work mapping the zoonotic niche of EVD utilised a bat distribution covariate layer. While attempts were made to replicate this approach for MVD, the lack of sufficiently detailed data available from the Global Biodiversity Information Facility to allow for differentiation between roosting and foraging sites meant that the niche modelling approaches were unable to produce reliable results and therefore these outputs could not be included in the final analysis.

Marburg distribution modelling

An ensemble boosted regression trees model was used to define areas environmentally suitable for zoonotic MARV transmission. The model requires both presence and background information to generate a prediction, the latter of which is often hard to collect systematically and in an unbiased manner. As a result, randomly generated background records are often supplied. For this study, a background record dataset was generated by randomly sampling 10 000 locations across Africa, biased towards more populous areas as a proxy for reporting bias.[22] This sampling allows for comparison of factors influencing presence and likely absence locations for MVD by the model. In total, 500 submodels were used. Each submodel was fitted using the gbm.step subroutine in the dismo package in the R statistical programming environment.[28,35,36] Given the limited number of records available, we reduced the number of cross-validation folds used to fit the model to three, from the default of 10. All other tuning parameters of the algorithm were held at their default values (tree complexity=4, learning rate=0.005, bag fraction=0.75, step size=10). For each polygon in the occurrence dataset, one point was randomly selected from within the defined area for each submodel. This Monte Carlo procedure enabled the model to efficiently integrate over the environmental uncertainty associated with imprecise geographic data. A bootstrap sample was then taken from each of these datasets and used to train the BRT model. For each submodel, weightings were applied to the background dataset so that the sum of the weighted background data equalled the weighted sum of the occurrence records.[37] This was done in order to improve the discrimination capacity of the model. Each submodel predicts environmental suitability on a continuous scale from 0 to 1. An ensemble final prediction map was generated by combining the predictions from these submodels, calculating the mean prediction as well as the 5% and 95% confidence intervals around this for each 5 km × 5 km pixel.

Two models were constructed. Model 1 used only records of human index cases and model 2 used both human index cases and reported infection in animals. This was done in order to augment the relatively small number of index case records available and to evaluate the influence of including animal data on the model.

The area under the curve (AUC) statistic was used to assess model accuracy. The statistic was calculated for each submodel using a three-fold cross validation, and then summarised across all the submodels to generate a mean and standard deviation for this value. This procedure divided the dataset into three subsets that had approximately equal numbers of presence records and background data. Due to the small number of presence records used to train each submodel, this approach represents a very thorough test of the model's predictive ability. In order to prevent inflation of the accuracy statistics due to spatial sorting bias, a pairwise distance sampling procedure was used.[38] As a result, these AUC statistics are lower than typical outputs, but give a more realistic evaluation of the ability of the model to predict for different regions.[39] Uncertainty in the prediction was evaluated by considering the difference between the 5% and 95% confidence intervals.

The final outputs represent the environmental suitability for zoonotic transmission of MARV for each 5 km × 5 km pixel which allows for relative comparison of risk across Africa.

Population living in areas of environmental suitability for zoonotic transmission

Estimates of population living in areas at risk of zoonotic transmission were derived by converting the continuous surface of transmission risk into a binary at-risk/not-at-risk classification for each pixel. The threshold for this classification was based upon the minimum environmental suitability value at the locations of the occurrence records. To calculate this value, the risk estimate for each point occurrence and the mean probability of each area/polygon occurrence were assessed. Countries were classified into two categories of risk. Set 1 are countries where index cases of MVD have been reported and set 2 are countries where no index cases have been previously reported and have more than 100 pixels (i.e., approximately 2500 km^2) at risk. The number of people living in these pixels was calculated from existing population surfaces for Africa.[40,41]

Contiguous areas of risk within each country were visually identified and the latitude and longitude for the approximate midpoint for these areas were recorded, suggesting areas of potential interest for further prospective epidemiological investigation.

The R code used for all of the analysis is freely available via https://github.com/SEEG-Oxford/marburg_zoonotic.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

Results

Reported infections in humans and animals

A total of ten distinct outbreaks of MVD were identified, ranging in size from single reported cases to community-wide outbreaks with hundreds of reported cases (Table 1). Five countries have confirmed or suspected instances of animal-to-human zoonotic transmission, namely Kenya, Uganda, Zimbabwe, Angola and the DRC (Figure 3). For the majority of these outbreaks, caves or mines have been singled out as the likely venue for their spill-over events. Some records were of individuals who had subsequently travelled elsewhere before becoming symptomatic, for which efforts were made to identify the original site of infection.[3,9,10]

All available animal infection records were from bat populations, often sampled in response to human cases, with the exception of one reported infection in grivets (the same animals responsible for the 1967 laboratory-based outbreak) (Table 2). These monkeys were trapped near Kidera and Namsale in Uganda where they were assumed to have been originally infected.[46] Where epidemiological surveys of nearby potential animal reservoirs were undertaken during, or shortly after outbreaks in humans, PCR identification of MARV was often performed.[17,18] A serological survey of Gabonese bat populations reported positivity in Egyptian rousettes and other bats.[47] This is the only evidence of animal infection occurring outside the recognised range of reported human populations, all other animal infections have been reported in the vicinity of human outbreaks (Figure 3).

Predicted environmental suitability for zoonotic transmission of marburgviruses

Due to the relative paucity of data, two model variants were used in order to test various assumptions about the poorly understood MVD epidemiology. Model 1, which only included human index case data, identified geological features (elevation and distance to Karst formation) and vegetation indices (both EVI mean and range) as the main predictors of suitability for zoonotic transmission (Table 3). Model 2, which included the entire dataset of MARV infections, implied a broader spatial extent, with environmental factors (EVI, LST and PET) playing a more important role in prediction compared to elevation. The AUC values were 0.64 ± 0.12 and 0.62 ± 0.08 for models 1 and 2, respectively, indicating that both the models demonstrated similar predictive skill. Note however that as these statistics were calculated using different evaluation datasets, they are not directly comparable. Uncertainty maps for the predicted surfaces for MVD are presented in Supplementary Figures 1 and 2.

Both models predict high suitability for zoonotic transmission in the set 1 countries. In total, model 2 predicts 27 countries to be at-potential-risk (set 1 and 2) of zoonotic transmission of MARV with 105 million people living in at-risk areas. Model 1 predicts 19 countries at risk with 75 million individuals living in at-risk areas. These 19 countries are consistently predicted to be at-risk in both models 1 and 2.

Discussion

This work utilises all known outbreaks of MVD in humans and reported infections in animals in order to understand the nature of risk posed by this disease (Figures 4 and 5). Previous assessments have indicated that a much broader region is at-risk of zoonotic transmission than those countries that have reported transmission to-date.[24] Our analysis, reinforced by new outbreak reports and environmental covariate information, is in concordance with previous ecological modelling investigations of MVD, identifying temperature and vegetation indices as key determinants of its spatial distribution.[23,24,50] In addition, we identify the potential importance of geological features in influencing areas of potential MARV risk. The majority of at-risk populations live in areas that have previously reported outbreaks, mainly Uganda, Kenya and the DRC. Amongst countries yet to see human infection (set 2), the most notable are Ethiopia, Cameroon and Zambia, in which large areas are predicted to be at-risk.

