Climate change will drive novel cross-species viral transmission

Colin J. Carlson$^{1,2,†}$, Gregory F. Albery$^{1,3,†}$, Cory Merow$^4$, Christopher H. Trisos$^5$, Casey M. Zipfel$^1$, Evan A. Eskew$^3$, Kevin J. Olival$^3$, Noam Ross$^3$, and Shweta Bansal$^1$

$^1$Department of Biology, Georgetown University, Washington, D.C., USA.
$^2$Center for Global Health Science & Security, Georgetown University, Washington, D.C., USA.
$^3$EcoHealth Alliance, New York, NY, USA.
$^4$Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA.
$^5$African Climate and Development Initiative, University of Cape Town, Cape Town, South Africa.

†These authors share equal authorship.

January 24, 2020
Abstract

Between 10,000 and 600,000 species of mammal virus are estimated to have the potential to spread in human populations, but the vast majority are currently circulating in wildlife, largely undescribed and undetected by disease outbreak surveillance. In addition, changing climate and land use drive geographic range shifts in wildlife, producing novel species assemblages and opportunities for viral sharing between previously isolated species. In some cases, this will inevitably facilitate spillover into humans—a possible mechanistic link between global environmental change and emerging zoonotic disease. Here, we map potential hotspots of viral sharing, using a phylogeographic model of the mammal-virus network, and projections of geographic range shifts for 3,870 mammal species under climate change and land use scenarios for the year 2070. Shifting mammal species are predicted to aggregate at high elevations, in biodiversity hotspots, and in areas of high human population density in Asia and Africa, sharing novel viruses between 3,000 and 13,000 times. Counter to expectations, holding warming under 2°C within the century does not reduce new viral sharing, due to greater range expansions—highlighting the need to invest in surveillance even in a low-warming future. Most projected viral sharing is driven by diverse hyperreservoirs (rodents and bats) and large-bodied predators (carnivores). Because of their unique dispersal capacity, bats account for the majority of novel viral sharing, and are likely to share viruses along evolutionary pathways that could facilitate future emergence in humans. Our findings highlight the urgent need to pair viral surveillance and discovery efforts with biodiversity surveys tracking range shifts, especially in tropical countries that harbor the most emerging zoonoses.
In the face of rapid environmental change, survival for many species depends on moving to track shifting climates. Even in a best case scenario, many species are projected to shift a hundred kilometers or more in the next century\textsuperscript{9,10}. In the process, many animals will bring their parasites and pathogens into new environments\textsuperscript{4,11}, creating new evolutionary opportunities for host jumps\textsuperscript{8}. Most conceptual frameworks for cross-species transmission revolve around how these host jumps facilitate the spillover of new zoonotic pathogens into humans\textsuperscript{12,13,14}, but viral evolution is an undirected process\textsuperscript{15}, in which humans are only one of over 5,000 mammal species with over 12 million possible pairwise combinations\textsuperscript{16}. Despite their indisputable significance, zoonotic emergence events are just the tip of the iceberg; almost all cross-species transmission events will occur among wild mammals, largely undetected and mostly inconsequential for public health.

Of the millions of possible pairwise viral exchanges, the vast majority are biologically implausible, as host species’ geographic ranges currently do not overlap. However, as ranges shift, a small fraction of possible interactions will occur, of which a subset will lead to viral establishment in a novel host. Which subset results in establishment depends on opportunity and compatibility\textsuperscript{14,17,18}, analogous to exposure and susceptibility within populations, and both dimensions pose an important predictive challenge. The ability of species to track shifting habitats in a changing climate will determine which pairs of species encounter each other for the first time\textsuperscript{4,19}. Habitat selection and behavioral differences can further limit contact, even if species are nominally sympatric\textsuperscript{19}. Some viruses may spread environmentally between spatially-proximate species with no direct behavioral contact\textsuperscript{20}, but generally, sharing is more likely among species with more ecological overlap\textsuperscript{21}. Even among species in close contact, most spillovers are still a dead end; progressively smaller subsets of viruses can infect novel host cells, proliferate, and transmit onward in a new host\textsuperscript{18}. Their ability to do so is determined by compatibility between viral structures, host cell receptors, and host immunity\textsuperscript{6}. Because closely related species share both ecological and immunological traits through identity by descent, phylogeny is a strong predictor of pathogen sharing\textsuperscript{17,22}, as well as susceptibility to invasion by new viruses\textsuperscript{23,24,25}. In a changing world, these factors should continue to mediate the impact of ecosystem turnover on the mammalian virome.

