HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2019 November 14.

Published in final edited form as:

Science. 2019 January 04; 363(6422): 74–77. doi:10.1126/science.aau9343.

Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak

A full list of authors and affiliations appears at the end of the article.

Abstract

The 2018 Nigerian Lassa fever season saw the largest ever recorded upsurge of cases, raising concerns over the emergence of a strain with increased transmission rate. To understand the molecular epidemiology of this upsurge we performed, for the first time at the epicenter of an unfolding outbreak, metagenomic nanopore sequencing directly from patient samples, an approach dictated by the highly variable genome of the target pathogen. Genomic data and phylogenetic reconstructions were communicated immediately to Nigerian authorities and the WHO to inform the public health response. Real-time analysis of 36 genomes, and subsequent confirmation using all 120 samples sequenced in-country, revealed extensive diversity and phylogenetic intermingling with strains from previous years, suggesting independent zoonotic transmission events; allaying concerns of an emergent strain or extensive human-to-human transmission.

Lassa fever is an acute viral hemorrhagic illness, first described in 1969 in the town of Lassa, Nigeria (1). It is contracted primarily through exposure to urine or feces of infected *Mastomys* spp. rodents or, less frequently, through the bodily fluids of infected humans. Lassa virus (LASV) is endemic in parts of West Africa including Nigeria, Benin, Côte d'Ivoire, Mali, Sierra Leone, Guinea and Liberia (2). The upsurge of Lassa Fever cases during the 2018 endemic season in Nigeria — referred to here as the 2018 Lassa fever

*Corresponding author: Stephan Günther, Bernhard-Nocht-Institute for Tropical Medicine, Bernhard-Nocht-Str. 74, 20359 Hamburg, Germany, Phone: +49 40 42818 940, Fax: +49 40 42818 931, guenther@bni.uni-hamburg.de.

Author contributions: LEK, SG, SD, STP, PL, SG conceptualized the study; LEK, STP, PL set up the methodology; LEK, JH, AT, SD, DUE performed sequencing and data validation; LEK, PL, MAS, STP, DS, FK, JM, SLo performed the formal sequencing data analysis; LEK, SD, JH, AT, MP, LO performed sample selection, data collection and organisation of sequencing datasets; DMW, KE, DS, FK, JM set up and assisted with bioinformatics pipeline; MAS, DUO, MP, LO, YI, DIA, TO, EO, RO, JAg, BE, JAi, PE, BO, SE, PA, MA, RE, EM, RG, AE, GI, GOd, GOk, RE, JO, EOY, IO, CA, MO, RA, ET, DA, NA, POO, MOR, KOI, COI, PA, CE, GA, EI performed diagnostic analysis; LEK, STP, PL, SD visualized data presentation; LEK, STP, PL, SD wrote the manuscript; all authors reviewed and edited the manuscript; SG, MWC, JAH, RH, RV supervised the study; MP, RV, AT, CI, PF, DN, SO, EOE, SG, SD, SLu performed project administration and implementation; SG, PL, MWC, RV, RH, JAH, LEK, DUE were involved in funding acquisition.

†These authors contributed equally to this work

Competing interests: CI is member of the WHO Strategic Technical Advisory Group on Infectious Diseases; DA serves as expert for WHO blueprint; SG is member of the Scientific Advisory Group (SAG) to advise WHO on the implementation of the WHO R&D Blueprint for action to prevent epidemics (the Blueprint), including a plan for international coordination of the R&D effort in case of an epidemic of highly infections pathogen; SO serves as expert for WHO blueprint. All other authors declare no competing interests.

Data and materials availability: LASV sequences from 2018 are deposited in GenBank, under BioProject PRJNA482058 (Data S1); sequences from 2012 to 2017 are deposited under BioProjects PRJNA482054 and PRJNA482058 (Data S2, to be published elsewhere). Alignments, trees, and BEAST xml files are available at: https://github.com/ISTH-BNITM-PHE/LASVsequencing.

SUPPLEMENTARY MATERIALS:
Materials and Methods
Figures S1-S10
Supplementary Data S1 and S2
References (16–29)

outbreak — has been the largest on record reaching 1,495 suspected cases, 376 confirmed cases, and affecting over 18 states by March 18th (Fig. S1). This notably exceeds the 102 confirmed cases reported during the same period in 2017 (Fig. S1) (3). The unprecedented scale of the outbreak raised fears of the emergence of a strain with a higher rate of transmission. Due to these concerns, on February 28th the Nigeria Centre for Disease Control (NCDC) and the WHO urgently requested sequencing information and preliminary results from our pilot-scale study, which employed in-country, mid-outbreak, viral genome sequencing directly from clinical samples using a metagenomic approach on the Oxford Nanopore MinION device (Oxford Nanopore Technologies, UK). This instigated a major upscale in sequencing efforts, leading to sequencing of 120 samples.

