# Methods

## Rodent reservoirs for the New World Arenaviruses (NWA)

A systematic literature search revealed potential host species of arenaviruses that act as reservoirs for spillover of NWA in humans 1–8. Since involvement of certain rodent species as reservoirs for rare NWA are ambiguous or underreported in literature 7, we selected six well-established rodent species that act as reservoirs for NWA and their zoonotic spillover (Table 1). These six species, namely, *Zygodontomys brevicauda* and *Sigmodon alstoni* for GTOV, *Calomys callosus* for MACV, *Calomys musculinus, Calomys laucha* and *Oligoryzomys flavescens* for JUNV, have been established to be associated with outbreaks of their respective NWA in human populations (Table 1). The distribution of these rodent reservoirs over the geographical boundaries of the reported outbreaks of NWA in the past was modeled using species distribution models (SDMs).

Table 1. New World Arenaviruses (NWA) that were reported to cause zoonotic outbreaks in humans with their reported rodent reservoirs

|  |  |  |  |
| --- | --- | --- | --- |
| **Common name of virus** | **Common name of the disease** | **Reported rodent reservoir** | **Outbreaks reported** |
| **Junin virus (JUNV)** | Argentine Hemorrhagic fever | *Calomys musculinus*  *(*Dryland Vesper mouse*)* | 1950s, 1990-2018 |
|  |  | *Calomys laucha*  *(*Small Vesper mouse*)* |  |
|  |  | *Oligoryzomys flavescens*  *(Yellow pygmy rice rat)* |  |
| **Machupo virus**  **(MACV)** | Bolivian Hemorrhagic fever | *Calomys callosus*  *(Large Vesper mouse)* | 1960s |
| **Guanarito virus**  **(GTOV)** | Venezuelan Hemorrhagic fever | *Zygodontomys brevicauda*  *(Short-tailed Cane mouse)* | 1989-2010 |
|  |  | *Sigmodon alstoni*  *(Alston's Cotton rat)* |  |

## Species Distribution Model (SDM)

Modeling frameworks such as Species Distribution Models (SDMs), which can incorporate bioclimatic and environmental factors to estimate the habitats of the rodent reservoirs could be effectively employed in predicting transmission risk for associated diseases 9. Since the modeled association between climate change and species distribution were not expected to be monotonic 10 but rather complex and subtle, modifying the SDM framework to employ ensemble tree-based techniques such as random forest (rf), extra trees (et), extreme gradient boost (xgb) and light gradient boosting machine (lgbm) is warranted 11. We followed a similar protocol to the one we used previously for modeling the species distribution of *Calomys musculinus* in Argentina 12. The protocol was adapted to each of the rodent species selected for this study (Table 1). Here is a summary of the SDM protocol (Figure 6) used to model the rodent reservoirs of NWA following the guidelines established for reporting Ecological Niche Models (ENMs) or SDMs to maximize reproducibility as closely as possible pertaining to our study 13.

### Data

Occurrence data on all six rodent species were sourced from Global Biodiversity Information Facility (GBIF) database (presence-only) 14 using R-package dismo 15 (Derived datasets 16). This data was cleaned to remove duplicates and restricted to years 1990 to the last reported record of occurrence in the GBIF database. All occurrences in the South American continent were retained regardless of the country. Presence-only data was preprocessed to be converted to 1:1 presence-absence data using habitat suitability analysis with R-package *USE::paSampling()* utilizing a uniform approach17. This was done to avoid sample location bias that commonly occurs with random sampling of a geographical extent without considering the environmental conditions as discussed by Da Re et al. (see Supplementary Section C3.4 for details) 18. Environmental and ecological data was sourced from open-source datasets, such as 19 Bioclimatic variables from World Clim (*Bioclim*), Normalized Differential Vegetation Index (*NDVI)* and Digital Elevation Model (*DEM*) from *Moderate-Resolution Imaging Spectroradiometer (MODIS) satellite database and 5 Land Use datasets from Land-Cover and Land-Use Change* (LCLUC) program maintained jointly by NASA and University of Maryland 19–23. These datasets were resampled to the highest common resolution and clipped to the extent of species occurrence data mentioned above for each individual rodent reservoir. In total, three sets of 24 to 26 rasters and six occurrence datasets were used for the SDMs (Supplementary Table S2.1).

