Article

Climate change increases cross-species viral transmission risk

https://doi.org/10.1038/s41586-022-04788-w

Received: 24 January 2020

Accepted: 21 April 2022

Published online: 28 April 2022

Check for updates

Colin J. Carlson[1,2,7]✉, Gregory F. Albery[1,3,7]✉, Cory Merow[4], Christopher H. Trisos[5], Casey M. Zipfel[1], Evan A. Eskew[3,6], Kevin J. Olival[3], Noam Ross[3] & Shweta Bansal[1]

At least 10,000 virus species have the ability to infect humans but, at present, the vast majority are circulating silently in wild mammals[1,2]. However, changes in climate and land use will lead to opportunities for viral sharing among previously geographically isolated species of wildlife[3,4]. In some cases, this will facilitate zoonotic spillover—a mechanistic link between global environmental change and disease emergence. Here we simulate potential hotspots of future viral sharing, using a phylogeographical model of the mammal–virus network, and projections of geographical range shifts for 3,139 mammal species under climate-change and land-use scenarios for the year 2070. We predict that species will aggregate in new combinations at high elevations, in biodiversity hotspots, and in areas of high human population density in Asia and Africa, causing the cross-species transmission of their associated viruses an estimated 4,000 times. Owing to their unique dispersal ability, bats account for the majority of novel viral sharing and are likely to share viruses along evolutionary pathways that will facilitate future emergence in humans. Notably, we find that this ecological transition may already be underway, and holding warming under 2 °C within the twenty-first century will not reduce future viral sharing. Our findings highlight an urgent need to pair viral surveillance and discovery efforts with biodiversity surveys tracking the range shifts of species, especially in tropical regions that contain the most zoonoses and are experiencing rapid warming.

In the face of rapid environmental change, survival for many species depends on moving to track shifting climates. Even in a best-case scenario, the geographical ranges of many species are projected to shift a hundred kilometres or more in the next century[5,6]. In the process, many animals will bring their parasites and pathogens into new environments. This poses a measurable threat to global health, particularly given several recent epidemics and pandemics of viruses that originate in wildlife (zoonotic viruses, or zoonoses)[7,8]. Most frameworks for predicting cross-species transmission therefore focus on the steps that enable animal pathogens to make the leap to human hosts (a process called spillover)[8,9]. However, zoonotic viruses make up a small fraction of total viral diversity, and viral evolution is an undirected process, in which humans are only one of at least around 6,500 mammalian hosts with over 21 million possible pairwise combinations (not including the other four classes of vertebrates, which have a much greater fraction of undescribed viral diversity). If those host species track shifting climates, they will share viruses not only with humans, but also with each other, for the very first time[3,4]. Despite their importance, spillover events are probably just the tip of the iceberg; by numbers alone, most cross-species transmission events that are attributable to climate change will probably occur among wildlife hosts, potentially threatening wildlife populations and largely undetected by zoonotic disease surveillance.

The scale of this process will depend on opportunity and compatibility[10,11], and both dimensions pose an important predictive challenge. As only a few species are common worldwide, most hosts have no opportunity to exchange pathogens: of all possible pairs of mammal species, only around 7% share any geographical range, and only about 6% are currently known to host one or more of the same virus species (hereafter, viral sharing)[10]. As host geographical ranges shift, some interactions will become possible for the first time, and a subset will lead to viral establishment in a previously inaccessible host (novel viral sharing). The potential ability of species to track shifting climate and habitat conditions will determine which pairs of species encounter each other for the first time. Once the ranges of species nominally overlap, habitat selection and behavioural differences can further limit contact. Although some viruses spread environmentally or by arthropod vectors between spatially proximate species with no direct behavioural contact, sharing is more likely on average among species with more ecological overlap. Even among species in close contact, most cross-species transmission events are still a dead end. Progressively smaller subsets of viruses can infect the cells of a new host, proliferate, cause disease and transmit onwards in a new host. Their ability to do so is determined by the compatibility between viral structures, host cell receptors and host immunity. As closely related species share both ecological and immunological traits through identity by descent, phylogeny is a strong

[1]Department of Biology, Georgetown University, Washington, DC, USA. [2]Center for Global Health Science & Security, Georgetown University, Washington, DC, USA. [3]EcoHealth Alliance, New York, NY, USA. [4]Eversource Energy Center, University of Connecticut, Storrs, CT, USA. [5]African Climate and Development Initiative, University of Cape Town, Cape Town, South Africa. [6]Department of Biology, Pacific Lutheran University, Tacoma, WA, USA. [7]These authors contributed equally: Colin J. Carlson, Gregory F. Albery. ✉e-mail: colin.carlson@georgetown.edu; gfalbery@gmail.com

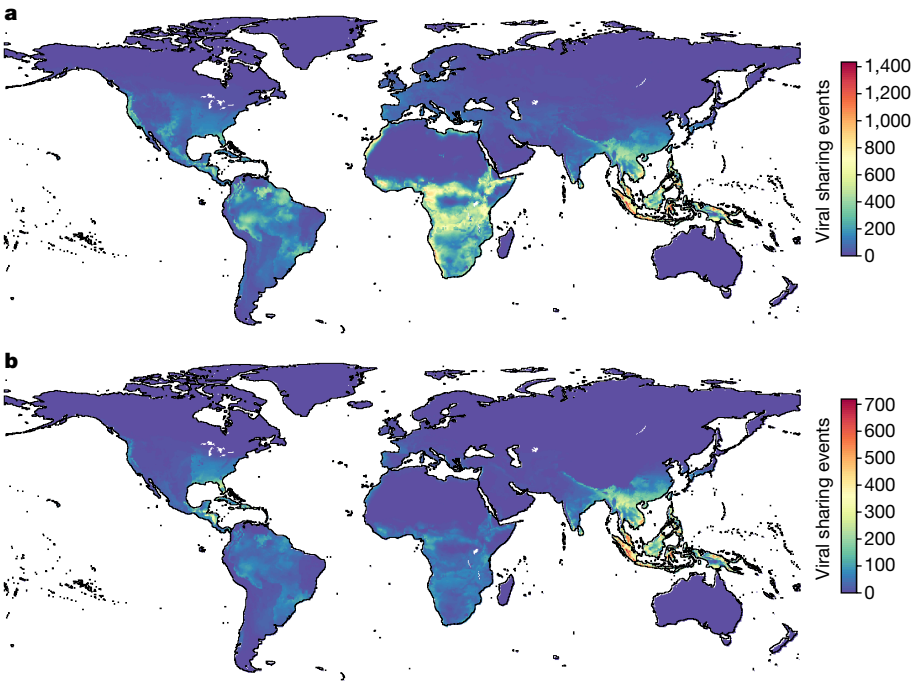


Fig. 1 | Climate change will drive novel viral sharing among mammal species. a,**b**, The projected number of novel viral sharing events among mammal species in 2070 based on geographical range shifts of the 3139 host species from changes in climate and land use (SSP 1–RCP 2.6) without dispersal limits (**a**) and with dispersal limitation (**b**). Results are averaged across nine GCMs.

predictor of pathogen sharing and of susceptibility to invasion by new viruses[11]. In a changing world, these mechanisms can help to predict how ecosystem turnover could affect the global virome.

Although several studies have mapped current hotspots of emerging diseases[2,7,12], few have forecasted them in the context of global change. With the global reassortment of animal biodiversity due to changes in climate and land use, it is unknown whether bats and rodents will still have a central role in viral emergence[2] (Extended Data Fig. 1) or whether hotspots of viral emergence will stay in tropical rainforests[12], which currently contain most of the undiscovered viruses[2]. Here, by projecting newly suitable habitat (which a species may or may not migrate to) and applying mechanistic biological rules for cross-species transmission, we predicted how and where global change could potentially create new opportunities for viral sharing, with particular attention to the potential connections between these risks and human health. We focused on mammals because they have some of the most complete biodiversity data, the highest proportion of viral diversity described[1], and the greatest downstream relevance to human health and zoonotic disease emergence of any vertebrate class. We built species distribution models (SDMs) for 3,870 placental mammal species, and projected potential shifts in geographical range on the basis of four paired scenarios for changes in climate (Representative Concentration Pathways (RCPs)) and land use (Shared Socioeconomic Pathways (SSPs)) by 2070. These scenarios characterize alternative futures for global environmental change, from sustainable land-use change and a high chance of keeping global warming under 2 °C (SSP 1–RCP 2.6), to a high chance of more than 4 °C warming, continued fossil fuel reliance, and rapid land degradation and change (SSP 5–RCP 8.5; a detailed explanation is provided in the Methods). We present the results for SSP 1–RCP 2.6 in the main text because this scenario is most in line with the goals of the Paris Agreement to keep global warming well below 2 °C (ref. [13]). We quantified model uncertainty in projected climate futures using nine global climate models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6). As many species are unlikely to be biologically suited for rapid range shifts and will therefore move slower than the local velocity of climate change, we constrained the speed of range shifts on the basis of the inferred allometric scaling of animal movement[14], and compared scenarios that assumed limited dispersal against full dispersal (that is, no dispersal limitation).

We used projections of newly suitable habitat to identify where new range overlap among currently non-overlapping species could happen in the future (hereafter, first encounters). We then used a recently developed viral sharing model to predict the probability of a novel viral sharing event—here defined as the future cross-species transmission of at least one virus species, in this case between a pair of hosts during first encounters—on the basis of new geographical overlap and host phylogenetic similarity[10], a first-order approximation of opportunity and compatibility (Extended Data Fig. 2). This model framework has previously provided insights into viral macroecology and zoonotic risk based on the ~1% of the global mammalian virome that has been described[1,2,10]. On the basis of the total number and distribution of first encounters among a subset of 3,139 species (Methods), we used cumulative viral sharing probabilities to estimate the total number of new sharing events that are expected (each of which describes the cross-species transmission of at least one virus). Using this approach, we tested the hypothesis that environmental change should alter mammal communities in ways that expose hosts to new viruses, altering the structure of the whole mammal–virus network.

The effects of changes in climate and land use

If shifts in species range can keep pace with climate change, we predict that the vast majority of mammal species will overlap with at least one unfamiliar species somewhere in their potential future range, regardless of emissions scenario (mean ± s.d. across GCMs here and after; RCP 2.6: 98.6 ± 0.2%; RCP 8.5: 96.6 ± 0.8%). At the global level, geographical range shifts would permit over 300,000 first encounters in every climate scenario (SSP 1–RCP 2.6: 316,426 ± 1,719; SSP 5–RCP 8.5: 313,973 ± 2,094; Fig. 1 and Extended Data Fig. 3). Compared with a present-day baseline, in which we calculated 345,850 current

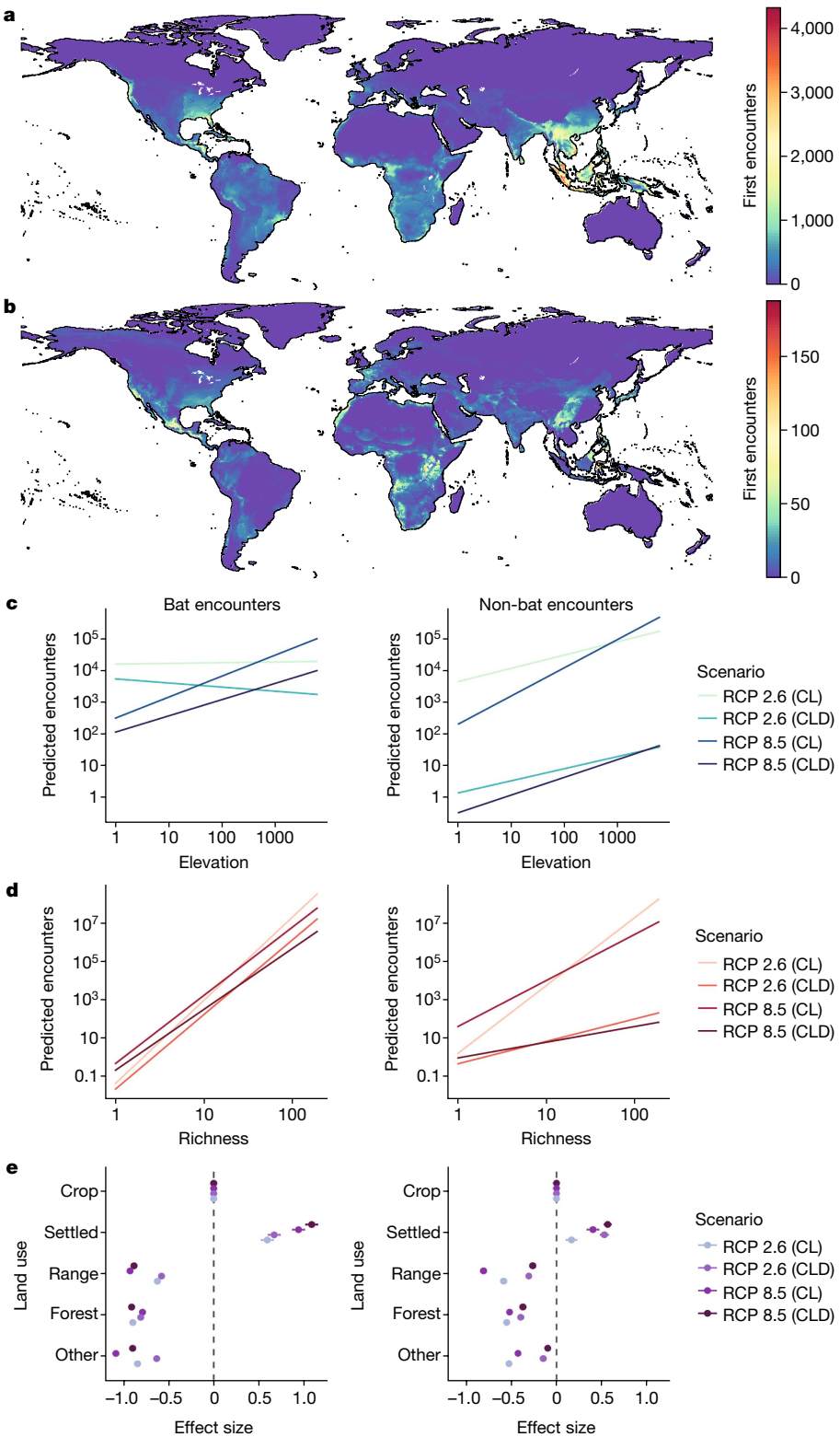


Fig. 2 | Bats disproportionately drive future novel viral sharing. **a**,**b**, The spatial pattern of first encounters (in SSP 1–RCP 2.6) differs among range-shifting mammal pairs, including bat–bat and bat–non-bat encounters (**a**) and encounters among only non-bats (**b**). **c**–**e**, Using a linear model, we show that elevation (**c**), species richness (**d**) and land use (**e**) influence the number of new overlaps for bats and non-bats across scenarios (RCPs paired with SSPs as described in the Methods). $n = 9$ GCM replicates. Slopes for the elevation effect were generally steeply positive: a $\log_{10}$-transformed increase in elevation was associated with a 0.4–1.41 $\log_{10}$ increase in first encounters. The results were averaged across nine GCMs. CL, climate and land use change; CLD, climate and land use change with dispersal limits. The error bars in **e** are the s.e. of the model estimate.

pairwise overlaps among the 3,870 species (~7%), this essentially represents a doubling of potential species contact. These first encounters between mammal species will occur everywhere in the world, but are concentrated in tropical Africa and southeast Asia (Extended Data Fig. 4). This result was counter to expectations that species might aggregate at higher latitudes, given that previous research has anticipated a link

