Study protocol: Production of Arenavirus and Hantavirus host-pathogen database.

David Simons

Ricardo Rivero

Harry Gordon

Anna Martinez-Checa

David Redding

Stephanie Seifert

2024-09-11

Table of contents

Target journals:

1. Scientific Data unclear if they accept protocols but may be worth a try?
2. Gigabyte?
3. Open Research Europe/F1000 or similar

Word count: ~4,000

# 1. Motivation

Current host-pathogen association datasets provide synthesised information on hosts and their pathogens but do not contain temporal or geographic information. These resources often provide linking information to publications reporting the association but information including accession numbers of archived sequences, number of individuals tested and measures of prevalence among sampled populations are not immediately retrievable. Using these resources for inference beyond host-pathogen associations is therefore limited. Here, we aim to produce a database of host-pathogen associations for two viral families of small mammals which contain several known zoonoses, namely Arenaviridae and Hantaviridae. This database can be used to explore the distribution of small mammal hosts of known and suspected pathogens and the extent to which they have been sampled. Further, linkage to sequence data of known zoonoses will support analysis of the risk of viral reassortment between viral species within geographically co-located host species.

# 2. Introduction

Arenaviruses and Hantaviruses are globally distributed pathogens primarily infecting rodents and other small mammals (Childs and Peters 1993; Jonsson, Figueiredo, and Vapalahti 2010). Several of these pathogens sporadically spillover into human populations where they may be responsible for substantial morbidity and mortality within their endemic regions, but often the human health impact of these pathogens remains unknown (Smith et al. 2024; Vapalahti et al. 2003). Geographically constrained arenaviruses, including *Mammarenavirus lassaense* (LASV) and *Mammarenavirus juninense* (JUNV), cause viral haemorrhagic fevers such as Lassa fever in West Africa and Argentine haemorrhagic fever (Gibb et al. 2017; Gallo, López, and Loureiro 2022). In contrast, Mammarenavirus choriomeningitidis has a global distribution and is known to cause Lymphocytic choriomeningitis, an aseptic meningitis (Vilibic-Cavlek et al. 2021). Hantaviruses, such as *Orthohantavirus puumalaense* (PUUV) and *Orthohantavirus seoulense* (SEOV) cause haemorrhagic fever with renal syndrome, while in the Americas *Orthohantavirus sinnombreense* (SNV) is the causative agent of hantavirus pulmonary syndrome (Makary et al. 2010; Park 2019; Goodfellow et al. 2021). The distribution and human health impact of hantaviruses in Africa is even more poorly understood (Kang et al. 2014).

Novel Arenaviruses and Hantaviruses continue to be described, with new species evolving and emerging in the Anthropocene (Yanagihara et al. 2014; McMahon, Morand, and Gray 2018). Deforestation, urbanization, and agricultural expansion bring humans into closer contact with rodent reservoirs, increasing the likelihood of spillover events (Plowright et al. 2017; Dearing and Dizney 2010). Changes in climate, land use, rodent ecology, and human behaviour alter human-reservoir contact, which is a component of a complex human-animal-environment nexus driving spillover and maintenance of novel pathogens (Keesing et al. 2010; Ecke et al. 2022). Although these pathogens remain endemic diseases with limited human-to-human transmission, there is concern about the potential emergence of novel strains that could sustain human-to-human transmission (i.e., Disease X), leading to LASV’s classification as a priority pathogen by the WHO (Organization et al. 2015).

A broader understanding of the economic and societal impacts of these pathogens is also required. The economic burden of arenavirus and hantavirus infections extends beyond healthcare costs to include long-term productivity losses and the economic strain on already vulnerable populations (Smith et al. 2024). Outbreaks of novel or emerging rodent-borne zoonosis such as HPS during the Four Corners Outbreak (1993) or Yosemite outbreak (2012) can lead to significant social and economic disruption in affected communities, demonstrating the potentially far-reaching consequences of these diseases (Van Hook 2018; Bach 2017). Lasting impacts on an individuals health and wellbeing following infection may also occur. While infection with LASV is only presumed to be symptomatic in 20% of infections, the true scale of symptomatic disease or long term sequelae are unknown, for example sensorineural hearing defects (Ficenec et al. 2020). Difficulty in diagnosing acute arenavirus and hantavirus infections likely means most cases are not detected through passive disease surveillance, likely leading to substantial under-reporting in endemic regions.