Similarly, for predicting the species distribution in the future, analogous raster data for Bioclimactic variables 24, NDVI (data from 2020 for NDVI 20), DEM 21 and land use 25 were downloaded for Shared Socioeconomic Pathways 6 (SSP) scenarios of Moderate (SSP 2-4.5) and Extreme (SSP 5-8.5) predictions for climate change in the future as established by Coupled Model Intercomparison Project v6 (CMIP6) for years 2041 to 2060 26. These datasets were also preprocessed and resampled to the highest common resolution and averaged over the 20-year time period.

### Modeling algorithms

A diagram of different species distribution

AI-generated content may be incorrect.

Figure 6. Schematic illustration of the framework for deriving the impact of climate change on the spillover risk for the three New World Arenaviruses (NWAs) in South America. (Abbrev: ML: Machine Learning, CMIP6: Coupled Model Intercomparison Project phase 6, FOI: Force-of-Infection, H: Human Population/ unit area, R: Rodent Population/ unit area, β: transmission rate parameter)

SDMs were developed by fitting four ensemble tree-based classifier algorithms, namely, random forest (rf), extra trees (et), extreme gradient boost (xgb) and light gradient boosting machine (lgbm). Data was split into train-test ratio of 4:1 (80% train; 20% test) and missing values were imputed using *SimpleImputer()* from *scikit-learn* python library 27. Each algorithm was trained with 5-fold cross validation and the predictions for training and the test set were generated to extract cross-validation accuracy, precision, recall and the F1 score based on (i) 5-fold Cross Validation and (ii) test set confusion matrices. This process was iterated 100 times to add to the robustness of the models using 100 different presence-absence resamples from the occurrence data.

### Hypertuning and Feature selection

Since hyper tuning did not yield significant improvements to any of the four classifier algorithms (results not shown), the default hyper parameters were used. We expected a high level of collinearity between the rasters based on spatial correlation analysis (results not shown), given the nature of geospatial data. To counter this, a recursive feature selection (RFE) was performed in each iteration to improve the accuracy of the model without having potential issues with multicollinearity. In each iteration, 10 features were selected per classifier algorithm based on stratified cross validation of RFE performed before the iterative model training using *RFECV()* from scikit-learn27. Final fitted models were taken as average of all the iterations and of the four classifier algorithms.

### Model fitting, interpolation and projection

Based on the fitted models, the current distribution probabilities of each rodent species were imputed to the geographical extent of their occurrence data using *pyimpute* library in python 28. Similarly, the future distributions of the rodents in response to SSP 2-4.5 and SSP 5-8.5 scenarios of CMIP6 climate change were imputed based on the fitted models. Changes in the distribution probabilities were mapped on the raster for each of the scenarios by subtracting current probabilities from the future probabilities of species distribution. The process is illustrated in Figure 5.

## Zoonotic risk for spillover and force of infection

Based on the distribution probabilities of the SDM for each rodent reservoir, a risk profile for human outbreak, i.e. zoonotic spillover of NWA, was modeled using infection dynamics simulation. The current and CMIP6 scenario-based projections of human population density were overlaid on the SDM probabilities to generate a risk profile based on the force of infection for NWA spillover. Force-of-infection (FOI) was calculated based on the contact rate between humans and the rodent reservoirs and the possibility of the rodent testing positive as reservoir for the NWA.

The mechanistic model used for estimating FOI was a density dependent contact rate model with binomial sampling as detailed in Eq 1.

Eq (1)

Where, is the force of infection defined as the contact rate between susceptible humans and infectious rodents resulting in the successful transmission of infection, is the transmission rate parameter derived from review of analogous viral transmission dynamics studies (see Appendix: Force of Infection), is the population of susceptible humans set at 0.95 times that of total human population in the same geospatial coordinates of SDMs (author’s expertise and from study performed on Lassa Fever in Nigeria 29) and is the infectious proportion of the rodent population based on the binomial sampling between 1 and 15 rodents per grid cell, adjusted with the probability of presence of rodents in the given geospatial coordinates based of the SDMs. The denominator represents the total density of the interacting populations of human hosts and rodent reservoirs. Eq (1) was adapted from a generalized formula used in similar transmission studies of vector and rodent borne infections with and without the effect of climate change 30–32.