Table 1. Locations of natural outbreaks of Marburg virus disease in humans

Outbreak	Date range	Countries	Location	Cases/deaths	Reference
1[a]	Feb 1975	Zimbabwe	Chinoyi caves	3/1	3,24
2	Jan 1980	Kenya	Nzoia	2/1	20
3	Aug 1987	Kenya	Kitum cave	1/1	19
4	Oct 1998-Aug 2000	DRC	Durba	154/128	4,42
5	Oct 2004-Jul 2005	Angola	Uige province	252/227	7,43,44,45
6	Jun 2007-Sept 2007	Uganda	Kitaka gold mine	4/1	8
7[a]	Dec 2007-Jan 2008	Uganda	Python cave	1/0	9
8[a]	Jul 2008	Uganda	Python cave	1/1	10
9	Aug 2012-Oct 2012	Uganda	Ibanda district	15/14	11
10	Sep 2014	Uganda	Mpigi/Mawokota district	1/1	12

DRC: Democratic Republic of the Congo.
[a] Indicates case imported elsewhere.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

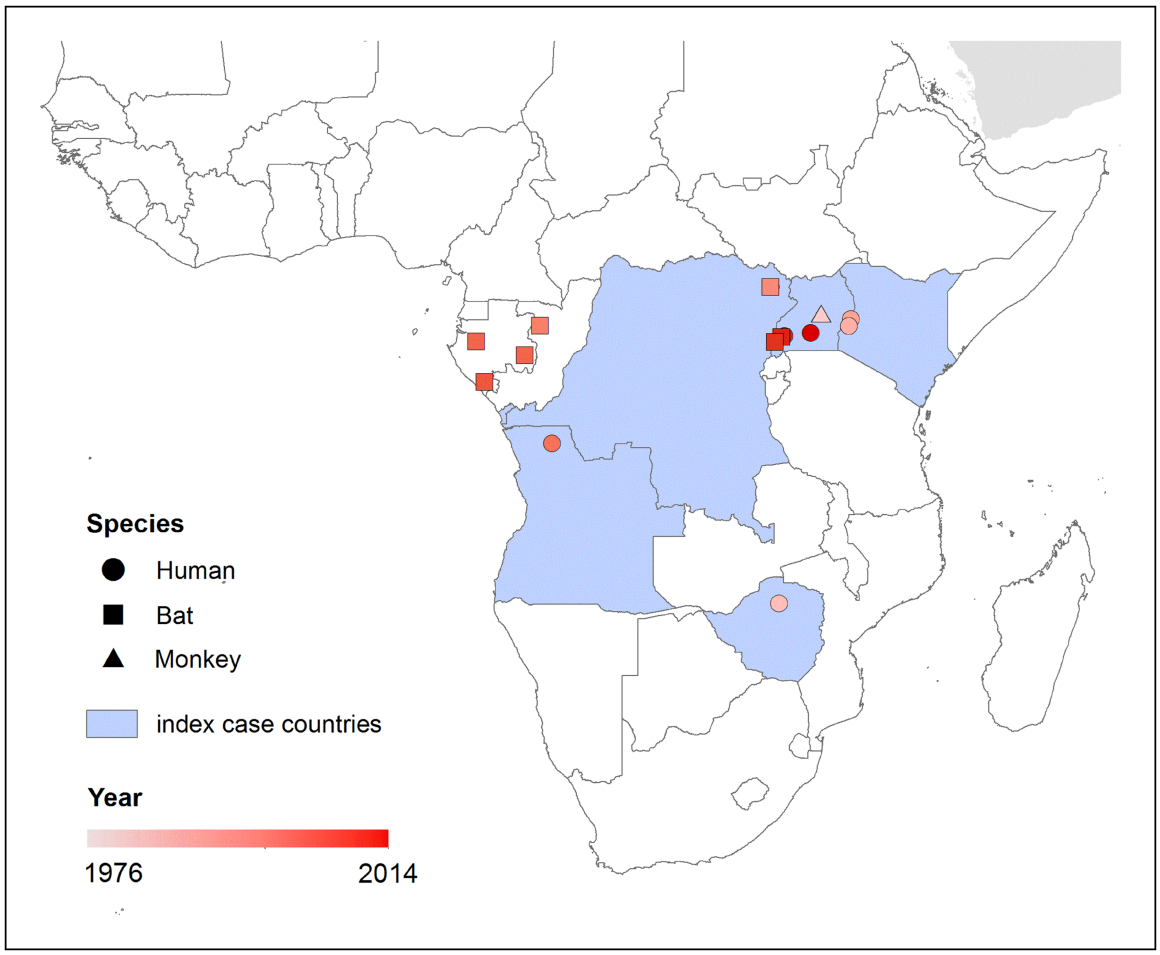


Figure 3. The locations of marburgvirus disease outbreaks in humans and reported animal infections across Africa. This figure is available in black and white in print and in colour at Transactions online.

Table 2. Locations of reported infections with Marburg virus in animals

Site	Date range	Country	Location	Species	Diagnosis	Reference
1	Aug-Sep 1967	Uganda	Kidera and Namasale	*Chlorocebus aethiops*	Serology	42
2	May-Oct 1999	DRC	Durba	*Miniopterus inflatus*	PCR	16
3	May-Oct 1999	DRC	Durba	*Rhinolophus eloquens*	PCR	16
4	May-Oct 1999	DRC	Durba	*Rousettus aegyptiacus*	PCR	16
5	Jun 2003-May 2006	ROC	Mbomo district	Various bat species	Serology	43
6	Feb 2005-Mar 2008	Gabon	Haut-Ogooue district	Various bat species	Serology	43
7	Apr 2005	Gabon	Moyen-Ogooue district	Various bat species	Serology	43
8	Feb 2006	Gabon	Nyanga district	Various bat species	Serology	43
9	Aug 2007	Uganda	Kitaka gold mine	*Rousettus aegyptiacus*	PCR	15
10	Aug 2007	Uganda	Kitaka gold mine	*Hipposideros* spp.	PCR	15
11	Apr 2008	Uganda	Kitaka gold mine	*Rousettus aegyptiacus*	PCR	15
12	Aug 2009	Uganda	Python cave	*Rousettus aegyptiacus*	PCR	48
13	Nov 2009	Uganda	Python cave	*Rousettus aegyptiacus*	PCR	48
14	Nov 2012	Uganda	Kitaka gold mine	*Rousettus aegyptiacus*	PCR	49

DRC: Democratic Republic of the Congo; ROC: Republic of Congo.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

Table 3. Summary statistics for model outputs

Statistic	Model 1: human data	Model 2: human and animal data
AUC (± standard deviation)	0.64±0.12	0.62±0.08
1st predictor	Elevation: 49.1%	Mean EVI: 49.1%
2nd predictor	Mean EVI: 30.1%	Night-time mean LST: 19.5%
3rd predictor	EVI range: 7.8%	Elevation: 10.9%
4th predictor	Distance to Karst: 4.9%	Mean PET: 7.5%
5th predictor	Night-time mean LST: 3.1%	Day-time mean LST: 5.0%

Relative contributions for each of the top five predictors are reported as a percentage.
AUC: area under the curve; EVI: enhanced vegetation index; LST: land surface temperature; PET: potential evapotranspiration.

