Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 646–655 doi: 10.1111/2041-210X.12549

SPECIAL FEATURE: 5TH ANNIVERSARY OF *METHODS IN ECOLOGY AND EVOLUTION*

Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever

David W. Redding[1]*, Lina M. Moses[2], Andrew A. Cunningham[3], James Wood[4] and Kate E. Jones[1,3]*

[1]*Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK;* [2]*Department of Microbiology and Immunology, Tulane University, New Orleans, Louisiana, USA;* [3]*Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, UK; and* [4]*Department of Veterinary Medicine, Disease Dynamics Unit, University of Cambridge, Cambridge, UK*

Summary

1. Human infectious diseases are a significant threat to global human health and economies (e.g. Ebola, SARs), with the majority of infectious diseases having an animal source (zoonotic). Despite their importance, the lack of a quantitative predictive framework hampers our understanding of how spillovers of zoonotic infectious diseases into the human population will be impacted by global environmental stressors.

2. Here, we create an environmental-mechanistic model for understanding the impact of global change on the probability of zoonotic disease reservoir host–human spillover events. As a case study, we focus on Lassa fever virus (LAS). We first quantify the spatial determinants of LAS outbreaks, including the phylogeographic distribution of its reservoir host Natal multimammate rat (*Mastomys natalensis*) (LAS host). Secondly, we use these determinants to inform our environmental-mechanistic model to estimate present-day LAS spillover events and the predicted impact of climate change, human population growth and land use by 2070.

3. We find phylogeographic evidence to suggest that LAS is confined to only one clade of LAS host (Western clade *Mastomys natalensis*) and that the probability of its occurrence was a major determinant of the spatial variation in LAS historical outbreaks (69·8%), along with human population density (20·4%). Our estimates for present-day LAS spillover events from our environmental-mechanistic model were consistent with observed patterns, and we predict an increase in events per year by 2070 from 195 125 to 406 725 within the LAS endemic western African region. Of the component drivers, climate change and human population growth are predicted to have the largest effects by increasing landscape suitability for the host and human–host contact rates, while land-use change has only a weak impact on the number of future events.

4. LAS spillover events did not respond uniformly to global environmental stressors, and we suggest that understanding the impact of global change on zoonotic infectious disease emergence requires an understanding of how reservoir host species respond to environmental change. Our environmental-mechanistic modelling methodology provides a novel generalizable framework to understand the impact of global change on the spillover of zoonotic diseases.

Key-words: climate change, haemorrhagic disease, infectious disease, land-use change, *Mastomys natalensis*, spillover events, West Africa

Introduction

There is growing interest in how human health and well-being will be impacted by future global change (Millennium Ecosystem Assessment 2005). Climate change, human population growth or degraded ecosystems with a reduced ability to produce goods and services (e.g. air and water quality, pollination) may severely negatively impact human health (Carpenter *et al.* 2009; IPCC 2014). One area of particular recent interest is in understanding how global change may impact the emergence and spread of human infectious diseases (e.g. Ebola, SARs, rabies) (Keesing *et al.* 2010). Infectious diseases are significant threats to global human health and economies (Grace *et al.* 2012). The majority of human infectious diseases are zoonotic, that is they have a wild or domestic animal origin or reservoir (Taylor, Latham & Woolhouse 2001). It might be expected then that anthropogenic changes impacting the

*Correspondence authors. E-mails: dwredding@gmail.com and kate.e.jones@ucl.ac.uk

This article was first published online on 13 June 2016 under a standard license. A CC-BY license has now been signed, and the copyright line and license statement was therefore updated on 13 July 2016.

distribution and abundance of these hosts and vectors may also affect the probability of emergence and spread of human zoonotic infectious diseases (Keesing *et al*. 2010).

Supporting evidence for how the loss of biological diversity generally is linked to the emergence and spread of infectious disease is contradictory and controversial (Keesing *et al*. 2010; Randolph & Dobson 2012; Civitello *et al*. 2015). Global spatial analyses of the initial emergence of zoonotic infectious disease have linked high risk (controlling for reporting effort) to areas of both high biodiversity and human densities (Jones *et al*. 2008). However, other evidence suggests that the spread of infectious disease (once emerged) is suppressed by areas of high biodiversity (the dilution effect) (Keesing *et al*. 2010; Civitello *et al*. 2015), but this remains controversial (Randolph & Dobson 2012). It is unlikely that zoonotic diseases, as a whole, will respond in a consistent way to biodiversity loss or environmental stressors. We need to first understand how zoonotic disease is mediated by the impacts of environmental change on reservoir host species. As species differ in their responses to environmental change, therefore the impact on zoonotic disease might depend on reservoir host species identity and their life-history characteristics.

One of the ways commonly used to predict impacts of global change on biodiversity is to model species' occurrence data with associated habitats and environments to develop a probabilistic distribution or niche model (Phillips, Dudik & Schapire 2004). Initial habitats and environments can then be altered in line with global predictions of change, to understand how species distributions and populations might be impacted in future (Pearson *et al*. 2013). On the other hand, to understand disease systems, epidemiological mechanistic or process-based models are often used which allow for conceptually straightforward predictions of disease transmission when the underlying conditions of the models are changed. If the impact of future environmental stressors on reservoir host populations could be modelled and added into disease mechanistic models, then this would be a powerful framework to predict future spillover and subsequent disease outcomes in humans. Solely using complex correlative approaches to estimate zoonotic disease risk, while useful for the analyses of present-day patterns (Mylne *et al*. 2015), is not ideal for making future predictions, as it is unclear whether the inferred underlying processes (and interactions between these inferred processes) will remain stable in future environments (Alexander *et al*. 2012). Alternatively, fully mechanistic host–human epidemiological models would be both data and computationally expensive and would be difficult to construct for all but the best known diseases.

Here, we develop an environmental-mechanistic framework to model the impacts of global change on zoonotic spillover using a case study of a western African haemorrhagic zoonotic disease, Lassa fever virus (LAS), and its rodent reservoir host species, the Natal multimammate rat (*Mastomys natalensis*). LAS is the only Old World arenavirus (OWA) known to be a major cause of human mortality. At present, it is thought to infect between 100 000 and a million people per year in western sub-Saharan Africa (McCormick *et al*. 1987; Richmond & Baglole 2003). LAS and its sister taxa Gbagroube are thought to have speciated approximately 1800 years ago, having split from Mopeia virus, which is a non-pathogenic virus also restricted to the rodent reservoir host species *M. natalensis* (Coulibaly-N'Golo *et al*. 2011). *M. natalensis* is a common household and agricultural rodent pest found in almost every country in sub-Saharan Africa (Fichet-Calvet & Rogers 2009). Despite the pan-continental distribution of its reservoir host, fever outbreaks associated with LAS are reported in only a small number of countries in western Africa: Sierra Leone, Guinea, Liberia, Nigeria and southern Mali (Safronetz *et al*. 2010). Within western Africa, human antibody prevalence studies have also demonstrated the circulation of LAS or a cross-reactive arenavirus in Senegal, Cote d'Ivoire, Ghana, Benin and Burkina Faso (Saluzzo *et al*. 1988; Emmerich, Gunther & Schmitz 2008).