Although several studies have mapped current hotspots of emerging diseases\textsuperscript{3,26,27}, few have modeled them in the context of global change. With the global reassortment of animal biodiversity, it is unknown whether bats and rodents will still play a central role in viral emergence\textsuperscript{3,28} (ED Figure 1), or whether hotspots of viral emergence will stay in tropical rainforests\textsuperscript{27,29} which currently harbor most undiscovered viruses\textsuperscript{3,30}. Here, by projecting geographic range shifts and applying fundamental biological rules...
for cross-species transmission, we predicted how and where global change could create novel opportunities for viral sharing. We built species distribution models for 3,870 mammal species, and projected geographic range shifts based on four paired scenarios of climate change (representative concentration pathways, RCPs) and land use change (shared socioeconomic pathways, SSPs) by 2070. We treated dispersal potential as an additional layer of biological realism, inferring these limits for species based on allometric scaling, and compared predictions with and without dispersal constraints. We used these projections to identify where novel range overlap among unfamiliar species (“first encounters”) could happen, and used a recently-developed model to predict the probability of viral sharing based on geographic overlap and host phylogenetic similarity. (ED Figure 2) This model framework allows powerful inference based on the ∼1% of the global mammalian virome that has been described. Using this approach, we tested the hypothesis that environmental change should drive biotic homogenization of mammal communities, exposing mammals to novel viruses, and altering the structure of mammal-virus interactions.

Most mammals are projected to undergo rapid range shifts in the next half century. If range shifts can keep pace with the velocity of climate change, we predict that the vast majority of mammal species (89%–98%) will overlap with at least one unfamiliar species somewhere in their future range, regardless of emissions scenario. At the global level, community turnover would permit almost 300,000 novel species interactions (ED Figure 3). These “first encounters” between mammal species will occur everywhere in the world, but are concentrated in tropical Africa and southeast Asia (ED Figure 4). This result was surprising, and counter to our expectation that species might aggregate at higher latitudes, given that most research has focused on poleward range shifts, and previous work has anticipated a link between climate change, range shifts, and parasite host-switching in the Arctic. However, our findings show that communities tend to shift along latitudinal gradients together, with species rarely encountering new conspecifics. In contrast, species will track thermal optima along elevational gradients and aggregate in novel combinations in mountain ranges, especially in tropical areas with the highest baseline diversity, matching prior predictions.

This global re-organization of mammal assemblages is projected to dramatically impact the structure of the mammalian virome. Accounting for geographic opportunity and phylogenetic compatibility, we projected that a total of 279,427 first encounters in RCP 2.6 would lead to nearly 12,000 novel sharing events. Assuming that spillover will be localized to areas of novel 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 (Figure 1A). If species survive a changing climate by aggregating in high elevation refugia, this suggests emerging viruses may be an increasing problem for...
their conservation\textsuperscript{40,41}. Across scenarios, the spatial signal of expected sharing events is nearly identical, and dominated more by the extent of range shifts than by underlying community phylogenetic structure (ED Figure 5); at least in our framework, opportunity drives spatial patterns more than compatibility.