A similar methodology was followed using the projected SDM probabilities and the projected human population under the SSP 2-4.5 and SSP 5-8.5 scenarios of CMIP6 33. For generating these maps, the FOI of each reservoir species of the same virus were combined where was the rodent reservoir and .

The FOI of GTOV was formulated as .

The FOI of JUNV was similarly formulated as . The FOI for MACV was .

A similar process was repeated for and .

The differences ( between the future climate change scenarios (SSP 2-4.5 and SSP 5-8.5) and the current scenario of FOI ( and ) were calculated and scaled to be from -100% to +100%.

The map depicting FOI for each of the three viruses under study was further reclassified in 10% quantiles and the geospatial zones which had FOI values in the top 10th percentile (>= 90% of the FOI estimated range across the geographical bounds of the raster) were depicted as potential hotspot zones for outbreak of zoonotic arenaviruses.

## Association of changing risk with climate change

We modeled the associations between the changes in FOI and the changes in 24 bioclimatic and land-use features that were used to predict the species distribution patterns of three NWAs. *NDVI* and *DEM* features were dropped from these analyses due to a high number of missing values in the established resolution (0.042 degrees). The features were converted from raster data to tabular data for current and SSP scenarios. The difference between values was taken for each cell representing the features and the differences were scaled and centered to the mean. A similar process was repeated for the FOI rasters and a dataset representing the changes in the features and FOI was generated. The difference ( between future climate change scenarios (SSP 2-4.5 and SSP 5-8.5) and the current scenario of FOI ( and ) estimates and the features ( were used as inputs in a random forest regressor model for each of the three viruses. A random forest regressor model was fitted to the resulting datasets with and serving as outcomes of the model and and serving as the features.

For each random forest model, the feature importances based on mean impurity reduction were plotted and the top three most important features were extracted for partial dependence plots (*PDPlots*). The random forest models were initially developed using the ranger() package from R 4.1.1 34 and finally fit in *sci-kit learn* python library 27 by using *RandomForestRegressor()* function. *PDPlots* of the top three features for each of the six models (three NWAs X two SSP scenarios) were developed from average grid values and predicted outcomes using functions from scikit-learn, matplotlib and seaborn packages in python 27,35,36.

# Methodology discussion

In our study, we restricted the focus to three NWAs--GTOV, MACV and JUNV--which have all been associated with outbreaks of substantial magnitude 37. This restriction excludes nine other potential NWAs out of 19 isolated and distinctly identified mammarenaviruses 38. Several of these NWAs have been shown to be emergent, such as Sabia virus in Brazil, Chapare virus in Bolivia, Pirital virus and Tacaribe virus in Venezuela and Guyana. But not many cases of these lesser-known NWAs have been reported with a confirmed diagnosis or with acute deaths investigated 37. This is further complicated by inconsistent case reporting caused by socio-economic concerns, lack of detection capabilities, etc. Examples may include Venezuela, which stopped reporting hemorrhagic fever cases in 2008 due to socio-political upheaval, or Bolivia which has not reported cases consistently since the 2017 BHF outbreak due to agricultural concerns and lack of resources 37. These gaps in the reporting could be ameliorated through strategic cooperation by developed countries in terms of public health resources, exchange of knowledge and creating awareness.