Available LAS case data are limited: of an estimated 200 000 cases per year, only approximately 400 have been accurately recorded (CDC 2012). Additionally, these data are likely to contain reporting biases; for example, many cases have been reported near the diagnostic laboratory at Kenema Hospital in Sierra Leone (Fichet-Calvet & Rogers 2009). Known LAS case data therefore represent a limited data set to interpolate the true spatial spread of reservoir host–human spillover events across West Africa. However, we can use this data set to determine the most important drivers of known LAS disease events by creating a correlative LAS spatial distribution or niche model. This then can inform our environmental-mechanistic spillover model, to predict the current and, importantly, the likely future LAS spillover events across the region. Reservoir host–human spillover for LAS is not fully understood, but is believed to be mediated by bushmeat hunting, inhaling rodent faeces and eating faeces-contaminated food (CDC 2012). Although interactions between these factors at a fine scale are likely to be complex and variable (Fichet-Calvet & Rogers 2009), at the macro and temporal scale we focus on here, our environmental-mechanistic modelling results suggest that we can reasonably approximate these processes for LAS. Our model is constructed such that it can take advantage of forecasts of climate change, alongside forecasts of human future population density and land-use change, and be effectively employed to provide a simulation of LAS spillover events in these future environments.

Methods

SPATIAL DETERMINANTS OF LAS OUTBREAK RISK

We collated and georeferenced human LAS cases from 1967 to 2012 from the literature ($n = 408$) (Table S1, Supporting information) to examine the most important spatial determinants of historical spillovers. We collated spatial environmental and habitat variables for continental Africa (Table S2, Supporting information) and used those spatial variables linked to LAS outbreaks, or those thought to explain higher incidences of disease more generally, to create a LAS distribution or niche model using boosted-regression trees (BRT) (Elith *et al*. 2006). For instance, the reservoir host for Lassa fever virus (LAS host) *M. natalensis* is known to be an agricultural pest (Fichet-Calvet & Rogers 2009), and we included yield estimates for all possible crop types in an attempt to define more specifically what agricultural prac-

tices are linked to LAS outbreaks. Similarly, by including human ethnic groupings, we attempted to control for some gross behavioural differences between populations in that region. Distances to mining locations were used due to reports that LAS outbreaks had occurred near mines (Kaslow, Stanberry & Le Duc 2014). Temperature and rainfall were included as these were identified as among the most important explanatory factors in previous work (Fichet-Calvet & Rogers 2009). Poverty was also included as this may correspond to the lack of healthcare provision and may be correlated with outbreaks, giving a data set consisting of (i) Human population density (2010) from the Gridded Population of the World v.3 (CIESIN 2005); (ii) Crop Yields from HarvestChoice (HarvestChoice 2014); (iii) Mining Locations from United States Geological Survey (United States Geological Survey 2010); (iv) Ethnic groups from GREG (Weidmann, Rød & Cederman 2010); and (v) Gross Domestic Product from Global Poverty Map (Elvidge *et al.* 2009). We also estimated (vi) LAS host *M. natalensis* probability of occurrence $p(H)$, through a species distribution or niche models (see below). We also investigated the impact of environmental variables independent of the influence of host distribution by creating two further variables: (vii) BIO12 Annual Precipitation residuals and (viii) BIO1 Annual Mean Temperature residuals. Residuals were created in each case from a linear model with LAS host occurrence probability, $p(H)$. We created an average probability surface of LAS outbreak risk from 25 replicates of boosted-regression trees (BRT) models (using sets of 500 randomly created absence points across western Africa). In all cases, the models were trained and tested using different random subsets of the location data, and the ability of models to successfully predict a testing data set was assessed using the area under operating curve (AUC) and true-skill statistic (TSS). For both, a value of 1 represents a perfect ability to predict the testing subset when using models created with the training subsets, and values close to 0 represents essentially random niche estimation with respect to the testing data set. The highest scoring model was used to create a spatial prediction across LAS endemic region. All analyses were conducted in R v.3.0.2 (R Development Core Team 2009).

LAS HOST (*MASTOMYS NATALENSIS*) OCCURRENCE PROBABILITY $P(H)$

Phylogeographic limits of Mastomys natalensis

Recent mitochondrial genetic evidence (cytochrome *b*) suggests that *Mastomys natalensis* may not have a continuous pan-continental population and contain geographically distinct phylogenetic clades (Colangelo *et al.* 2013). This has important implications on the potential distribution of Lassa fever virus (LAS), because, for example, if there are phylogenetically and geographically distinct reservoir host populations, LAS might not be present in all of them. To investigate the phylogeographic limits of *M. natalensis*, we collated genetic sequence data from across the extent of its range. We used 3 mitochondrial genes (16S ribosomal RNA, cytochrome *b* and cytochrome oxidase subunit 1) for 373 sequences of *M. natalensis* and, as an outgroup, 3 sequences from its sister species *M. erythroleucus*, downloaded from GenBank (Benson *et al.* 2005) accessed on 01/03/2013 (Table S3, Supporting information). Only genetic sequences with locations were used. Place names were extracted using BioEdit (v.1.7.10) (Hall 1999) and assigned to coordinates using an online service (Zwiefelhofer 2013). In addition, we also used material collected from 71 individuals from 6 villages from Eastern Sierra Leone in January–February and July–August 2009. DNA was purified from liver stored in 95% ethanol using a 'DNeasy Blood & Tissue kit' (Qiagen Inc., Valencia, CA, USA). Cytochrome *b* was amplified using primers L14723 and 15915, and samples were sequenced by a service provider using Sanger dideoxynucleotide sequencing (Table S4, Supporting information).