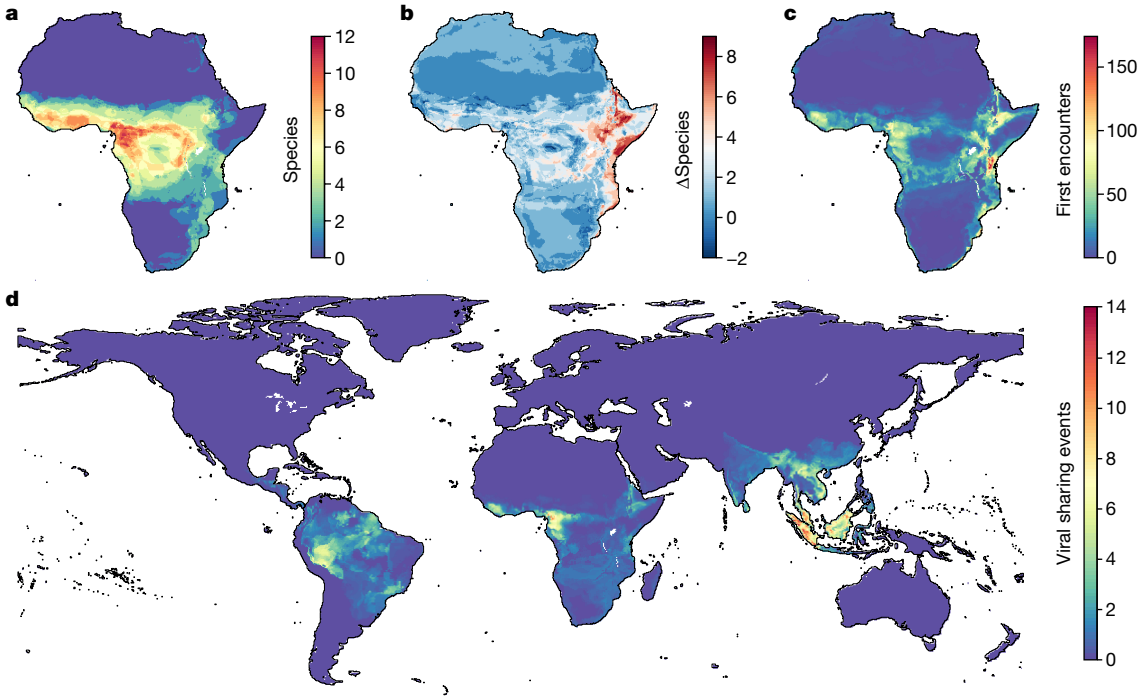


Fig. 3 | Range expansions will expose naive hosts to zoonotic reservoirs. **a**, The predicted distribution of known African hosts of ZEBOV. **b**, The change in richness of these hosts as a result of range shifts (SSP 1–RCP 2.6). **c**, Projected first encounters with non-Ebola hosts. **d**, Bat–primate first encounters are projected to occur globally, producing new sharing events. The results were averaged across nine GCMs.

between climate change, range shifts and parasite host switching in the Arctic[15,16]. However, we found that, when species shift along latitudinal gradients, they travel in the same direction as others that are already included in their assemblage, leading to few first encounters. By contrast, when species track thermal optima along elevational gradients (allowing them to come from different directions, that is, mountains force species to cluster), they will aggregate in the most new combinations in mountain ranges, especially in tropical areas with the highest baseline diversity, matching prior predictions[17]. This pattern was robust to climate model uncertainty (Supplementary Figs. 3–11) and to differences in dispersal ability (Fig. 2c). The most notable model variation is in the Amazon basin, as well as a small portion of the central African basin, Botswana and parts of the Indian subcontinent (Extended Data Fig. 5). These areas become essentially devoid of first encounters in the most sensitive climate models and warmest pathways, presumably because all are high-endemism basins of homogenous climate that may warm too much for species to escape into high-elevation refugia (a fairly well-documented pattern[18–20]).

This global reorganization of mammal assemblages is projected to substantially affect the structure of the mammalian virome. Accounting for geographical opportunity and phylogenetic compatibility, we project that a total of 316,426 (±1,719) first encounters in RCP 2.6 would lead to 15,311 new sharing events (±140)—that is, a minimum of at least ~15,000 cross-species transmission events of at least one new virus (but potentially many more) between a pair of naive host species. Assuming that viral sharing will initially be localized to areas of new host overlap, we mapped expected viral sharing events, and found again that most sharing should occur in high-elevation, species-rich ecosystems in Africa and Asia (Fig. 1a). If species survive a changing climate by aggregating in high-elevation refugia, this suggests that emerging viruses may be an increasing problem for their conservation[21,22]. Across scenarios, the spatial pattern of expected sharing events was nearly identical, and was dominated more by the extent of potential range shifts than by underlying community phylogenetic structure (Extended Data Fig. 6 and Supplementary Figs. 12–20). Although previous research has suggested that the phylogenetic structure of mammal communities might drive continental hotspots of pathogen sharing and emergence[23], in our framework, opportunity drives spatial patterns more than compatibility. Given that phylogeny is a strong determinant of viral sharing in the underlying model, this difference from previous studies can probably be explained by evolutionary scale, where previous work focused on primates, and our study includes all mammals. At this broader scale, predicted viral sharing patterns mostly track total richness (see Fig. 3b of ref. [10]) and, at finer scales, phylogeny has a stronger effect.

Dispersal drives the importance of bats

The intrinsic dispersal ability of species is likely to constrain their ability to move to newly suitable locations, and therefore to limit novel viral sharing (as are harder to predict extrinsic factors, such as landscape connectivity or facilitated migration by conservation efforts). We limited the dispersal potential of flightless species based on an established allometric scaling with body size, trophic rank and generation time[14]. Dispersal limits caused substantial decreases in predicted potential range expansions across all scenarios, especially for higher warming scenarios and therefore drove a reduction in first encounters and novel viral sharing. Even in RCP 2.6 (the scenario with the least warming), limiting dispersal reduced the number of first encounters by 61% (±0.3%) and reduced the associated viral sharing events by 70% (±0.1%) to 4,584 (±52) projected viral sharing events. As trophic position and body size determine dispersal ability, carnivores account for a slightly disproportionate number of first encounters, whereas ungulates and rodents have slightly fewer first encounters than expected at random (Extended Data Fig. 7). Spatial patterns also changed considerably when dispersal constraints were added, with the majority of first encounters and cross-species viral transmission events occurring in southeast Asia (Fig. 1b and Extended Data Figs. 4 and 6). This viral sharing hotspot is driven disproportionately by bats, because their dispersal was left unconstrained within continents; we

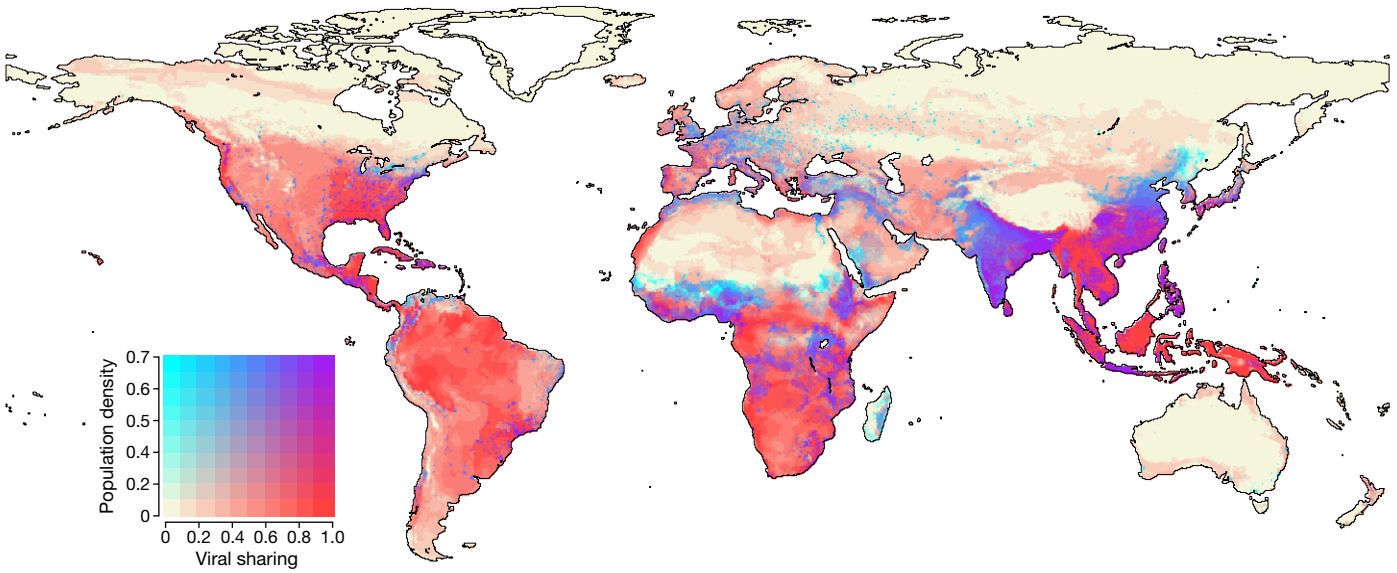


Fig. 4 | Novel viral sharing events coincide with human population centres. In 2070 (SSP 1–RCP 2.6; climate only), human population centres in equatorial Africa, south China, India and southeast Asia will overlap with projected hotspots of cross-species viral transmission in wildlife. Both variables were linearly rescaled to 0 to 1. The results were averaged across nine GCMs.

made this choice given their exclusion from previous research characterizing the dispersal ability of range-shifting mammals[14], genetic evidence that flight enables bats—and their viruses—to often circulate at continental levels[24,25] and data suggesting that bat distributions are already undergoing disproportionately rapid shifts[26–34]. Bats account for nearly 90% of first encounters after constraining dispersal in any climate scenario (RCP 2.6: 88% ± 0.1%; RCP 8.5: 89% ± 0.5%) and dominate the spatial pattern, with most of their first encounters restricted to southeast Asia (Fig. 2).

The unique ability of bats for flight could be an important and previously unconsidered link between climate-driven range shifts and future changes in the mammalian virome. Even non-migratory bats can regularly travel hundreds of kilometres within a lifetime, far exceeding what small mammals might be able to cover in 50 years; half of all bat population genetic studies have failed to find any evidence for isolation by distance[35]. This unique dispersal ability has inevitable epidemiological implications, with recent evidence suggesting that continental panmixia may be common for zoonotic reservoirs, enabling viral circulation at comparable scales[24,25,36]. Several studies have also identified ongoing rapid range expansions in bat species around the world[26–34], with little mention in the broader climate change or emerging disease literature. If flight does enable bats to undergo more rapid range shifts than other mammals, we expect that they should drive the majority of new cross-species viral transmission, and probably bring zoonotic viruses into new regions. However, the ability of bats to move rapidly might be attenuated by the other biotic constraints to species distributions (for example, social behaviour and food availability, which are unaddressed by the current approach). This uncertainty adds an important new dimension to the ongoing debate about whether bats are unique in their higher viral richness, higher proportion of zoonotic viruses or immune adaptations compared with other mammals[2,11].

Zoonotic emergence and human health

The effects of climate change on mammalian viral sharing patterns are likely to cascade in the future emergence of zoonotic viruses. Among the thousands of expected viral sharing events, some of the highest-risk zoonoses or potential zoonoses are likely to find new hosts. This may eventually pose a threat to human health—the same general rules for cross-species transmission explain spillover patterns for emerging zoonoses[11], and the viral species that make successful jumps across wildlife species have the highest propensity for zoonotic emergence[2]. Just as the simian immunodeficiency virus making a host jump from monkeys to chimpanzees and gorillas facilitated the origins of HIV, or SARS-CoV spillover into civets enabled a bat virus to reach humans, these kinds of wildlife-to-wildlife host jumps may be evolutionary stepping stones for the approximately 10,000 potentially zoonotic viruses that are currently circulating in mammalian hosts[1].

To illustrate this problem at the scale of a single pathogen's 'sharing network' (the set of all hosts that are known or suspected to host the virus, and likely to share with those known hosts), we constructed a subnetwork of 13 possible hosts of the Zaire ebolavirus (ZEBOV) in Africa, and projected possible first encounters involving these species (Fig. 3a–c and Extended Data Fig. 8). We project these 13 species to encounter 3,695 (±49) new mammals in RCP 2.6, with a modest reduction to 2,627 (±44) species when accounting for dispersal limitation, and little variation among climate scenarios (RCP 8.5: 3,529 ± 47 encounters without dispersal limits; 2,455 ± 88 with dispersal limits). Even with dispersal limits, these first encounters are predicted to produce almost one hundred new viral sharing events (RCP 2.6: 96 ± 2; RCP 8.5: 86 ± 4) that might include ZEBOV and which cover a much broader part of Africa than the current zoonotic niche of Ebola[37]. Human spillover risk aside, this could expose several new wildlife species to a deadly virus that is historically responsible for sizable primate die-offs[38]. Moreover, for zoonoses like ZEBOV without known reservoirs, future host jumps—and, therefore, the emergence of a larger pool of potential reservoirs covering a greater geographical area (such as the potential introduction of ZEBOV to east African mammals)—would only complicate ongoing efforts to trace the sources of spillover and anticipate future emergence. Ebola is far from unique, with 8,429 ± 228 first encounters in RCP 2.6 between bats and primates, leading to an expected 110 ± 4 new viral sharing events even with dispersal limits (Fig. 3d; RCP 8.5: 7,326 ± 667 first encounters, 90 ± 8 sharing events)—many potential zoonoses are likely to experience new evolutionary opportunities due to climate change.

Future hotspots of new mammal assemblages and viral evolution are projected to coincide with areas of high human population density, further increasing vulnerability to potential zoonoses. Potential first encounters are disproportionately likely to occur in areas that

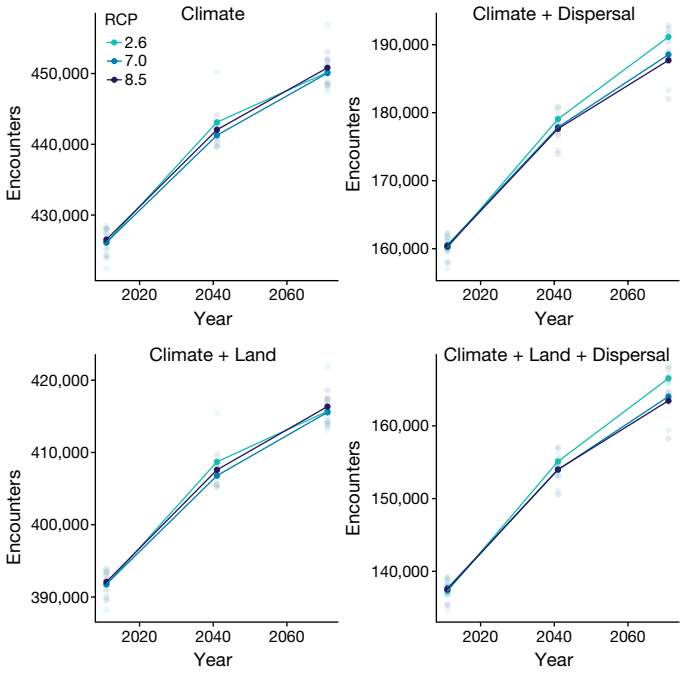


Fig. 5 | Projected timing of first encounters. By tracking the identity of each species pair, we simulated the first encounters as cumulative (the same pair cannot meet for the first time at two timepoints). The results are broken down by whether each climate (C) scenario includes land-use change (L) and dispersal limits (D). The light points show 1 out of 15 climate scenarios (5 GCMs with 3 RCPs each); and the dark points are the mean values for each RCP. Points are plotted at the start of each 30 year window.

are projected to be either human settled or used as cropland and less likely to occur in forests (Fig. 2e), despite current literature suggesting that forests contain most emerging and undiscovered viruses[12] (Fig. 4). This finding is consistent for bats and non-bats, and may be an accident of geography, but more likely represents the tendency of human settlements to aggregate on continental edges and around biodiversity hotspots[39]. Regardless of the mechanism, we predict that tropical hotspots of novel viral sharing will broadly coincide with areas of high population density in 2070, especially in the Sahel, the Ethiopian highlands and the Rift Valley, India, eastern China, Indonesia and the Philippines (Fig. 4). Some European population centres also land in these hotspots; recently emergent pathogens in this region like the Usutu virus highlight that these populations can still be vulnerable, despite greater surveillance and healthcare access. If range-shifting mammals create ecological release for undiscovered zoonoses, populations in any of these areas are likely to be vulnerable, and some viruses will be able to spread globally from any of these population centres.