Small mammals, primarily rodents (Rodentia), but also shrews (Soricidae) are reservoirs of zoonotic arenaviruses and hantaviruses (Jonsson, Figueiredo, and Vapalahti 2010). Traditionally these pathogens, particularly hantaviruses have been considered single reservoir species but this paradigm is changing with increased pathogen surveillance in endemic settings (Reusken and Heyman 2013). Multiple potential hosts have been identified for pathogens including LASV, SEOV and SNV (Simons et al. 2023; Clement et al. 2019; Goodfellow et al. 2021). Similarly, a single host species may be a host of multiple arenaviruses and hantaviruses throughout its geographic distribution, for example *Mastomys natalensis*, the primary host of LASV is the host of at least 7 distinct arenaviruses throughout its range (De Bellocq et al. 2020).

The ecology of these hosts, including their population dynamics, behaviour, habitat preferences, and interactions within their communities, directly influences pathogen transmission dynamics (Voutilainen et al. 2016). Linking ecological and genomic data allows for the identification of genetic markers associated with increased virulence or transmissibility, providing critical insights into the evolutionary trajectories of arenaviruses and hantaviruses (Kozakiewicz et al. 2018; Näpflin et al. 2019). Small-mammal sampling to quantify these parameters, is vital to better understand ecological factors driving pathogen spillover (Childs and Gordon 2009; Vora et al. 2023). Existing data on rodent-pathogen associations from these studies is typically at a local scale and often lack comprehensive geographic and temporal information which hampers research into the full scope of host-pathogen interactions which requires synthesised and standardised data sources (Simons et al. 2023).

Additionally, human disease data is often reported at national or subnational scales, to preserve patient anonymity, making it challenging to directly link ecological findings with human infection risk. A dataset that includes detailed spatial and temporal data could serve as a critical tool for connecting local ecological data with broader-scale public health patterns, enhancing the ability to identify hotspots of transmission risk and inform targeted interventions.

Research efforts have also been unevenly distributed, with high-risk but understudied regions often lacking sufficient sampling intensity, particularly in Africa and Southeast Asia (Lachish and Murray 2018; Wille, Geoghegan, and Holmes 2021; Simons et al. 2023). Identifying these undersampled areas is essential for guiding future viral discovery and research prioritisation. This requires a comprehensive understanding of global sampling efforts within and between different host and pathogen taxa, helping to pinpoint gaps in knowledge and inform future research priorities.

Local-scale, high-intensity research can provide valuable insights into host-pathogen dynamics, but data fragmentation limits the ability to connect this high-resolution information to broader-scale patterns (i.e., human infection). This impairs efforts to identify larger-scale trends that could inform public health strategies and disease management. For example, a better linkage of ecological and genomic data could facilitate a better understanding of the drivers of viral evolution, reassortment, and the emergence of novel pathogens with zoonotic potential (Arruda et al. 2023). A unified dataset could enable better predictive modelling and outbreak response strategies, addressing current data fragmentation and improving risk assessments and public health interventions.

A detailed dataset on the current state of sampling for arenaviruses and hantaviruses in wild caught small-mammals is therefore urgently needed. Synthesising spatial, temporal and genomic data from previously published small-mammal sampling studies will provide the most comprehensive index of host-pathogen associations for these viral families in addition to identifying locations and potential hosts that have been relatively undersampled and which may benefit from targeted viral discovery. This dataset is expected to provide a key resource for the development of targeted public health interventions, risk mapping and predictive modelling of future outbreaks.

Following Open Science principles, we will share research tools, data extraction methods, and processing code on suitable platforms, ensuring adherence to the FAIR principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al. 2016). The dataset will be accessible to researchers, policymakers, and public health officials worldwide, fostering evidence-based decision-making and international collaboration in disease surveillance. It will be submitted to a suitable data repository (e.g., the Global Biodiversity Information Facility) (Wilkinson et al. 2016; GBIF Secretariat 2024).

# 3. Method

## 3.1 Search strategy

We searched NCBI PubMed and Clarivate Web of Science. We did not use any limits for publication date, language or geographic locations.

*I think I should re-run the search before we submit this. I’m a bit concerned the expanded terms aren’t as good as picking words up as I thought they were.*

Table 1: Search results **needs updating**

| Search term | PubMed | Web of Science |
| --- | --- | --- |
| 1. rodent\* | 175,084 | 175,258 |
| 2. shrew | - | - |
| 3. arenavir\* | 2,506 | 1,941 |
| 4. hantavir\* | 4,274 | 5,016 |
| 5. 1 OR 2 | - | - |
| 6. 2 OR 3 | 6,689 | 6,809 |
| 7. 1 AND 4 | 1,956 | 2,064 |

Citations returned by the searches were downloaded and de-duplicated by matching on titles, authors, publication identifiers or digital object identifiers. Search terms 1-4 included expanded terms. The initial search (conducted 2023-10-16) resulted in 2,755 unique publications. These citations were uploaded to the Rayyan platform to assess against inclusion and exclusion criteria (Ouzzani et al. 2016).