Species’ dispersal capacity is likely to constrain range shifts, and therefore to limit novel viral exchange. We limited the dispersal potential of flightless species further to the restrictions placed on the SDM projections, based on an established allometric scaling with body size, trophic rank, and generation time\textsuperscript{42}. Dispersal limits caused significant reductions in range expansions across all scenarios, especially warmer ones, and therefore drove a reduction in novel interactions. Even in RCP 2.6 (the mildest scenario), limiting dispersal reduced the number of first encounters by 60%, and reduced the associated viral sharing events by 69%—to a still-staggering 3,600–3,800 projected viral sharing events. Because trophic position and body size determine dispersal capacity, carnivores account for a disproportionate number of first encounters, while ungulates and rodents have slightly fewer first encounters than expected at random (ED Figure 6) Spatial patterns also changed dramatically when dispersal constraints were added, with the majority of first encounters and cross-species viral transmission events occurring in southeast Asia (Figure 1B, ED Figures 4, 5). This viral sharing hotspot is driven disproportionately by bats, because their dispersal was left unconstrained; we made this choice given their exclusion from the original study\textsuperscript{31}, genetic evidence that flight allows bats—and their viruses—to circulate at continental levels\textsuperscript{43,44}, and data suggesting that bat distributions are already undergoing disproportionately rapid shifts\textsuperscript{45}. Bats account for 87% of first encounters after constraining dispersal, and dominate the spatial pattern, with most of their first encounters restricted to southeast Asia (Figure 2).

Bats’ unique capacity for flight could be an important and previously unconsidered link between climate-driven range shifts and future changes in the mammal virome. Even non-migratory bats can regularly travel hundreds of kilometers 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\textsuperscript{46}. This unique dispersal capacity has inevitable epidemiological implications, with recent evidence suggesting that continental panmixia may be common for zoonotic reservoirs, and allow viral circulation at comparable scales\textsuperscript{43,44,47}. We found that a staggering number of studies have also identified ongoing rapid range expansions in bat species around the world\textsuperscript{45,48,49,50,51,52,53,54,55}, with little mention in the broader climate change or emerging disease literature. If flight does allow bats to undergo more rapid range shifts than other mammals, we expect they should drive the majority of novel cross-species viral transmission, and likely bring zoonotic viruses into new regions. This could add an important new dimension to ongoing debate about whether bats are “special”...
due to their higher viral richness, higher proportion of zoonotic viruses, and potentially unique immune adaptations. More broadly, climate-driven changes in the mammalian virome are likely to cascade in future emergence of zoonotic viruses. Among the tens of thousands of expected viral host jumps, some of the highest-risk zoonoses or potential zoonoses are likely to find new hosts. This may pose a threat to human health down the road: the same general rules for cross-species transmission explain spillover patterns for emerging zoonoses, and the viral species that make successful jumps across wildlife species have the highest propensity for zoonotic emergence. Just as simian immunodeficiency virus emergence in chimpanzees and gorillas facilitated the origin of HIV, or SARS-CoV spillover into civets allowed a bat virus to reach humans, these wildlife-to-wildlife host jumps may be evolutionary stepping stones for the ~10,000 to 600,000 potentially zoonotic viruses that are currently circulating in mammal hosts.

To illustrate this problem, we constructed a sub-network of 13 possible Zaire ebolavirus hosts in Africa, and projected possible first encounters involving these species (Figure 3A-C). We project these 13 species to encounter 3,604 new mammals in RCP 2.6, with a modest reduction to 2,586 species by dispersal limits. These first encounters are predicted to produce 87 new viral sharing events that might include ZEBOV, and which cover a much broader part of Africa than the current zoonotic niche of Ebola. Human spillover risk aside, this could expose several new wildlife species to a deadly virus, historically responsible for sizable primate die-offs. Moreover, for zoonoses like Zaire ebolavirus without known reservoirs, future host jumps would only complicate urgent efforts to trace the source of spillover and anticipate future emergences. Ebola is far from unique: with 5,762–11,122 first encounters between bats and primates alone leading to an expected 57–181 new viral sharing events across scenarios (Figure 3D), many potential zoonoses are likely to experience new evolutionary opportunities because of climate change.