As with any model-based approach, an awareness of the limitations of the data and the assumptions made by the model is important. Limited datasets, particularly those where definitive identification of zoonotic transmission sites is unlikely, will hinder predictive capability. However, this study attempts to be as comprehensive as possible by including all reports of infections, as well as considering uncertainty in geopositioning ability. Further information can only help to improve these predictions. Similarly, the model is only able to assess environmental suitability for MVD, therefore in order to translate this into true outbreak risk, additional information on how humans and animal reservoirs interact, as well as how the disease is transmitted within these populations is required. Bearing these caveats in mind, we hope that these results will act as a springboard for further research to better understand the epidemiology and characterise the risk of this disease.

The two model iterations ('human only' versus 'animal and human data') illustrate the need for further research into MARV hosts and their potential for zoonotic transmission to humans. Areas predicted at-risk in model 1 are consistently identified at-risk in model 2 (although the absolute probability of transmission is altered); the inclusion of animal data however, expands the areas of potential risk to include countries across western and central Africa not indicated as being at-risk by model 1. Figure 6 visualises the differences between these two models. As a result, given the limited data availability, model 2 would currently be the most sensible option when discussing the potential risk posed by MVD. In addition, while no reported cases of MVD have been recorded in set 2 countries, a number have seen serological evidence of past exposure in humans.[29] Seropositive individuals have been reported in locations identified as at-risk in model 2 in West Africa, Cameroon, Central African Republic, Nigeria and South Sudan.[48,49,51–56]

Since many MVD spillover events have only resulted in a handful of cases, the likelihood of outbreaks going unrecognised is a possibility.[57] It is therefore also possible that spatial variation in the probability of cases being identified may have biased our models. While we strived to account for such an observation bias in our analysis by weighting pseudo-absence records to areas where infection might be more likely to be detected, we cannot rule out the presence of residual bias. The true nature of zoonotic transmission potential within these countries can only be elucidated by additional surveys.

Knowledge on the animal reservoir for MVD is limited. Egyptian rousettes have consistently been identified as PCR positive for the virus, however animals of a number of other species have also been seropositive.[15,43,58,59] The maps presented here can be used to target key sites for future surveys of bats to better understand the true nature of risk within those areas where no previous outbreak has been reported.

There is considerable overlap between the reported distribution of Egyptian rousettes (Figure 7) and areas of highest risk. Evidence suggests that there are various subspecies of Egyptian rousette across Africa.[33] All but one outbreak of MVD in humans occurred within the known range of members of *R. aegyptiacus leachi*; the outbreak in Uige Province, Angola, however occurred outside the range of bats of this subspecies, but was within the reported range of *R. aegyptiacus unicolor*. It remains unclear whether these populations differ in disease transmission cycles and the nature of the connectivity between bats of these two potential subspecies has important implications for potential disease transmission, either restricting the likely areas of risk to eastern and southern Africa, or including much of central and west Africa (Figure 7B). Similarly, it is possible that bats of subspecies present in north Africa and the Middle East could also be potential reservoirs for MARV. The inclusion of bat distributions in future models would allow for a better understanding of the relationship between MVD and Egyptian rousettes, with the possibility of identifying regions where other bats may be more likely reservoir hosts.

In addition, further surveys for MARV infection in bats in these regions therefore would not only help to better understand the ecology of these bats but also the nature of the risk posed to human populations. As with EVD, spillover of MARV into humans is rare and infection in bat populations also appears uncommon. Understanding the nature of infection within these bat and other potential reservoirs, is crucial in identifying the true nature of risk to human populations, not just for MVD, but also a variety of other viral pathogens.[61] Table 4 identifies the main regions within each at-risk country where such surveillance activities would be of greatest benefit. The output maps, to allow for national survey placement, are freely available from the following link: http://goo.gl/0qTOfe

Our risk maps provide a baseline estimate for the extent of the zoonotic niche of MVD, which can subsequently be enhanced through more specific research. While an area may have the potential for zoonotic transmission of MVD, if humans rarely interact with these animal hosts, spillover events are unlikely to occur. As a result, in spite of a large number of individuals living in areas where transmission is possible, a considerably smaller number will be at-risk of encountering an infected reservoir and subsequently being infected. Surveys and ethnographic assessments can help better understand the true nature of risk within these regions, particularly important if a quantitative

