




# Genome Sequences of Five Arenaviruses from Pygmy Mice (*Mus minutoides*) in Sierra Leone

Matej Vučak,<sup>a</sup> James Bangura,<sup>b</sup> Bruno M. Gherzi,<sup>c</sup> Jenna Nichols,<sup>a</sup> Joseph Hughes,<sup>a</sup> Ana da Silva Filipe,<sup>a</sup> Alexandre Tremeau-Bravard,<sup>c</sup> David J. Wolking,<sup>c</sup> Emmanuel Amara,<sup>b</sup> Abdulai Bangura,<sup>b</sup> Marilyn C. Kanu,<sup>b</sup> Osman T. Kanu,<sup>b</sup> Dickson Kargbo,<sup>b</sup> Edwin G. Lavalie,<sup>b</sup> Victor Lungay,<sup>b</sup> Willie Robert,<sup>b</sup> Mohamed Turay,<sup>b</sup> Steven Fornie,<sup>d</sup> Thomas T. Samba,<sup>d</sup> Bankolay B. Sesay,<sup>d</sup> Patrick Swaray,<sup>d</sup> Mohamed A. Vandi,<sup>d</sup> Mohamed Alpha Bah,<sup>e</sup> Andrew A. Mansaray,<sup>e</sup> Brian H. Bird,<sup>c</sup>  Andrew J. Davison<sup>a</sup>

<sup>a</sup>MRC–University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

<sup>b</sup>University of Makeni and University of California, Davis One Health Program, Makeni, Sierra Leone

<sup>c</sup>One Health Institute, School of Veterinary Medicine, University of California, Davis, Davis, California, USA

<sup>d</sup>Ministry of Health and Sanitation, Freetown, Sierra Leone

<sup>e</sup>Ministry of Agriculture and Forestry, Freetown, Sierra Leone

**ABSTRACT** The genome sequences of five strains of a mammarenavirus were assembled from metagenomic data from pygmy mice (*Mus minutoides*) captured in Sierra Leone. The nearest fully sequenced relatives of this virus, which was named Seli virus, are lymphocytic choriomeningitis virus, Lunk virus, and Ryukyu virus.

Members of genus *Mammarenavirus* in family *Arenaviridae* have genomes consisting of two single-stranded ambisense RNA segments (L and S) and most have rodents as hosts (1). Among a collection of rodents captured in Sierra Leone that had tested positive for the mammarenavirus Lassa virus (LASV), we used visual recognition, DNA barcoding of the *MT-CO1* gene (2), and sequencing of the *MT-CO1* and *MT-CYB* genes to identify five apparently healthy animals from Koinadugu District as pygmy mice (*Mus minutoides*), in which LASV has been reported occasionally (3). An additional 18 pygmy mice in this collection were LASV negative. The five mice had tested positive for LASV in two assays for the L gene (encoding the RNA-directed RNA polymerase) but not in an assay for the GPC gene (encoding the glycoprotein precursor) (Table 1, see footnotes b and c for assay details). We determined the genome sequences of the mammarenaviruses in these mice.

Blood, oral swab, and urogenital swab samples were obtained from each mouse. Nucleic acids were extracted using Applied Biosystems MagMAX CORE and pathogen kits with a KingFisher Duo Prime purification system and treated with Ambion DNase I. RNA was purified using the Beckman Coulter RNAClean XP kit and depleted of rRNA using an Illumina Ribo-Zero Plus kit. A metagenomic method (G-Meta) (4) was used to prepare cDNA and sequencing libraries, which were analyzed on an Illumina NextSeq 550 instrument using a high-output 300-cycle cartridge.