We modeled known rodent reservoirs for NWAs, such as *Z. brevicauda* and *S. alstoni* for GTOV, *C.callosus* for MACV and *C. musculinus, C. laucha* and *O. flavescens* for JUNV. However, rodent species that are reservoirs for one virus might also be reservoirs for other similarly transmitted viruses and are thus non-specific reservoirs. For example, reservoirs of VHF caused by Guanarito have also been implicated in carrying viruses from the *Bunyaviridae* family as well as other Hantaviruses 39. The issue of putative rodent reservoirs is further complicated by the contentious classification of rodent species in the New World. Lendino et al. reported that the identification and classification of rodent reservoirs, especially those belonging to *Sigmodontinae* sub-family (New World rats and mice) 37. The continuously expanding species diversity and evolving classification of these species (especially around the Andes foothills 40), combined with different nomenclature and local names, makes it difficult to study as well as target reservoirs species of NWAs. To circumvent this issue, we applied a species distribution model for rodent reservoirs that was not limited to the extent of historically endemic zones of NWAs but rather by the geographical extent of reported suitable? habitat zones of the rodents. A macro-ecological outlook that is reservoir-centric rather than disease outbreak-centric could be more helpful given the lack of clarity on reservoir species diversity and their classification 41. Whether multiple reservoirs exist for the same virus, or the same reservoir for multiple viruses or multiple reservoirs carrying multiple viruses, the risk of zoonotic spillover of NWAs heavily depends on the presence, density and habitat patterns of reservoir hosts 41,42. In further studies, these risks may be predicted with increasing accuracy by combining granular data on human and reservoir populations and their interaction along with climate data.

# References for Methods

1. Milazzo, M. L. et al. Transmission of Guanarito and Pirital Viruses among Wild Rodents, Venezuela - Volume 17, Number 12—December 2011 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid1712.110393.

2. Tesh, R. B. et al. Field studies on the epidemiology of Venezuelan hemorrhagic fever: implication of the cotton rat Sigmodon alstoni as the probable rodent reservoir. Am. J. Trop. Med. Hyg. **49**, 227–235 (1993).

3. Weaver, S. C. et al. Guanarito Virus (Arenaviridae) Isolates from Endemic and Outlying Localities in Venezuela: Sequence Comparisons among and within Strains Isolated from Venezuelan Hemorrhagic Fever Patients and Rodents. Virology **266**, 189–195 (2000).

4. Salazar-Bravo, J., Ruedas, L. A. & Yates, T. L. Mammalian Reservoirs of Arenaviruses. in Arenaviruses I: The Epidemiology, Molecular and Cell Biology of Arenaviruses (ed. Oldstone, M. B. A.) 25–63 (Springer, Berlin, Heidelberg, 2002). doi:10.1007/978-3-642-56029-3\_2.

5. Mills, J. N. et al. A Longitudinal Study of Junin Virus Activity in the Rodent Reservoir of Agrentine Hemorrhagic Fever. Am. J. Trop. Med. Hyg. **47**, 749–763 (1992).

6. Morales, M. A. et al. Evaluation of an enzyme-linked immunosorbent assay for detection of antibodies to Junin virus in rodents. J. Virol. Methods **103**, 57–66 (2002).

7. Salazar-Bravo, J. et al. Natural nidality in Bolivian hemorrhagic fever and the systematics of the reservoir species. Infect. Genet. Evol. **1**, 191–199 (2002).

8. González-Ittig, R. E. et al. Molecular systematics and biogeographic insights of the Calomys callosus complex (Rodentia, Cricetidae). Zool. Scr. **51**, 498–521 (2022).

9. Elith, J. & Leathwick, J. R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annu. Rev. Ecol. Evol. Syst. **40**, 677–697 (2009).

10. Ebi, K. L., Ziska, L. H. & Yohe, G. W. The shape of impacts to come: lessons and opportunities for adaptation from uneven increases in global and regional temperatures. Clim. Change **139**, 341–349 (2016).

11. Predictive performance of presence‐only species distribution models: a benchmark study with reproducible code - Valavi - 2022 - Ecological Monographs - Wiley Online Library. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecm.1486.

12. Flores-Pérez, N., Kulkarni, P., Uhart, M. & Pandit, P. S. Climate Change Impact on Human-Rodent Interfaces: Modeling Junin Virus Reservoir Shifts. EcoHealth (2025) doi:10.1007/s10393-025-01723-z.

13. Feng, X. et al. A checklist for maximizing reproducibility of ecological niche models. Nat. Ecol. Evol. **3**, 1382–1395 (2019).

14. GBIF. What is GBIF? https://www.gbif.org/what-is-gbif.

15. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. dismo: Species Distribution Modeling. 1.3-16 https://doi.org/10.32614/CRAN.package.dismo (2010).

16. Pranav S. Kulkarni (uid: vetpsk). Derived dataset for Kulkarni et al. The Global Biodiversity Information Facility https://doi.org/10.15468/DD.6U8HUK (2025).

