

Rodent trapping studies can improve our understanding of the risk of endemic zoonoses spillover and novel zoonotic pathogen emergence.

David Simons^{1,2,3}

Lauren A. Attfield²

Kate E. Jones²

Deborah Watson-Jones^{3, 4}

Richard Kock¹

¹ Centre for Emerging, Endemic and Exotic Diseases, The Royal Veterinary College

² Centre for Biodiversity and Environment Research, University College London

³ Clinical Research Department, London School of Hygiene and Tropical Medicine

⁴ Mwanza Intervention Trials Unit, National Institute for Medical Research, Mwanza, Tanzania

Abstract

Rodents are important globally distributed reservoirs of known and novel zoonotic pathogens. Ongoing anthropogenic land use change is altering the composition of host species assemblages and modifying the risk of endemic zoonoses spillover events. These changes mean that an understanding of the current distribution of rodent species is vital for accurately describing disease hazard and managing risk. However, available species distribution and host-pathogen association datasets such as International Union for Conservation of Nature (IUCN) distribution maps, Global Biodiversity Information Facility (GBIF) species occurrence data, or the CLOVER host-pathogen dataset are biased by the inclusion of limited data sources and spatial biases in sampled locations. Here, we synthesise data from West Africa from 127 rodent trapping studies, conducted between 1964–2021, including over 76,000 individual rodent detections to more comprehensively characterise the range and presence of important zoonotic pathogen host species in this region. We identify that rodent trapping studies, although biased towards human dominated landscapes across West Africa, can usefully complement current rodent species distribution datasets and calculate the discrepancies between sampling in GBIF and included studies. For five regionally important zoonotic pathogens (*Arenaviridae* spp., *Borrelia* spp., *Lassa mammarenavirus*, *Leptospira* spp. and *Toxoplasma gondii*) we identify host-pathogen associations that are not reported in the consolidated CLOVER dataset. These host-pathogen association omissions may bias studies investigating the current risk of endemic zoonoses and the effect of land use and climate change on future novel pathogen emergence. Finally, we find that for these five pathogen groups, the proportion of a rodent hosts range that has been sampled remains small. A priority of future rodent trapping studies should be to

sample rodent hosts across a greater geographic range to better characterise current and future risk. In the interim, future studies must incorporate a measure of the current sampling biases. The current synthesis of contextually rich rodent trapping data enriches available information from IUCN, GBIF and CLOVER which can support a more complete understanding of the hazard of zoonotic spillover events.

Introduction

There is increasing awareness of the global health and economic impacts of novel zoonotic pathogen spillover, evidenced by the ongoing SARS-CoV-2 pandemic and previous HIV/AIDs and Spanish Influenza pandemics (Bernstein *et al.*, no date). The number of zoonotic disease spillover events and the frequency of the emergence of novel zoonotic pathogens are predicted to increase under intensifying anthropogenic pressure driven by increased human populations, urbanisation, intensification of agriculture, climate change and wildlife defaunation (Allen *et al.*, 2017; Hassell *et al.*, 2017; McMahon, Morand and Gray, 2018; García-Peña *et al.*, 2021). The impact of endemic zoonoses meanwhile remains underestimated (Maudlin, Eisler and Welburn, 2009). Endemic zoonoses disproportionately affect those in the poorest sections of society, those living in close contact with their animals and those with limited access to healthcare (Molyneux *et al.*, 2011; Halliday *et al.*, 2015; Judson and Rabinowitz, 2021). West Africa is a nexus for these processes and is identified as a hotspot for increased risk of large epidemics of known zoonotic diseases and novel zoonotic pathogen emergence (Grace *et al.*, 2012; Han, Kramer and Drake, 2016; Allen *et al.*, 2017). Two taxonomic groups - rodents (*Rodentia*) and bats (*Chiroptera*) - contribute the greatest number of predicted novel zoonotic pathogens and known zoonoses (Han *et al.*, 2015; Gibb, Gregory F. Albery, *et al.*, 2021b). Of 2,220 extant rodent species, 244 (10.7%) are described as reservoirs of 85 zoonotic pathogens (Han *et al.*, 2015). Specifically, West Africa has been identified as a region at increased hazard for rodent-borne zoonotic disease spillover events under different projected future land-use change scenarios (García-Peña *et al.*, 2021). Currently within West Africa, rodents are involved in the transmission of multiple endemic zoonoses with large burdens on human health, these pathogens include, Lassa fever, Schistosomiasis, Leptospirosis and Toxoplasmosis (Meerburg, Singleton and Kijlstra, 2009; Galeh *et al.*, 2020).

Rodent species form diverse assemblages, which provide important and beneficial ecosystem services including pest regulation and seed dispersal (Fischer *et al.*, 2018). The role of rodent species in zoonotic disease spillover or novel zoonotic pathogens emergence are examples of ecosystem disservices. Rodents typically demonstrate "fast" life histories (Dobson and Oli, 2007) with traits such as early maturation and short gestation times (<4 days) further associated with being zoonotic reservoirs (Han *et al.*, 2015; Albery and Becker, 2021). Rodent species with "fast" life histories thrive in human dominated landscapes, displacing species less likely to be reservoirs of zoonotic pathogens (Gibb, Redding, *et al.*, 2020). The widespread occurrence of reservoir species and their proximity to human activity make the description of rodent species assemblages and host-pathogen associations vitally important to understanding the hazard of zoonotic disease spillover and novel zoonotic pathogen emergence (Han, Kramer and Drake, 2016).

Despite the importance of understanding these complex systems, current evidence on host-pathogen associations is considerably affected by taxonomic and geographical sampling biases

(Gibb, Franklino, *et al.*, 2020; Gibb, Gregory F. Albery, *et al.*, 2021b). Curated biodiversity datasets such as the Global Biodiversity Information Facility (GBIF) and resources produced by the International Union for Conservation of Nature suffer from well described sampling biases (Boakes *et al.*, 2010; Bowler *et al.*, no date). These sampling biases can gravely distort produced species distribution models (Beck *et al.*, 2014). Datasets on host-pathogen associations also suffer from biases introduced from literature selection criteria and taxonomic discrepancies a recently produced consolidated dataset (CLOVER) has been produced to reduce this (Gibb, Carlson and Farrell, 2021). Studies identifying potential geographic hotspots of zoonotic disease spillover and novel pathogen emergence are typically based on these datasources for modelling host species distributions and host-pathogen associations which can result in biased hazard estimates (Plowright *et al.*, 2019; Carlson *et al.*, 2021). For example, systematically increased sampling, over-representation of certain habitats and clustering around areas of high human population could lead to an apparent association between locations and hazard that is driven by these factors rather than underlying host-pathogen associations (Redding *et al.*, 2017; Wille, Geoghegan and Holmes, 2021; Gibb, Gregory F. Albery, *et al.*, 2021b). While other regions remain systematically undersampled (e.g., areas of sparse human populations, endemic zoonoses), potentially resulting in a reduced hazard of zoonotic spillover events or novel zoonotic pathogen emergence being attributed (Beck *et al.*, 2014). Predictions of zoonotic disease spillover and novel zoonotic pathogen emergence must account for these biases to understand the future hazard of zoonotic diseases (Carlson *et al.*, 2021).

Rodent trapping studies provide contextually rich information on when, where and under what conditions rodents were trapped, potentially enriching consolidated datasets and providing additional data to minimise inherent biases (Bovendorp, McCleery and Galetti, 2017). Rodent trapping studies describe rodent population assemblages, their geographic distribution and host-pathogen associations beyond what is currently available in consolidated datasets. Studies have been conducted in West Africa to investigate the distribution of rodent species, their species assemblages, the prevalence of endemic zoonoses within rodent hosts (e.g., Lassa fever, Schistosomiasis) and to identify emerging and novel zoonotic pathogens (Fichet-Calvet *et al.*, 2009; Catalano *et al.*, 2020; USAID, 2021). However, individual level data from these studies have not previously been synthesised for inclusion in assessments of zoonotic disease spillover and novel zoonotic pathogen emergence.

Here, we synthesise rodent trapping studies conducted across West Africa from a search of literature between 1964-2022. First, we use this dataset to investigate the geographic sampling biases in relation to human population density and land use classification. Second, we compare this to curated host datasets (IUCN and GBIF) to understand differences in reported host geographic distributions. Third, we compare identified host-pathogen associations with a consolidated dataset (CLOVER) to explore discrepancies in rodent host-pathogen associations and report the proportion of positive assays for pathogens of interest. Finally, within our dataset we investigate the spatial extent of current host-pathogen sampling to identify areas of sparse sampling of pathogens within their host ranges. We expect that rodent trapping studies provide an important additional source of high-resolution data that can be used to enrich available consolidated datasets to better understand the hazard of zoonotic disease spillover and novel zoonotic pathogen emergence across West Africa.

Methods

Data sources

Host and pathogen trapping data

To identify relevant literature, we conducted a search in Ovid MEDLINE, Web of Science (Core collection and Zoological Record), JSTOR, BioOne, African Journals Online, Global Health and the pre-print servers, BioRxiv and EcoEvoRxiv for the following terms as exploded keywords: (1) Rodent OR Rodent trap* AND (2) West Africa, no date limits were set. We also searched other resources including the UN Official Documents System, Open Grey, AGRIS FAO and Google Scholar using combinations of the above terms. Searches were run on 2022-05-01, and returned studies conducted between 1964-2021.

For further analysis, we included studies if they met all of the following inclusion criteria; i) reported findings from trapping studies where the target was a small mammal, ii) described the type of trap used or the length of trapping activity or the location of the trapping activity, iii) included trapping activity from at least one West African country, iv) recorded the genus or species of trapped individuals, and v) were published in a peer-reviewed journal or as a pre-print on a digital platform or as a report by a credible organisation. We excluded studies if they met any of the following exclusion criteria: i) reported data that were duplicated from a previously included study, ii) no full text available, iii) not available in English. One author screened titles, abstracts and full texts against the inclusion and exclusion criteria. At each stage; title screening, abstract screening and full text review, a random subset (10%) was reviewed by a second author.

We extracted data from eligible studies using a standardised tool that was piloted on 5 studies (Supplementary Table 1.). Data was abstracted into a Google Sheets document, which was archived on completion of data extraction [ref to zenodo on finalised]. We identified the aims of included studies, for example, whether it was conducted as a survey of small mammal species or specifically to assess the risk of zoonotic disease spillover. To understand potential limitations in the representativeness of rodent trapping we extracted data on study methodology, such as, the number of trap nights, the type of traps used and whether the study attempted to estimate abundance. For studies not reporting number of trap nights we used imputation based on the number of trapped individuals, stratified by the habitat type from which they were obtained. This was performed by multiplying the total number of trapped individuals within that study site by the median trap success for study sites with the same reported habitat type. Stratification was used as trap success varied importantly between traps placed in or around buildings (13%, IQR 6-24%) compared with other habitats (3%, IQR 1-9%).

The method of taxonomic identification applied within a study was recorded, reported species identification was assumed to be accurate. The number of individuals of these species or genera was extracted with species and genus names mapped to GBIF taxonomy (GBIF: The Global Biodiversity Information Facility, 2021b) for internal consistency. We expanded species detection and non-detection records by explicitly specifying non-detection at a trap site if a species was recorded as detected at other trapping locations within the same study.

Geographic locations of trapping studies was extracted using GPS locations for the most precise location presented. GPS coordinates for locations not reported were produced using the National Geospatial-Intelligence Agency GEOnet Names Server (National Geospatial-Intelligence Agency, no date) based on place-names and maps presented in the study. All locations were converted to decimal degrees for internal consistency. The year of rodent trapping was extracted alongside the length of the trapping activity to understand seasonal representativeness of trapping activity. The habitats of trapping sites were mapped to the IUCN Habitat Classification Scheme (Version 3.1). For studies reporting multiple habitat types for a single trap, trap-line or trapping grid, a higher order classification of habitat type was recorded.

