# Title: Impact of global change on future Ebola emergence and epidemic potential in Africa

**Authors:** D. W. Redding[1*], P. M. Atkinson[2], A. A. Cunningham[3], G. Lo Iacono[4,5], L. M. Moses[6], J. Wood[4], and K. E. Jones[1,3]*

**Affiliations:**

[1] Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom.

[2]Geography and Environment, University of Southampton, Southampton SO17 1BJ, United Kingdom.

[3]Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom.

[4]Department of Veterinary Medicine, Disease Dynamics Unit, University of Cambridge, Cambridge, United Kingdom.

[5]Environmental Change Department, Public Health England, Didcot, United Kingdom.

[6]Department of Global Community Health and Behavioral Sciences, Tulane University, New Orleans, Louisiana, United States of America.

*__Correspondence to__: D. Redding (d.redding@ucl.ac.uk); K. E. Jones (kate.e.jones@ucl.ac.uk).

**Keywords**: bats, climate, ebola virus disease, environmental-mechanistic disease model, global change, land-use.

**Abstract**: Animal-borne or zoonotic human diseases (e.g., SARS, Rabies) represent major health and economic burdens throughout the world, disproportionately impacting poor communities. In 2013-2016, an outbreak of the Ebola virus disease (EVD), a zoonotic disease spread from animal reservoirs caused by the Zaire Ebola virus (EBOV), infected approximately 30,000 people, causing considerable negative social and economic impacts in an unexpected geographical location (Sierra Leone, Guinea, and Liberia). It is not known whether the spatial distribution of this outbreak and unprecedented severity was precipitated by environmental changes and, if so, which areas might be at risk in the future. To better address the major health and economic impacts of zoonotic diseases we develop a system-dynamics approach to capture the impact of future climate, land use and human population change on Ebola (EVD). We create future risk maps for affected areas and predict between a 1.75 and 3.2-fold increase in EVD outbreaks per year by 2070. While the best case future scenarios we test saw a reduction in the likelihood of epidemics, other future scenarios with high human population growth and low rates of socioeconomic development saw a fourfold increase in the risk of epidemics occurring and almost 50% increase in the risk of catastrophic epidemics. As well as helping to target where health infrastructure might be further developed or vaccines best deployed, our modelling framework can be used to target global interventions and forecast risk for many other zoonotic diseases.

**Significance Statement**: Despite the severe health and economic impacts of outbreaks of diseases like SARS or Zika, there has been surprisingly little progress in predicting where and when human infectious disease outbreaks will occur next. By modelling the impacts of future climate, land use and human population change on one particular disease Ebola, we develop future risk maps for the affected areas and predict 1.7-3.2 times as many human Ebola outbreaks per year by 2070, and a 50% increase in the chance that these outbreaks will become epidemics. As well as helping to target where health infrastructure might be further

developed or vaccines deployed, our approach can also be used to target actions and predict risk hotspots for many other infectious diseases.

**Introduction:** Little is known about how the majority of human infectious diseases will be affected by predicted future global environmental changes (such as climate, land use, human societal and demographic change) (1-5). Importantly, two thirds of human infectious diseases are animal-borne (zoonotic) (6) and these diseases form a major, global health and economic burden, disproportionately impacting poor communities (7, 8). Many zoonotic diseases are poorly understood, and global health responses to them are chronically underfunded (9). The 2013-2016 Ebola outbreak was unprecedented in terms of size, financial cost, and geographical location (10, 11); a stark illustration of our knowledge gaps, and demonstrating that it is imperative we develop quantitative approaches to better forecast zoonotic disease risk.

Ebola virus disease (EVD) was first identified in 1976, and since then there have been approximately 23 recognized outbreaks (12), predominantly within central Africa. EVD is caused by any one of four pathogenic strains of Ebola virus: Zaire (EBOV), Sudan (SUDV), Taï Forest (TAFV), and Bundibugyo (BDBV). It presents as a non-specific febrile illness that can cause haemorrhagic fever, often with a high case fatality rate in diagnosed patients (13). Some Old World fruit bat species (Family Pteropodidae) have been suggested as reservoir hosts (14), however, while there is limited direct evidence, they are strong candidates to play a key role either as an reservoir or amplifying host (15, 16). In areas with EVD, there are frequent direct and indirect human-bat interactions, e.g., via bush meat hunting and during fruit harvesting (17), presenting numerous opportunities for bat-to-human pathogen spill-overs to occur. Additionally, a third of known zoonotic spill-overs have been connected to contact with great apes and duikers, although there is no evidence that these species act as reservoir hosts (10). It is clear, however, that once spill-over occurs human social factors

such as movement and healthcare responses greatly influence the cumulative outcome of an outbreak (18). For instance, previous work has highlighted the importance of family interactions (19), funeral practices (20) and differential transmission rates in hospitalized individuals (18).

Many attempts to understand Ebola outbreak dynamics have focused on mechanistic modelling approaches of human-to-human transmission post spill-over from animal hosts (13, 18, 19, 21-24). Mechanistic, or process-based, models are ideal for capturing epidemiological characteristics of diseases and, importantly, testing how disease outbreaks might be impacted by intervention efforts (25). One downside is that mechanistic models rarely incorporate spatially heterogeneous ecological and environmental information (26), such as the known high variance of bat abundance and pathogen sero-prevalence across widespread individuals (27). In this context, correlative, or pattern-based, models (e.g. MaxEnt, Boosted-regression trees) have been used to simultaneously capture the spatial risk of both zoonotic spill-over and subsequent human-to-human infection (12). For some spatially-explicit analyses, there have been attempts to incorporate spatial patterns of human populations (28), while other have included air transportation networks (29), but no studies that we are aware of have considered whole-system analyses for major epidemic zoonoses, such as Ebola. Like other rare or poorly-sampled diseases, Ebola suffers from limited data availability, meaning pattern-finding, correlative analytical techniques are at a disadvantage (30).

In 2014 a spill-over in Gueckedou district, Guinea of Ebola-Zaire virus led to an EVD outbreak approximately 100 times larger than any of the previous 21 known outbreaks (31). Such epidemics have a disproportionate impact on the affected societies. For example, the World Bank estimates a cost of US$2.2 billion to the three most affected countries (32) due to, amongst other drivers, widespread infrastructure breakdown, mass migration, crop

abandonment and a rise in endemic diseases due to overrun healthcare systems. Recent work has uncovered the human-to-human transmission patterns underlying this outbreak, using case (33) and genomic data (31) to demonstrate that EVD spread can be successfully predicted by a dispersal model that is weighted by both geographic distance and human population density. Attempting to understand zoonotic epidemic risk, however, using a human-only transmission model and without incorporating host ecology would inevitably lead to areas with high human density and connectivity being identified as the regions with the highest risk, despite some areas of these lacking competent hosts. Therefore, to model both the spatial variation in spill-over risk and, concurrently, the likely progression of subsequent outbreaks in human populations, we need to take a system-dynamics modelling approach (1, 34). Key non-linear feedbacks can also be captured, for example, the trade-off between increasing human populations and loss of reservoir host species through anthropogenic land-use conversion, and using this to design the optimal roll-out of vaccinations (35) and other interventions in a changing landscape.

Here, we use a disease system-dynamics approach (Fig. 1) to extend a discrete-time, stochastic epidemiological compartmental model incorporating spatial environmental variability (Environmental-Mechanistic Model or EMM, Fig. 2) to simulate present day spill-over and subsequent human-to-human transmission of the Zaire Ebola virus (EBOV) (the strain responsible for the 2013-2016 outbreak in West Africa). We model the impact of future anthropogenic changes on the occurrence and spread of this disease in 2070 (36-38) under a variety of possible integrated global change scenarios (39). We use a combination of three Representative Concentration Pathways (RCP) scenarios of increasing greenhouse gas concentrations: RCP4.5, RCP6, and RCP8.5 (40), and three possible socio-economic development scenarios (Shared Socio-economic Pathways or SSP), ordered by increasing human population density and reduced regional socio-economic cooperation: SSP1, SSP2

and SSP3. Finally, we compare the changes to spatial patterns of risk and chances of outbreaks and epidemics occurring across Africa.

**Results**: Our EMM simulation for present day EBOV-EVD risk correctly identified areas of observed outbreaks as high risk, such as Democratic Republic of Congo, Gabon and the 2013-2016 outbreak in West Africa, but also identified some areas where EVD has not been reported, such as Nigeria and Ghana (Fig. 3A). As a result, our model suggests that the at-risk area for EBOV-EVD is much larger than the areas known to have experienced disease outbreaks thus far. Our risk map also identified areas that are endemic for the other EVD strains, likely due to similar transmission pathways and reservoir host characteristics (Fig. 3A). Although the index case risk map (Fig. 3B) shows a similar spatial pattern to all cases, high risk spill-over areas are constrained to more distinct hot-spots. Importantly, the locations of index cases that resulted in epidemics were even more geographically constrained, with Ghana, Sierra Leone, Liberia, Kenya Uganda and Cameroon all having medium risk but Nigeria is the focus of the highest potential for epidemic spill-over (Fig. 3C). Comparing the mean number of spill-overs per year gave higher results for present day simulations with 2.464 spill-overs per year (95% CI 2.361-2.567) compared to the mean historical number over the last 40 years: 0.75 (95% CI 0.695-0.905). High risk of Ebola case importation using the current network of airline flights was seen in China, Russia, India, the United States as well as many high-income European countries (Fig. 4). Especially high importation risk, however, was seen in Italy and Germany.

Similar to historic data (Fig. S3), the distribution of the final size of the simulated outbreaks was multimodal with distinct peaks at very low numbers (less than 3 cases) and medium outbreaks (approximately 3-1500 cases) (Fig. S3). Through extensive simulations we were able to explore the lower probability areas of the distribution effectively and, unique to the simulation data, there is a third peak of outbreaks (here we term ‘epidemics’) with high, to

very high, numbers of cases (1500-100,000,000 cases). This threshold of assigning an outbreak with greater than 1500 cases as an epidemic also corresponds to the top 1 percentile of a log-normal distribution approximating the variation in pre-2016 observed outbreak sizes (~1538 cases per year). Of the ~2500 simulation runs for present day conditions, epidemics (>1500) occurred approximately in 5.8% of the yearly simulations, with catastrophic epidemics (>2,000,000) occurring in around 2.3% of simulations, or once every 43.5 years given current conditions. From the sensitivity testing, the key parameters that affected outbreak size were illness length and $R_0$, which positively increased case numbers (Fig. S4a), whereas the annual spill-over rate (Fig. S4b) was most impacted by the spill-over rate constant (strongly positive), shape of the poverty/spill-over curve (weakly positive), and by host movement distance (weakly negative).

Our future EMM simulations estimate an annual increase in maximum area impacted by the disease from 3.45 million $km^2$ to 3.8 million $km^2$ under the scenario by 2070, with increases in maximum area seen under all future scenarios (Fig. 5A,D,G). The maximum areas where spill-overs could occur, however, increased by just 1% under the RCP4.5 SSP1 (Fig. 5B: 2.01 million $km^2$), when compared to present day (Fig. 3B: 1.99 million $km^2$), but increased by 14.7% under the RCP8.5 SSP3 (Fig. 5H: 2.29 million $km^2$) scenario. Conversely, the total area where epidemics could start decreased under the RCP4.5 SSP1 by 47% (Fig. 5C: 0.444 million $km^2$), when compared to present day (Fig. 3C: 0.836 million $km^2$), but again increases under RCP6 SSP2 this time by 20.5%, and by 34% under the RCP8.5 SSP3 scenario (Fig. 5F,I).

The increases seen in the area affected is mirrored by greater total numbers of spill-overs experienced in future scenarios, with the greatest increase seen under the RCP8.5 SSP3 scenario at 7.92 (CI 7.62-8.19) spill-overs per year. Spill-over numbers increased with greenhouse gas concentrations (represented here by the RCP value) with a mean 0.257 spill

over a year increase between the RCP4.5 SSP2 and RCP6 SSP2 scenarios, and a mean 0.343 spill over a year increase between the RCP6 SSP3 and RCP8.5 SSP3 scenarios. Greater increases were seen, however, with SSP change, with a mean 1.297 spill over a year increase between RCP4.5 SSP1 and RCP4.5 SSP2 and a mean 1.475 spill over a year increase between RCP6 SSP2 and RCP6-SSP3. In general, the probability of the index cases resulting in small outbreaks reduced in future environments, whereas the chance of epidemics increased (Fig. 6). For instance, the proportion of epidemics per year (>1500 cases) decreased in the RCP4.5 SSP1 to 3.43% (from 5.8% in present day) but increased in all others, with RCP6 SSP3 gaining the greatest number, with epidemics in 9.5% of all simulations. The number of catastrophic epidemics (>2,000,000), generally increased with both RCP and SSP values up to 3.43% and 3.54% for the RCP6 and RCP8.5 SSP3 scenarios respectively, but again saw a decrease from the present day level (2.3 %) to 1.19% for just the single 'best case' future scenario (RCP4.5 SSP1).