## 3.2 Study screening

We screened studies against the following inclusion criteria. Studies were included if they reported:

1. Rodent OR Shrew sampling AND
2. Direct or indirect detection of microorganisms in the Arenaviridae and Hantaviridae families AND
3. Information on the sampling location of small mammals

Studies were excluded if they did not contain information on the host species of the sample for direct or indirect detection of Arenaviridae or Hantaviridae. Studies reporting experimental infections or solely human infections were also excluded. Studies reported as abstract or conference proceedings were also excluded if they did not provide sufficient description of the method of animal sampling or microorganism detection.

Direct detection was defined as PCR or culture of virus and indirect detection was defined as detection of antibodies against an Arenaviridae or Hantaviridae.

We first screened study title and abstracts, for manuscripts not published in English we used the search datasets translation of titles and abstracts to assess against the inclusion and exclusion criteria. Assessments were made by two members of the study team, studies deemed to meet inclusion criteria by at least one member proceeded to full text review. For articles published in languages other than English automated translation services (i.e., Google Translate) were used prior to assessing the full text manuscript for inclusion.

Reference chaining was performed on studies considered for inclusion at the full text screening stage and relevant publications were added as manually identified relevant articles. The PRISMA flow chart for the search process is shown in Figure [Figure 1](#fig-prisma), we have indicated the the current status of the *pilot* data extraction. We intend to re-run the search when data extraction from the current search has been completed to capture more recently published articles.

|  |
| --- |
| **Figure** 1: PRISMA flow diagram for records identified in the initial search run on 2023-10-16. Studies excluded at full text review included those with incomplete data on one or multiple of pathogen, host or location data. Studies excluded for duplicated data were revisited to ensure the study containing the highest resolution data was retained. |

## 3.3 Data extraction

We aimed to extract included study meta-data and to produce three linked datasets that could used to assess sampling for a) small-mammal hosts, b) viral prevalence and c) genetic variability. We have developed and refined the current data extraction tools through data extraction from 10% of the studies included at title and abstract stage (917 articles). Progress towards final dataset development will be shared through a dedicated RShiny web-application (Simons 2024).

### 3.3.1 Included studies

Included study meta-data was abstracted to ensure appropriate attribution for the underlying data that would be extracted from the study. Each included study was provided an internal study\_id unique identifier which provides linkage to the full text manuscript citation. We extracted reported sampling effort, where available. For articles not reporting sampling effort at the study level we inferred total sampling effort from individual trapping session effort. Studies not reporting any measure of sampling effort we recorded a value of not reported. Finally, we record the level of data aggregation of rodent or pathogen sampling, classified as individual (data were provided at the level of individual small-mammals) or summarised (data were aggregated, e.g., at site or sampling session level).

Table 2: Study information extraction sheet

| Column name | Description |
| --- | --- |
| study\_id | An internal unique identifier for the included study |
| full\_text\_id | The internal ID for the full text manuscript and its associated citation |
| datasetName | The title of the manuscript, report or book section |
| sampling\_effort | A free text entry to capture the effort of sampling, ideally in trap nights for rodent studies |
| data\_access | Whether data is available for individual small mammals or whether it is aggregated |
| data\_resolution | The level of aggregation available |
| linked\_manuscripts | The DOI or weblink to other studies including the same dataset either in its entirety or a subset, this will be used to attempt to de-duplicate data |
| notes | Details that may be important for interpreting the extracted data |

### 3.3.2 Small-mammal sampling

Data is extracted into a rodent sheet. A record is produced at the highest level of temporal or spatial resolution for detected small-mammals. For example a study reporting on the detection of a single species at four spatially distinct sampling sites would be associated with four records. If this same study reported detections from four distinct trapping sessions at each of these sites there would be 16 records associated with the study. Studies presenting data on *N* individual rodents will have *N* records with each record associated with the detection of a single rodent.

Identification of a detected small mammal was recorded as reported within the article, we will not systematically assess the method of identification and therefore assume accurate identification as reported within the articles. Where identification is to genus or higher taxonomic level we report to this level (i.e., *Rattus spp.*, Rodentia). See data processing below for the method used to harmonise of species names.

