Rodent trapping studies as an overlooked information source for understanding endemic and novel zoonotic spillover.

Rodent trapping to understand zoonotic spillover risk

David Simons[1, 2, 3]

Lauren A. Attfield[3, 5]

Kate E. Jones[3]

Deborah Watson-Jones[2, 4]

Richard Kock[1]

[1] Centre for Emerging, Endemic and Exotic Diseases, The Royal Veterinary College, London, UK

[2] Clinical Research Department, London School of Hygiene and Tropical Medicine, London, UK

[3] Centre for Biodiversity and Environment Research, University College London, London, UK

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

[5] Department of Life Sciences, Imperial College London, London, UK.

1. Abstract

Rodents a diverse, globally distributed and ecologically important order of mammals are nevertheless important reservoirs of known and novel zoonotic pathogens. Ongoing anthropogenic land use change is altering the composition of these species' abundance and distribution which among zoonotic host species may alter the risk of zoonoses spillover events. These changes demand that a better understanding of the current distribution of rodent species is vital for mitigating potential zoonotic disease hazard and managing risk. However, available species distribution and host-pathogen association datasets (e.g. IUCN, GBIF, CLOVER) are often taxonomically and spatially biased. Here, we synthesise data from West Africa from 127 rodent trapping studies, published between 1964-2022, as an additional source of information to characterise the range and presence of rodent species and identify the subgroup of species that are potential or known pathogen hosts. We identify that these rodent trapping studies, although biased towards human dominated landscapes across West Africa, can usefully complement current rodent species distribution datasets and we calculate the discrepancies between these datasets. For five regionally important zoonotic pathogens (Arenaviridae spp., Borrelia spp., *Lassa mammarenavirus*, Leptospira spp. and *Toxoplasma gondii*), we identify host-pathogen associations that have not been previously reported in host-association datasets. Finally, for these five pathogen groups, we find that the proportion of a rodent hosts range that has been sampled remains small with geographic clustering. A priority should be to sample rodent hosts across a greater geographic range to better characterise current and future risk of zoonotic spillover events. In the interim, studies of spatial pathogen risk informed by rodent distributions 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.

2. Author Summary

Emerging and endemic zoonotic diseases are projected to have increasing health impacts, particularly under changing climate and land-use scenarios. Rodents, an ecologically vital order of mammals carry a disproportionate number of zoonotic pathogens and are abundant across West Africa. Prior modelling studies rely on large, consolidated data sources which do not incorporate high resolution spatial and temporal data from rodent trapping studies. Here, we synthesise these studies to quantify the bias in the sampling of rodent hosts and their pathogens across West Africa. We find that rodent trapping studies are complementary to these datasets and can provide additional, high-resolution data on the distribution of hosts and their pathogens. Further, rodent trapping studies have identified additional potential host-pathogen associations than are recorded in consolidated host-pathogen association datasets. This can help to understand the risk of zoonotic diseases based on host distributions. Finally, we quantify the current extent of known rodent presence and pathogen sampling within a species range, highlighting that current knowledge is limited across much of the region. We hope that this will support work to study rodent hosts and their pathogens in currently under sampled regions to better understand the risk of emerging and endemic zoonoses in West Africa.

56 ## 3. Introduction

57 There is increasing awareness of the global health and economic impacts of novel zoonotic
58 pathogen spillover, driven by the ongoing SARS-CoV-2 pandemic and previous HIV/AIDs and
59 Spanish Influenza pandemics [1]. The number of zoonotic disease spillover events and the
60 frequency of the emergence of novel zoonotic pathogens from rodents are predicted to increase
61 under intensifying anthropogenic pressure driven by increased human populations,
62 urbanisation, intensification of agriculture, climate change leading to altered rodent species
63 assemblages [2–5]. The impact of endemic zoonoses meanwhile remains underestimated [6].
64 Endemic zoonoses disproportionally affect those in the poorest sections of society, those living
65 in close contact with their animals and those with limited access to healthcare [7–9].

66 Rodents along with bats contribute the greatest number of predicted novel zoonotic pathogens
67 and known endemic zoonoses [10,11]. Of 2,220 extant rodent species, 244 (10.7%) are
68 described as reservoirs of 85 zoonotic pathogens [10]. Most rodent species do not provide a
69 direct risk to human health and all species provide important and beneficial ecosystem services
70 including pest regulation and seed dispersal [12]. Increasing risks of zoonotic spillover events
71 are driven by human actions rather than by rodents, for example, invasive rodent species being
72 introduced to novel ranges through human transport routes. Rodents typically demonstrate
73 "fast" life histories which allow them to exploit opportunities provided by anthropogenic
74 disturbance [13]. Within rodents, species level traits such as early maturation and short
75 gestation times are associated with increased probabilities of being zoonotic reservoirs [10,14].
76 Rodent species with these traits are able to thrive in human dominated landscapes, displacing
77 species less likely to be reservoirs of zoonotic pathogens [15]. The widespread occurrence of
78 reservoir species and their proximity to human activity make the description of rodent species

79 assemblages and host-pathogen associations vitally important to understanding the hazard of
80 zoonotic disease spillover and novel zoonotic pathogen emergence [16].

81 Despite the importance of understanding these complex systems, current evidence on host-
82 pathogen associations is considerably affected by taxonomic and geographical sampling biases
83 [11,17]. Curated biodiversity datasets such as the Global Biodiversity Information Facility
84 (GBIF) and resources produced by the International Union for Conservation of Nature (IUCN)
85 suffer from well described spatial and temporal sampling biases [18,19]. These data are
86 typically obtained from museum specimen collections and non-governmental organisation
87 surveys. These sampling biases can importantly distort produced species distribution models
88 that are used to infer risk of zoonotic disease risk [20]. Datasets on host-pathogen associations
89 (i.e. CLOVER) also can suffer from biases introduced from literature selection criteria and
90 taxonomic discrepancies resulting in differential likelihood of accurate host-pathogen
91 attribution by host species. These biases are important because identification of potential
92 geographic hotspots of zoonotic disease spillover and novel pathogen emergence are often
93 produced from these types of host species distributions and host-pathogen associations
94 [21,22]. For example, systematically increased sampling, over-representation of certain
95 habitats and clustering around areas of high human population could lead to an apparent
96 association between locations and hazard that is driven by these factors rather than underlying
97 host-pathogen associations [11,23,24]. Predictions of zoonotic disease spillover and novel
98 zoonotic pathogen emergence must account for these biases to understand the future hazard of
99 zoonotic diseases [22].

100 West Africa has been identified as a region at increased risk for rodent-borne zoonotic disease
101 spillover events, the probability of these events are predicted to increase under different

102 projected future land-use change scenarios [4,25]. Currently within West Africa, some rodent
103 species are known to be involved in the transmission of multiple endemic zoonoses with large
104 burdens on human health, these pathogens include Lassa fever, Schistosomiasis, Leptospirosis
105 and Toxoplasmosis [26,27]. The presence of other species within shared habitats may mitigate
106 the spread of these pathogens through the "dilution effect", ongoing loss of biodiversity may
107 further increase the risk to human populations [5]. Understanding of the distribution of these
108 zoonoses are limited by biases in consolidated datasets. Rodent trapping studies provide
109 contextually rich information on when, where and under what conditions rodents were
110 trapped, potentially enriching consolidated datasets [28]. Studies have been conducted in West
111 Africa to investigate the distribution of rodent species, their species assemblages, the
112 prevalence of endemic zoonoses within rodent hosts (e.g., Lassa fever, Schistosomiasis) and to
113 identify emerging and novel zoonotic pathogens [29–31]. However, individual level data from
114 these studies have not previously been synthesised for inclusion in assessments of zoonotic
115 disease spillover and novel zoonotic pathogen emergence.

116 Here, we synthesise rodent trapping studies conducted across West Africa published between
117 1964-2022. First, we use this dataset to investigate the geographic sampling biases in relation
118 to human population density and land use classification. Second, we compare this to curated
119 host datasets (IUCN and GBIF) to understand differences in reported host geographic
120 distributions. Third, we compare identified host-pathogen associations with a consolidated
121 dataset (CLOVER) to explore discrepancies in rodent host-pathogen associations and report the
122 proportion of positive assays for pathogens of interest. Finally, within our dataset we
123 investigate the spatial extent of current host-pathogen sampling to identify areas of sparse
124 sampling of pathogens within their host ranges. We expect that rodent trapping studies provide

125 an important additional source of high-resolution data that can be used to enrich available
126 consolidated datasets to better understand the hazard of zoonotic disease spillover and novel
127 zoonotic pathogen emergence across West Africa.

128 4. Methods

129 4.1. Data sources

130 4.1.1. Host and pathogen trapping data

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

138 We included studies for further analysis if they met all of the following inclusion criteria; i)
139 reported findings from trapping studies where the target was a small mammal, ii) described the
140 type of trap used or the length of trapping activity or the location of the trapping activity, iii)
141 included trapping activity from at least one West African country, iv) recorded the genus or
142 species of trapped individuals, and v) were published in a peer-reviewed journal or as a pre-
143 print on a digital platform or as a report by a credible organisation. We excluded studies if they
144 met any of the following exclusion criteria: i) reported data that were duplicated from a

145 previously included study, ii) no full text available, iii) not available in English. One author
146 screened titles, abstracts and full texts against the inclusion and exclusion criteria. At each
147 stage; title screening, abstract screening and full text review, a random subset (10%) was
148 reviewed by a second author.

149 We extracted data from eligible studies using a standardised tool that was piloted on 5 studies
150 (Supplementary Table 1.). Data was abstracted into a Google Sheets document, which was
151 archived on completion of data extraction [32]. We identified the aims of included studies, for
152 example, whether it was conducted as a survey of small mammal species or specifically to
153 assess the risk of zoonotic disease spillover. we extracted data on study methodology, such as,
154 the number of trap nights, the type of traps used and whether the study attempted to estimate
155 abundance. For studies not reporting number of trap nights we used imputation based on the
156 number of trapped individuals, stratified by the habitat type from which they were obtained.
157 This was performed by multiplying the total number of trapped individuals within that study
158 site by the median trap success for study sites with the same reported habitat type.
159 Stratification was used as trap success varied importantly between traps placed in or around
160 buildings (13%, IQR 6-24%) compared with other habitats (3%, IQR 1-9%)

161 We also recorded how species were identified within a study and species identification was
162 assumed to be accurate. The number of individuals of these species or genera was extracted
163 with taxanomic names mapped to GBIF taxonomy [33]. We expanded species detection and
164 non-detection records by explicitly specifying non-detection at a trap site if a species was
165 recorded as detected at other trapping locations within the same study.

166 Geographic locations of trapping studies were extracted using GPS locations for the most
167 precise location presented. Missing locations were found using the National Geospatial-

168 Intelligence Agency GEOnet Names Server [34] based on placenames and maps presented in the
169 study. All locations were converted to decimal degrees. The year of rodent trapping was
170 extracted alongside the length of the trapping activity to understand seasonal
171 representativeness of trapping activity. The habitats of trapping sites were mapped to the IUCN
172 Habitat Classification Scheme (Version 3.1). For studies reporting multiple habitat types for a
173 single trap, trap-line or trapping grid, a higher order classification of habitat type was recorded.

174 For included studies with available data we extracted information on all microorganisms and
175 known zoonotic pathogens tested and the method used (e.g., molecular or serological
176 diagnosis). Where assays were able to identify the microorganism to species level this was
177 recorded, non-specific assays higher order attribution was used (e.g. to family level). A broad
178 definition of known zoonotic pathogen was used, a species of microorganism carried by an
179 animal that may transmit to humans and cause illness [35]. We do not include evolved
180 pathogens acquired originally through zoonotic pathways in our definition (i.e. HIV). The term
181 microorganism is used where either the microorganism is not identified to species level, in
182 which case it remains unclear whether it is a zoonotic pathogen (i.e. Arenaviridae), or the
183 species is not known to be a zoonotic pathogen (i.e. *Candidatus Ehrlichia senegalensis*). We
184 recorded the species of rodent host tested, the number of individuals tested and the number of
185 positive and negative results. For studies reporting summary results all testing data were
186 extracted, this may introduce double counting of individual rodents, for example, if a single
187 rodent was tested using both molecular and serological assays. Where studies reported
188 indeterminate results, these were also recorded.