Nanopore sequencing is an emerging technology with significant potential. The MinION is a small and robust sequencing device suited for the genetic analysis of pathogens in remote or resource-limited settings (4). Nanopore sequencing of PCR-amplicons of Ebola virus genomes provided important data from the field in real-time during the 2014–2016 Ebola virus disease outbreak in West Africa (5) and a more sophisticated multiplex amplicon sequencing methodology (6) has been used to great effect during recent Zika and Yellow fever outbreaks in Brazil (7, 8). However such an amplicon-based approach is extremely challenging for highly variable pathogens such as LASV. Due to an inter-strain nucleic acid sequence variation of up to 32% and 25% for the L (large segment encoding the RNA polymerase and the zinc binding protein) and S (small segment encoding the glycoprotein and the nucleoprotein) segments respectively (9), even PCR-based laboratory diagnosis poses a significant challenge. Designing targeted whole-genome sequencing approaches, such as PCR amplicons or bait/capture probes, without prior knowledge of the targeted LASV lineage is therefore cumbersome. Random reverse-transcription and amplification by Sequence-Independent Single Primer Amplification (SISPA) for metagenomic sequencing to identify RNA viruses has been demonstrated to work on the MinION (10) and our previous work highlighted the feasibility of retrieving complete viral genomes directly from patient samples at clinically relevant viral titers using this approach for Dengue and Chikungunya viruses (11). We describe here the application of field metagenomic sequencing of LASV at the Irrua Specialist Teaching Hospital (ISTH), Edo State, during the 2018 Lassa fever season.

A total of 120 LASV positive samples were sequenced during a seven-week mission, selected based on Cycle threshold (Ct) value and location of the 341 cases reported by ISTH between 1st January and 18th March 2018 (Figs. S1 and S2). The majority of samples originated from Edo state followed by Ondo and Ebonyi (Fig. S2). Samples selected covered the wide range of clinical viral loads observed, including several samples testing negative in one of the two real-time RT-PCR assays used (Fig. S3 and Data S1). Up to six samples were run in multiplex per MinION flow cell, along with a negative control. To produce high-confidence consensus sequences for phylogenetic inference we chose to map both basecalled reads and raw signal data to a reference sequence and call variants using Nanopolish software, as developed for the West African Ebola virus disease outbreak (5); basecalled reads were then remapped to the consensus and a further round of correction was applied (Fig. S4). Owing to the diversity of LASV, selection of an individual reference genome for read alignment was required for each sample. To select the closest existing LASV reference

genome, non-human reads from each sample were assembled *de novo* using Canu (12). A significant proportion of reads generated per sample were LASV at an average frequency of 4.26% with a maximum of 42.9% allowing for sufficient genomic sequence (>70%) for phylogenetic comparison of at least one segment in 91 of the samples tested (Figs. S3–6).

Additionally, sequences were validated by Illumina re-sequencing of 14 SISPA preparations which matched with their Nanopore counterpart with little to no divergence between them confirming the accuracy of the Nanopore approach (Table S1).

Metagenomic classification using the Centrifuge software system (13) identified 0.10% of reads from sample 110 as originating from Hepatitis A virus; providing 74% genome coverage at 20-fold depth. LASV accounted for 0.83% in the same sample, providing 96% genome coverage. This demonstrates the potential of this simple approach to identify multiple RNA viruses, including those present as co-infections. In all other samples tested, LASV was the sole pathogen identified despite a small number reads classified as other viruses (Fig. S7 and Data S1).

To dissect the molecular epidemiology of the 2018 Lassa fever outbreak in Nigeria, we performed phylogenetic analysis of all newly generated LASV sequences together with unpublished sequences from previous years (Data S2) and sequences available in GenBank. We use this as a frame of reference to document how the genomic data generated in real-time (made publicly available as posted on virological.org) provided valuable epidemiological insights into the unfolding outbreak dynamics.