The effects of climate-change mitigation

Whereas most studies agree that climate-change mitigation through reducing greenhouse gas emissions will prevent extinctions and minimize harmful ecosystem effects[40], our results suggest that mitigation alone cannot reduce the likelihood of climate-driven viral sharing. Instead, the mildest scenarios for global warming appear likely to produce at least as much or even more cross-species viral transmission: when warming is slower, species can successfully track shifting climate optima, leading to more potential for range expansion and more first encounters. Accounting for dispersal limits, species are projected to experience a median potential loss (± SD) of 0.3% (±2.5%) of their range in RCP 2.6, with 49.8% (±3.8%) experiencing a net potential increase in range; by contrast, species were predicted to experience a 26.2% (±13.2%) median potential loss in RCP 8.5, and only 30.8% (±5.45%) potentially gained any range (Extended Data Fig. 3a). In fact, in RCP 8.5, we projected that 261 (±76) species could lose their entire range, with 162 (±53) attributable to dispersal limits alone. As a result, there were 5.4% (±1.7%) fewer potential first encounters in RCP 8.5 compared with in RCP 2.6 and, unexpectedly, a 1.9% (±0.3%) predicted reduction in the connectivity of the future global viral sharing network (Extended Data Fig. 3b, d). Overall, our results indicate that a mild perturbation of the climate system could create thousands of new opportunities for viruses to find new hosts.

We caution that these results should not be interpreted as a justification for inaction, or as a possible upside to unmitigated warming, which will be accompanied by mass defaunation, devastating disease emergence, and unprecedented levels of human displacement and global instability[40]. Rather, our results highlight the urgency of better wildlife disease surveillance systems and public health infrastructure as a form of climate-change adaptation, even if mitigation efforts are successful and global warming stays below +2 °C above preindustrial levels.

The timing of ecological opportunity

As a final analysis, we examined the potential timing of climate-change impacts. We expected that most first encounters would occur later in the twenty-first century, given the time required for species' habitats to shift (especially with dispersal constraints). To test this hypothesis, we reproduced our analysis with an entirely new climate product (CHELSA v2.1; Methods) that enabled us to set a baseline in the recent past (1981–2010) and examine three time intervals of future impacts (2011–2040, 2041–2070 and 2071–2100). Projecting species distributions under the mean conditions in each interval, we identified the cumulative number of unique first encounters in each. Notably, we found that the majority of first encounters occur by the 2011–2040 period (Fig. 5), although steady and sizeable increases continue through the rest of the century. Differences in the number of first encounters between time intervals (for example, RCP 2.6, 2011–2040, climate + land use: 391,716 ± 2,322; versus RCP 2.6, 2071–2100, climate + land use: 415,703 ± 3,537) are substantially greater than differences driven by climate scenarios and climate model uncertainty (for example, climate + land use, 2071–2100: lowest GCM–RCP combination: 413,120 in RCP 2.6, GCM GFDL-ESM4; versus highest GCM–RCP combination: 421,917 in RCP 2.6, GCM IPSL-CM6ALR), indicating that first encounters continue to non-trivially accumulate over time. However, those differences are still substantially less than the difference created by dispersal assumptions (for example, RCP 2.6, 2071–2100, climate + land use: 415,703 ± 3,537; versus RCP 2.6, 2071–2100, climate + land use + dispersal: 166,527 ± 1,905), and the proportion of future first encounters is much higher when dispersal maxima are restricted (Fig. 5). Put more succinctly, species continue to meet throughout the century, and our simulations indicate that how fast species move will matter more to the timing and magnitude of first encounters than how fast suitable habitat moves or has already moved. Moreover, the geography of first encounters remains consistent across all points in time (Extended Data Figs. 9 and 10).

Overall, these findings suggest that, in a world that has already passed global warming of +1 °C, the majority of climate-related opportunities for novel viral sharing may already have been realized—if and only if species' dispersal has kept pace with shifting habitat suitability. That premise—and, particularly, our simulation of bats as unconstrained by dispersal limits—is more tenuous over smaller timescales; research is urgently needed that estimates the real-time signal of climate-attributable range shifts. Even if these opportunities exist, the timing of cross-species transmission itself remains uncertain and unpredictable; our viral sharing model is

trained on an equilibrium level of connectivity, and we expect some degree of lag between interspecific contact and viral establishment. It is certainly possible or even likely that climate change is already reshaping the mammalian virome, and as warming continues over the next half-century, we predict that both the opportunities created for ecological novelty and the resulting effects on viral assemblages will begin to saturate.

Conclusions

Our study establishes a macroecological link between climate change and cross-species viral transmission. The patterns that we describe are probably further complicated by uncertainties in the species distribution modelling process, including local adaptation or plasticity in response to changing climates, or lack of landscape connectivity preventing dispersal. The projections that we make are also likely to be complicated by several ecological factors, including the temperature sensitivity of viral host jumps[41]; potential independence of vector or non-mammal reservoir range shifts; or the possibility that defaunation especially at low elevations might interact with disease prevalence through biodiversity dilution and amplification effects not captured by our models[42]. Future work can expand the scope of our findings to other host–parasite systems[4]. Our approach, which combines viral sharing models with species distribution modelling approaches for thousands of species, is readily applied to other datasets. Birds have the best documented virome after mammals, and account for the majority of non-mammalian reservoirs of zoonotic viruses[43]; changing bird migration patterns in a warming world may be especially important targets for prediction. Similarly, with amphibians facing disproportionately high extinction rates due to a global fungal panzootic, and emerging threats such as ranavirus causing conservation concern, pathogen exchange among amphibians may be especially important for conservation practitioners to understand[44]. Finally, marine mammals are an important target given their exclusion here, especially after a recent study implicating reduced Arctic sea ice in novel viral transmission between pinnipeds and sea otters—a result that may be the first proof of concept for our proposed climate–disease link[45].

Our study provides a template for how surveillance could target future hotspots of viral emergence in wildlife. In the next decade alone, it could cost at least a billion dollars to comprehensively identify and counteract zoonotic threats before they spread from wildlife reservoirs into human populations[46]. These efforts will be undertaken during the greatest period of global ecological change recorded in human history and, in a practical sense, the rapid movement of species (and their viromes) poses an unexpected challenge for virological research. Although several studies have addressed how range shifts in zoonotic reservoirs might expose humans to novel viruses, few have considered the fact that most new exposures will be among wildlife species. The relevance of this process is reinforced by the COVID-19 pandemic, which began only weeks after the completion of this study; the progenitor of SARS-CoV-2 probably originated in southeast Asian horseshoe bats (*Rhinolophus* sp.), and may have spread to humans through an as-yet-unknown bridge host[47]. Although we caution against overinterpreting our results as explanatory of the current pandemic[48], our findings suggest that climate change could easily become the dominant anthropogenic force in viral cross-species transmission, which will undoubtedly have a downstream effect on human health and pandemic risk. Tracking viral spillover into humans is paramount, but so is the monitoring of viral transmission among wildlife species. Targeting surveillance in future hotspots of cross-species transmission, such as southeast Asia, and developing norms of open data sharing for the global scientific community, will help researchers to identify host jumps early on, ultimately improving our ability to respond to potential threats.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04788-w.

1. Carlson, C. J., Zipfel, C. M., Garnier, R. & Bansal, S. Global estimates of mammalian viral biodiversity accounting for host sharing. *Nat. Ecol. Evol.* **3**, 1070–1075 (2019).
2. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
3. Hoberg, E. P. & Brooks, D. R. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philos. Trans. R. Soc. B* **370**, 20130553 (2015).
4. Morales-Castilla, I. et al. Forecasting parasite sharing under climate change. *Philos. Trans. R. Soc. B* **376**, 20200360 (2021).
5. Burrows, M. T. et al. Geographical limits to species-range shifts are suggested by climate velocity. *Nature* **507**, 492–495 (2014).
6. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026 (2011).
7. Jones, K. E. et al. Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
8. Plowright, R. K. et al. Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
9. Lloyd-Smith, J. O. et al. Epidemic dynamics at the human–animal interface. *Science* **326**, 1362–1367 (2009).
10. Albery, G. F., Eskew, E. A., Ross, N. & Olival, K. J. Predicting the global mammalian viral sharing network using phylogeography. *Nat. Commun.* **11**, 2260 (2020).
11. Albery, G. F. et al. The science of the host–virus network. *Nat. Microbiol.* **6**, 1483–1492 (2021).
12. Allen, T. et al. Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124 (2017).
13. O'Neill, B. C. et al. The scenario model intercomparison project (scenariomip) for cmip6. *Geosci. Model Dev.* **9**, 3461–3482 (2016).
14. Schloss, C. A., Nuñez, T. A. & Lawler, J. J. Dispersal will limit ability of mammals to track climate change in the western hemisphere. *Proc. Natl Acad. Sci. USA* **109**, 8606–8611 (2012).
15. Davidson, R. et al. Arctic parasitology: why should we care? *Trends Parasitol.* **27**, 239–245 (2011).
16. Hoberg, E. P. et al. Arctic systems in the quaternary: ecological collision, faunal mosaics and the consequences of a wobbling climate. *J. Helminthol.* **91**, 409–421 (2017).
17. Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C. & Longino, J. T. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322**, 258–261 (2008).
18. Sales, L., Ribeiro, B. R., Chapman, C. A. & Loyola, R. Multiple dimensions of climate change on the distribution of amazon primates. *Perspect. Ecol. Conserv.* **18**, 83–90 (2020).
19. Trisos, C. H., Merow, C. & Pigot, A. L. The projected timing of abrupt ecological disruption from climate change. *Nature* **580**, 496–501 (2020).
20. Newbold, T. Future effects of climate and land-use change on terrestrial vertebrate community diversity under different scenarios. *Proc. R. Soc. B* **285**, 20180792 (2018).
21. Pauchard, A. et al. Non-native and native organisms moving into high elevation and high latitude ecosystems in an era of climate change: new challenges for ecology and conservation. *Biol. Invasions* **18**, 345–353 (2016).
22. Atkinson, C. T. & LaPointe, D. A. Introduced avian diseases, climate change, and the future of hawaiian honeycreepers. *J. Avian Med. Surg.* **23**, 53–63 (2009).
23. Pedersen, A. B. & Davies, T. J. Cross-species pathogen transmission and disease emergence in primates. *EcoHealth* **6**, 496–508 (2009).
24. Peel, A. J. et al. Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses. *Nat. Commun.* **4**, 2770 (2013).
25. Riesle-Sbarbaro, S. A. et al. The Gambian epauletted fruit bat shows increased genetic divergence in the Ethiopian highlands and in an area of rapid urbanization. *Ecol. Evol.* **8**, 12803–12820 (2018).
26. Wu, J. Detection and attribution of the effects of climate change on bat distributions over the last 50 years. *Clim. Change* **134**, 681–696 (2016).
27. Ancillotto, L., Santini, L., Ranc, N., Maiorano, L. & Russo, D. Extraordinary range expansion in a common bat: the potential roles of climate change and urbanisation. *Sci. Nat.* **103**, 15 (2016).
28. Ancillotto, L. et al. What is driving range expansion in a common bat? Hints from thermoregulation and habitat selection. *Behav. Processes* **157**, 540–546 (2018).
29. Geluso, K., Mollhagen, T. R., Tigner, J. M. & Bogan, M. A. Westward expansion of the eastern pipistrelle (*Pipistrellus subflavus*) in the United States, including new records from New Mexico, South Dakota, and Texas. *West. N. Am. Naturalist* **65**, 405–409 (2005).
30. Kurta, A., Winhold, L., Whitaker, J. O. & Foster, R. Range expansion and changing abundance of the eastern pipistrelle (Chiroptera: Vespertilionidae) in the central Great Lakes region. *Am. Midland Naturalist* **157**, 404–412 (2007).
31. Lundy, M., Montgomery, I. & Russ, J. Climate change-linked range expansion of Nathusius' pipistrelle bat, *Pipistrellus nathusii* (Keyserling & Blasius, 1839). *J. Biogeogr.* **37**, 2232–2242 (2010).
32. McCracken, G. F. et al. Rapid range expansion of the Brazilian free-tailed bat in the southeastern United States, 2008–2016. *J. Mammal.* **99**, 312–320 (2018).
33. Roberts, B. J., Catterall, C. P., Eby, P. & Kanowski, J. Latitudinal range shifts in Australian flying-foxes: a re-evaluation. *Austral Ecol.* **37**, 12–22 (2012).

34. Uhrin, M. et al. Status of Savi's pipistrelle *Hypsugo savii* (Chiroptera) and range expansion in Central and south-eastern Europe: a review. *Mammal Rev.* **46**, 1–16 (2016).
35. Olival, K. in *Evolutionary History of Bats: Fossils, Molecules and Morphology* (eds Gunnell, G. F. & Simmons, N. B.) 267–316 (Cambridge Univ. Press, 2012).
36. Olival, K. J. et al. Population genetics of fruit bat reservoir informs the dynamics, distribution and diversity of Nipah virus. *Mol. Ecol.* **29**, 970–985 (2019).
37. Pigott, D. M. et al. Updates to the zoonotic niche map of Ebola virus disease in Africa. *eLife* **5**, e16412 (2016).
38. Bermejo, M. et al. Ebola outbreak killed 5000 gorillas. *Science* **314**, 1564–1564 (2006).
39. Williams, J. N. Humans and biodiversity: population and demographic trends in the hotspots. *Popul. Environ.* **34**, 510–523 (2013).
40. IPCC *Climate Change 2022: Impacts, Adaptation, and Vulnerability* (eds. Pörtner, H.-O. et al.) (Cambridge Univ. Press, 2022).
41. Roberts, K. E., Hadfield, J. D., Sharma, M. D. & Longdon, B. Changes in temperature alter the potential outcomes of virus host shifts. *PLoS Pathog.* **14**, e1007185 (2018).
42. Faust, C. L. et al. Null expectations for disease dynamics in shrinking habitat: dilution or amplification? *Philos. Trans. R. Soc. B* **372**, 20160173 (2017).
43. Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl Acad. Sci. USA* **117**, 9423–9430 (2020).
44. Cunningham, A. Infectious disease threats to amphibian conservation. In *Proc. The Amphibians and Reptiles of Scotland.* (eds McInerny, C. J. & Wilkie, I.) 81–90 (Glasgow Natural History Society, Glasgow, 2018).
45. VanWormer, E. et al. Viral emergence in marine mammals in the North Pacific may be linked to Arctic sea ice reduction. *Sci. Rep.* **9**, 15569 (2019).
46. Carroll, D. et al. The global virome project. *Science* **359**, 872–874 (2018).
47. Becker, D. J. et al. Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe* https://doi.org/10.1016/S2666-5247(21)00245-7 (2022).
48. Beyer, R. & Manica, A. Range sizes of the world's mammals, birds and amphibians from 10,000 BC to 2100 AD. *Nat. Commun.* **11**, 5633 (2020).

Methods

Here we developed global maps for terrestrial mammals characterizing their habitat use and their ecological niche as a function of climate. We projected these into paired climate–land use futures for 2070, with dispersal limitations set by biological constraints for each species. For a final subset of 3,139 species, we predicted the probability of viral sharing among species pairs using a model of the mammalian viral sharing network that is trained on phylogenetic relatedness and current geographical range overlaps. With that model, we mapped the projected hotspots of new viral sharing in different futures. All analyses were conducted in R (v.4.1.3). Analysis and visualization code is available at GitHub (https://github.com/viralemergence/iceberg).