Future hotspots of novel assemblages and viral evolution are projected to coincide areas of high human population density, further increasing vulnerability to potential zoonoses. First encounters are disproportionately likely to occur in areas that are projected to be either human settled or used as cropland, and surprisingly less likely to occur in forests, which current literature highlights as producing most emerging diseases (Figure 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. Regardless of mechanism, we predict that tropical hotspots of novel viral sharing will broadly coincide with high population density areas in 2070, especially in the Sahel, the Ethiopian highlands and the Rift Valley, India, eastern China, Indonesia, and the Philippines (Figure 4). Some
European population centers also land in these hotspots; recent emergences in this region like 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 these areas are likely to be the most vulnerable.

Whereas most studies agree that climate change mitigation through reducing greenhouse gas emissions will prevent extinctions and minimize harmful ecosystem impacts, our results suggest that mitigation cannot reduce the likelihood of climate-driven viral sharing. Instead, the mildest, slowest scenarios for biotic homogenization appear likely to produce the most cross-species viral transmission: when climate velocity is lowest, species can successfully track shifting climate optima, leading to more range expansion, and more first encounters. Accounting for dispersal limits, species gained an average of 75% range in the mildest pathway (RCP 2.6); in comparison, only 28% of species experienced a net expansion in the most extreme pathway (RCP 8.5), for an average of 21% range gain. (ED Figure 3A) In fact, in the warmest scenario, up to 326 species lost their entire range, with 168 attributable to dispersal limits alone. As a result, there were 5% fewer first encounters in RCP 8.5 compared to RCP 2.6, and unexpectedly, a 2% reduction in the connectivity of the future global sharing network. (ED Figure 3B,D)

Overall, our results indicate that a mild perturbation of the climate system could create thousands of new eco-evolutionary opportunities for viruses. We caution that this does not imply a possible upside to catastrophic warming, which will be accompanied by mass defaunation, devastating disease emergence, and unprecedented levels of human displacement and global instability. Rather, our results highlight the urgency of better wildlife surveillance systems and health infrastructure as a form of climate change adaptation, even if mitigation efforts are successful and global temperatures stay under +2°C.

Our study establishes a macroecological link between climate change and cross-species viral transmission. In practice, the patterns we describe are likely to be complicated by several ecological factors, including the temperature sensitivity of viral host jumps; the possibility that defaunation especially at low elevations might interact with disease prevalence through biodiversity dilution and amplification effects, not captured by our models; or temporal heterogeneity in exposure (hosts might exchange viruses in passing but not overlap by 2070, especially in warmer scenarios). Future work can also expand the scope of our findings to other host-parasite systems; our novel approach, which combines viral sharing models with massive species distribution modeling pipelines, is readily applied to other datasets. Birds have the best documented virome after mammals, and changing migration targets in a warming world may be especially important targets for prediction. With amphibians facing disproportionately high extinction rates due to a global fungal panzootic, and emerging threats like ranavirus causing conservation
concern, viral exchange among amphibians may be especially important information for conservation practitioners\textsuperscript{70}. Finally, marine mammals are an important target given their exclusion here, especially after a recent study implicating reduced Arctic sea ice in viral sharing among sympatric pinnipeds and sea otters—a result that may be the first proof of concept for our proposed climate-disease link\textsuperscript{71}.

Because hotspots of cross-species transmission are predictable, our study provides the first template for how surveillance could target future hotspots of viral emergence in wildlife. In the next decade alone, over a billion dollars could be spent on a proposed global effort to identify zoonotic threats before they spread from wildlife reservoirs into human populations\textsuperscript{2}. These efforts are being undertaken during the greatest period of global ecological change recorded in human history, and in a practical sense, the rapid movement of species and formation of no-analog communities poses an unexpected challenge for virological research. While 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 mammal species. Tracking spillover into humans is paramount, but so is tracking of viral sharing in wildlife, and targeting surveillance in hotspots of future sharing may help researchers identify host jumps early on.
Methods

In this study, we develop global maps for terrestrial mammals that model their eco-
logical niche as a function of climate and habitat use. We project these into paired
climate-land use futures for 2070, with dispersal limitations set by biological constraints
for each species. We predict the probability of viral sharing among species pairs us-
ing a model of the mammalian viral sharing network that is trained on phylogenetic
relatedness and current geographic range overlaps. With that model, we map the pro-
jected hotspots of new viral sharing in different futures. All analysis code is available at
github.com/cjcarlson/iceberg.