Maximum likelihood phylogenetic reconstruction of the S segment sequences indicates that all 2018 viruses fall within the Nigerian LASV diversity, specifically within genotype II and III, and they are phylogenetically interspersed with Nigerian LASV sequences from previous years (Fig. 1). This phylogenetic pattern is mimicked by the L segment reconstruction (Fig. S8). Only seven viruses in the entire complete genome data set (n = 348) were identified as clustering significantly differently in the L and S segment (Supplementary Methods), in line with the small number of potential LASV reassortments identified previously (9). The phylogenetic pattern clearly implicates independent spill-over from rodent hosts as the major driver of Lassa fever incidence during the outbreak (Figs. 1 and S8).

However, a number of sequences from the 2018 outbreak clustered as pairs in the phylogenetic reconstructions, raising concerns over human-to-human transmissions. We illustrate such cluster pairs in a Bayesian time-measured tree estimated from genotype II S (Fig. 2) and L segment sequences (Fig. S9). These analyses resulted in highly similar evolutionary rate estimates for both segments (mean around 1.2×10^{-3} subst./site/year, Figs. 2 and S9–S10), in agreement with previous estimates (9). We used these rate estimates together with an estimate of the time between successive cases in a transmission chain to assess how many substitutions can be expected between directly linked infections. We compare conservative to more liberal expectations, the latter accommodating an independent upper estimate of potential sequencing errors (Figs. 2 and S9). In the S segment, for example, more than 2 substitutions between sequences from directly linked infections is highly unlikely ($P<0.01$ and $P=0.03$ respectively for the conservative and liberal probability

estimates). This expectation is consistent with the low number of substitutions observed in the coding region of human-to-human LASV transmission (14). Four clusters of sequences showing ≤4 and ≤12 nucleotide differences in S and L segment, respectively, were identified (035–045, 035–058, 137–138, and 053-089-106; for some of them only S or L segment sequence was available). Retrospective tracing revealed that the sequences for pairs 137–138 and 035–058, respectively, were in fact derived from the same patients. Epidemiological investigation of the remaining clusters did not provide evidence for transmission chains, though direct linkage cannot be excluded. In conclusion, even when applying liberal assumptions for the number of mutations during human-to-human transmission, the vast majority of cases during the 2018 outbreak resulted from spill-over from the natural reservoir.

A request for information on circulating strains was made on 28th February at the height of the outbreak, within 10 days our pilot research study was expedited, and the initial analysis completed. The fact that the 2018 outbreak was fueled by the circulating LASV diversity and not by transmission of a new or divergent lineage was already evident from the first seven genomes generated by 10th March (Fig. S1). This information was promptly communicated to the NCDC forming the basis of their report “Early Results of Lassa Virus Sequencing & Implications for Current Outbreak Response in Nigeria” released on March 12th 2018 (15). While this small sample was restricted to genotype II, the final collection of 36 LASV genome sequences generated on-site also included a representative of genotype III (Figs. 1 and S9), further supporting the spill-over of longstanding LASV diversity in the outbreak. The conclusions drawn from the first set of genome sequences immediately removed fears of extensive human-to-human transmission and allowed public health resources to be allocated appropriately. The response was focused on intensified community engagement on rodent control, environmental sanitation, and safe food storage. Further research is needed to evaluate if improved diagnostics and disease awareness and/or ecological and climate factors promoting transmission are the drivers behind the changing epidemiology of Lassa fever in Nigeria.

Portable metagenomic sequencing of genetically diverse RNA viruses on the MinION, direct from patient samples without the need to export material outside of the country of origin and with no pathogen-specific enrichment, is shown to be a feasible methodology enabling a real-time characterization of potential outbreaks in the field.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

L. E. Kafetzopoulou[1,2,3], S. T. Pullan[1,2], P. Lemey[4], M. A. Suchard[5], D. U. Ehichioya[3,6], M. Pahlmann[3,6], A. Thielebein[3,6], J. Hinzmann[3,6], L. Oestereich[3,6], D. M. Wozniak[3,6], K. Efthymiadis[7], D. Schachten[3], F. Koenig[3], J. Matjeschk[3], S. Lorenzen[3], S. Lumley[1], Y. Ighodalo[8], D. I. Adomeh[8], T. Olokor[8], E. Omomoh[8], R. Omiunu[8], J. Agbukor[8], B. Ebo[8], J. Aiyepada[8], P. Ebhodaghe[8], B. Osiemi[8], S.

