

ArHa: Synthesising Arena- and Hantavirus data from rodents to understand current known host distributions and viral pathogens

David Simons and Steph Seifert

MEZAP lab:

Ricardo Rivero and Grant Rickard

Redding lab:

Dave Redding, Greg Milne, Harry Gordon, Ana Martinez-Checa Guiote

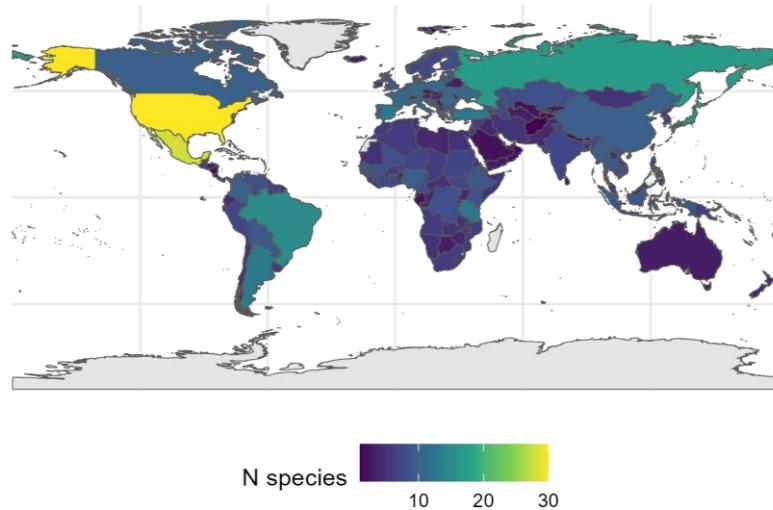


PennState

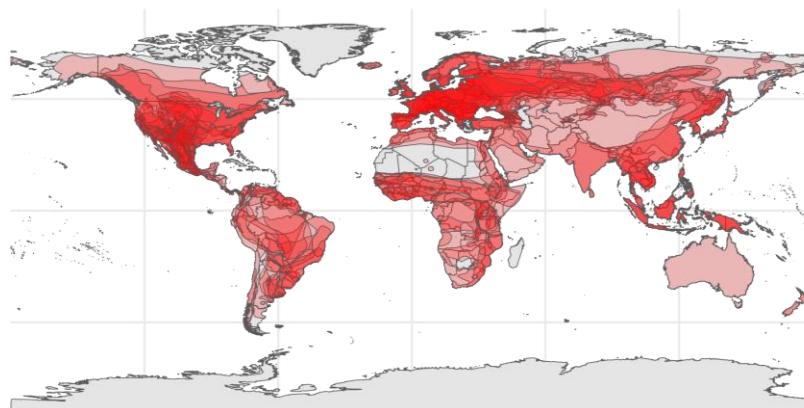
College of the Liberal Arts

Arenaviruses

Number of Arenavirus rodent host species within a country



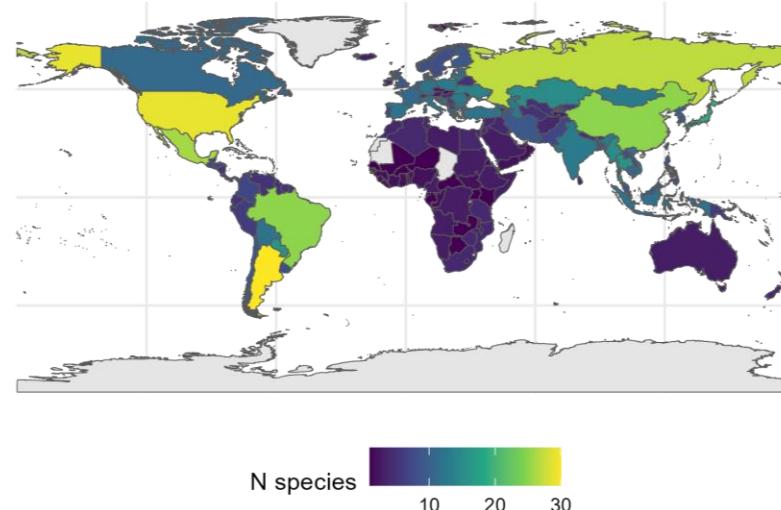
IUCN ranges of Arenavirus rodent host species