Data

Mammal virus data. Our understanding of viral sharing patterns is based on a previously published dataset[49]. The dataset describes 2,805 known associations between 754 species of mammalian host and 586 species of virus, scraped from the taxonomic data stored in the International Committee on Taxonomy of Viruses (ICTV) database. These data have previously been used in several studies modelling global viral diversity in wildlife[1,50], including a previous study that developed the model of viral sharing we use here[10]. As that model is reproduced exactly in our study, we have made no further modifications to the data, and more detailed information on data management (such as the exclusion of *Homo sapiens* from that analysis) can be found in a previous publication[10].

Biodiversity data. We downloaded Global Biodiversity Informatics Facility (GBIF: https://www.gbif.org/) occurrence records for all mammals based on taxonomic names resolved by the IUCN Red List. Records spanned the years 1658 to 2019 and were intended to provide a comprehensive baseline of species distributional data for as many mammal species as possible. These records were filtered to those within a 10 km buffer of species' IUCN range maps, and further filtered using the Grubbs outlier test (see below); this procedure is likely to eliminate most signal of climate-driven geographical range shifts from recent sightings. If any of that signal persists after these steps, there are two possible effects on models: extending the geographical extent attributed to a species' current range and, potentially, expanding the climatic envelope that a model identifies as suitable for its presence. Both of these would increase the predicted rate of overlap among species pairs in their current (modelled) distributions, and reduce the predicted number of first encounters, making our predictions more conservative.

We developed SDMs for all 3,870 species with at least three unique terrestrial presence records at a 0.25° spatial resolution (approximately 25 km by 25 km at the equator). To focus on species occurrence, we retained one unique point per 0.25° grid cell. This spatial resolution was chosen to match the available resolution of land-use change projections (see below). Spatial and environmental outliers were removed on the basis of Grubbs outlier tests[51]. To implement the Grubbs outlier tests for a given species, we defined a distance matrix between each record and the centroid of all records (in both environmental or geographical space, respectively) and determined whether the record with the largest distance was an outlier with respect to all other distances, at a given statistical significance ($P = 1 \times 10^{-3}$, to exclude only extreme outliers). If an outlier was detected, it was removed and the test was repeated until no additional outliers were detected.

Climate and land-use data. We compiled climate and land-use data from WorldClim 2[52] and the Land Use Harmonization 2 (LUH2) project[53], respectively, for both baseline conditions (operationalized as 1970–2000 for the climate data, 2015 for land use and 2020 for dispersal limits; see 'The effect of recent warming' for an interrogation of the difference between climate baselines and actual present-day climate) and a half-century in the future (operationalized as 2061–2080 for climate, 2070 for land use, and 2070 for dispersal).

The WorldClim dataset is widely used in ecology, biodiversity and agricultural projections of potential climate-change impacts. WorldClim makes data available for current and future climates in the form of 19 preprocessed bioclimatic variables (Bioclim: BIO1–19). To reduce collinearity among climate variables in the SDMs, we selected five Bioclim variables from the full set of 19 Bioclim variables: mean annual temperature (BIO1), temperature seasonality (BIO4), annual precipitation (BIO12), precipitation seasonality (coefficient of variation; BIO15) and precipitation of the driest quarter (BIO17). This is the largest set of Bioclim variables possible that keeps their correlation over a global extent suitably low ($r < 0.7$). The Bioclim variables for the historical climate are the mean from 1970–2000, and those for the future climate are the mean from 2060–2080.

To account for model uncertainty in climate projections, we used projections for future climates from all nine GCMs that are currently available on WorldClim 2 and participating in CMIP6—the most recent generation of climate models: BCC-CSM2-MR, CNRM-CM6-1, CNRM-ESM2-1, CanESM5, GFDL-ESM4, IPSL-CM6A-LR, MIROC-ES2L, MIROC6 and MRI-ESM2-0. These nine GCMs encompass a wide range of effective climate sensitivities from 2.6 K (MIROC6) to 5.6 K (CanESM5) compared with a range of 1.8–5.6 K across 27 CMIP6 models and 2.1–4.7 K for CMIP5[54]. Temperature and precipitation for future climates have been downscaled and bias-corrected by WorldClim 2 using a change factor approach. The multi-year average of the GCM output for minimum temperature, maximum temperature and total precipitation is calculated for each month of the simulated historical and future period, and the absolute (for temperature) or proportional (for precipitation) difference in these values is then calculated, resulting in climate anomalies that are then applied to the 10 min spatial resolution observed historical dataset[52,55]. WorldClim 2 then calculates Bioclim variables on the basis of these downscaled and bias-corrected data. This approach makes the assumption that the change in climate is relatively stable across space (that is, has high spatial autocorrelation). We downloaded the five preprocessed Bioclim variables for all nine GCMs at 10 min spatial resolution from WorldClim 2[52], and aggregated with bilinear interpolation to 0.25° spatial resolution (approximately 25 km at the equator) to match with the LUH2 land-use data resolution.

Historical land-use data for 2015 and projected land-use data for 2070 were obtained from the Land Use Harmonization 2 (LUH2) project at 0.25° spatial resolution[53,56]. The LUH2 data reconstructs and projects changes in land use among 12 categories: primary forest, non-forested primary land, potentially forested secondary land, potentially non-forested secondary land, managed pasture, rangeland, cropland (four types) and urban land. To capture species' habitat preferences, we downloaded data for all 3,870 mammal species from the IUCN Habitat Classification Scheme (v.3.1) and mapped the 104 unique IUCN habitat classifications onto the 12 land-use types present in the LUH2 dataset following ref. [57] (Supplementary Table 1).

Finally, we downloaded global population projections from the SEDAC Global 1 km Downscaled Population Base Year and Projection Grids based on the SSPs v.1.0[58], and selected the year 2070 for RCP 2.6 (see the 'Climate and land-use futures' section). These data are downscaled to 1 km from a previous dataset at 7.5 arcminute resolution[59]. We aggregated 1 km grids up to 0.25° grids for compatibility with other layers, again using bilinear interpolation.

Additional data. A handful of smaller datasets were incidentally used throughout the study. These included the IUCN Red List, which was used to obtain species taxonomy, range maps and habitat preferences[60]; the US Geological Survey Global Multi-resolution Terrain Elevation Data 2010 dataset, which was used to derive a gridded elevation in metres at around 25 km resolution; and a literature-derived list of suspected hosts of Ebola virus[61].

Mapping species distributions

We developed SDMs for a total of 3,870 species in this study, divided into two modelling pipelines on the basis of data availability (Supplementary Figs. 1 and 2).

Poisson point process models. For 3,088 species with at least 10 unique presence records, Poisson point process models, a method that is closely related to maximum entropy SDMs, were fitted using regularized downweighted Poisson regression[62] with 20,000 background points, using the R package glmnet[63,64]. The spatial domain of predictions was chosen on the basis of the continent(s) in which a species occurred in their IUCN range map. As a final error check, species ranges were constrained to a 1,000 km buffer around their IUCN ranges. We trained SDMs on current climate data using the WorldClim 2 dataset[52], using the five previously specified Bioclim variables.

To reduce the possibility of overfitting patterns due to spatial aggregation, we used spatially stratified cross-validation. Folds were assigned by clustering records on the basis of their coordinates and splitting the resulting dendrogram into 25 groups. These groups were then randomly assigned to five folds (if a species had fewer than 25 records, a smaller number of groups was used based on sample size, and these were split into five folds). This flexible approach accounts for variation in the spatial scale of aggregation among species by using the cluster analysis. By splitting into 25 groups initially (rather than 5), we obtain better environmental coverage (at least on average) within a fold and minimize the need to extrapolate for withheld predictions.

Linear (all species), quadratic (species with >100 records) and product (species with >200 records) features were used. Positive coefficients of quadratic features are not allowed (that is, all have an upper bound of 0 in the model-fitting process), to avoid the undesirable effect of increasing suitability predictions at range edges. The regularization parameter was determined on the basis of fivefold cross-validation with each fold, choosing a value 1 s.d. below the minimum deviance[65]. This resulted in five models per species that were then combined into an unweighted ensemble. Continuous predictions of the ensemble were converted to binary presence/absence predictions by choosing a threshold based on the fifth percentile of the ensemble predictions at training presence locations.

When models were projected into the future, we limited extrapolation to 1 s.d. beyond the data range of the presence locations for each predictor. This decision balances a small amount of extrapolation based on patterns in a species niche with limiting the influence of monotonically increasing marginal responses, which can lead to statistically unsupported (and probably biologically unrealistic) responses to climate.

Range-bagging models. For an additional 783 poorly sampled species (3 to 9 unique points on the 25 km grid), we produced SDMs with a simpler range-bagging algorithm, a stochastic hull-based method that can estimate climate niches from an ensemble of underfit models[66,67] that is therefore well suited for smaller datasets. From the full collection of presence observations and environmental variables range-bagging proceeds by randomly sampling a subset of presences (proportion p) and a subset of environmental variables (d). From these, a convex hull around the subset of points is generated in environmental space. The hull is then projected onto the landscape with a location considered part of the species range if its environmental conditions fall within the estimate hull. The subsampling is replicated N times, generating N 'votes' for each cell on the landscape. One can then choose a threshold for the number of votes required to consider the cell as part of the species' range to generate the binary map used in our downstream analyses. On the basis of the guidelines in ref. [66], we chose $p = 0.33$, $d = 2$ and $N = 100$. We then chose the voting threshold to be 0.165 (=0.33/2) because this implies that the cell is part of the range at least half the time for each subsample. After visual inspection, this generally led to predictions that were very conservative about inferring that unsampled locations were part of a species distribution. The same environmental predictors and ecoregion-based domain selection rules were used for range-bagging models as were used for the point process models discussed above. This hull-based approach is particularly valuable for poorly sampled species which may suffer from sampling bias because bias within niche limits has little effect on range estimates.

Model validation and limitations. PPM models performed well, with a mean test area under the receiver operator curve (AUC) under fivefold cross-validation (using spatial clustering to reduce inflation) of 0.78 (s.d., 0.14). The mean partial AUC evaluated over a range of sensitivity relevant for SDM (0.8–0.95) was 0.81 (s.d., 0.09). The mean sensitivity of binary maps used to assess range overlap (based on the 5% training threshold used to make a binary map) was 0.90 (s.d., 0.08). Range-bagging models were difficult to meaningfully evaluate because they were based on extremely small sample sizes (3–9). The mean training AUC (we did not perform cross-validation due to small sample size) was 0.96 (s.d., 0.09). The binary maps had perfect sensitivity (1) because the threshold used to make them was chosen sufficiently low to include the handful of known presences for each species. One way to assess how well we inferred the range for these species is to quantify how much of the range was estimated based on our models, on the basis of the number of 10 km cells predicted to be part of the species range even when it was not observed there. The mean number of cells inferred to contain a presence was 254 (s.d., 503); however, the distribution is highly right skewed with a median of 90. This indicates that the range-bagging models were typically relatively conservative about inferring ranges for poorly sampled species.

Although our models performed well, we note that researchers should approach the interpretation of SDMs with a certain degree of caution. Even adhering to best practices, many SDM methods are sensitive to subjective user-end choices that influence model performance, transferability and interpretability. Some of those choices may have marginally affected the patterns that we document in this study. For example, to quantify the resilience of our results to the choice of threshold, we constructed pairwise overlaps for the current range estimates of all species across three habitat suitability thresholds (1%, 5% and 10%). We did this using the climate projections, the IUCN-clipped climate projections, and the land-use and IUCN-clipped projections (see the sections below), such that there were nine total replicates, only one of which (IUCN and land use-clipped 5% threshold) was used in our main analyses. We fitted the proportional overlap between each species pair across all nine replicates in a linear mixed model with the identity of the species pair and the thresholding replicate as random effects to quantify the variance associated with the choice of processing pipeline compared with the variance associated with the species pair itself. We also examined the mean proportional overlap across the nine replicates. Our linear mixed model examining the variance associated with thresholding pipeline found that thresholding accounted for only 2.2% of the variance in proportional overlap, in contrast to the 72.3% accounted for by the identity of the species pair. Furthermore, there was very little difference observed in the mean proportional overlap and the number of overlapping species across thresholds. These results demonstrate that the choice of thresholding had an effect on the results of our analysis, but an extremely marginal one, and we expect that similar results would be found for other choices such as variable set reduction, model calibration, the resolution of predictor data and the processing of point occurrence data.

Finally, we note that—although many factors besides climate are ignored by our models, such as biotic interactions or animal social behaviour—our models are tailored to our aim: predicting hotspots of elevated risk under climate change. In our application, correctly predicting presences is more important than incorrect prediction of absences, because we are focused on the potential for new species overlap. We cannot say whether that overlap will happen, owing to

the multiple factors besides climate that influence distributions and range shifts, but we can say with confidence based on robust current niche estimates, validated with spatially stratified cross-validation and biologically grounded estimates of dispersal ability where risk would be elevated in accordance with our simulations.

Habitat range and land use. To capture species' habitat preference, we collated data for all 3,870 mammal species from the IUCN Habitat Classification Scheme (v.3.1). We next mapped 104 unique IUCN habitat classifications onto the 12 land-use types present in the LUH2 dataset. For 962 species, no habitat data were available, or no correspondence existed between a land type in the IUCN scheme and our land-use data; for these species, land-use filters were not used. Filtering on the basis of habitat was performed as permissively as possible: species were allowed in current and potential future ranges to exist in a pixel if any non-zero percent was assigned a suitable habitat type; almost all pixels contain multiple habitats. In some scenarios, human settlements cover at least some of a pixel for most of the world, allowing synanthropic species to persist throughout most of their climatically suitable range. For those with habitat data, the average reduction in range from habitat filtering was 7.6% of pixels.

Predicting future species distributions

We modelled a total of 136 future scenarios, produced by the four paired climate–land-use change pathways replicated across nine GCMs (with one, GFDL-ESM4, available for only two climate scenarios, RCP 2.6 and RCP 7.0; see below), modified by two optional filters on species ranges (habitat preferences and dispersal limits). The full matrix of possible scenarios captures a combination of scenario uncertainty about global change and epistemological uncertainty about how best to predict species' range shifts. By filtering potential future distributions on the basis of climate, land-use and dispersal constraints, we aimed to maximize realism; our predictions were congruent with extensive previous literature on climate- and land-use-driven range loss[57,68,69].

Climate and land-use futures. We considered four possible scenarios for the year 2070 each based on a pairing of the RCPs and the SSPs. RCP numbers (that is, 2.6 or 4.5) represent W m^{-2} of additional radiative forcing by the end of the century, whereas SSPs describe alternative possible pathways of socioeconomic development and demographical change. As pairs, SSP–RCP scenarios describe alternative futures for global socioeconomic and environmental change. Not all SSP–RCP scenario combinations in the 'scenario matrix' are realistically possible[70]. For example, in the vast majority of integrative assessment models, decarbonization cannot be achieved fast enough in the SSP 5 scenario to achieve RCP 2.6.

We used four SSP–RCP combinations: SSP 1–RCP 2.6, SSP 2–RCP 4.5, SSP 3–RCP 7.0, and SSP 5–RCP 8.5. We selected these four scenarios because they span a wide range of plausible global change futures, and serve as the basis for climate model projections in the Scenario Model Intercomparison Project for the newest generation of GCMs (CMIP6)[13]. SSP 1–RCP 2.6 is a scenario with low population growth, strong greenhouse gas mitigation and land-use change (especially an increase in global forest cover), which makes global warming probably less than 2 °C above preindustrial levels by 2100; SSP 2–RCP 4.5 has moderate land-use change and greenhouse gas mitigation with global warming of around 2.5 °C by 2100; SSP 3–RCP 7.0 has high population growth, substantial land-use change (especially a decrease in global forest cover) and very weak greenhouse gas mitigation efforts with global warming of around 4 °C by 2100; and SSP 5–RCP 8.5 is the highest warming scenario with less decrease in forest cover compared with SSP 3 but more substantial increases in coal and other fossil fuel use, leading to more than 4 °C warming by 2100[13,71–73].