For relevant studies we extracted data on all microorganisms and zoonotic pathogens tested and the method used (e.g., molecular or serological diagnosis). Where assays were able to identify the microorganism to species level this was recorded, non-specific assays higher order attribution was used (e.g. to family level). We recorded the species of rodent host tested, the number of individuals tested and the number of positive and negative results. For studies reporting summary results all testing data were extracted, this may introduce double counting of individual rodents, for example, if a single rodent was tested using both molecular and serological assays. Where studies reported indeterminate results, these were also recorded.

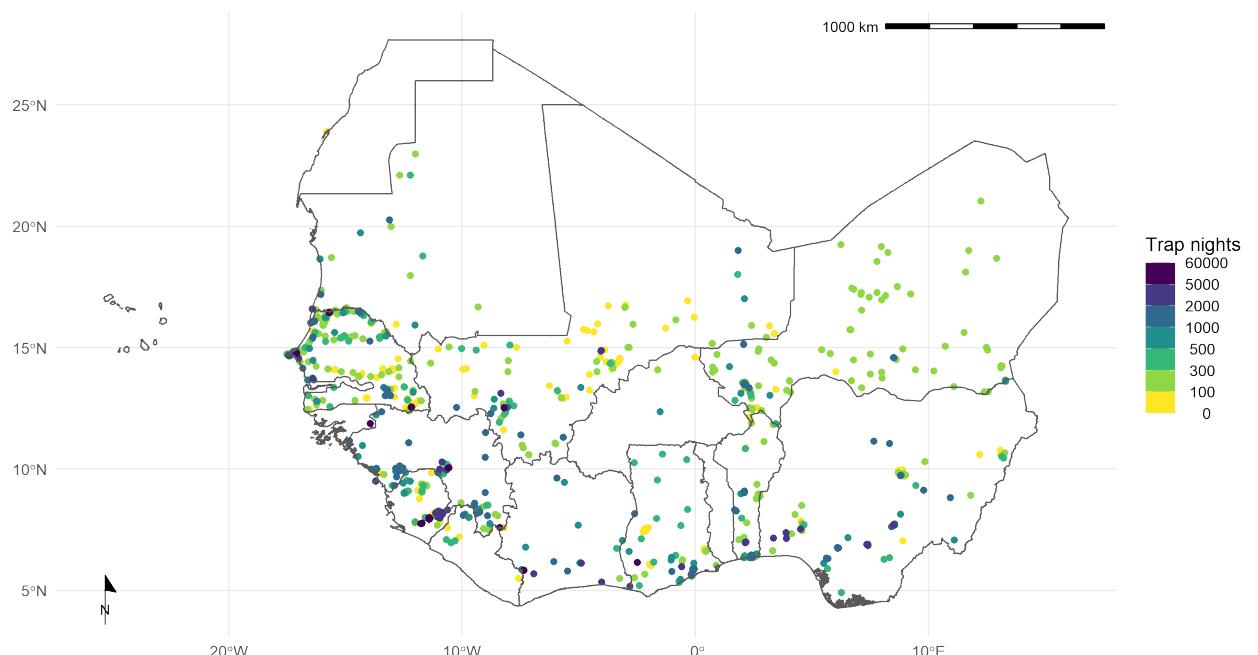
We summarised the number of studies, the year in which trapping occurred and the country in which they were conducted. We identified 4,692 relevant citations, with 127 rodent trapping studies included (Supplementary Table 2.). The earliest trapping studies were conducted in 1964, with a trend of increasing numbers of studies being performed annually since 2000. The median year of first trapping activity was 2007, with the median length of trapping activity being 1 year (IQR 0-2 years) (Supplementary Figure 1.). Studies were conducted in 14 West African countries, with no studies reported from The Gambia or Togo, at 1,611 trap sites (Figure 1A.).

Included studies explicitly reported on 601,184 trap nights, a further 341,445 trap nights were imputed from studies with no recording of trapping effort based on trap success. A total of 942,629 trap nights were estimated (Figure 1B.). A minority of studies trapped at a single study site (30, 24%), with 46 (36%) trapping at between two and five sites, the remaining 51 studies (40%) trapped at between six and 93 study sites.

In total 76,275 small mammals were trapped with 65,628 (90%) identified to species level and 7,439 (10%) identified to genus, with the remaining classified to higher taxonomic level. The majority of the 132 identified species were Rodentia (102, 78%), of which Muridae (73, 72%) were the most common family. Soricomorpha were the second most identified order of small mammals (28, 21%). 57 studies tested for 32 microorganisms, defined to species or genus level that are known or potential pathogens. Most studies tested for a single microorganism (48, 84%). The most frequently assayed microorganisms were *Lassa mammarenavirus* or Arenaviridae (21, 37%), *Borrelia* sp. (9, 16%), *Bartonella* sp. (4, 7%) and *Toxoplasma gondii* (4, 7%). Most studies used Polymerase Chain Reaction (PCR) to detect microorganisms (37, 65%), with fewer studies using serology-based tests (11, 19%) or histological or direct visualisation assays (11, 21%). From 32,920 individual rodent samples we produced 351 host-pathogen pairs. With *Rattus rattus*, *Mus musculus*, *Mastomys erythroleucus*, *Mastomys natalensis* and *Arvicanthis niloticus* being assayed for at least 18 microorganisms. Further descriptive information from the included studies including geolocated trapping of species, their detection

and non-detection alongside microorganism data is available as supplementary material (Simons, 2022).

A



B

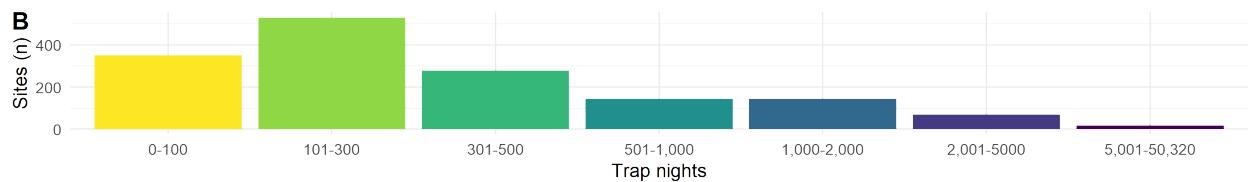


Figure 1: A) The location of trapping sites in West Africa. No sites were recorded from Togo or The Gambia. Heterogeneity is observed in the coverage of each country by trap night (colour) and location of sites. For example, Senegal, Mali and Sierra Leone have generally good coverage compared to Guinea and Burkina Faso. B) Histogram of trap nights performed at each study site, a median of 248 trap nights (IQR 116-500) was performed at each site

Analysis

What is the extent of spatial bias in the rodent trapping data?

To investigate the risk to human health from endemic zoonoses and novel pathogen emergence using synthesised data from rodent trapping data understanding of sampling biases is vital. Sampling of rodents has not been performed randomly across space in West Africa (Figure 1.) and at varying intensity, we aimed quantify this bias as the relationship between trapping effort

(measured as trap night density (TN density)), human population density and land use type. We define this as relative trapping effort bias.

To address this question, we calculated TN density within each West African level-2 administrative region, administrative boundaries were obtained from GADM 4.0.4 (Database of Global Administrative Areas, 2022 of Global Administrative Areas). We assigned trap nights to specific regions using GPS coordinates obtained from trapping studies as described above. The sf package in the R statistical language was used to handle shapefiles and geographic data (Pebesma, 2018). Density was calculated by dividing the number of trap nights by the area of a level-2 administrative area. For studies not reporting trap nights imputation was used as previously described. Human population density was obtained for the closest year (2005) to the median year of trapping (2007) from Socioeconomic Data and Applications Center (SEDAC) gridded population of the world v4 at ~ 1km resolution (Socioeconomic Data and Applications Center, 2021). Median population density was then calculated for each level-2 administrative region. Land cover classification was obtained from the Copernicus climate change service at ~300m resolution (European Space Agency Climate Change Initiative, 2022). The proportion of cropland, shrubland, tree cover and urban land cover within a level-2 administrative region in 2005 was calculated.

We investigated the role of these covariates on the spatial bias of TN density by comparing to a spatial model (equation 1), as follows:

$$\text{TN density} \sim \text{Tweedie}(\text{Longitude} * \text{Latitude}) \quad (\text{equation 1})$$

$$\text{TN density} \sim \text{Tweedie}(\text{population density} + (\text{Longitude} * \text{Latitude})) \quad (\text{equation 2})$$

$$\begin{aligned} \text{TN density} \sim \\ \text{Tweedie}(\text{population density} + \text{level 2 administrative region area} + (\text{Longitude} * \text{Latitude})) \end{aligned} \quad (\text{equation 3})$$

$$\begin{aligned} \text{TN density} \sim \\ \text{Tweedie}(\text{proportion cropland} + \text{proportion tree cover} + \text{proportion urban} + (\text{Longitude} * \text{Latitude})) \end{aligned} \quad (\text{equation 4})$$

$$\begin{aligned} \text{TN density} \sim \\ \text{Tweedie}(\text{proportion urban} + \text{population density} + \text{level 2 administrative region area} + (\text{Longitude} * \text{Latitude})) \end{aligned} \quad (\text{equation 5})$$

We formulated these models as a Generalised Additive Models (GAM) incorporating a spatial interaction term (longitude by latitude) (Pedersen *et al.*, 2019). These models investigated the association between TN density, proportion of urban, cropland, tree cover and human population density. The models were constructed in the *mgcv* package using the R programming language (Wood, 2017; R Core Team, 2020). Selection of the most parsimonious model was based on Deviance explained and the Akaike information criterion for each model. We performed sensitivity analysis by removing sites with imputed trapping effort and performing a pixel-based analysis.

What is the difference in rodent host distributions between curated datasets and rodent trapping studies?

Curated datasets (i.e., IUCN and GBIF) are commonly used in assessments of biodiversity in response to land use and climate change in addition to being used to explore zoonotic pathogen host distributions. To understand any biases introduced from these sources into these analyses we aimed to assess the concordance of curated rodent host distributions from IUCN and GBIF with observed rodent presence and absence from rodent trapping studies. We obtained rodent species distribution maps as shapefiles from the IUCN red list and translated these to a ~20km resolution raster (IUCN, 2021). Distributions were cropped to the study region for globally distributed rodent species. We obtained rodent presence locations from GBIF as point data limited to the study region (GBIF: The Global Biodiversity Information Facility, 2021a). Presence locations were associated to cells of raster with a ~20km resolution produced for the study region.

We assessed the taxonomic coverage of the trapped taxa and for the seven species with the most trap locations (*M. natalensis*, *R. rattus*, *M. erythroleucus*, *M. musculus*, *A. niloticus*, *Praomys daltoni* and *Cricetomys gambianus*) we mapped detection and non-detection, associating each location with a ~20km resolution raster. For these seven species we calculated the area of the IUCN expected range. We calculated the percentage of this range covered by presence detections in GBIF, this process was repeated for presence detections from rodent trapping studies. For both GBIF data and trapping studies we calculated the area of detections outside of the IUCN expected range. For rodent trapping studies only, we were able to calculate the area of non-detections within the IUCN expected area. Finally, we calculated the combined area of detection when combining GBIF and rodent trapping study data.

Are rodent trapping derived host-pathogen associations present in a consolidated zoonoses dataset?

Studies investigating the risk of endemic zoonoses and novel pathogen emergence generally rely on consolidated datasets of host-pathogen associations. To examine the usefulness of rodent trapping studies as an additional source of data we compared identified host-pathogen associations with a consolidated zoonoses dataset (CLOVER) (Gibb, Gregory F. Albery, *et al.*, 2021b). CLOVER is a synthesis of four host-pathogen datasets (GMPD2, EID2, HP3 and Shaw, 2020) and was released in 2021, it contains more than 25,000 host-pathogen associations for Bacteria, Viruses, Helminth, Protozoa and Fungi. An archived version of the CLOVER dataset was obtained and subset for host species trapped within included studies, spatial filtering of CLOVER for host-pathogen associations limited to the study region was not possible (Gibb, Carlson and Farrell, 2021).