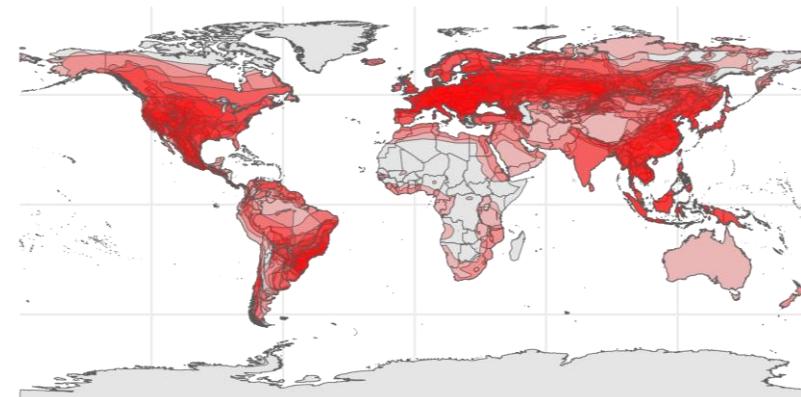
Hantaviruses

Data from CLOVER, Gibb, et al. 2021

Number of Hantavirus rodent host species within a country



IUCN ranges of Hantavirus rodent host species



PennState
College of the Liberal Arts

Motivation

- Current host-pathogen association datasets do not typically contain temporal or geographic information.
- Linkage to specific pathogens or hosts is through referencing underlying publications limiting usability.
- Sampling data, including:
 - Sampling effort
 - Number of individuals assayed
 - Pathogens tested
 - Measures of prevalence within defined populationsAre not immediately retrievable often limiting the inference that can be drawn.

Aims

- Produce a geographically and temporally rich database of small-mammal sampling for Arena- and Hantaviruses.
- Enrich GenBank metadata to link sequences to the hosts from which they originated.
- Explore geographic and temporal sampling biases of potential zoonoses.
- Develop a dataset that can be used to test several hypothesis:
 - To what extent does niche overlap facilitate viral reassortment and evolution?
 - What are the relative contributions of biotic and molecular factors on cross-species transmission?
 - How does pathogen prevalence vary spatially within host ranges (the abundant-centre hypothesis)?
 - Which host traits are associated with greater pathogen prevalence, incorporating sampling bias and sampling locations within host ranges?