Ehikhametalor[8], P. Akhilomen[8], M. Airende[8], R. Esumeh[8], E. Muoebonam[8], R. Giwa[8], A. Ekanem[8], G. Igenegbale[8], G. Odigie[8], G. Okonofua[8], R. Enigbe[8], J. Oyakhilome[8], E. O. Yerumoh[8], I. Odia[8], C. Aire[8], M. Okonofua[8], R. Atafo[8], E. Tobin[8], D. Asogun[8,9], N. Akpede[8], P. O. Okokhere[8,9], M. O. Rafiu[8], K. O. Iraoyah[8], C. O. Irolagbe[8], P. Akhideno[8], C. Erameh[8], G. Akpede[8,9], E. Isibor[8], D. Naidoo[10], R. Hewson[1,2], J. A. Hiscox[2,11,12], R. Vipond[1,2], M. W. Carroll[1,2], C. Ihekweazu[13], P. Formenty[10], S. Okogbenin[8,9], E. Ogbaini-Emovon[8,†], S. Günther[3,6,*,†], S. Duraffour[3,6,†]

Affiliations

[1]Public Health England, National Infections Service, Porton Down, UK. [2]National Institute of Health Research (NIHR), Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, UK [3]Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany [4]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, Leuven, Belgium [5]Departments of Biomathematics, Biostatistics and Human Genetics, University of California, Los Angeles, USA. [6]German Center for Infection Research (DZIF), partner site Hamburg-Lübeck-Borstel-Riems, Germany [7]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium [8]Irrua Specialist Teaching Hospital, Irrua, Nigeria [9]Faculty of Clinical Sciences, College of Medicine, Ambrose Alli University, Ekpoma, Nigeria [10]World Health Organization, Geneva, Switzerland. [11]Singapore Immunology Network, Agency for Science, Technology and Research (A*STAR), Singapore. [12]Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, UK. [13]Nigeria Centre for Disease Control, Abuja, Nigeria.

ACKNOWLEDGMENTS

We thank the health authorities of Nigeria for their excellent cooperation during the outbreak response.

Funding: LK, STP, RH, RV, MWC, JAH acknowledge funding by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England. LEK has received travel expenses and accommodation from Oxford Nanopore to speak at conferences regarding this work. LEK has received some reagents free of charge from Oxford Nanopore in support of her PhD project. MWC has received reagents free of charge from Oxford Nanopore in support of previous projects not related to the work presented in this manuscript. LEK and MWC have not receive other financial compensation or hold shares. PL and MAS acknowledge funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 725422-ReservoirDOCS) and from the Wellcome Trust Collaborative Award, 206298/Z/17/Z. PL acknowledges support by the Special Research Fund, KU Leuven ('Bijzonder Onderzoeksfonds', KU Leuven, OT/14/115), and the Research Foundation -- Flanders ('Fonds voor Wetenschappelijk Onderzoek – Vlaanderen', G066215N, G0D5117N and G0B9317N). MAS acknowledges support under National Science Foundation grant DMS 1264153. This study was supported by the German Federal Ministry of Health through support of the WHO Collaborating Centre for Arboviruses and Hemorrhagic Fever Viruses at the Bernhard-Nocht-Institute for Tropical Medicine (agreements ZMV I 1–2517WHO005 and ZMV I 1–2517WHO010) and through the Global Health Protection Program (agreement number ZMVI1-2517-GHP-704), the German Federal Ministry for Economic Cooperation and Development through the Rapid Deployment Expert Group to Combat Threats (SEEG), the European Union's Horizon 2020 research and innovation program to SG (grant agreement no. 653316-EVAg), and the German Research Foundation (DFG) to SG and DUE (GU 883/4–1). DUE acknowledges fellowships from Alexander von Humboldt Foundation and Kirmser Foundation. The funders had no role in the design and interpretation of the data and preparation of the manuscript.