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

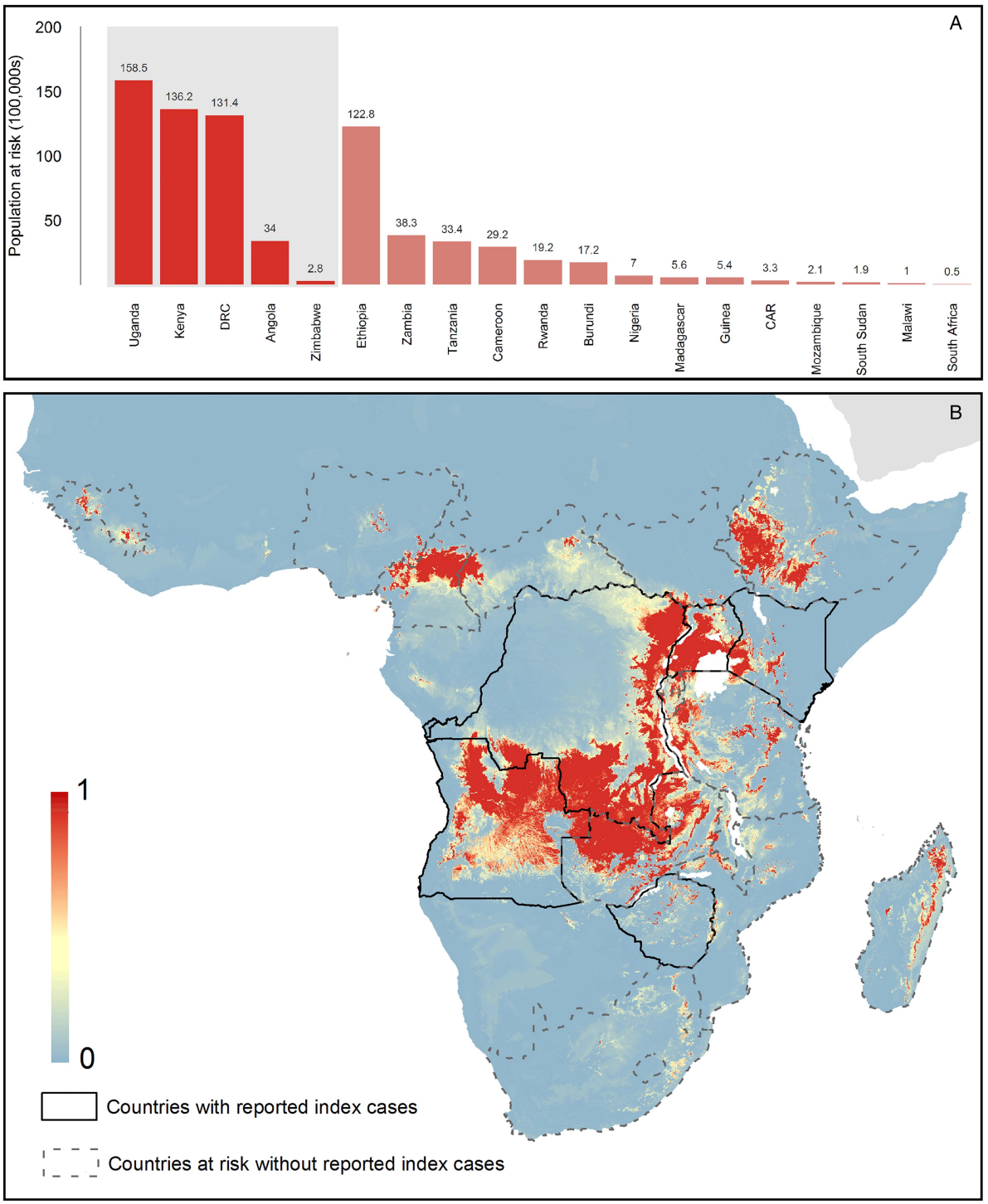


Figure 4. Predicted geographical distribution of the zoonotic niche for marburgviruses using model 1 – human index cases only. Panel A shows the total populations living in areas of risk of zoonotic transmission for each at-risk country. The grey rectangle highlights countries in which index cases of disease have been reported (set 1); the remainder are countries in which risk of zoonotic transmission is predicted, but in which index cases of Marburg virus disease have not been reported and have more than 100 at-risk pixels (set 2). These countries are ranked by population-at-risk within each set. The population-at-risk figure in 100 000 s is given above each bar. Panel B shows the predicted distribution of zoonotic marburgviruses. The scale reflects the relative probability that zoonotic transmission of marburgviruses could occur at these locations; areas closer to 1 (red) are more likely to harbour zoonotic transmission than those closer to 0 (blue). Countries with borders outlined are those which are predicted to contain at-risk areas for zoonotic transmission based on a thresholding approach (see Methods). The area under the curve statistic, calculated under a stringent cross-validation procedure is 0.64±0.12. Solid lines represent set 1 whilst dashed lines delimit set 2. Areas covered by major lakes have been masked white.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

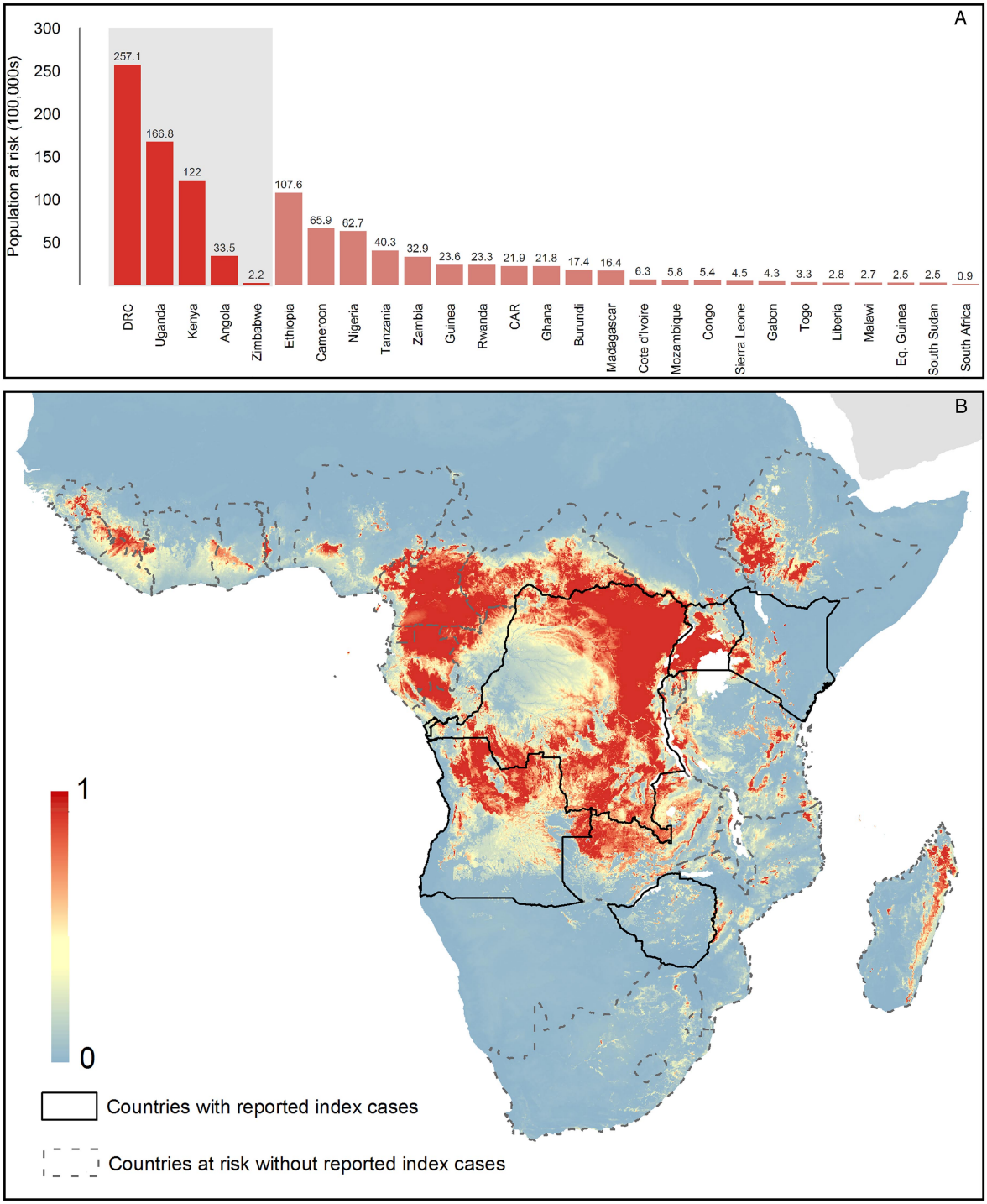


Figure 5. Predicted geographical distribution of the zoonotic niche for marburgviruses using model 2–both human index cases and infections in animals. Panel A shows the total populations living in areas of risk of zoonotic transmission for each at-risk country. The grey rectangle highlights countries in which index cases of Marburg virus disease have been reported (set 1); the remainder are countries in which risk of zoonotic transmission is predicted, but in which index cases of Marburg have not been reported and have more than 100 at-risk pixels (set 2). These countries are ranked by population-at-risk within each set. The population-at-risk figure in 100 000 s is given above each bar. Panel B shows the predicted distribution of zoonotic marburgviruses. The scale reflects the relative probability that zoonotic transmission of marburgviruses could occur at these locations; areas closer to 1 (red) are more likely to harbour zoonotic transmission than those closer to 0 (blue). Countries with borders outlined are those which are predicted to contain at-risk areas for zoonotic transmission based on a thresholding approach (see Methods). The area under the curve statistic, calculated under a stringent cross-validation procedure, is 0.62 ± 0.08. Solid lines represent set 1 whilst dashed lines delimit set 2. Areas covered by major lakes have been masked white.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