Mapping species distributions

We developed species distribution models for a total of 3,870 species in this study, divided
into two modeling pipelines based on data availability (ED Figures 8, 9).

Data Collection

We scraped the Global Biodiversity Informatics Facility (GBIF) for mammal occurrence
records, and developed species distribution models for all 3,870 species with at least 3
unique terrestrial presence records on a 25 km by 25 km grid (one unique point per grid
cell). This grain was chosen based on the availability of future land use projections (see
below). Spatial and environmental outliers were removed based on Grubb outlier tests
(p-value of 1e-3)\textsuperscript{72}.

Poisson point process models

For 3,088 species with at least 10 unique presence records, Poisson point process models
(closely related to Maxent) were fit using regularized downweighted Poisson regression\textsuperscript{73}
with 20,000 background points fit with the R package \texttt{glmnet}\textsuperscript{74,75,74}. The spatial do-
main of predictions was chosen based on the continent(s) where a species occurred in
their IUCN range map. We trained species distribution models on current climate data
using the WorldClim 2 data set\textsuperscript{76}, using mean annual temperature, mean diurnal temper-
ature range, annual precipitation, precipitation seasonality, and precipitation in warmest
quarter/ (precipitation in warmest quarter + precipitation in coldest quarter). These
predictors were chosen based on having global correlations <0.7 among one another.
These candidate predictors were further filtered on a species-by-species basis, retaining
the maximum number of predictors with correlation <0.7 within the domain where the
model was fit.
Models were fit with 5-fold cross validation, where folds were assigned based on spatial clusters to remove the influence of spatial autocorrelation on cross-validation performance statistics. Linear (all species), quadratic (species with >100 records), and product (species with >200 records) features were used. The regularization parameter was determined based on 5-fold cross-validation with each fold, choosing a value 1 standard deviation below the minimum deviance. This resulted in five models per species which were then combined in an unweighted ensemble. Continuous predictions of the ensemble were converted to binary presence/absence predictions by choosing a threshold based on the 5th percentile of the ensemble predictions at training presence locations.

When models were projected into the future, we limited extrapolation to 1 standard deviation beyond the data range of 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 likely biologically unrealistic) responses to climate.

Range bagging models

For an additional 783 rare species (3 to 9 unique points on the 25 km grid), we produced species distribution models with a simpler range bagging algorithm, a stochastic hull-based method that can estimate climate niches from an ensemble of underfit models\textsuperscript{78,79}, and 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. Based on general guidelines in\textsuperscript{78} 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. Upon visual inspection, this generally lead 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

PPM models performed well, with a mean test AUC under 5 fold cross-validation (using spatial clustering to reduce inflation) of 0.77 (s.d. 0.13). The mean partial AUC evaluated over a range of sensitivity relevant for SDM (0.8-0.95) was 0.8 (s.d. 0.08). The mean sensitivity of binary maps used to assess range overlap (based on the 5% training threshold used to make a binary map) was 0.89 (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 much we inferred the range for these species is to quantify how much of the range was estimated based on our models, based on the number of (10km) 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 253 (s.d. 448); however, the distribution is highly right skewed with a median of 94. This indicates that the range bagging models were typically relatively conservative about inferring ranges for poorly sampled species.