Included studies were expected to variably report location data associated with sampling. We extract locality as the highest resolution of location data that could be matched to administrative levels within a country (e.g., city, county), verbatimLocality is used where higher resolution spatial data is available which may not map to administrative levels (e.g., Site A).

Table 3: Rodent information extraction sheet

| Column name | Description |
| --- | --- |
| rodent\_record\_id | An internal unique identifier for the rodent record, the resolution of a record is dependent on the level of aggregation reported in the study |
| study\_id | A unique identifier to link a record to the study from which it originated |
| scientificName | The species name of the small-mammal as reported in the study, ideally binomial names |
| eventDate | The period in which sampling was conducted |
| locality | The location of sampling effort, reported to the highest spatial resolution available |
| country | Country where trapping occurred, for multinational studies where counts are not disaggregated by country, all countries sampled are reported |
| verbatimLocality | High level habitat type will be recorded here at the scale for which trapping is recorded, particularly useful when locality does not discriminate between multiple sampling sites |
| coordinate\_resolution | The description of coordinate levels provided in the study |
| decimalLatitude | Latitude in decimal format, converted from coordinates reported as required |
| decimalLongitude | Longitude in decimal format, converted from coordinates reported as required |
| individualCount | The number of detected individuals associated with a record |
| trapEffort | Trap effort, recorded in trap nights associated with a record |
| trapEffortResolution | The resolution of trapping effort for the record |

### 3.3.3 Pathogen sampling

Assays conducted for arenaviruses and hantavrisues are extracted in the pathogen sheet. A record is produced at the highest resolution of pathogen sampling. Each pathogen record is associated with a rodent record, one-to-many matching at this stage is possible. For example, a single rodent may be tested for a pathogen using both antibody and PCR based assays, hence a single sample may be associated with multiple pathogen records.

The species or family targeted by the assay was recorded as reported within the relevant article, we did not make any assessment of the suitability, sensitivity or specificity of an assay. We recorded the number of samples tested using the assay for a specific pathogen. This may differ from the individual count of the small-mammals sampled in the associated record as not all samples may not have been suitable for testing, or the study authors may have decided to subset available samples. The number of positive and negative samples are related to this tested number. Where reported we extracted the number of inconclusive samples. Similarly to above, we recorded positives and negatives as reported by study authors and make no assessment of these classifications.

Table 4: Pathogen information extraction sheet

| Column name | Description |
| --- | --- |
| pathogen\_record\_id | A unique identifier for the group of samples from the same rodent species, at a specific location or timepoint, tested for the same pathogen using the same method |
| associated\_rodent\_record\_id |  |
| study\_id | A unique identifier to link a study to the descriptive sheet entry for that study |
| associatedTaxa | The scientificName of the rodent from the associated\_rodent\_record\_id |
| family | The family of the pathogen, either Arenaviridae or Hantaviridae |
| scientificName | The species name of the pathogen assayed for, if available. Some assays are not specific to a species. |
| assay | Whether the assay is attempting to detect antibody, direct detection of virus (i.e. pcr), or other |
| number\_tested | The number of distinct samples tested |
| number\_negative | The number of reported negative samples |
| number\_positive | The number of reported positive samples |
| number\_inconclusive | The number of samples with inconclusive results |
| note | Notes relevant to this record |

### 3.3.4 Sequences

If studies include complete or partial sequences of hosts or viruses archived in NCBI they will be linked through the sequences sheet. A record will be produced for each accession available. A sequence\_record\_id will be associated with each accession and depending on whether the sequence relates to a pathogen or host each record will be associated with one or both of these. Many-to-one matching of sequences may occur for several reasons, first, hantaviruses and arenaviruses contain multiple segments and so three or two records respectively will be produced for each acutely infected small mammal. Second, reporting of pathogen sampling may be aggregated with multiple sequences obtained from a single reported assay. Pathogen sequences from human infections will be extracted but will only be linked at study level.

Table 5: Pathogen sequences extraction sheet

| Column name | Description |
| --- | --- |
| sequence\_record\_id | A unique identifier for the sequence record |
| sequenceType | One of Pathogen or Host |
| associated\_pathogen\_record\_id | An associated pathogen record for this sequence |
| associated\_rodent\_record\_id | An associated rodent record for this sequence |
| study\_id | Study ID associated with this sequence |
| associatedTaxa | Species name of small mammal sampled/sequenced |
| scientificName | Species name of pathogen sequenced |
| accession\_number | NCBI nucleotide accession |
| note | Notes relevant to this record |

We are not integrating NCBI with our dataset beyond incorporating the NCBI accession. We will develop a method and share code to integrate the enhanced metadata produced with current metadata available in the GenBank NCBI format.