We produced host-pathogen pairs from rodent trapping studies investigating zoonoses. For each host and pathogen, with both identified to species level, we compared our host-pathogen pair network with one produced using CLOVER, reporting concordance and discordance. We describe these networks and report differences in network composition. For host-pathogen pairs with assay results consistent with acute or prior infection we calculate the proportion positive and identify those absent from CLOVER. We expand the analysis to host-pathogen pairs with pathogens identified to genus level in Supplementary Figure 4.

What is the spatial extent of pathogen testing within a hosts range?

Sampling of zoonotic pathogens may also be biased within their hosts ranges, which may hinder inference across the entire range of the species. To quantify this, we focused on the five pathogen groups and species sampled at the most locations (Arenaviridae, Borreliaceae, Leptospiraceae, *Lassa mammarenavirus* and *Toxoplasma gondii*). We geolocated individuals of host species tested for these pathogens and calculated the area sampled. We adopt a similar approach to above by mapping sampled hosts to ~20km pixels and calculating the proportion of their IUCN range in which sampling has occurred. We also present the area covered of all areas in which the host has been detected in this study to produce a measure of relative completeness of sampling within the included rodent trapping studies.

Data and code implemented in the R statistical language has been made available alongside this manuscript to reproduce these analyses [ref_zenodo].

Results

What is the extent of spatial bias in the rodent trapping data?

We found a non-random spatial distribution of rodent trapping locations across the study region, suggestive of underlying bias in the sampling or rodents across West Africa. Trap sites were situated in 256 of 1,450 (17.6%) level-2 administrative regions in 14 West African nations. The regions with highest TN density included the capitals and large cities of Niger (Niamey), Nigeria (Ibadan), Ghana (Accra), Senegal (Dakar), Ghana (Accra) and Benin (Cotonou). Outside of these cities, regions in, Northern Senegal, Southern Guinea, Edo and Ogun States in Nigeria and Eastern Sierra Leone had the greatest TN density (Figure 1A.).

Trapping sites were biased towards human modified landscapes (i.e., cropland, grassland and mosaic habitats), with under-representation of forest and bare habitats. However, urban habitats were comparatively undersampled (Supplementary Figure 2.).

The most parsimonious GAM model (adjusted R² = 0.3, Deviance explained = 48.7%) included smooth terms for geographic coordinates, human population density, proportion of urban landscape, and the area of the region. All smooth terms demonstrate significant non-linear associations ($p < 0.002$) with TN density (Supplementary Table 3.5). The spatial smooth term adjusted for population density, proportion of urban landscape and region area is shown in Figure 2. The final model identified relative trapping effort bias, with a bias towards increased trapping effort in North West Senegal, North and East Sierra Leone, Central Guinea and coastal regions of Nigeria, Benin and Ghana; in contrast South East Nigeria, Northern Nigeria and Burkina Faso had an observed bias towards a reduced trapping effort (Figure 2). Model summaries informing model selection are shown in Supplementary Tables 3.1-3.5. In sensitivity analysis, excluding sites with imputed trap nights reduced contributing data; however, Mauritania, Northern Senegal and Sierra Leone remained as regions trapped at higher rates, with Nigeria being trapped at lower than expected rates (Supplementary Figure 3a.). In pixel based sensitivity analysis spatial coverage was reduced with similar patterns of bias observed (Supplementary Figure 3b.).

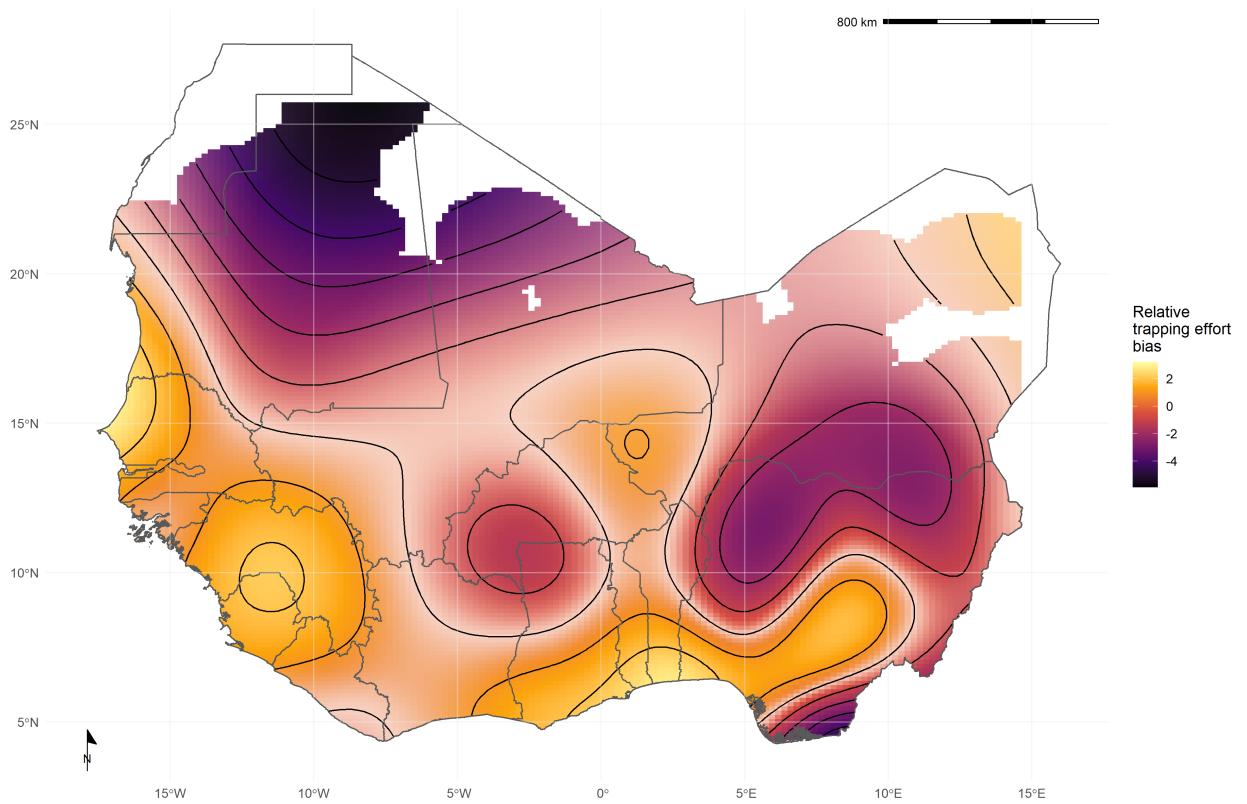


Figure 2. Relative trapping effort bias across West Africa from included studies adjusted for human population density, proportion urban land cover and area of the administrative region.

Uncertainty in the point estimate of relative trapping effort bias is represented by colour transparency/pallor (e.g., Eastern Liberia and Central Mali). Yellow regions represent areas with a bias towards increased trapping effort (e.g., North West Senegal), purple regions represent areas with a bias towards reduced trapping effort (e.g., Northern Nigeria).

What is the difference in rodent host distributions between curated datasets and rodent trapping studies?

We found that rodent trapping studies could provide a useful supplement to curated data describing rodent host distributions. For the seven rodent species of interest (*M. natalensis*, *R. rattus*, *M. erythroleucus*, *M. musculus*, *A. niloticus*, *P. daltoni* and *C. gambianus*), except *C. gambianus*, trapping studies provided more distinct locations of detection and non-detection than were available from GBIF. For the endemic rodent species (*M. natalensis*, *M. erythroleucus*, *A. niloticus*, *P. daltoni* and *C. gambianus*) IUCN ranges had good concordance to both trapping studies and GBIF, however, individuals of *A. niloticus* and *P. daltoni* were detected outside of

IUCN ranges. In contrast, the non-native species *R. rattus* and *M. musculus* were detected across much greater ranges than were expected from IUCN distributions (Figure 3A. and 3B.).

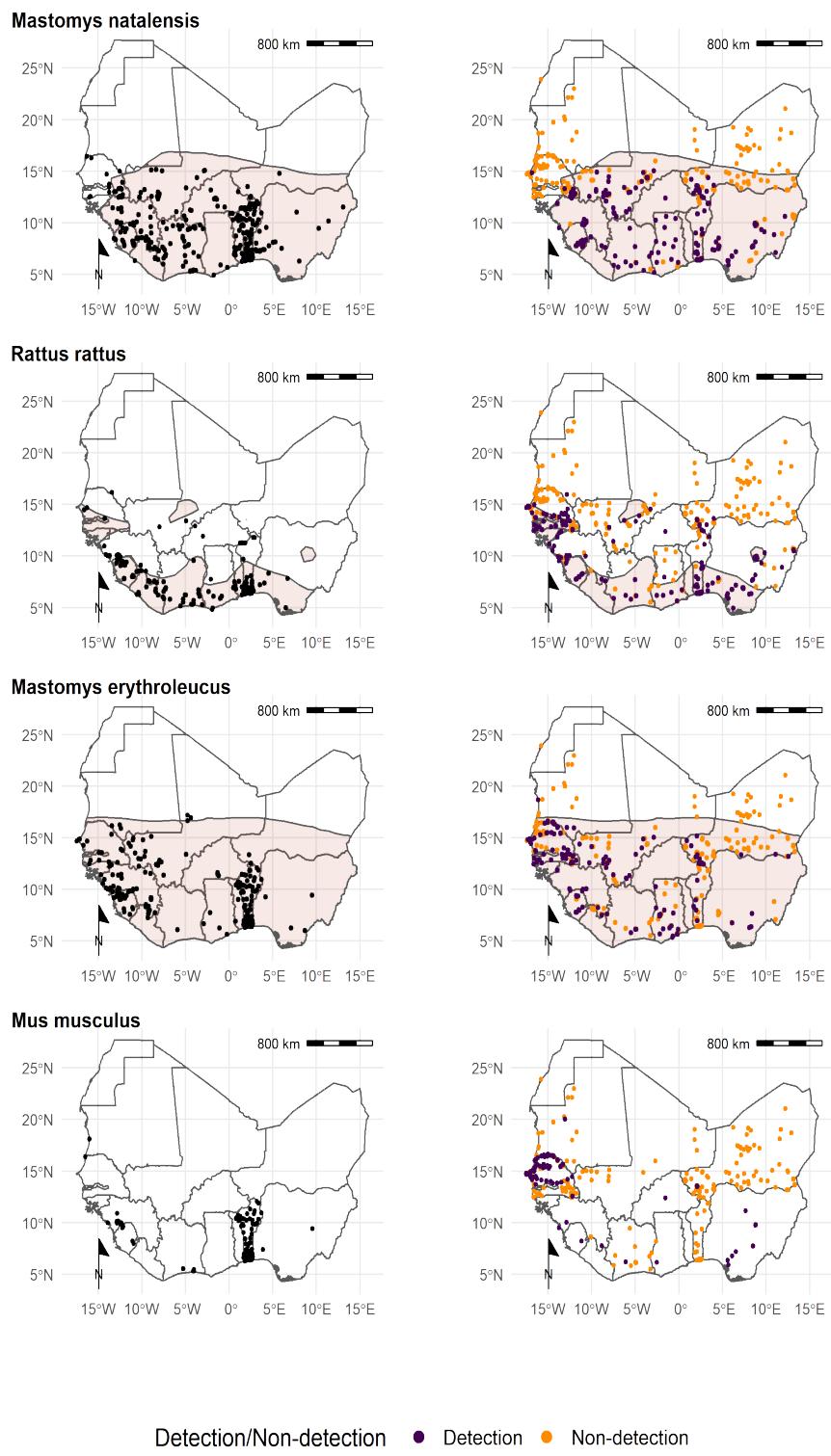
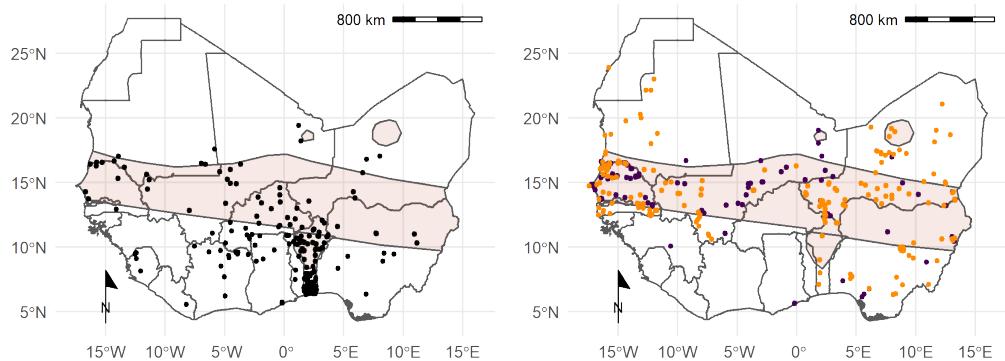
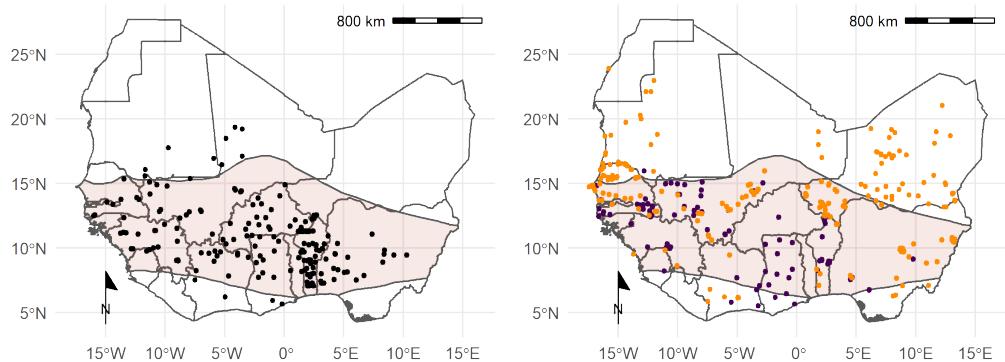