Climate model uncertainty. To identify the contribution of climate model uncertainty and its propagation through our analysis, we used all nine selected GCMs from CMIP6 and produced multimodel averages for Figs. 1–5. For all statistical analysis in the main text, we present each multimodel mean with a s.d. across the nine GCMs. We also compared the first encounters from the two models with the highest (CanESM5) and lowest (MIROC6) effective climate sensitivity in the available CMIP6 set on WorldClim[54] (Extended Data Fig. 5) . We also present the map of first encounters and novel viral sharing in each GCM run for each RCP, accounting for both climate and land-use change, with the full dispersal and limited dispersal scenario (Supplementary Figs. 3–20).

Limiting dispersal ability. Not all species can disperse to all environments, and not all species have equal dispersal ability, in ways that are likely to covary with viral sharing properties. We follow a rule proposed in ref. [14], in which Schloss described an approximate formula for mammal range shift ability on the basis of body mass and trophic position. For carnivores, the maximum distance travelled in a generation is given as $D = 40.7M^{0.81}$, where D is distance in kilometres and M is body mass in kilograms. For herbivores and omnivores, the maximum is estimated as $D = 3.31M^{0.65}$.

We used mammalian diet data from the EltonTraits database[74], and used the same cut-off as ref. [14] to identify carnivores as any species with 10% or less plants in their diet. We used body mass data from EltonTraits in the Schloss formula to estimate maximum generational dispersal, and converted estimates to annual maximum dispersal rates by dividing by generation length, as previously estimated by another comprehensive mammal dataset[75]. We multiply by 50 years (from 2020 as the present to 2070) and use the resulting distance as a buffer around the original range map, and constrain possible range shifts within that buffer. For 420 species with missing data in one of the required sources, we interpolated dispersal distance based on the closest relative in our supertree with a dispersal velocity estimate.

Qualified by the downsides of assuming full dispersal[76], we excluded bats from the assumed scaling of dispersal limitations. The original study[14] chose to omit bats entirely, and subsequent work has not proposed any alternative formula. Moreover, the Schloss formula performs notably poorly for bats. For example, it would assign the largest bat in our study, the Indian flying fox (*Pteropus giganteus*), a dispersal ability of lower than that of the grey dwarf hamster (*Cricetulus migratorius*). Bats were instead given full dispersal in all scenarios: given significant evidence that some bat species regularly cover continental distances[24,25], and that isolation by distance is uncommon within many bats' ranges[35], we felt this was a defensible assumption for modelling purposes. Moving forwards, the rapid range shifts already observed in many bat species (see 'Dispersal drives the importance of bats' in the main text) could provide an empirical reference point to fit a new allometric scaling curve (after standardizing those results for the studies' many different methodologies). A different set of functional traits probably govern the scaling of bat dispersal, chiefly the aspect ratio (length:width) of wings, which is a strong predictor of population genetic differentiation[35]. Migratory status would also be important to include as a predictor, although here we exclude information on long-distance migration for all species (owing to a lack of any real framework for adding that information to SDMs in the literature).

Explaining spatial patterns

To examine the geography of new assemblages, we used linear models that predicted the number of first encounters (new overlap of species pairs) at the 25 km level (N = 258, 539 grid cells). Explanatory variables included: richness (number of species inhabiting the grid cell in our predicted current ranges for the given scenario); elevation in metres (derived from the US Geological Survey Global Multi-resolution Terrain Elevation Data 2010 dataset); and the predominant land cover type for the grid cell. We simplified the classification scheme for land-use types into five categories for these models (human settlement, cropland, rangeland and pasture, forest and unforested wildland), and assigned pixels a single land-use type based on the maximum probability from

the land-use scenarios. We fitted a model for each scenario and pair of biological assumptions; owing to the large effect that bats had on the overall pattern, we retrained these models on subsets of encounters with and without a bat species involved. To help model fitting, we log[x + 1]-transformed the response variable (number of overlaps in the pixel) and both continuous explanatory variables (metres of elevation above the lowest point and species richness). As some elevation values were lower than 0 (that is, below sea level), we treated elevation as metres above the lowest terrestrial point rather than metres above sea level to allow us to log-transform the data.

Viral sharing models

Criteria for species inclusion. Of the 3,870 species for which we generated distribution models, 103 were aquatic mammals (cetaceans, sirenians, pinnipeds and sea otters) and 382 were not present in the mammalian supertree that we used for phylogenetic data[77]. These species, and the associated SDMs, were excluded from the analysis. Aquatic species were removed using a two-filter approach, by first cross-referencing with Pantheria[78], and second by checking that no species had only aquatic habitat use types (see the 'Habitat range and land use' section). We also excluded 246 monotremes and marsupials because the shape of the supertree prevented us from fitting satisfactory generalized additive mixed model (GAMM) smooths to the phylogeny effect, leaving 3,139 non-marine placental mammals with associated phylogenetic data.

Generalized additive mixed models. We used a previously published model of the phylogeography of viral sharing patterns to make predictions of future viral sharing[10]. This model was based on an analysis of 510 viruses shared between 682 mammal species[2], and predicted the probability that a pair of mammal species will share a virus given their geographical range overlap and phylogenetic relatedness. The original study uncovered strong, nonlinear effects of spatial overlap and phylogenetic similarity in determining viral sharing probability, and simulating the unobserved global network using these effect estimates capitulated multiple macroecological patterns of viral sharing.

In the original study, a GAMM was used to predict virus sharing as a binary variable on the basis of (1) geographical range overlap; (2) phylogenetic similarity; and (3) species identity as a multimembership random effect. The phylogeographical explanatory variables were obtained from two broadly available, low-resolution data sources: pairwise phylogenetic similarity was derived from a mammalian supertree previously modified for host-pathogen studies[2,77], with similarity defined as the inverse of the cumulative branch length between two species, scaled to between 0 and 1. Geographical overlap was defined as the area of overlap between two species' IUCN range maps, divided by their cumulative range size[79].

We first retrained the GAMMs from[10] on the pairwise overlap matrix of SDMs generated for this study, such that present predictions would be comparable with potential future distributions. Of the 3,139 species in our reduced dataset, 544 had viral records in our viral sharing dataset and shared with at least one other mammal, and were used to retrain the GAMM from ref. [10]. To check the performance of the GAMM, we predicted sharing patterns with (1) only random effects, (2) only fixed effects and (3) with both. To extend predictions to the full set of mammals, we generated random effects for out-of-sample species by drawing from the fitted distribution of species-level effects (predicting without these random effects underestimates species variance, resulting in mean sharing of 0.02 rather than the observed 0.06). The mean sharing value across these predictions closely approximated observed sharing probability (~0.06).

Note that this model uses citation counts to correct for sampling bias, an imperfect method but one that leads to strong validation performance on an independently compiled dataset of host–virus associations, which carries a different set of biases. However, it is still possible that sampling bias in host–virus datasets could artificially inflate the signal of phylogeography in viral sharing, if researchers investigating a noteworthy viral detection then preferentially sample closely related host species in the immediate area. It is improbable that these effects would bias our results in a particular direction, but accounting for these biases should at least involve some acknowledgement that cross-species transmission is challenging to predict (see the discussion in ref. [10] for a more in-depth treatment of sampling bias effects).

Model validation and limits. Compared with the current viral sharing matrix, the model performs well with only fixed effects (AUC = 0.80) and extremely well with both fixed and random effects (AUC = 0.93). The model explained a very similar proportion of the deviance in viral sharing to that in ref. [10] (44.5% and 44.8%, respectively).

In practice, several unpredictable but confounding factors could affect the reliability of this model as a forecasting tool, including temperature sensitivity of viral evolution in host jumps[41], or increased susceptibility of animals with poorer health in lower-quality habitat or unfavourable climates. Moreover, once viruses can produce an infection, their ability to transmit within a new species is an evolutionary race between mutation and recombination rates in viral genomes, host innate and adaptive immunity, virulence-related mortality, and legacy constraints of coevolution with previous hosts and vectors[80,81]. However, data cataloguing these precise factors are hardly comprehensive for the hundreds of zoonotic viruses, let alone for the thousands of undescribed viruses in wildlife. Moreover, horizontal transmission is not necessary for spillover potential to be considered significant; for example, viruses such as rabies or West Nile virus are not transmitted within human populations but humans are still noteworthy hosts.

Mapping opportunities for sharing. We used the GAMM effect estimates to predict viral sharing patterns across the 3,139 mammals with associated geographical range and phylogenetic data, for both the present and future scenarios. By comparing current and future sharing probabilities for each of the four global change scenarios, we estimated which geographical and taxonomic patterns of viral sharing would probably emerge. We separately examined patterns of richness, patterns of sharing probability and their change (that is, future sharing probability − current sharing probability, giving the expected probability of a new sharing event).

A subset of the mammals in our dataset were predicted to encounter each other for the first time during range shifts. For each of these pairwise first encounters, we extracted the area of overlap in every future scenario, and assigned each overlap a probability of sharing from the mean GAMM predictions and mapped the mean and cumulative probability of a new sharing event happening in a given geographical pixel.

Case study on ZEBOV. For a case study in possible significant cross-species transmission, we compiled a list of known hosts of ZEBOV, a zoonosis with potentially high host breadth that has been known to cause wildlife die-offs, but has no known definitive reservoir. Hosts were taken from two sources: the training dataset on host–virus associations[2] and an additional dataset of filovirus testing in bats[61]. In the latter case, any bats that have been reported to be antibody positive or PCR-positive for ZEBOV were included. A total of 19 current known hosts were selected. We restricted our analysis to the 13 hosts from Africa, because there is no conclusive evidence that ZEBOV actively circulates outside Africa; although some bat species outside Africa have tested positive for antibodies to ZEBOV, this is probably due to cross-reactivity with other undiscovered filoviruses[61,82,83]. We used the 13 African hosts to predict possible first encounters in all scenarios (Extended Data Fig. 8), and mapped the current richness of ZEBOV hosts, the change in host richness by 2070 and the number of first encounters (Fig. 3).

Overlap with human populations. To examine the possibility that hotspots of cross-species transmission would overlap with human populations, we used SEDAC's global population projections v.1.0 for the year 2070[58]. We aggregated these to native resolution for each of the four

SSPs paired with the native RCP–SSP pairing for the SDMs. In Fig. 4, we present the population projections for SSP 1, which pairs with RCP 2.6.

Timing of opportunity

As a final supplementary analysis, we examined the potential timing of first encounters throughout a century of species' movements. To do so, we reproduced our entire modelling pipeline using CHELSA v.2.1, a climate product for CMIP6 with five general circulation models (GFDLESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0 and UKESM1-0-LL) projected in three scenarios (SSP 1–RCP 2.6, SSP 3–RCP 7.0 and SSP 5–RCP 8.5) and averaged in four time slices (1981–2010, 2011–2040, 2041–2070 and 2071–2100). Given the focus on timing, species occurrence records were filtered to span the years 1900 to 2019; this enabled us to build 2,948 point process models and 903 range-bagging models (3,850 models) for this analysis. We treated the first time interval as the baseline for SDMs, and projected species distributions into each of those three future intervals; we then identified the first encounters in each time interval as a cumulative process (for example, two species that meet in 2011–2040 for the first time cannot also have a first encounter in 2041–2070).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

No original data were generated during our study. All raw datasets are available online, including the GBIF database of biodiversity occurrence data (https://www.gbif.org/), the IUCN Red List (https://www.iucnredlist.org/), the WorldClim climate dataset (https://worldclim.org/)[52], the CHELSA climate dataset (https://chelsa-climate.org/cmip6/)[84], the LUH2 land-use dataset (https://luh.umd.edu/)[56], the USGS GMTED 2010 elevation dataset (https://www.usgs.gov/coastal-changes-andimpacts/gmted2010), the HP3 dataset of host–virus associations (https://github.com/ecohealthalliance/HP3; https://doi.org/10.5281/zenodo.596810) and a dataset of filovirus testing in bats[61].

Code availability

Code to reproduce the study is deposited on Zenodo (https://doi.org/10.5281/zenodo.6463429) and is available at GitHub (https://github.com/viralemergence/iceberg). Additional code to generate the generalized additive mixed models used in this study, reused from ref. [10], are also available at GitHub (https://github.com/gfalbery/ViralSharingPhylogeography).

49. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017). Data from.
50. Washburne, A. D. et al. Taxonomic patterns in the zoonotic potential of mammalian viruses. *PeerJ* **6**, e5979 (2018).
51. Grubbs, F. E. et al. Sample criteria for testing outlying observations. *Ann. Math. Stat.* **21**, 27–58 (1950).
52. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
53. Hurtt, G. et al. Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev* **13**, 5425–5464 (2018).
54. Zelinka, M. D. et al. Causes of higher climate sensitivity in cmip6 models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
55. Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A. & Ramirez-Villegas, J. High-resolution and bias-corrected cmip5 projections for climate change impact assessments. *Sci. Data* **7**, 7 (2020).
56. Hurtt, G. C. et al. 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. *Clim. Change* **109**, 117–161 (2011).
57. Powers, R. P. & Jetz, W. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nat. Clim. Change* **9**, 323–329 (2019).
58. Gao, J. *Downscaling Global Spatial Population Projections From 1/8-Degree to 1-km Grid Cells* NCAR Technical Note NCAR/TN-537+STR (National Center for Atmospheric Research, 2017).
59. Jones, B. & O'Neill, B. C. Spatially explicit global population scenarios consistent with the shared socioeconomic pathways. *Environ. Res. Lett.* **11**, 084003 (2016).
60. *The IUCN Red List of Threatened Species* http://www.iucnredlist.org/ (IUCN, 2019).
61. Han, B. A. et al. Undiscovered bat hosts of filoviruses. *PLoS Negl.Trop. Dis.* **10**, e0004815 (2016).
62. Renner, I. W. et al. Point process models for presence-only analysis. *Methods Ecol. Evol.* **6**, 366–379 (2015).
63. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
64. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
65. Hastie, T., Tibshirani, R. & Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn (Springer, 2009).
66. Drake, J. M. Range bagging: a new method for ecological niche modelling from presence-only data. *J. R. Soc. Interface* **12**, 20150086 (2015).
67. Drake, J. M. & Richards, R. L. Estimating environmental suitability. *Ecosphere* **9**, e02373 (2018).
68. Jetz, W., Wilcove, D. S. & Dobson, A. P. Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biol.* **5**, e157 (2007).
69. Pecl, G. T. et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017).
70. van Vuuren, D. P. et al. The shared socio-economic pathways: trajectories for human development and global environmental change. *Global Environ. Change* **42**, 148–152 (2017).
71. Riahi, K. et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environ. Change* **42**, 153–168 (2017).
72. Popp, A. et al. Land-use futures in the shared socio-economic pathways. *Global Environ. Change* **42**, 331–345 (2017).
73. Samir, K. & Lutz, W. The human core of the shared socioeconomic pathways: population scenarios by age, sex and level of education for all countries to 2100. *Global Environ. Change* **42**, 181–192 (2017).
74. Wilman, H. et al. Eltontraits 1.0: species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027–2027 (2014).
75. Pacifici, M. et al. Generation length for mammals. *Nat. Conserv.* **5**, 89–94 (2013).
76. Bateman, B. L., Murphy, H. T., Reside, A. E., Mokany, K. & VanDerWal, J. Appropriateness of full-, partial-and no-dispersal scenarios in climate change impact modelling. *Divers. Distrib.* **19**, 1224–1234 (2013).
77. Fritz, S. A., Bininda-Emonds, O. R. & Purvis, A. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol. Lett.* **12**, 538–549 (2009).
78. Jones, K. E. et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648 (2009).
79. Araújo, M. B., Rozenfeld, A., Rahbek, C. & Marquet, P. A. Using species co-occurrence networks to assess the impacts of climate change. *Ecography* **34**, 897–908 (2011).
80. Geoghegan, J. L., Senior, A. M., Di Giallonardo, F. & Holmes, E. C. Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl Acad. Sci. USA* **113**, 4170–4175 (2016).
81. Walker, J. W., Han, B. A., Ott, I. M. & Drake, J. M. Transmissibility of emerging viral zoonoses. *PLoS ONE* **13**, e0206926 (2018).
82. Olival, K. J. et al. Ebola virus antibodies in fruit bats, Bangladesh. *Emerg. Infect. Dis.* **19**, 270–273 (2013).
83. Yang, X.-L. et al. Genetically diverse filoviruses in *Rousettus* and *Eonycteris* spp. bats, China, 2009 and 2015. *Emerg. Infect. Dis.* **23**, 482–486 (2017).
84. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).