## 3.4 Data processing

We describe the processes in place to clean and harmonise extracted data below. These processes have been developed on a pilot sample of 10% of included articles. All data processing will be conducted in R with scripts retained in a version controlled git repository (R Core Team 2023; Simons and Seifert 2024). Raw data will be downloaded from Google Sheets using the googledrive API in R, with date stamped files stored locally (D’Agostino McGowan and Bryan 2023).

### 3.4.1 Species name harmonisation

Small mammal and rodent taxonomy has and will continue to change over time. To standardise reported species we will match reported names to both the GBIF backbone taxonomy and the NCBI taxonomy database (GBIF Secretariat 2024; Schoch et al. 2020). The taxize R package is used to query the APIs of these platforms (Scott Chamberlain and Eduard Szocs 2013). Unmatched names are retained as reported by study authors, corrected to a matching name or allocated to a higher taxa (e.g., “*Mus/Nannomys*” becoming identified at the genus level as *Mus*).

### 3.4.2 Location of sampling

For studies not reporting geographic coordinates of sampling but including some information that describes the location (i.e., the name of the village sampled) we will locate coordinates through several methods. Locations will be searched for in Google Maps, Wikipedia or the Geographic Names Server provided by the National Geospatial-Intelligence Agency USA. Locations that represent administrative divisions will be matched using the Database of Global Administrative Areas (GADM) accessed through the geodata R package (Hijmans et al. 2023).

The coordinate resolution for coordinates labelled, site or study site will be set as 100 meters. For locations associated with an administrative area but no higher resolution data we will use the radius of the administrative region to represent uncertainty in these coordinates.

### 3.4.3 Imputed non-detections

Non-detection of a small mammal species in a location it may be expected to be is of interest to researchers. We will enrich detection only data from included studies by imputing non-detections where they are not given. We restrict this imputation to species that have been detected elsewhere in the study. Imputed non-detections will be labelled as imputed in the final data product.

# 4. Discussion

This novel dataset provides spatially and temporally explicit data on small mammal sampling for arenaviruses and hantaviruses, offering a more comprehensive view of host-pathogen associations than is currently available. By including sampling effort and explicit spatial data, this dataset enhances the ability to quantify sampling biases, assess geographic gaps, and better understand the spatial ecology of these pathogens. The insights gained from this dataset could improve our understanding of how environmental changes, such as habitat fragmentation and urbanisation, influence pathogen dynamics and spillover risks.

Importantly, this resource could serve as a foundational tool for predictive modelling, helping identify areas at heightened risk for emerging zoonosis spillover. The ability to link ecological data with human health outcomes could inform targeted public health interventions, enhancing outbreak preparedness and response. For instance, identifying hotspots of high pathogen prevalence in rodents could guide targeted development of ecologically-based rodent control measures in regions identified as vulnerable.

Existing host-pathogen datasets lack detailed spatial or temporal information, limiting their use in ecological and epidemiological modelling (Gibb et al. 2021). By addressing these gaps, our dataset provides information about host-pathogen interactions that span multiple spatial and temporal scales. Moreover, the explicit reporting of sampling effort allows for more robust analyses of detection probability, which is crucial for understanding the true distribution of pathogens (Baele et al. 2016; Didelot et al. 2017). Current global host-pathogen datasets are unable to account for sampling biases due to their method of synthesis.

This dataset also offers the unique benefit of explicitly reporting the extent of sampling for pathogens within their host species, providing critical insights into spatial sampling biases and detection efforts. For instance, while it is expected that Mastomys natalensis sampling occurs throughout its range, the detection of specific pathogens such as Lassa virus (LASV) or Morogoro virus may be confined to only a subset of that range. By detailing both presence and absence data, the dataset allows for a more nuanced understanding of pathogen distribution within host species, beyond mere occurrence records. This capability to quantify the detection effort for specific pathogens is vital for assessing spatial sampling biases, identifying under-sampled regions, and refining ecological and epidemiological models.

Despite its strengths, this dataset has several limitations. First, the inherent variability in the quality and detail of reported data from included studies means that some records may lack critical information, such as exact coordinates or specific sampling dates. The imputation of non-detections, introduces assumptions that could affect the interpretation of absence data. Additionally, the dataset is inherently static and will become outdated as new data emerge.