**Discussion:** We show that changes in future expected disease incidence are likely to be related to the rate of global environmental change. According to our study, EVD mitigation attempts would be best placed in efforts to reduce both population growth, increase socioeconomic development and ameliorate climate change, such that global change most closely tracks the RCP4.5 SSP1 scenario. Global binding commitments to reducing climate change may act to slow the effects, but evidence (41) suggests a wholesale change is difficult. Expected decreases in poverty and a concomitant increase in healthcare resources, therefore, would appear to be the most realistic approaches to reduce the future EVD disease burden. While vaccinations may be effective, the sporadic nature of spill-over events mean it is unclear where vaccination should be targeted and whether it would be cost-effective at this time (35). More generally, increasing health care provision and poverty reduction efforts in West Africa would not only reduce the potential effects of EVD but also other diseases,

including those that have yet to emerge in earnest, such as Marburg virus disease (42), Lassa fever (43), and Nipah/Hendra virus infection (44). This, in turn, could limit disease emergence to local outbreaks, preventing nosocomial infections and acting to prevent subsequent epidemics.

Changes to SSP scenarios, which control levels of poverty and human population size in our models, had a greater impact than changing the climate and land-use change (here mediated via RCP scenario). This is not surprising as poverty reduction increases the presumed EVD-EBOV healthcare response in our simulations, and many of the countries in the endemic region are expected to have substantial reductions in poverty levels by 2070 (37). Similarly, contact rates in our simulation (both between humans and between humans and wildlife) depend linearly on human population growth, whereas climate change increases EVD-EBOV cases through more complex interactions. Species distribution models indicate that the presumed wildlife hosts prefer warm and wet conditions (Figs. S1-2), which are expected to increase in these regions according to the HADGem3-AO climate model (38) (Fig. S5). This expansion of the optimal conditions for presumed the wildlife host species effectively increases the at-risk human population by including more of the northern, eastern and southern areas of Africa (Fig. 3A). Predicted future anthropogenic land-use changes, however, reduces the optimal wildlife host habitat, thereby reducing human-wildlife interactions.

We identify Nigeria as, not only a key area for epidemics to be initiated, but also an area with potential for many small outbreaks. This might indicate that our model has not correctly balanced the impact of healthcare infrastructure on disease spread, regional behavioural barriers to infection or regional differences in contact rates between both humans and hosts. Until these additional factors are explicitly tested, the high human density and known

presence of putative wildlife hosts mean that this area should be consider at high risk of initiating epidemics.

There is a pressing need to better understand the spatial variation in other key disease transmission parameters. For instance, bush-meat hunting is an important process by which human populations come into contact with large bats resource (45) and the spatial variation in bush-meat extraction is likely a component of spill-over variation. Little is known, however, about bush-meat hunting outside a few specific studies but there is potential to use spatial interpolation techniques to make reasonable predictions in un-sampled areas. Our model does not incorporate this data or test its impact and, similarly due to lack of data resources, we do not use information about local differences in funeral practices. Hospital compartments are thought to be useful to understand quarantine and super-spreading events but there is very limited data on the quality and geographic reach of small health clinics. Some other important behavioural trends are not captured in our model, such as the post-outbreak behavioural reactions of human populations e.g. mass migration away from affected regions. Recent findings regarding the persistence of Ebola virus in semen of convalescent men may also help explain the intermittent spatiotemporal patterns of infections in endemic areas (46, 47). Future work incorporating such data, may further improve the spatial resolution and accuracy of risk estimates.

Our approach demonstrates not only an important framework understanding Ebola but also for other diseases. Analysing diseases singly cannot be an effective approach for policy making at a large geopolitical scale, particularly in regions with multi-disease burden and limited healthcare resources. Net disease risk patterns, when summed across a wide variety of zoonoses, will be an emergent property of the distribution of very different wildlife host species and their respective responses to increasing anthropogenic land-use conversion and climate change. Any lack of data in the short-term does not reduce the obvious importance of

246 understanding future disease trends. Attempts, such as ours, establish a first heuristic step on
247 a pathway to building intervention measures aimed at reducing overall future disease burden.

248 **Materials and Methods:**

249 Environmental Mechanistic Model (EMM) EBOV

250 Using our discrete-time, stochastic epidemiological compartmental model incorporating
251 spatial environmental variability (30), we extended the approach to not just simulate
252 pathogen spill-over but also subsequent human-to-human transmission, focusing on the Zaire
253 Ebola virus (EBOV) (Fig.2). Within grid cells (0.0416°) covering continental Africa, we used
254 a Susceptible, Exposed, Infectious, Funeral and Removed (SEIFR) EVD-EBOV disease
255 compartmental model (following 13, 19, 23) to estimate the number of individuals per
256 compartment, in each time step *t*, for present day bioclimatic, land use and demographic
257 conditions. Although some previous compartmental models for EBOV have included a
258 Hospital compartment (48), adding this complexity was not feasible over large and poorly
259 known geographical areas. Without knowing more about the spatial variation in health
260 seeking behaviour, exactly which grid cells contain clinics, and the variation of healthcare
261 resources in these clinics, adding in this compartment would not likely significantly improve
262 our model's ability to predict the progression of outbreaks. Furthermore, hospital
263 interventions had the least impact controlling EVD outbreaks in a recent meta-analysis (24).
264 All analyses were carried out in R v.3.2.2 (49). Each stage of the EMM simulation is
265 discussed in more detail below:

266 Stage 1: SEIFR compartmental model within grid cells

267 We used starting EBOV transmission characteristics of incubation time = 7 days, onset of
268 symptoms to resolution = 9.6 days, case fatality rate (CFR) $\sigma$ = 0.78, and burial time = 2 days
269 (23) to parameterize the SEIFR compartmental model to determine transition rates $\alpha$
270 (between Exposed to Infectious compartments), $\gamma_\sigma$ (Infectious to Funeral), $\gamma_{1-\sigma}$ (Infectious

to Removed), and $\gamma_F$ (Funeral to Removed) (Fig. 2). To incorporate sensitivity around these transmission parameters, we allowed values to vary for each simulation run by sampling from a Gaussian distribution where the mean was their initial value and standard deviation was fifth of the mean, to give a reasonable spread of values. For each time step *t*, the number of individuals moving between all compartments was estimated by drawing randomly from a binomial distribution (Section S1 Equation 1), parameterized using the respective compartmental transition rates. Transition rates for compartments were assumed to be the same in all grid cells except for the transition between Susceptible to Exposed. The per grid cell Susceptible to Exposed transition rates were determined by the force of zoonotic infection $\lambda_z$, and the force of infection $\lambda$ (Fig. 2) and these were calculated as follows:

(a) Force of Zoonotic Infection, $\lambda_z$. The force of infection for zoonotic transmission $\lambda_z$, per time step *t*, was estimated as the product of the probability of host presence *H*, and spill-over rate $\kappa$ (Section S1 Equation 2). Without any evidence to the contrary (15, 50), we parameterized *H* by calculating the spatial probability of the presence of the most likely EBOV reservoir host species based on available data (Old World fruit bat species *Epomophorus gambianus gambianus*, *Epomops franqueti*, *Hypsignathus monstrosus*, and *Rousettus aegyptiacus* see Table S1) within each grid cell (0.0416°) across the African continent using species distribution models (SDMs) (51) and assuming constant pathogen prevalence. We also calculated the spatial probability of the presence of other species which are known to provide an alternative route of infection, but likely do not act as reservoirs (*Gorilla spp.*, *Pan spp.*, *and Cephalophus spp*.) (12). SDMs for each species were inferred using boosted regression trees (BRT) using distribution data from the Global Biodiversity Information Facility (GBIF) (52) and 11 present day bioclimatic and land use variables (Table S2). Data with coarse scale GBIF spatial coordinates (decimal degree coordinates with less than four decimal places) were filtered out of the analysis. To reduce spatial

296 autocorrelation and duplicate records, any records that co-occurred in the same grid cell were
297 removed. Lastly, GBIF records older than 1990 were discarded to ensure samples more
298 closely matched the current landscapes. BRT tree complexity was set at 5 reflecting the
299 suggested value and the learning rate was adjusted until >1000 trees were selected (53). A
300 total of 25 models were estimated for each species using four fifths of the distribution data as
301 a training dataset and one fifth as a testing dataset, chosen randomly for each model. Those
302 with the highest predictive ability (high area under operating curve, AUC and true-skill
303 statistic, TSS values) were selected as the best model for each species (Fig. S1). The most
304 important spatial variables determining distributions across the different reservoir host
305 species were BIO7 Temperature Annual Range, BIO13 Precipitation of Wettest Month, BIO2
306 Mean Diurnal Temperature Range and Land Use-Land Cover (Fig. S2). The outputs from all
307 putative reservoir (bat) species were combined into a single value representing the probability
308 of any reservoir species being present and a similar approach was taken for the non-reservoir
309 host species. The reservoir and non-reservoir host layers were then combined, but since only
310 a third of index cases were attributed to non-reservoir host spill-overs (10), we down-
311 weighted the probability of the non-reservoir occurrence by two thirds and reservoir
312 occurrence by one third when combining the layers. The final resulting probability was
313 bounded by zero and one. Additionally, as EBOV presence in non-reservoir host species is
314 impossible without the presence of reservoir hosts, cells with a reservoir host probability of
315 zero were given a value of zero irrespective of the non-reservoir host score. For
316 computational simplicity, we assume that all human individuals have equal chance of
317 exposure to infected host species. The initial value used for spill-over rate $\kappa$, per time step $t$,
318 was estimated from the number of historic outbreaks $O$ (defined here as distinct clusters of
319 cases) (taken from empirical EBOV outbreak data 12), and the number of historically
320 susceptible individuals $S_h$ (inferred from human population estimates from 1976 to 2015 from

37) (see Section S1 Equation 3). During each simulation run, κ was allowed to vary using the same method as the compartmental transmission parameters above.

(b) Force of Infection, λ. The force of infection for human-to-human transmission $\lambda$ per time step *t*, was estimated as the product of the effective contact rate $\beta$, and the number of individuals that can transmit the disease in each relevant compartment (*Infectious and Funeral)* per grid cell (0.0416°) (Section S1 Equation 4). We assumed that $\beta$ for the Infectious and Funeral compartments was equivalent, due to the contact rates of moving individuals in the Infectious compartment being offset by large aggregations of individuals at funerals. We estimated the effective contact rate $\beta$, as the basic reproduction number $R_0$ divided by the product of the total number of individuals *N*, and infectious duration *D* (the sum of Infectious and Funeral compartment time, 11 days taken from 23). As a starting value for $R_0$ we used a value of 1.7 (54) and this was allowed to vary per simulation run using the same method as the compartmental transmission parameters above. As per previous research(30), we incorporated spatial variance in contact rates among grid cells using a weighting factor *m*, whereby the effective contact rate in grid cells with greater than expected contact rates was increased and decreased where fewer contacts were predicted (Section S1 Equation 5). We estimated *m* by creating an ideal free gas model of human movement within each grid cell and approximated collision frequency per person per day, using the following: the total individuals in each grid cell (estimated from Gridded Population of the World v3 55), an individual interaction sphere of radius 0.5 m, and using per person, daily walking distances in meters $v\Delta t$, where *v* is walking velocity, and $\Delta t$ equals time period (Section S1 Equation 6). To capture geographic variation in human movement patterns, each grid cell was assigned a value for per person daily walking distance, based on the empirical relationship between daily walking distances and per person per country Gross Domestic Product (measured as Purchasing Power Parity from 37) (Table S3). As the availability of mass transit

as alternative to walking tends to be centrally controlled, we assumed that grid cells in each country had the same value.

Under real conditions, the effective reproduction number $R_e$ decays over time as both efforts are made to control disease spread and as the pool of susceptible reduces, which results in $R_0$ being equal to $R_e$ only when time step $t$ is zero. Therefore, to calculate effective contact rate $\beta$, we allowed $R_e$ to decay per time step $t$ (Section S1 Equations 7, 8 and 9). However, countries that can invest more in health infrastructure (e.g., barrier nursing, surveillance) should see a more rapid reduction in $R_e$ over time compared to countries that do not have such infrastructure and also a concomitantly, a decrease in CFR. Therefore we derived an empirical estimate of the relationship between wealth (measured using GDP-PPP per capita) and both the relative rate of decay of $R_e$ over time (Section S1 Equation 10) and CFR (Section S1 Equation 11), and using a spatially disaggregated poverty data layer (56) we weighted the per grid cell per time step $R_e$ reduction and CFR accordingly to the values in each grid cell. While we found the relationship between wealth and both $R_e$ and CFR reduction over time to be best described using curves with exponents of -0.08 and -0.02, respectively, this was inferred using relatively few data points (Table S4). In our simulation runs, therefore, we allowed these exponents to vary similarly to the parameters above, to allow either more linear declines or deeper curves to best estimate the true impact of this relationship.