Sequences were aligned using MAFFT (version 7) (Katoh, Kuma & Hiroyuki 2005) and then edited by eye in BioEdit (Hall 1999). Sequences that could not be aligned were not used in further analysis. Gene topologies for each gene were inferred using MrBayes (version 3.2) (Ronquist & Huelsenbeck 2005), on each occasion using a strict rather than relaxed molecular clock as it produced trees with higher likelihoods. For the cytochrome *b* gene tree, branch lengths were calculated in years using per-year nucleotide substitution rates of 0·0078 nucleotides per year calculated from closely related rodent species (*Mus indutus*, *Rattus norvegicus*, *Apodemus sylvaticus*, *Apodemus mystacinus*, *Gerbillus nigeriae*, *Gerbillus nanus*) (Nabholz, Glemin & Galtier 2008). Given the variation in substitution rates in the reported literature, we used the mean, minimum and maximum values to calculate branch lengths to test the sensitivity of our estimates.

The resulting separate phylogenetic analyses of the three genes (Fig. 1a–c) suggest that *M. natalensis* can be split into three well-supported monophyletic clades (>99% posterior probability). The location of these sequences suggests that the clades are also distinct geographically and correspond to 'Western', 'Central' and 'Southern' distributions (Fig. 1d). There is some evidence for a fourth clade or a hybrid zone between the Central and Southern *M. natalensis* clades (Fig. 1c), as a few sequences from Mozambique, Tanzania and Kenya did not fit well into the three clades. The Western clade (from the eastern Nigerian border westwards) consists of, and entirely contains, individuals from areas known to have LAS outbreaks. We estimate that the Western clade *M. natalensis* split off from the other clades 250 000 years ago (525 000–60 000 maximum and minimum values, respectively) (Fig. 1c). Despite the large range of dates estimated using our data, this split is at least twenty times older than the estimated date when LAS split from the Lassa-Mopeia ancestral strain (Coulibaly-N'Golo *et al.* 2011). It is therefore reasonable to suggest that LAS evolved to become the pathogenic virus seen today, within the Western clade of *M. natalensis*. Consequently, we base our following analyses of the LAS reservoir's host's environmental and habitat requirements on this Western clade.

Environmental and habitat requirements of Mastomys natalensis

To understand how the spatial distribution of the reservoir host species is limited by climatic factors and habitat coverage, we collated 88 occurrence records of Western clade *M. natalensis* from our collected material and the Global Biodiversity Information Facility data base (Global Biodiversity Information Facility 2013). Western clade *M. natalensis* records were defined as those from Nigeria's eastern border westwards. Records were cleaned manually by removing erroneous geolocations and points of low spatial resolution which could not be confidently spatially assigned. From our collated geospatial environmental and habitat variables for continental Africa (Table S2, Supporting information), we used the following bioclimatic variables from Hijmans *et al.* (2005) for analysis: (i) BioClim BIO1 Annual Mean Temperature; (ii) BioClim BIO5 Maximum Temperature of Warmest Month; (iii) BioClim BIO6 Minimum Temperature of Coldest Month; (iv) BioClim BIO7 Temperature Annual Range; (v) BioClim BIO10 Mean Temperature of Warmest Quarter; (vi) BIO11 BioClim Mean Temperature of Coldest Quarter; (vii) BioClim BIO12 Annual Precipitation; (viii) BioClim BIO13 Precipitation of Wettest Month;