REFERENCES AND NOTES

1. Frame JD, Baldwin JM Jr, Gocke DJ, Troup JM, Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. Am. J. Trop. Med. Hyg 19, 670–676 (1970). [PubMed: 4246571]
2. Asogun DA et al., Molecular diagnostics for lassa fever at Irrua specialist teaching hospital, Nigeria: lessons learnt from two years of laboratory operation. PLoS Negl. Trop. Dis 6, e1839 (2012). [PubMed: 23029594]
3. WHO | Lassa Fever – Nigeria (2018) (available at http://www.who.int/csr/don/23-march-2018-lassa-fever-nigeria/en/).
4. Jain M, Olsen HE, Paten B, Akeson M, The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17, 239 (2016). [PubMed: 27887629]
5. Quick J et al., Real-time, portable genome sequencing for Ebola surveillance. Nature. 530, 228–232 (2016). [PubMed: 26840485]
6. Quick J et al., Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc 12, 1261–1276 (2017). [PubMed: 28538739]
7. Faria NR et al., Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 546, 406–410 (2017). [PubMed: 28538727]
8. Faria NR et al., Genomic and epidemiological monitoring of yellow fever virus transmission potential. Science. 361, 894–899 (2018) [PubMed: 30139911]
9. Andersen KG et al., Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. Cell. 162, 738–750 (2015). [PubMed: 26276630]
10. Greninger AL et al., Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med 7, 99 (2015). [PubMed: 26416663]
11. Kafetzopoulou LE et al., Assessment of Metagenomic MinION and Illumina sequencing as an approach for the recovery of whole genome sequences of chikungunya and dengue viruses directly from clinical samples. bioRxiv (2018), p. 355560.
12. Koren S et al., Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. Genome Res 27, 722–736 (2017). [PubMed: 28298431]
13. Kim D, Song L, Breitwieser FP, Salzberg SL, Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 26, 1721–1729 (2016). [PubMed: 27852649]
14. Whitmer SLM et al., New Lineage of Lassa Virus, Togo, 2016. Emerg. Infect. Dis 24, 599–602 (2018). [PubMed: 29460758]
15. Nigeria Centre for Disease Control, (available at https://ncdc.gov.ng/news/121/early-results-of-lassa-virus-sequencing-%26-implications-for-current-outbreak-response-in-nigeria).
16. Nikisins S et al., International external quality assessment study for molecular detection of Lassa virus. PLoS Negl. Trop. Dis 9, e0003793 (2015). [PubMed: 25996783]
17. Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv: 1303.3997v2 [q-bio.GN] (2013).
18. Loman NJ, Quick J, Simpson JT, A complete bacterial genome assembled de novo using only nanopore sequencing data (2015),, doi:10.1101/015552.
19. Penedos AR, Myers R, Hadef B, Aladin F, Brown KE, Assessment of the Utility of Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks. PLoS One. 10, e0143081 (2015). [PubMed: 26569100]
20. Edgar RC, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5, 113 (2004). [PubMed: 15318951]
21. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B, RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1, vev003 (2015). [PubMed: 27774277]
22. Stamatakis A, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312–1313 (2014). [PubMed: 24451623]
23. Rambaut A, Lam TT, Max Carvalho L, Pybus OG, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2, vew007 (2016). [PubMed: 27774300]

24. Trovão NS et al., Host ecology determines the dispersal patterns of a plant virus. Virus Evol 1, vev016 (2015). [PubMed: 27774287]

25. Suchard MA et al., Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4, vey016 (2018). [PubMed: 29942656]

26. Edo-Matas D et al., Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient hypothesis testing through hierarchical phylogenetic models. Mol. Biol. Evol 28, 1605–1616 (2011). [PubMed: 21135151]

27. Minin VN, Bloomquist EW, Suchard MA, Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol 25, 1459–1471 (2008). [PubMed: 18408232]

28. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A, Relaxed phylogenetics and dating with confidence. PLoS Biol 4, e88 (2006). [PubMed: 16683862]

29. Baele G, Lemey P, Rambaut A, Suchard MA, Adaptive MCMC in Bayesian phylogenetics: an application to analyzing partitioned data in BEAST. Bioinformatics. 33, 1798–1805 (2017). [PubMed: 28200071]