Habitat range and land use

We used the Land Use Harmonization version 2.0 (LUH2) gridded dataset to capture global patterns in land cover for the present and future\textsuperscript{80}. These data are derived from an integrative assessment model that pairs land use scenarios with representative concentration pathways. For the current models, we used historical land-use maps (LUH2 v2h), which are intended for use over the period 850 to 2015 C.E.\textsuperscript{81}. To capture species’ habitat preference, we collated data for all 3,870 mammal species from the IUCN Habitat Classification Scheme version 3.1. We then mapped 104 unique IUCN habitat classifications onto the eight land use types present in the LUH dataset. For 962 species, no habitat data was 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 based on habitat was done conservatively: species were allowed in current and 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.
Refining the dataset

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\textsuperscript{82}. These species were excluded. Aquatic species were removed using a two-filter approach, by cross-referencing with Pantheria\textsuperscript{83}. These results were verified by checking no species only had marine habitat use types (see ‘Habitat range and land use’). We also excluded 246 monotremes and marsupials because the shape of the supertree prevented us from fitting satisfactory GAMM smooths to the phylogeny effect, leaving 3,139 non-marine Eutherian mammals with associated phylogenetic data.

Predicting future species distributions

We modeled a total of 16 possible futures, produced by four paired climate-land use change pathways and 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 possible future distributions based on climate, land use, and dispersal constraints, we aimed to maximize realism; our predictions were congruent with extensive prior literature on climate- and land use-driven range loss\textsuperscript{84,85,86}.

Climate and land use futures

Species distribution models were projected for 2070 using climate models, and then spatially filtered by land use projections. Climate and land-use future pathways are coupled by the Land Use Harmonization 2.0 integrative assessment model\textsuperscript{87,81}, such that every future has a representative concentration pathway (RCP) for climate and a shared socioeconomic pathway (SSP) for land use. For climate we used the HadGEM2 Earth System Model projections for 2070, with the four standard RCPs: 2.6, 4.5, 6.0, and 8.5 (where the values represent added W/m\textsuperscript{2} of solar radiation by the end of the century due to greenhouse gas emissions). These were respectively paired with SSP 1 (‘Sustainability’); SSP 2 (‘Middle of the Road’); SSP 4 (‘Inequality’); and SSP 5 (‘Fossil-Fueled Development’).

These pairings can be thought of as a gradient of scenarios of global change with different levels of severity and sustainability. Not all scenarios are possible; the four we selected are drawn as some of the most representative from an underlying ‘scenario matrix’ that includes every possible parameterization, some of which are entirely incompatible\textsuperscript{88}. (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.) As pairs, SSP-RCP
narratives can be merged to create overall narratives about how global change could look.
For example, in SSP 1-RCP 2.6, a global transition to renewable energy and mitigation of
climate change corresponds to sustainable population growth and economic development.
Driven by international cooperation on climate agreements, afforestation and bioenergy
cropland become major land uses, while tropical deforestation is strongly reduced. In
contrast, in SSP 5-RCP 8.5, business-as-usual development leads to catastrophic levels
of warming, unsustainable population growth and increasing poverty, and massive land
conversion.¹⁸⁻⁹⁰

**Limiting dispersal capacity**

Not all species can disperse to all environments, and not all species have equal disper-
sal capacity—in ways likely to covary with viral sharing properties. We follow a rule
proposed by Schloss et al.³¹, who described an approximate formula for mammal range
shift capacity based on body mass and trophic position. For carnivores, the maximum
distance traveled in a generation is given as \( D = 40.7M^{0.81} \), where \( D \) is distance in kilo-
maters 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, and used the same
cutoff as Schloss 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.⁹² We multiply by 50 years 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,³³ we excluded bats from the
assumed scaling of dispersal limitations. The original study by Schloss et al.³¹ 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 disper-
sal capacity lower than that of the gray dwarf hamster \( (Cricetulus migratorius) \). Bats
were instead given full dispersal in all scenarios: given significant evidence that some bat
species regularly cover continental distances,⁴³,⁴⁴ and that isolation by distance is uncom-
mon within many bats’ ranges,⁴⁶ we felt this was a defensible assumption for modeling
purposes. Moving forward, the rapid range shifts already observed in many bat species
(see 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 likely govern the scaling of bat dispersal, chiefly the aspect ratio (length:width) of wings, which is a strong predictor of population genetic differentiation. Migratory status would also be important to include as a predictor although here, we exclude information on long-distance migration for all species (due to a lack of any real framework for adding that information to species distribution models in the literature).