Bioinformatic tools were used with default parameters unless specified otherwise. The genome sequences were determined using a pipeline built around SPAdes v3.13 (5) and DIAMOND v2.0.9 (6) ([https://github.com/mvvucak/LASV\\_Sample\\_Classification/](https://github.com/mvvucak/LASV_Sample_Classification/)). The 15 data sets (each consisting of 5,076,546 to 34,278,974 paired-end reads) were quality filtered using Trim Galore v0.4.0 ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). The three data sets for each mouse were combined and assembled *de novo*, and contigs related to mammarenavirus sequences were identified. The contigs were joined into genome sequences manually by iteratively incorporating reads sharing short sequences near their ends. Individual and combined data sets were then mapped to the genome sequences using Bowtie 2 v2.3.1 (7) with the `–local` parameter. The alignments were visualized using

**Editor** Jelle Matthijssens, KU Leuven

**Copyright** © 2022 Vučak et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Brian H. Bird, [bhbird@ucdavis.edu](mailto:bhbird@ucdavis.edu), or Andrew J. Davison, [andrew.davison@glasgow.ac.uk](mailto:andrew.davison@glasgow.ac.uk).

The authors declare no conflict of interest.

**Received** 11 February 2022

**Accepted** 11 March 2022

**Published** 7 April 2022

**TABLE 1** Data on SELV in this study<sup>a</sup>

| Animal | Location <sup>b</sup>      | qPCR LASV load by sample type (C <sub>T</sub> ) |                 |                 |                 | Size (nt) of:  |                |                | G + C (%) content of: |             |             | NCBI SRA accession no. by sample type |          |          | NCBI GenBank accession no. of: |          |          |          |
|--------|----------------------------|---|-----------------|-----------------|-----------------|----------------|----------------|----------------|-----------------------|-------------|-------------|---------------------------------------|----------|----------|--------------------------------|----------|----------|----------|
|        |                            | BL <sup>c</sup>                                 | OS <sup>c</sup> | US <sup>c</sup> | BL <sup>d</sup> | L <sup>e</sup> | S <sup>e</sup> | S <sup>e</sup> | L                     | S           | L           | S                                     | L        | S        | L                              | S        | L        | S        |
| 00094  | Sengbeh, Gbenikoro         | 31.7  | 36.4            | ND              | 27.8            | 7,173 (7,160)  | 3,390 (3,377)  | 45             | 47                    | SRX13144664 | SRX13144665 | SRX13144671                           | OL364850 | OL364851 | OL415139                       | OL415140 | OL441084 | OL441085 |
| 00127  | Wara Wara Bafodia, Bafodia | 38.6  | ND              | ND              | 27.2            | 7,210 (7,171)  | 3,389 (3,384)  | 44             | 47                    | SRX13144672 | SRX13144673 | SRX13144674                           | OL364852 | OL364853 | OL415140                       | OL441085 |          |          |
| 00207  | Diang, Badala              | 27.5  | 31.7            | 37.6            | 21.7            | 7,180 (7,180)  | 3,389 (3,384)  | 44             | 47                    | SRX13144675 | SRX13144676 | SRX13144677                           | OL364854 | OL364855 | OL415141                       | OL441086 |          |          |
| 00367  | Wara Wara Bafodia, Bafodia | 31.6  | 32.3            | 31.6            | 27.4            | 7,173 (7,173)  | 3,389 (3,389)  | 45             | 47                    | SRX13144678 | SRX13144666 | SRX13144667                           | OL364856 | OL364857 | OL415142                       | OL441087 |          |          |
| 00414  | Diang, Badala              | 28.6  | 28.9            | 35.7            | 21.5            | 7,181 (7,176)  | 3,388 (3,363)  | 45             | 47                    | SRX13144668 | SRX13144669 | SRX13144670                           | OL364858 | OL364859 | OL415143                       | OL441088 |          |          |

<sup>a</sup> NCBI BioProject no. PRJNA780418. C<sub>T</sub> cycle threshold; BL, blood sample; OS, oral swab sample; US, urogenital swab sample; ND, not detected; nt, nucleotide; L, L segment; S, S segment; SRA, sequence read archive; MT-CO1, mitochondrial gene encoding cytochrome c oxidase I; MT-CYB, mitochondrial gene encoding cytochrome b.