Furthermore, the reliance on published literature introduces publication bias, as studies with significant findings are more likely to be published than those with negative results. This could skew the dataset towards areas and hosts with known or previously detected pathogens, potentially under representing regions where sampling has occurred but without positive detections. We are not routinely contacting study authors to obtain data reported within publications. If data are not available within the article or as supplementary material we are not incorporating it into this dataset. An associated project is currently being planned where we hope to contact individual study authors to obtain, standardise and support submission to available repositories. To address limitations associated with extracting data without access to the raw data we encourage scientists with ongoing field studies and pathogen surveillance programmes, to submit their data to dynamic repositories (e.g., Pharos and GBIF (*The Pathogen Harmonized Observatory (PHAROS)* 2024; GBIF Secretariat 2024)).

Developing predictive models based on this dataset could provide valuable insights into the factors driving zoonotic spillover, enabling targeted surveillance and control measures. Such models could also be used to forecast the emergence of novel strains with epidemic potential, guiding resource allocation and intervention strategies. Beyond research applications, the dataset could serve as a tool for policymakers, helping to identify priority areas for viral discovery and public health interventions.

# 5. Conclusion

Overall, this dataset will be a valuable resource for understanding the ecology of arenaviruses and hantaviruses in their natural reservoirs. By bridging the gap between local-scale ecological studies and broader public health needs, it has the potential to enhance our ability to predict and mitigate the risks posed by these emerging pathogens. Continued efforts to update and expand this resource will be crucial for maintaining its utility in a rapidly changing epidemiological landscape.

# 6. References

Arruda, Liã Bárbara, Hayley Beth Free, David Simons, Rashid Ansumana, Linzy Elton, Najmul Haider, Isobella Honeyborne, et al. 2023. “Current Sampling and Sequencing Biases of Lassa Mammarenavirus Limit Inference from Phylogeography and Molecular Epidemiology in Lassa Fever Endemic Regions.” *PLOS Global Public Health* 3 (11): e0002159.

Bach, Michelle. 2017. “Social Implications of the Yosemite Hantavirus Outbreak.” *Pathways: Stanford Journal of Public Health (SJPH)* 6: 23–26.

Baele, Guy, Marc A. Suchard, Andrew Rambaut, and Philippe Lemey. 2016. “Emerging Concepts of Data Integration in Pathogen Phylodynamics.” *Systematic Biology* 66 (1): e47–65. <https://doi.org/10.1093/sysbio/syw054>.

Childs, James E, and Elizabeth R Gordon. 2009. “Surveillance and Control of Zoonotic Agents Prior to Disease Detection in Humans.” *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine* 76 (5): 421–28.

Childs, James E, and Clarence J Peters. 1993. “Ecology and Epidemiology of Arenaviruses and Their Hosts.” In *The Arenaviridae*, 331–84. Springer.

Clement, Jan, James W LeDuc, Graham Lloyd, Jean-Marc Reynes, Lorraine McElhinney, Marc Van Ranst, and Ho-Wang Lee. 2019. “Wild Rats, Laboratory Rats, Pet Rats: Global Seoul Hantavirus Disease Revisited.” *Viruses* 11 (7): 652.

D’Agostino McGowan, Lucy, and Jennifer Bryan. 2023. *Googledrive: An Interface to Google Drive*. <https://CRAN.R-project.org/package=googledrive>.

De Bellocq, Joëlle Goüy, Anna Bryjová, Aleksey A Martynov, and Leonid A Lavrenchenko. 2020. “Dhati Welel Virus, the Missing Mammarenavirus of the Widespread Mastomys Natalensis.” *Journal of Vertebrate Biology* 69 (2): 20018–11.

Dearing, M Denise, and Laurie Dizney. 2010. “Ecology of Hantavirus in a Changing World.” *Annals of the New York Academy of Sciences* 1195 (1): 99–112.

Didelot, Xavier, Christophe Fraser, Jennifer Gardy, and Caroline Colijn. 2017. “Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks.” *Molecular Biology and Evolution* 34 (4): 997–1007.

Ecke, Frauke, Barbara A Han, Birger Hörnfeldt, Hussein Khalil, Magnus Magnusson, Navinder J Singh, and Richard S Ostfeld. 2022. “Population Fluctuations and Synanthropy Explain Transmission Risk in Rodent-Borne Zoonoses.” *Nature Communications* 13 (1): 7532.