Explaining spatial patterns

To explore the geography of novel assemblages, we used linear models which predicted the number of first encounters (novel overlap of species pairs) at the 25km 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 meters (derived from the US Geological Service 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; because of the large effect 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 (meters of elevation above the lowest point and species richness). Because some elevation values were lower than 0 (i.e., below sea level), we treated elevation as meters above the lowest terrestrial point rather than meters above sea level to allow us to log-transform the data.

Viral sharing models

Generalized Additive Mixed Models

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

In the original study, a Generalized Additive Mixed Model (GAMM) was used to
predict virus sharing as a binary variable, based on (1) geographic range overlap; (2) phy-
logenetic similarity; and (3) species identity as a multi-membership random effect. The
phylogeographic 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\(^2\)\(^3\), with similarity defined as
the inverse of the cumulative branch length between two species, scaled to between 0
and 1. Geographic overlap was defined as the area of overlap between two species’ IUCN
range maps, divided by their cumulative range size\(^9\)\(^4\).

We first retrained the GAMMs from\(^1\)\(^7\) on the pairwise overlap matrix of species distribu-
tion models generated for this study, so that present predictions would be comparable
with 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\(^1\)\(^7\). To check the performance of the GAMM, we
predicted sharing patterns with a) only random effects, b) only fixed effects, and c) with
both. Although species-level random effects had a mean effect of \(~0\), excluding them en-
tirely resulted in a substantial underestimation of the mean viral sharing rates across the
network (mean sharing \(\approx 0.02\) compared to \(\approx 0.06\)). Therefore to ensure that the model
recapitulated traits of the observed network, we simulated 1,000 binary sharing networks
when predicting with only fixed effects, randomly drawing species-level random effects
in each iteration. The mean sharing value across these iterations closely approximated
observed sharing probability (\(~0.06\)).

Model validation and limits

Compared to 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 Albery et al.\(^1\)\(^7\) (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\(^6\)\(^8\), or increased susceptibility of animals with poorer health in lower-quality
habitat or unfavorable 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 prior hosts and vectors\(^6\)\(^0\)\(^,\)\(^6\)\(^1\).
But data cataloging 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 like 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 geographic 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 geographic and taxonomic patterns of viral sharing would likely emerge. We separately examined patterns of richness, patterns of sharing probability, and their change (i.e., future sharing probability - current sharing probability, giving the expected probability of a novel 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 geographic pixel.

Case study on Zaire ebolavirus

For a case study in possible significant cross-species transmission, we compiled a list of known hosts of Zaire ebolavirus (ZEBOV), a zoonosis with 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, and an additional dataset of filovirus testing in bats. In the latter case, any bats that have been reported antibody positive or PCR-positive for ZEBOV were included. A total of 13 current "known hosts" in Africa were used to predict current possible hosts, and first encounters in all scenarios. We restricted our analysis to Africa because there is no published evidence that Zaire ebolavirus actively circulates outside Africa; although some bat species outside Africa have tested positive for antibodies to ZEBOV, this is likely due to cross-reactivity with other undiscovered filoviruses.

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 version 1.0 for the
year 2070. We aggregated these to native resolution, for each of the four SSP paired with the native RCP/SSP pairing for the species distribution models. In Figure 4 we present the population projections for SSP 1, which pairs with RCP 2.6.
Acknowledgements

This paper is the culmination of several years of idea development and owes special thanks to many people, including the entire Bansal Lab, Laura Ward Alexander, Kevin Burgio, Eric Dougherty, Romain Garnier, Wayne Getz, Peta Hitchens, Christine Johnson, and Isabel Ott. We especially thank Laura Alexander for sharing bat filovirus testing sources used to compile the Ebola sub-network. Thanks are also extended to José Hidasi-Neto for publicly-available data visualization code. This research was supported by the George-town Environment Initiative and the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1639145. C.M. acknowledges funding from National Science Foundation grant DBI-1913673.