Figure 3A. Each row corresponds to a single rodent species. L) Presence recorded in GBIF overlaid on IUCN species range (red-shaded area). R) Detection and non-detection from rodent trapping studies overlaid on IUCN species ranges. *M. musculus* has no IUCN West African range.

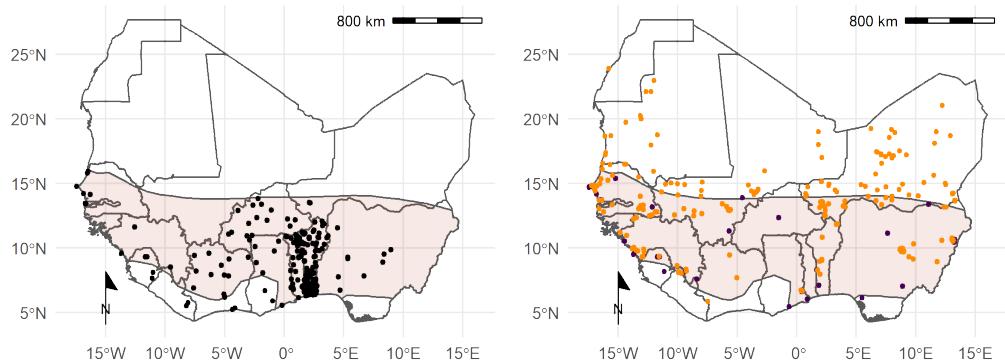
Arvicanthis niloticus



Praomys daltoni



Cricetomys gambianus



Detection/Non-detection • Detection • Non-detection

Figure 3B. Each row corresponds to a single rodent species. L) Presence recorded in GBIF overlaid on IUCN species range (red-shaded area). R) Detection (purple) and non-detection (orange) from rodent trapping studies overlaid on IUCN species ranges.

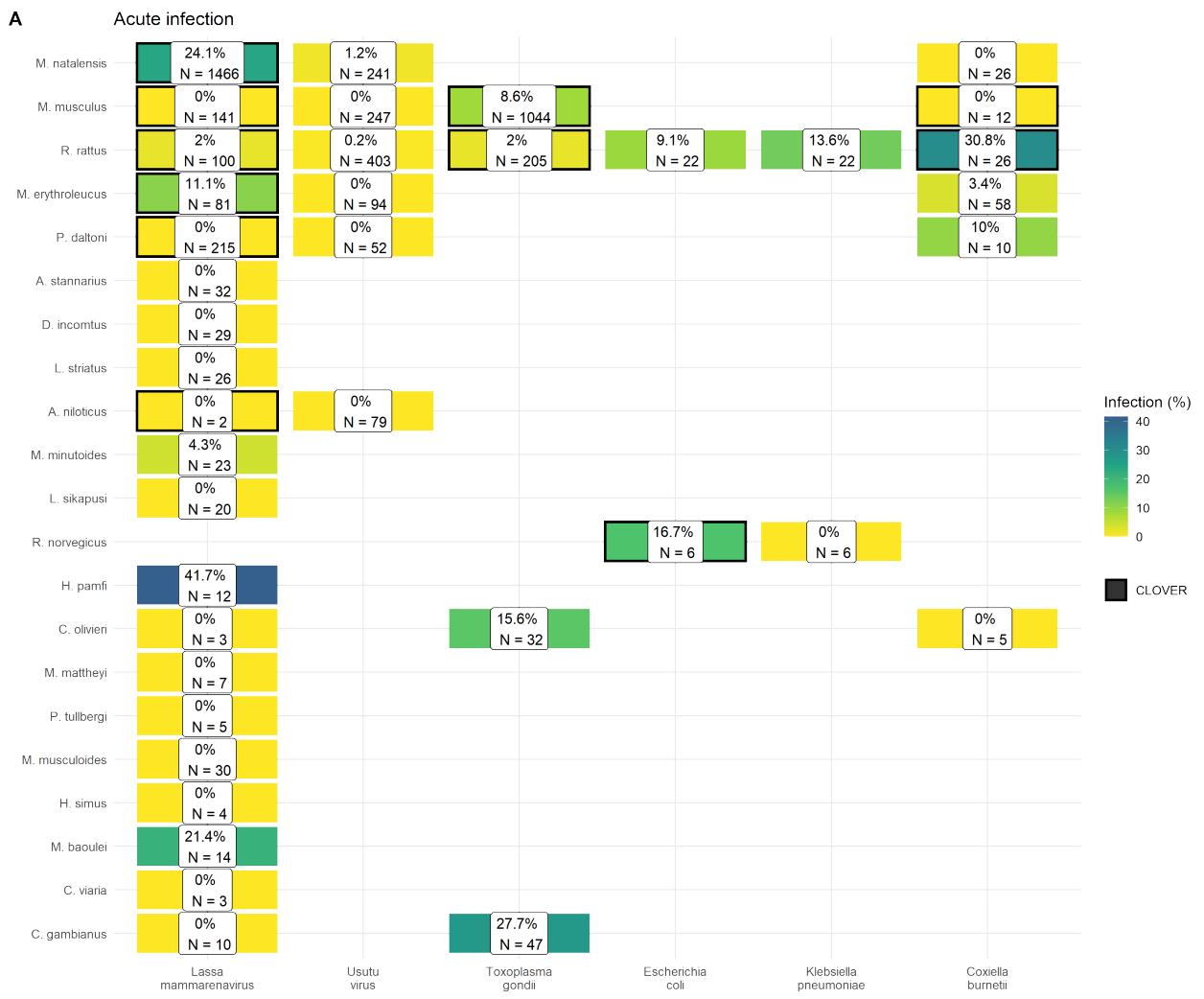
To further understand the sampling coverage of curated datasets and included rodent trapping studies we identified the proportion of a species IUCN range in which detections and non-detections occurred for the seven species most detected in the included studies (Table 1.). Apart from *M. musculus* these species had IUCN ranges greater than 1 million km², with the Mastomys species having the largest geographic ranges. Sampling locations of these species within GBIF covered between 0.09-0.26% of expected ranges with detections occurring outside IUCN ranges for all species, most noticeably for *A. niloticus* and *R. rattus*. Sampling locations within rodent trapping studies covered 0.03-0.24% of expected ranges with detections occurring outside IUCN ranges most noticeably for *A. niloticus* and *R. rattus*. Combining GBIF and rodent trapping data increased the sampled area by a mean of 65% (SD 36.5%), suggesting limited overlap between the locations sampled. Non-detection of species occurred across species ranges (mean = 0.11%, SD = 0.03%), suggestive of heterogeneity of presence across IUCN ranges.

Table 1: Comparison of IUCN, GBIF and rodent trapping ranges for the 7 species trapped at the most sites.

	IUCN	GBIF		Trapping studies			Combined
Species	Range (1,000 km ²)	Area inside range (1,000 km ²)	Area outside range (1,000 km ²) (%) of IUCN)	Detection area inside range (1,000 km ²) (%) of IUCN)	Area outside range (1,000 km ²) (%) of IUCN)	Non- detection area inside range (1,000 km ²) (% of IUCN)	Detection area inside range (1,000 km ²) (% of IUCN)
<i>Mastomys natalensis</i>	3,257.11	6.83 (0.21%)	0.19	4.4 (0.14%)	0.17	3.12 (0.1%)	12.73 (0.33%)
<i>Rattus rattus</i>	1,018.71	2.61 (0.26%)	0.52	2.42 (0.24%)	1.21	1.3 (0.13%)	5.72 (0.48%)
<i>Mastomys erythroleucus</i>	3,735.48	4.48 (0.12%)	0.04	3.24 (0.09%)	0.12	4.35 (0.12%)	11 (0.2%)
<i>Mus musculus</i>			2.15		1.85		3.94
<i>Arvicanthis niloticus</i>	1,829.14	1.69 (0.09%)	2.41	1.98 (0.11%)	0.34	3.09 (0.17%)	5.96 (0.2%)
<i>Praomys daltoni</i>	2,657.77	4.03 (0.15%)	0.29	2.03 (0.08%)	0.15	2.78 (0.1%)	8.21 (0.22%)
<i>Cricetomys gambianus</i>	2,475.97	5 (0.2%)	0.17	0.75 (0.03%)	0.06	2.99 (0.12%)	8.37 (0.23%)

Are rodent trapping derived host-pathogen associations present in a consolidated zoonoses dataset?

We found potentially important differences between the host-pathogen networks produced from included rodent trapping studies and the consolidated CLOVER dataset. When limited to taxonomic classification of both pathogen and host to species level we identified 25 host-pathogen pairs among 14 rodent and 6 pathogen species (Figure 4.). We identified negative associations (non-detection through specific assays) for 45 host-pathogen pairs among 35 rodent and 7 pathogen species. CLOVER contained 10 (40%) of our identified host-pathogen associations, the remaining 15 (60%) were not found to be present in CLOVER, additionally CLOVER recorded positive associations for 4 (9%) of the negative associations produced from the rodent trapping studies.