Ficenec, Samuel C, Jeffrey Percak, Sara Arguello, Alison Bays, Augustine Goba, Michael Gbakie, Jeffrey G Shaffer, Susan D Emmett, John S Schieffelin, and Daniel Bausch. 2020. “Lassa Fever Induced Hearing Loss: The Neglected Disability of Hemorrhagic Fever.” *International Journal of Infectious Diseases* 100: 82–87.

Gallo, Giovanna Lucrecia, Nora López, and Marı́a Eugenia Loureiro. 2022. “The Virus–Host Interplay in Junı́n Mammarenavirus Infection.” *Viruses* 14 (6): 1134.

GBIF Secretariat. 2024. “Global Biodiversity Information Facility.” <https://www.gbif.org/>.

Gibb, Rory, Gregory F Albery, Daniel J Becker, Liam Brierley, Ryan Connor, Tad A Dallas, Evan A Eskew, et al. 2021. “Data Proliferation, Reconciliation, and Synthesis in Viral Ecology.” *Bioscience* 71 (11): 1148–56.

Gibb, Rory, Lina M Moses, David W Redding, and Kate E Jones. 2017. “Understanding the Cryptic Nature of Lassa Fever in West Africa.” *Pathogens and Global Health* 111 (6): 276–88.

Goodfellow, Samuel M, Robert A Nofchissey, Kurt C Schwalm, Joseph A Cook, Jonathan L Dunnum, Yan Guo, Chunyan Ye, et al. 2021. “Tracing Transmission of Sin Nombre Virus and Discovery of Infection in Multiple Rodent Species.” *Journal of Virology* 95 (23): 10–1128.

Hijmans, Robert J., Márcia Barbosa, Aniruddha Ghosh, and Alex Mandel. 2023. *Geodata: Download Geographic Data*. <https://CRAN.R-project.org/package=geodata>.

Jonsson, Colleen B, Luiz Tadeu Moraes Figueiredo, and Olli Vapalahti. 2010. “A Global Perspective on Hantavirus Ecology, Epidemiology, and Disease.” *Clinical Microbiology Reviews* 23 (2): 412–41.

Kang, Hae Ji, William T Stanley, Jacob A Esselstyn, Se Hun Gu, and Richard Yanagihara. 2014. “Expanded Host Diversity and Geographic Distribution of Hantaviruses in Sub-Saharan Africa.” *Journal of Virology* 88 (13): 7663–67.

Keesing, Felicia, Lisa K Belden, Peter Daszak, Andrew Dobson, C Drew Harvell, Robert D Holt, Peter Hudson, et al. 2010. “Impacts of Biodiversity on the Emergence and Transmission of Infectious Diseases.” *Nature* 468 (7324): 647–52.

Kozakiewicz, Christopher P, Christopher P Burridge, W Chris Funk, Sue VandeWoude, Meggan E Craft, Kevin R Crooks, Holly B Ernest, Nicholas M Fountain-Jones, and Scott Carver. 2018. “Pathogens in Space: Advancing Understanding of Pathogen Dynamics and Disease Ecology Through Landscape Genetics.” *Evolutionary Applications* 11 (10): 1763–78.

Lachish, Shelly, and Kris A Murray. 2018. “The Certainty of Uncertainty: Potential Sources of Bias and Imprecision in Disease Ecology Studies.” *Frontiers in Veterinary Science* 5: 90.

Makary, P, Mari Kanerva, J Ollgren, MJ Virtanen, O Vapalahti, and O Lyytikäinen. 2010. “Disease Burden of Puumala Virus Infections, 1995–2008.” *Epidemiology & Infection* 138 (10): 1484–92.

McMahon, Barry J, Serge Morand, and Jeremy S Gray. 2018. “Ecosystem Change and Zoonoses in the Anthropocene.” *Zoonoses and Public Health* 65 (7): 755–65.

Näpflin, Kathrin, Emily A O’Connor, Lutz Becks, Staffan Bensch, Vincenzo A Ellis, Nina Hafer-Hahmann, Karin C Harding, et al. 2019. “Genomics of Host-Pathogen Interactions: Challenges and Opportunities Across Ecological and Spatiotemporal Scales.” *PeerJ* 7: e8013.

Organization, World Health et al. 2015. “Blueprint for r&d Preparedness and Response to Public Health Emergencies Due to Highly Infectious Pathogens.” *World Health Organization*.