Acknowledgements This paper is the culmination of several years of idea development. We thank many people, including the entire Bansal laboratory, L. W. Alexander, K. Burgio, E. Dougherty, R. Garnier, W. Getz, P. Hitchens, C. Johnson and I. Ott; L. W. Alexander for sharing bat filovirus testing sources used to compile the Ebola subnetwork; and J. Hidasi-Neto for publicly available data-visualization code. C.J.C. was supported by the Georgetown Environment Initiative and the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1639145. C.J.C., G.F.A. and E.A.E. were supported by funding to the Verena Consortium including NSF BII 2021909 and a grant from Institut de Valorisation des Données (IVADO). C.M. acknowledges funding from National Science Foundation grant DBI-1913673. E.A.E., K.J.O. and N.R. were supported by the United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT project.

Author contributions C.J.C. and G.F.A. conceived the study. C.M., C.J.C. and C.H.T. developed SMDs. G.F.A., E.A.E., K.J.O. and N.R. developed the generalized additive models. G.F.A., C.J.C. and C.M.Z. integrated the predictions of species distributions and viral sharing patterns and designed visualizations. All of the authors contributed to writing the manuscript.

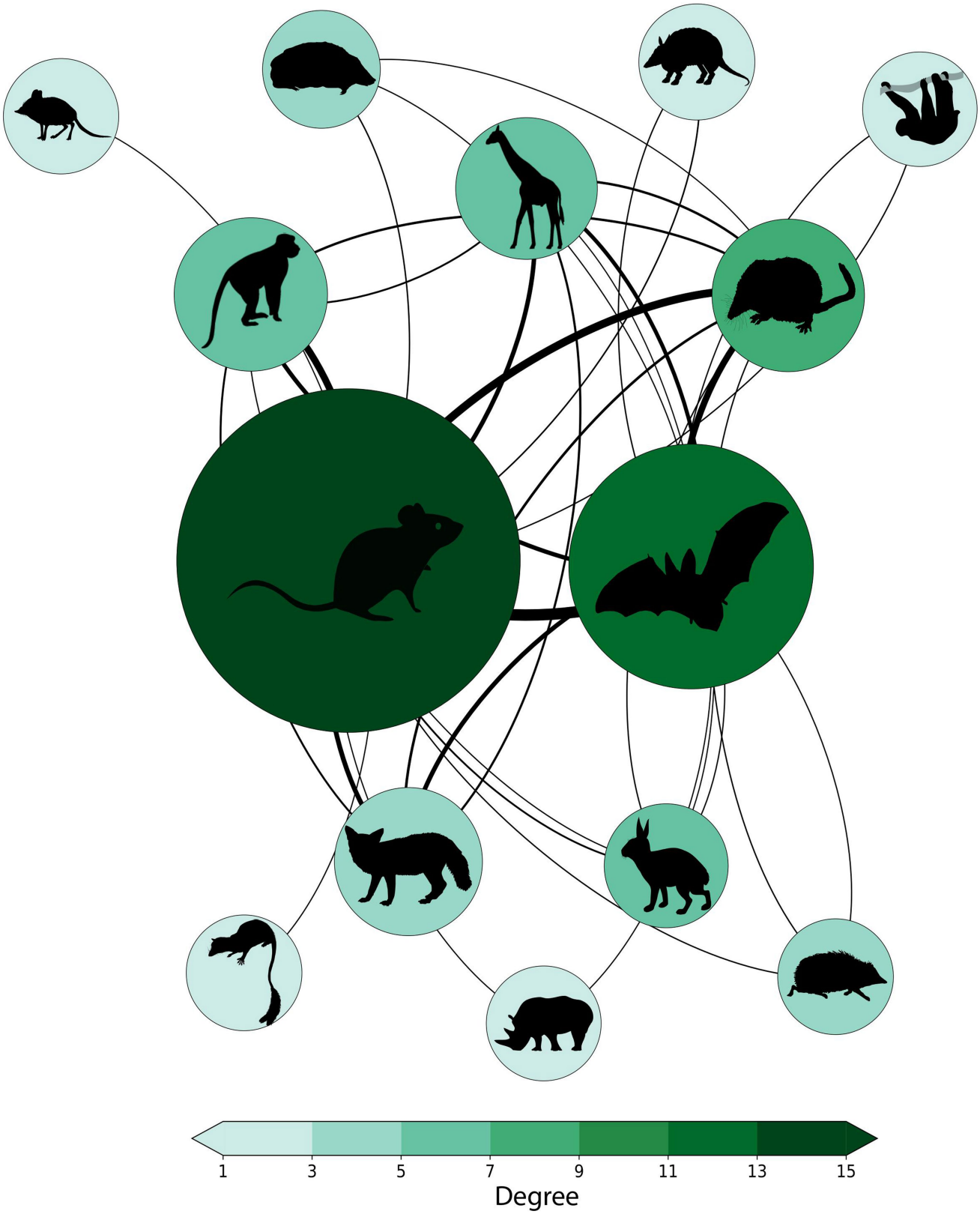
Competing interests The authors declare no competing interests.

Additional information

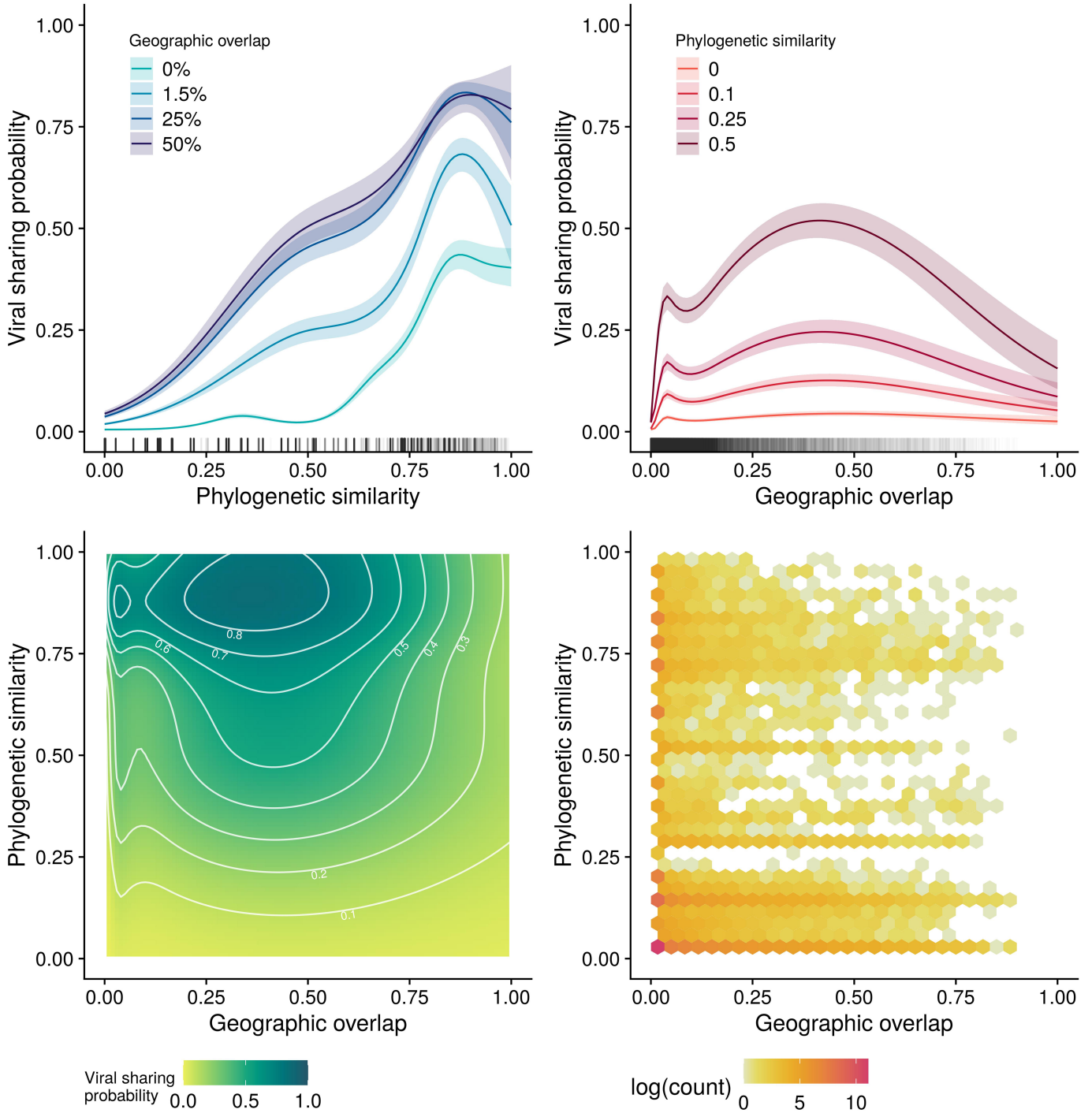
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04788-w.

Correspondence and requests for materials should be addressed to Colin J. Carlson or Gregory F. Albery.

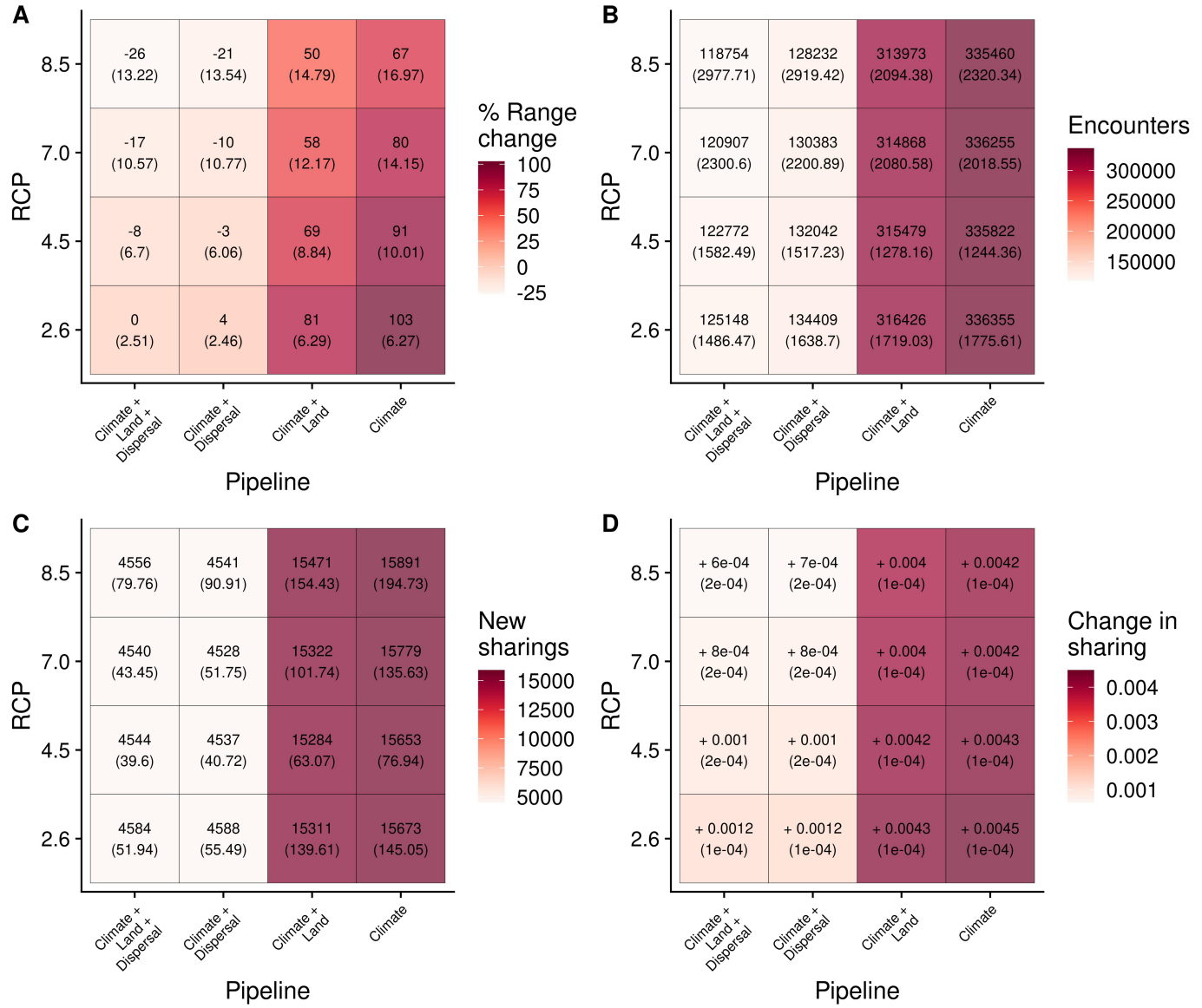
Peer review information *Nature* thanks Jonathan Dushoff, Tom Matthews and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