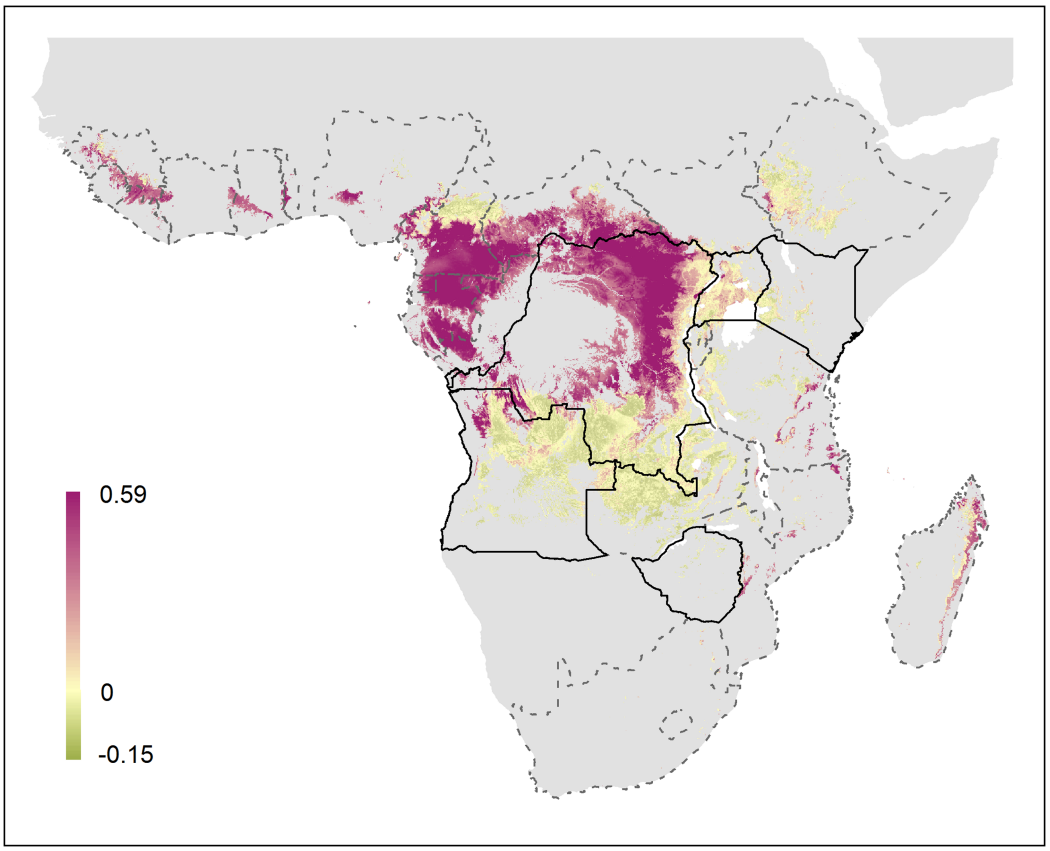


Figure 6. Difference between model predictions with animal data omitted. The difference between outputs for model 2 and model 1 are presented. Pixels in purple represent those regions predicted at higher risk in model 2; regions in green indicate areas where model 1 predicts higher risk. Yellow pixels represent areas with consistent probabilities. Pixels predicted not to be at-risk are in grey.

assessment of outbreak likelihood is wanted. While environmental triggers have been linked to outbreaks of MVD, equally important in determining outbreak potential is an understanding of the dynamics of the virus within reservoir populations, which has also been shown to be highly variable.[44,58,62] In attempting to predict outbreaks it is therefore crucial to understand the interplay between environmental factors, human pressures and reservoir host dynamics.[63]

The differences in areas predicted to be at-risk of infection by models 1 and 2 may in fact reflect the manner in which humans interact with bats. Since the majority of human outbreaks have arisen from contact in caves or other underground systems (rather than in the forest foraging sites of the bats, where animal infections have been reported), the risk map derived from model 1 could be a spatial representation of this transmission risk, as opposed to reflecting the broader distribution of infection in animal populations. Further infection surveys and ethnographic research can help to elucidate and map the risk of animal-human transmission within the at-risk region we have identified. Such an analysis would be particularly important in order to produce an absolute, rather than relative estimate of the likelihood of an outbreak in humans.

Nevertheless, while the true nature of risk to humans is likely to be a function of a variety of different factors, it is still important to gauge how and where potential spillover events could occur. The west African outbreak of EVD has shown that it is critical to understand the potential for such outbreaks in geographically distinct areas, and the subsequent need for other causes to be included in the differential diagnosis to facilitate rapid detection. This is all the more important where the potential causes of disease have varying potential for nosocomial transmission, as is the case with viral haemorrhagic fevers. Failure to rapidly and accurately diagnose these diseases can lead to uncontrolled chains of secondary transmission in certain scenarios.[64] Maps such as ours can therefore be used to shape clinical recommendations for diagnosing haemorrhagic fever cases presenting in hospital. MVD has seen a number of significant geographic translocation of cases, with individuals becoming symptomatic far from the original infection site.[9,10] The most recent outbreak of EVD in west Africa has demonstrated the role that both local and global connectivity can play in causing disease importation, and as connectivity continues to increase, the likelihood of widespread secondary cases occurring will also increase, particularly if infection reaches densely populated areas.[65–67] Accounting for a range of possible aetiological agents can therefore reduce the risk of further secondary transmission amongst humans in these settings.[3,26]