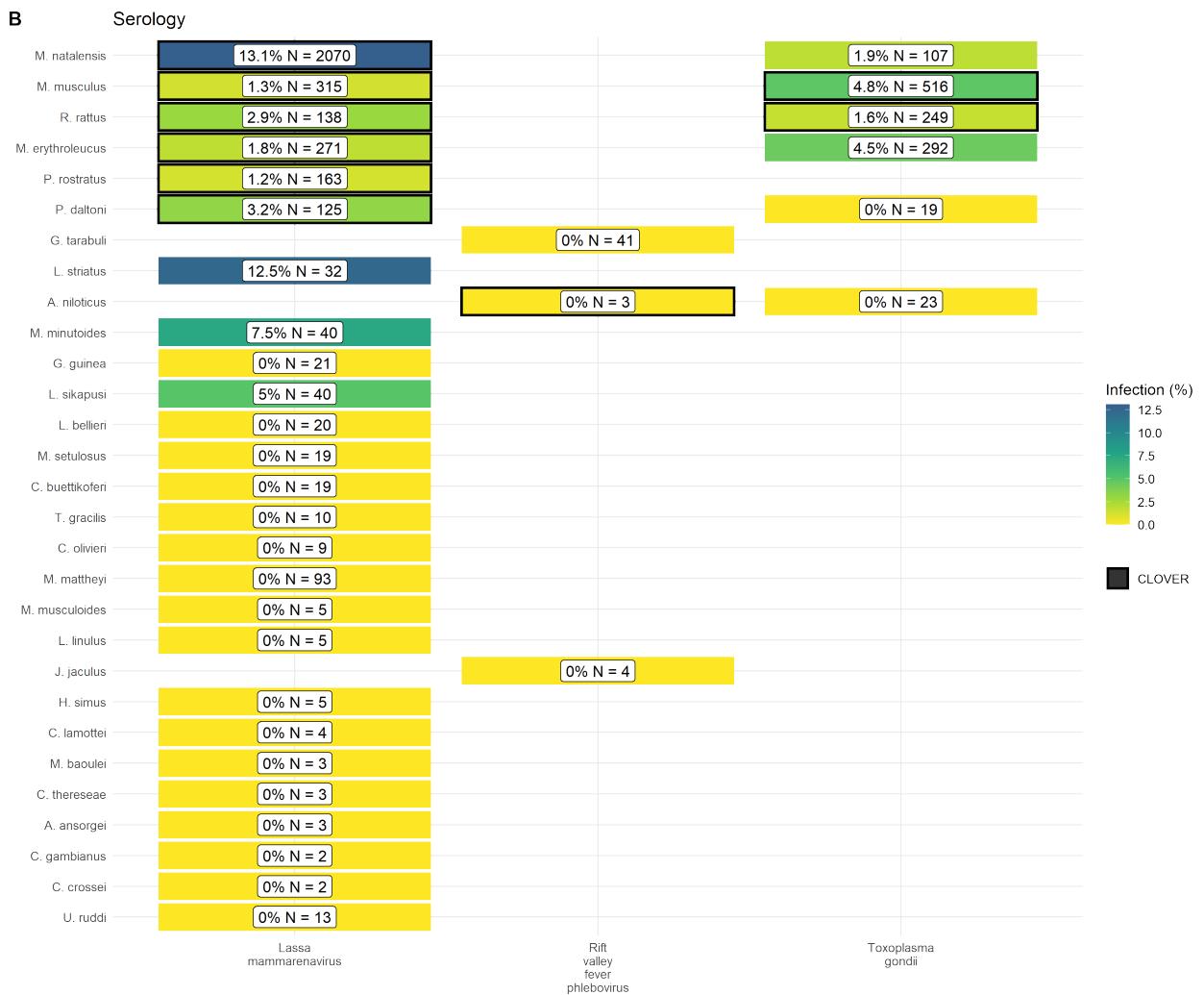


Figure 4. A) Identified species level host-pathogen associations through detection of acute infection (i.e. PCR, culture). Percentages and colour relate to the proportion of all assays that were positive. Associations with a black border are present in the CLOVER dataset. **B)** Identified species level host-pathogen associations through serological assays (i.e. ELISA). Percentages and colour relate to the proportion of all assays that were positive. Associations with a black border are present in the CLOVER dataset.

CLOVER included an additional 492 host-pathogen associations we do not observe in rodent trapping studies. The majority of these 392 (80%) pairs are from species with global distributions (*M. musculus*, *R. rattus* and *R. norvegicus*), or from those with wide ranging distributions in sub-Saharan Africa (38, 8%) (i.e., *A. niloticus*, *M. natalensis* and *Atelerix albiventris*). For pathogens not identified to species level (i.e. family or higher taxa only), we identified 148 host-pathogen pairs among 32 rodent species and 25 pathogen families (Supplementary Figure 4.), with CLOVER containing 66 (45%) of these associations. Rodent trapping studies identified additional rodent host species for six pathogens; *Lassa mammarenavirus* (5), *Toxoplasma gondii* (4), Usutu virus (2), *Coxiella burnetii* (2), *Escherichia coli* and *Klebsiella pneumoniae* (both 1), that were not present in this consolidated host-pathogen association dataset.

What is the spatial extent of pathogen testing within a host's range?

The five most widely sampled pathogen species/families in included studies were Arenaviridae, Borreliaceae, *Lassa mammarenavirus*, Leptospiraceae and *Toxoplasma gondii* (Table 2.). Assays to identify Arenaviridae infection were performed in 44 rodent species with evidence of viral infection in 15 species. *Lassa mammarenavirus* was specifically tested for in 43 species with 10 showing evidence of viral infection. The most commonly infected species for both Arenaviridae, generally, and *Lassa mammarenavirus* specifically, were *M. natalensis* and *M. erythroleucus*. Assays were conducted on individuals trapped within less than a third of the locations in which the species were trapped (Table 2.).

Infection with species of Borreliaceae was assessed in 42 species, with evidence of infection in 17 rodent species. Among species with more than 500 samples the highest rates of infection were among *A. niloticus* (16%), *Mastomys huberti* (11%) and *M. erythroleucus* (9%). The coverage of sampling within the trapping locations was slightly greater than that for Arenaviridae but remains a small proportion (<1%) of the host species IUCN range. Infection with species of Leptospiraceae and *Toxoplasma gondii* was assessed in 8 species, with evidence of infection in 5 and 6 rodent species respectively. Testing for these pathogens was more limited with few individuals tested and reduced coverage of host ranges.

*Table 2: Comparison of pathogen sampling ranges for the 5 most widely sampled pathogens and the 5 most sampled potential rodent host species. * no IUCN range in West African*

Pathogen	Species	Tested	Positive	Pathogen testing area (1,000 km ²)	Pathogen testing area within trapped area (%)	Pathogen testing area within IUCN range (%)
Arenaviridae sp.						
	<i>Mastomys natalensis</i>	2,841	104 (4%)	0.61	13.45%	0.02%
	<i>Praomys daltoni</i>	854	6 (1%)	0.42	19.43%	0.02%
	<i>Mastomys erythroleucus</i>	398	20 (5%)	0.40	11.97%	0.01%
	<i>Rattus rattus</i>	396	4 (1%)	0.38	10.5%	0.04%
	<i>Praomys rostratus</i>	310	5 (2%)	0.13	12.53%	0.02%
Borrelia sp.						
	<i>Mastomys erythroleucus</i>	1,586	140 (9%)	1.14	33.94%	0.03%
	<i>Arvicantis niloticus</i>	1,551	253 (16%)	0.66	28.48%	0.03%
	<i>Mastomys natalensis</i>	733	54 (7%)	0.69	15.08%	0.02%
	<i>Mastomys huberti</i>	731	83 (11%)	0.23	29.83%	0.04%
	<i>Mus musculus</i>	686	26 (4%)	0.45	24.54%	*
Lassa mammarenavirus						
	<i>Mastomys natalensis</i>	3,199	580 (18%)	1.03	22.65%	0.03%
	<i>Mastomys erythroleucus</i>	352	14 (4%)	0.36	10.63%	0.01%
	<i>Rattus rattus</i>	177	2 (1%)	0.34	9.26%	0.03%
	<i>Praomys rostratus</i>	163	2 (1%)	0.27	27.02%	0.04%

Pathogen	Species	Tested	Positive	Pathogen testing area (1,000 km ²)	Pathogen testing area within trapped area (%)	Pathogen testing area within IUCN range (%)
Leptospira sp.	<i>Mus musculus</i>	147	0 (0%)	0.04	2.29%	*
	<i>Rattus rattus</i>	646	65 (10%)	0.40	11.1%	0.04%
	<i>Arvicanthis niloticus</i>	221	10 (5%)	0.02	0.9%	<0.01%
	<i>Crocidura olivieri</i>	141	14 (10%)	0.34	25.16%	*
	<i>Mastomys natalensis</i>	136	26 (19%)	0.36	7.91%	0.01%
	<i>Rattus norvegicus</i>	79	19 (24%)	0.21	40.08%	*
Toxoplasma gondii	<i>Mus musculus</i>	1,548	115 (7%)	0.62	33.64%	*
	<i>Rattus rattus</i>	428	8 (2%)	0.36	9.77%	0.03%
	<i>Mastomys erythroleucus</i>	292	13 (4%)	0.37	11.06%	0.01%
	<i>Mastomys natalensis</i>	107	2 (2%)	0.08	1.83%	<0.01%
	<i>Cricetomys gambianus</i>	47	13 (28%)	0.06	7.6%	<0.01%

Discussion

Endemic rodent zoonoses and novel pathogen emergence from rodent hosts are predicted to have an increasing burden in West Africa and globally (Han *et al.* 2015). Here we have synthesised data from 126 rodent trapping studies containing information on more than 76,000 rodents across 1,611 trap sites producing an estimated 942,669 trap nights across 14 West African countries. Locations studied are complementary to curated datasets (e.g. IUCN,

GBIF), incorporation of our synthesised dataset when assessing zoonosis risk based on host distributions could counteract some of the biases inherent to these curated datasets (Boakes *et al.*, 2010). We identified 25 host-pathogen pairs reported from included studies, 15 of these were not included in a consolidated host-pathogen dataset. However, pathogen sampling within identified species occurrence locations was limited, highlighting a further component of sampling bias supporting calls for further pathogen sampling across diverse species in zoonotic hotspots (Harvey and Holmes, 2022).

We found that rodent trapping studies, similar to biodiversity data, showed important spatial biases (Beck *et al.*, 2014). Relative trapping effort was biased towards increased effort in Benin, Guinea, Senegal and Sierra Leone driven by long-standing research collaborations investigating the invasion of non-native rodent species (*M. musculus* and *R. rattus*) and the risk of endemic zoonosis outbreaks (e.g. *Lassa mammarenavirus*). Much of West Africa remains relatively understudied, particularly Burkina Faso, Côte d'Ivoire, Ghana and Nigeria, despite these countries facing many of the same challenges. For example, annual outbreaks of Lassa fever are reported in Nigeria and there are potentially 60,000 unrecognised cases of Lassa fever each year in Côte d'Ivoire and Ghana (Basinski *et al.*, 2021). Rodent sampling should be targeted towards regions with minimal data to reduce this bias and assist in understanding the risk from endemic and novel zoonoses.

Rodent trapping studies provide geographic and temporally contextualised data on both species detection and non-detection which are not available from curated datasets. Non-detection data can improve models of species distributions, unfortunately, high levels of missing data on trapping effort will continue to confound the allocations of non-detections as true absences (Václavík and Meentemeyer, 2009). Models of host species occurrence and abundance, improved by incorporating species absence, are important to assess the effect of land use and climate change on endemic zoonosis spillover to human populations and direct limited public health resources towards regions at greatest risk (Plowright *et al.*, 2019).

Currently available consolidated datasets on host-pathogen associations (e.g. CLOVER, EID2 and GMPD2) do not include spatial or temporal components (Gibb, Gregory F. Albery, *et al.*, 2021a). The current synthesis of rodent trapping studies has highlighted that pathogens have been sparsely sampled within a hosts range. Current zoonosis risk models dependent on these sources of data are therefore not able to incorporate spatial heterogeneity in pathogen prevalence across the host range. Additional uncertainty in current models of zoonotic disease risk arises from observed host-pathogen associations that have not been reported in these consolidated datasets. For example, *Hylomyscus pamfi* infected with *Lassa mammarenavirus* and *R. rattus* infected with *Coxiella burnetii*, will not be included when solely dependent on consolidated host-pathogen datasets. Further, detection of zoonotic pathogens in multiple co-occurring host species supports the adoption of multi-species to better understand the potential range of endemic zoonoses (Wilkinson *et al.*, 2019).

Few studies stratified detection and non-detection of hosts or pathogen prevalence by time, therefore limiting inference of changes in host and pathogen dynamics. This limitation prevents calculation of incidence of infection and the abundance of infectious rodents which potentially varies by both time and space (Fichet-Calvet *et al.*, 2016). Understanding of temporal changes in viral burden and shedding for endemic zoonoses is required to accurately predict current and future risk of pathogen spillover.