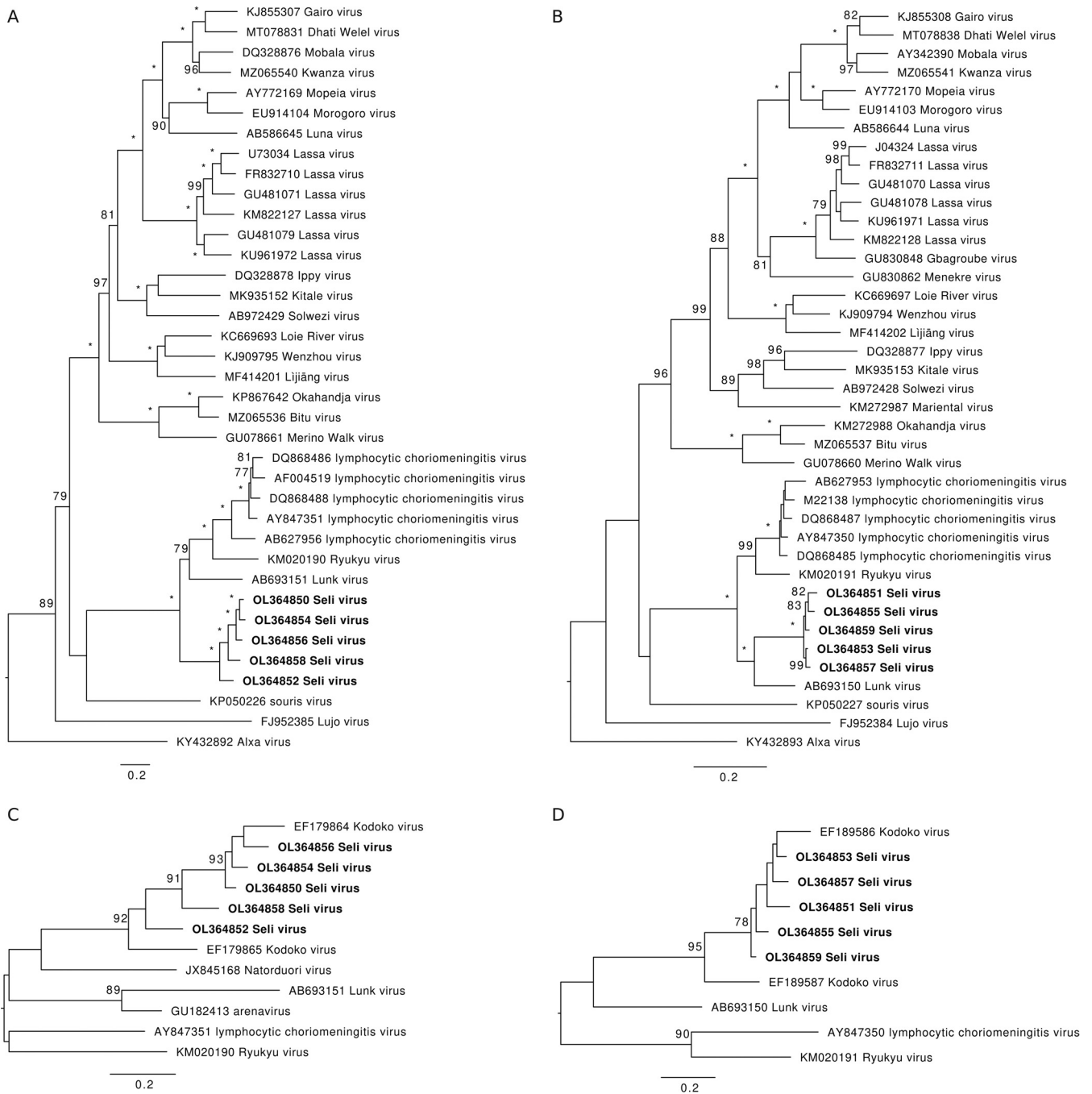
<sup>b</sup> Chiefdom and site in Koinadugu District.

<sup>c</sup> Using a method targeting the L gene (16).