189 **4.1.2. Description of included studies**

190 Out of 4,692 relevant citations, we identified 127 rodent trapping studies (Supplementary
191 Table 2.). Of these, 55 (43%) were conducted to investigate rodent-borne zoonoses, with the
192 remaining 77 (57%) conducted for ecological purposes (i.e. population dynamics, distribution)
193 in rodents, including those known to be hosts of zoonotic pathogens. The earliest trapping
194 studies were conducted in 1964, with a trend of increasing numbers of studies being performed
195 annually since 2000. The median year of first trapping activity was 2007, with the median
196 length of trapping activity being 1 year (IQR 0-2 years) (Supplementary Fig 1.). Studies were
197 conducted in 14 West African countries, with no studies reported from The Gambia or Togo, at
198 1,611 trap sites (Fig 1A.).

199 Included studies explicitly reported on 601,184 trap nights, a further 341,445 trap nights were
200 imputed from studies with no recording of trapping effort based on trap success, leading to an
201 estimate of 942,629 trap nights (Fig 1B.). A minority of studies trapped at a single study site
202 (30, 24%), with 46 (36%) trapping at between two and five sites, the remaining 51 studies
203 (40%) trapped at between six and 93 study sites.

204 In total 76,275 small mammals were trapped with 65,628 (90%) identified to species level and
205 7,439 (10%) identified to genus, with the remaining classified to higher taxonomic level. The
206 majority of the 132 identified species were Rodentia (102, 78%), of which Muridae (73, 72%)
207 were the most common family. Soricomorpha were the second most identified order of small
208 mammals (28, 21%). 57 studies tested for 32 microorganisms, defined to species or genus level
209 that are known or potential pathogens. Most studies tested for a single microorganism (48,
210 84%). The most frequently assayed microorganisms were *Lassa mammarenavirus* or

211 Arenaviridae (21, 37%), *Borrelia sp.* (9, 16%), *Bartonella sp.* (4, 7%) and *Toxoplasma gondii* (4,
212 7%). Most studies used Polymerase Chain Reaction (PCR) to detect microorganisms (37, 65%),
213 with fewer studies using serology-based tests (11, 19%) or histological or direct visualisation
214 assays (11, 21%). From 32,920 individual rodent samples we produced 351 host-pathogen
215 pairs. With *Rattus rattus*, *Mus musculus*, *Mastomys erythroleucus*, *Mastomys natalensis* and
216 *Arvicanthis niloticus* being assayed for at least 18 microorganisms.

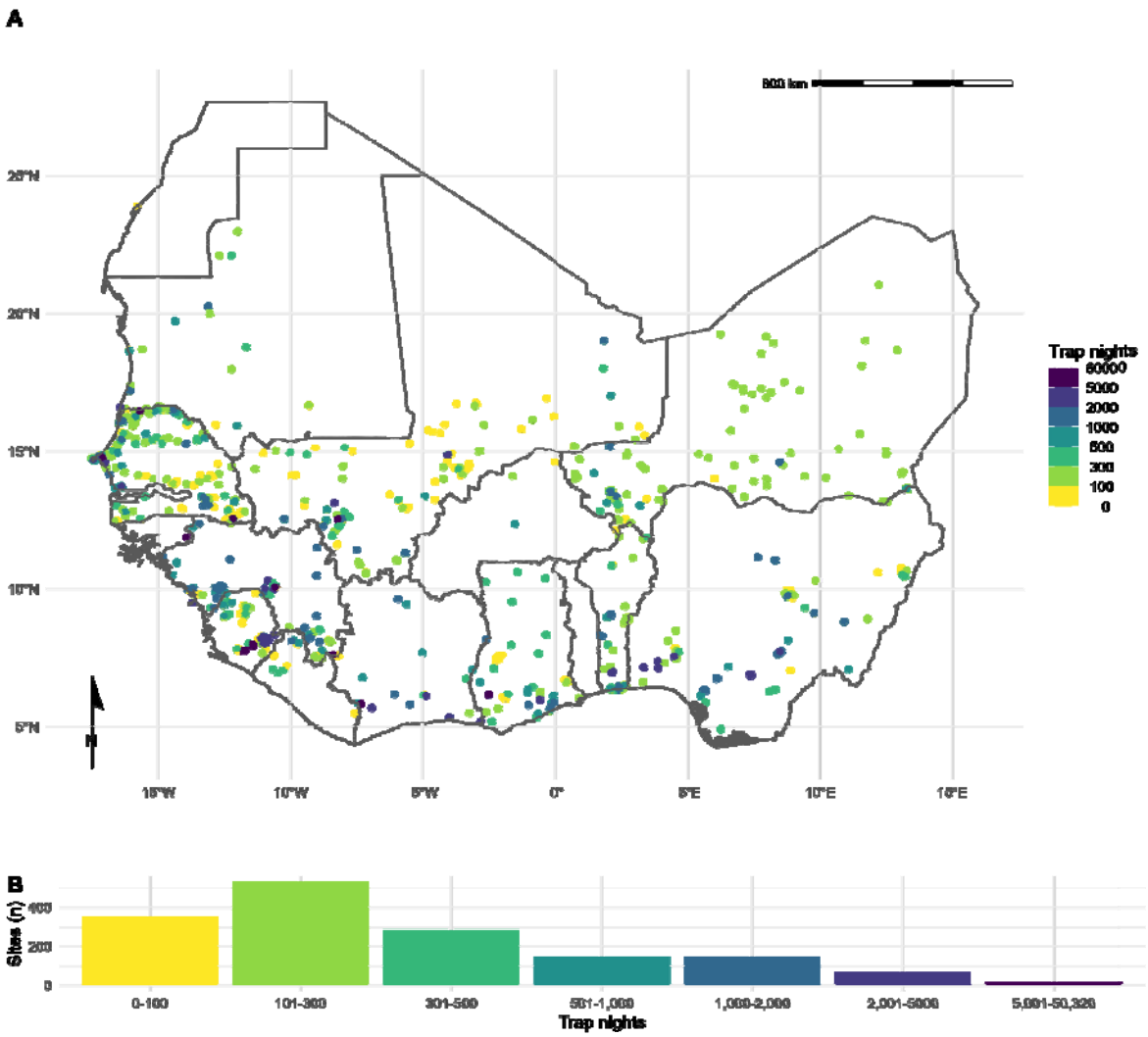


Fig 1: Rodent trapping sites across West Africa. A) The location of trapping sites in West Africa. No sites were recorded from Togo or The Gambia. Heterogeneity is observed in the coverage of each country by trap night (colour) and location of sites. For example, Senegal, Mali and Sierra Leone have generally 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. A labelled map of the study region is attached in Supplementary Figure 5.

4.2. Analysis

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

To investigate the extent of spatial bias in the rodent trapping data, we calculated trap-night (TN) density within each West African level-2 administrative region. The `sf` package in the R statistical language (R version 4.1.2) was used to manipulate geographic data, administrative boundaries were obtained from GADM 4.0.4 [36–38]. Trap-night density ($TN_{density}$) was calculated by dividing the number of trap nights by the area of a level-2 administrative area (R_{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 ($P_{density}$) [39]. 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 [40]. The proportion of cropland, shrubland, tree cover (ψ_{tree}) and urban land cover (ψ_{urban}) within a level-2 administrative region in 2005 was calculated.

We investigated the association between relative trapping effort, measured as TN density, and the proportion of urban, cropland, tree cover and human population density using Generalised Additive Models (GAM) incorporating a spatial interaction term (longitude and latitude) [41]. Spatial aggregation of relative trapping effort was modelled using an exponential dispersion distribution (Tweedie) [42]. The models were constructed in the `mgcv` package [43]. Selection of the most parsimonious model was based on Deviance explained and the Akaike Information Criterion for each model (Equations 1-5 below). Relative trapping effort was then predicted

247 across West Africa using these covariates. We performed two sensitivity analyses, first, by
248 removing sites with imputed trapping effort, second, by associating trap locations to ~1km
249 pixels rather than level-2 administrative areas.

$$\mathrm{TN}_{\mathrm{density}} \sim \mathrm{Tweedie}(X * Y) \ \#(1)$$

$$\mathrm{TN}_{\mathrm{density}} \sim \mathrm{Tweedie}\left(\mathrm{P}_{\mathrm{density}} + (X * Y)\right) \ \#(2)$$

$$\mathrm{TN}_{\mathrm{density}} \sim \mathrm{Tweedie}\left(\mathrm{P}_{\mathrm{density}} + \mathrm{R}_{\mathrm{area}} + (X * Y)\right) \ \#(3)$$

$$\mathrm{TN}_{\mathrm{density}} \sim \mathrm{Tweedie}\left(\mathrm{P}_{\mathrm{density}} + \psi_{\mathrm{tree}} + \psi_{\mathrm{urban}} + (X * Y)\right) \ \#(4)$$

$$\mathrm{TN}_{\mathrm{density}} \sim \mathrm{Tweedie}\left(\mathrm{P}_{\mathrm{density}} + \mathrm{R}_{\mathrm{area}} + \psi_{\mathrm{urban}} + (X * Y)\right) \ \#(5)$$

250

251 **4.2.2. What is the difference in rodent host distributions between curated**
252 **datasets and rodent trapping studies?**

253 We assessed the concordance of curated rodent host distributions from IUCN and GBIF with
254 observed rodent detection and non-detection from rodent trapping studies for seven species
255 with the most trap locations (*M. natalensis*, *R. rattus*, *M. erythroleucus*, *M. musculus*, *A. niloticus*,
256 *Praomys daltoni* and *Cricetomys gambianus*). We obtained rodent species distribution maps as
257 shapefiles from the IUCN red list and translated these to a ~20km resolution raster [44].
258 Distributions were cropped to the study region for globally distributed rodent species. We
259 obtained rodent presence locations from GBIF as point data limited to the study region [45].

260 Presence locations were associated to cells of raster with a ~20km resolution produced for the
261 study region.

262 For each of the seven species, we first calculated the area of the IUCN expected range, and then
263 the percentage of this range covered by presence detections in GBIF, and from detections in the
264 rodent trapping data. We then calculated the area of both types of both GBIF and rodent
265 trapping detections outside of the IUCN expected range. For rodent trapping data, we
266 additionally calculated the area of non-detections within the IUCN expected area. Finally, we
267 calculated the combined area of detection from both GBIF and rodent trapping data.

268 4.2.3. Are rodent trapping derived host-pathogen associations present in a 269 consolidated zoonoses dataset?

270 To examine the usefulness of rodent trapping studies as an additional source of data we
271 compared identified host-pathogen associations from trapping studies investigating zoonoses
272 with a consolidated zoonoses dataset (CLOVER) [11,46]. CLOVER is a synthesis of four host-
273 pathogen datasets (GMPD2, EID2, HP3 and Shaw, 2020) and was released in 2021, it contains
274 more than 25,000 host-pathogen associations for Bacteria, Viruses, Helminth, Protozoa and
275 Fungi. We compared the host-pathogen networks across the two datasets, where the CLOVER
276 data was subset for host species present in the rodent trapping data.

277 For host-pathogen pairs with assay results consistent with acute or prior infection, we
278 calculated the proportion positive and identify those absent from CLOVER. We expand the
279 analysis to host-pathogen pairs with pathogens identified to genus level in Supplementary Fig
280 4.

281 #### 4.2.4. What is the spatial extent of pathogen testing within host ranges?

282 We use the sampled area of three pathogen groups and two pathogens (Arenaviridae,
283 Borreliaceae, Leptospiraceae, *Lassa mammarenavirus* and *Toxoplasma gondii*) to quantify the
284 bias of sampling within their hosts ranges. For each pathogen, we first describe the number of
285 host species assayed, for the five most commonly tested species we associate the locations of
286 sampled individuals to ~20km pixels and calculate the proportion of the IUCN range of the host
287 in which sampling has occurred. We compare this figure to the total area in which the host has
288 been detected to produce a measure of relative completeness of sampling within the included
289 rodent trapping studies.

290 Data and code to reproduce all analyses are available in an archived Zenodo repository [32].

291 ## 5. Results

292 ### 5.1. What is the extent of spatial bias in the rodent trapping data?

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

300 The most parsimonious GAM model (adjusted R2 = 0.3, Deviance explained = 48.7%) reported
301 significant non-linear associations between relative trapping effort bias and human population
302 densities (Effective Degrees of Freedom (EDF) = 7.13, $p < 0.001$), proportion of urban landscape
303 (EDF = 1.92, $p < 0.002$) and region area (EDF = 3.63, $p < 0.001$), alongside significant spatial
304 associations (EDF = 27.3, $p < 0.001$) (Supplementary table 3.1). Greatest trapping effort bias
305 peaked at population densities between 5,000-7,500 individuals/km^2, proportion of urban
306 landscape >10% and region areas < 1,000km^2. Increased trapping effort was found in North
307 West Senegal, North and East Sierra Leone, Central Guinea and coastal regions of Nigeria, Benin
308 and Ghana; in contrast South East Nigeria, Northern Nigeria and Burkina Faso had an observed
309 bias towards a reduced trapping effort (Fig 2). In sensitivity analysis, excluding sites with
310 imputed trap nights, Mauritania, Northern Senegal and Sierra Leone remained as regions
311 trapped at higher rates, with Nigeria being trapped at lower than expected rates
312 (Supplementary Fig 3a.). In pixel-based sensitivity analysis spatial coverage was reduced with
313 similar patterns of bias observed to the primary analysis (Supplementary Fig 3b.).

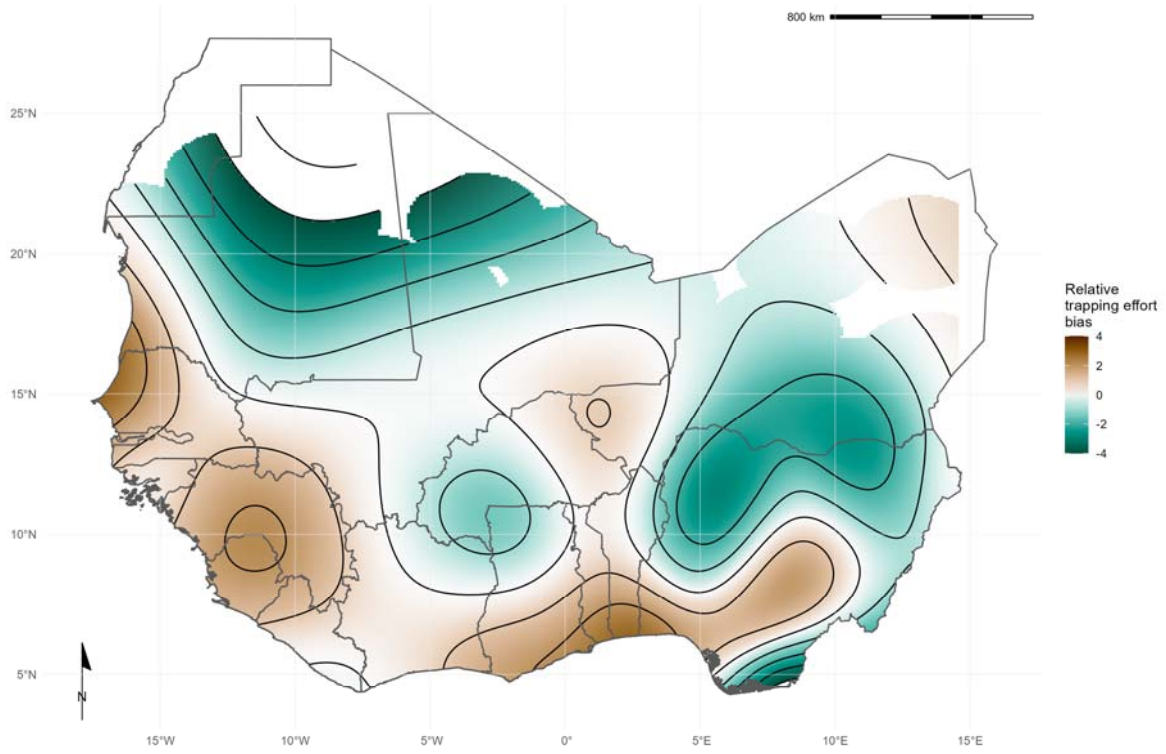
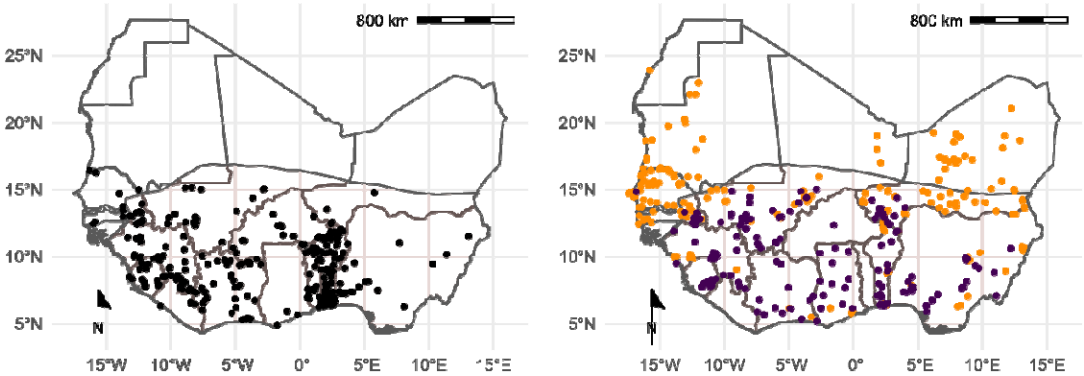


Fig 2. Relative trapping effort bias across West Africa. Modelled relative trapping effort bias adjusted for human population density, proportion urban land cover and area of the administrative region. Brown regions represent areas with a bias towards increased trapping effort (e.g., North West Senegal), Green regions represent areas with a bias towards reduced trapping effort (e.g., Northern Nigeria).

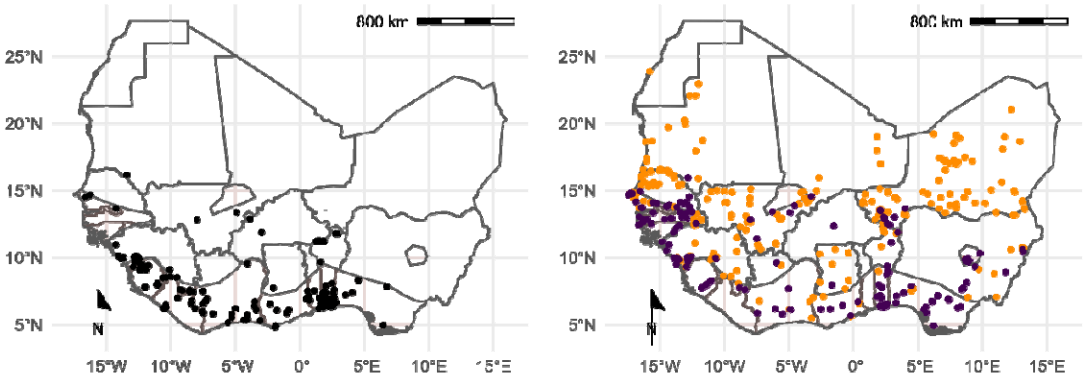
320 ## 5.2. What is the difference in rodent host distributions between
321 curated datasets and rodent trapping studies?

322 We found that for six of the seven most detected rodent species (*M. natalensis*, *R. rattus*, *M.*
323 *erythroleucus*, *M. musculus*, *A. niloticus* and *P. daltoni*), trapping studies provided more distinct
324 locations of detection and non-detection than were available from GBIF. For the endemic
325 rodent species (*M. natalensis*, *M. erythroleucus*, *A. niloticus*, *P. daltoni* and *C. gambianus*) IUCN
326 ranges had good concordance to both trapping studies and GBIF, however, individuals of *A.*
327 *niloticus* and *P. daltoni* were detected outside of IUCN ranges. In contrast, the non-native
328 species *R. rattus* and *M. musculus* were detected across much greater ranges than were
329 expected from IUCN distributions. Comparisons for *M. natalensis*, *R. rattus* and *M. musculus* are
330 shown in Fig 3, the remaining species are shown in Supplementary Fig 4.

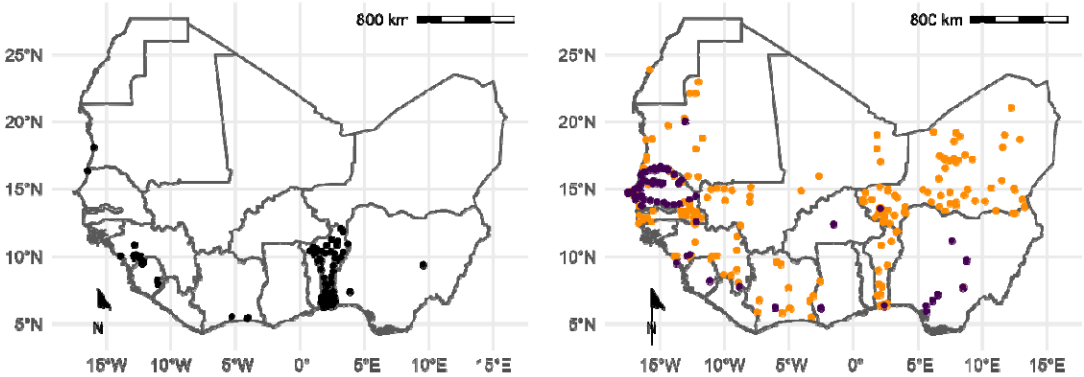
Mastomys natalensis



Rattus rattus



Mus musculus



Detection/Non-detection ● Detection ● Non-detection

Fig 3. Locations of detection and non-detection sites for rodent species in West Africa

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

Comparison of the proportion of a species IUCN range in which detections and non-detections occurred showed that sampling locations of these seven species within GBIF covered between 0.09-0.26% of expected ranges (Table 1.), compared to 0.03-0.24% for rodent trapping data. Detections occurred outside IUCN ranges for all species in both the GBIF and rodent trapping data, most noticeably for *A. niloticus* and *R. rattus*. Combining GBIF and rodent trapping data increased the sampled area by a mean of 1.6 times compared to the GBIF area alone, demonstrating limited overlap between the locations providing information to either dataset. Non-detection of a species occurred across species ranges (mean = 0.11%, SD = 0.03%), suggestive of spatial heterogeneity of presence within IUCN ranges.

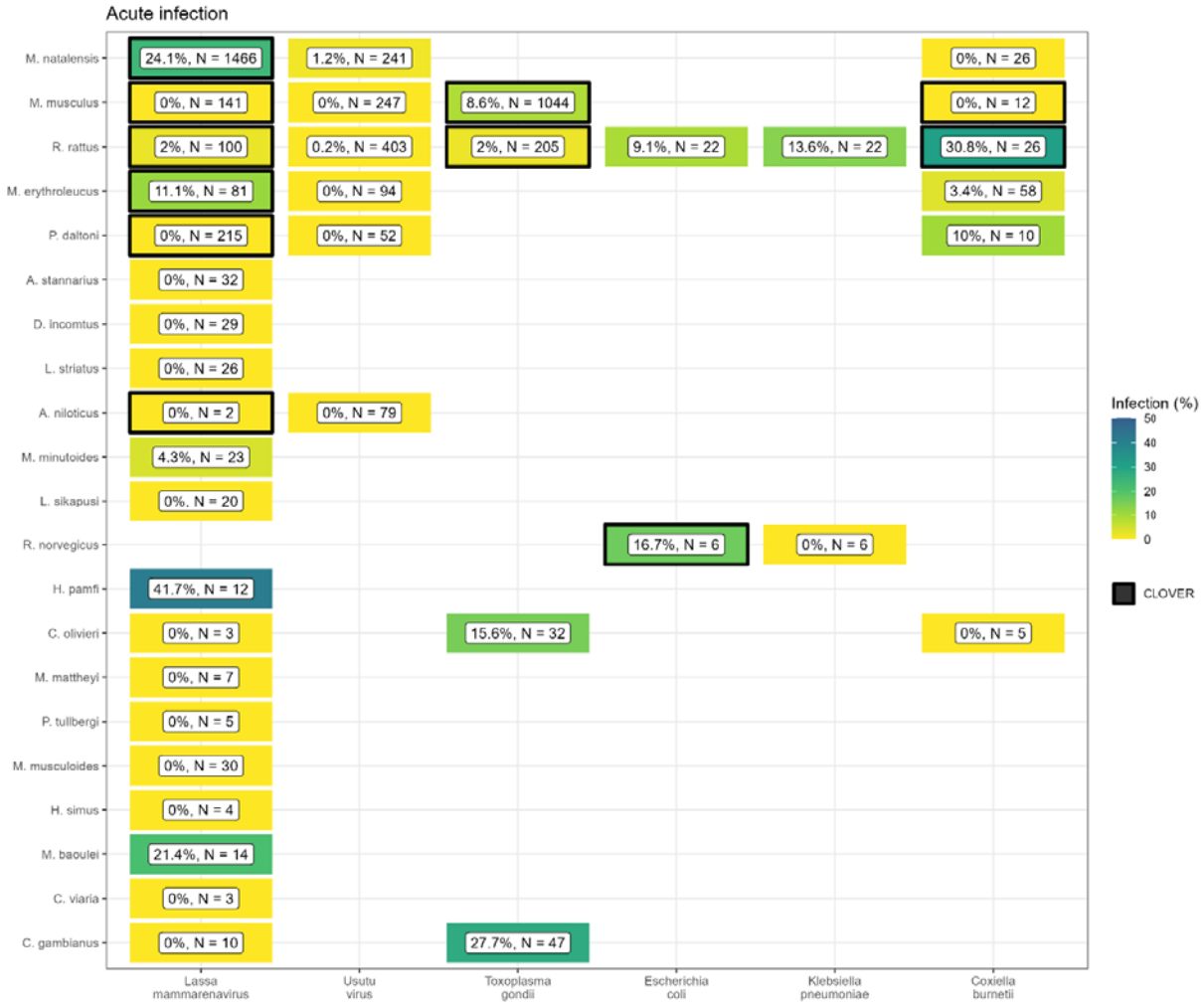
Table 1: Comparison of IUCN, GBIF and rodent trapping ranges for the 7 most detected rodent species.

	IUCN	GBIF		Trapping studies			Combined
Species	Range (1,000 km^2)	Area inside range (1,000 km^2) (% of IUCN range)	Area outside range (1,000 km^2)	Detection area inside range (1,000 km^2) (% of IUCN range)	Area outside range (1,000 km^2) (% of IUCN range)	Non-detection area inside range (1,000 km^2) (% of IUCN range)	Detection area (1,000 km^2) (% of IUCN range)
Mastomys natalensis	3,257	6.83 (0.21%)	0.19	4.4 (0.14%)	0.17	3.12 (0.1%)	10.66 (0.33%)
Rattus rattus	1,019	2.61 (0.26%)	0.52	2.42 (0.24%)	1.21	1.3 (0.13%)	4.88 (0.48%)
Mastomys erythroleucus	3,735	4.48 (0.12%)	0.04	3.24 (0.09%)	0.12	4.35 (0.12%)	7.5 (0.2%)
Mus musculus			2.15		1.85		3.94
Arvicanthis niloticus	1,829	1.69 (0.09%)	2.41	1.98 (0.11%)	0.34	3.09 (0.17%)	3.59 (0.2%)
Praomys daltoni	2,658	4.03 (0.15%)	0.29	2.03 (0.08%)	0.15	2.78 (0.1%)	5.83 (0.22%)
Cricetomys gambianus	2,476	5 (0.2%)	0.17	0.75 (0.03%)	0.06	2.99 (0.12%)	5.69 (0.23%)

5.3. 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 (Fig 4. and Fig 5.). We identified

354 negative associations (non-detection through specific assays) for 45 host-pathogen pairs
355 among 35 rodent and 7 pathogen species. CLOVER contained 10 (40%) of our identified host-
356 pathogen associations, the remaining 15 (60%) were not found to be present in CLOVER,
357 additionally CLOVER recorded positive associations for 4 (9%) of the negative associations
358 produced from the rodent trapping data.

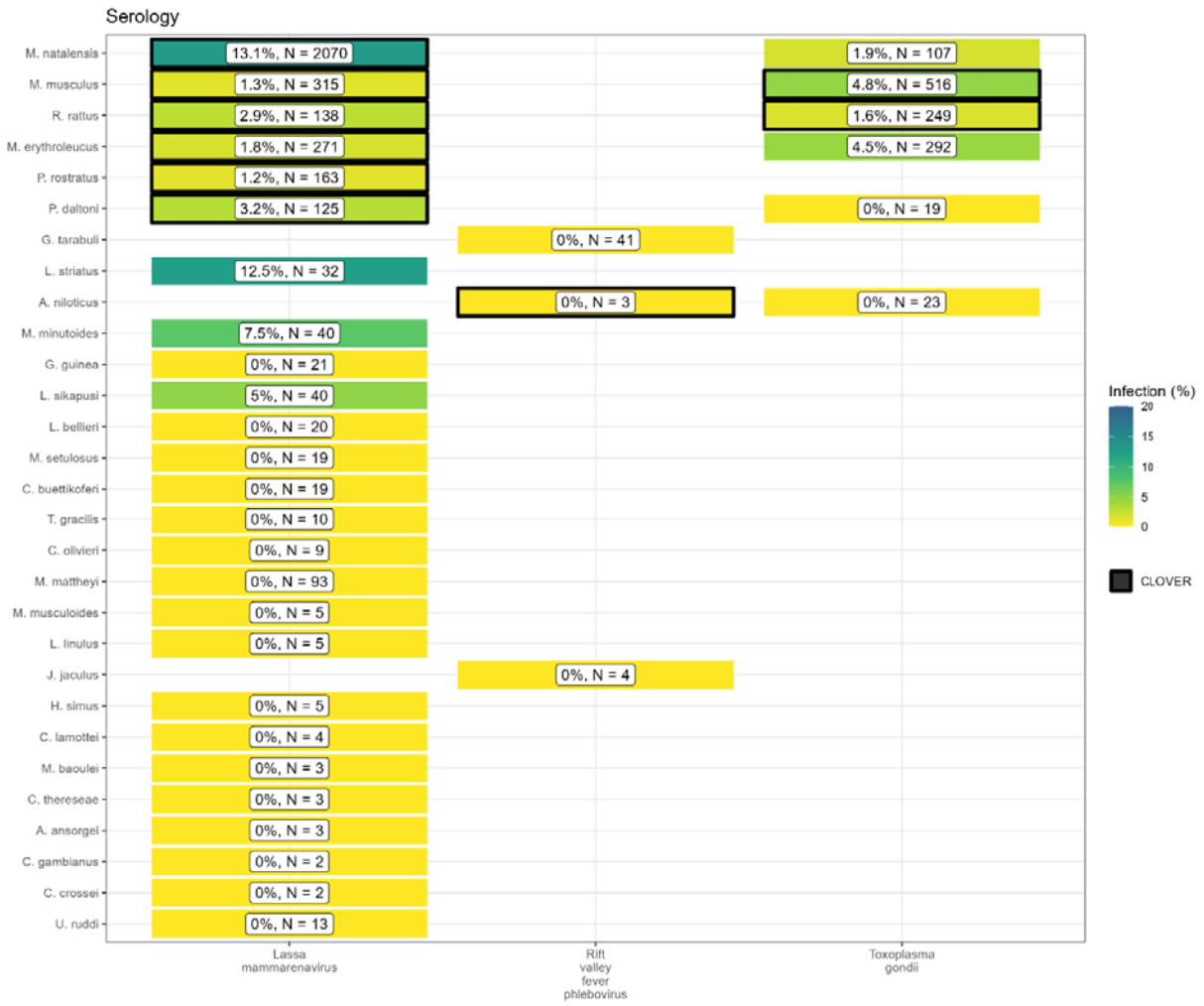


359

360 **Fig 4. Host-Pathogen associations detected through acute infection.** A) Identified species
361 level host-pathogen associations through detection of acute infection (i.e. PCR, culture).
362 Percentages and colour relate to the proportion of all assays that were positive, the number of

363 individuals tested for the pathogen is labelled N. Associations with a black border are present in

364 the CLOVER dataset.



365

366 **Fig 5. Host-Pathogen associations detected through evidence of prior infection.** B)

367 Identified species level host-pathogen associations through serological assays (i.e. ELISA).

368 Percentages and colour relate to the proportion of all assays that were positive, the number of

369 individuals tested for the pathogen is labelled N. Associations with a black border are present in

370 the CLOVER dataset.

371 CLOVER included an additional 492 host-pathogen associations we do not observe in rodent
372 trapping studies. The majority of these 392 (80%) pairs are from species with global
373 distributions (*M. musculus*, *R. rattus* and *R. norvegicus*), or from those with wide ranging
374 distributions in sub-Saharan Africa (38, 8%) (i.e., *A. niloticus*, *M. natalensis* and *Atelerix*
375 *albiventris*).

376 For pathogens not identified to species level (i.e. family or higher taxa only), we identified 148
377 host-pathogen pairs among 32 rodent species and 25 pathogen families (Supplementary Fig 4.),
378 with CLOVER containing 66 (45%) of these associations.

379 Rodent trapping studies identified additional rodent host species for six pathogens; *Lassa*
380 *mammarenavirus* (5), *Toxoplasma gondii* (4), Usutu virus (2), *Coxiella burnetii* (2), *Escherichia*
381 *coli* and *Klebsiella pneumoniae* (both 1), that were not present in this consolidated host-
382 pathogen association dataset.

383 5.4. What is the spatial extent of microorganism testing within a 384 host's range?

385 The five most widely sampled microorganism species/families in included studies were
386 Arenaviridae, Borreliaceae, *Lassa mammarenavirus*, Leptospiraceae and *Toxoplasma gondii*
387 (Table 2.). Assays to identify Arenaviridae infection were performed in 44 rodent species with
388 evidence of viral infection in 15 species. Studies that reported Arenaviridae infection did not
389 identify the microorganism to species level and were distinct from those reporting *Lassa*
390 *mammarenavirus* infection. *Lassa mammarenavirus* was specifically tested for in 43 species
391 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*. These species were assayed across between 10-20% of their trapped area, equating to ~0.02% of their IUCN range (Table 2.).

Infection with species of Borreliaceae was assessed in 42 species, with evidence of infection in 17 rodent species. The greatest rates of infection were among *A. niloticus* (16%), *Mastomys huberti* (11%) and *M. erythroleucus* (9%). Testing was more widespread than for Arenviruses with coverage between 15-34% of their trapped area, however, this remains a small area in relation to their IUCN ranges (<0.05%). Leptospiraceae and *Toxoplasma gondii* was assessed in 8 species, with evidence of infection in 5 and 6 rodent species respectively. The spatial coverage of testing for these microorganisms was more limited within IUCN host species ranges (~0.01%).

Table 2: Comparison of microorganism sampling ranges for the 5 most widely sampled microorganisms and the 5 most sampled rodent host species (* no IUCN range in West African).

Microorganism	Host species	Tested (N)	Positive (N (%))	***Microorganism*** testing area (1,000 km^2)	***Microorganism*** testing area within trapped area (%)	***Microorganism*** testing area within IUCN range (%)
Arenaviridae sp.						
	Mastomys natalensis	2,841	104 (4%)	0.61	13.45%	0.02%
	Praomys daltoni	854	6 (1%)	0.42	19.43%	0.02%
	Mastomys erythroleucus	398	20 (5%)	0.40	11.97%	0.01%

Microorganism	Host species	Tested (N)	Positive (N (%))	***Microorganism*** testing area (1,000 km^2)	***Microorganism*** testing area within trapped area (%)	***Microorganism*** testing area within IUCN range (%)
Borrelia sp.	*Rattus rattus*	396	4 (1%)	0.38	10.5%	0.04%
	Praomys rostratus	310	5 (2%)	0.13	12.53%	0.02%
	Mastomys erythroleucus	1,586	140 (9%)	1.14	33.94%	0.03%
	Arvicanthis niloticus	1,551	253 (16%)	0.66	28.48%	0.03%
	Mastomys natalensis	733	54 (7%)	0.69	15.08%	0.02%
	Mastomys huberti	731	83 (11%)	0.23	29.83%	0.04%
	Mus musculus	686	26 (4%)	0.45	24.54%	*
Lassa mammarenavirus	*Mastomys natalensis*	3,199	580 (18%)	1.03	22.65%	0.03%
	Mastomys erythroleucus	352	14 (4%)	0.36	10.63%	0.01%
	Rattus rattus	177	2 (1%)	0.34	9.26%	0.03%
	Praomys rostratus	163	2 (1%)	0.27	27.02%	0.04%
	Mus musculus	147	0 (0%)	0.04	2.29%	*
Leptospira sp.	*Rattus rattus*	646	65 (10%)	0.40	11.1%	0.04%

Microorganism	Host species	Tested (N)	Positive (N (%))	***Microorganism*** testing area (1,000 km^2)	***Microorganism*** testing area within trapped area (%)	***Microorganism*** testing area within IUCN range (%)
Toxoplasma gondii	*Arvicanthis niloticus*	221	10 (5%)	0.02	0.9%	<0.01%
	Crocidura olivieri	141	14 (10%)	0.34	25.16%	*
	Mastomys natalensis	136	26 (19%)	0.36	7.91%	0.01%
	Rattus norvegicus	79	19 (24%)	0.21	40.08%	*
	Mus musculus	1,548	115 (7%)	0.62	33.64%	*
	Rattus rattus	428	8 (2%)	0.36	9.77%	0.03%
	Mastomys erythroleucus	292	13 (4%)	0.37	11.06%	0.01%
	Mastomys natalensis	107	2 (2%)	0.08	1.83%	<0.01%
	Cricetomys gambianus	47	13 (28%)	0.06	7.6%	<0.01%

6. Discussion

Endemic rodent zoonoses and novel pathogen emergence from rodent hosts are predicted to have an increasing burden in West Africa and globally [10]. Here we have synthesised data from 126 rodent trapping studies containing information on more than 72,000 rodents, from at least 132 species of small mammals, 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

413 risk based on host distributions could counteract some of the biases inherent to these curated
414 datasets [18]. Most assayed rodents were not found to be hosts of known zoonotic pathogens.
415 We identified 25 host-pathogen pairs reported from included studies, 15 of these were not
416 included in a consolidated host-pathogen dataset. Generally, the number of different species
417 tested for a microorganism and the spatial extent of these sampling locations were limited.
418 These findings highlight a number of sampling bias, supporting calls for further microorganism
419 sampling across diverse species in zoonotic hotspots [47].

420 We found that rodent trapping data, like biodiversity data, showed important spatial biases
421 [20]. Relative trapping effort bias was greater in Benin, Guinea, Senegal and Sierra Leone driven
422 by long-standing research collaborations investigating the invasion non-native rodent species
423 (*M. musculus* and *R. rattus*) and the hazard of endemic zoonosis outbreaks (e.g., *Lassa*
424 *mammarenavirus*). In addition to identifying point locations of prior rodent and pathogen
425 sampling (Figure 1.), additional information on the trapping effort (density of trap-nights),
426 human population density and land use type have been incorporated to produce a value of
427 relative effort that will assist researchers in identifying specific locations where predictions
428 based on these underlying data sources may suffer from increased effects of trapping effort
429 bias. This approach improves the ease of identifying under sampled locations, for example,
430 Figure 1. may suggest that South East Senegal, Southern Mali and Southern Niger are well
431 sampled based on locations of trapping sites. When the number of trap nights, human
432 population density and land use of these regions are taken into account (Figure 2.) and
433 compared with better sampled locations (i.e. Western Senegal, Eastern Sierra Leone) these
434 areas are found to be relatively under sampled and would benefit from further sampling effort.
435 This contrasts to North West Nigeria where no trapping has occurred (Figure 1.), the modelling

436 approach has perhaps highlighted this beyond a simple map of currently trapped locations as
437 an immediate priority for sampling of rodents and their pathogens given high human
438 population densities and a human dominated landscape.

439 Much of West Africaremains relatively under sampled, particularly Burkina Faso, Côte d'Ivoire,
440 Ghana and Nigeria, despite these countries facing many of the same challenges. For example,
441 annual outbreaks of Lassa fever are reported in Nigeria and there are potentially 60,000
442 unrecognised cases of Lassa fever every year in Côte d'Ivoire and Ghana [48]. Our estimates of
443 the proportion of a rodent species range that have been sampled along with pathogen testing
444 within their sampled range are sensitive to our choice of raster cell size, smaller area cells will
445 reduce the reported coverage while larger cells will have the opposite effect. Despite this, the
446 observed patterns are unlikely to importantly change, with the finding of sparse sampling of
447 both rodents and their pathogens remaining present across cell scales. Rodent sampling should
448 be targeted towards currently under sampled regions to reduce the potential impact of this bias
449 and improve our understanding of both the distribution of rodent hosts and the prevalence of
450 pathogens within their populations. This will allow for better estimation of risk from endemic
451 and novel zoonoses.

452 Rodent trapping studies provide geographic and temporally contextualised data on both
453 species detection and non-detection which are not available from curated datasets. Non-
454 detection data can improve models of species distributions, unfortunately, high levels of
455 missing data on trapping effort will continue to confound the allocations of non-detections as
456 true absences [49]. Models of host species occurrence and abundance, improved by
457 incorporating species absence, are important to assess the effect of land use and climate change

458 on endemic zoonosis spillover to human populations and direct limited public health resources
459 towards regions at greatest risk [50,51].

460 Currently available consolidated datasets on host-pathogen associations (e.g. CLOVER, EID2
461 and GMPD2) do not include spatial or temporal components [52]. The current synthesis of
462 rodent trapping studies has highlighted that pathogens have been sparsely sampled within a
463 host's range. Current zoonosis risk models dependent on these sources of data are therefore
464 not able to incorporate spatial heterogeneity in pathogen prevalence across the host range.
465 Additional uncertainty in current models of zoonotic disease risk arises from host-pathogen
466 associations that have not been reported in these consolidated datasets. For example,
467 *Hylomyscus pamfi* infected with *Lassa mammarenavirus* and *R. rattus* infected with *Coxiella*
468 *burnetii*, will not be included when solely based on consolidated host-pathogen datasets.
469 Further, detection of zoonotic pathogens in multiple co-occurring host species supports the
470 adoption of multi-species to better understand the potential range of endemic zoonoses [53].

471 Few studies stratified detection and non-detection of hosts or pathogen prevalence by time,
472 therefore limiting inference of changes in host and pathogen dynamics. This limitation prevents
473 calculation of incidence of infection and the abundance of infectious rodents which potentially
474 varies by both time and space [54]. Understanding of temporal changes in viral burden and
475 shedding for endemic zoonoses is required to accurately predict current and future risk of
476 pathogen spillover.

477 Finally, due to data sparsity, we were unable to account for temporal change over the six
478 decades of rodent trapping studies. Land use change and population density have changed
479 dramatically over this period in West Africa [55]. We attempted to mitigate against this by
480 using the median year of trapping to understand the spatial and land use biases in trapping

481 activity. It is possible that land use and population density at trapping sites varied importantly
482 between when rodent trapping was conducted and the conditions in 2005. Despite this
483 limitation, the finding that trapping is biased towards high density, human dominated
484 landscapes is unlikely to substantially change.

485 We have shown that synthesis of rodent trapping studies to supplement curated rodent
486 distributions can counteract some of the inherent biases in these data and that they can add
487 further contextual data to host-pathogen association data. Together this supports their
488 inclusion in efforts to model endemic zoonotic risk and novel pathogen emergence.
489 Contribution of rodent trapping studies as data sources can be improved by adopting reporting
490 standards and practices consistent with Open Science, namely sharing of disaggregated
491 datasets alongside publication [56].

492 Future rodent trapping studies should be targeted towards regions that are currently under-
493 studied. Further information on rodent presence and abundance across West Africa will aid the
494 modelling of changing endemic zoonosis risk and the potential for novel pathogen emergence.
495 Sharing of disaggregated data alongside research publications should be promoted with
496 adoption of data standards to support ongoing data synthesis. Specifically, inclusion of exact
497 locations of trapping sites, trapping effort and the dates at which trapping occurred would
498 support more detailed inference of the spatio-temporal dynamics of host populations and the
499 risk of endemic zoonosis spillover events. Despite these challenges we propose that rodent
500 trapping studies can provide an important source of data to supplement curated datasets on
501 rodent distributions to quantify the risk of endemic zoonosis spillover events and the hazard of
502 novel pathogen emergence.

503 ## 7. Author contributions

504 DS – Conceptualisation, Data Curation, Formal Analysis, Writing

505 LAA - Data Curation, Validation, Writing – Review and Editing

506 KEJ – Conceptualisation, Supervision, Writing – Review and Editing

507 DWJ & RK – Supervision, Funding Acquisition, Writing – Review and Editing

508 ## 8. Data availability

509 All data and code to reproduce this analysis is available in an archived Zenodo repository [32].

510 ## 9. References

511 1. Bernstein AS, Ando AW, Loch-Temzelides T, Vale MM, Li BV, Li H, et al. The costs and
512 benefits of primary prevention of zoonotic pandemics. Science Advances. 2022;8: eabl4183.
513 doi:10.1126/sciadv.abl4183

514 2. Allen T, Murray KA, Zambrana-Torrelio C, Morse SS, Rondinini C, Di Marco M, et al.
515 Global hotspots and correlates of emerging zoonotic diseases. Nature Communications. 2017;8:
516 1124. doi:10.1038/s41467-017-00923-8

517 3. Hassell JM, Begon M, Ward MJ, Fèvre EM. Urbanization and disease emergence:
518 Dynamics at the wildlife–livestock–human interface. Trends in Ecology & Evolution. 2017;32:
519 55–67. doi:10.1016/j.tree.2016.09.012

520 4. García-Peña GE, Rubio AV, Mendoza H, Fernández M, Milholland MT, Aguirre AA, et al.
521 Land-use change and rodent-borne diseases: Hazards on the shared socioeconomic pathways.
522 Philosophical Transactions of the Royal Society B: Biological Sciences. 2021;376: 20200362.
523 doi:10.1098/rstb.2020.0362

524 5. McMahon BJ, Morand S, Gray JS. Ecosystem change and zoonoses in the anthropocene.
525 Zoonoses and Public Health. 2018;65: 755–765. doi:10.1111/zph.12489

526 6. Maudlin I, Eisler MC, Welburn SC. Neglected and endemic zoonoses. Philos Trans R Soc
527 Lond B Biol Sci. 2009;364: 2777–2787. doi:10.1098/rstb.2009.0067

528 7. Halliday JEB, Allan KJ, Ekwem D, Cleaveland S, Kazwala RR, Crump JA. Endemic zoonoses
529 in the tropics: A public health problem hiding in plain sight. Vet Rec. 2015;176: 220–225.
530 doi:10.1136/vr.h798

531 8. Molyneux D, Hallaj Z, Keusch GT, McManus DP, Ngowi H, Cleaveland S, et al. Zoonoses
532 and marginalised infectious diseases of poverty: Where do we stand? Parasites & Vectors.
533 2011;4: 106. doi:10.1186/1756-3305-4-106

534 9. Judson SD, Rabinowitz PM. Zoonoses and global epidemics. Current Opinion in Infectious
535 Diseases. 2021;34: 385–392. doi:10.1097/QCO.0000000000000749

536 10. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases.
537 PNAS. 2015;112: 7039–7044. doi:10.1073/pnas.1501598112

538 11. Gibb R, Albery GF, Mollentze N, Eskew EA, Brierley L, Ryan SJ, et al. Mammal virus
539 diversity estimates are unstable due to accelerating discovery effort. Biology Letters. 2021;18:
540 20210427. doi:10.1098/rsbl.2021.0427

541 12. Fischer C, Gayer C, Kurucz K, Riesch F, Tscharntke T, Batáry P. Ecosystem services and
542 disservices provided by small rodents in arable fields: Effects of local and landscape
543 management. Journal of Applied Ecology. 2018;55: 548–558. doi:10.1111/1365-2664.13016

544 13. Dobson FS, Oli MK. Fast and slow life histories of rodents. Rodent societies: an ecological
545 and evolutionary perspective. 2007; 99–105. Available:
546 https://www.jstor.org/stable/42902037

547 14. Albery GF, Becker DJ. Fast-lived hosts and zoonotic risk. Trends in Parasitology.
548 2021;37: 117–129. doi:10.1016/j.pt.2020.10.012

549 15. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic
550 host diversity increases in human-dominated ecosystems. Nature. 2020;584: 398–402.
551 doi:10.1038/s41586-020-2562-8

552 16. Han BA, Kramer AM, Drake JM. Global patterns of zoonotic disease in mammals. Trends
553 in Parasitology. 2016;32: 565–577. doi:10.1016/j.pt.2016.04.007

554 17. Gibb R, Franklinos LHV, Redding DW, Jones KE. Ecosystem perspectives are needed to
555 manage zoonotic risks in a changing climate. BMJ. 2020;371: m3389. doi:10.1136/bmj.m3389

556 18. Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, et al. Distorted
557 views of biodiversity: Spatial and temporal bias in species occurrence data. PLOS Biology.
558 2010;8: e1000385. doi:10.1371/journal.pbio.1000385

559 19. Bowler DE, Callaghan CT, Bhandari N, Henle K, Benjamin Barth M, Koppitz C, et al.
560 Temporal trends in the spatial bias of species occurrence records. Ecography. 2022;n/a:
561 e06219. doi:10.1111/ecog.06219

562 20. Beck J, Böller M, Erhardt A, Schwanghart W. Spatial bias in the GBIF database and its
563 effect on modeling species' geographic distributions. Ecological Informatics. 2014;19: 10–15.
564 doi:10.1016/j.ecoinf.2013.11.002

565 21. Plowright RK, Becker DJ, McCallum H, Manlove KR. Sampling to elucidate the dynamics
566 of infections in reservoir hosts. Philos Trans R Soc Lond B Biol Sci. 2019;374: 20180336.
567 doi:10.1098/rstb.2018.0336

568 22. Carlson CJ, Farrell MJ, Grange Z, Han BA, Mollentze N, Phelan AL, et al. The future of
569 zoonotic risk prediction. Philosophical Transactions of the Royal Society B: Biological Sciences.
570 2021;376: 20200358. doi:10.1098/rstb.2020.0358

571 23. Wille M, Geoghegan JL, Holmes EC. How accurately can we assess zoonotic risk? Dobson
572 AP, editor. PLoS Biol. 2021;19: e3001135. doi:10.1371/journal.pbio.3001135

573 24. Redding DW, Lucas TCD, Blackburn TM, Jones KE. Evaluating bayesian spatial methods
574 for modelling species distributions with clumped and restricted occurrence data. PLOS ONE.
575 2017;12: e0187602. doi:10.1371/journal.pone.0187602

576 25. Grace D, Mutua F, Ochungo P, Kruska R, Jones K, Brierley L, et al. Mapping of poverty and
577 likely zoonoses hotspots. 2012.

578 26. Meerburg B, Singleton G, Kijlstra A. Rodent-borne diseases and their risks for public
579 health. Critical reviews in microbiology. 2009;35: 221–70. doi:10.1080/10408410902989837

580 27. Galeh TM, Sarvi S, Montazeri M, Moosazadeh M, Nakhaei M, Shariatzadeh SA, et al. Global
581 status of toxoplasma gondii seroprevalence in rodents: A systematic review and meta-analysis.

582 Frontiers in Veterinary Science. 2020;7. Available:
583 https://www.frontiersin.org/article/10.3389/fvets.2020.00461

584 28. Bovendorp RS, MCCleery RA, Galetti M. Optimising sampling methods for small mammal
585 communities in neotropical rainforests. Mammal Review. 2017;47: 148–158.
586 doi:10.1111/mam.12088

587 29. Fichet-Calvet E, Lecompte E, Veyrunes F, Barrière P, Nicolas V, Koulémou K. Diversity
588 and dynamics in a community of small mammals in coastal guinea, west africa. Belgian Journal
589 of Zoology. 2009;139: 93–102. Available: http://www.vliz.be/nl/open-marien-
590 archief?module=ref&refid=204614

591 30. Catalano S, Leger E, Fall CB, Borlase A, Diop SD, Berger D, et al. Multihost transmission of
592 schistosoma mansoni in senegal, 2015-2018. Emerging Infectious Diseases. 2020;26: 1234–
593 1242. doi:10.3201/eid2606.200107

594 31. USAID. One health surveillance. PREDICT project. 2021 [cited 27 Sep 2021]. Available:
595 https://p2.predict.global/surveillance

596 32. Simons D. DidDrog11/scoping_review: Repository to accompany submission. Zenodo;
597 2022. doi:10.5281/zenodo.7022637

598 33. GBIF: The Global Biodiversity Information Facility. The global biodiversity information
599 facility. 2021 [cited 25 Apr 2021]. Available: https://www.gbif.org/

600 34. National Geospatial-Intelligence Agency. NGA: GNS home. [cited 22 Mar 2021]. Available:
601 https://geonames.nga.mil/gns/html/

35. CDC. Emerging and Zoonotic Infectious Disease Laws. 2022 [cited 31 Oct 2022]. Available: https://www.cdc.gov/phlp/publications/topic/zoonotic.html

36. Pebesma E. Simple features for r: Standardized support for spatial vector data. The R Journal. 2018;10: 439–446. Available: https://journal.r-project.org/archive/2018/RJ-2018-009/index.html

37. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: https://www.R-project.org/

38. Database of Global Administrative Areas. GADM. 2022 [cited 25 Apr 2021]. Available: https://gadm.org/index.html

39. Socioeconomic Data and Applications Center. Gridded population of the world (GPW), v4 SEDAC. 2021 [cited 4 Feb 2021]. Available: https://sedac.ciesin.columbia.edu/data/collection/gpw-v4

40. European Space Agency Climate Change Initiative. Land cover classification gridded maps from 1992 to present derived from satellite observations. 2022 [cited 3 Mar 2022]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=overview

41. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: An introduction with mgcv. PeerJ. 2019;7: e6876. doi:10.7717/peerj.6876

42. Kendal W. Spatial aggregation of the Colorado potato beetle described by an exponential dispersion model. Ecological Modelling. 2002;151: 261-269. Doi: 10.1016/S0304-3800(01)00494-X

623 43. Wood SN. Generalized additive models: An introduction with r. 2nd ed. Chapman;
624 Hall/CRC; 2017.

625 44. IUCN. The IUCN red list of threatened species. 2021. Available: www.iucnredlist.org

626 45. GBIF: The Global Biodiversity Information Facility. Occurrence download. The Global
627 Biodiversity Information Facility; 2021. doi:10.15468/DL.S52MEH

628 46. Gibb R, Carlson CJ, Farrell MJ. Viralemergence/clover: Preprint + zenodo. Zenodo; 2021.
629 doi:10.5281/zenodo.4435128

630 47. Harvey E, Holmes EC. Diversity and evolution of the animal virome. Nat Rev Microbiol.
631 2022;20: 321–334. doi:10.1038/s41579-021-00665-x

632 48. Basinski AJ, Fichet-Calvet E, Sjodin AR, Varrelman TJ, Remien CH, Layman NC, et al.
633 Bridging the gap: Using reservoir ecology and human serosurveys to estimate lassa virus
634 spillover in west africa. Wesolowski A, editor. PLoS Comput Biol. 2021;17: e1008811.
635 doi:10.1371/journal.pcbi.1008811

636 49. Václavík T, Meentemeyer RK. Invasive species distribution modeling (iSDM): Are
637 absence data and dispersal constraints needed to predict actual distributions? Ecological
638 Modelling. 2009;220: 3248–3258. doi:10.1016/j.ecolmodel.2009.08.013

639 50. Zeimes CB, Olsson GE, Ahlm C, Vanwambeke SO. Modelling zoonotic diseases in humans:
640 Comparison of methods for hantavirus in sweden. Int J Health Geogr. 2012;11: 39.
641 doi:10.1186/1476-072X-11-39

642 51. Judson SD, LeBreton M, Fuller T, Hoffman RM, Njabo K, Brewer TF, et al. Translating
643 predictions of zoonotic viruses for policymakers. EcoHealth. 2018;15: 52–62.
644 doi:10.1007/s10393-017-1304-3

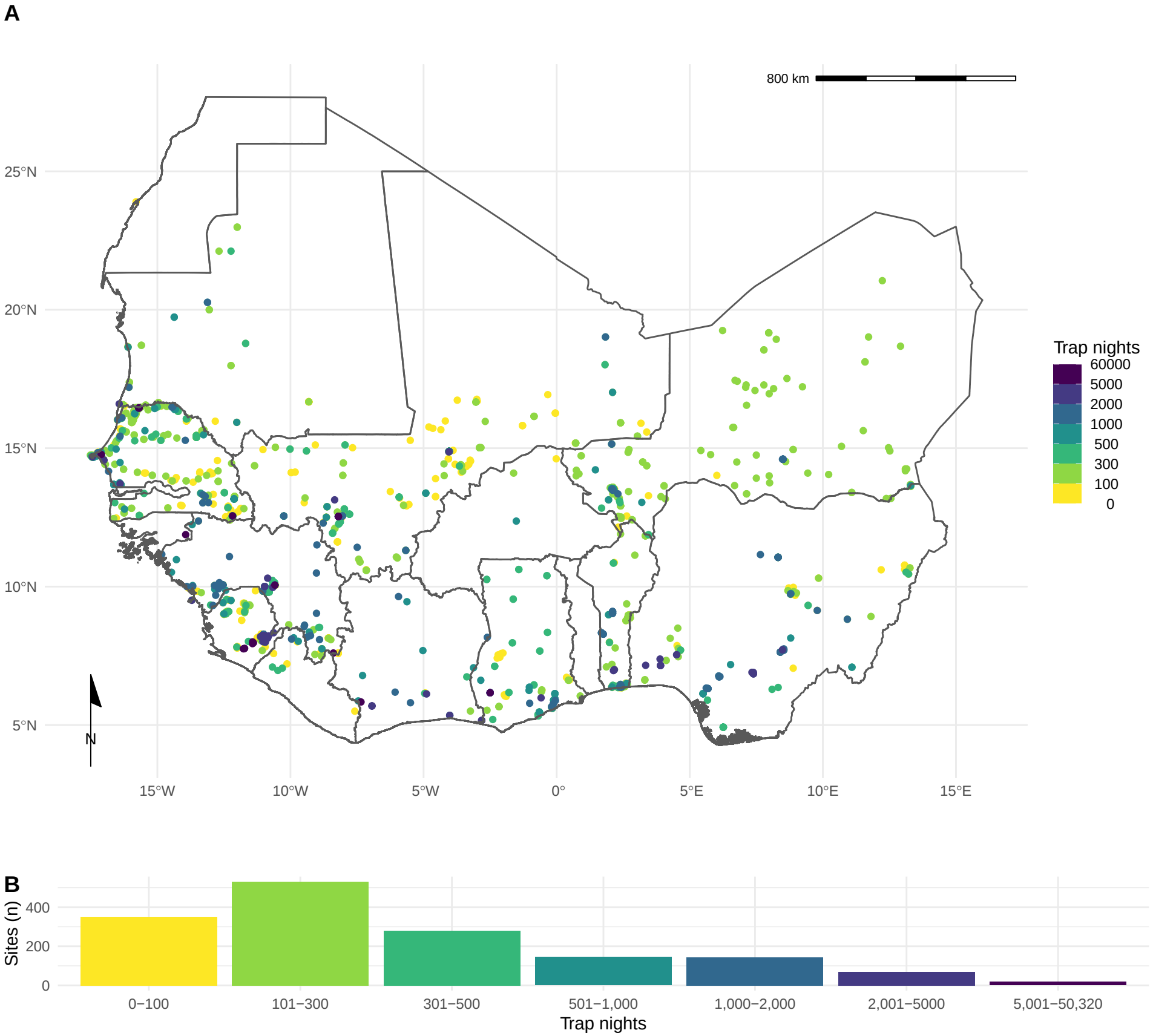
645 52. Gibb R, Albery GF, Becker DJ, Brierley L, Connor R, Dallas TA, et al. Data proliferation,
646 reconciliation, and synthesis in viral ecology. BioScience. 2021;71: 1148–1156.
647 doi:10.1093/biosci/biab080

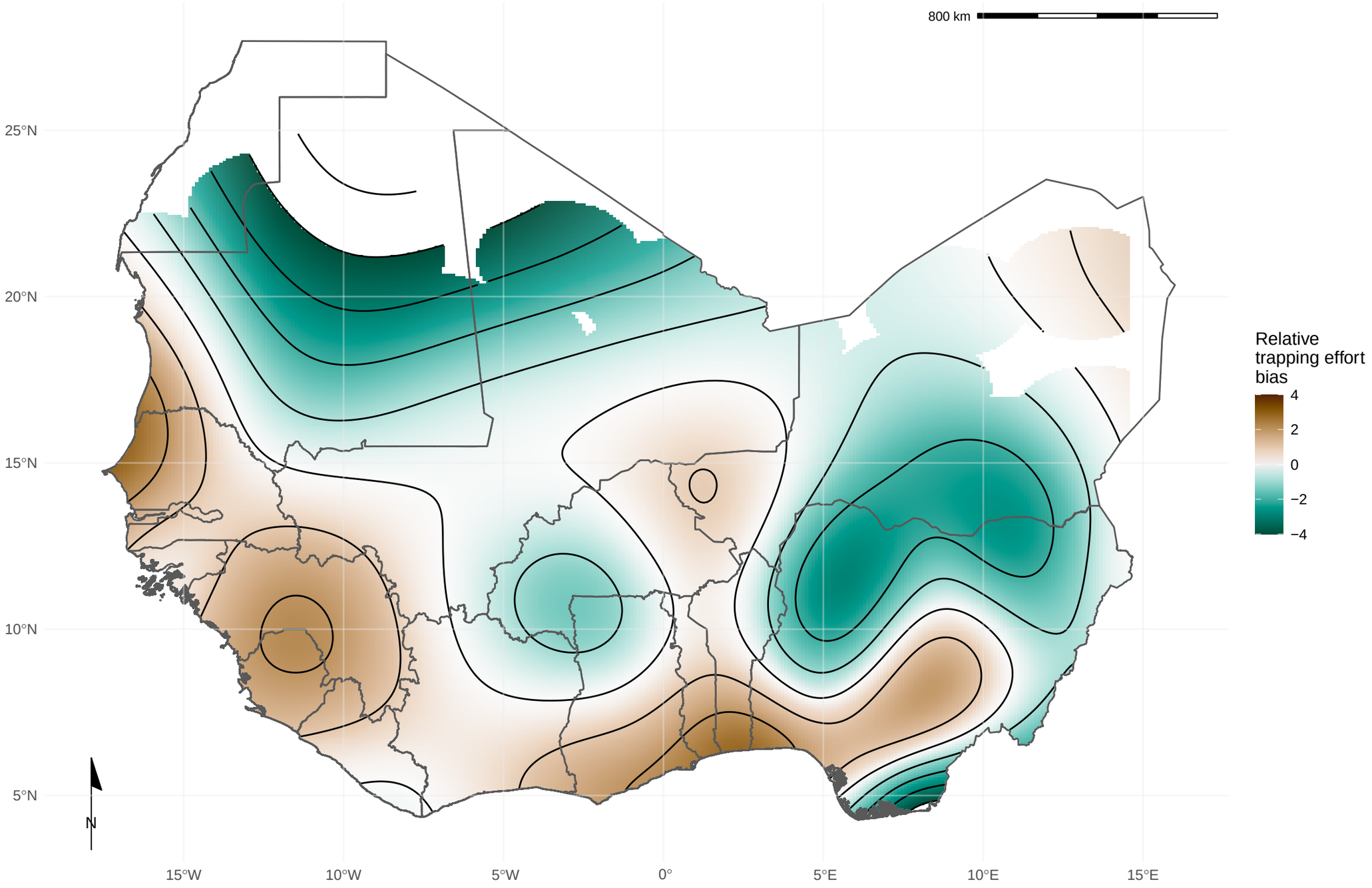
648 53. Wilkinson DP, Golding N, Guillera-Arroita G, Tingley R, McCarthy MA. A comparison of
649 joint species distribution models for presence–absence data. Methods in Ecology and Evolution.
650 2019;10: 198–211. doi:10.1111/2041-210X.13106

651 54. Fichet-Calvet E, Ölschläger S, Strecker T, Koivogui L, Becker-Ziaja B, Camara AB, et al.
652 Spatial and temporal evolution of lassa virus in the natural host population in upper guinea. Sci
653 Rep. 2016;6: 21977. doi:10.1038/srep21977

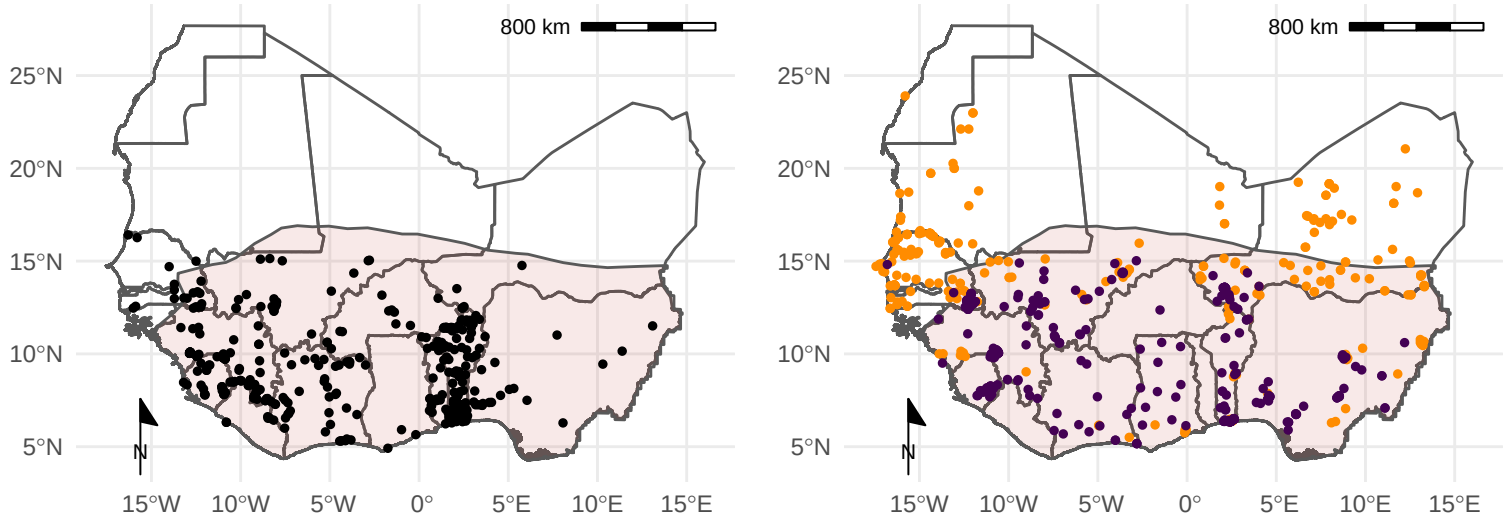
654 55. Herrmann SM, Brandt M, Rasmussen K, Fensholt R. Accelerating land cover change in
655 west africa over four decades as population pressure increased. Commun Earth Environ.
656 2020;1: 1–10. doi:10.1038/s43247-020-00053-y

657 56. Foster ED, Deardorff A. Open science framework (OSF). J Med Libr Assoc. 2017;105:
658 203–206. doi:10.5195/jmla.2017.88

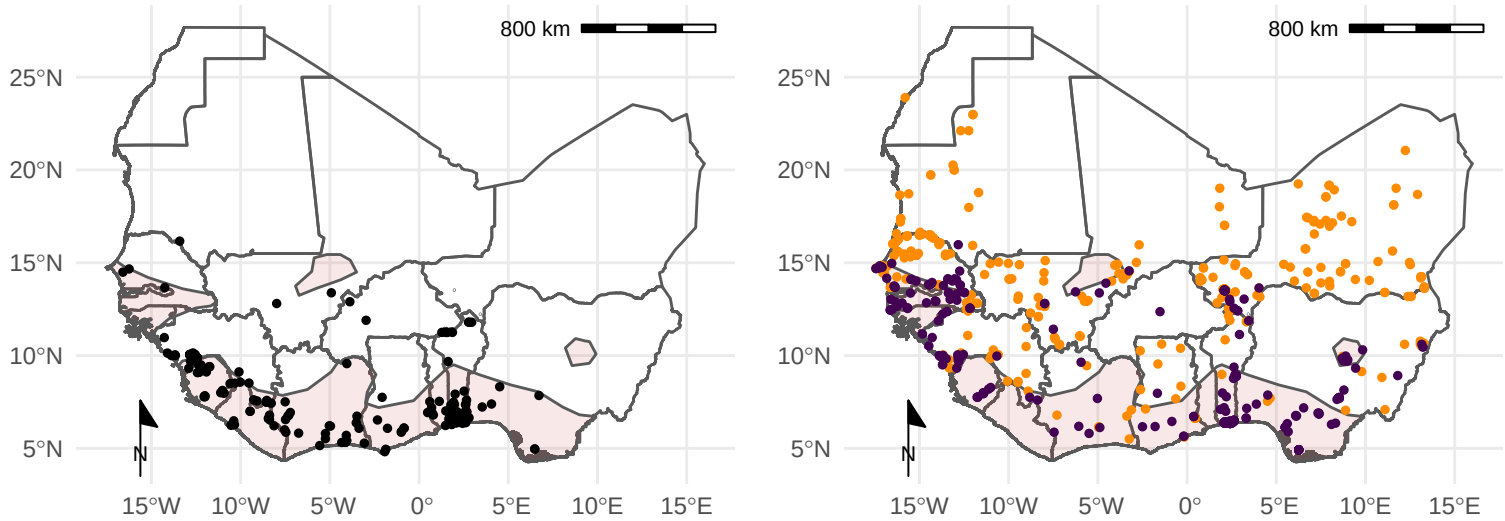

A
800 km
25°N
20°N
15°N
10°N
5°N
N
15°W
10°W
5°W
0°
5°E
10°E
15°E
Trap nights
60000
5000
2000
1000
500
300
100
0
B
Sites (n)
400
200
0
0–100
101–300
301–500
501–1,000
1,000–2,000
2,001–5000
5,001–50,320
Trap nights


800 km
25°N
20°N
15°N
10°N
5°N
15°W
10°W
5°W
0°
5°E
10°E
15°E
N
Relative trapping effort bias
4
2
0
−2
−4

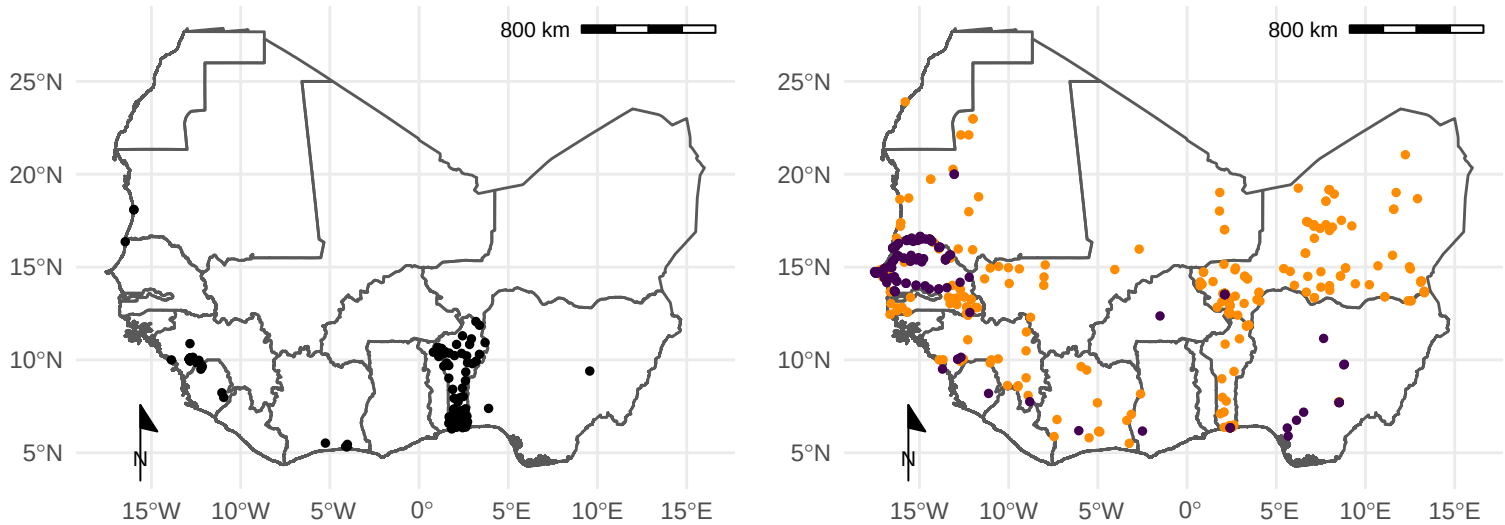
Mastomys natalensis



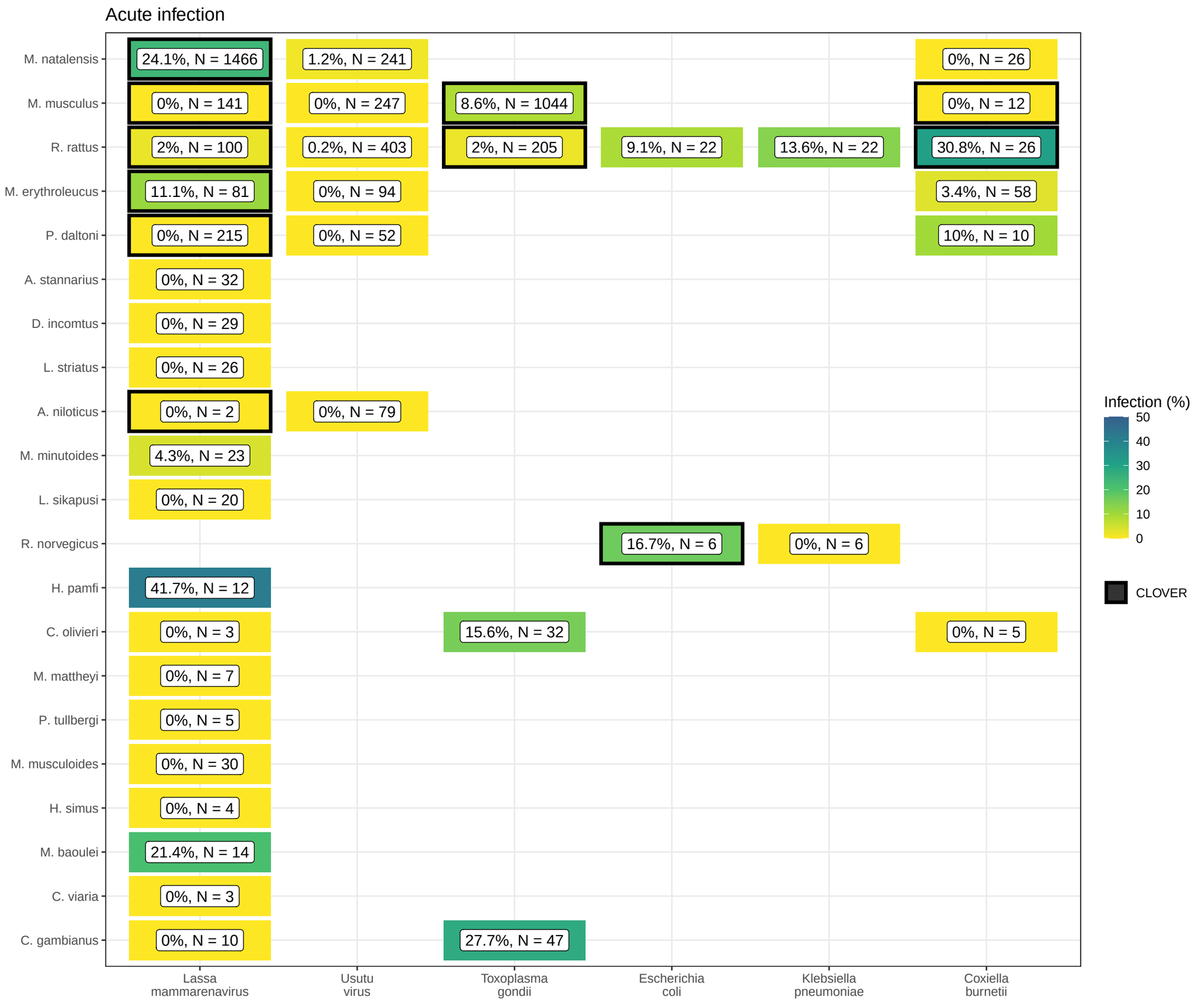
Rattus rattus

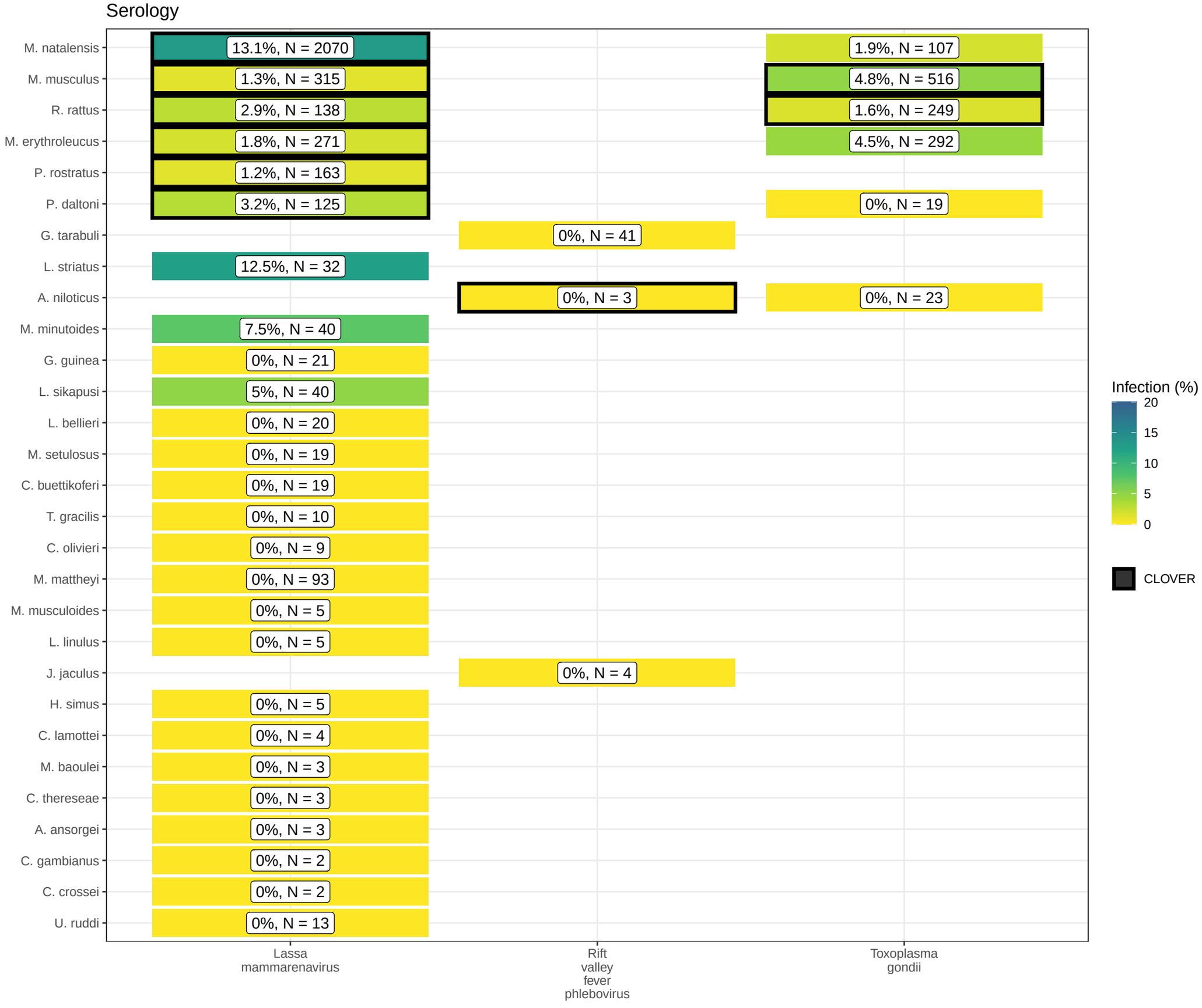


Mus musculus



Detection/Non−detection ● Detection ● Non−detection


Acute infection
M. natalensis
M. musculus
R. rattus
M. erythroleucus
P. daltoni
A. stannarius
D. incomtus
L. striatus
A. niloticus
M. minutoides
L. sikapusi
R. norvegicus
H. pamfi
C. olivieri
M. mattheyi
P. tullbergi
M. musculoides
H. simus
M. baoulei
C. viaria
C. gambianus
24.1%, N = 1466
1.2%, N = 241
0%, N = 26
0%, N = 141
0%, N = 247
8.6%, N = 1044
0%, N = 12
2%, N = 100
0.2%, N = 403
2%, N = 205
9.1%, N = 22
13.6%, N = 22
30.8%, N = 26
11.1%, N = 81
0%, N = 94
3.4%, N = 58
0%, N = 215
0%, N = 52
10%, N = 10
0%, N = 32
0%, N = 29
0%, N = 26
0%, N = 2
0%, N = 79
4.3%, N = 23
0%, N = 20
16.7%, N = 6
0%, N = 6
41.7%, N = 12
0%, N = 3
15.6%, N = 32
0%, N = 5
0%, N = 7
0%, N = 5
0%, N = 30
0%, N = 4
21.4%, N = 14
0%, N = 3
0%, N = 10
27.7%, N = 47
Lassa mammarenavirus
Usutu virus
Toxoplasma gondii
Escherichia coli
Klebsiella pneumoniae
Coxiella burnetii
Infection (%)
50
40
30
20
10
0
CLOVER


Serology
M. natalensis
M. musculus
R. rattus
M. erythroleucus
P. rostratus
P. daltoni
G. tarabuli
L. striatus
A. niloticus
M. minutoides
G. guinea
L. sikapusi
L. bellieri
M. setulosus
C. buettikoferi
T. gracilis
C. olivieri
M. mattheyi
M. musculoides
L. linulus
J. jaculus
H. simus
C. lamottei
M. baoulei
C. thereseae
A. ansorgei
C. gambianus
C. crossei
U. ruddi
13.1%, N = 2070
1.3%, N = 315
2.9%, N = 138
1.8%, N = 271
1.2%, N = 163
3.2%, N = 125
12.5%, N = 32
7.5%, N = 40
0%, N = 21
5%, N = 40
0%, N = 20
0%, N = 19
0%, N = 19
0%, N = 10
0%, N = 9
0%, N = 93
0%, N = 5
0%, N = 5
0%, N = 5
0%, N = 4
0%, N = 3
0%, N = 3
0%, N = 3
0%, N = 2
0%, N = 2
0%, N = 13
0%, N = 41
0%, N = 3
0%, N = 4
1.9%, N = 107
4.8%, N = 516
1.6%, N = 249
4.5%, N = 292
0%, N = 19
0%, N = 23
Lassa mammarenavirus
Rift valley fever phlebovirus
Toxoplasma gondii
Infection (%)
20
15
10
5
0
CLOVER