Finally, due to data sparsity, we were unable to account for temporal change over the six decades of rodent trapping studies. Land use change and population density have changed dramatically over this period in West Africa (Herrmann *et al.*, 2020). We attempted to mitigate against this by using the median year of trapping as a reference point to understand the spatial and land use biases in trapping activity. It is possible that land use and population density at trapping sites varied importantly between when the rodent trapping study was conducted and the land use and human population variables in 2005. Despite this limitation, the finding that trapping is biased towards high density, human dominated landscapes is unlikely to substantially change (Beck *et al.*, 2014).

We have shown that synthesis of rodent trapping studies to supplement curated rodent distributions can counteract some of the inherent biases in these data and that they can add further contextual data to host-pathogen association data. Together this supports their inclusion in efforts to model endemic zoonotic risk and novel pathogen emergence. Contribution of rodent trapping studies as data sources can be improved by adopting reporting standards and practices consistent with Open Science, namely sharing of disaggregated datasets alongside publication (Foster and Deardorff, 2017).

Future rodent trapping studies should be targeted towards regions that are currently under-studied. Further information on rodent presence and abundance across West Africa will assist in modelling changing endemic zoonosis risk and the potential for novel pathogen emergence. Sharing of disaggregated data alongside research publications should be supported with adoption of data standards to support further data synthesis. Inclusion of exact locations of trapping sites, trapping effort and the dates at which trapping occurred would support more detailed understanding of the spatio-temporal dynamics of host populations and the risk of endemic zoonosis spillover events. Despite these challenges we propose that rodent trapping studies can provide an important source of data to supplement curated datasets on rodent distributions to quantify the risk of endemic zoonosis spillover events and the hazard of novel pathogen emergence.

References

- Albery, G. F. and Becker, D. J. (2021) 'Fast-lived Hosts and Zoonotic Risk', *Trends in Parasitology*, 37(2), pp. 117–129. doi: [10.1016/j.pt.2020.10.012](https://doi.org/10.1016/j.pt.2020.10.012).
- Allen, T. *et al.* (2017) 'Global hotspots and correlates of emerging zoonotic diseases', *Nature Communications*, 8(1), p. 1124. doi: [10.1038/s41467-017-00923-8](https://doi.org/10.1038/s41467-017-00923-8).
- Basinski, A. J. *et al.* (2021) 'Bridging the gap: Using reservoir ecology and human serosurveys to estimate Lassa virus spillover in West Africa', *PLOS Computational Biology*. Edited by A. Wesolowski, 17(3), p. e1008811. doi: [10.1371/journal.pcbi.1008811](https://doi.org/10.1371/journal.pcbi.1008811).
- Beck, J. *et al.* (2014) 'Spatial bias in the GBIF database and its effect on modeling species' geographic distributions', *Ecological Informatics*, 19, pp. 10–15. doi: [10.1016/j.ecoinf.2013.11.002](https://doi.org/10.1016/j.ecoinf.2013.11.002).
- Bernstein, A. S. *et al.* (no date) 'The costs and benefits of primary prevention of zoonotic pandemics', *Science Advances*, 8(5), p. eabl4183. doi: [10.1126/sciadv.abl4183](https://doi.org/10.1126/sciadv.abl4183).

- Boakes, E. H. *et al.* (2010) 'Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data', *PLOS Biology*, 8(6), p. e1000385. doi: [10.1371/journal.pbio.1000385](https://doi.org/10.1371/journal.pbio.1000385).
- Bovendorp, R. S., McCleery, R. A. and Galetti, M. (2017) 'Optimising sampling methods for small mammal communities in Neotropical rainforests', *Mammal Review*, 47(2), pp. 148–158. doi: [10.1111/mam.12088](https://doi.org/10.1111/mam.12088).
- Bowler, D. E. *et al.* (no date) 'Temporal trends in the spatial bias of species occurrence records', *Ecography*, n/a(n/a), p. e06219. doi: [10.1111/ecog.06219](https://doi.org/10.1111/ecog.06219).
- Carlson, C. J. *et al.* (2021) 'The future of zoonotic risk prediction', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1837), p. 20200358. doi: [10.1098/rstb.2020.0358](https://doi.org/10.1098/rstb.2020.0358).
- Catalano, S. *et al.* (2020) 'Multihost Transmission of Schistosoma mansoni in Senegal, 2015–2018', *Emerging Infectious Diseases*, 26(6), pp. 1234–1242. doi: [10.3201/eid2606.200107](https://doi.org/10.3201/eid2606.200107).
- Database of Global Administrative Areas (2022) 'GADM'. Available at: <https://gadm.org/index.html> (Accessed: 25 April 2021).
- Dobson, F. S. and Oli, M. K. (2007) 'Fast and slow life histories of rodents', *Rodent societies: an ecological and evolutionary perspective*, pp. 99–105. Available at: <https://www.jstor.org/stable/42902037>.
- European Space Agency Climate Change Initiative (2022) 'Land cover classification gridded maps from 1992 to present derived from satellite observations'. Available at: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=overview> (Accessed: 3 March 2022).
- Fichet-Calvet, E. *et al.* (2009) 'Diversity and dynamics in a community of small mammals in coastal Guinea, West Africa', *Belgian Journal of Zoology*, 139(2), pp. 93–102. Available at: <http://www.vliz.be/nl/open-marien-archief?module=ref&refid=204614>.
- Fichet-Calvet, E. *et al.* (2016) 'Spatial and temporal evolution of Lassa virus in the natural host population in Upper Guinea', *Scientific Reports*, 6, p. 21977. doi: [10.1038/srep21977](https://doi.org/10.1038/srep21977).
- Fischer, C. *et al.* (2018) 'Ecosystem services and disservices provided by small rodents in arable fields: Effects of local and landscape management', *Journal of Applied Ecology*, 55(2), pp. 548–558. doi: [10.1111/1365-2664.13016](https://doi.org/10.1111/1365-2664.13016).
- Foster, E. D. and Deardorff, A. (2017) 'Open Science Framework (OSF)', *Journal of the Medical Library Association : JMLA*, 105(2), pp. 203–206. doi: [10.5195/jmla.2017.88](https://doi.org/10.5195/jmla.2017.88).
- Galeh, T. M. *et al.* (2020) 'Global Status of Toxoplasma gondii Seroprevalence in Rodents: A Systematic Review and Meta-Analysis', *Frontiers in Veterinary Science*, 7. Available at: <https://www.frontiersin.org/article/10.3389/fvets.2020.00461> (Accessed: 2 March 2022).
- García-Peña, G. E. *et al.* (2021) 'Land-use change and rodent-borne diseases: Hazards on the shared socioeconomic pathways', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1837), p. 20200362. doi: [10.1098/rstb.2020.0362](https://doi.org/10.1098/rstb.2020.0362).

GBIF: The Global Biodiversity Information Facility (2021a) 'Occurrence Download'. The Global Biodiversity Information Facility. doi: [10.15468/DL.S52MEH](https://doi.org/10.15468/DL.S52MEH).

GBIF: The Global Biodiversity Information Facility (2021b) 'The Global Biodiversity Information Facility'. Available at: <https://www.gbif.org/> (Accessed: 25 April 2021).

Gibb, R., Franklino, L. H. V., et al. (2020) 'Ecosystem perspectives are needed to manage zoonotic risks in a changing climate', *BMJ*, 371, p. m3389. doi: [10.1136/bmj.m3389](https://doi.org/10.1136/bmj.m3389).

Gibb, R., Redding, D. W., et al. (2020) 'Zoonotic host diversity increases in human-dominated ecosystems', *Nature*, 584(7821), pp. 398–402. doi: [10.1038/s41586-020-2562-8](https://doi.org/10.1038/s41586-020-2562-8).

Gibb, R., Albery, Gregory F., et al. (2021a) 'Data Proliferation, Reconciliation, and Synthesis in Viral Ecology', *BioScience*, 71(11), pp. 1148–1156. doi: [10.1093/biosci/biab080](https://doi.org/10.1093/biosci/biab080).

Gibb, R., Albery, Gregory F., et al. (2021b) 'Mammal virus diversity estimates are unstable due to accelerating discovery effort', *Biology Letters*, 18(1), p. 20210427. doi: [10.1098/rsbl.2021.0427](https://doi.org/10.1098/rsbl.2021.0427).

Gibb, R., Carlson, C. J. and Farrell, M. J. (2021) 'Viralemergence/clover: Preprint + Zenodo'. Zenodo. doi: [10.5281/zenodo.4435128](https://doi.org/10.5281/zenodo.4435128).

Grace, D. et al. (2012) 'Mapping of poverty and likely zoonoses hotspots'.

Halliday, J. E. B. et al. (2015) 'Endemic zoonoses in the tropics: A public health problem hiding in plain sight', *The Veterinary Record*, 176(9), pp. 220–225. doi: [10.1136/vr.h798](https://doi.org/10.1136/vr.h798).

Han, B. A. et al. (2015) 'Rodent reservoirs of future zoonotic diseases', *Proceedings of the National Academy of Sciences*, 112(22), pp. 7039–7044. doi: [10.1073/pnas.1501598112](https://doi.org/10.1073/pnas.1501598112).

Han, B. A., Kramer, A. M. and Drake, J. M. (2016) 'Global Patterns of Zoonotic Disease in Mammals', *Trends in Parasitology*, 32(7), pp. 565–577. doi: [10.1016/j.pt.2016.04.007](https://doi.org/10.1016/j.pt.2016.04.007).

Harvey, E. and Holmes, E. C. (2022) 'Diversity and evolution of the animal virome', *Nature Reviews Microbiology*, 20(6), pp. 321–334. doi: [10.1038/s41579-021-00665-x](https://doi.org/10.1038/s41579-021-00665-x).

Hassell, J. M. et al. (2017) 'Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface', *Trends in Ecology & Evolution*, 32(1), pp. 55–67. doi: [10.1016/j.tree.2016.09.012](https://doi.org/10.1016/j.tree.2016.09.012).

Herrmann, S. M. et al. (2020) 'Accelerating land cover change in West Africa over four decades as population pressure increased', *Communications Earth & Environment*, 1(1), pp. 1–10. doi: [10.1038/s43247-020-00053-y](https://doi.org/10.1038/s43247-020-00053-y).

IUCN (2021) 'The IUCN Red List of Threatened Species'. Available at: www.iucnredlist.org.

Judson, S. D. and Rabinowitz, P. M. (2021) 'Zoonoses and global epidemics', *Current Opinion in Infectious Diseases*, 34(5), pp. 385–392. doi: [10.1097/QCO.0000000000000749](https://doi.org/10.1097/QCO.0000000000000749).

- Maudlin, I., Eisler, M. C. and Welburn, S. C. (2009) 'Neglected and endemic zoonoses', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530), pp. 2777–2787. doi: [10.1098/rstb.2009.0067](https://doi.org/10.1098/rstb.2009.0067).
- McMahon, B. J., Morand, S. and Gray, J. S. (2018) 'Ecosystem change and zoonoses in the Anthropocene', *Zoonoses and Public Health*, 65(7), pp. 755–765. doi: [10.1111/zph.12489](https://doi.org/10.1111/zph.12489).
- Meerburg, B., Singleton, G. and Kijlstra, A. (2009) 'Rodent-borne diseases and their risks for public health', *Critical reviews in microbiology*, 35, pp. 221–70. doi: [10.1080/10408410902989837](https://doi.org/10.1080/10408410902989837).
- Molyneux, D. *et al.* (2011) 'Zoonoses and marginalised infectious diseases of poverty: Where do we stand?', *Parasites & Vectors*, 4(1), p. 106. doi: [10.1186/1756-3305-4-106](https://doi.org/10.1186/1756-3305-4-106).
- National Geospatial-Intelligence Agency (no date) 'NGA: GNS Home'. Available at: <https://geonames.nga.mil/gns/html/> (Accessed: 22 March 2021).
- Pebesma, E. (2018) 'Simple Features for R: Standardized Support for Spatial Vector Data', *The R Journal*, 10(1), pp. 439–446. Available at: <https://journal.r-project.org/archive/2018/RJ-2018-009/index.html> (Accessed: 24 May 2022).
- Pedersen, E. J. *et al.* (2019) 'Hierarchical generalized additive models in ecology: An introduction with mgcv', *PeerJ*, 7, p. e6876. doi: [10.7717/peerj.6876](https://doi.org/10.7717/peerj.6876).
- Plowright, R. K. *et al.* (2019) 'Sampling to elucidate the dynamics of infections in reservoir hosts', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1782), p. 20180336. doi: [10.1098/rstb.2018.0336](https://doi.org/10.1098/rstb.2018.0336).
- R Core Team (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Redding, D. W. *et al.* (2017) 'Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data', *PLOS ONE*, 12(11), p. e0187602. doi: [10.1371/journal.pone.0187602](https://doi.org/10.1371/journal.pone.0187602).
- Simons, D. (2022) 'Rodents and potential zoonotic diseases in West Africa'. Available at: https://diddrog11.shinyapps.io/scoping_review_app/ (Accessed: 2 March 2022).
- Socioeconomic Data and Applications Center (2021) 'Gridded Population of the World (GPW), v4 SEDAC'. Available at: <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4> (Accessed: 4 February 2021).
- USAID (2021) 'One Health Surveillance', *PREDICT Project*. Available at: <https://p2.predict.global/surveillance> (Accessed: 27 September 2021).
- Václavík, T. and Meentemeyer, R. K. (2009) 'Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions?', *Ecological Modelling*, 220(23), pp. 3248–3258. doi: [10.1016/j.ecolmodel.2009.08.013](https://doi.org/10.1016/j.ecolmodel.2009.08.013).