<sup>d</sup> Using a RealStar LASV RT-PCR 2.0 kit targeting the L gene (Altona Diagnostics). LASV was not detected in these samples using a RealStar LASV RT-PCR 2.0 kit targeting the GPC gene.

<sup>e</sup> Including undetermined nucleotides at the segment termini. The size without these nucleotides is given in parentheses.

<sup>f</sup> Sequence determined from the combined data set for each animal and by Illumina sequencing of OS DNA from each animal.



**FIG 1** Phylogenetic analyses of SELV. The complete amino acid sequences of the L (A) and the NP (B) proteins of Old World mammarenaviruses and partial nucleotide sequences of the L (C) and the NP (D) genes of SELV and its closer relatives were aligned using MAFFT v7.475 (13). The alignments were inspected visually in JalView v2.11 (14). Maximum likelihood phylogenies for each alignment were reconstructed using IQ-Tree v2.1.3 (15) by instituting the best fitting model selected by ModelFinder according to the Bayesian information criterion. The trees in A and B were rooted using a New World mammarenavirus, Junin virus (not shown; GenBank accession no. [AY358022](#) and [AY358023](#), respectively). The trees in C and D were rooted using Lassa virus (not shown; GenBank accession no. [U73034](#) and [J04324](#), respectively). Branch support was assessed using 1,000 bootstrap replicates, with values above 75% shown at the nodes and values of 100% indicated by asterisks. The viruses are denoted by GenBank accession no. and name, and SELV strains are indicated by bold font. The scale bars represent substitutions per site.

Tablet v1.21.02.08 (8), and final adjustments to the consensus sequences were made. Published information on other mammarenaviruses was used to locate the segment ends when possible and to annotate the genome sequences.

The analysis indicated that a single virus, in some instances with a few single-nucleotide variants, was present in each mouse. Phylogenetic analyses showed that these

viruses represent strains of a mammarenavirus, which we have named Seli virus (SELV) after a river in Sierra Leone (Fig. 1). Thus, among fully sequenced mammarenaviruses, SELV is most closely related to Lunk virus, which was isolated from a pygmy mouse in Zambia (9); Ryukyu virus, which was isolated from a Ryukyu mouse (*Mus caroli*) in China (10); and lymphocytic choriomeningitis virus, which is widespread in house mice (*Mus musculus*) (1) (Fig. 1A and B). SELV is sufficiently distinct to meet the formal criteria for assignment to a new viral species (1), to which additional phylogenetic analyses based on short sequences suggest that Kodoko virus, from pygmy mice in Guinea (11), may also belong (Fig. 1C and D). Further mapping experiments using Bowtie 2 as described above showed that reads originating from recognized LASV clades (12) were completely absent from all 15 data sets. This finding suggests that the L gene assays used are capable of detecting SELV as well as LASV.

The work was conducted in partnership with the Government of Sierra Leone Ministry of Agriculture and Forestry representatives working under a ministry-approved wildlife specimen collection permit, and with further approvals from the University of California, Davis Institutional Animal Care and Use Committee.

**Data availability.** Sequence information and accession numbers are provided in Table 1.

## ACKNOWLEDGMENTS

The work described in this report was funded by DARPA grant no. D18AC00028 in support of the PREventing EMerging Pathogenic Threats (PREEMPT) program. Information on the University of California, Davis PREEMPT Project Consortium, including a list of Consortium members, is available online at <https://www.preemptproject.org>. The funder had no role in study design, data collection and analysis, preparation of the manuscript, or decision to publish.

We thank colleagues in Sierra Leone, including Vice-Chancellor Father Joseph Turay, Mercy Mwaura, and the project administration finance team and drivers at the University of Makeni (UNIMAK); James Squire and Joseph Kanu of the National Disease Surveillance Program at the Ministry of Health and Sanitation; and Momoh Massaquoi, Mohamed K Marrah, and Mohamed Turay of the Wildlife Division of the Ministry of Agriculture and Forestry.