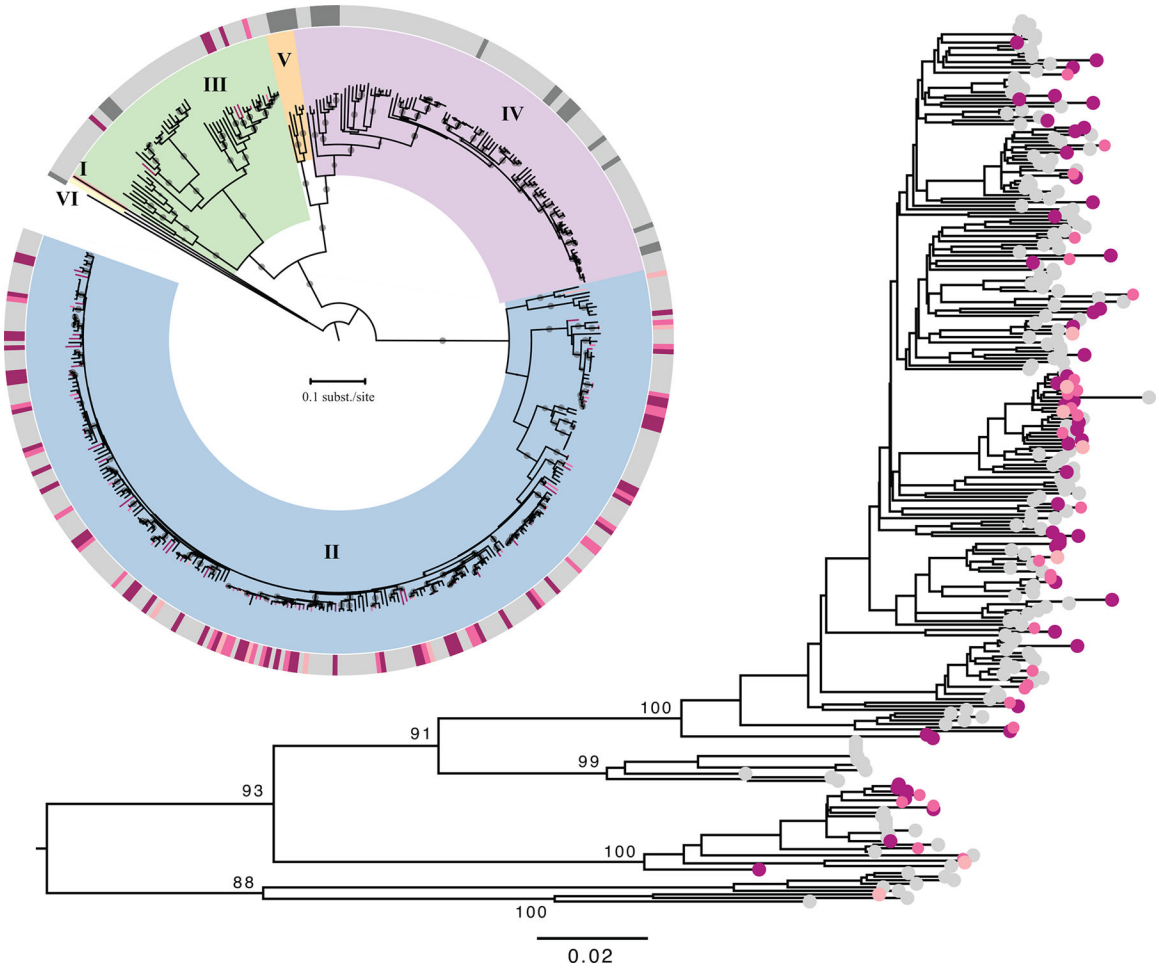

Figure 1. Phylogenetic reconstruction of the S segment data.

The circular tree includes 96 sequences from 2012 to 2017, 88 sequences from 2018 and sequences available from GenBank. The rectangular tree focuses on the genotype II clade (in blue in the circular tree) that includes most of the 2018 sequences. The six genotypes are indicated with different colors and roman symbols. Bootstrap support >90% is indicated with a small grey circle at the middle of their respective branches. The color strip highlights the human LASV sequences obtained from previous years (light grey), sequences obtained from rodent samples (dark grey) and 2018 sequences as light pink for the first seven sequences generated in Nigeria, magenta for the remaining 28 sequences analysed on-site, and purple for the remaining finalised in Europe. The same color code is used in the genotype II rectangular tree. Bootstrap values >80% are shown for the major genotype II lineages.