Wilkinson, D. P. *et al.* (2019) 'A comparison of joint species distribution models for presence-absence data', *Methods in Ecology and Evolution*, 10(2), pp. 198–211. doi: [10.1111/2041-210X.13106](https://doi.org/10.1111/2041-210X.13106).

Wille, M., Geoghegan, J. L. and Holmes, E. C. (2021) 'How accurately can we assess zoonotic risk?', *PLOS Biology*. Edited by A. P. Dobson, 19(4), p. e3001135. doi: [10.1371/journal.pbio.3001135](https://doi.org/10.1371/journal.pbio.3001135).

Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R*. 2nd edn. Chapman; Hall/CRC.

Supplementary

Supplementary Table 1

Supplementary Table 2

Supplementary Table 3

*Supplementary Table 3.1: GAM model Trap night density ~ Tweedie(Longitdue * Latitude)*

Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	-2.062	0.126	-16.403	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(x,y)	30.431	39.000	7.290	***

Signif. codes: 0 <= **** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < " < 1

Adjusted R-squared: 0.0261, Deviance explained 0.349

fREML : 4252.745, Scale est: 11.991, N: 1450

*Supplementary Table 3.2: GAM model Trap night density ~ Tweedie(Population density + (Longitdue * Latitude))*

Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	-2.230	0.119	-18.704	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(pop_2005)	2.818	11.000	7.231	***
	s(x,y)	27.640	39.000	4.545	***

Signif. codes: 0 <= **** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < " < 1

Adjusted R-squared: 0.161, Deviance explained 0.422

fREML : 3924.187, Scale est: 9.836, N: 1450

*Supplementary Table 3.3: GAM model Trap night density ~ Tweedie(Population density + Region area + (Longitdue * Latitude))*

Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	-2.371	0.120	-19.699	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(pop_2005)	3.182	11.000	9.055	***
	s(area_km2)	3.052	9.000	2.294	***
	s(x,y)	26.819	39.000	4.843	***

Signif. codes: 0 <= **** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < " < 1

Adjusted R-squared: 0.148, Deviance explained 0.443

fREML : 3941.346, Scale est: 9.432, N: 1450

*Supplementary Table 3.4: GAM model Trap night density ~ Tweedie(Proportion cropland + Proportion shrubland + Proportion tree cover + Proportion urban + (Longitdue * Latitude))*

Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	-1.775	0.123	-14.379	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(cropland)	0.000	9.000	0.000	
	s(shrubland)	0.000	8.000	0.000	
	s(tree_cover)	1.579	9.000	0.795	**
	s(urban)	4.772	9.000	19.703	***
	s(x,y)	1.837	19.000	1.499	***

Signif. codes: 0 <= **** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1 < " < 1

Adjusted R-squared: 0.0599, Deviance explained 0.306

fREML : 4198.018, Scale est: 12.895, N: 1450

*Supplementary Table 3.5: Final GAM model Trap night density ~ Tweedie(Population density + Region area + Proportion urban + (Longitdue * Latitude))*

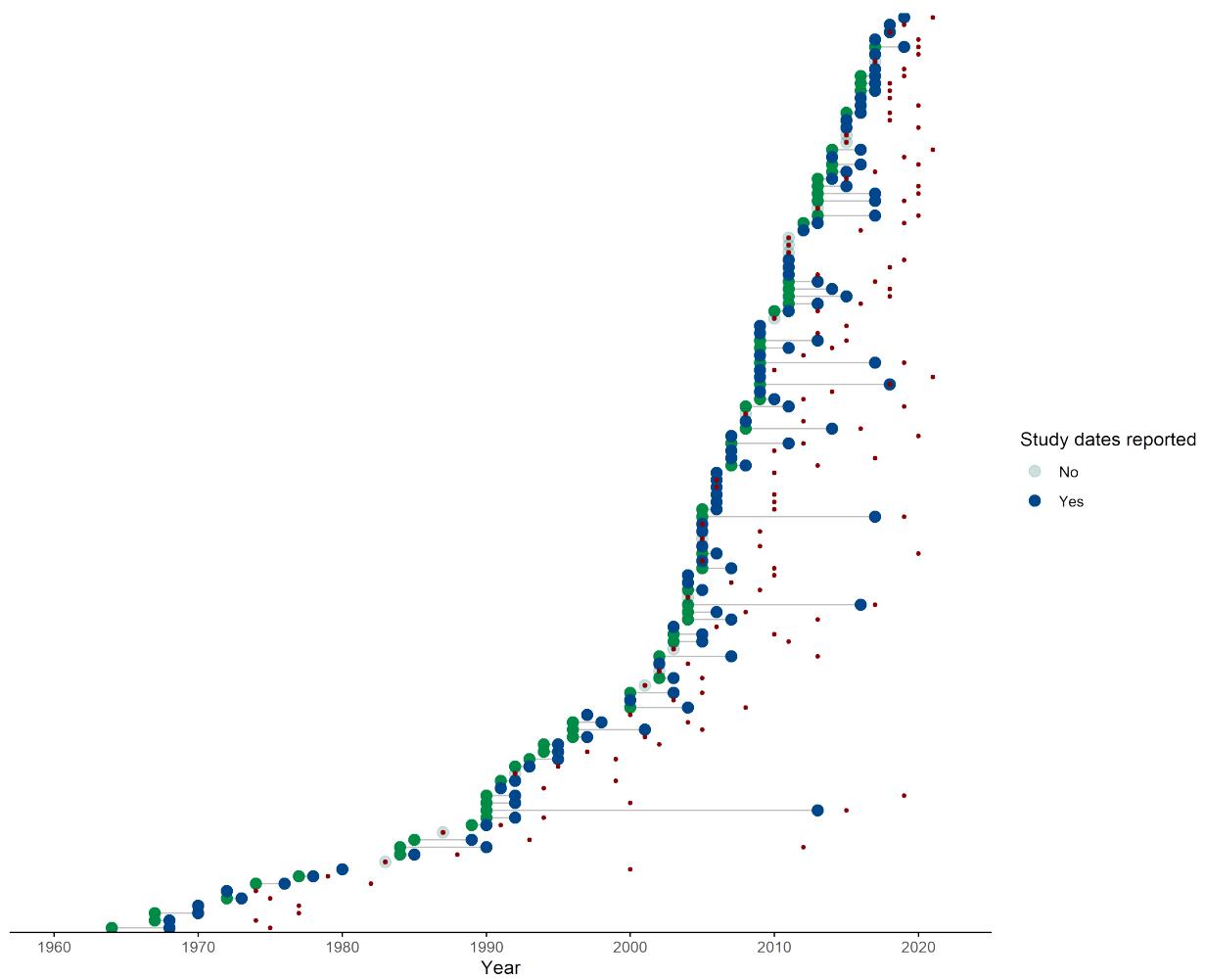
Component	Term	Estimate	Std Error	t-value	p-value
A. parametric coefficients	(Intercept)	-2.599	0.123	-21.112	***
Component	Term	edf	Ref. df	F-value	p-value
B. smooth terms	s(pop_2005)	7.131	11.000	6.398	***
	s(area_km2)	3.629	9.000	3.450	***
	s(urban)	1.921	9.000	1.235	**
	s(x,y)	27.252	39.000	4.563	***

Signif. codes: 0 <= '****' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < " < 1

Adjusted R-squared: 0.301, Deviance explained 0.487

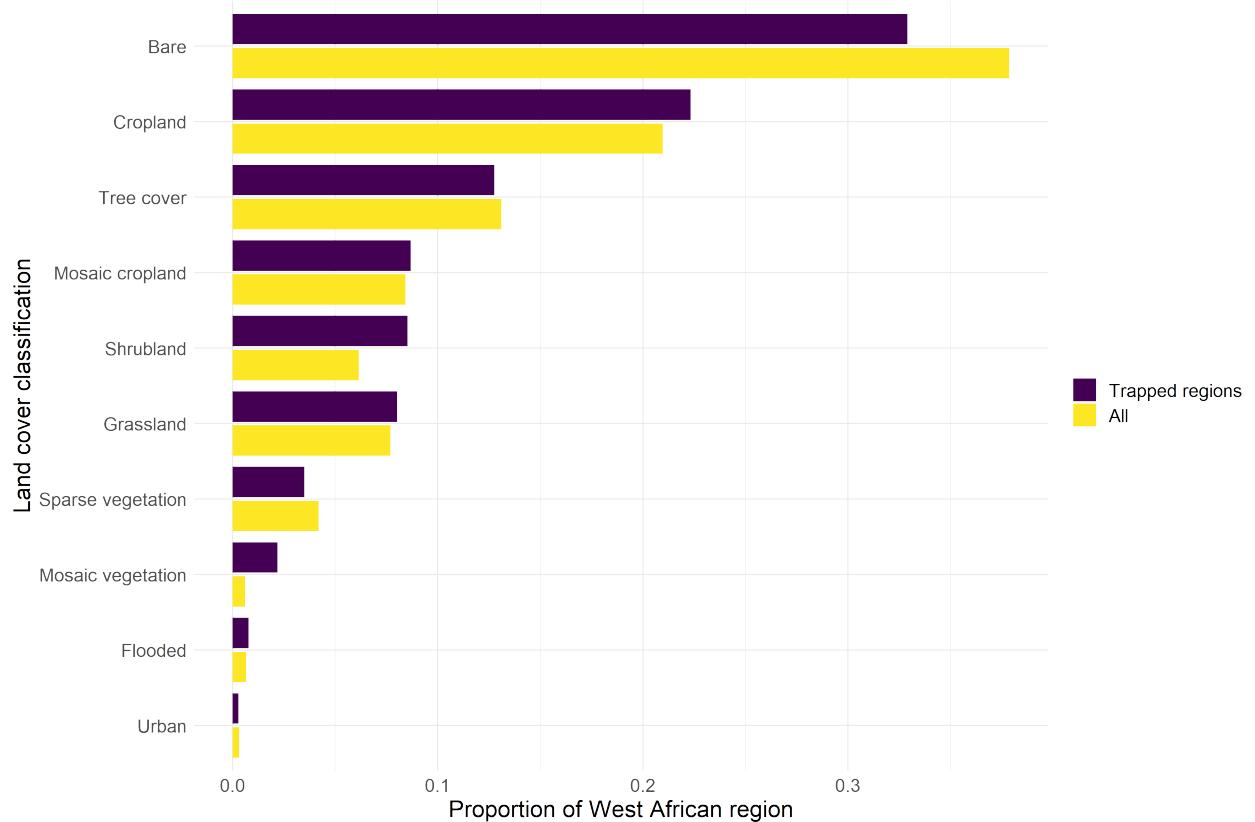
fREML : 3907.563, Scale est: 8.408, N: 1450