## REFERENCES

- Radoshitzky SR, Buchmeier MJ, Charrel RN, Clegg JCS, Gonzalez JJ, Günther S, Hepojoki J, Kuhn JH, Lukashevich IS, Romanowski V, Salvato MS, Sironi M, Stenglein MD, de la Torre JC, ICTV Report Consortium. 2019. ICTV virus taxonomy profile: *Arenaviridae*. *J Gen Virol* 100:1200–1201. <https://doi.org/10.1099/jgv.0.001280>.
- Townzen JS, Brower AVZ, Judd DD. 2008. Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. *Med Vet Entomol* 22:386–393. <https://doi.org/10.1111/j.1365-2915.2008.00760.x>.
- Kenmoe S, Tchatchouang S, Ebogo-Belobo JT, Ka'e AC, Mahamat G, Guiamdjo SR, Bowo-Ngandji A, Demeni EC, Che E, Tchami ND, Amougou-Atsama M, Nzukui ND, Mbongue MC, Mbaga DS, Kenfack S, Kingue BS, Amvongo AN, Efietngab AN, Tazokong HR, Modiyinji AF, Kengne-Nde C, Sadeuh-Mba SA, Njouom R. 2020. Systematic review and meta-analysis of the epidemiology of Lassa virus in humans, rodents and other mammals in sub-Saharan Africa. *PLoS Negl Trop Dis* 14:e0008589. <https://doi.org/10.1371/journal.pntd.0008589>.
- Thomson E, Ip CL, Badhan A, Christiansen MT, Adamson W, Ansari MA, Bibby D, Breuer J, Brown A, Bowden R, Bryant J, Bonsall D, Da Silva Filipe A, Hinds C, Hudson E, Klenerman P, Lythgow K, Mbisa JL, McLauchlan J, Myers R, Piazza P, Roy S, Trebes A, Sreenu VB, Witteveldt J, STOP-HCV Consortium, Barnes E, Simmonds P. 2016. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. *J Clin Microbiol* 54:2470–2484. <https://doi.org/10.1128/JCM.00330-16>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* 14:193–202. <https://doi.org/10.1093/bib/bbs012>.
- Ishii A, Thomas Y, Moonga L, Nakamura I, Ohnuma A, Hang'ombe BM, Takada A, Mweene AS, Sawa H. 2012. Molecular surveillance and phylogenetic analysis of Old World arenaviruses in Zambia. *J Gen Virol* 93:2247–2251. <https://doi.org/10.1099/vir.0.044099-0>.
- Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, Jiang J, Yang J, Dong J, Sun L, Zhu Y, Li Y, Zheng D, Zhang C, Su H, Zheng Y, Zhou H, Zhu G, Li H, Chmura A, Yang F, Daszak P, Wang J, Liu Q, Jin Q. 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome* 6:178. <https://doi.org/10.1186/s40168-018-0554-9>.
- Lecompte E, ter Meulen J, Emonet S, Daffis S, Charrel RN. 2007. Genetic identification of Kodoko virus, a novel arenavirus of the African pigmy mouse (*Mus Nannomys minutoides*) in West Africa. *Virology* 364:178–183. <https://doi.org/10.1016/j.virol.2007.02.008>.
- Welch SR, Scholte FEM, Albariño CG, Kainulainen MH, Coleman-McCray JD, Guerrero LW, Chakrabarti AK, Klerna JD, Nichol ST, Spengler JR, Spiropoulou CF. 2018. The S genome segment is sufficient to maintain pathogenicity in intra-clade Lassa virus reassortants in a guinea pig

- model. *Front Cell Infect Microbiol* 8:240. <https://doi.org/10.3389/fcimb.2018.00240>.
13. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
  14. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
  15. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
  16. Nikisins S, Rieger T, Patel P, Müller R, Günther S, Niedrig M. 2015. International external quality assessment study for molecular detection of Lassa virus. *PLoS Negl Trop Dis* 9:e0003793. <https://doi.org/10.1371/journal.pntd.0003793>.