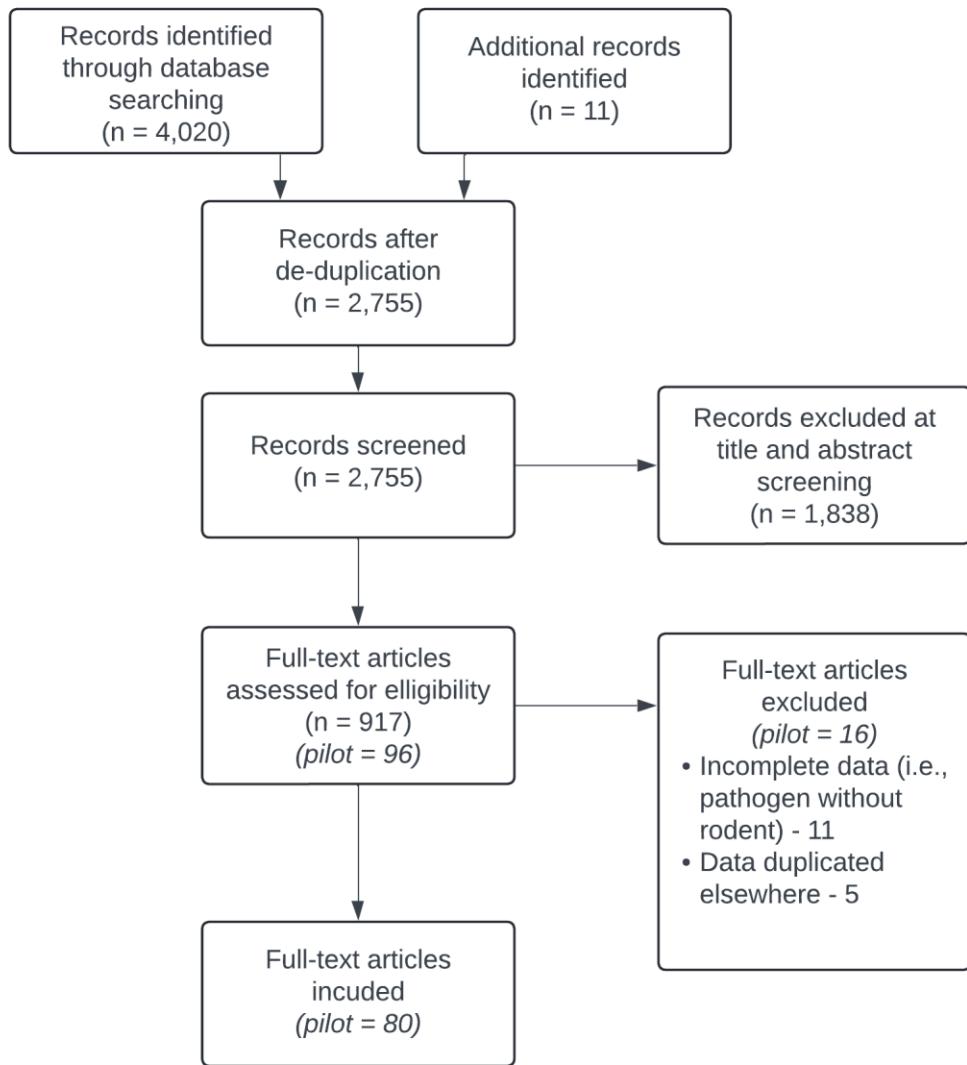
Producing the dataset

Identification

Screening

Eligibility

Included



- Search term:

1. rodent* OR shrew*

AND

2. arenavir* OR hantavir*

- Inclusion criteria:

- Contain data on rodent/shrew sampling AND pathogen sampling (Arena or Hantaviridae)