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

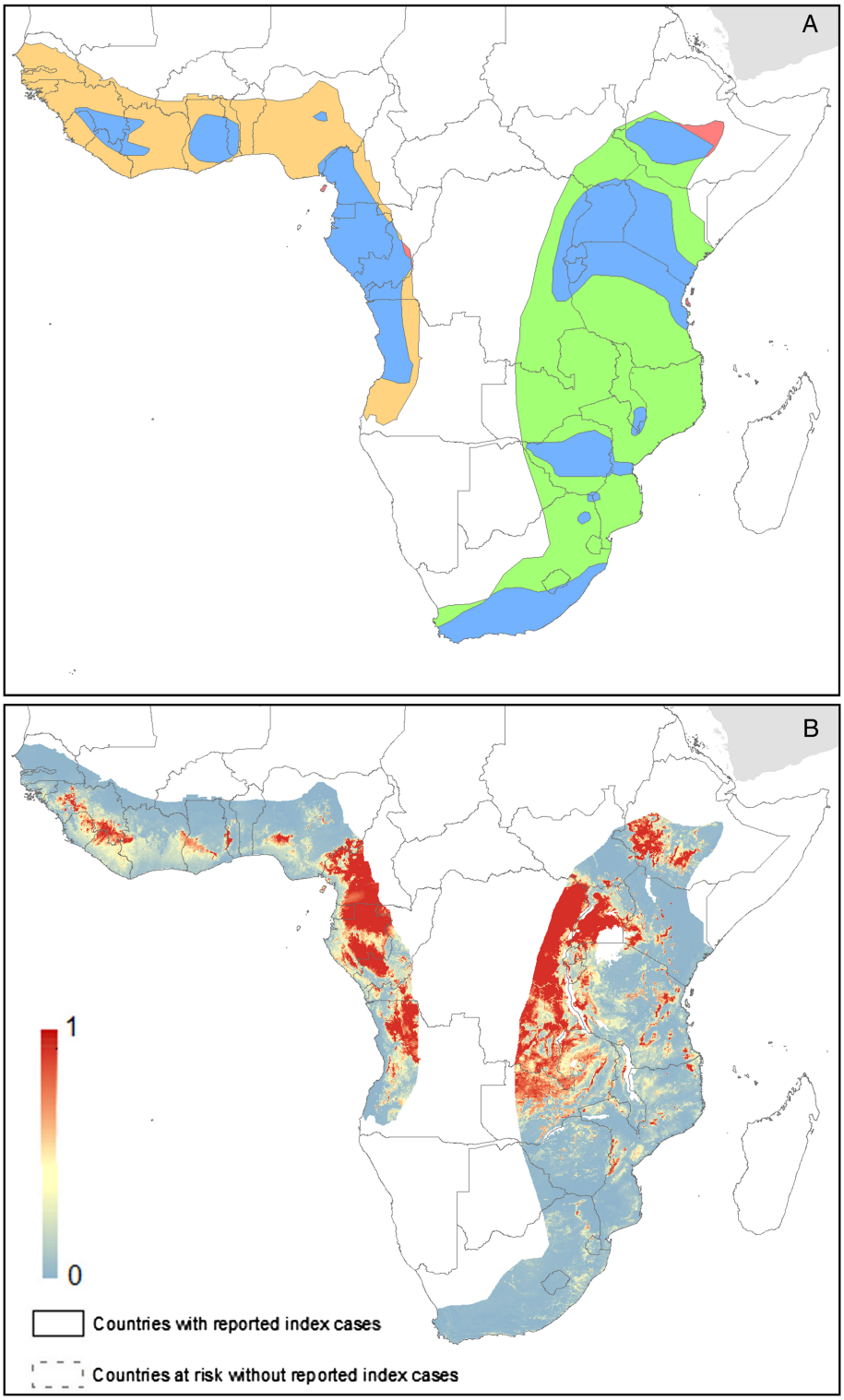


Figure 7. Expert opinion maps for the range of Egyptian rousettes. Panel A is derived from the IUCN and Kwiecinski et al.[33,60] Blue regions are those where both depict Egyptian rousette populations. Red areas are those only indicated in the IUCN dataset. Orange sections are where bats of the subspecies *R. aegyptiacus unicolor* are thought to be present; green shows the distribution of bats of the subspecies *R. aegyptiacus leachi*. Panel B shows the predicted values from model 2, masked by the bat layer.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

Table 4. Identification of potential survey sites in at-risk countries

Country	Region	Latitude	Longitude
Angola	Northern Angola	−7.83	16.07
	Moxico	−12.28	23.29
Burundi	Western Burundi	−3.94	29.44
Cameroon	Central and Southern Cameroon	4.73	12.76
CAR	South CAR	5.95	18.91
Côte d'Ivoire	Dix-Huit Montagnes	7.62	−7.89
	Bondoukou	7.96	−3.03
DRC	Eastern Congo Basin	2.04	27.72
	Katanga	−10.70	26.23
	Kasai	−6.42	21.95
	Western DRC	−7.03	18.30
Equatorial Guinea	Kie-Ntem and Wele-Nzas Provinces	1.55	10.83
Ethiopia	Western Oromia	8.29	35.77
	Bale Mountains and Harenna Forest	6.07	39.14
Gabon	Northern Gabon	1.07	12.36
	Southern Gabon	−1.83	11.92
Ghana	Ashanti Uplands and Kwahu Plateau	6.99	−1.61
	Akwapim-Togo Range	8.01	0.51
Guinea	Moyenne Guinea	10.71	−12.47
	Guinea Highlands	8.23	−9.09
Kenya	Lake Victoria	−0.70	34.99
	Mount Kenya	−0.43	37.49
Liberia	Guinea Highlands–Wologizi and Wonegizi Ranges	8.15	−9.87
	Guinea Highlands–Nimba Range	7.41	−8.59
Madagascar	Tsaratanana and Marojejy Nature Reserves	−14.21	49.38
	Ambohijanahay Reserve	−18.38	45.45
Malawi	Lake Malawi	−11.89	33.90
Mozambique	Maeda Plateau	−11.44	39.63
	Mount Mabu	−16.30	36.54
	Inyanga Mountains	−19.64	33.29
Nigeria	Ekiti	7.76	5.09
	Gashaka Gumti National Park	7.36	11.57
ROC	Niari and Lekoumou	−2.92	13.19
	Sangha	1.56	14.53
	Likouala	3.17	16.92
Rwanda	Eastern Rwanda	−2.14	30.37
Sierra Leone	Loma Mountains	8.99	−10.98
South Africa	Woodbush Forest Reserve and Motlatse Canyon	−23.94	30.12
South Sudan	Imatong Mountains	3.95	32.64
	Yambio	4.57	28.29
Tanzania	Tanga	−5.94	37.57
	Mahale Mountain National Park	−6.49	30.30
	Kigoma	−4.49	30.64
	Morogoro	−9.58	35.60
Togo	Akwapim-Togo Range	7.59	0.79
Uganda	Great Lakes Region	0.18	31.01
Zambia	Northwestern Zambia	−12.80	24.76
	Northern Zambia	−9.49	29.23
	Muchinga Escarpment	−11.89	31.84
Zimbabwe	Chinoyi	−17.36	30.13

Sites are identified by country. The latitude and longitude represent the centroid of the proposed survey area.
CAR: Central African Republic; DRC: Democratic Republic of Congo; ROC: Republic of Congo.

Authors' disclaimer: Funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Authors' contributions: SIH conceived the work; DMP, AM, OJB and MUGK collected the data; DMP and NG implemented the modelling; DJW generated the environmental covariate layers; ZH geopositioned the outbreak data and provided all maps. All authors read and approved the final manuscript. DMP is the guarantor of the paper.

Acknowledgments: We thank Dr. Jens Kuhn for his comments and Maria Devine, Joshua Longbottom and Freya Shearer for proofreading.

Funding: DMP is funded by a Sir Richard Southwood Graduate Scholarship from the Department of Zoology at the University of Oxford; NG is funded by a grant from the Bill & Melinda Gates Foundation [OPP1053338]; ZH is funded by the Bill & Melinda Gates Foundation [OPP1106023]; DJW is funded by the Bill & Melinda Gates Foundation [OPP1068048]; OJB is funded by a BBSRC studentship; MUGK is funded by the German Academic Exchange Service (DAAD) through a graduate scholarship; SIH is funded by a Senior Research Fellowship from the Wellcome Trust [095066] that also supports AM and a grant from the Bill & Melinda Gates Foundation [OPP1093011]. SIH would also like to acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Competing interests: None declared.

Ethical approval: Not required.