17. USE it: Uniformly sampling pseudo‐absences within the environmental space for applications in habitat suitability models - Da Re - 2023 - Methods in Ecology and Evolution - Wiley Online Library. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.14209.

18. Da Re, D. et al. USE it: Uniformly sampling pseudo-absences within the environmental space for applications in habitat suitability models. Methods Ecol. Evol. **14**, 2873–2887 (2023).

19. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. J. R. Meteorol. Soc. **25**, 1965–1978 (2005).

20. Didan, K. MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center https://doi.org/10.5067/MODIS/MOD13A1.061 (2021).

21. Earth Resources Observation And Science (EROS) Center. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. U.S. Geological Survey https://doi.org/10.5066/F7PR7TFT (2017).

22. Potapov, P. et al. The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results. Front. Remote Sens. **3**, 856903 (2022).

23. Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. bioclim : the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. Divers. Distrib. **20**, 1–9 (2014).

24. Presentation and Evaluation of the IPSL‐CM6A‐LR Climate Model - Boucher - 2020 - Journal of Advances in Modeling Earth Systems - Wiley Online Library. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS002010.

25. Hurtt, G. et al. Harmonization of Global Land Use Change and Management for the Period 2015-2300. Earth System Grid Federation https://doi.org/10.22033/ESGF/INPUT4MIPS.10468 (2019).

26. Masson-Delmotte, V., P. Zhai, A. Pirani, S.L., Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K., & Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC\_AR6\_WGI\_SPM\_final.pdf (2021).

27. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. Mach. Learn. PYTHON.

28. Matthew Perry. pyimpute: Utilities for applying scikit-learn to spatial datasets. (2022).

29. Barua, S., Dénes, A. & Ibrahim, M. A. A seasonal model to assess intervention strategies for preventing periodic recurrence of Lassa fever. Heliyon **7**, (2021).

30. Begon, M. et al. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. Epidemiol. Infect. **129**, 147–153 (2002).

31. Medone, P., Ceccarelli, S., Parham, P. E., Figuera, A. & Rabinovich, J. E. The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. Philos. Trans. R. Soc. B Biol. Sci. **370**, 20130560 (2015).

32. Ferrari, M. J., Perkins, S. E., Pomeroy, L. W. & Bjørnstad, O. N. Pathogens, Social Networks, and the Paradox of Transmission Scaling. Interdiscip. Perspect. Infect. Dis. **2011**, 267049 (2011).

33. Jones, B. & O’Neill, B. C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. Environ. Res. Lett. **11**, 084003 (2016).

34. Wright, M. N. & Ziegler, A. **ranger** : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. **77**, (2017).

35. Waskom, M. seaborn: statistical data visualization. J. Open Source Softw. **6**, 3021 (2021).

36. Hunter, J. D. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. **9**, 90–95 (2007).

37. Lendino, A., Castellanos, A. A., Pigott, D. M. & Han, B. A. A review of emerging health threats from zoonotic New World mammarenaviruses. BMC Microbiol. **24**, 115 (2024).

38. Tapia-Ramírez, G. et al. A Review of Mammarenaviruses and Rodent Reservoirs in the Americas. EcoHealth **19**, 22–39 (2022).

39. Rodriguez-Morales, A. J., Castañeda-Hernández, D. M., Escalera-Antezana, J. P. & Alvarado-Arnez, L. E. Organisms of Concern But Not Foodborne or Confirmed Foodborne: Bolivian Hemorrhagic Fever Virus (Machupo Virus)☆. in Reference Module in Food Science (Elsevier, 2019). doi:10.1016/B978-0-08-100596-5.22639-5.

40. Vallejos-Garrido, P. et al. The importance of the Andes in the evolutionary radiation of Sigmodontinae (Rodentia, Cricetidae), the most diverse group of mammals in the Neotropics. Sci. Rep. **13**, 2207 (2023).

41. Han, B. A., O’Regan, S. M., Paul Schmidt, J. & Drake, J. M. Integrating data mining and transmission theory in the ecology of infectious diseases. Ecol. Lett. **23**, 1178–1188 (2020).

42. Dobson, A. Population Dynamics of Pathogens with Multiple Host Species. Am. Nat. **164**, S64–S78 (2004).