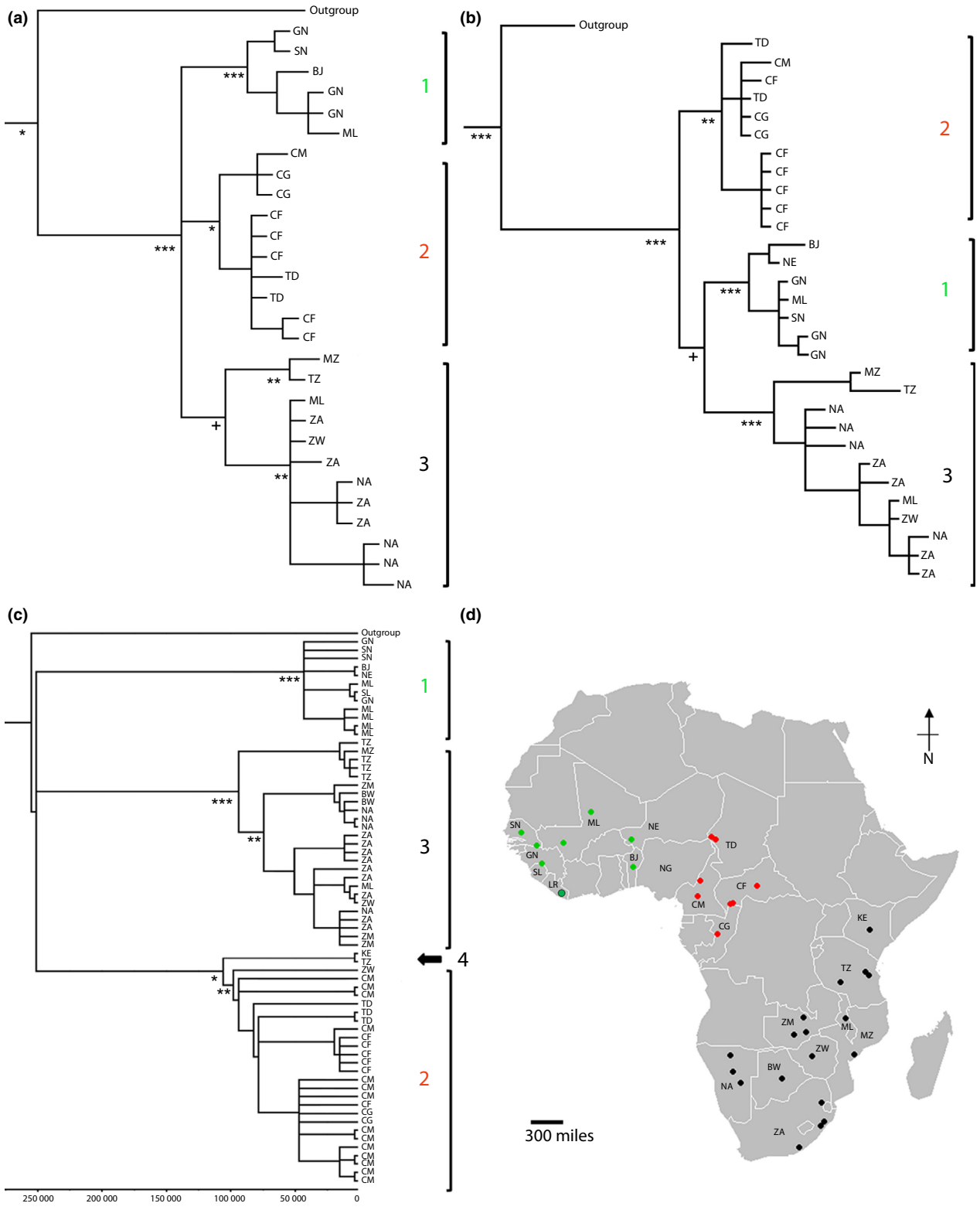

Fig. 1. Bayesian phylogenetic analysis of three mitochondrial genes of *Mastomys natalensis* using the outgroup *M. erythroleucus* for (a) 16S ribosomal RNA, (b) cytochrome oxidase subunit 1 and (c) cytochrome *b*. Cytochrome *b* sequence locations are shown in (d). Branch lengths represent expected substitutions per site in (a) and (b) and are proportional to time (years) in (c). Tip labels represent country of sequence origin (see Table S3, Supporting information for country codes). *** represents 100% posterior probability; ** >95%; * >75%; and + >50% posterior probability. 1 or green filled circles represents the 'Western Clade'; 2 or red filled circles, 'Central Clade'; and 3 or black filled circles, 'Southern Clade'; and 4 or black filled circles, hybrid zone between 2 and 3.

(ix) BioClim BIO14 Precipitation of Driest Month. The bioclimatic variables chosen were the nine most orthogonal (<75% correlation). We also used additional variables of (x) Distance to Water Bodies from Global Lakes and Wetlands data base (Lehner & Doll 2004); (xi) Elevation from Digital Elevation Model (Jarvis *et al.* 2008); (xii) Soil Type from Harmonized World Soil Database (Fischer *et al.* 2008); and (xiii) Land Use from Harmonized Land Cover (Chini, Hurtt & Frolking 2014) (percentage cultivated, urban, pastoral, primary and secondary). For analysis, all variables were reduced in latitudinal extent to 85°N 58°S and resampled to a 0·0416 degree grid cell size using a World Geodetic System 84 projection using 'raster' (Hijmans & van Etten 2012).

We used boosted-regression trees (BRT) models (Elith *et al.* 2006) to examine the relationship between occurrence records of Western clade *M. natalensis* and our ecological and environmental data to create niche models. We created an average probability surface from ten replicates of BRT models (using sets of 500 randomly created absence points) to represent LAS host occurrence probability $p(H)$. In all cases, niche models were trained and tested using different random subsets of the location data, and the ability of niche models to successfully predict a testing data set was assessed using AUC and TSS (Elith *et al.* 2006). All analyses were conducted in R v.3.0.2 (R Development Core Team 2009). The BRT model for LAS host occurrence was strongly supported as AUC and TSS values were high (AUC 0·96, TSS 0·29), suggesting that the training data were able to predict the test data effectively, with the lower TSS score due to using pseudo-absences reducing model specificity. Western clade *M. natalensis* were found to be associated most strongly with areas of high annual precipitation (28·3% of the variance), with a higher proportion of cultivated land (25·4%) and with areas that had the highest precipitation in the wettest month (17·4%) (Figs S1, S2, Supporting information).

LAS ENVIRONMENTAL-MECHANISTIC ZOONOTIC SPILLOVER MODEL

We constructed an environmental-mechanistic model of LAS spillover by quantifying the contact rate between LAS host and humans in 0·0416° grid cells across the West African LAS endemic region (18°E to 16°W and 1°N to 18·5°S). However, instead of just assuming that contact rates could be quantified by homogenous, random mixing of infectious reservoir hosts and susceptible human individuals, we chose to further incorporate a spatially variable weighting factor to describe the increased contact rates in areas of the world where there is less non-mechanized transport (thereby increasing on-the-ground movement and likelihood of direct and indirect contact). Therefore, in each 0·0416° grid cell j, we calculated the force of infection for zoonotic host-to-human transmission λ_z, as follows:

$$\lambda z_j = \frac{c_j}{\bar{c}} \kappa S_j \qquad \text{eqn 1}$$

where c_j = the number of human–reservoir host contacts in a grid cell j per day, $\bar{c}$ = mean number of human–reservoir host contacts across all grid cells per day, κ = spillover risk and S_j = the number of susceptible people per grid cell j.

To estimate the spatially variable weighted contact rate, we first approximated per grid cell reservoir host–human contact rate c_j, using a two-dimensional, two-particle ideal free gas model (Hutchinson & Waser 2007). While an ideal free gas model may overestimate the real contact rate compared to, for instance, correlated movement patterns, it is also likely to underestimate contact rate due to human-seeking behaviour of *M. natalensis* (Fichet-Calvet & Rogers 2009) and therefore was selected as a compromise. We calculated contact rates per grid cell j, as follows:

$$c_j = p_1 p_2 q a \sqrt{(d_1)^2 + (d_2)^2} \qquad \text{eqn 2}$$

where p_1 = density of people calculated for 2010 from the Gridded Population of the World v.3 (CIESIN 2005), p_2 = density of LAS reservoir hosts (hereafter called LAS hosts), q = interaction sphere radius of radius (0·5 m), a = area of each grid cell (5·6 km^2), d_1 = daily non-mechanized movement distance (km) of people calculated from an empirical positive relationship between daily movement distances and per person, per country gross national income (GNI) (Table S5, Supporting information; based on Purchasing Power Parity World Bank 2014) (this relationship was applied to generate a value for each grid cell, with grid cells in each country having the same value), and d_2 = daily distance travelled of LAS hosts derived from mean of daily distance values from closely related species to *M. natalensis* with the same body size (0·6 km) (Carbone *et al.* 2005).

The density of LAS hosts p_2, per grid cell, was calculated as follows:

$$p_2 = p(H) \max_n z \qquad \text{eqn 3}$$

where $p(H)$ = probability of LAS host occurrence informed from our previous analysis above, z = densities of the host with n samples calculated from the literature (Table S6, Supporting information), using the maximum value so that in optimum conditions ($p(H) = 1$) the grid cell is given the maximum density, and for less than optimum conditions the grid cells contain proportionally fewer.

Spillover risk κ was calculated as follows:

$$\kappa = \frac{O}{\sum_n^1 c_j S_j} \qquad \text{eqn 4}$$

where O = estimated cases per year from (CDC 2012) (200 000), c_j = human–reservoir host contact rate per grid cell, S_j = the number of susceptible people inferred from human population estimates from 2014 (based on Purchasing Power Parity World Bank 2014), and n = the total number of grid cells in the area of interest.