- 2,755 unique citations

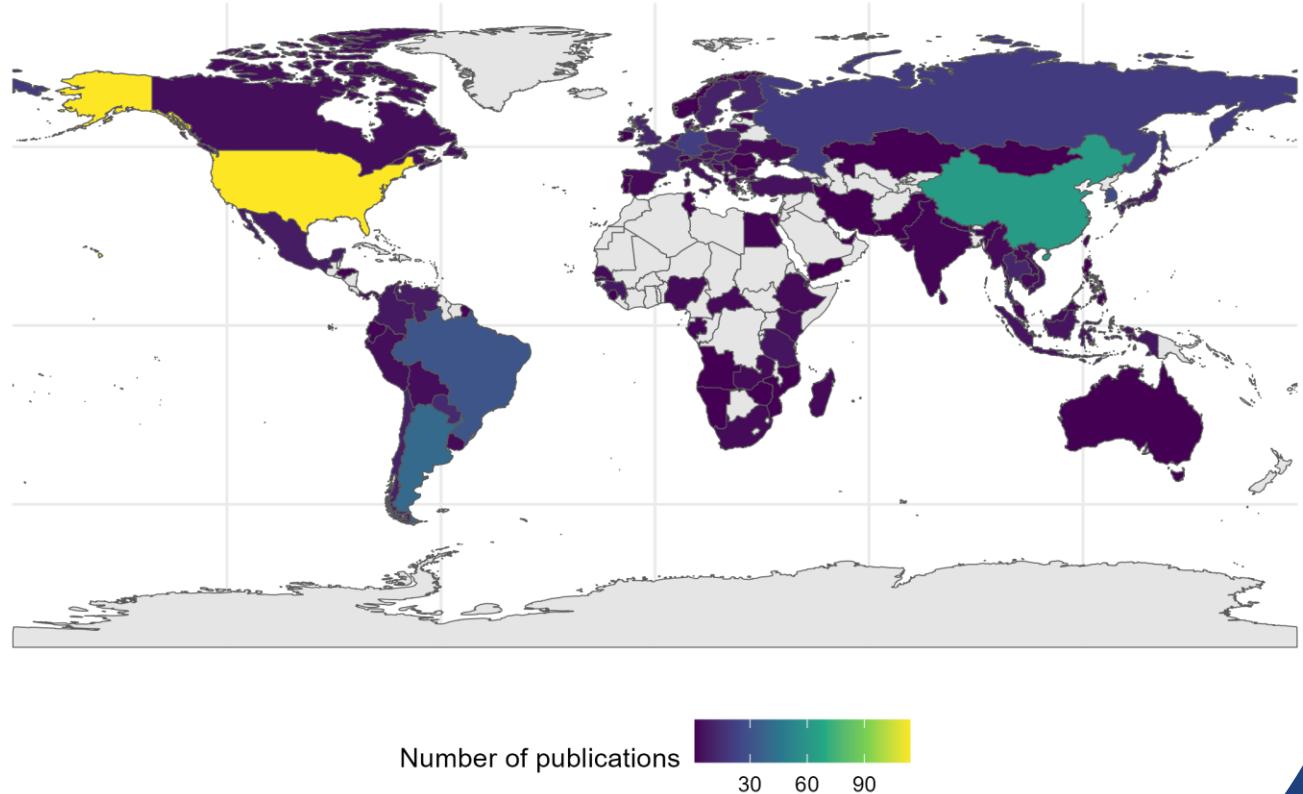
- 1,838 excluded on title and abstracts

- 917 full texts to be screened and data extracted

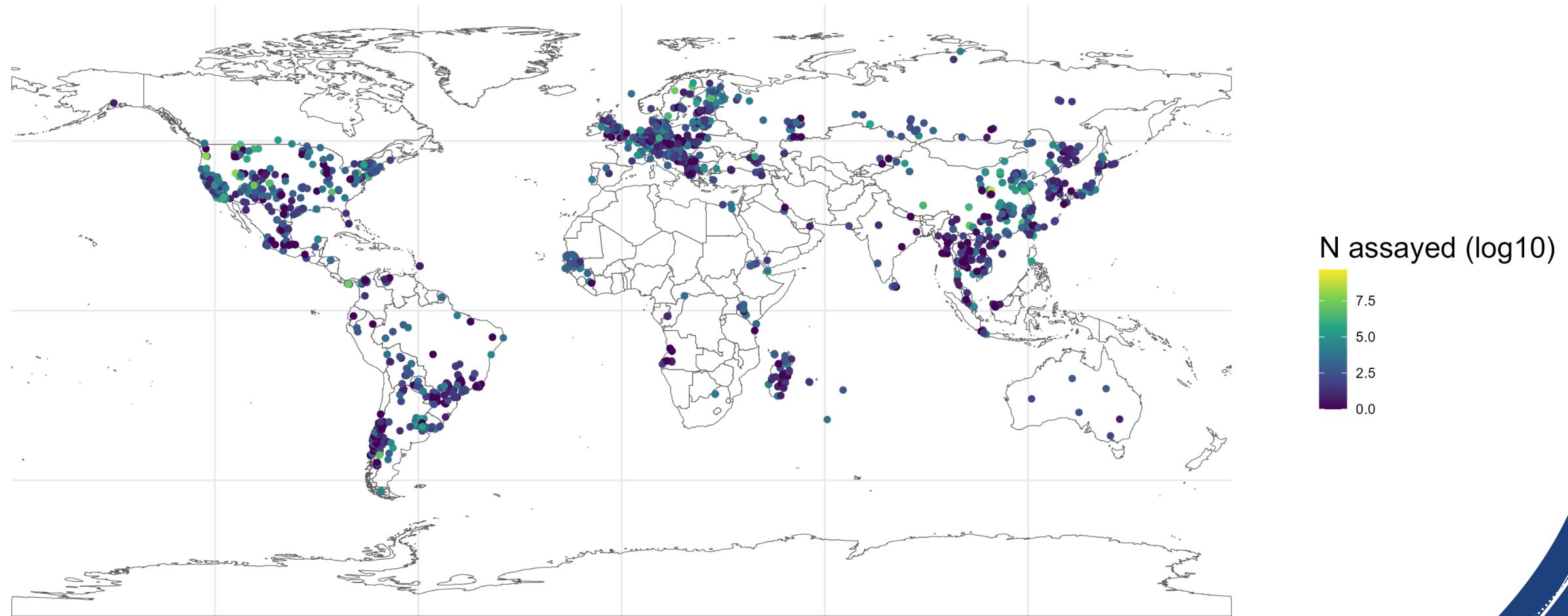
The current dataset

- 148 excluded studies (16%)
- 585 included studies (62%)
- 207 remaining (22%)
- 44,944 small-mammal records
 - 490,810 small-mammal detections
 - 6,179 distinct sampling locations/times
- 56,369 pathogen records
 - 584,391 pathogen assays
 - 48,542 positive assays (8.3%)
- 4,998 sequences able to be enriched with host/pathogen data
 - 4,248 pathogen sequences linked to hosts
 - 749 host sequences