Stage 2: SEIFR compartmental model between grid cells

We allowed those individuals that had contracted EBOV to travel between grid cells, specifically individuals in Exposed and Infectious (but not Funeral) compartments (Fig. 2), but assumed for simplicity that the overall net movement of susceptible individuals between cells was zero. As previously supported with empirical data, we employed a movement model that was weighted by both geographic distance and human density (31, 33) and was also geographically constrained to known transportation routes. The transmission rate $\varepsilon$, of

individuals between target compartments of different grid cells was estimated by two different methods: between grid cells along road networks $\varepsilon_r$, and along flight routes $\varepsilon_f$. We sampled randomly, from a binomial distribution, the number of travellers per grid cell and time step $t$ (Section S1 Equation 1) with the probability of travel by road per day $\varepsilon_r$, being proportional to the distance to the nearest road using the Global Roads Open Access Data Set (Global Roads Open Access Data Set from 57). Global roads dataset contains in total 585413 routes from tracks to multi-lane highways and has been extensively validated for Africa (58). We allowed travellers to move freely (agnostic to any particular transportation method or country boundary) across the continent up to 10 road junctions in any direction from the centroid of the starting cell along the road network (Global Roads Open Access Data Set from 57), giving a potential of up to 500 km of linear travel per time step. Each proposed travel end point was given an individual probability from the daily distance travelled probability curve from (Fig. 2(f) of 59), which is derived from transport data and validated against mobile phone data. For air travel, we set the potential pool of travellers as the individuals in grid cells containing airports across the world (from Open Flights Airport Database 60) plus all the Exposed individuals in the 8 grid cells surrounding each airport grid cell. We sampled randomly from a binomial distribution the number of travellers per grid cell and time step $t$ (Section S1 Equation 1) with the probability of travel by air per day $\varepsilon_f$, being proportional to the total number of flights per day divided by the population of that country (37). We allowed travellers to move up to 2 edges on the current airline routes from airport origin using the (from Open Flights Airport Database 60). This approximates a traveller taking either a one or two-legged journey. Final destinations were sampled at random, based on all potential air routes having equal priority, but in most cases potential destinations were located nearby which by default meant that more distance travel was less likely than travel to a nearby location. For both road and air travellers, individuals were then added to the correct

compartment of their final destination in the new grid cell and removed from the same compartment from the original source grid cell.

Stage 3: Impact of future anthropogenic change

(a) Future force of zoonotic infection $\lambda_z$. We recalculated values of the force of zoonotic infection $\lambda_z$, by estimating the probability of EBOV host presence, $H_{2070}$ under several different future integrated scenarios that incorporate projections of bioclimatic and land use variables (Table S2). Estimates of bioclimatic variables for 2070 were based on the HADGem3-AO climate model (61) under three Representative Concentration Pathways: RCP4.5, RCP6, and RCP8.5 (RCP45, RCP60 and RCP85 40). To estimate host presence probability in the future we needed to predict fine-scale future habitat data under the RCP scenarios. As only coarse categorisations are currently available (62), we therefore separately empirically estimated future land use-land cover (LULC) change (using MODIS data 36). For each grid cell we calculated the probability of each possible LULC change within the 2001-2012 MODIS dataset within a surrounding 5x5 cell grid using satellite data from 20. Based on these probabilities we simulated yearly LULC change across the region of interest for each grid cell from 2012 until 2070, and ran this simulation 100 times to create a bank of future possible landscapes, which were then summarized into three consensus landscapes representing low (with anthropogenic changes rejected where possible) , medium (by choosing the majority consensus across all 100 runs) and high anthropogenic change, (anthropogenic changes were chosen if available across the landscape) and we aligned these three scenarios to SSP1, SSP2 and SPP3 respectively.

(b) Future force of infection $\lambda$. Using predicted human demographic variables and poverty levels for 2070, we recalculated values for the force of infection $\lambda$, by estimating the number of individuals per grid cell, $n$ and effective reproduction number, $R_e$. We inferred human population estimates per grid cell for 2070 by using the Gridded Population of the World v4

(55) for present day and multiplying each cell by the expected future proportional change over that time period predicted by three Shared Socio-economic Pathways: SSP1, SSP2 and SSP3. Future poverty estimates per country were similarly inferred using a spatially-disaggregated GDP layer (63) multiplied by the expected change in per country GDP over the time period as predicted by the SSP integrated scenario. We note that as our travel probability is defined per person, increasing future populations will see a proportion increase in the amount of both road and air travel.

(c) Comparison of simulation runs. We reran the EMM simulations under 5 plausible combinations of 2070 future environmental-socioeconomic scenarios of global change and greenhouse gas concentrations: RCP4.5/SSP1, RCP4.5/SSP2, RCP6/SSP2, RCP6/SSP3, RCP8.5/SSP3 (64). These different input data options were, specifically: (i) RCP 4.5 - stabilization scenario in which total radiative forcing is stabilized shortly after 2100, (ii) RCP 6 - stabilization scenario in which total radiative forcing is stabilized shortly after 2100, without overshoot, by the application of a range of technologies and strategies for reducing greenhouse gas emissions (iii) RCP 8 – worsening scenarios with increasing greenhouse gas emissions over time, leading to high greenhouse gas concentration levels, (iv) SSP1 – high regional cooperation, low population growth due high education and high GDP growth, (v) SSP2 – a 'processes as usual' scenario with ongoing levels of population growth and wealth, with medium estimates for both these by 2070, and (vi) SSP3 – regional antagonism, high population growth, unsustainable resource extraction and low economic growth. For each of the six scenarios we aimed for 2500 runs of 365 days, each day measuring the number of spill-overs, the number of secondary cases associated with each spill-over, and the geographical areas affected. This allowed us to measure likelihood of spill-overs leading to small, medium and very large outbreaks, and also to determine the geographical areas with

445 the highest risk of experiencing cases. We also noted the destination of any flights out of
446 Africa that contained infected people.

447 **Acknowledgments:** This work, Dynamic Drivers of Disease in Africa Consortium, NERC
448 project no. NE-J001570-1 was funded with support from the Ecosystem Services for Poverty
449 Alleviation Programme (ESPA). The ESPA programme is funded by the Department for
450 International Development (DFID), the Economic and Social Research Council (ESRC) and
451 the Natural Environment Research Council (NERC). AAC is additionally supported by a
452 Royal Society Wolfson Research Merit Award. We thank Prabu Sivasubramaniam for
453 technical assistance, and A., Jones, M. Wilson, G. Mace, M. Leach, and C. Watts for
454 comments on previous versions of the manuscript. All simulation data are available on
455 figshare (figshare.com/s/c41c50a0675311e5b6b306ec4bbcf141) and the supplementary
456 materials contain all other data. D.W.R. and K.E.J. developed the overall study design.
457 D.W.R. carried out the modelling and data processing with assistance from K.E.J. All authors
458 contributed to writing the manuscript. The authors declare no competing financial interests.

459 **References:**

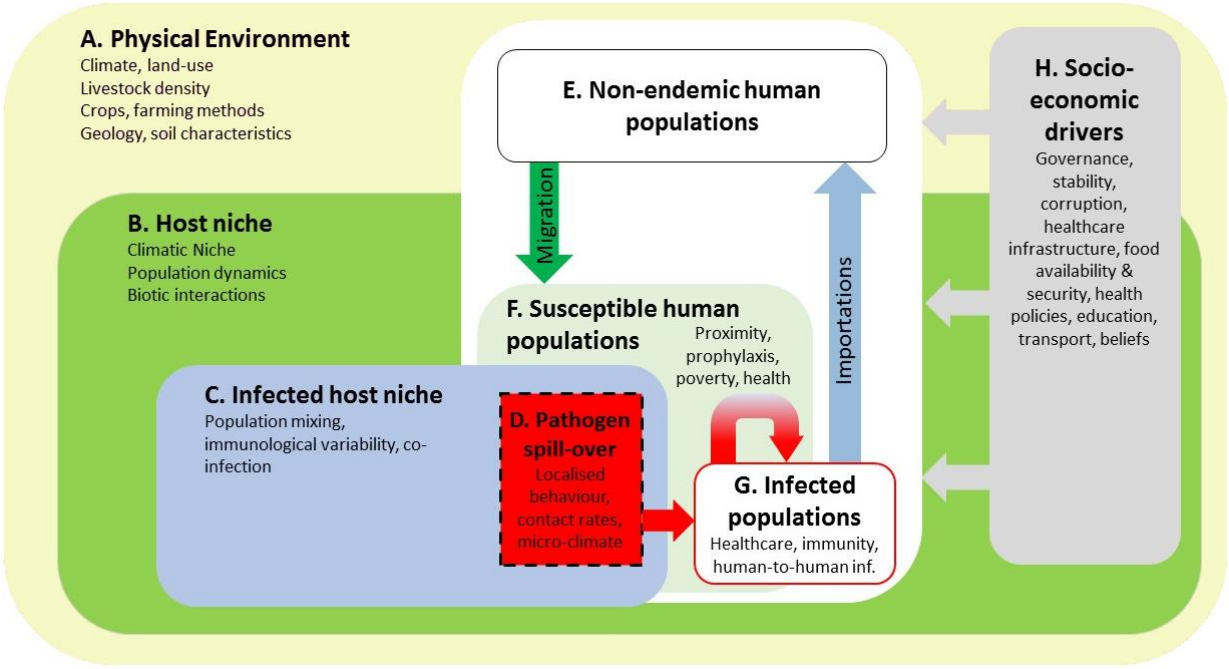
460 1. Whitmee S, *et al.* (2015) Safeguarding human health in the Anthropocene epoch:
461 report of The Rockefeller Foundation–Lancet Commission on planetary health. *The*
462 *Lancet* 386: 1973-2028.
463 2. Civitello DJ, *et al.* (2015) Biodiversity inhibits parasites: Broad evidence for the
464 dilution effect. *Proceedings of the National Academy of Sciences* 112:8667-8671.
465 3. Costello A, *et al.* (2009) Managing the health effects of climate change. *The Lancet*
466 373(9676):1693-1733.
467 4. Lafferty KD (2009) Calling for an ecological approach to studying climate change
468 and infectious diseases. *Ecology* 90:932-933.
469 5. Keesing F, *et al.* (2010) Impacts of biodiversity on the emergence and transmission of
470 infectious diseases. *Nature* 468:647-652.
471 6. Taylor LH, Latham SM, & Woolhouse ME (2001) Risk factors for human disease
472 emergence. *Philosophical Transactions of the Royal Society London, Series B:*
473 *Biological Sciences* 29:983-989.
474 7. Grace D, *et al.* (2012) Mapping of poverty and likely zoonoses hotspots.
475 (International Livestock Research Institute, Nairobi, Kenya).
476 8. Hotez PJ, Fenwick A, Savioli L, & Molyneux DH (2009) Rescuing the bottom billion
477 through control of neglected tropical diseases. *The Lancet* 373(9674):1570-1575.

9. Hotez PJ, *et al.* (2007) Control of neglected tropical diseases. *New England Journal of Medicine* 357(10):1018-1027.
10. Alexander KA, *et al.* (2015) What factors might have led to the emergence of Ebola in West Africa? *PLoS Neglected Tropical Diseases* 9:e0003652.
11. Centre for Disease Control (2015) 2014 Ebola Outbreak in West Africa.
12. Pigott DM, *et al.* (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* 3:e04395.
13. Legrand J, Grais RF, Boelle P-Y, Valleron A-J, & Flahault A (2007) Understanding the dynamics of Ebola epidemics. *Epidemiology and Infection* 135:610-621.
14. Leroy EM, *et al.* (2005) Fruit bats as reservoirs of Ebola virus. *Nature* 438:575-576.
15. Leendertz SAJ (2016) Testing New Hypotheses Regarding Ebolavirus Reservoirs. *Viruses* 8(2):30.
16. Pigott DM, *et al.* (2016) Updates to the zoonotic niche map of Ebola virus disease in Africa. e*Life* 5:e16412.
17. Mickleburgh SP, Hutson AM, & Racey PA (1992) Old World fruit bats. *An action plan for their conservation. Gland, Switzerland: IUCN*.
18. 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.
19. Kiskowski MA (2014) A three-scale network model for the early growth dynamics of 2014 west Africa ebola epidemic. PLoS Currents. doi: 10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070.
20. Weitz JS & Dushoff J (2015) Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control. *Scientific Reports* 5:8751.
21. Althaus CL, Low N, Musa EO, Shuaib F, & Gsteiger S (2015) Ebola virus disease outbreak in Nigeria: transmission dynamics and rapid control. *Epidemics* 11:80-84.
22. Nishiura H & Chowell G (2014) Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. *Eurosurveillance* 19:20894.
23. Valdez LD, Aragão Rêgo HH, Stanley HE, & Braunstein LA (2015) Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. *Scientific Reports* 5:12172.
24. Li S-L, *et al.* (2017) Essential information: Uncertainty and optimal control of Ebola outbreaks. *Proceedings of the National Academy of Sciences* 114(22):5659-5664.
25. Alexander KA, *et al.* (2015) What factors might have led to the emergence of Ebola in West Africa? *PLoS Neglected Tropical Diseases* 9(6):e0003652.
26. Iacono GL, *et al.* (In Review) The environmental limits of Rift Valley Fever revealed using eco-epidemiological mechanistic models. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES*.
27. Pourrut X, *et al.* (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* 9:159.
28. Bogoch II, *et al.* (2016) Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study. *The Lancet Infectious Diseases* 16(11):1237-1245.
29. Messina JP, *et al.* (2016) Mapping global environmental suitability for Zika virus. *Elife* 5:e15272.
30. Redding DW, Moses LM, Cunningham AA, Wood J, & Jones KE (2016) Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever. *Methods in Ecology and Evolution* 7(6):646-655.

31. Dudas G, *et al.* (2017) Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544(7650):309-315.
32. Mayhew, M. (2017) World Bank Group Ebola Response Fact Sheet. (http://www.worldbank.org/en/topic/health/brief/world-bank-group-ebola-fact-sheet)
33. Kramer AM, *et al.* (2016) Spatial spread of the West Africa Ebola epidemic. *Royal Society Open Science* 3(8):160294.
34. Garchitorena A, *et al.* (2017) Disease ecology, health and the environment: a framework to account for ecological and socio-economic drivers in the control of neglected tropical diseases. *Philosophical Transactions of the Royal Society London, Series B: Biological Sciences* 372(1722):20160128.
35. Henao-Restrepo AM, *et al.* (2015) Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *The Lancet* 386(9996):857-866.
36. Friedl MA, *et al.* (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterisation of new datasets. *Remote Sensing of the Environment* 114:168-182.
37. World Bank (2014) *World Development Indicators* (World Bank, Washington DC, USA).
38. Bellouin N, *et al.* (2011) The HadGEM2 family of met office unified model climate configurations. *Geoscientific Model Development* 4(3):723-757.
39. Riahi, K. (2013). Preliminary IAM scenarios based on the RCP/SSP framework. Climate Change Impacts/Integrated Assessment XIX.
40. van Vuuren D, *et al.* (2011) The representative concentration pathways: an overview. *Climatic Change* 109(1-2):5-31.
41. Tompkins EL & Adger WN (2005) Defining response capacity to enhance climate change policy. *Environmental Science & Policy* 8(6): 562-571.
42. MacNeil A & Rollin PE (2012) Ebola and marburg hemorrhagic fevers: Neglected Preliminary IAM scenarios based on the RCP/SSP framework tropical diseases? *PLoS Neglected Tropical Diseases* 6:1-7.
43. Ogbu O, Ajuluchukwu E, & Uneke CJ (2007) Lassa fever in West African sub-region: an overview. *Journal of Vector Borne Diseases* 44:1-11.
44. Daszak P, *et al.* (2013) Interdisciplinary approaches to understanding disease emergence: the past, present, and future drivers of Nipah virus emergence. *Proceedings of the National Academy of Sciences* 110:3681-3688.
45. Mickleburgh S, Waylen K, & Racey P (2009) Bats as bushmeat: a global review. *Oryx* 43(02):217-234.
46. Deen GF, *et al.* (2015) Ebola RNA persistence in semen of Ebola virus disease survivors—preliminary report. *New England Journal of Medicine* doi: 10.1056/NEJMoa1511410
47. Mackay IM & Arden KE (2015) Ebola virus in the semen of convalescent men. *The Lancet Infectious Diseases* 15(2):149-150.
48. Rivers CM, Lofgren ET, Marathe M, Eubank S, & Lewis BL (2014) Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia. *arXiv preprint arXiv:1409.4607*.
49. R Development Core Team (2009) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).
50. Leendertz SAJ, 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(1):18-25.

51. Phillips S, Anderson RP, & Schapire R (2006) Maximum entropy modelling of species geographic distribution. *Ecological Modelling* 190:231-259.
52. Global Biodiversity Information Facility (2013) Global Biodiversity Information Facility. (http://www.gbif.org/).
53. Elith J, *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
54. Althaus CL (2014) Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *arXiv preprint arXiv:1408.3505*.
55. CIESIN (2005) Gridded Population of the World Version 3 (GPWv3): Population Grids. (NASA Socioeconomic Data and Applications Center (SEDAC), Palisades NY).
56. Ghosh T, *et al.* (2010) Shedding light on the global distribution of economic activity. *The Open Geography Journal* 3(1).
57. CIESIN (2013) Global Roads Open Access Data Set, Version 1 (gROADSv1). . (NASA Socioeconomic Data and Applications Center (SEDAC), Palisades NY).
58. CIESIN (2013) Methods Used in the Development of the Global Roads Open Access Data Set (gROADS), version 1. (NASA Socioeconomic Data and Applications Center (SEDAC), Palisades NY).
59. Merler S & Ajelli M (2010) The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences* 277:557-565.
60. OpenFlights (2012) OpenFlights Airports Database. (http://openflights.org).
61. Hewitt HT, *et al.* (2011) Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. *Geoscientific Model Development* 4:223-253.
62. Hurtt GC, *et al.* (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic change* 109(1-2):117.
63. Elvidge CD, *et al.* (2009) A global poverty map derived from satellite data. *Computers & Geosciences* 35(8):1652-1660.
64. Kriegler E, *et al.* (2014) A new scenario framework for climate change research: the concept of shared climate policy assumptions. *Climatic Change* 122(3):401-414.
65. Hutchinson J & Waser PM (2007) Use, misuse and extensions of "ideal gas" models of animal encounter. *Biological Reviews* 3:335-359.
66. Pigott DM, *et al.* (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* 3:e04395.
67. Wilson DE & Reeder DM eds (2005) *Mammal Species of the World. A Taxonomic and Geographic Reference* (John Hopkins University Press, Baltimore), 3rd Edition Ed, Vol 1.
68. Weiß aus Karlsruhe S (2013) Identification and characterisation of emerging viruses in free-ranging bats from sub-Saharan Africa. EngD Thesis (Technical University of Berlin, Berlin, Germany).
69. Hayman DTS, *et al.* (2010) Long-Term Survival of an Urban Fruit Bat Seropositive for Ebola and Lagos Bat Viruses. *PloS One* 5:e11978.
70. Hayman DTS, *et al.* (2012) Ebola virus antibodies in fruit bats, ghana, west Africa. *Emerging Infectious Diseases* 18:1207-1209.
71. 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.

72. Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org).
73. Barreira TV, Rowe DA, & Kang M (2010) Original Research Parameters of Walking and Jogging in Healthy Young Adults. *International Journal of Exercise Science* 3(1):4-13.
74. Hallal PC, *et al.* (2012) Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet* 380(9838):247-257.
75. Filmer D (2004) *If You Build It, Will They Come? School Availability and School Enrollment in 21 Poor Countries* (The World Bank) p 25.
76. Cook I, Alberts M, & Lambert EV (2008) Relationship between adiposity and pedometer-assessed ambulatory activity in adult, rural African women. *International Journal of Obesity* 32(8):1327-1330.
77. Marlowe F (2010) *The Hadza: Hunter-gatherers of Tanzania* (University of California Press).
78. BASSETT DRJ, WYATT HR, THOMPSON H, PETERS JC, & HILL JO (2010) Pedometer-Measured Physical Activity and Health Behaviors in U.S. Adults. *Medicine & Science in Sports & Exercise* 42(10):1819-1825.
79. Chowell G & Nishiura H (2014) Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC medicine* 12(1):196.
80. Lefebvre A, *et al.* (2014) Case fatality rates of Ebola virus diseases: a meta-analysis of World Health Organization data. *Médecine et Maladies Infectieuses* 44(9):412-416.
81. Kucharski AJ & Edmunds WJ (2014) Case fatality rate for Ebola virus disease in west Africa. *The Lancet* 384(9950):1260.
82. Fasina FO, *et al.* (2014) Transmission dynamics and control of Ebola virus disease outbreak in Nigeria, July to September 2014. *European Surveillance* 19(40):20920.
83. Rouquet P, *et al.* (2005) Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001–2003. *Emerging Infectious Diseases* 11(2):283.
84. Team WER (2014) Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *New England Journal of Medicine* 2014(371):1481-1495.
85. Towers S, Patterson-Lomba O, & Castillo-Chavez C (2014) Temporal variations in the effective reproduction number of the 2014 West Africa Ebola outbreak. *PLoS Currents* doi: 10.1371/currents.outbreaks.9e4c4294ec8ce1adad283172b16bc908
86. Althaus CL (2014) Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Currents* doi: 10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288




667

668 **Fig. 1. System-dynamics model of zoonotic disease transmission.** Letters A-H indicate

669 major system components, arrows showing links, and key sub-components in smaller font.

670

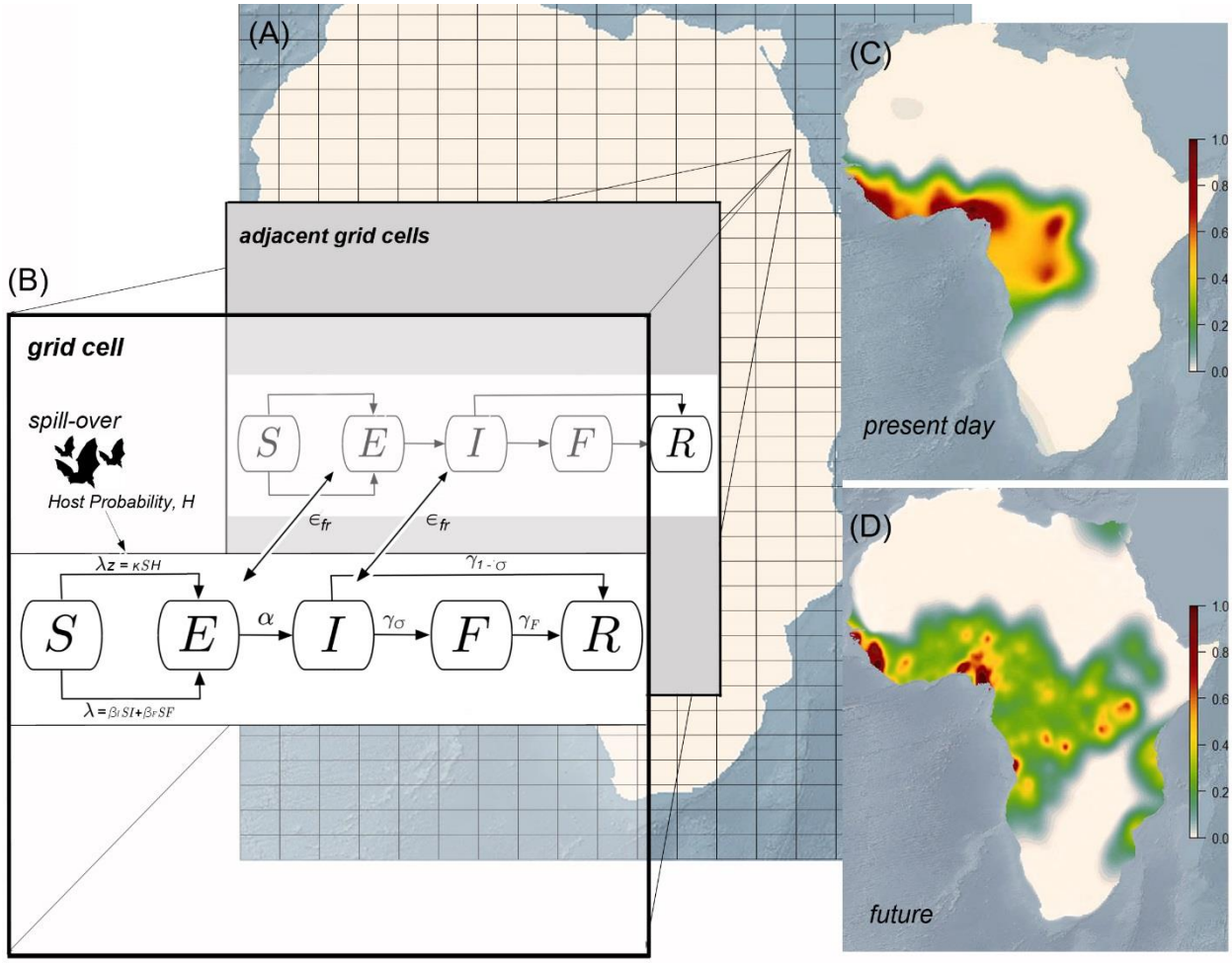



**Fig. 2. Predictive Integrated Zoonotic Model (EMM) EBOV Simulation Schematic**. Within 0.0416° grid cells across the globe (**A**), we used a *SEIFR* (*Susceptible, Exposed, Infectious, Funeral* and *Removed*) disease compartmental model (**B**), to estimate the number of people in each compartment. *S*-E transmission rate was determined for each grid cell by calculating the force of zoonotic infection (between hosts and humans) $\lambda_z$, and within human populations $\lambda$ (see Materials and Methods). Travel of exposed or infectious individuals between grid cells occurred across existing road and flight transport networks, with transmission rate $\varepsilon_{fr}$. Mean transition rates used as the starting parameters for simulations were as follows: $\alpha$ for *E*-*I* was calculated as the reciprocal of incubation time in days ($\alpha$= 1/7), $\gamma_\sigma$ (*I*-*F* transition rate) was the product of the probability of the reciprocal of days infectious ($\gamma$=1/9.6) and poverty-weighted case fatality rate ($\sigma$=0.78), $\gamma_{1-\sigma}$ (*I*-*R* transition

683 rate) was the product of the probability of the reciprocal of days infectious ($\gamma$=1/9.6) and
684 probability of recovering (1-σ), and$\gamma_F$ (F-R transition rate) was the reciprocal of the burial
685 time of 2 days. Each simulation was run 2500 times for 365 days in each grid cell containing
686 a human population. The total number of people in each compartment per grid cell, per day
687 from each simulation was then used to calculate the total number of index and secondary
688 cases and mapped spatially (**C**). Bioclimatic, land use and demographic conditions were then
689 changed to predicted values for 2070 to estimate changes to $\lambda$ and $\lambda_z$, and the simulations
690 repeated to investigate impacts of global change on disease (**D**).

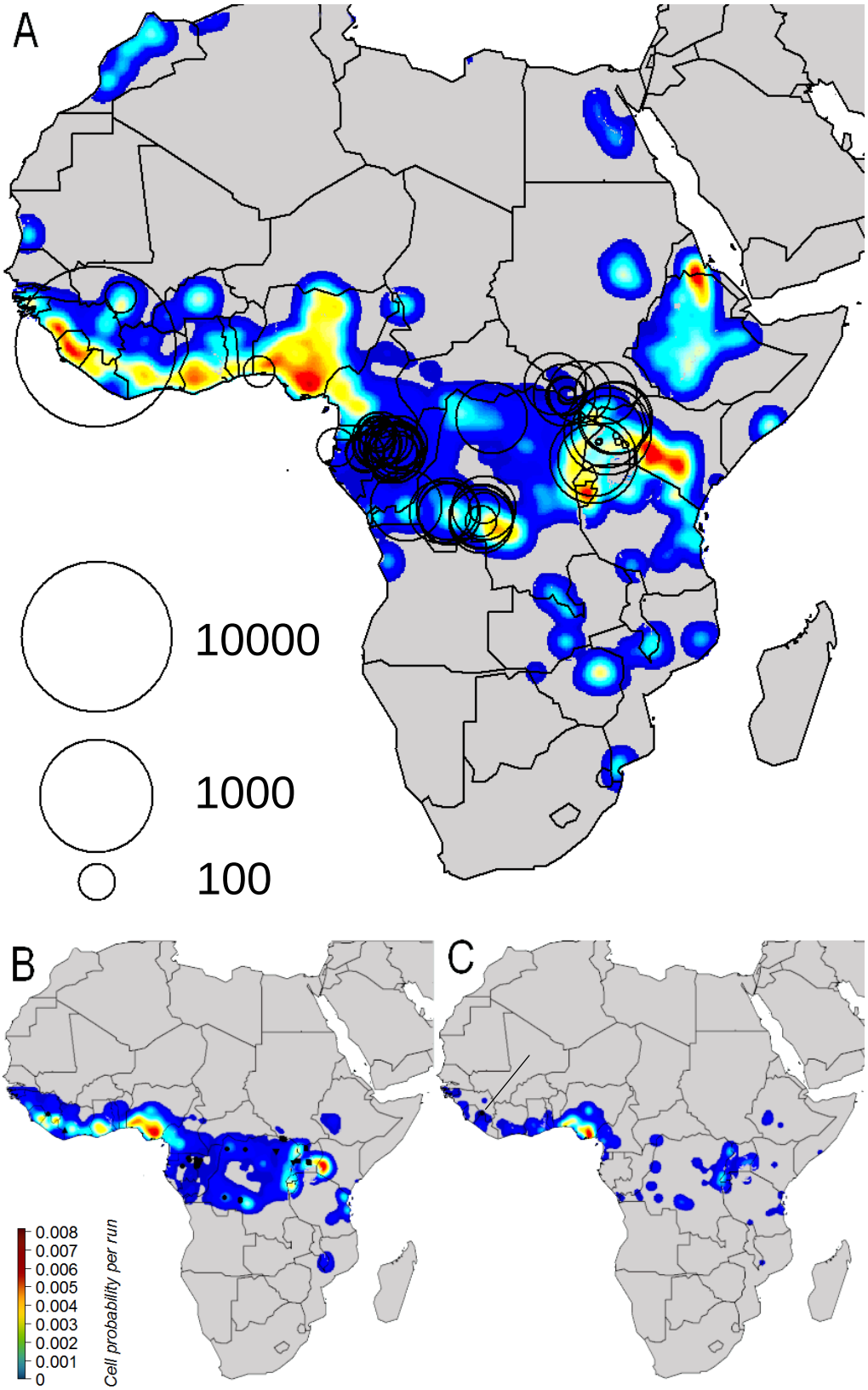



**Fig. 3. Present day risk for Zaire Ebola virus (EBOV) from EMM simulations**. Maps represent the proportion of times between zero (dark blue) and 0.01 (red) when a EVD-

EBOV case was present in a grid cell (0.0416°) across 2500, 365 day simulation runs for the present day, where (**A**) shows all cases (both index and secondary), (**B**) index cases only, and (**C**) index cases from epidemics (1500+ cases). Black open circles in (**A**) represent log outbreak size with the location of the index case at the centre of the circle. Black symbols in (**B**) represent all locations of known EVD index cases from different viral strains, where circles represent Zaire (EBOV), square Sudan (SUDV), triangles Taï Forest (TAFV), and tetrahedrons Bundibugyo (BDBV). Single black circle in (**C**) shows the only known site where an epidemic has occurred, with the black line highlighting its location.

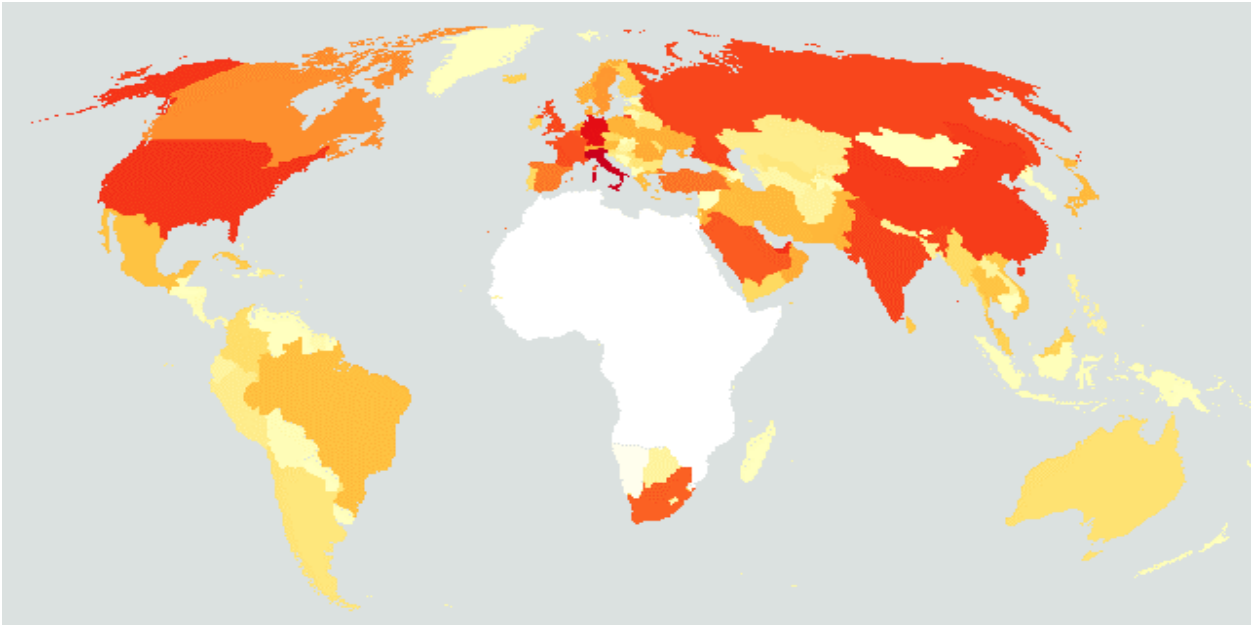
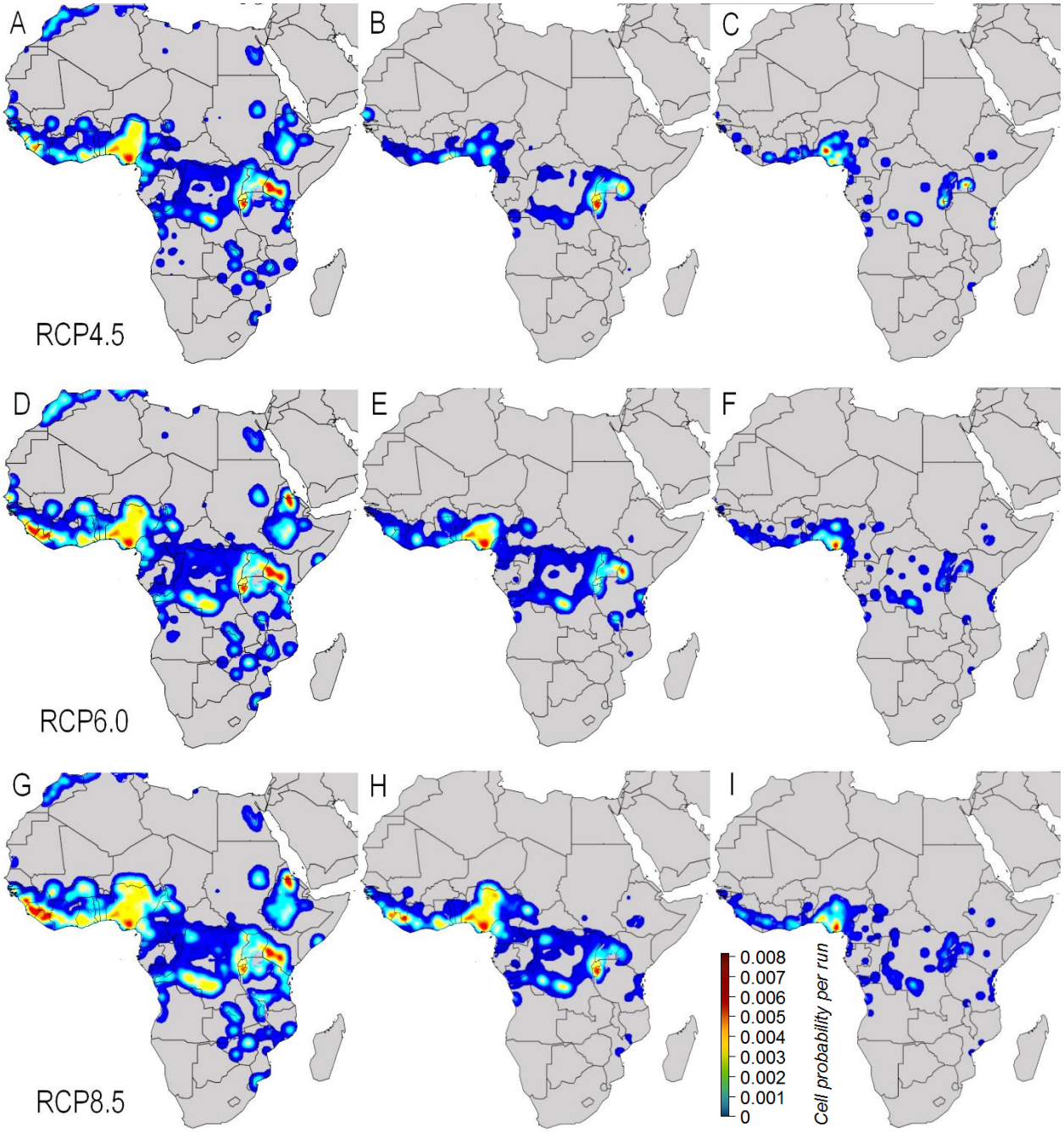
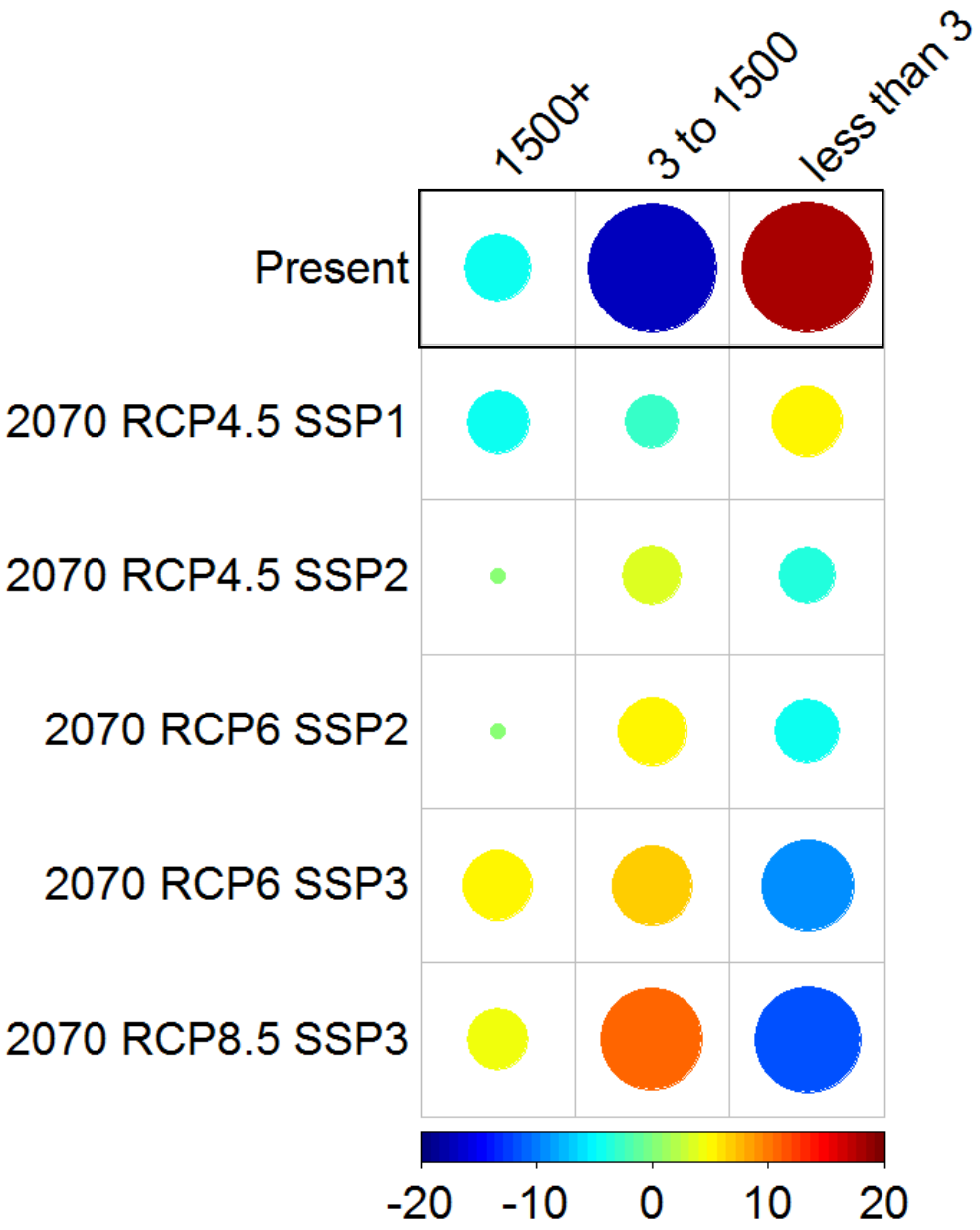


**Fig. 4. Most common country locations for importation of EBOV infected individuals.** Map shows the countries that received, by airline flights, the most EBOV infected individuals (Red) with paler, orange and then yellow coloured countries having proportional fewer importations and white showing the EVD endemics area. Data come from 2500 simulations of EVD outbreaks under present data climate, land-use, demographic and transportation conditions.




727

728 **Fig. 5**. **Future risk for Zaire Ebola virus (EBOV) from EMM simulations for 2070**. Maps
729 represent the proportion of times between zero (dark blue) and 0.01 (red) when a EVD-
730 EBOV case was present in a grid cell (0.0416°), where (**A, D, G**) show all cases (both index
731 and secondary), (**B, E, H**) index cases only, and (**C, F, I**) index cases from epidemics (1500+
732 cases), with data from EMM simulations for 2070, where rows show three different scenarios
733 of global change (RCP4.5/SSP1, RCP6.0/SSP2, RCP8.5/SSP3).




734

735 **Fig. 6. Comparison of 2070 EMM simulation scenarios by EVD-EBOV final epidemic**
736 **size**. Circles represents standardized residuals from a chi-squared test of association between
737 simulation scenario and final outbreak size category. More orange/red colours show greater
738 than expected number of outbreaks in a cell (for any given scenario and final outbreak size),
739 with more blue colours representing fewer than expected outbreaks. Size of circle indicates
740 the quantity greater or less than expected, with large circle more different than expected from

741 random allocation of simulation runs among grid cells and small circles close to the expected
742 number.

## Supplementary Information:

## Section S1

EMM compartmental transition algorithm.

For each time step *t*, the number of individuals moving through disease compartments both within and between grid cells (see Fig. 2) was estimated using disease transmission parameters. We predicted the likely movement between disease compartments per time step, by drawing randomly from a binomial distribution. We describe this process below, using as an example the movement of individuals moving from Exposed to Infectious compartments within grid cells.

1. We determined the probability that a number of individuals were likely to move from the Exposed to Infectious compartments as:

$$p(k_i \; infections \; within \; E_t) = \binom{E_t}{k_i} \alpha_t^{k_i} (1 - \alpha_t)^{E_t - k_i} \qquad \text{(Equation 1)}$$

   where $k_i$ represents the number of individuals that enter the Infectious compartment, $E_t$ the number of Exposed individuals at time *t*, and $\alpha_t$ the transition probability at time *t*.
2. Using Equation 1, we determined for any value of $k_i$ the probability of $k_i$ individuals that move into the Infectious compartment, i.e., we computed the probability of the number of people, between 0 and the total number of individuals in the Exposed compartment, entering the Infectious compartment at time step *t*. We then drew randomly from this probability distribution to choose $k_i$ individuals that moved into the Infectious compartment, thereby weighting the choice towards the more likely outcomes given $\alpha$.
3. Once the number of people that will be infected in the next time step $k_i$ was determined, then $k_i$ individuals were removed from the Exposed compartment and added to the Infectious compartment.
4. This process continued (per time step) until the number of individuals in the Exposed compartment equaled zero.

The same process was applied to every compartment change using the respective transition probabilities (i.e., substituting $\alpha$ in the above example). Movement of individuals between respective Exposed and Infectious compartments between grid cells was also

775 modelled similarly, but stopping movements if the exposed or infectious number dropped to
776 zero but with no change to susceptible numbers. Due to the high morbidity from this disease,
777 individuals in the Infectious compartment were deemed less likely to travel and were
778 awarded a travel probability that was half of the expected rate for non-symptomatic
779 individuals.
780
781 Force of zoonotic infection, $\lambda_z$ algorithms.

782 The force of infection for zoonotic host to human transmission, $\lambda_z$ was estimated per grid
783 cell, per time step *t*, as follows:

784 $$\lambda z_t = \kappa H \quad \text{(Equation 2)}$$

785

786 where $\kappa$ = spill-over risk, and $H$ = probability of zoonotic host presence per grid cell. Spill-
787 over event probability, $\kappa$ per person, per time step is given by:

788

789 $$\kappa = \left(\frac{O}{S_h T}\right) \quad \text{(Equation 3)}$$

790

791 where $O$ = number of historic outbreaks, $S_h$ = number of historically susceptible individuals,
792 and $T$ = total time when infections could have occurred. Note: Above we are estimating the
793 probability of an individual being involved in a spill-over event directly from an animal host,
794 which is distinct from the overall risk of contracting the disease.
795
796 Force of infection, $\lambda$ algorithms.

797 The force of infection for human-to-human transmission, per grid cell and per time step
798 *t*, was estimated as:

799 $$\lambda_t = \beta I_t + \beta F_t \quad \text{(Equation 4)}$$

800

801 where $\beta$ = effective contact rate, $I_t$ = number of individuals in Infectious compartment at time
802 step *t*, and $F_t$ = number of individuals in Funeral compartment at time step *t*. For simplicity
803 we assumed that $\beta I$ and $\beta F$ were the same (hereafter referred to as $\beta$). When $t = 0$, $\beta$ is given
804 by:

805

806 $$\beta = m * \left(\frac{R_0}{ND}\right) \quad \text{(Equation 5)}$$

807

where $R_0$ = basic reproduction number, $m$ = mobility, $N$ = population size per time step, and $D$ = duration in days that an individual is infectious. In this context, $m$ was used to modify the ideal free gas model of human movement with distances travelled which are spatially variable across the landscape. We calculated a two-dimensional collision frequency $c$, per person per grid cell(65) as follows:

$$c = nv\Delta tq2 \quad \text{(Equation 6)}$$

where $n$ = number of individuals, $v$ = walking velocity, $\Delta t$ = time period and $q$ = interaction sphere radius. In the context of our simulation, $v\Delta t$ represents daily walking distance. Then we defined $m$ as the inverse deviation from a mean of $c$ such that areas with more movement have a higher effective contact rate. However, when $t > 0$ we redefined $\beta$ as follows:

$$\beta = m * \left(\frac{R_e}{ND}\right) \quad \text{(Equation 7)}$$

where $R_e$ = effective reproduction number, $m$ = mobility, $N$ = population size per time step, and $D$ = duration in days that an individual is infectious. $R_e$ is related to $R_0$ but due to changes in human behaviour and health care responses, $R_e$ may be lower and decline over time, in addition to the implicit reduction in $R$ as the pool of susceptibles decreases during an outbreak. We make the assumption that the effective reproduction number reduces on a daily basis due to increasingly strong health care responses over time.

So initially, when $t$ = 1:

$$R_e = aR_0 \quad \text{(Equation 8)}$$

where $R_e$ = effective reproduction number at $t$ = 1, $a$ = decay rate, and $R_0$ = basic reproduction number. However, when $t$ >1:

$$R_e^{t+1} = aR_e^t \quad \text{(Equation 9)}$$

where $R_e^t$= effective reproduction number at time $t$, and $a$ = decay rate. We define decay rate $a$ per grid cell, from the empirical relationship between wealth and health outcomes. Using either direct or derived empirical estimates of the gradient of the change in $R_e$ over time from (13, 19, 21, 22), we fitted an exponential decay curve between estimates of per captia Gross

Domestic Product measured as Purchasing Power Parity (from 37) and the gradient of $R_e$ change per day. The starting $R_e$ decay value $a$ per grid cell, was given by:

$$a = 1.024\ x\ GDP^{-w2} \quad \text{(Equation 10)}$$

where the best estimate for exponent w2 was -0.848, *GDP* = Gross Domestic Product from (63), pseudo $r^2 = 0.76$, and $n = 8$.

The poverty-weighted Case Fatality Rate (*w*CFR) per grid cell, was given by:

$$wCFR = 0.21 \ln(\frac{1}{GDP})^{-w1} \quad \text{(Equation 11)}$$

where the best estimate for exponent w1 was -0.0239, GDP = Gross Domestic Product from (63), pseudo $r^2 = 0.9081$, and $n = 20$.

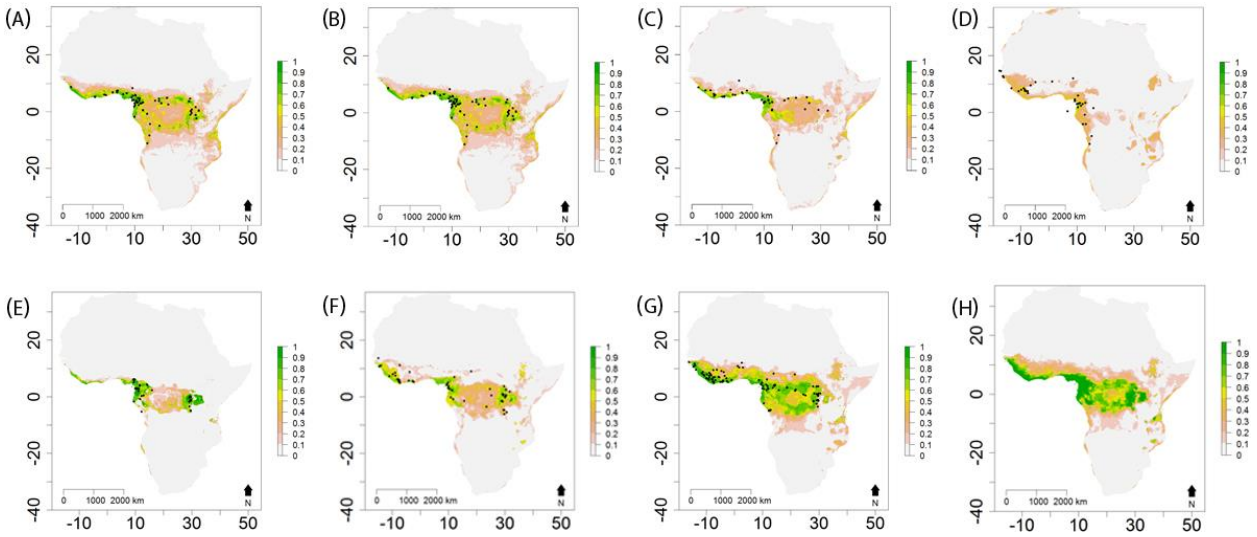



**Fig. S1**| **Maps of present day occurrence probability, H of EBOV host and other infection source species estimated from boosted-regression trees (BRT) models.** Probability of species occurrence per grid cell (0.0416°) is represented on a linear color scale where green is most suitable (p(*H*) = 1) and white unsuitable (p(*H*) = 0) where (**A**) *Epomophorus gambianus gambianus*; (**B**) *Epomops franqueti*; (**C**) *Hypsignathus monstrosus*; (**D**) *Rousettus aegyptiacus*; (E) *Gorilla spp*.; (F) *Pan spp*.; (G) *Cephalophus spp*.; and (H) all species combined. Axis labels indicate degrees in a World Geodetic System 84 projection. Filled black circles represent GBIF (52) occurrence records.

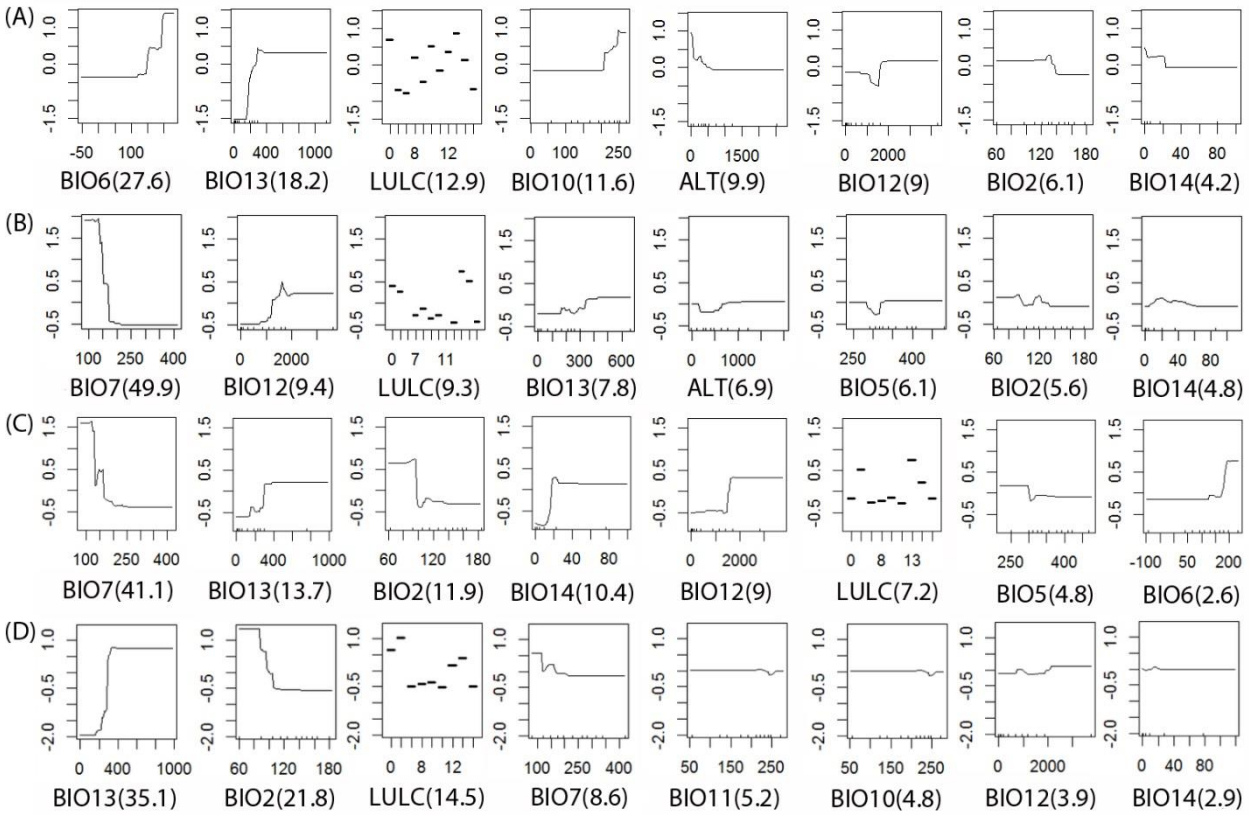



**Fig. S2| Response curves from boosted-regression trees (BRT) models of EBOV host species occurrences.** Each plot represents the shape of the normalized fitted functions for each variable where (**A**) *Epomophorus gambianus gambianus*; (**B**) *Epomops franqueti*; (**C**) *Hypsignathus monstrosus*; and (**D**) *Rousettus aegyptiacus*. The relative percentage contribution of each variable to the model in terms of variance explained is given in parenthesis, where only the top eight variables of the model are included for each species. Variable abbreviations are defined in table S2.

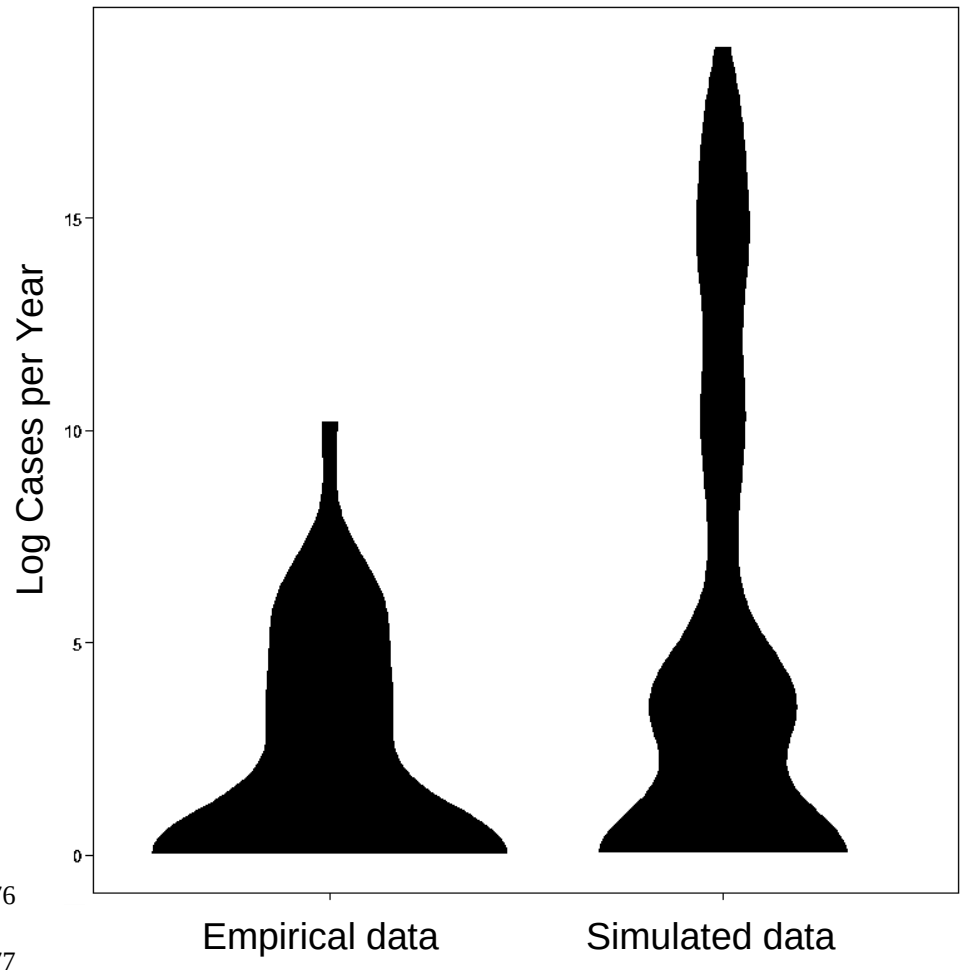



**Fig. S3| Distributions of relative frequency of EBOV cases per year.** Violin plots represent the empirical observed (n=23 outbreaks) data of log total number of cases per year from 1967-2016 (66), and log total number of cases per year (n=2500 runs) from EMM simulations for present day environmental and demographic conditions.

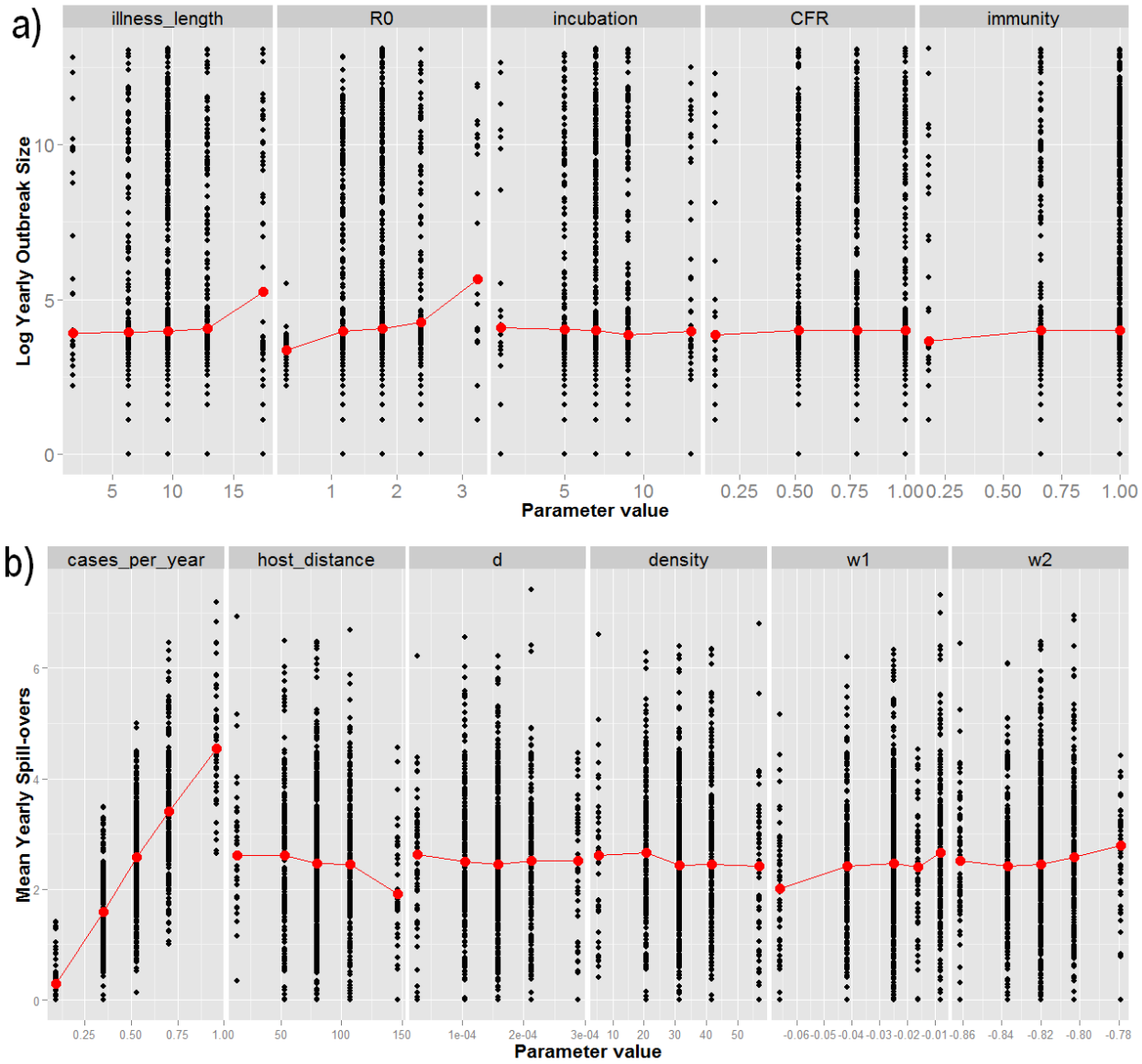



**Fig. S4| Sensitivity plots of input parameters for (a) total number of annual log EVD cases, and (b) mean annual spill-overs.** Black dots show the response values per simulation and are jittered for greater clarity. Red dots represent the median values for each parameter value, and red lines join the medians to aid interpretation of any trend. Parameters are as follows: illness length - mean number of days in the infectious compartment; $R_0$ – basic reproductive number; incubation - mean number of days in the exposed compartment; CFR - mean case fatality rate per illness; immunity - mean immunity to re-infection where 1 is totally immune; cases per year - mean spill-over rate constant; host distance - mean daily distance (m) travelled by host reservoir species; density - mean number of reservoir host individuals per grid cell; w2 - shape of the effective reproductive number ($R_e$) decay curve, where low values represent a less curved, more linear negative relationship; and w1 - shape of the CFR~poverty curve, where lower values represent a less curved, more linear negative relationship.