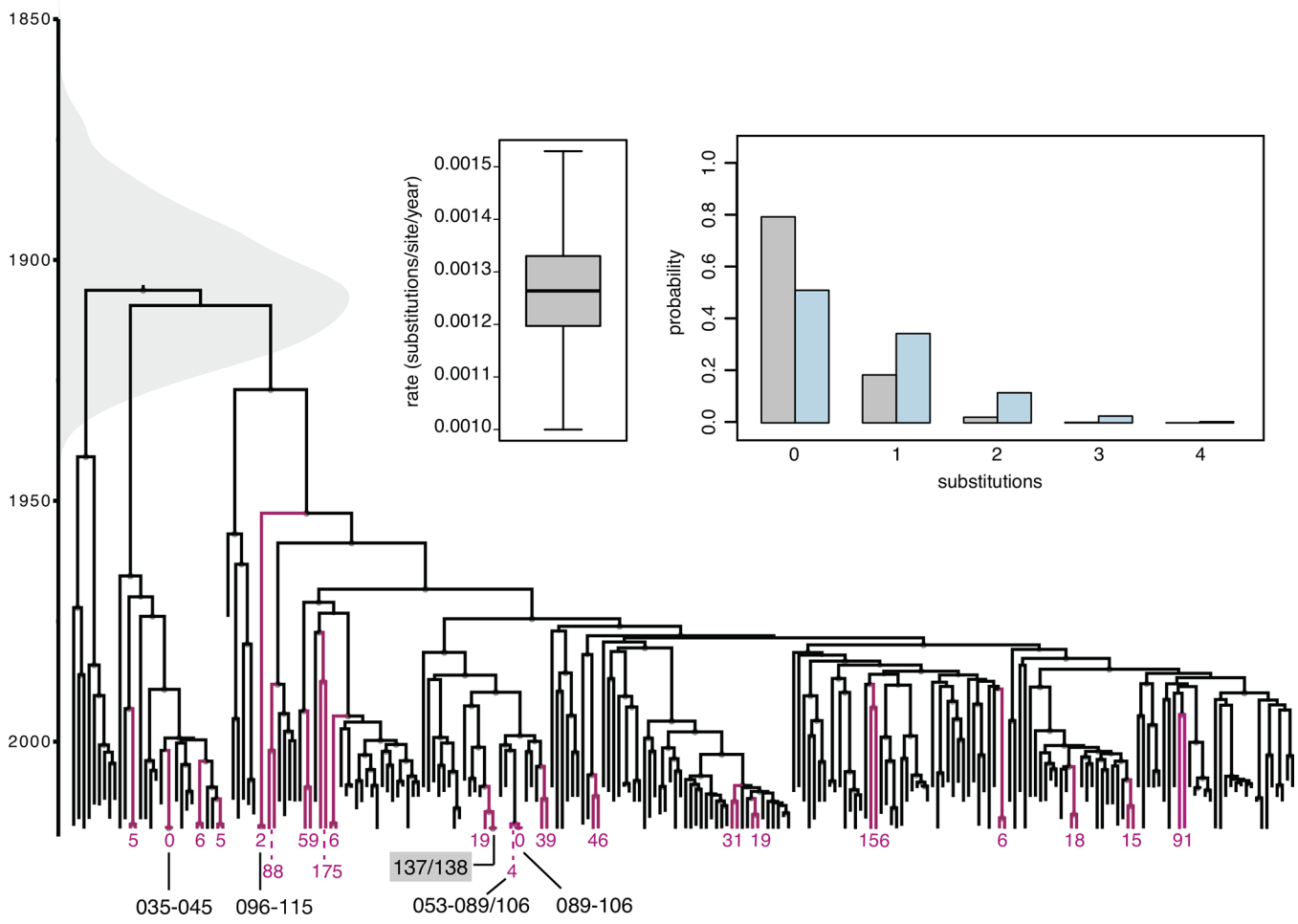

Figure 2. Assessing the potential for direct linkage between pairs of 2018 sequences in the S segment.

The maximum clade credibility tree summarizes a Bayesian evolutionary inference for the genotype II sequences in the S segment. A time scale and a marginal posterior distribution for the time to the most recent common ancestor are shown to the left. The size of the internal node circles reflects posterior probability support values. 2018 sequences clustering as pairs are indicated in purple; the number of substitutions between them is indicated at their respective tips. A posterior estimate of the evolutionary rate and probability distributions for observing a given number of substitutions during a human-tohuman transmission event are shown as insets. The distribution represented by grey bars is based on the mean evolutionary rate estimate and a mean estimate for the generation time whereas the light blue distribution is based on upper estimates and also incorporates an upper estimate for the MinION sequencing error (Supplementary Methods). At the bottom, clusters of sequences for which human-to-human transmission cannot be excluded according to the upper estimates of generation time are indicated. A pair of identical sequences (137–138) that was retrospectively found to be derived from the same patient is marked with a grey box. One pair (096–115) was still disregarded as potential transmission chain due to 21

differences in L segment (Fig. S9). The temporal signal prior to BEAST inference was explored in Fig. S10.