**Current sampling and sequencing biases of Lassa mammarenavirus limit inference from phylogeography and molecular epidemiology in Lassa Fever endemic regions.**

**Authors**

Liã Bárbara Arruda1b#, Hayley Free2a§, David Simons2§, Rashid Ansumana3, Linzy Elton1, Najmul Haider2, Isobella Honeyborne1, Richard Kock2, Timothy D McHugh1, Francine Ntoumi4, Alimuddin Zumla1,5

**Affiliations**

1 Centre for Clinical Microbiology, Division of Infection and Immunity, University College London, London, UK

2 The Royal Veterinary College, University of London, Hatfield, UK.

3 School of Community Health Sciences, Njala University, Bo, Sierra Leone

4 Fondation Congolaise pour la Recherche Médicale (FCRM), Brazzaville, Republic of Congo; Institute for Tropical Medicine, University of Tübingen, Germany

5 NIHR Biomedical Research Centre, UCLHospitals NHS Foundation Trust, London, UK

a Current affiliation Oxford Brookes University, Oxford, UK

b Current affiliation Wellcome Connecting Science, Hinxton, UK

§ Both authors contributed equality to this work

# Corresponding author

**Abstract**

The viral haemorrhagic infection caused by Lassa virus (LASV) is an important endemic zoonotic disease in West Africa with evidence for increasing outbreak sizes. Phylogeographic and molecular epidemiology methods have been used to project expansion of the Lassa Fever endemic zone in the context of future global change. The Natal multimammate mouse (*Mastomys natalensis*) is the predominant viral reservoir, although few studies have investigated the role of other animal species. To investigate host sequencing biases, all LASV nucleotide sequences and associated metadata (n = 2,298) available on GenBank were retrieved. Most data originated from Nigeria (56%), Guinea (20%) and Sierra Leone (14%). Data from non-human hosts (n = 703) were limited, only 69 sequences encompassed complete genes. Spatial modelling of sequencing effort adjusted for the number of suspected and confirmed human cases highlighted current biases in locations of available sequences. Increased sequencing effort was observed in Southern Guinea and Southern Nigeria with undersampling occurring primarily in Northern Nigeria and Liberia. Phylogenetic analyses showed geographic clustering of LASV lineages, suggested isolated events of human-to-rodent transmission and the emergence of currently circulating strains of LASV from the year 1498 in Nigeria. Overall, the current study highlights significant geographic limitations in LASV surveillance, particularly, in non-human species. Further investigation of the non-human reservoir of this virus, alongside improved surveillance in other endemic countries, are required for further characterisation of the historic emergence and dispersal of LASV. Accurate assessment on viral circulation in non-human hosts is vital to guide public health interventions to respond to expected Lassa fever outbreaks.

**Key-words**

Lassa virus; Phylogeography; Metadata

1. **Introduction**

Lassa fever (LF) is a viral haemorrhagic disease, caused by *Lassa mammarenavirus* (LASV) that causes several thousand deaths in West Africa annually (Asogun et al, 2019). The WHO assigns endemicity to eight West African countries including Benin, Ghana, Guinea, Liberia, Mali, Sierra Leone, Togo and Nigeria (Supplementary figure 1) (World Health Organisation, 2022a).

There is limited epidemiological data on LF and making accurate estimates of its true burden remains challenging. Many individuals infected with LASV do not seek healthcare with up to 80% of infections assumed asymptomatic or presenting as mild illness (McCormick *et al.*, 1987). Estimates based on longitudinal serological surveys in Sierra Leone in the early 1980’s indicated that 100,000 to 300,000 infections of Lassa fever occurred annually in West Africa, with more recent estimates being up to 900,000 infections (McCormick *et al.*, 1987; Basinski *et al.*, 2021). Identification of symptomatic cases is further confounded by overlapping symptoms with other diseases (e.g., malaria) and lack of available diagnostic methods (Asogun *et al.*, 2019; Takah *et al.*, 2019; Nnaji *et al.*, 2021; Ashcroft *et al.*, 2022). Access to diagnostic assays varies spatially, increased availability at centers of excellence in Lassa fever treatment and research such as the Irrua Specialist Teaching Hospital, Nigeria and Kenema General Hospital, Sierra Leone results in a spatial bias of reported cases from these locations. Phylogenetic analysis and molecular dating of sequence clinical and research samples suggest a westward route of dispersal of LASV lineages, from the most recent common ancestor in Nigeria. (Bowen *et al.*, 2000; Andersen *et al.*, 2015; Manning, Forrester and Paessler, 2015; Olayemi *et al.*, 2016, 2020; Whitmer *et al.*, 2018; Okoro *et al.*, 2020). These estimates have been used to estimate the potential for Lassa Fever to extend beyond the current endemic zone (Klitting *et al.*, 2021).

The Natal multimammate mouse (*Mastomys natalensis*) is the primary reservoir of LASV, however, other rodents have been found to be infected; *Mastomys erythroleucus, Hylomyscus pamfi,* *Mus baoulei* and *Rattus rattus* (Wulff, Fabiyi and Monath, 1975; Lecompte *et al.*, 2006; Olayemi *et al.*, 2016; Yadouleton *et al.*, 2019; Forni and Sironi, 2020; Bangura *et al.*, 2021). Humans become infected with LASV upon contact with or inhalation of excretions from the rodent species (Andersen *et al.*, 2015; Oti, 2018). Although human-to-human transmission has been reported – typically associated with nosocomial outbreaks – these are rare events when compared with spillover from rodent hosts (Lo Iacono *et al.*, 2015).

LASV is a bisegmented ssRNA- virus of the family *Arenaviridae* (Günther and Lenz, 2004; Hallam *et al.*, 2018). Based on the genomic analysis of the large (L) and small segments (S) LASV has been classified into seven lineages which demonstrate spatial segregation across the endemic range (Welch *et al.*, 2018). The high nucleotide variability (25-32%) of these lineages introduces complexity into assays to detect LASV infection. Here, we compiled a comprehensive dataset of publicly available full-segment LASV sequences, spanning West Africa and host species, to inform our understanding of the phylogeny of LASV dispersal. We identified substantial variability in the origin of available sequences and completeness of records. We show strong geographic clustering among lineages supporting prior hypotheses of radiation from both Nigeria and a subsequent introduction into Liberia (Klitting *et al.*, 2021). The synthesis of available metadata highlights important gaps in currently available data, including spatial bias in the sequencing of samples and should be used to inform the design of epidemiological programmes going forward. A better understanding of LASV phylogeography would improve and support effective implementation of measures to prevent an expected increase in the size of the endemic region due to projected climate, human population and land-use change.

1. **Methods**

**2.1 Data Collection and Processing**

LASV nucleotide and protein sequences were obtained from the National Centre for Biotechnology Information (NCBI) GenBank (National Center for Biotechnology Information, 2022). The search query run on 24 Sep 2021 was for “Lassa mammarenavirus” in the organism field of the NCBI nucleotide dataset. Data were obtained using the NCBI Entrez API (National Center for Biotechnology Information, 2022) with analysis conducted using the “genbankr” package (Becker and Lawrence, 2021) and the R statistical programme (R Core Team, 2021). Associated citations were manually retrieved to identify missing metadata for sequences including hosts and geographic location of samples. Sequences with large portions (10% missing compared to reference sequences, NC\_004296.1 and NC\_004297.1 for S and L segments respectively) of missing nucleotide data on the L- or S-segment or lacking associated metadata (collection year, host species, country, and geographical region of sampling) were excluded from phylogenetic analysis. Nucleotide sequences were aligned using the ‘map to reference’ tool on Geneious Prime 20201.2. Alignment, visual inspection and manual editing were performed, and entries that contained >100 continuous ambiguous nucleotide calls were excluded.

**2.2 Sequencing bias**

To understand the bias of sequenced samples at a sub-national level the origin of a sequenced sample was geocoded using the Google Geocoding API using the “ggmap” package (Kahle and Wickham, 2013). Sequence locations were associated with level-1 administrative regions and data were separated into human and rodent sources of samples to visualise the heterogeneity of spatial sampling. To measure sampling effort bias, the number of samples obtained within a level-1 administrative region was associated with the centroid of the region. The number of suspected and confirmed clinical cases reported from these regions in the previous 25 years was obtained (Supplementary table 1). The number of cases within a region was divided by the human population count to produce the number of suspected and confirmed cases per 100,000 individuals. The number of sequences was used as the response variable in a spatial Generalised Additive Model, with geographic coordinates and cases per 100,000 individuals used as covariates. This model was constructed using the “mgcv” package (Wood, 2017).

* 1. **Phylogenetic Analysis**

Phylogenetic analysis was undertaken through Bayesian Markov Chain Monte Carlo (MCMC) method using BEAST.v1.10.4 (Suchard *et al.*, 2018). In BEAUTi, the parameters were a substitution model as a generalised time reversible plus gamma site heterogeneity, with codon partition positions 1, 2, 3. A strict clock and a coalescent tree prior with a constant size population was used. Each analysis consisted of 20 million MCMC steps and trees were sampled every 20,000 generations. Sample collection dates from the metadata were used as tip dates to fit to a molecular clock, and country of sample collection was incorporated as a discrete state (Olayemi and Fichet-Calvet, 2020; Olayemi *et al.*, 2020). To assess the log files of the output TRACER.v.1.7.1 was used. Maximum-clade credibility trees were generated through TreeAnnotator v1.8.4 and visualised in FigTree.v1.4.4 (Rambaut *et al.*, 2018).

1. **Results**

**3.1 Compiled Dataset**

The initial dataset comprised 2,298 records (from samples obtained 1969-2019), including nucleotide sequences and associated metadata. Sequences lacking country information (n = 134) and incomplete gene sequences (n = 906) were removed from phylogenetic analyses. Therefore, 680 sequences of complete S segment and 578 sequences of partial L segment (L protein only) were used. Accession numbers of included and excluded sequences are available in Supplementary table 2.

**3.2 Descriptive Analysis**

Year of collection was available for 2,108 records, with the oldest sequence dating from 1969 and latest from 2019 Among these records, most sequences (n = 2,063) have been obtained since 2000. Human-derived LASV sequences comprised most of the available records (67%), other host species include *Mastomys natalensis* (29%) and *Mastomys sp.* (3%), while *Mastomys erythroleucus (n=18)*, *Mus baoulei (n=9)* and *Hylomyscus pamfi (n=10)* represent < 1% each. The species sampled was not documented in 107 records. The majority of sequences were produced from samples collected in Nigeria (56%), followed by Guinea (20%), Sierra Leone (14%), Liberia (4%) and Cote d’Ivoire (3%) with the remainder obtained from, Benin, Ghana, Mali and Togo (Figure 1).

Sequences for human derived samples were clustered in Edo State, Nigeria and Eastern Province, Sierra Leone with 51 samples from the remaining endemic countries. Sequences from rodent samples were most commonly obtained from Faranah, Guinea and Eastern Province, Sierra Leone with 136 samples from the remaining endemic countries.

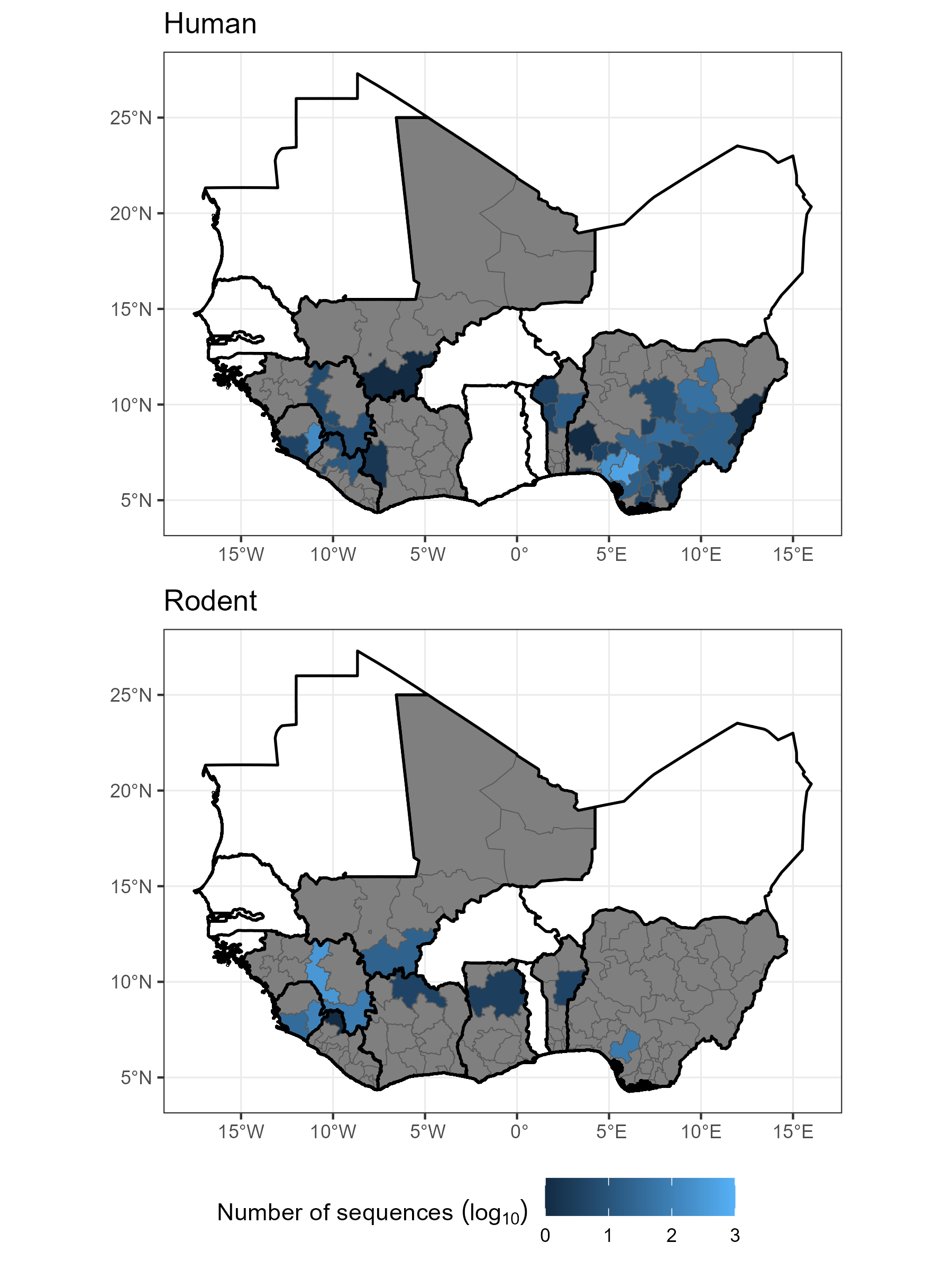


Figure 1 – The number of sequences, shown on a log10scale, retrieved from NCBI GenBank with associated location and host for human samples (top, N = 1070) and rodent samples (bottom, N = 656). Grey regions represent level-1 administrative areas with no sequences within countries that have at least one available sequence. White countries are West African countries with no available *Lassa mammarenavirus* sequences. See Supplementary Figure 1 for country names.

* 1. **Sequencing bias**

Combining both human and rodent-derived samples, relative sequencing effort, adjusted for the number of suspected and confirmed cases per 100,000 individuals, was found to be greatest in Southwest Nigeria, centred over Edo State and the border region of Guinea, Sierra Leone and Liberia centred over the Faranah. Nzérékoré regions of Guinea, Eastern Province of Sierra Leone and Nimba district of Liberia (Figure 2). There was a positive, non-linear association between the number of suspected and confirmed human cases with the number of available rodent and human derived LASV sequences (edf = 2.29, deviance explained = 14%, p < 0.001).

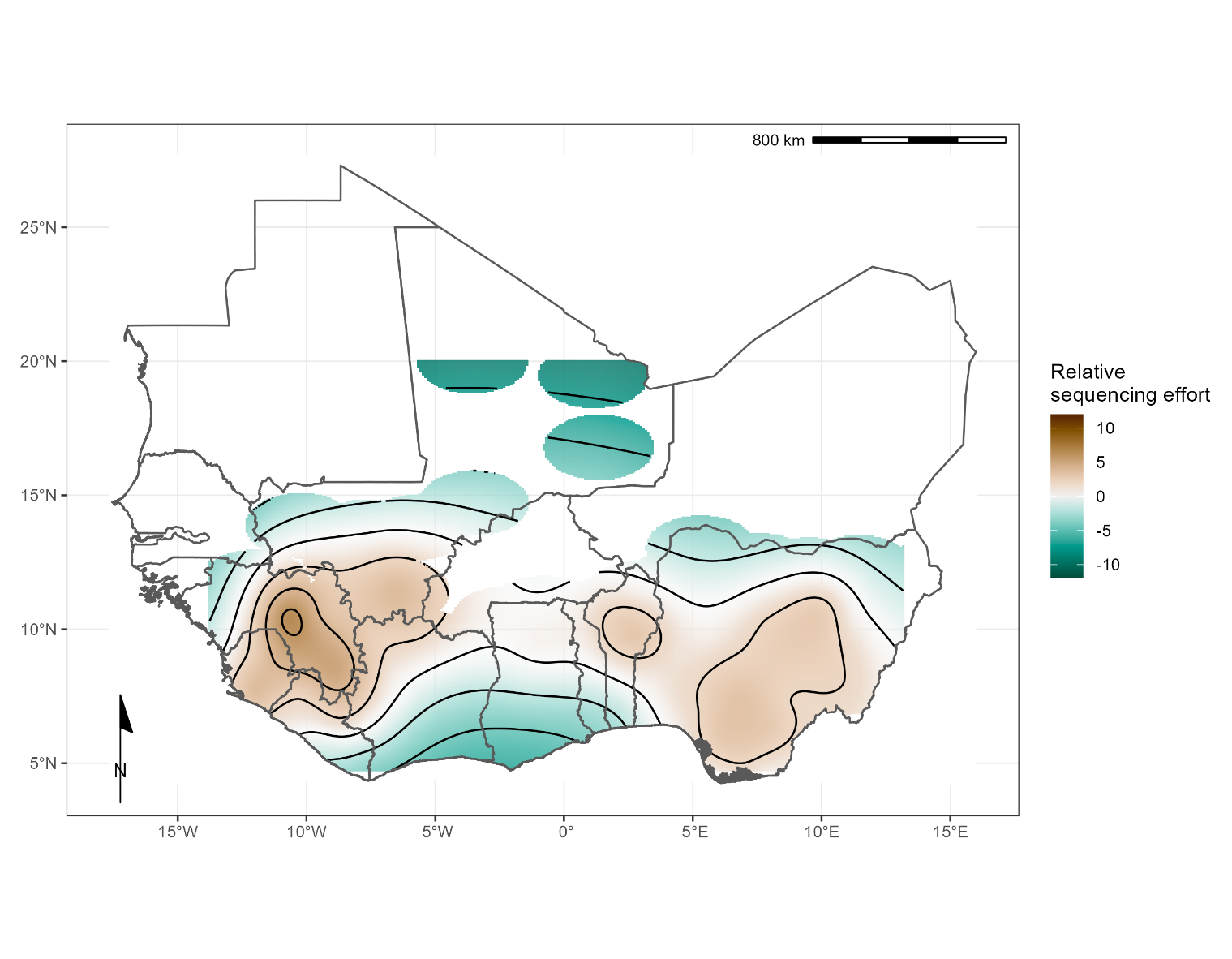


Figure 2 – Modelled relative sequencing effort derived from both human and rodent samples. Greatest sequencing effort coincides with areas where sampling in humans (Edo, Nigeria and Kenema, Sierra Leone) and rodents (Faranah, Guinea) have historically been focussed.

**3.4 Phylogenetic Analysis**

Sequences for each segment of LASV showed clustering according to previously documented lineages I-VII alongside geographical clustering with lineages I-III and VI present in Nigeria, lV in Liberia, Guinea and Sierra Leone, V in Mali and VII in Togo (Supplementary figure 2). In this analysis only L segment sequences of lineage V from Cote d’Ivoire were included due to quality control exclusion criteria. The phylogeny of the L segment indicates an older emergence of LASV in the human population, with the most recent common ancestor (MRCA) predicted in the year 828 in Nigeria, inference based on the S segment indicates the emergence in the year 1350 (Table 1).

Table 1 - The most recent common ancestor (MRCA) stratified by host and country of collection of Lassa mammarenavirus (LASV) S and L segments. Samples were collected between 1969-2018.

|  |  |  |  |
| --- | --- | --- | --- |
| **Host species** | **Country** | **S segment MRCA** | **L segment MRCA** |
| Homo sapiens (n=1181) | Benin | 1995 | 1989 |
| Guinea | 1895 | 1871 |
| Liberia | 1895 | 1627 |
| Nigeria | 1681 | 1498 |
| Sierra Leone | 1901 | 1874 |
| Togo | 2016 | 2014 |
| Hylomyscus pamfi (n=2) | Nigeria | 1681 | 1498 |
| Mastomys erythroleucus (n=18) | Guinea | 1975 | 2010 |
| Nigeria | 2008 | 2006 |
| Mastomys natalensis (n=36) | Guinea | 1938 | 1997 |
| Mali | 1951 | 2007 |
| Sierra Leone | 1909 | 1979 |

There was a lack of sequence information from lineage I and VI, however, phylogeny suggests these lineages are basal to others in Nigeria (Supplementary figure 2). Lineage VII in Togo is most closely related to Nigerian isolates and potentially diverged between 500-900 years ago. The divergence of lineage III and IV is predicted to have occurred between the years 1332-1551. Introduction to countries west of Nigeria appears to be by dispersal initially to Liberia, followed by Guinea in the 1700s, followed by Sierra Leone and Mali approximately 100 years later. A lack of full segment sequences from lineage V limits calculation of divergence from the most recent common ancestor from lineage IV (approximately 200 years). Regional-level data were available for sequences obtained from Nigeria (Figure 3 and Supplementary figure 3). The lineages circulating in Nigeria also tend to form regional clusters, with lineage II dispersed in the southern region and III across the central region of the country (data not shown).

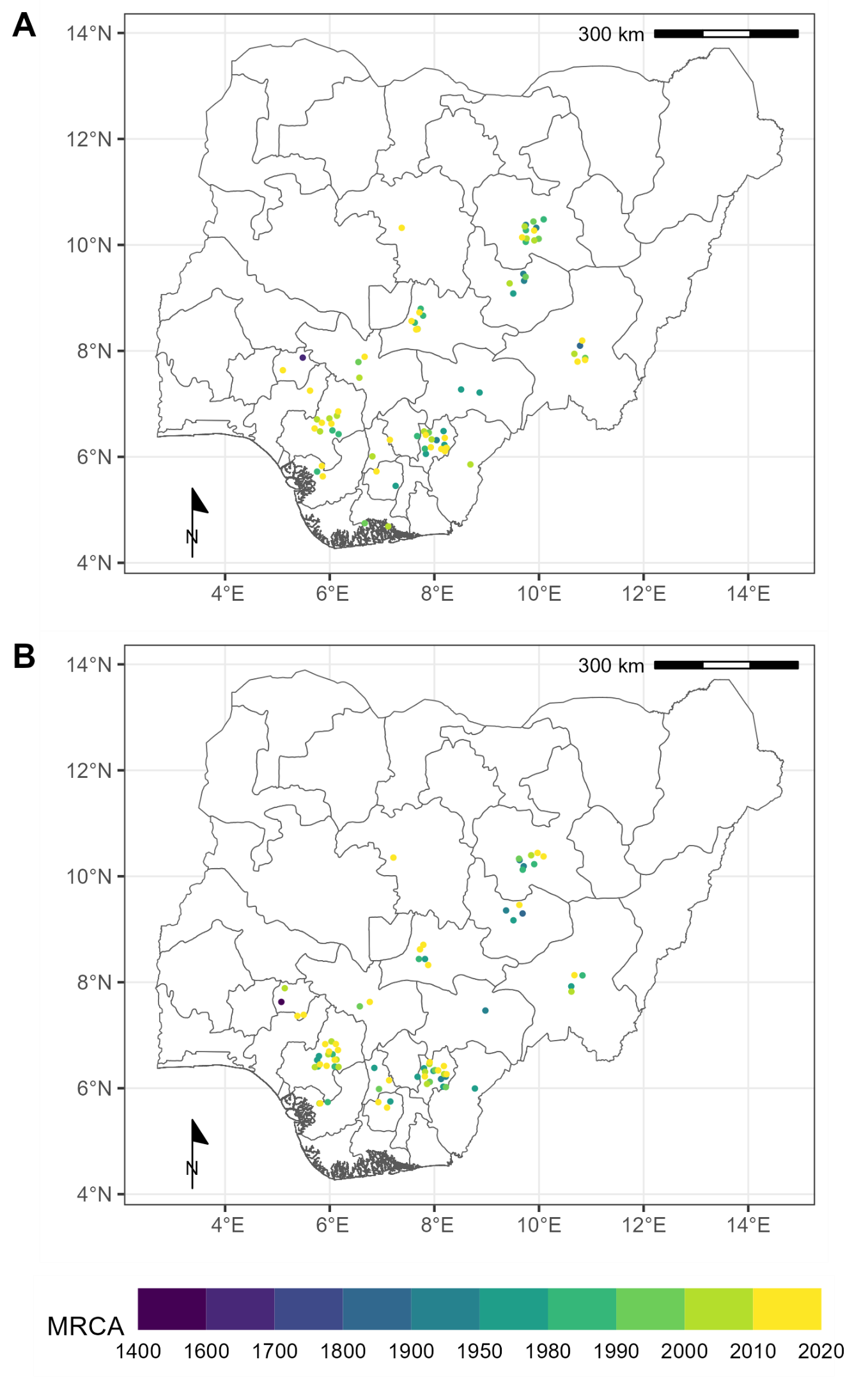


Figure 3 –Most recent common ancestor (MRCA) for included samples in phylogenetic analysis from Nigerian states for the S segment (**A**) and L segment (**B**). Colours of points correspond to year of MRCA for each individual sample. Points are jittered within the level-1 administrative region to reduce overplotting.

1. **Discussion**

Our analyses of 2,298 LASV sequences obtained from GenBank highlights the spatial biases in the availability of sequence data that may limit our understanding of the dispersal of LASV lineages in West Africa. First, sequence data was typically obtained from three countries: Nigeria (56%), Guinea (20%) and Sierra Leone (14%). Sequence data from other countries in the region with detected human cases or rodent infections are required to increase confidence in the currently inferred westward expansion. The overrepresentation of data from these tree countries has been mapped as relative sequencing effort to identify regions where increased *Lassa mammarenavirus* sequencing are required to counteract current sequencing biases. Second, geographic clustering of LASV lineages, suggest isolated events of human-to-rodent transmission and the emergence of LASVdating from 1498 in Nigeria. Similarly, Olayemi *et al.* report evidence of human-to-rodent transmission in Nigeria (Olayemi *et al.*, 2020) Comparatively limited data from non-human hosts with limited genome coverage, (69/703 sequences encompassed complete genes) produce important uncertainty around the observation of human-to-rodent transmission. Taken together, this data highlight limited surveillance among animal species, necessitating further investments in data acquisition and sharing to accurately define the spatiotemporal expansion of LASVin West Africa.

The phylogenetic analysis of LASV stratified by host species supports spatial evolution, as opposed to intra-host viral evolution (Supplementary figure 3). For instance, LASV sequences from *M. erytholeucus* sampled in Nigeria and Guinea clustered within lineages III and IV, respectively. Interestingly, these isolates appear to occur after the emergence of the most recent common ancestor virus circulating among humans and *M. natalensis* in these countries (Table 1), suggesting introduction of LASV into *M. erythroleucus* populations was a consequence of pathogen circulation in human and *M. natalensis* populations. Sequences from *M. natalensis* in Sierra Leone exhibit minimal clustering, and were interspersed with sequences from humans, potentially representing isolated events of pathogen introduction into human populations with spillback into commensal rodent populations (i.e., reverse zoonosis). The most recent common ancestor of LASV sequences from *M. natalensis* in Sierra Leone suggest a later emergence of the virus in this country. Our findings corroborate those of Andersen et al., that within Sierra Leone LASV appears to have emerged in human hosts before rodents (Andersen *et al.*, 2015). However, this data must be caveated by the limited information from rodent species in these locations.

There is a lower coverage of rodent-derived LASV sequences, with those from the primary reservoir *M. natalensis* forming fewer than one-third of the sequences (n = 609), with substantially lower sampling of other possible rodent hosts, including other *Mastomys* species. Rodent sampling has not increased at the same rate as human samples despite increased sampling effort since 2008 (Lecompte *et al.*, 2005, 2006; Olayemi *et al.*, 2016). There is substantial heterogeneity in the locations in which rodent and human samples are available. For example, a relatively high number of rodent samples have been obtained from Guinea while few human sequences are available from these locations. The inverse is true of Nigeria where most human derived sequences are obtained but fewer than 80 rodent sequences are available, and all of these from a single state. The number of suspected and reported cases was found to be positively but non-linearly associated with the number of available sequences. This is suggestive of a consolidation of research and focus of sampling in areas historically with high numbers of human cases but has led to a paucity of sequences from elsewhere in the endemic region. The limited number of full segment sequences from rodents, from few geographic locations, limits our understanding of viral radiation in rodent hosts, particularly from species which are not considered the primary reservoir, e.g., *H. pamfi.* The most recent common ancestor for the viral sequence obtained from *H. pamfi* is estimated to be in the late 1600s, it is therefore possible lineage VI and/or *H. pamfi* as a reservoir of LASV has gone undetected due to lack of sufficient sampling (Olayemi *et al.*, 2016).

Interpreting available LASV sequences is challenging for several reasons. A large proportion of available sequences (70%) have been obtained within Lassa fever research programs, representing spatial ascertainment bias (Khan *et al.*, 2008; Ehichioya *et al.*, 2010; Townsend Peterson, Moses and Bausch, 2014). In addition to these spatial biases’ temporal biases are apparent. Since 2016 there has been a substantial increase in the number of LASV sequences available in NCBI GenBank, reflecting increasing research effort, availability of sequencing platforms and increased data collection during Lassa fever epidemics, such as in the 2018 Nigeria Lassa fever outbreak (Control, 2018). There are notably fewer recorded sequences of LASV from Benin, Togo, and Ghana, potentially suggesting a potential a gap in surveillance and research capacity in these locations or a lack of circulating LASV, despite several reported outbreaks (Yadouleton *et al.*, 2020; World Health Organisation, 2022b; Ghana Health Services, 2023). Phylogenetic analysis on 60% of our initial dataset, following removal of sequences due to incompleteness (n = 906) or missing geographic information (n = 134) demonstrated geographic clustering of LASV lineages, supporting prior analyses (Lalis *et al.*, 2013; Manning, Forrester and Paessler, 2015; Olayemi *et al.*, 2016, 2020; Wiley *et al.*, 2019; Ibukun, 2020; Olayemi and Fichet-Calvet, 2020; Yadouleton *et al.*, 2020). Increased data availability from Nigeria following increased LASV surveillance allowed regional analysis of phylogeny for lineages II and III supporting previous findings of expansion of these lineages from North-East Nigeria to the South-West of the country (Bowen *et al.*, 2000; Ehichioya *et al.*, 2011; Naidoo and Ihekweazu, 2020).

A substantial number (n = 906) of the sequences retrieved corresponded to short fragments derived from PCR products used for diagnostic purposes rather than for viral genomic surveillance. LASV is a segmented virus, and it was not possible to identify complete genome sequences since both S and L segments are reported separately on the sequence’s repository. The molecular clock analyses from L protein indicated an earlier emergence of LASV when compared to S segment analysis (828 and 1350 respectively), potentially because the viral RNA polymerase (L protein) is less affected by selective pressure than the S segment (Andersen *et al.*, 2015; Hastie and Saphire, 2018; Ibukun, 2020).

Despite these challenges, this study has synthesised currently available data on LASV sequences to investigate the location and period of sampling to reconstruct the dispersal of viral lineages across the endemic region. We identify strong lineage and geographic clustering of LASV sequences, supporting the role of the rodent reservoir to sustain the endemic cycle. Despite the regionalisation of LF being driven by rodent-to-human transmission, there remains scarce LASV genomic data from non-human hosts. We have mapped the locations of relative under sampling to guide targeted efforts to counteract biases in currently available data for both rodent and human derived sequences. Expanded sampling of LASV from animal species within the endemic region will improve our current understanding of LASV evolution and ecology and improve confidence in current estimates of westward expansion of Lassa fever in humans. Further understanding of the viral evolution dynamics of LASV and spatial expansion of current lineages will be vital to ensure adequate diagnostic tools are available to respond to the expected sporadic outbreaks of Lassa Fever across the region.

**Supplementary material**

Supplementary table 1 presents the number of suspected and confirmed cases of Lassa Fever reported from countries between 2008 and 2023 at a subnational level that were used to calculate the number of cases per 100,000 people. References for the reports used to produce this dataset are included.

Supplementary table 2 presents the GenBank accession number of analysed sequences, including available data about host, country, region, year, sequence length, genome segment (L or S) and predicted MRCA.

Supplementary figure 1 displays a map of West Africa with country names for reference against Figure 1 and Figure 2.

Supplementary figure 2 displays the produced time-calibrated phylogeny for both the small segment (S) and large segment (L) from included LASV sequences.

Supplementary figure 3 displays the produced time-calibrated phylogeny for both the small segment (S) and large segment (L) from LASV sequences obtained in Nigeria.

**Funding**

Pan-African Network for Rapid Research, Response and Preparedness for Infectious Diseases Epidemics – PANDORA-ID-NET ([www.pandora-id.net](http://www.pandora-id.net)), funded through the European and Developing Countries Clinical Trials Partnership (EDCTP) (grant number RIA2016E-1609). DS is supported by a PhD studentship from the UK Biotechnology and Biological Sciences Research Council [BB/M009513/1]. AZ is in receipt of UK NIHR Senior Investigator Award.

**Author contributions**

Conceptualisation: DS and LBA; Methodology: HF, DS and LBA; Formal Analyses: HF, DS and LBA; Investigation: HF, DS and LBA; Supervision: LBA; Data Curation: HF and DS; Writing – original draft preparation: HF, DS, LBA; Writing – Review and Editing: IH, LE, NH, RA, RK, FN AZ and TMcH; Funding acquisition: AZ and FN.

**Data availability and reproducibility**

All data used in these analyses are publicly available from GenBank. The accession numbers of records used and code to reproduce the metadata analyses are available as an archived Git release on Zenodo (<https://doi.org/10.5281/zenodo.6340162>)

**Conflict of interests**

The authors declare no conflict of interests

**References**

Andersen, K.G. *et al.* (2015) ‘Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus’, *Cell* [Preprint]. Available at: https://doi.org/10.1016/j.cell.2015.07.020.

Ashcroft, J.W. *et al.* (2022) ‘Pathogens that Cause Illness Clinically Indistinguishable from Lassa Fever, Nigeria, 2018’, *Emerging Infectious Diseases*, 28(5), pp. 994–997. Available at: https://doi.org/10.3201/eid2805.211153.

Asogun, D.A. *et al.* (2019) ‘Lassa Fever: Epidemiology, Clinical Features, Diagnosis, Management and Prevention. [Review]’, *Infectious Disease Clinics of North America*, 33(4), pp. 933–951. Available at: https://doi.org/10.1016/j.idc.2019.08.002.

Bangura, U. *et al.* (2021) ‘Lassa Virus Circulation in Small Mammal Populations in Bo District, Sierra Leone’, *BIOLOGY-BASEL*. ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND: MDPI. Available at: https://doi.org/10.3390/biology10010028.

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. Available at: https://doi.org/10.1371/journal.pcbi.1008811.

Becker, G. and Lawrence, M. (2021) ‘genbankr: Parsing GenBank files into semantically useful objects’.

Bowen, M.D. *et al.* (2000) ‘Genetic Diversity among Lassa Virus Strains’, *Journal of Virology*, 74(15), pp. 6992–7004. Available at: https://doi.org/10.1128/jvi.74.15.6992-7004.2000.

Control, N.C. for D. (2018) *Lassa fever Situation Report*.

Ehichioya, D.U. *et al.* (2010) ‘Lassa fever, Nigeria, 2005-2008’, *Emerging Infectious Diseases* [Preprint]. Available at: https://doi.org/10.3201/eid1606.100080.

Ehichioya, D.U. *et al.* (2011) ‘Current molecular epidemiology of Lassa virus in Nigeria’, *Journal of Clinical Microbiology* [Preprint]. Available at: https://doi.org/10.1128/JCM.01891-10.

Forni, D. and Sironi, M. (2020) ‘Population Structure of Lassa Mammarenavirus in West Africa’, *Viruses*, 12(4), p. 437.

Ghana Health Services (2023) ‘Lassa Fever Press Release Ghana 2023’. Available at: https://osf.io/ft2gy/ (Accessed: 19 April 2023).

Günther, S. and Lenz, O. (2004) ‘Lassa virus’, *Critical Reviews in Clinical Laboratory Sciences* [Preprint]. Available at: https://doi.org/10.1080/10408360490497456.

Hallam, S.J. *et al.* (2018) ‘Review of mammarenavirus biology and replication’, *Frontiers in Microbiology*, 9(AUG), pp. 1–8. Available at: https://doi.org/10.3389/fmicb.2018.01751.

Hastie, K.M. and Saphire, E.O. (2018) ‘Lassa virus glycoprotein: stopping a moving target’, *Current Opinion in Virology* [Preprint]. Available at: https://doi.org/10.1016/j.coviro.2018.05.002.

Lo Iacono, G. *et al.* (2015) ‘Using Modelling to Disentangle the Relative Contributions of Zoonotic and Anthroponotic Transmission: The Case of Lassa Fever’, *PLoS Neglected Tropical Diseases* [Preprint]. Available at: https://doi.org/10.1371/journal.pntd.0003398.

Ibukun, F.I. (2020) ‘Inter-lineage variation of lassa virus glycoprotein epitopes: A challenge to lassa virus vaccine development’, *Viruses* [Preprint]. Available at: https://doi.org/10.3390/v12040386.

Kahle, D. and Wickham, H. (2013) ‘ggmap: Spatial Visualization with ggplot2’, *The R Journal*, 5(1), pp. 144–161. Available at: https://doi.org/10.32614/RJ-2013-014.

Khan, S.H. *et al.* (2008) ‘New opportunities for field research on the pathogenesis and treatment of Lassa fever’, *Antiviral Research* [Preprint]. Available at: https://doi.org/10.1016/j.antiviral.2007.11.003.

Klitting, R. *et al.* (2021) *Predicting the evolution of Lassa Virus endemic area and population at risk over the next decades*. preprint. Microbiology. Available at: https://doi.org/10.1101/2021.09.22.461380.

Lalis, A. *et al.* (2013) ‘The Impact of Human Conflict on the Genetics of Mastomys natalensis and Lassa Virus in West Africa’, *PLoS ONE*, 7(5). Available at: https://doi.org/10.1371/journal.pone.0037068.

Lecompte, E. *et al.* (2005) ‘Molecular identification of four cryptic species of Mastomys (Rodentia, Murinae)’, *Biochemical Systematics and Ecology* [Preprint]. Available at: https://doi.org/10.1016/j.bse.2004.12.015.

Lecompte, E. *et al.* (2006) ‘Mastomys natalensis and Lassa fever, West Africa’, *Emerging Infectious Diseases* [Preprint]. Available at: https://doi.org/10.3201/eid1212.060812.

Manning, J.T., Forrester, N. and Paessler, S. (2015) ‘Lassa virus isolates from Mali and the Ivory Coast represent an emerging fifth lineage’, *Frontiers in Microbiology* [Preprint]. Available at: https://doi.org/10.3389/fmicb.2015.01037.

McCormick, J.B. *et al.* (1987) ‘A prospective study of the epidemiology and ecology of Lassa fever.’, *The Journal of infectious diseases*, 155(3), pp. 437–44. Available at: https://doi.org/10.1093/infdis/155.3.437.

Naidoo, D. and Ihekweazu, C. (2020) ‘Nigeria’s efforts to strengthen laboratory diagnostics – Why access to reliable and affordable diagnostics is key to building resilient laboratory systems’, *African Journal of Laboratory Medicine*, 9(2). Available at: https://doi.org/10.4102/ajlm.v9i2.1019.

National Center for Biotechnology Information (2022) *National Center for Biotechnology Information*. Available at: https://www.ncbi.nlm.nih.gov/ (Accessed: 3 February 2022).

Nnaji, N.D. *et al.* (2021) ‘The deuce-ace of Lassa Fever, Ebola virus disease and COVID-19 simultaneous infections and epidemics in West Africa: clinical and public health implications’, *Tropical Medicine and Health*, 49(1), p. 102. Available at: https://doi.org/10.1186/s41182-021-00390-4.

Okoro, O.A. *et al.* (2020) ‘Descriptive epidemiology of Lassa fever in Nigeria, 2012-2017’, *The Pan African Medical Journal*, 37, p. 15. Available at: https://doi.org/10.11604/pamj.2020.37.15.21160.

Olayemi, A. *et al.* (2016) ‘New Hosts of The Lassa Virus’, *Scientific Reports* [Preprint]. Available at: https://doi.org/10.1038/srep25280.

Olayemi, A. *et al.* (2020) ‘Determining ancestry between rodent-and human-derived virus sequences in endemic foci: Towards a more integral molecular epidemiology of lassa fever within West Africa’, *Biology* [Preprint]. Available at: https://doi.org/10.3390/biology9020026.

Olayemi, A. and Fichet-Calvet, E. (2020) ‘Systematics, ecology, and host switching: Attributes affecting emergence of the Lassa virus in rodents across western Africa’, *Viruses* [Preprint]. Available at: https://doi.org/10.3390/v12030312.

Oti, V.B. (2018) ‘A Reemerging Lassa Virus: Aspects of Its Structure, Replication, Pathogenicity and Diagnosis’, in Alfonso J. Rodriguez-Morales (ed.) *Current Topics in Tropical Emerging Diseases and Travel Medicine*. BoD – Books on Demand.

R Core Team (2021) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rambaut, A. *et al.* (2018) ‘Posterior summarization in Bayesian phylogenetics using Tracer 1.7’, *Systematic Biology* [Preprint]. Available at: https://doi.org/10.1093/sysbio/syy032.

Suchard, M.A. *et al.* (2018) ‘Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10’, *Virus Evolution* [Preprint]. Available at: https://doi.org/10.1093/ve/vey016.

Takah, N.F. *et al.* (2019) ‘Sensitivity and specificity of diagnostic tests for Lassa fever: a systematic review’, *BMC Infectious Diseases*, 19(1), p. 647. Available at: https://doi.org/10.1186/s12879-019-4242-6.

Townsend Peterson, A., Moses, L.M. and Bausch, D.G. (2014) ‘Mapping transmission risk of lassa fever in West Africa: The importance of quality control, sampling bias, and error weighting’, *PLoS ONE* [Preprint]. Available at: https://doi.org/10.1371/journal.pone.0100711.

Welch, S.R. *et al.* (2018) ‘The S Genome Segment Is Sufficient to Maintain Pathogenicity in Intra-Clade Lassa Virus Reassortants in a Guinea Pig Model’, *Frontiers in Cellular and Infection Microbiology*, 8. Available at: https://www.frontiersin.org/article/10.3389/fcimb.2018.00240 (Accessed: 3 February 2022).

Whitmer, S.L.M. *et al.* (2018) ‘New lineage of lassa virus, Togo, 2016’, *Emerging Infectious Diseases*, 24(3), pp. 599–602. Available at: https://doi.org/10.3201/eid2403.171905.

Wiley, M.R. *et al.* (2019) ‘Lassa virus circulating in Liberia: a retrospective genomic characterisation’, *The Lancet Infectious Diseases* [Preprint]. Available at: https://doi.org/10.1016/S1473-3099(19)30486-4.

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

World Health Organisation (2022a) *Lassa fever*. Available at: https://www.who.int/health-topics/lassa-fever#tab=tab\_1 (Accessed: 22 February 2022).

World Health Organisation (2022b) *Lassa Fever – Togo*. Available at: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON362 (Accessed: 24 November 2022).

Wulff, H., Fabiyi, A. and Monath, T.P. (1975) ‘Recent isolations of Lassa virus from Nigerian rodents’, *Bulletin of the World Health Organization*, 52(4–6), pp. 609–613. Available at: https://doi.org/PMC2366652.

Yadouleton, A. *et al.* (2019) ‘Lassa virus in pygmy mice, Benin, 2016-2017’, *Emerging Infectious Diseases* [Preprint]. Available at: https://doi.org/10.3201/eid2510.180523.

Yadouleton, A. *et al.* (2020) ‘Lassa fever in Benin: description of the 2014 and 2016 epidemics and genetic characterization of a new Lassa virus’, *Emerging Microbes & Infections*, pp. 1–23.