As spillover is a binary state (either a person is infected or not), we can approximate the probability of i spillover events for each grid cell j, using the following binomial model:

$$p(k_i \text{ infections within } S) = \binom{S}{k_i} \lambda z_j^{k_i} \left(1 - \lambda z_j\right)^{\lambda z - k_i} \qquad \text{eqn 5}$$

where k_i = the number of individuals infected, S = the number of susceptible individuals and λz = force of zoonotic infection. Using the probability distribution function derived from these distributions (dbinom R Development Core Team 2009), we calculated the number of spillover events per year, alongside the expected variation in cases, for the whole landscape given present-day conditions. By altering the model inputs in LAS host occurrence probability $p(H)$, human population densities to those predicted in 2070, we then calculated how the total numbers and spatial patterns in spillover events were likely to change in future. Last, we reran the present-day model without the spatially variable weighting per grid cell for contact rate, that is with force of infection being weighted by host numbers only:

$$\lambda z_j = \frac{p_{2j}}{\bar{p}_2} \kappa S_j \qquad \text{eqn 6}$$

where p_{2j} = density of LAS hosts, κ = spillover risk and S_j = the number of susceptible people per grid cell j, with κ in this case being derived by substituting host numbers p_2 for contact rate c_j in eqn 4.

We chose to model host–human spillover specifically, rather than include human-to-human transmission as it is unclear how much impact this transmission pathway has across the endemic landscape, though it clearly plays a role in the few recorded severe, often nosocomial, outbreaks when contaminated medical equipment and fluids could be playing a significant role (Lo Iacono *et al.* 2015). The final absolute case number would likely be higher than our predictions if human-to-human transmission plays a major role in day-to-day infections.

INCORPORATING GLOBAL CHANGE

Future LAS host occurrence

To infer the probability of future host occurrence $p(H)_{2070}$ for each grid cell, we first substituted the values of the BioClim variables in the present-day BRT model with those predicted from the Hadley atmospheric ocean-coupled climate model (HADGem3) for 2070, under the three carbon emission scenarios: low (2·5 ppm), medium (6·5 ppm) and severe (8·5 ppm). Secondly, we estimated future land cover–land use (LULC) change by using the changes seen per grid cell from 2000 to 2012 to estimate the probability of the type of LULC change in 2070 using the data from MODIS (Friedl *et al.* 2010). We calculated a probability for each grid cell based on the pattern of LULC transitions seen since 2000–2012 within a grid of 16 surrounding cells, where 5 scenarios were created producing landscapes with 0·1, 0·25, 0·5, 0·75 and 1 of the background rate of LULC change seen previously. We updated the LULC of each grid cell yearly from 2012 until 2070 for the 5 scenarios and ran each scenario 100 times to create a bank of future possible landscapes (Table S7, Supporting information). Each scenario was summarized by choosing the modal class per grid cell. We then substituted these predicted LULC values into the present-day BRT models and, along with the climatic variables, used these predicted variables to recalculate the probability of LAS host occurrence for 2070. We undertook this simulation as previous work to date (e.g. Chini, Hurtt & Frolking 2014) has not differentiated habitat cover beyond a few broad categories. The future predictions, when visually inspected, produced reasonable future extensions of previously observed patterns of land-use change.

Future human populations

Human population estimates per grid cell for 1995, 2000 and 2005 from Gridded Population of the World v3 (CIESIN 2005) were used to establish a per-year, per grid cell slope of population change using a linear model. This model was then used to forecast human populations per cell in 2070. To estimate future human on-the-ground mobility, we predicted future GNI for each country based on current trends (World Bank 2014) and awarded grids cells within each country their corresponding new estimate of movement (Table S5, Supporting information).

We estimated present-day LAS spillover per grid cell in the endemic regions and then made a prediction for the year 2070, using the model and data sets outlined above. To examine the impact of each of the drivers of change in our model (i.e. climate change, human population density, land cover change and human mobility), we ran a series of exploratory scenarios where all drivers were held at present-day levels except one which we substituted for the future values.

Results

SPATIAL DETERMINANTS OF LAS OUTBREAK RISK

The BRT niche model for LAS outbreak risk was strongly supported (AUC 0·98, TSS 0·51) and LAS outbreaks were most strongly correlated with the spatial occurrence of LAS host (explaining 69·8% of the model variance), followed by human population density (20·4%), annual precipitation residuals (4·4%) and rice yields (1·9%). Other variables (e.g. mine sites and ethnic groups) showed no ability to explain disease variation. The spatial risk of LAS outbreaks across the endemic region in western Africa (Fig. 2) showed a high correlation with areas most suitable for the host (*M. natalensis*) (Fig. S2, Supporting information). In most cases, the relationship of LAS outbreak risk with the different variables was best summarized as a positive, but asymptotic one, except for LAS host occurrence probability that had an approximately linear relationship with LAS outbreak risk (Fig. S3, Supporting information).

ENVIRONMENTAL-MECHANISTIC MODEL OF PRESENT AND FUTURE LAS SPILLOVER

The spatial distribution of the number of LAS spillover events from the simulation for present day was consistent with historic outbreaks within the LAS endemic region (Fig. 3a). An evaluation using locations of known outbreaks and random background points to create a confusion matrix resulted in high/medium AUC and TSS scores of 0·71 and 0·58, respectively. The basic contact model (eqn 6) gave poorer AUC and TSS scores (0·64 and 0·26, respectively). Qualitative differences from the correlative model (Fig. 2) are influenced by biased clusters of disease cases near, for instance, diagnostic laboratories. An increase in the distributional range of LAS spillover events is predicted by 2070, especially in the northern and eastern parts of the present simulated distribution (Fig. 3b), with the biggest impact seen under medium to high climate impact

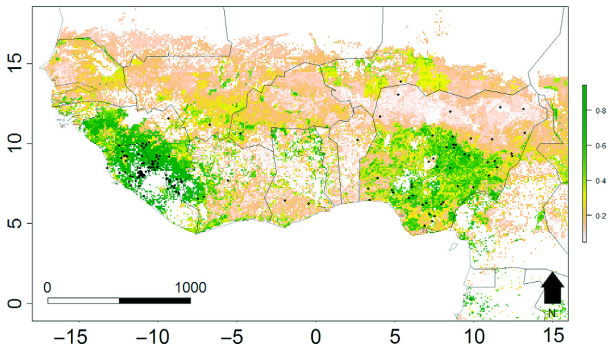

Fig. 2. Spatial distribution of LAS risk across its endemic region in West Africa estimated from outbreak records from 1967–2012 using BRT models. Risk probability per grid (0·0416°) cell is represented on a linear colour scale from 1-0, where green is most suitable and grey unsuitable. Axis labels indicate degrees, in a World Geodetic System 84 projection. Filled black circles represent locations of historic LAS outbreaks (sources, see Table S1, Supporting information).