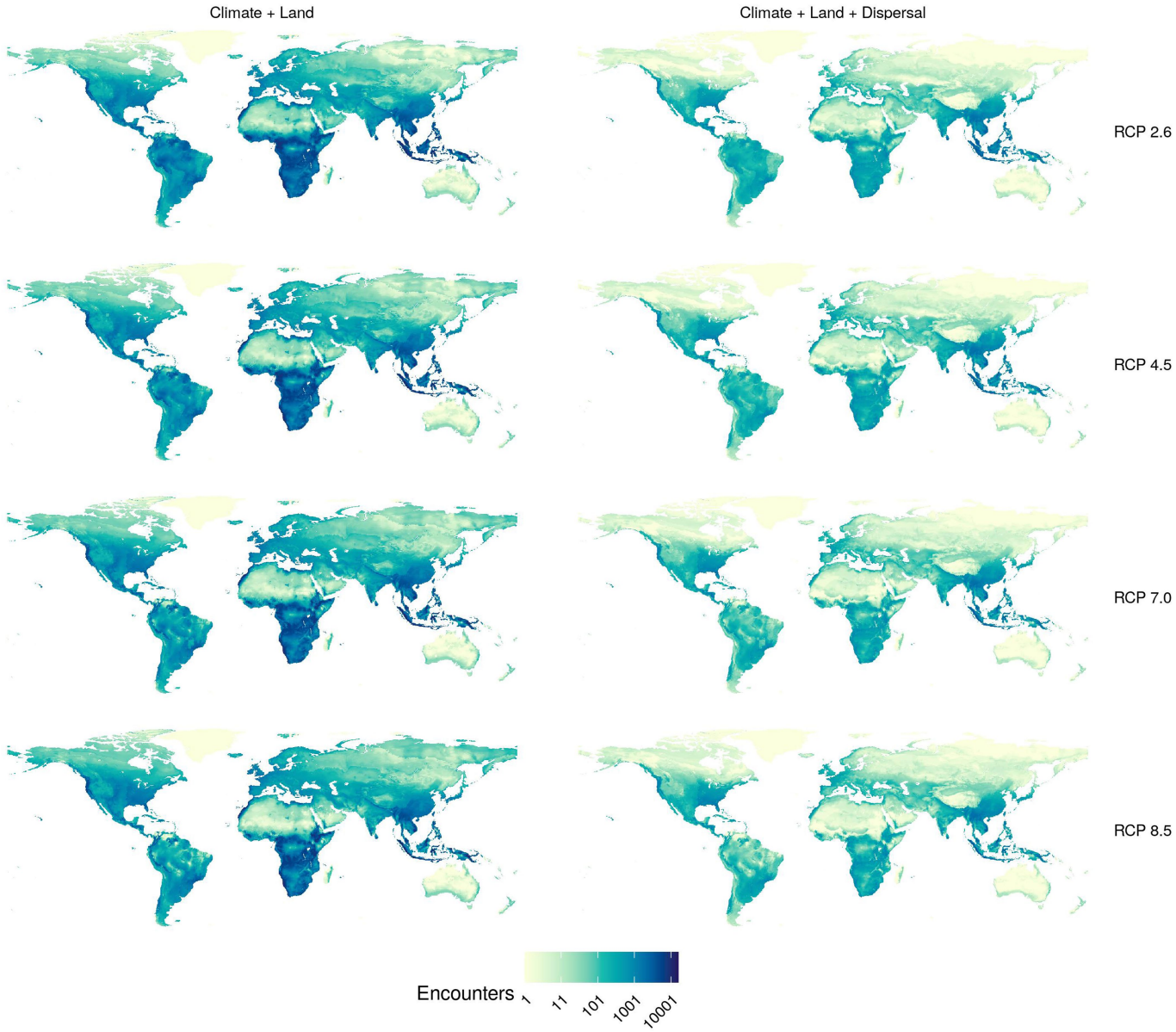
Extended Data Fig. 1 | The mammal-virus network. The present-day viral sharing network by mammal order inferred from modelled pairwise predictions of viral sharing probabilities. Edge width denotes the expected number of shared viruses (the sum of pairwise species-species viral sharing probabilities), with most sharing existing among the most speciose and closely related groups. Edges shown in the network are the top 25% of links. Nodes are sized by total number of species in that order in the host-virus association dataset, colour is scaled by degree. Silhouettes are from http://phylopic.org under Creative Commons license (creativecommons.org/licenses/by/3.0).



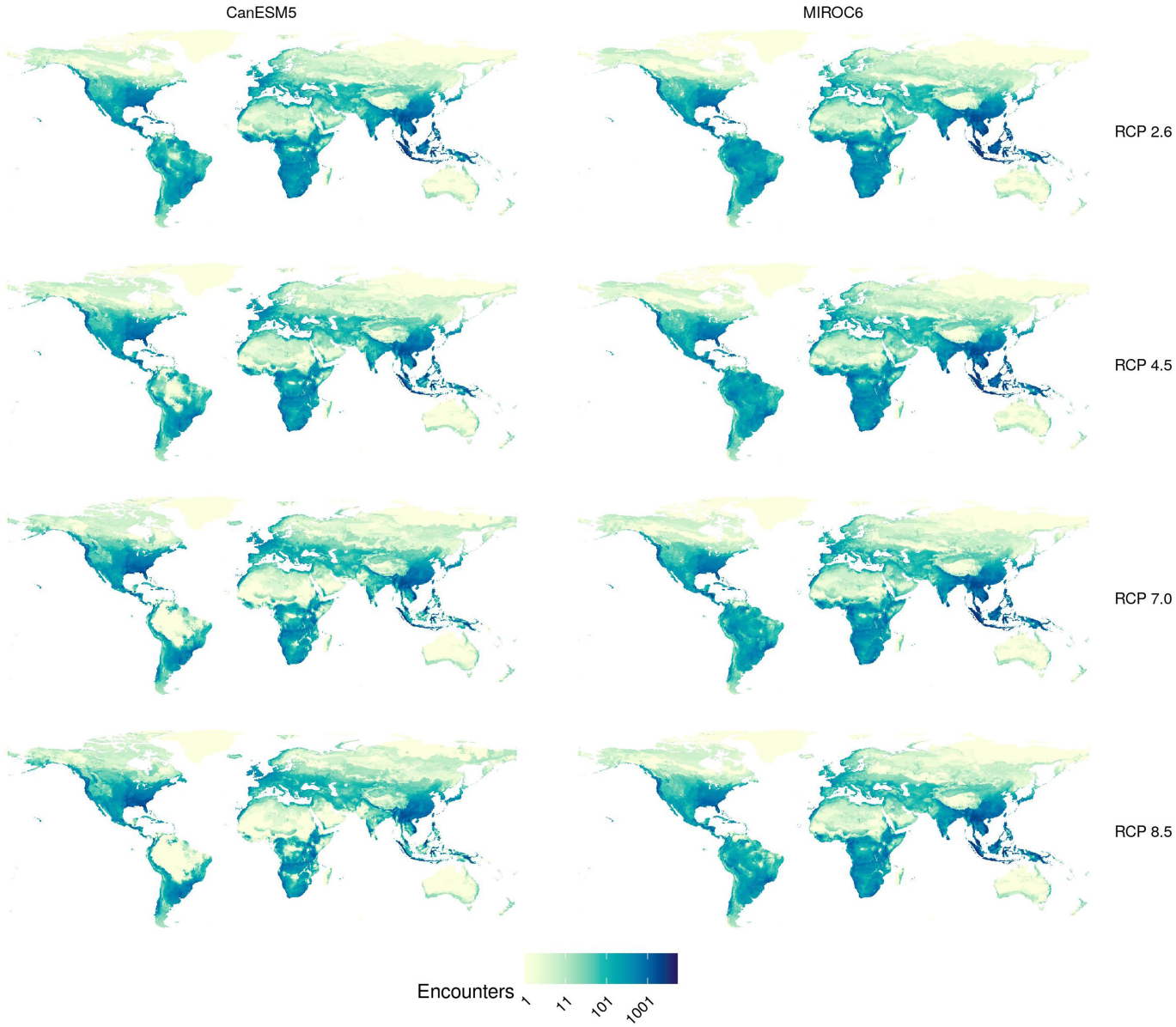
Extended Data Fig. 2 | Predicted phylogeographical structure of viral sharing. Phylogeographical prediction of viral sharing using a generalized additive mixed model. Viral sharing increases as a function of phylogenetic similarity (upper left) and geographical overlap (upper right), which have strong nonlinear interactions, shown in the contour map of joint effects (bottom left). Error bars are the 95% confidence interval for the estimated response. White contour lines denote 10% increments of sharing probability. Declines at high values of overlap may be an artefact of model structure and low sampling in the upper levels of geographical overlap, shown in a hexagonal bin chart of the raw data distribution (bottom right).



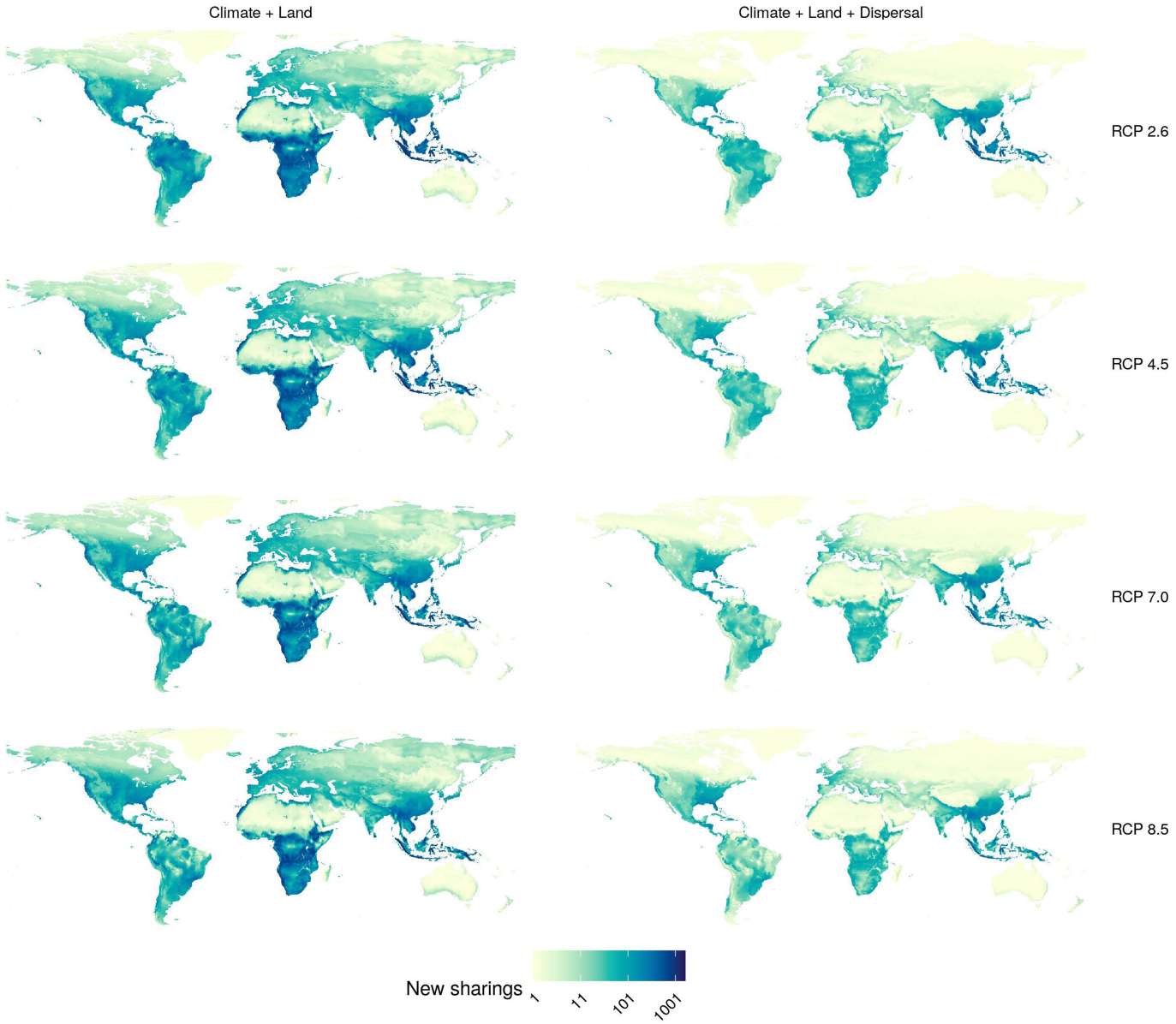
Extended Data Fig. 3 | Outcomes by model formulation and climate change scenario. Heatmaps displaying predicted changes across model formulations. (A) Range expansions were highest in non-dispersal-limited scenarios and in scenarios with lower levels of global warming. (B) The number of predicted first encounters was higher in non-dispersal-limited scenarios and in scenarios with lower levels of global warming. (C) The number of expected new viral sharing events was higher in non-dispersal-limited scenarios and in more severe RCPs. (D) The overall change in sharing probability (connectance) across the viral sharing network between the present day and the future scenarios; absolute changes may appear small, but an 0.4% increase in connectivity is notable on the scale of millions of possible pairwise combinations of species. Change is positive across all scenarios, being greatest in non-dispersal-limited scenarios and in scenarios with lower levels of global warming. Results are averaged across nine global climate models, with standard deviation indicated in parentheses underneath main statistics.



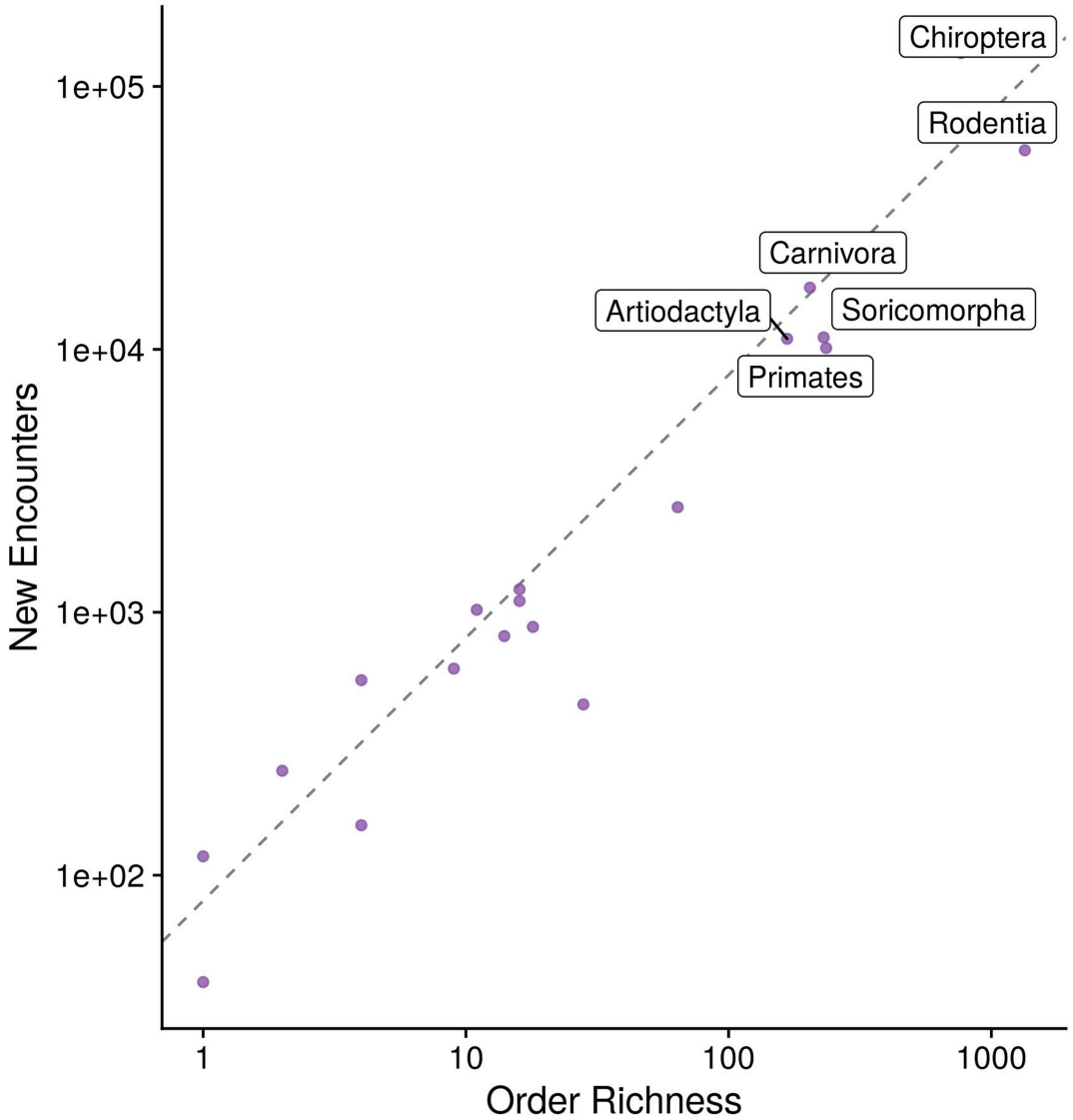
Extended Data Fig. 4 | Geographical distribution of first encounters. Predictions were carried out for four representative concentration pathways (RCPs), accounting for climate change and land use change, without (left) and with dispersal limits (right). Darker colours correspond to greater numbers of first encounters in the pixel. Results are averaged across nine global climate models.