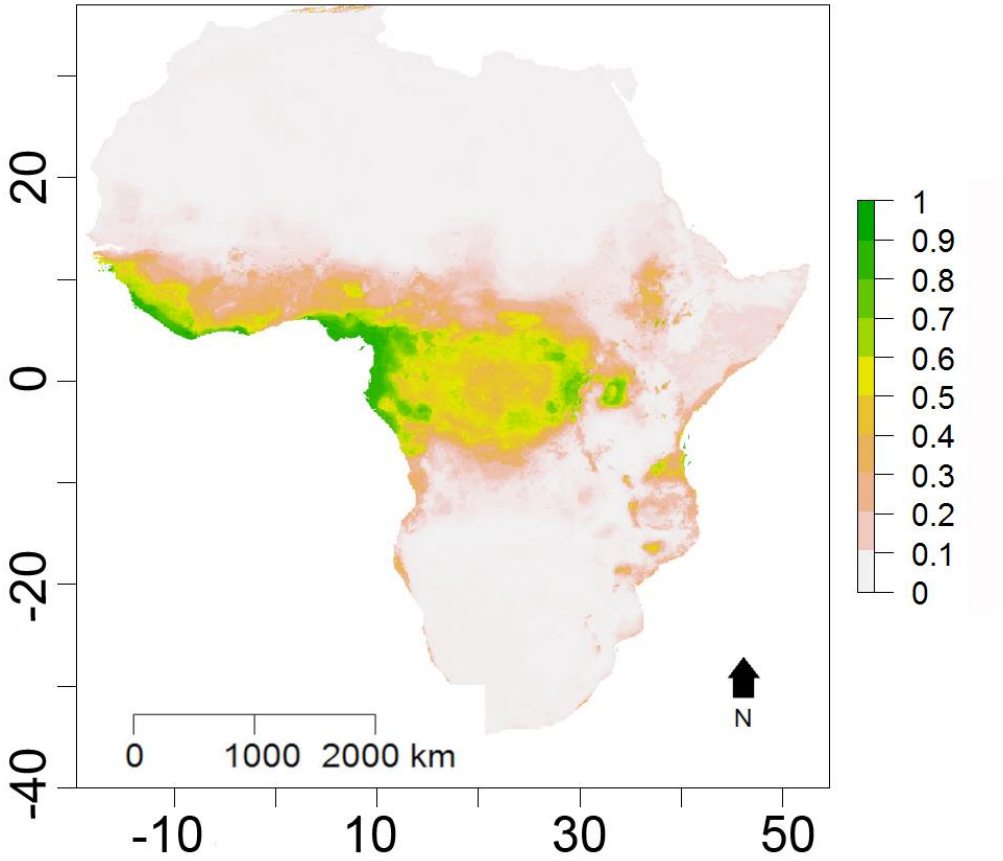



**Fig. S5| Map of future occurrence probability, $H_{2070}$ of EBOV host and other infection source species estimated from boosted-regression trees (BRT) models under the medium outlook RCP6 scenario**. Probability of species occurrence per grid cell (0.0416°) is represented on a linear color scale where green is most suitable (p(*H*) = 1) and white unsuitable (p(*H*) = 0) for all species combined. Axis labels indicate degrees in a World Geodetic System 84 projection.

**Table S1| Seroprevalance of EBOV in reservoir host species.** Species assignments followed the taxonomy in (67). Prevalence was measured as the proportion of positive results per sample and raw prevalence data was transformed to a rank within each study. Direct prevalence comparisons were not possible due to methodological differences. We estimated the most important EBOV host species as those that appear as the top two ranks in all sources. We identified four candidate bat species hosts: *Epomops franqueti*, *Epomophorus gambianus gambianus*, *Hypsignathus monstrosus*, and *Rousettus aegyptiacus*. N represents sample size; Hipp Hipposideridae; Molo Molossidae; Ptero Pteropodidae; CI Côte d'Ivoire; SL Sierra Leone; LR Liberia; GH Ghana; CG Congo; and GA Gabon.

| Family | Species | Country | N | Prevalence | Rank | Source |
|---|---|---|---|---|---|---|
| Hipp | *Hipposideros sp.* | CG, SL, LR | 98 | 0.04 | 4 | (68) |
| Molo | *Mops condylurus* | CI, SL, LR, CG | 37 | 0.05 | 4 | (68) |
| Ptero | *Eidolon helvum* | GH | 252 | 0.004 | - | (69) |
| Ptero | *Epomophorus gambianus gambianus* | GH | 37 | 0.38 | 2 | (70) |
| Ptero | *Epomops franqueti* | GH | 27 | 0.37 | 2 | (70) |
| Ptero | *Epomops franqueti* | GA, CG | 11 | 0.07 | 2 | (14) |
| Ptero | *Epomops franqueti* | GA, CG | 805 | 0.04 | 2 | (27) |
| Ptero | *Epomops franqueti* | CI, SL, LR, CG | 62 | 0.08 | 3 | (68) |
| Ptero | *Hypsignathus monstrosus* | GH | 16 | 0.44 | 1 | (70) |
| Ptero | *Hypsignathus monstrosus* | GA, CG | 17 | 0.24 | 1 | (14) |
| Ptero | *Hypsignathus monstrosus* | GA, CG | 125 | 0.07 | 1 | (27) |
| Ptero | *Hypsignathus monstrosus* | CI, SL, LR, CG | 70 | 0.16 | 2 | (68) |
| Ptero | *Micropteropus pusillus* | GA, CG | 197 | 0.02 | 4 | (27) |
| Ptero | *Micropteropus sp.* | CG | 40 | 0.03 | 4 | (68) |
| Ptero | *Myonycteris torquata* | GA, CG | 58 | 0.07 | 3 | (14) |
| Ptero | *Myonycteris torquata* | GA, CG | 573 | 0.03 | 3 | (27) |
| Ptero | *Myonycteris torquata* | CI, SL, LR, CG | 307 | 0.01 | 5 | (68) |
| Ptero | *Nanonycteris veldkampii* | GH | 4 | 0.25 | 3 | (70) |
| Ptero | *Rousettus aegyptiacus* | GA, CG | 307 | 0.08 | 1 | (27) |
| Ptero | *Rousettus aegyptiacus* | CG | 2 | 1.00 | 1 | (68) |

**Table S2| Details of bioclimatic and land use variables used to estimate probability of EBOV host presence, *H*.** Nine most orthogonal (<75% correlation) bioclimatic variables were chosen from (71). For analysis, all variables were reduced in latitudinal extent to 85° N, 58° S and resampled to a 0.0416° grid cell size using a World Geodetic System 84 projection. LULC is a categorical dataset where the most predominant land use-land cover type in each grid cell is given within the following categories: Evergreen needle leaf forest; Evergreen broadleaf forest; Deciduous needle leaf forest; Deciduous broadleaf forest; Mixed forest; Closed shrublands; Open shrublands; Woody savannah; Grassland; Permanent wetlands; Cropland; Urban and built-up; Cropland/natural vegetation mosaic; Snow and ice; Barren or sparsely vegetated; and Water bodies.

| No. | Variable Description | Original Spatial Extent | Original Spatial Resolution (cell size at equator) | Temporal Resolution | Source |
|---|---|---|---|---|---|
| 1 | BIO2 Mean Diurnal Temperature Range | Global | 1km | 2012 | (71) |
| 2 | BIO5 Maximum Temperature of Warmest Month | Global | 1km | 2012 | (71) |
| 3 | BIO6 Minimum Temperature of Coldest Month | Global | 1km | 2012 | (71) |
| 4 | BIO7 Temperature Annual Range | Global | 1km | 2012 | (71) |
| 5 | BIO10 Mean Temperature of Warmest Quarter | Global | 1km | 2012 | (71) |
| 6 | BIO11 Mean Temperature of Coldest Quarter | Global | 1km | 2012 | (71) |
| 7 | BIO12 Annual Precipitation | Global | 1km | 2012 | (71) |
| 8 | BIO13 Precipitation of Wettest Month | Global | 1km | 2012 | (71) |
| 9 | BIO14 Precipitation of Driest Month | Global | 1km | 2012 | (71) |
| 10 | ALT Digital Elevation Model | Global | 1km | 2008 | (72) |
| 11 | LULC Land Use-Land Cover | Global | 500m | 2001-2012 | (36) |

**Table S3| Estimates of global daily walking distances, vΔt.** Estimates of daily walking distances were collected from the literature per country. Daily step numbers were converted to distance (km) using an average step length of 1.41m (73). As studies have suggested that daily walking distance is stratified among income categories (74), countries were assigned to income bands based on per capita Gross Domestic Product (GDP) (measured as Purchasing Power Parity from 37) such that the poorest countries were given a value of 1 and the richest 4. A mean estimate of walking distance was calculated for each band. Countries were then assigned a walking distance corresponding to their GDP band. No estimates were found for band 3 ($1600 - $35000), so countries in this band were given daily walking distances halfway between bands 2 and 4.

| Country | Steps | Distance (km) | GDP band | GDP PPP Per capita (lower bound) $ | GDP PPP Per capita (upper bound) $ | Mean km Per GDP Band | Source |
|---|---|---|---|---|---|---|---|
| Niger | - | 7 | 1 | 0 | 1600 | 9.6 | (75) |
| Central African Republic | - | 8 | 1 | 0 | 1600 | 9.6 | (75) |
| Chad | - | 15 | 1 | 0 | 1600 | 9.6 | (75) |
| Mali | - | 13.2 | 1 | 0 | 1600 | 9.6 | (75) |
| Niger | - | 4.8 | 1 | 0 | 1600 | 9.6 | (75) |
| South Africa | 12471 | 8.85 | 2 | 1600 | 13000 | 8.5 | (76) |
| Tanzania | - | 8.3 | 2 | 1600 | 13000 | 8.5 | (77) |
| Australia | 9695 | 6.88 | 4 | 35000 | 128530 | 5.6 | (78) |
| Japan | 7168 | 5.08 | 4 | 35000 | 128530 | 5.6 | (78) |
| Switzerland | 9650 | 6.85 | 4 | 35000 | 128530 | 5.6 | (78) |
| United States | 5117 | 3.63 | 4 | 35000 | 128530 | 5.6 | (78) |

945 **Table S4| Collated epidemiological data on EBOV outbreaks.** Data on 19 locations that
946 have experienced EBOV outbreaks or importations and have data on either Case Fatality Rate
947 (CFR) (13, 18, 54, 79-83) or on Effective Reproductive Number change (21, 54, 79, 84-86)
948 ($R_e$ gradient per week). The latter data was either taken directly from tables or text from
949 within literature sources or estimated (Spain, United Kingdom, Nigeria, United States) from
950 descriptions of outbreak events detailed in the sources. Child mortality data for the year of
951 outbreak is taken from World Bank Development Indicators (37)

952

| Location | County | Year | ln GDP per capita for year | CFR | $R_e$ gradient per week |
|---|---|---|---|---|---|
| United States | Texas | 2014 | 4.74 | 0.3 | 0.5 |
| Guinea | | 2014 | 3.09 | 0.707 | 0.113636 |
| Sierra Leone | | 2014 | 3.31 | 0.69 | 0.076923 |
| Liberia | | 2014 | 2.99 | 0.723 | 0.04 |
| Germany | | 2014 | 4.66 | 0 | |
| Spain | Madrid | 2014 | 4.53 | 0 | 0.5 |
| United Kingdom | London | 2014 | 4.59 | 0 | 3 |
| Nigeria | | 2014 | 3.77 | 0.666667 | 0.533333 |
| Mali | | 2014 | 3.24 | 0.75 | |
| Congo, Dem. Rep. | | 1976 | 2.72 | 0.88 | 0.105 |
| Gabon | | 1994 | 4.14 | 0.61 | |
| Congo, Dem. Rep. | | 1995 | 2.72 | 0.81 | |
| Gabon | | Early-1996 | 4.17 | 0.68 | |
| Gabon | | Late-1996 | 4.17 | 0.75 | |
| Gabon | | 2001–2002 | 4.15 | 0.82 | |
| Congo, Rep. | | 2001–2002 | 3.58 | 0.76 | |
| Congo, Rep. | | Early-2003 | 3.6 | 0.89 | |
| Congo, Rep. | | Late-2003 | 3.6 | 0.83 | |
| Congo, Rep. | | 2005 | 3.65 | 0.75 | |

953
954
955
956