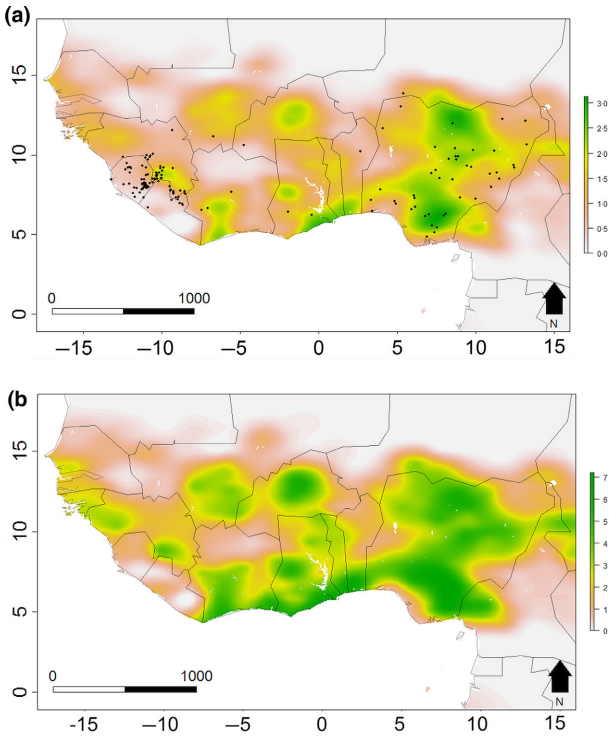

Fig. 3. Spatial distribution of simulated LAS spillover events across its endemic region in western Africa for (a) present day and (b) projected for 2070 under a medium climate and full land cover change scenario. Values represent the expected number of spillover events per grid cell per year (0·0416°) and are represented on a linear colour scale from green to grey. Axis labels indicate degrees, in a World Geodetic System 84 projection. Filled black circles represent locations of historic LAS outbreaks (sources, see Table S1, Supporting information).

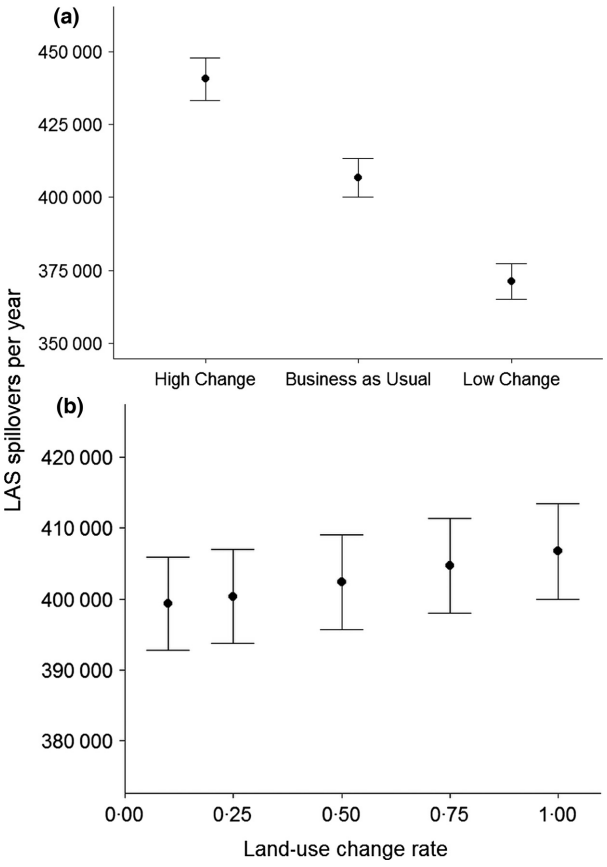

Fig. 4. Effect of different (a) climate and (b) land-use change scenarios on expected LAS spillover events in 2070. Where (a) represents three different climate scenarios ('high change', 'business as usual' and 'low change') and (b) five scenarios of land-use change, representing 0·1, 0·25, 0·5, 0·75 and 1 times the rate of land-use change seen in the region in last few decades. Error bars represent 95% confidence intervals around the mean.

scenarios (Fig. 4a) and greater land cover change scenarios (Fig. 4b). For a medium climate change and full land cover change scenario, the model predicts an average 2·1-fold increase (expected range 2·0-2·1) in the number of spillover events per year by 2070 (Fig. 5), with the largest increases seen under higher impact scenarios (i.e. full land cover change and high levels of climate change Fig. 4a, b). Of the component drivers, changes to climate and human population densities have the largest effect on increasing spillover events, with land cover change having a weaker positive impact. Changes in human mobility patterns acted independently to reduce the number of LAS spillover events (Fig. 5).

Discussion

LAS is a serious, high burden disease within West Africa that is currently still poorly understood. We show that, from our spillover environmental-mechanistic model, more areas of West Africa are at risk of reservoir host–human spillover that current case data suggest. Our model suggests that in future, it is likely to become a greater burden on local communities spreading to more areas with approximately twice as many spillover events predicted by 2070. The spillover increase in our model in future is attributable to a large extent to climate change, which appears to increase future disease burden, because the future climatic conditions in this region (hotter and wet) provide more suitable habitat for the Western clade *M. natalensis*. Regional predicted increases in human populations increase the number of susceptible individuals, thereby increasing the contact rate between reservoir hosts and humans and the number of spillover events. Landscape change has a less powerful effect on spillovers. LAS host (*M. natalensis*) is a common household and agricultural rodent pest (Fichet-Calvet & Rogers 2009), and our model of the habitat requirements of the Western clade *M. natalensis* shows that this species occurs in areas of high rainfall and cultivated land and savanna (Figs S2, S3, Supporting information). The largest landscape change predicted by 2070 is the conversion of savanna and grasslands to croplands (Table S7, Supporting information). This conversion is from one land-use type that is suitable to the reservoir host to another, providing only limited net positive effects on total spillover events. Nevertheless, this suggests that a greater anthropogenic transformation of the landscape across western Africa may have the potential to involve a trade-off between agricultural production and LAS spillover events. Current predictions of changes to national wealth in the endemic regions (assuming future mobility patterns map well to these expected changes) suggest that reduced daily