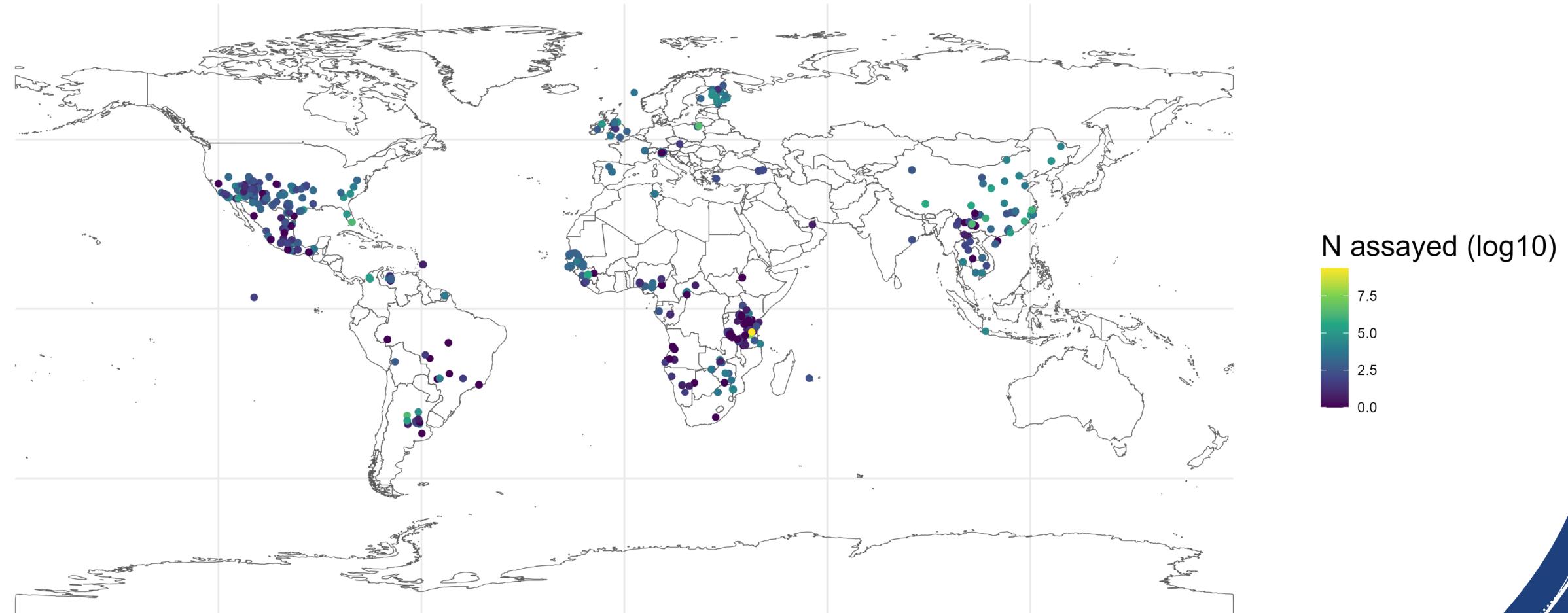
Countries with included studies



Hantaviridae



Mammarenaviridae

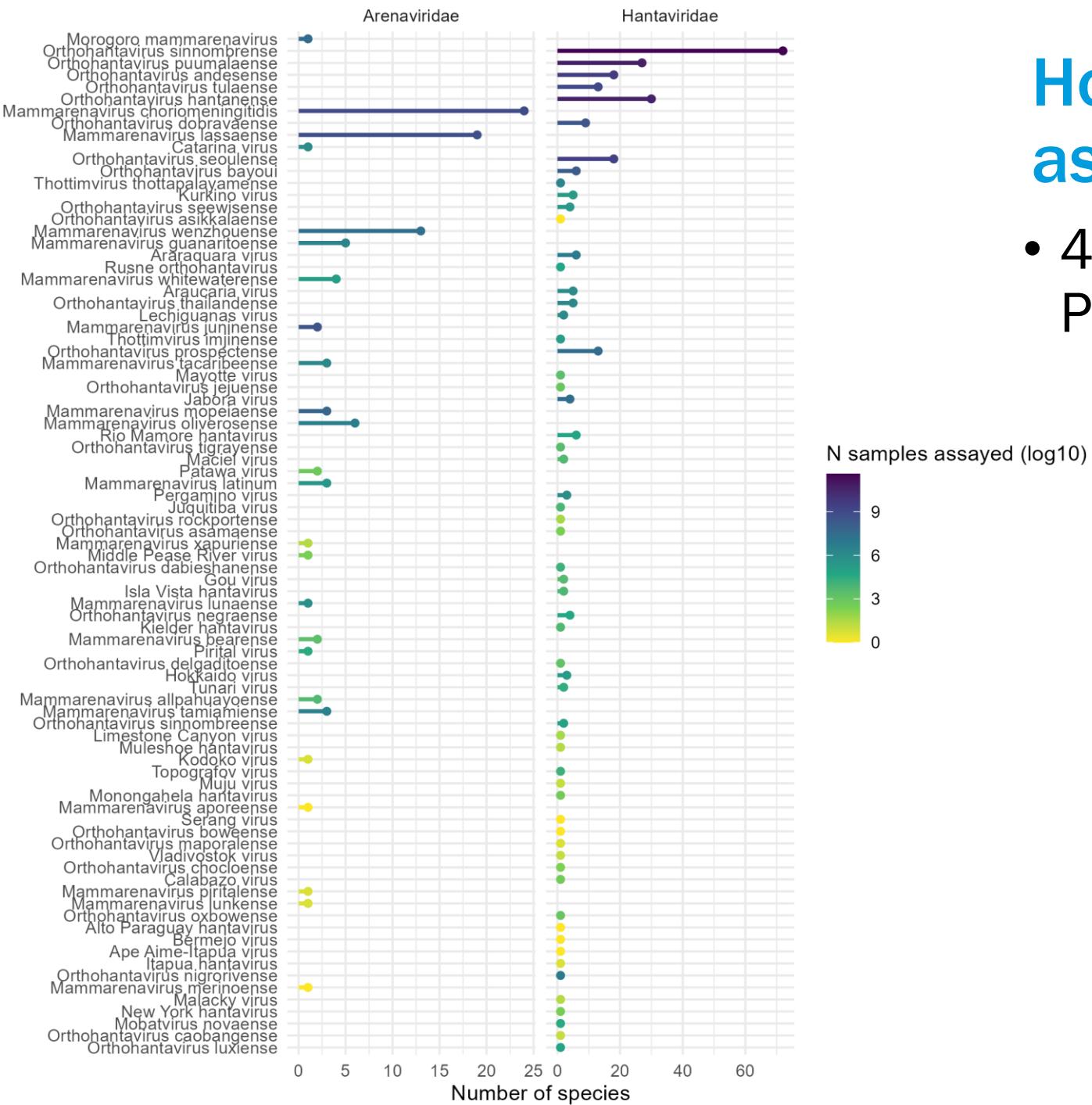


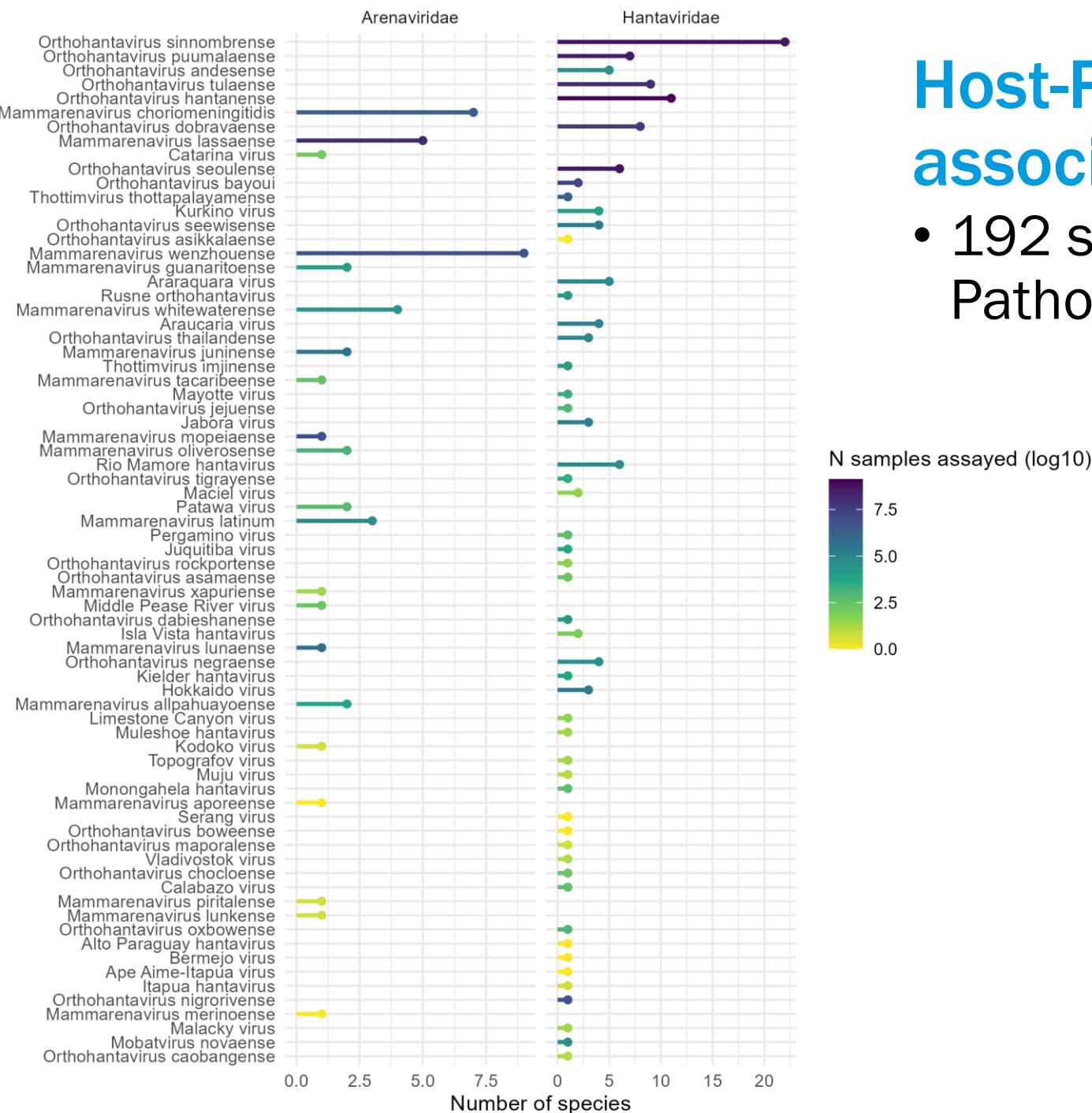
Harmonising data

- Host taxa matching
 - 1,252 reported species/genus/order names matched to 760 names resolved against GBIF with associated IDs using the `taxize` package.
- Pathogen taxa matching
 - 290 pathogen names matched to 92 NCBI Taxonomy Browser records (currently manual).
- Imputing non-detections of small-mammals
 - Expand non-detections based on whether the small-mammal was detected elsewhere in the study.
- To Do
 - Function to download and extract NCBI genbank data to incorporate sequence data and currently available metadata.
 - Mapping coordinate resolution (i.e., county, district, state) of sampling location to a contextually relevant distance.
 - Associating sampling locations with countries administrative levels to allow support future areal analyses

Host-Pathogen associations

- 406 species level Host-Pathogen associations





Host-Pathogen associations (acute only)