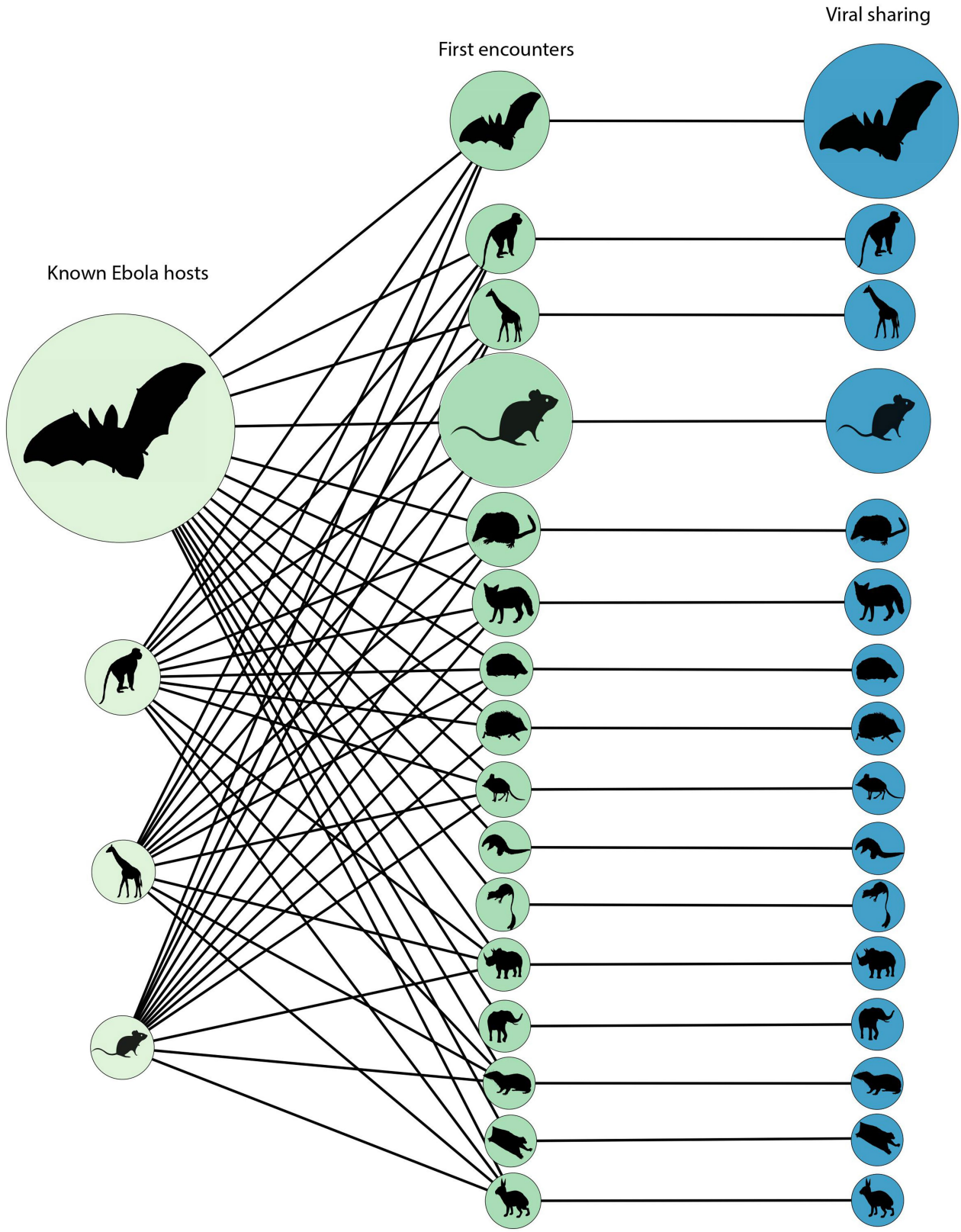
Extended Data Fig. 5 | Geographical distribution of first encounters in two global climate models. Predictions were carried out for four representative concentration pathways (RCPs), accounting for climate change and land use change through pairing with shared socioeconomic pathways (SSPs) as detailed in the Methods. The two models selected are those with the highest (CanESM5) and lowest (MIROC6) effective climate sensitivity in the available CMIP6 set on WorldClim[54]. Darker colours correspond to greater numbers of first encounters in the pixel.



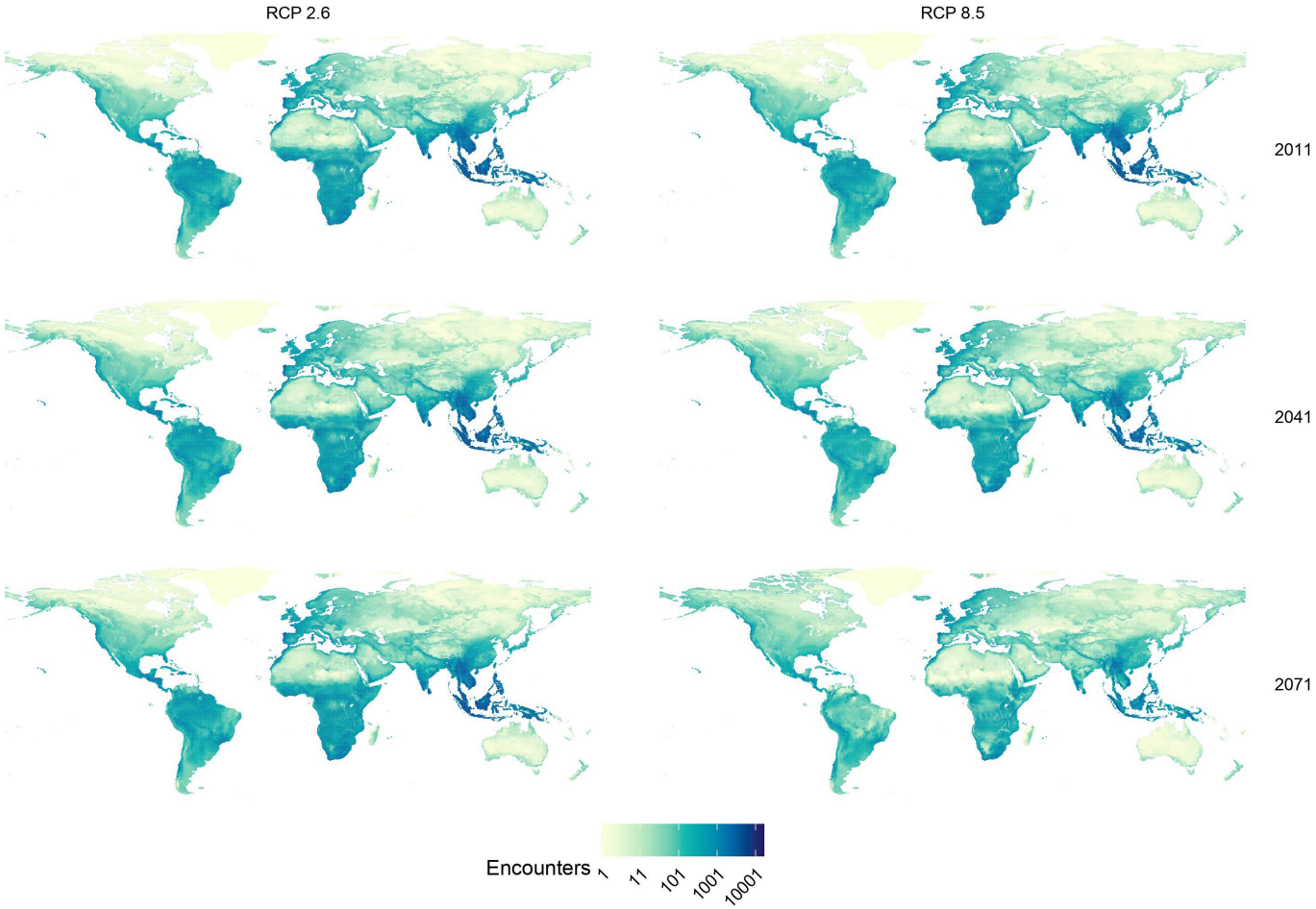
Extended Data Fig. 6 | Geographical distribution of expected viral sharing events from first encounters. Predictions were carried out for potential future distributions for four representative concentration pathways (RCPs), accounting for climate change and land use change, without (left) and with dispersal limits (right). Darker colours correspond to greater numbers of new viral sharing events in the pixel. Probability of new viral sharing was calculated by subtracting the species pair's present sharing probability from their future sharing probability that our viral sharing GAMMs predicted. This probability was projected across the species pair's range intersection, and then summed across all novel species pairs in each pixel. Results are averaged across nine global climate models.



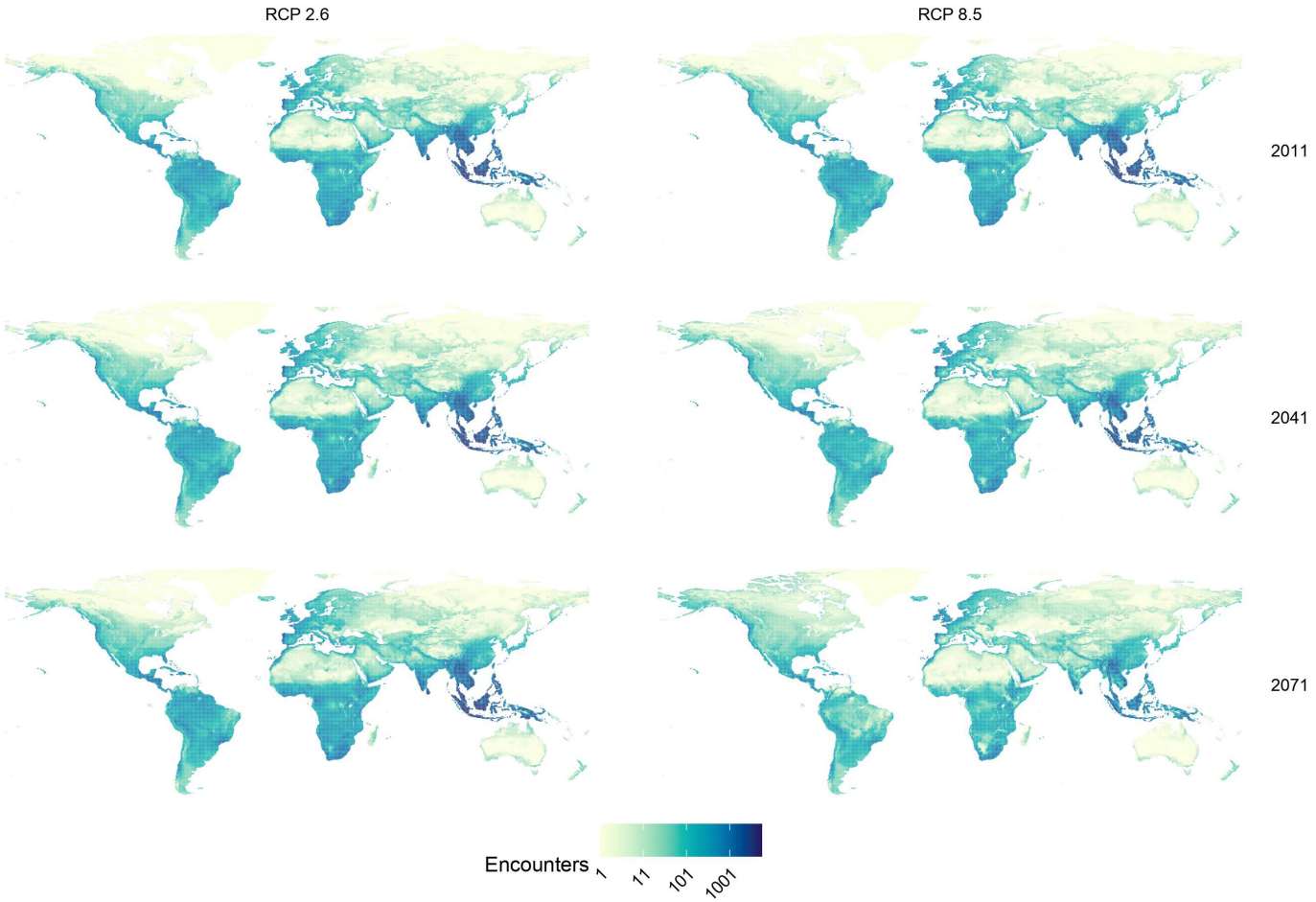
Extended Data Fig. 7 | Order-level heterogeneity in first encounters. Dispersal stratifies the number of first encounters (RCP 2.6 with all range filters), where some orders have more than expected at random, based on the mean number of first encounters and order size (line). Results are averaged across nine global climate models.



Extended Data Fig. 8 | Projected viral sharing from suspected Ebola reservoirs is dominated by bats. Node size is proportional to (left) the number of suspected Ebola host species in each order, which connect to (middle) first encounters with potentially naive host species; and (right) the number of projected viral sharing events in each receiving group. (Node size denotes proportions out of 100% within each column total). While Ebola hosts will encounter a much wider taxonomic range of mammal groups than current reservoirs, the vast majority of future viral sharing will occur disproportionately in bats. (First encounters are averaged across GCMs to capture the maximum range of taxonomic diversity). Silhouettes are from http://phylopic.org under Creative Commons license (creativecommons.org/licenses/by/3.0).



Extended Data Fig. 9 | Geographical distribution of first encounters over time without dispersal restrictions. We show the RCP with the least mitigation (RCP 2.6) and most mitigation (RCP 8.5). Projections are made based on future climate and land use. Years are the start of each interval (2011-2040; 2041-2070; 2071-2100). Darker colours correspond to greater numbers of first encounters in the pixel. Results are averaged across five global climate models from CHELSA v2.1.



Extended Data Fig. 10 | Geographical distribution of first encounters over time with dispersal restrictions. We show the RCP with the least mitigation (RCP 2.6) and most mitigation (RCP 8.5). Projections are made based on future climate and land use. Years are the start of each interval (2011-2040; 2041-2070; 2071-2100). Darker colours correspond to greater numbers of first encounters in the pixel. Results are averaged across five global climate models from CHELSA v2.1.

nature portfolio

Corresponding author(s): Colin J. Carlson

Last updated by author(s): Mar 24, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used R (v4.1.3) code to programatically download and extract data from GBIF (gbif.org), which is directly used in the study's species distribution modeling pipeline.
Data analysis	Two separate pipelines are used in this study, both based in R (v4.1.3). The species distribution modeling component uses the 'glmnet' R package version 4.1-3. Code for both species distribution models and the viral sharing simulation components of the study, and the code to produce all visualizations, has been made publicly available (github.com/viralemergence/iceberg).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

No original data is used in, generated by, or produced by our study. The Data Availability statement includes a full list of open data sources we use.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We report the sample size of species included in the study throughout the paper, including in Extended Data Figure 10. The decision to apply point process or range bagging models is based on a sample size cutoff, informed by literature on the performance of these models with different levels of data. Species with 3 to 9 unique georeferenced points on a 25km grid were entered into the range bagging pipeline, and species with 10 or more were subject to the point process modeling pipeline. Species with fewer points were omitted from the analyses. For climate datasets used in the study, we always used every available climate model at the time of pipeline initiation.
Data exclusions	Species with data limited by sample size (above) were excluded, as were those absent from the phylogeny, marine mammals (inappropriate to model with terrestrial species), and non-eutherian mammals (due to model calibration challenges for phylogenetic regression explained in Albery et al.'s original publication). Species occurrence points were also excluded based on a Grubb outlier test to identify spurious records of their distribution.
Replication	Our study is reproduced on both WorldClim2's CMIP6 models and the CHELSA CMIP6 ensemble, which each include several global climate models (GCMs) and multiple scenarios. Throughout the text, we report key statistics with a standard deviation across GCMs. The code pipeline made available with the study makes simultaneous replication across climate products possible.
Randomization	Not applicable (no experimental trials with a control and intervention; this is a simulation study).
Blinding	Not applicable (no experimental trials of any kind; this is a simulation study, which would be impossible to design "blinded").

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging