Open Peer Review on Qeios

A Dataset of Small-Mammal Detections in West Africa and their Associated Micro-Organisms

David Simons[1], Lauren A. Attfield[2], Kate E. Jones[2], Deborah Watson-Jones[3], Richard Kock[1]

1 Royal Veterinary College, RVC

2 University College London, University of London

3 Imperial College London

Funding: D.S. is supported by a PhD award from the UK Biotechnology and Biological Sciences Research Council [BB/M009513/1]. L.A.A. was funded by a PhD award from the QMEE CDT, funded by NERC grant number [NE/P012345/1]. K.E.J is supported by the Ecosystem Services for Poverty Alleviation Programme, Dynamic Drivers of Disease in Africa Consortium, NERC grant number [NE-J001570-1]. D.W-J. receives support from the PREVAC-UP, EDCTP2 programme supported by the European Union [RIA2017S-2014]. D.S. and R.K. are members of the Pan-African Network on Emerging and Re-emerging Infections (PANDORA-ID-NET) funded by the European and Developing Countries Clinical Trials Partnership the EU Horizon 2020 Framework Programme for Research and Innovation [RIA2016E-1609].

Potential competing interests: No potential competing interests to declare.

Abstract

Rodents are a diverse, globally distributed and ecologically important order of mammals that are known and predicted hosts of zoonotic pathogens. The sampling of rodents and their pathogens are taxonomically and spatially biased which limits inference of the hazard of spillover of zoonotic pathogens into human populations from rodent vectors. Data on the distribution and occurrence of rodent hosts are typically derived from consolidated databases (e.g., IUCN, GBIF) which suffer from these biases. Here, we synthesise data from 127 rodent trapping studies, identified through a comprehensive search of the published literature from 1964-2022 conducted in 14 West African countries to provide an additional source of information that can supplement consolidated databases to characterise the range and occurrence of rodent species. We combine these occurrence data with results from reported pathogen screening to produce a dataset containing detection/non-detection data for 65,628 individual small mammals identified to species level from at least 1,611 trap sites in addition to 32 microorganisms identified to species and genus level that are known or potential pathogens. The produced dataset is formatted to Darwin Core Standard with associated metadata. This dataset is expected to mitigate some of the spatial and taxonomic biases in current databases to improve analyses of rodent-borne zoonotic pathogen spillover hazard across West Africa.

David Simons[1,2,3,*], **Lauren A. Attfield**[2,4], **Kate E. Jones**[2], **Deborah Watson-Jones**[3,5], **Richard Kock**[1]

[1] *The Royal Veterinary College, London, United Kingdom*

[2] *University College London, London, United Kingdom*

[3] *The London School of Hygiene and Tropical Medicine, London, United Kingdom*

[4] *Imperial College London, London, United Kingdom*

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

[*] dsimons19@rvc.ac.uk

Keywords: Rodents; Zoonosis; Vector-Borne Diseases; West Africa.

Data Description

Context

Rodents are a diverse, globally distributed and ecologically important order of mammals. Along with bats, these two orders are proposed to contain the greatest number of host species of known and predicted of zoonotic pathogens. Of 2,220 extant rodent species, 244 (10.7%) are described as hosts of 85 zoonoses [1]. Importantly, rodent hosts of zoonoses are typically synanthropic and thrive in anthropogenically disturbed habitats leading to spatially heterogeneous risk of pathogen transmission [2]. Rodent borne endemic zoonoses are a significant public health threat across much of West Africa and include bacterial, viral and protozoan pathogens. Important endemic rodent-borne zoonoses in West Africa include Lassa fever (caused by *Lassa mammarenavirus*), Leptospirosis (caused by *Leptosira sp.*) and Toxoplasmosis (caused by *Toxoplasma gondii*) [3][4][5]. It is likely that other potential zoonoses are circulating in rodent populations in this region that have not been described or identified as causing human infections [6]. Currently described zoonotic pathogens are generally associated with multiple rodent species, although a single species may be the primary reservoir species, for example, *Mastomys natalensis* is considered the primary reservoir of Lassa fever although *Lassa mammarenavirus* infection has been associated with ten other rodent species[7]. For this reason understanding the structure of rodent communities, their spatial distribution and associated pathogens are vital to understand the hazard of endemic zoonotic disease spillover and novel zoonotic pathogen emergence [8].

Studies assessing the risk of outbreaks of endemic zoonoses and novel pathogen emergence often use consolidated datasets such as the Global Biodiversity Information Facility (GBIF) and International Union for Conservation of Nature (IUCN) Redlist to model host occurrence [9][10][11]. Despite the importance of understanding the true distribution of rodent hosts and their pathogens, curated biodiversity databases such as these are affected by taxonomic and geographical sampling biases [12][13]. These biases can subsequently limit inference from produced species distribution models that are used to quantify the hazard of zoonotic disease spillover into human populations and guide public health interventions [14]. Rodent trapping studies are also taxonomically and spatially biased [2][7]. Despite these biases, we found that combining data from rodent trapping studies conducted in West Africa with data from GBIF and IUCN has the potential to increase the sampled area for commonly occurring species by up to 160% and mitigate some of the effects of these biases when attempting to model the distribution of rodent vectors of zoonoses [7]. We found that rodent trapping studies were more likely to have been conducted in locations of relatively high human population density and include data on small-mammal species that are synanthropic [7].

The current dataset, a synthesis of 127 rodent trapping studies conducted within 17 African countries (but focussing on

studies conducted in West Africa), can aid the development of models based on rodent reservoir occurrence to estimate the potential for pathogen spillover into human populations by providing additional geographic locations of presence and absence. For example, a recent article developed a model of the risk of Lassa fever spillover based on both *M. natalensis* occurrence and pathogen prevalence in West Africa[11]. This study included 167 locations of *M. natalensis* detections which would be potentially expanded by up to an additional 337 locations of detections and 320 locations of non-detection from the addition of the current dataset, increasing the coverage of observations over the endemic region.

Methods

Search strategy

This dataset contains information on small mammal detections and non-detections obtained from rodent trapping studies conducted in West Africa between 1964 and 2022. Data have been extracted from published articles, biodiversity surveys and impact assessments. Studies were identified through a search conducted 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 using the following terms as exploded keywords:

1. Rodent OR Rodent trap*
   AND
2. West Africa

We used the UN definition for West Africa which includes the following countries (ISO 3166-1 alpha-2 codes are given in parenthesis): Benin (BJ), Burkina Faso (BF), Cape Verde (CV), Ivory Coast (CI), Gambia (GM), Ghana (GH), Guinea (GN), Guinea-Bissau (GW), Liberia (LR), Mali (ML), Mauritania (MR), Niger (NE), Nigeria (NG), Senegal (SN), Sierra Leone (SL) and Togo (TG).

Similar searches were conducted in additional resources, including the UN Official Documents System, Open Grey, AGRIS FAO and Google Scholar. Searches were run on 2022-05-01.

We included studies for further analysis if they met all of the following inclusion criteria;

1. Reported findings from trapping studies where the target was a small mammal.
2. Described the type of trap used or the length of trapping activity or the location of the trapping activity
3. Included trapping activity from at least one West African country.
4. Recorded the genus or species of trapped individuals.
5. 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:

1. Reported data that were duplicated from a previously included study.
2. No full text available.
3. Not available in English.

One author (DS) 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 (LAA). Supplementary Table 1 contains the year of publication, name of the first author, title of the study, publication and unique identifier of the included study.

Data were extracted from eligible studies using a standardised tool that was piloted on five randomly selected studies. Supplementary Table 2 contains the variable names and descriptors that were abstracted into three sheets. The first sheet, “Study data”, contained information on the included study, the purpose of the study, methodology of rodent sampling and species identification. The second sheet, “Rodent data”, contained information on the number of individuals of each species detected at a trapping location, alongside geographic coordinates of the sampling location and habitat type. Data for this section were expanded by adding non-detections if the rodent species was detected at other sampling sites within the study. Finally, the third sheet, “Pathogen data”, contained information on the testing of the individual rodent species for known and suspected zoonotic pathogens. Unprocessed data is archived in a Zenodo repository within the `data_raw` folder [15].

Data validation and quality control

Species identification was assumed to be accurate in included studies. For studies reporting genus level or multiple possible species names for a single trapped individual data were extracted as presented in the study. Species names were mapped to GBIF taxonomy to resolve changes in taxonomic classification using the `taxize` package (version 0.9.98) in the R statistical programming language (version 4.1.2) [16][17]. Geographic locations of trapping studies were extracted using GPS locations for the most precise location presented. Missing locations were found using the National Geospatial-Intelligence Agency GEOnet Names Server based on placenames and maps presented in the study (National Geospatial-Intelligence Agency, 2022). All locations were converted to decimal degrees in the EPSG:4326 coordinate reference system.

For included studies with available data we extracted information on all microorganisms and known 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, for non-specific assays higher order attribution was used (e.g., to family level). 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 included data released in pre-prints identified from our systematic search, in addition to studies conducted at the same locations over multiple time periods. We reviewed all occurrence data to ascertain that the geographic coordinates,

time period of sampling and number of identified individuals were unique to ensure we did not include duplicated data. Where duplicated data was identified the record with the greatest number of detections, or the most recent reporting was retained. For example, multiple published studies may have included updates of a longitudinal sampling design, in these cases only the most recent data for that location was retained.

Data processing and exploration

R code to process the raw data into the Darwin Core format, for rodent occurrence and associated pathogen detection data along with metadata has been archived as a Zenodo repository [18].

An RShiny web application has been produced to visualise the data contained in this release. The web-based application is available at this webpage. This application allows exploration of the location of sampling sites for both rodents and their pathogens within included studies alongside sampling effort reported by the study.

Reuse potential

This dataset of harmonised rodent species detections obtained from rodent trapping surveys conducted across West Africa will contribute to understanding rodent biodiversity across the region. It is envisaged that this dataset will be of particular interest to researchers investigating the risk of rodent borne zoonotic pathogen outbreaks and emergence in this region and beyond. This data will expand the geographical coverage of occurrence data within GBIF for most of the rodent species detected in the included rodent trapping studies with additional data on non-detections of these species across the region. Where possible, dates of rodent sampling have been included which may be of benefit to researchers investigating how occurrence patterns of rodent species may vary over time, which will be important to understanding changes in the context of climate, land use and population changes.

Statements and Declarations

Data availability

As part of this series of Data Release articles this dataset is available under a CC0 1.0 Universal license.

List of abbreviations

- GBIF – Global Biodiversity Information Facility
- IUCN – International Union for Conservation of Nature

Competing Interests

The authors declare that they have no competing interests.

Funding

D.S. is supported by a PhD award from the UK Biotechnology and Biological Sciences Research Council [BB/M009513/1]. L.A.A. was funded by a PhD award from the QMEE CDT, funded by NERC grant number [NE/P012345/1]. K.E.J is supported by the Ecosystem Services for Poverty Alleviation Programme, Dynamic Drivers of Disease in Africa Consortium, NERC grant number [NE-J001570-1]. D.W-J. receives support from the PREVAC-UP, EDCTP2 programme supported by the European Union [RIA2017S-2014]. D.S. and R.K. are members of the Pan-African Network on Emerging and Re-emerging Infections (PANDORA-ID-NET) funded by the European and Developing Countries Clinical Trials Partnership the EU Horizon 2020 Framework Programme for Research and Innovation [RIA2016E-1609].

Author's Contributions

- DS – Conceptualization (Equal), Data curation (Equal), Formal analysis (Lead), Methodology (Lead), Software (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Equal).
- LAA - Data curation (Equal), Writing – review & editing (Equal), Validation (Lead).
- KEJ – Conceptualization (Equal), Supervision (Equal), Writing – original draft (Supporting), Formal analysis (Supporting), Writing – review & editing (Lead).
- DWJ - Funding acquisition (Equal), Supervision (Equal), Writing – review & editing (Equal).
- RK - Funding acquisition (Equal), Supervision (Equal), Writing – review & editing (Equal)

Acknowledgements

We would like to thank the Health data team at the Global Biodiversity Information Facility for helpful guidance for formatting our data.

Other References

- National Geospatial-Intelligence Agency. NGA: GNS Home [Internet]. 2023 [cited 2021 Mar 22]. Available from: https://geonames.nga.mil/gns/html/

References

1. ^*Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases. PNAS. 2015 Jun 2;112(22):7039–44.*
2. a, b*Ecke F, Han BA, Hörnfeldt B, Khalil H, Magnusson M, Singh NJ, et al. Population fluctuations and synanthropy explain transmission risk in rodent-borne zoonoses. Nat Commun. 2022 Dec 7;13(1):7532.*
3. ^*Allan KJ, Biggs HM, Halliday JEB, Kazwala RR, Maro VP, Cleaveland S, et al. Epidemiology of Leptospirosis in Africa: A Systematic Review of a Neglected Zoonosis and a Paradigm for 'One Health' in Africa. PLoS Negl Trop Dis.*

2015 Sep 14;9(9):e0003899.

4. ^*Brouat C, Diagne CA, Ismail K, Aroussi A, Dalecky A, Ba K, et al. Seroprevalence of Toxoplasma gondii in commensal rodents sampled across Senegal, West Africa. Parasite. 2018;25(bz6, 9437094):32.*
5. ^*Garry RF. Lassa fever — the road ahead. Nat Rev Microbiol. 2023 Feb;21(2):87–96.*
6. ^*Suu-Ire R, Obodai E, Bel-Nono SO, Ampofo WK, Mazet JAK, Goldstein T, et al. Surveillance for potentially zoonotic viruses in rodent and bat populations and behavioral risk in an agricultural settlement in Ghana. One Health Outlook. 2022 Mar 8;4(1):6.*
7. a, b, c, d*Simons D, Attfield LA, Jones KE, Watson-Jones D, Kock R. Rodent trapping studies as an overlooked information source for understanding endemic and novel zoonotic spillover. PLOS Neglected Tropical Diseases. 2023;17(1):e0010772.*
8. ^*Han BA, Kramer AM, Drake JM. Global Patterns of Zoonotic Disease in Mammals. Trends in Parasitology. 2016 Jul;32(7):565–77.*
9. ^*Mylne AQN, Pigott DM, Longbottom J, Shearer F, Duda KA, Messina JP, et al. Mapping the zoonotic niche of Lassa fever in Africa. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2015 Aug 1;109(8):483–92.*
10. ^*Pigott DM, Millear AI, Earl L, Morozoff C, Han BA, Shearer FM, et al. Updates to the zoonotic niche map of Ebola virus disease in Africa. Jit M, editor. eLife. 2016 Jul 14;5:e16412.*
11. a, b*Basinski AJ, Fichet-Calvet E, Sjodin AR, Varrelman TJ, Remien CH, Layman NC, et al. Bridging the gap: Using reservoir ecology and human serosurveys to estimate Lassa virus spillover in West Africa. Wesolowski A, editor. PLoS Comput Biol. 2021 Mar 3;17(3):e1008811.*
12. ^*Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, et al. Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. PLOS Biology. 2010 Jun 1;8(6):e1000385.*
13. ^*Bowler DE, Callaghan CT, Bhandari N, Henle K, Benjamin Barth M, Koppitz C, et al. Temporal trends in the spatial bias of species occurrence records. Ecography. 2022 May 20;n/a (n/a):e06219.*
14. ^*Beck J, Böller M, Erhardt A, Schwanghart W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics. 2014 Jan 1;19:10–5.*
15. ^*Simons D. DidDrog11/scoping_review: To accompany re-submission following reviewer comments - files archived [Internet]. Zenodo; 2022 [cited 2023 Feb 22]. Available from: https://zenodo.org/record/4718374*
16. ^*Chamberlain S, Szoecs E, Foster Z, Arendsee Z, Boettiger C, Ram K, et al. taxize: Taxonomic information from around the web [Internet]. 2020. Available from: https://github.com/ropensci/taxize*
17. ^*R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/*
18. ^*Simons D. DidDrog11/data_for_gbif: Dataset v0.92 [Internet]. Zenodo; 2023 [cited 2023 Mar 7]. Available from: https://zenodo.org/record/7706777*