References

1 Slenczka WG. The Marburg virus outbreak of 1967 and subsequent episodes. Curr Top Microbiol Immunol 1999;235:49–75.

2 Siegert R, Shu H-L, Slenczka W et al. On the eitology of an unknown infection originating in monkeys [in German]. Dtsch med Wochenschr 1967;92:2341–3.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

3 Conrad JL, Isaacson M, Smith EB et al. Epidemiologic investigation of Marburg virus disease, Southern Africa, 1975. Am J Trop Med Hyg 1978;27:1210–5.

4 Bausch DG, Nichol ST, Muyembe-Tamfum JJ et al. Marburg hemorrhagic fever associated with multiple genetic lineages of virus. N Engl J Med 2006;355:909–19.

5 Mylne A, Brady OJ, Huang Z et al. A comprehensive database of the geographic spread of past human Ebola outbreaks. Sci Data 2014;1:e140042.

6 Gire SK, Goba A, Andersen KG et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 2014;345:1369–72.

7 Towner JS, Khristova ML, Sealy TK et al. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. J Virol 2006;80:6497–516.

8 Adjemian J, Farnon EC, Tschioko F et al. Outbreak of Marburg hemorrhagic fever among miners in Kamwenge and Ibanda districts, Uganda, 2007. J Infect Dis 2011;204:S796–9.

9 Fujita N, Miller A, Miller G et al. Imported case of Marburg hemorrhagic fever - Colorado, 2008. MMWR Morb Mortal Wkly Rep 2009;58: 1377–81.

10 Timen A, Koopmans MPG, Vossen AC et al. Response to imported case of Marburg hemorrhagic fever, the Netherlands. Emerg Infect Dis 2009;15:1171–5.

11 Mbonye A, Wamala J, Winyi-Kaboyo et al. Repeated outbreaks of viral hemorrhagic fevers in Uganda. Afr Health Sci 2012;12:579–83.

12 WHO. Marburg virus disease - Uganda. Geneva: World Health Organization; 2014. http://www.who.int/csr/don/10-october-2014-marburg/en/ [accessed 15 December 2014].

13 Bermejo M, Rodriguez-Teijeiro JD, Illera G et al. Ebola outbreak killed 5000 gorillas. Science 2006;314:1564.

14 Formenty P, Boesch C, Wyers M et al. Ebola virus outbreak among wild chimpanzees living in a rain forest of Côte d'Ivoire. J Infect Dis 1999;179:S120–6.

15 Groseth A, Feldmann H, Strong JE. The ecology of Ebola virus. Trends Microbiol 2007;15:408–16.

16 Laminger F, Prinz A. Bats and other reservoir hosts of Filoviridae. Danger of epidemic on the African continent? a deductive literature analysis [in German]. Wien Klin Wochenschr 2010;122:19–30.

17 Towner JS, Amman BR, Sealy TK et al. Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. PLoS Pathog 2009;5:e1000536.

18 Swanepoel R, Smit SB, Rollin PE et al. Studies of reservoir hosts for Marburg virus. Emerg Infect Dis 2007;13:1847–51.

19 Johnson ED, Johnson BK, Silverstein D et al. Characterization of a new Marburg virus isolated from a 1987 fatal case in Kenya. Arch Virol Suppl 1996;11:101–14.

20 Smith DH, Johnson KM, Saacson M et al. Marburg virus disease in Kenya. Lancet 1982;1:816–20.

21 Pigott DM, Golding N, Mylne A et al. Mapping the zoonotic niche of Ebola virus disease in Africa. eLife 2014;3:e04395.

22 Phillips SJ, Dudik M, Elith J et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 2009;19:181–97.

23 Peterson AT, Bauer JT, Mills JN. Ecologic and geographic distribution of filovirus disease. Emerg Infect Dis 2004;10:40–7.

24 Peterson AT, Lash RR, Carroll DS, Johnson KM . Geographic potential for outbreaks of Marburg hemorrhagic fever. Am J Trop Med Hyg 2006;75:9–15.

25 Weiss DJ, Atkinson PM, Bhatt S et al. An effective approach for gap-filling continental scale remotely sensed time-series. ISPRS J Photogramm Remote Sens 2014;98:106–18.

26 Bannister B. Viral haemorrhagic fevers imported into non-endemic countries: risk assessment and management. Br Med Bull 2010;95: 193–225.

27 De'ath G. Boosted trees for ecological modeling and prediction. Ecology 2007;88:243–51.

28 Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol 2008;77:802–13.

29 Kuhn JH. Filoviruses: a compendium of 40 years of epidemiological, clinical, and laboratory studies. Vienna: Springer-Verlag; 2008.

30 Wan ZM, Li ZL. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. IEEE Trans Geosci Remote Sens 1997;35:980–96.

31 ORNL DAAC Shuttle radar topography mission near-global digital elevation models. http://webmap.ornl.gov/wcsdown/wcsdown.jsp?dg_id=10008_1 [accessed 1 August 2014].

32 USF Geoportal and Data Repository Karst regions of the world. http://arcweb.forest.usf.edu/flex/KarstRegions/ [accessed 6 October 2014].

33 Kwiecinski GG, Griffiths TA. *Rousettus egyptiacus*. J Mammal 1999; 611:1–9.

34 Ford D, Williams P. Karst hydrogeology and geomorphology. Chichester: Wiley; 2007.

35 Hijmans RJ, Phillips S, Leathwick J, Elith J. Package dismo. R package. http://cran.r-project.org/web/packages/dismo/dismo.pdf [accessed 1 August 2014].

36 R Core Team R: a language and environment for statistical computing. www.R-project.org/ [accessed 1 November 2014].

37 Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 2012;3:327–38.

38 Hijmans RJ. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 2012;93:679–88.

39 Wenger SJ, Olden JD. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evol 2012;3:260–7.

40 WorldPop. WorldPop project. http://www.worldpop.org.uk/ [accessed 1 August 2014].

41 Linard C, Gilbert M, Snow RW et al. Population distribution, settlement patterns and accessibility across Africa in 2010. PLoS One 2012;7:e31743.

42 Borchert M, Muyembe-Tamfum JJ, Colebunders R et al. A cluster of Marburg virus disease involving an infant. Trop Med Int Health 2002;7:902–6.

43 Jeffs B, Roddy P, Weatherill D et al. The Médecins Sans Frontières intervention in the Marburg hemorrhagic fever epidemic, Uige, Angola, 2005. I. Lessons learned in the hospital. J Infect Dis 2007;196: S154–61.

44 Roddy P, Weatherill D, Jeffs B et al. The Médecins Sans Frontières intervention in the Marburg hemorrhagic fever epidemic, Uige, Angola, 2005. II. Lessons learned in the community. J Infect Dis 2007;196(Supp 2):S162–7.

45 Roddy P, Thomas SL, Jeffs B et al. Factors associated with Marburg hemorrhagic fever: analysis of patient data from Uige, Angola. J Infect Dis 2010;201:1909–18.

46 Henderson BE, Kissling RE, Williams MC et al. Epidemiological studies in Uganda relating to the Marburg agent. In: Martini GA, Siegert R, editors. Marburg virus disease. Berlin: Springer-Verlag; 1971.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

47 Pourrut X, Souris M, Towner JS et al. Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in *Rousettus aegyptiacus*. BMC Infect Dis 2009;9:159.