Supplementary Figure 1



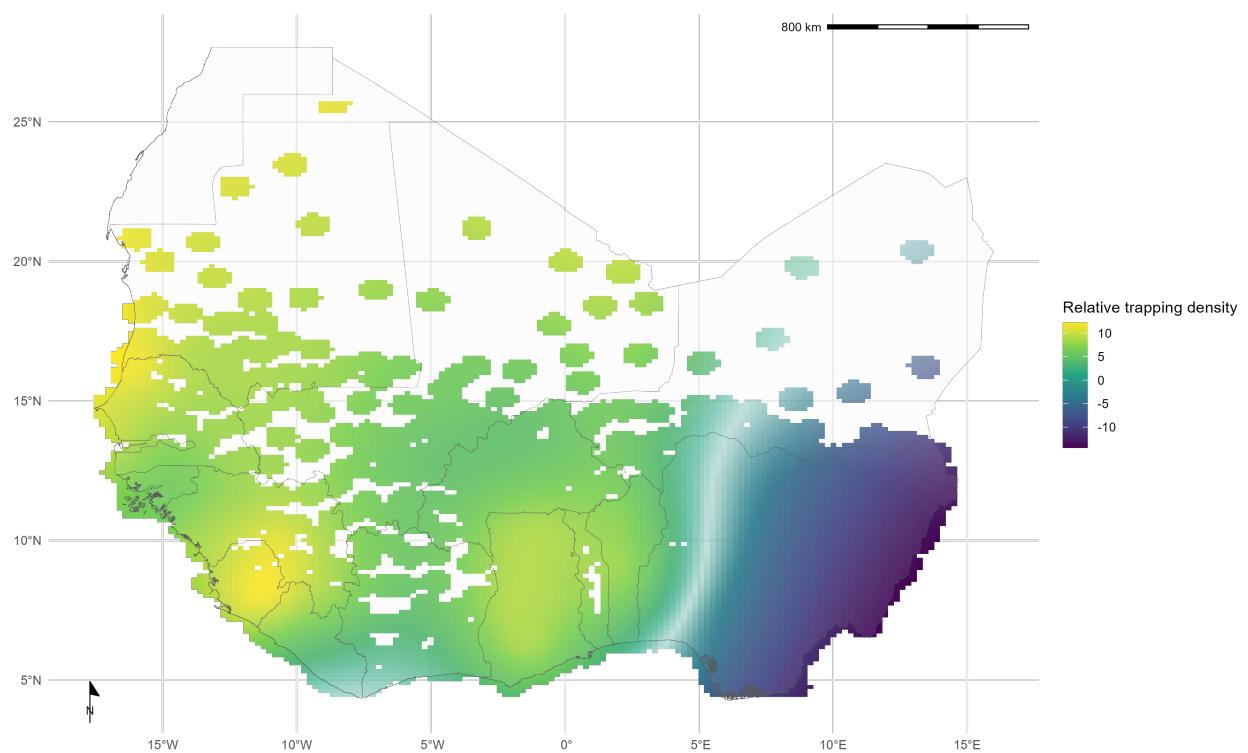
Supplementary Figure 1. Green points represent the start date of rodent trapping studies, blue points representing the final trapping activity. Red points indicate the publication of studies. Increasing numbers of studies have been published since 2000 with more studies being conducted over repeated visits.

Supplementary Figure 2

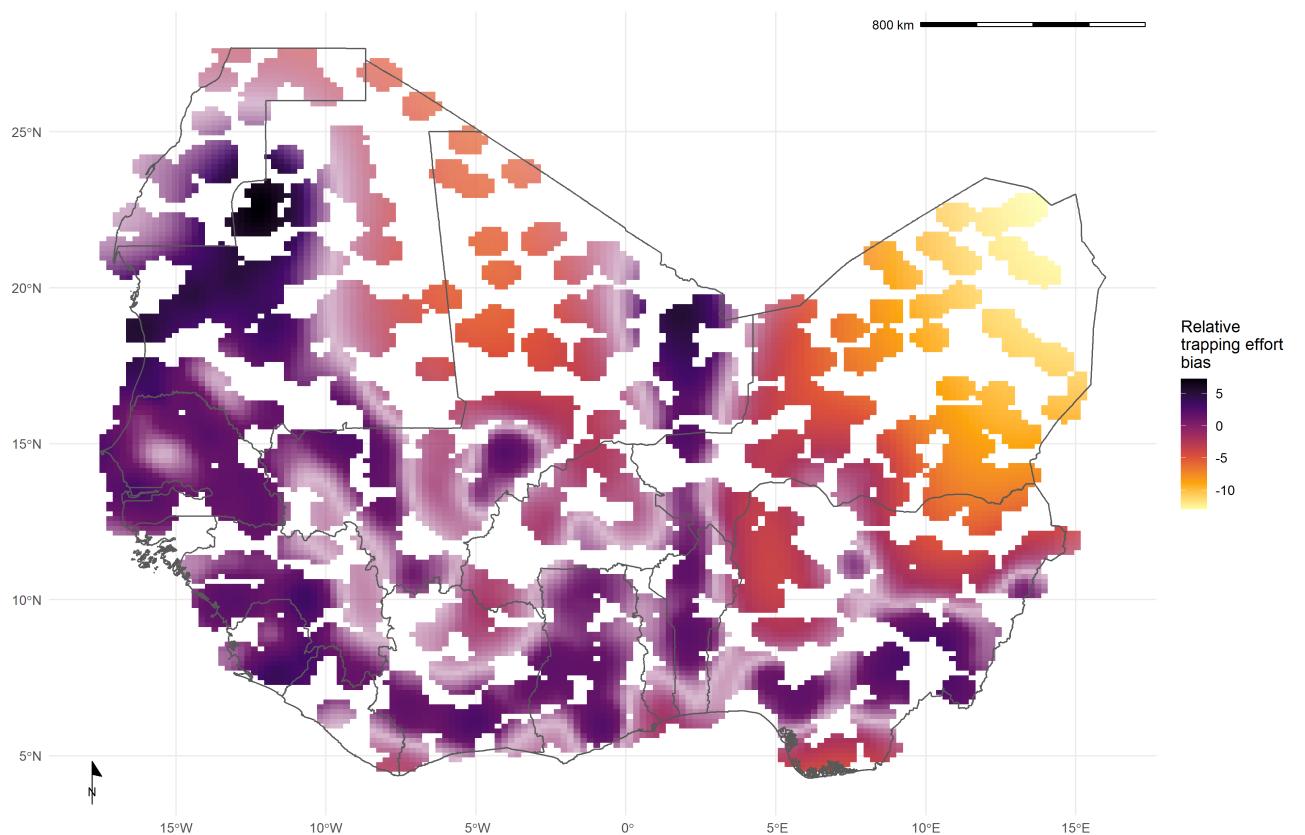


Supplementary Figure 2. Comparison between proportion of land use between regions trapped and all land use in West Africa. Cropland, Shrubland and Grassland were trapped at than representative rates while Bare habitats, Forested regions and Urban landscapes were trapped at lower than representative rates.

Supplementary Figure 3

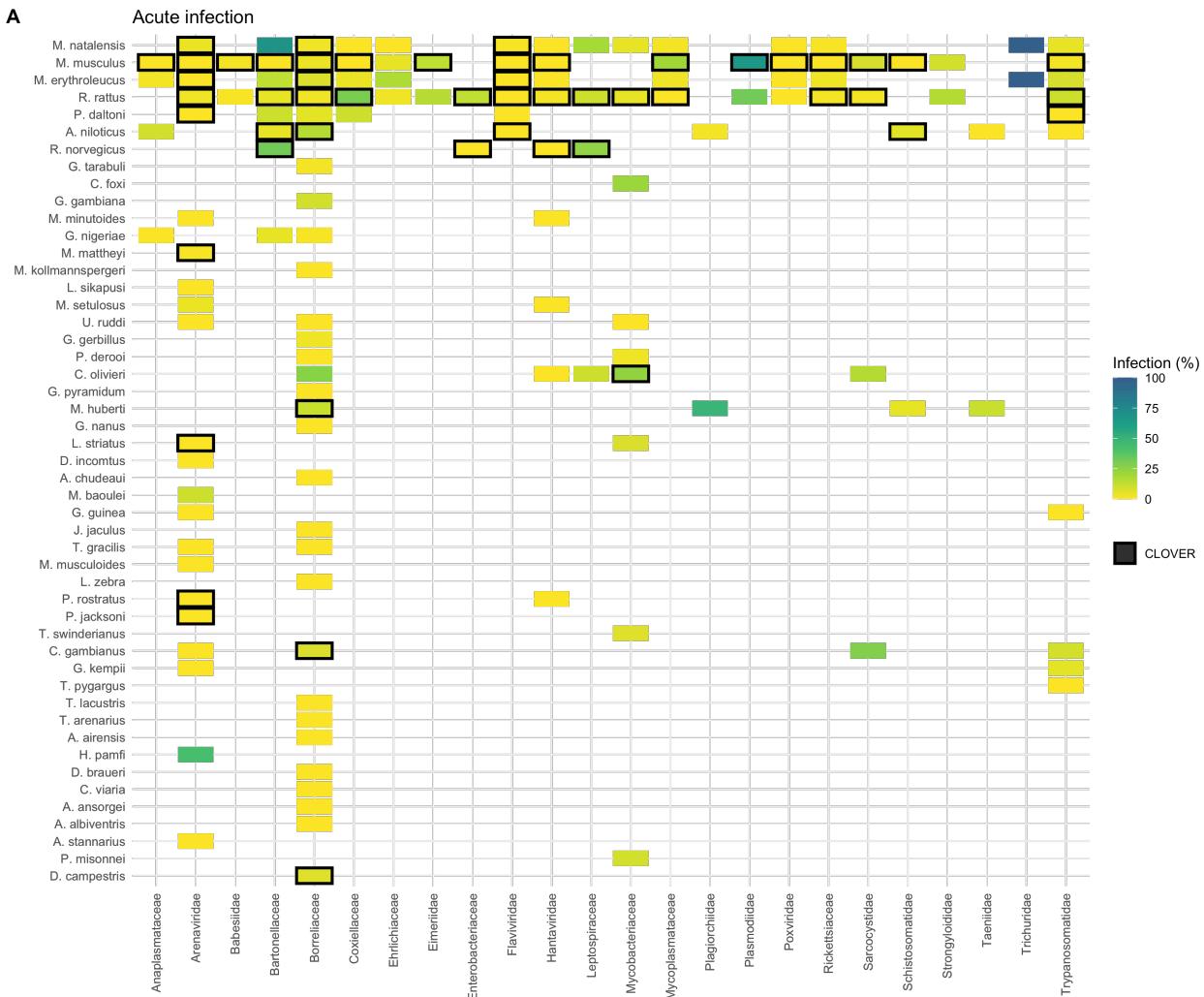


Supplementary Figure 3A. Relative trapping effort bias across West Africa from the subset of included studies reporting trapping effort adjusted for human population density, proportion urban land cover and area of the administrative region. Uncertainty in the estimate of bias is represented by colour transparency/pallor. Yellow regions represent areas with higher than expected trapping effort, Blue regions represent areas lower than expected trapping effort



Supplementary Figure 3B. Relative trapping effort bias across West Africa using a pixel based analysis adjusted for human population density, proportion urban land cover and area of the administrative region. Predictions are limited to areas around trap sites (coloured areas), uncertainty in the estimate of bias is represented by colour transparency/pallor. Purple regions represent areas with higher than expected trapping effort, yellow regions represent areas lower than expected trapping effort

Supplementary Figure 4





Supplementary Figure 4. A) Identified host-pathogen associations at pathogen family level through detection of acute infection (i.e. PCR, culture). Percentages and colour relate to the proportion of all assays that were positive. Associations with a black border are present in the CLOVER dataset. B) Identified host-pathogen associations at pathogen family level through serological assays (i.e. ELISA). Percentages and colour relate to the proportion of all assays that were positive. Associations with a black border are present in the CLOVER dataset.