Ouzzani, Mourad, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. “Rayyan—a Web and Mobile App for Systematic Reviews.” *Systematic Reviews* 5 (1): 210. <https://doi.org/10.1186/s13643-016-0384-4>.

Park, Yohan. 2019. “Epidemiologic Study on Changes in Occurrence of Hemorrhagic Fever with Renal Syndrome in Republic of Korea for 17 Years According to Age Group: 2001–2017.” *BMC Infectious Diseases* 19: 1–8.

Plowright, Raina K, Colin R Parrish, Hamish McCallum, Peter J Hudson, Albert I Ko, Andrea L Graham, and James O Lloyd-Smith. 2017. “Pathways to Zoonotic Spillover.” *Nature Reviews Microbiology* 15 (8): 502–10.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Reusken, Chantal, and Paul Heyman. 2013. “Factors Driving Hantavirus Emergence in Europe.” *Current Opinion in Virology* 3 (1): 92–99.

Schoch, Conrad L, Stacy Ciufo, Mikhail Domrachev, Carol L Hotton, Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe, et al. 2020. “NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools.” *Database* 2020: baaa062.

Scott Chamberlain, and Eduard Szocs. 2013. “Taxize - Taxonomic Search and Retrieval in r.” *F1000Research*. <https://f1000research.com/articles/2-191/v2>.

Simons, David. 2024. “Arenaviruses and Hantaviruses of Rodents.” <https://diddrog11.shinyapps.io/arenavirus_hantavirus_app/>.

Simons, David, Lauren A Attfield, Kate E Jones, Deborah Watson-Jones, and Richard Kock. 2023. “Rodent Trapping Studies as an Overlooked Information Source for Understanding Endemic and Novel Zoonotic Spillover.” *PLOS Neglected Tropical Diseases* 17 (1): e0010772.

Simons, David, and Stephanie N Seifert. 2024. “ArHa: Arenavirus and Hantavirus sampling in small mammals.” <https://github.com/DidDrog11/arenavirus_hantavirus>.

Smith, David RM, Joanne Turner, Patrick Fahr, Lauren A Attfield, Paul R Bessell, Christl A Donnelly, Rory Gibb, et al. 2024. “Health and Economic Impacts of Lassa Vaccination Campaigns in West Africa.” *Nature Medicine*, 1–10.

*The Pathogen Harmonized Observatory (PHAROS)*. 2024. The Viral Emergence Research Initiative. <https://pharos.viralemergence.org/about/>.

Van Hook, Charles J. 2018. “Hantavirus Pulmonary Syndrome—the 25th Anniversary of the Four Corners Outbreak.” *Emerging Infectious Diseases* 24 (11): 2056.

Vapalahti, Olli, Jukka Mustonen, Åke Lundkvist, Heikki Henttonen, Alexander Plyusnin, and Antti Vaheri. 2003. “Hantavirus Infections in Europe.” *The Lancet Infectious Diseases* 3 (10): 653–61.

Vilibic-Cavlek, Tatjana, Vladimir Savic, Thomas Ferenc, Anna Mrzljak, Ljubo Barbic, Maja Bogdanic, Vladimir Stevanovic, Irena Tabain, Ivana Ferencak, and Snjezana Zidovec-Lepej. 2021. “Lymphocytic Choriomeningitis—Emerging Trends of a Neglected Virus: A Narrative Review.” *Tropical Medicine and Infectious Disease* 6 (2): 88.

Vora, Neil M, Lee Hannah, Chris Walzer, Mariana M Vale, Susan Lieberman, Ashley Emerson, Jonathan Jennings, et al. 2023. “Interventions to Reduce Risk for Pathogen Spillover and Early Disease Spread to Prevent Outbreaks, Epidemics, and Pandemics.” *Emerging Infectious Diseases* 29 (3).

Voutilainen, Liina, Eva R Kallio, Jukka Niemimaa, Olli Vapalahti, and Heikki Henttonen. 2016. “Temporal Dynamics of Puumala Hantavirus Infection in Cyclic Populations of Bank Voles.” *Scientific Reports* 6 (1): 21323.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 1–9.

Wille, Michelle, Jemma L Geoghegan, and Edward C Holmes. 2021. “How Accurately Can We Assess Zoonotic Risk?” *PLoS Biology* 19 (4): e3001135.

Yanagihara, Richard, Se Hun Gu, Satoru Arai, Hae Ji Kang, and Jin-Won Song. 2014. “Hantaviruses: Rediscovery and New Beginnings.” *Virus Research* 187: 6–14.