48 Gonzalez JP, Nakoune E, Slenczka W et al. Ebola and Marburg virus antibody prevalence in selected populations of the Central African Republic. Microbes Infect 2000;2:39–44.

49 Tomori O, Fabiyi A, Sorungbe A et al. Viral hemorrhagic fever antibodies in Nigerian populations. Am J Trop Med Hyg 1998;38:407–10.

50 Lash RR, Brunsell NA, Peterson AT. Spatiotemporal environmental triggers of Ebola and Marburg virus transmission. Geocarto Int 2008;23:451–66.

51 Schoepp RJ, Rossi CA, Khan SH et al. Undiagnosed acute viral febrile illnesses, Sierra Leone. Emerg Infect Dis 2014;20:1176–82.

52 Neppert J, Gohring S, Schneider W, Wernet P. No evidence of LAV infection in the Republic of Liberia, Africa, in the year 1973. Blut 1986;53:115–7.

53 Bergmann JF, Bouree P. Haemorrhagic fever due to Ebola. A study of 1517 sera from Cameroon. Med Maladies Infect 1982;12:638–42.

54 Saluzzo JF, Gonzalez JP, Georges AJ. Anti-Marburg antibodies in the Republic of Central Africa. Ann Virol 1982;133:129–31.

55 Johnson ED, Gonzalez JP, Georges A. Haemorrhagic fever virus activity in equatorial Africa: distribution and prevalence of filovirus reactive antibody in the Central African Republic. Trans R Soc Trop Med Hyg 1993;87:530–5.

56 Woodruff PWR, Morrill JC, Burans JP et al. A study of viral and rickettsial exposure and causes of fever in Juba, southern Sudan. Trans R Soc Trop Med Hyg 1988;82:761–6.

57 Gire SK, Stremlau M, Andersen KG et al. Emerging disease or diagnosis? Science 2012;338:750–2.

58 Amman BR, Carroll SA, Reed ZD et al. Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. PLoS Pathog 2012;8: e1002877.

59 Amman BR, Nyakarahuka L, McElroy AK et al. Marburgvirus resurgence in Kitaka mine bat population after extermination attempts, Uganda. Emerg Infect Dis 2014;20:1761–4.

60 Schipper J, Chanson JS, Chiozza F et al. The status of the world's land and marine mammals: diversity, threat, and knowledge. Science 2008;322:225–30.

61 Wood JLN, Leach M, Waldman L et al. A framework for the study of zoonotic disease emergence and its drivers: spillover of bat pathogens as a case study. Philos Trans R Soc Lond B Biol Sci 2012; 367:2881–92.

62 Hayman DT. Biannual birth pulses allow filoviruses to persist in bat populations. Proc R Soc Lond B Biol Sci 2015;282:pii: 20142591.

63 Plowright RK, Eby P, Hudson PJ et al. Ecological dynamics of emerging bat virus spillover. Proc Biol Sci 2015;282:20142124.

64 Farrar JJ, Piot P. The Ebola emergency - immediate action, ongoing strategy. N Engl J Med 2014;371:1545–6.

65 Bogoch II, Creatore MI, Cetron MS et al. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. Lancet 2015;385:29–35.

66 Snyder RE, Marlow MA, Riley LW. Ebola in urban slums: the elephant in the room. Lancet Glob Health 2014;2:e685.

67 Wesolowski A, Buckee CO, Bengtsson L et al. Commentary: containing the Ebola outbreak - the potential and challenge of mobile network data. PLoS Curr 2014;6:pii: ecurrents.outbreaks.0177e7fcf52217b8b 634376e2f3efc5e.

Downloaded from http://trstmh.oxfordjournals.org/ by guest on July 26, 2016

News

Enhancing infectious disease mapping with open access resources

D M Pigott[1], M U Kraemer (moritz.kraemer@zoo.ox.ac.uk)[1]
1. Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, United Kingdom

Citation style for this article:
Pigott DM, Kraemer MU. Enhancing infectious disease mapping with open access resources . Euro Surveill. 2014;19(49):pii=20989. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20989

Article published on 11 December 2014

The Spatial Ecology and Epidemiology Group (SEEG) at the University of Oxford is currently carrying out several disease and vector mapping projects. The source code for many of these projects have been and will continue to be made openly available through our GitHub account. This archive includes a description of the R software package 'seegSDM', that allows for species distribution modelling using a boosted regression tree approach and a walkthrough tutorial. This method has been successfully applied to diseases such as dengue, ebola virus, leishmaniasis, and anopheline vectors.

The group has collated a number of globally comprehensive and up-to-date databases from published and unpublished sources. The databases compile geographically located records of occurrence, from three sources: (i) comprehensive PubMed searches, (ii) information from unpublished health surveys and entomological field studies made available by collaborators, and (iii) internet disease surveillance systems such as HealthMap. These occurrences, defined as unique spatio-temporal records of pathogen transmission or vector presence, serve as standardised input data used to generate global risk maps. Prior to modelling these databases all go through a number of technical validation steps to ensure their spatio-temporal accuracy and overall validity. This is facilitated by overlaying these occurrence records with independently derived evidence consensus maps that summarise the quality and diversity of disease information.

Disease and vector specific databases are then made openly available through online depositories which cover the diseases mentioned above. Similarly, occurrence records of the dominant vector species of malaria and more comprehensive prevalence datasets of the global distribution of *Plasmodium falciparum, Plasmodium vivax* [1] and related inherited blood disorders are downloadable through the Malaria Atlas Project. Details about collection methods, design and validation of the respective datasets can be found either in published research articles or within data descriptor publications [2–4].

Future efforts include developing the Atlas of Baseline Risk Assessment for Infectious Diseases (ABRAID), an automated mapping platform which integrates the framework described above to generate spatially comprehensive, iteratively improving, evidence based maps of disease risk at the global level for a prioritised number of infectious diseases [5].

1. Moyes CL, Temperley WH, Henry AJ, Burgert CR, Hay SI. Providing open access data online to advance malaria research and control. Malar J. Malaria Journal; 2013 Jan;12(1):161. http://dx.doi.org/10.1186/1475-2875-12-161
2. Mylne A, Brady OJ, Huang Z, Pigott DM, Golding N, Kraemer MUG, et al. A comprehensive database of the geographic spread of past human Ebola outbreaks. Sci Data. 2014 Oct 23;1:140042. http://dx.doi.org/10.1038/sdata.2014.42
3. Messina JP, Brady OJ, Pigott DM, Brownstein JS, Hoen AG, Hay SI. A global compendium of human dengue virus occurrence. Sci Data. 2014 May 27;1:1–6. http://dx.doi.org/10.1038/sdata.2014.4
4. Pigott DM, Golding N, Messina JP, Battle KE, Duda K a, Balard Y, et al. Global database of leishmaniasis occurrence locations, 1960–2012. Sci Data. 2014 Sep 30;1:140036. http://dx.doi.org/10.1038/sdata.2014.36
5. Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S, et al. Global mapping of infectious disease. Philos Trans R Soc Lond B Biol Sci. 2013;368(20120250).