Author Contributions

CJC and GFA conceived the study. CM, CJC, and CHT developed species distribution models; GFA, EAE, KJO, and NR developed the generalized additive models. CJC, GFA, and CMZ integrated the predictions of species distributions and viral sharing patterns and designed visualizations. All authors contributed to the writing of the manuscript.
Figure 1: Climate change will drive novel viral sharing among mammal species. The projected number of novel viral sharing events among mammal species in 2070 based on host species geographic range shifts from climate change (RCP 2.6) and land-use change (SSP 1), without dispersal limits (A) and with dispersal limitation (B).
Figure 2: Bats disproportionately drive future novel viral sharing. The spatial pattern of first encounters differs among range-shifting mammal pairs including bat-bat and bat-nonbat encounters (A) and only encounters among non-bats (B). Using a linear model, we show that elevation (C), species richness (D), and land use (E) together explain 57.7% of deviance in new overlaps for bats, and 25.8% for non-bats. Slopes for the elevation effect were generally steeply positive: a $\log_{10}$-increase in elevation was associated with between a 0.4-1.41 $\log_{10}$-increase in first encounters.
Figure 3: **Range expansions will expose naive hosts to zoonotic reservoirs.** (A) The predicted distribution of known African hosts of Zaire ebolavirus. (B) The change in richness of these hosts as a result of range shifts. (C) Projected first encounters with non-Ebola hosts. (D) Bat-primate first encounters are projected to occur globally, producing novel sharing events.
Figure 4: Novel viral sharing events coincide with population centers. In 2070 (RCP 2.6; climate only), human population centers in equatorial Africa, south China and southeast Asia will overlap with projected hotspots of cross-species viral transmission in wildlife. (Both variables are linearly rescaled to 0 to 1.)
Extended Data Figure 1: **The mammal-virus network.** The present-day viral sharing network by mammal order inferred from modeled 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, color is scaled by degree.
Extended Data Figure 2: **Predicted phylogeographic structure of viral sharing.**
Phylogeographic prediction of viral sharing using a generalized additive mixed model. Viral sharing increases as a function of phylogenetic similarity (A) and geographic overlap (B), fit together as a tensor interaction (C). 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 geographic overlap, shown in a hexagonal bin chart for raw data (D).
Extended Data Figure 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 milder RCPs. (B) The number of predicted first encounters was higher in non-dispersal-limited scenarios and in milder RCPs. (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 change is minimal but positive across all scenarios, being greatest in non-dispersal-limited scenarios and in milder RCPs.
Extended Data Figure 4: **Geographic 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.
Extended Data Figure 5: **Geographic distribution of expected viral sharing events from first encounters.** Predictions were carried out for 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 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.
Extended Data Figure 6: **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).
Extended Data Figure 7: **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 viral sharing will occur disproportionately in bats.
Extended Data Figure 8: **Data processing workflow.** Summary of species inclusion across the modeling pipeline for species distributions and viral sharing models. The final analyses in the main text use 3,139 species of Eutherian mammals across all scenarios.
Extended Data Figure 9: **Species distribution modeling workflow for a single species.** A focal species (the European red deer, *Cervus elaphus*) is displayed as an illustrative example. The present day climate prediction (top left) was clipped to the same continent according to the IUCN distribution (top right). This was then clipped according to *Cervus elaphus* land use (second row, left). The known dispersal distance of the red deer was used to buffer the climate distribution (second row, right). The future distribution predictions (RCP 2.6 shown as an example) are displayed in the bottom four panels, for each of the four pipelines: only climate (third row, left); climate + dispersal clip (third row, right); climate + land use clip (bottom row, left) and climate + land use + dispersal clip (bottom row, right). The four distributions clearly display the limiting effect of the dispersal filter (bottom right panels) in reducing the probability of novel species interactions (bottom left panels). The land use clip had little effect on this species as the entire distribution area was habitable for the red deer.
References