- 192 species level Host-Pathogen associations

Hantaviruses



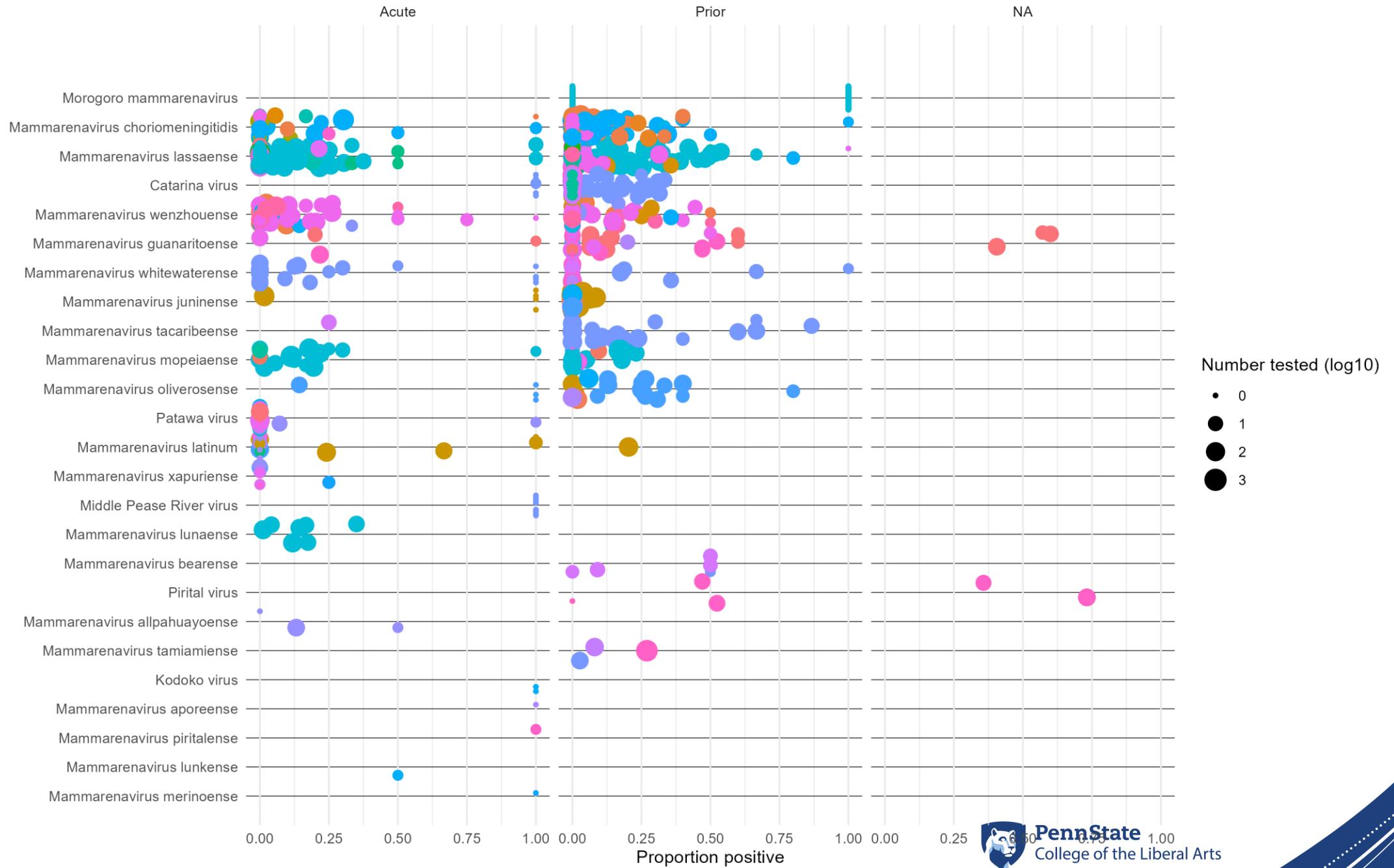
Proportion positive

Labels only shown for >25 assayed and >1 positive



PennState
College of the Liberal Arts

Arenaviridae



Follow along and get involved

- Protocol
 - <https://www.biorxiv.org/content/10.1101/2025.01.17.633514v1.abstract>
- Cleaning scripts, clean data
 - https://github.com/DidDrog11/arenavirus_hantavirus
- Shiny app for exploration (needs updating)
 - https://diddrog11.shinyapps.io/arenavirus_hantavirus_app/