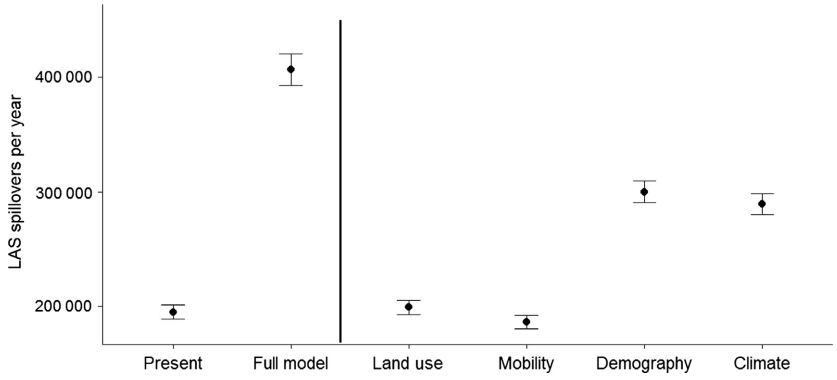

Fig. 5. Change in expected number of LAS spillover events per year from present to 2070 across its endemic region in western Africa (left-hand side of vertical line). The independent impact on LAS spillover events per month by human population, climate, land-use change and expected changes in human mobility patterns are projected for 2070 under a medium climate and full land cover change scenarios (right-hand side of vertical line). Present represents the present-day model; Future, 2070 model; Demography, 2070 model allowing only human population density to change; Climate, 2070 model allowing only climate to change; and Land use, 2070 model allowing only land use to change. Error bars represent 95% confidence intervals around the mean.

movement distance will also reduce human–host contact rates, thereby reducing spillover events. Taken together, our LAS case study suggests that zoonotic diseases are unlikely to respond uniformly to global environmental stressors.

Our environmental-mechanistic model provides a useful framework for understanding the impact of any interventions or change in national or international government policies. For example, although population growth is highly stable and unlikely to have sharp deflections away from historical trends (Lutz, Sanderson & Scherbov 2001), land cover change has perhaps a less predetermined future pattern. Our model shows that the severity of landscape conversion is proportional to the severity of the disease spillover events, suggesting that a reduction in the rate of land-use change may benefit future communities. However, to achieve a consensus position from which to petition changes to regional land-use policy, we need to move away from a single disease approach. For example, if one environmental stressor decreases one disease, but increases another, then the overall cost to community could be negative. The modelling framework we propose offers the potential to look at the impact of global change on multiple diseases at once, to understand these trade-offs. We can test, for instance, the relative impact of reducing reservoir host–human contact rates compared to mitigation measures aimed at reducing reservoir host populations. A future development of our environmental-mechanistic framework might be to model zoonotic disease transmission within human populations by including the impact of, for example, travel infrastructure, human-to-human contact rates and poverty. One disease that this approach would be relevant for is plague (*Yersinia pestis*) (Makundi, Massawe & Mulungu 2007), which is carried by *M. natalensis* alongside LAS, but differs in its transmission structure and way the disease presents in human cases. Extending our spillover model to model a disease within human populations would mean that it would be possible to determine the fate of predicted spillovers, with regard to the expected number of infections, amount of morbidity or resource cost to local communities.

Finally, the accuracy of our modelling approach is contingent on the greater understanding of both the phylogeography and ecology of the reservoir host. More information on the spatial variation of seroprevalence in the host would further add to the understanding of current disease patterns. Further work needs to be undertaken to see whether our model can be extended to other zoonotic human–reservoir host systems and used globally, as part of a forecasting system for major zoonoses. It is imperative that we begin to uncover and quantify these expected changes in response to future environmental stressors, from a multidisease and global perspective, to create effective and non-contradictory policy solutions.

Acknowledgements

We thank P. Sivasubramaniam and R. Gibb for technical assistance and C. Watts, C. Carbone, T. Lucas and R. Freeman for comments on previous versions of the manuscript. This work, Dynamic Drivers of Disease in Africa Consortium, NERC project no. NE-J001570-1, was funded with support from the Ecosystem Services for Poverty Alleviation Programme (ESPA). The ESPA is funded by the Department for International Development (DFID), the Economic and Social Research Council (ESRC) and the Natural Environment Research Council (NERC). AAC is additionally supported by a Royal Society Wolfson Research Merit Award.

Data accessibility

Land-use conversion simulations, raster layers of all future spillover scenarios and all R code are available on figshare (https://figshare.com/s/0a41d604f1b2c0bb5080).

References

Alexander, K.A., Lewis, B.L., Marathe, M., Eubank, S. & Blackburn, J.K. (2012) Modeling of wildlife-associated zoonoses: applications and caveats. *Vector Borne and Zoonotic Diseases (Larchmont, NY)*, **12**, 1005–1018.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2005) GenBank. *Nucleic Acids Research*, **33**, D34–D38.

Carbone, C., Cowlishaw, G., Isaac, N.J. & Rowcliffe, J.M. (2005) How far do animals go? Determinants of day range in mammals. *American Naturalist*, **165**, 290–297.

Carpenter, S.R., Mooney, H.A., Agard, J., Capistrano, D., DeFries, R.S., Díaz, S. *et al.* (2009) Science for managing ecosystem services: Beyond the Millen-

nium Ecosystem Assessment. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1305–1312.
CDC (2012) Lassa Fever - Fact Sheet. http://wwwcdcgov/ncidod/dvrd/spb/mnpages/dispages/Fact_Sheets/Lassa_Fever_Fact_Sheet.pdf (ed. C.f.D.C.a. Prevention).
Chini, L.P., Hurtt, G.C. & Frolking, S. (2014) *Harmonized Global Land Use for Years 1500–2100, V1. Data set*. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee.
CIESIN (2005) *Gridded Population of the World Version 3 (GPWv3): Population Grids*. SEDAC Columbia University, New York.
Civitello, D.J., Cohen, J., Fatima, H., Halstead, N.T., Liriano, J., McMahon, T.A. *et al.* (2015) Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences*, **112**, 8667–8671.
Colangelo, P., Verheyen, E., Leirs, H., Tatard, C., Denys, C. & Dobigny, G. (2013) A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society*, **108**, 901–916.
Coulibaly-N'Golo, D., Allali, B., Kousassi, S.K., Fichet-Calvet, E. & Becker-Ziaja, B. (2011) Novel Arenavirus Sequences in *Hylomyscus* sp and *Mus* (*Nannomys*) *setulosus* from Côte d'Ivoire: Implications for Evolution of Arenaviruses in Africa. *PLoS ONE*, **6**, e20893.
Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
Elvidge, C.D., Sutton, P.C., Ghosh, T., Tuttle, B.T., Baugh, K.E., Bhaduri, B. & Bright, E. (2009) A global poverty map derived from satellite data. *Computers & Geosciences*, **35**, 1652–1660.
Emmerich, P., Gunther, S. & Schmitz, H. (2008) Strain-specific response to Lassa virus in the local population in West Africa. *Journal of Clinical Virology*, **42**, 40–44.
Fichet-Calvet, E. & Rogers, D.J. (2009) Risk maps of lassa fever in West Africa. *PLoS Neglected Tropical Diseases*, **3**, e388.
Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L. & Wiberg, D. (2008) *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA and FAO, Laxenburg, Austria and Rome, Italy.
Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. & Huang, X. (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterisation of new datasets. *Remote Sensing of the Environment*, **114**, 168–182.
Global Biodiversity Information Facility (2013) Global Biodiversity Information Facility. http://www.gbif.org/.
Grace, D., Mutua, F., Ochungo, P., Kruska, R., Jones, K.E., Brierley, L. *et al.* (2012) *Mapping of poverty and likely zoonoses hotspots*. International Livestock Research Institute, Nairobi, Kenya.
Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
HarvestChoice (2014) Crop Production: SPAM. International Food Policy Research Institute, Washington, DC, and University of Minnesota, St. Paul, MN.
Hijmans, R.J. & van Etten, J. (2012) *Raster: Geographic Analysis and Modeling with Raster Data R Package*. R Foundation for Statistical Computing, Vienna, Austria.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
Hutchinson, J. & Waser, P.M. (2007) Use, misuse and extensions of "ideal gas" models of animal encounter. *Biological Reviews*, **3**, 335–359.
IPCC (2014) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects*. Cambridge University Press, Cambridge, UK and New York, NY USA.
Jarvis, A., Reuter, H.I., Nelson, A. & Guevara, E. (2008) Hole-filled SRTM for the globe. Version 4. CGIAR-CSI SRTM 90m Database.
Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.
Kaslow, R.A., Stanberry, L.R. & Le Duc, J.W. (2014) *Viral Infections of Humans: Epidemiology and Control*. Springer, New York, NY, USA.
Katoh, K., Kuma, K. & Hiroyuki, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignments. *Nucleic Acids Research*, **33**, 511–518.
Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D. *et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, **468**, 647–652.
Lehner, B. & Doll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296**, 1–22.
Lo Iacono, G., Cunningham, A.A., Fichet-Calvet, E., Garry, R.F., Grant, D.S., Khan, S.H. *et al.* (2015) Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: the case of lassa fever. *PLoS Neglected Tropical Diseases*, **9**, e3398.
Lutz, W., Sanderson, W. & Scherbov, S. (2001) The end of world population growth. *Nature*, **412**, 543–545.
Makundi, R.H., Massawe, A.W. & Mulungu, L.S. (2007) Reproduction and population dynamics of Mastomys natalensis Smith, 1834 in an agricultural landscape in the Western Usambara Mountains, Tanzania. *Integrative Zoology*, **2**, 233–238.
McCormick, J.B., Webb, P.A., Krebs, J.W., Johnson, K.M. & Smith, E.S. (1987) A prospective study of the epidemiology and ecology of Lassa fever. *Journal of Infectious Diseases*, **155**, 437–444.
Millennium Ecosystem Assessment (2005) *Ecosystems and Human Well-Being: Synthesis*. Island Press, Washington, DC.
Mylne, A.Q., Pigott, D.M., Longbottom, J., Shearer, F., Duda, K.A., Messina, J.P. *et al.* (2015) Mapping the zoonotic niche of Lassa fever in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **109**, 483–492.
Nabholz, B., Glemin, S. & Galtier, N. (2008) Strong Variations of Mitochondrial Mutation Rate across Mammals-the Longevity Hypothesis. *Molecular Biology and Evolution*, **25**, 120–130.
Pearson, R.G., Phillips, S.J., Loranty, M.M., Beck, P.S.A., Damoulas, T., Knight, S.J. & Goetz, S.J. (2013) Shifts in Arctic vegetation and associated feedbacks under climate change. *Nature Climate Change*, **3**, 673–677.
Phillips, S.J., Dudik, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*. pp. 655–662. ACM Press, New York.
R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Randolph, S.E. & Dobson, A.D. (2012) Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, **139**, 847–863.
Richmond, J.K. & Baglole, D.J. (2003) Lassa fever: epidemiology, clinical features, and social consequences. *British Medical Journal*, **327**, 1271–1275.
Ronquist, F. & Huelsenbeck, J.P. (2005) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
Safronetz, D., Lopez, J.E., Sogoba, N., Traore, S.F., Raffel, S.J., Fischer, E.R. *et al.* (2010) Detection of Lassa virus, Mali. *Emerging Infectious Diseases*, **16**, 1123–1126.
Saluzzo, J.F., Adam, F., McCormick, J.B. & Digoutte, J.P. (1988) Lassa fever virus in Senegal. *Journal of Infectious Diseases*, **157**, 605.
Taylor, L.H., Latham, S.M. & Woolhouse, M.E. (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society London, Series B: Biological Sciences*, **29**, 983–989.
United States Geological Survey (2010) Mineral operations outside the United States. (ed. G.S. (U.S.)). Reston, Virginia.
Weidmann, N.B., Rød, J.K. & Cederman, L.E. (2010) Representing ethnic groups in space: a new dataset. *Journal of Peace Research*, **47**, 1–9.
World Bank (2014) *World Development Indicators*. World Bank, Washington, DC, USA.
Zwiefelhofer, D. (2013) Batch Geocoding.

Received 19 October 2015; accepted 28 January 2016
Handling Editor: Robert Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1. Details of LAS outbreaks and sources from 1967–2012 (n = 408).

Table S2. Details of all environmental and habitat geospatial variables used for niche models and simulations.

Table S3. GenBank accession numbers (accessed 01/03/2013) and country of origin of sequences for three mitochondrial genes

(Cytochrome *b*, Cytochrome oxidase subunit 1, and 16S) of *Mastomys natalensis* (n = 373 specimens) and for the outgroup *Mastomys erythroleucus*.

Table S4. Locations of material collected from 71 individuals of *Mastomys natalensis* from six villages in eastern Sierra Leone in 2009 with collection and analysis details.

Table S5. Estimates of global daily non-mechanized movement distance, $v\Delta t$.

Table S6. Density estimates of *Mastomys natalensis* across Africa in a range of habitats.

Table S7. Comparison of present and estimated change in land cover-land use (LULC) in 2070 using MODIS data in the LAS endemic region in Western Africa measured by grid cell counts (0·0416°).

Fig. S1. Response curves from boosted-regression trees model of Western clade *Mastomys natalensis* spatial occurrences from GBIF.

Fig. S2. Map of Western Clade *Mastomys natalensis* occurrence probability $p(H)$ estimated from boosted-regression trees models using records from our collected records and GBIF.

Fig. S3. Response curves from boosted-regression trees model of historic LAS outbreaks from 1967–2012.