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

Kafetzopoulou LE1,2,3, Pullan ST1,2, Lemey P4, Suchard MA5, Ehichioya DU3,6, Pahlmann M3,6, Thielebein A3,6, Hinzmann J3,6, Oestereich L3,6, Wozniak DM3,6, Efthymiadis K7, Schachten D3, Koenig F3, Matjeschk J3, Lorenzen S3, Lumley S1, Ighodalo Y8, Adomeh DI8, Olokor T8, Omomoh E8, Omiunu R8, Agbukor J8, Ebo B8, Aiyepada J8, Ebhodaghe P8, Osiemi B8, Ehikhametalor S8, Akhilomen P8, Airende M8, Esumeh R8, Muoebonam E8, Giwa R8, Ekanem A8, Igenegbale G8, Odigie G8, Okonofua G8, Enigbe R8, Oyakhilome J8, Yerumoh EO8, Odia I8, Aire C8, Okonofua M8, Atafo R8, Tobin E8, Asogun D8,9, Akpede N8, Okokhere PO8,9, Rafiu MO8, Iraoyah KO8, Iruolagbe CO8, Akhideno P8, Erameh C8, Akpede G8,9, Isibor E8, Naidoo D10, Hewson R1,2,11,12, Hiscox JA2,13,14, Vipond R1,2, Carroll MW1,2, Ihekweazu C15, Formenty P10, Okogbenin S8,9, Ogbaini-Emovon E#8, Günther S#16,6, Duraffour S#3,6.

**Affiliations:**

1. Public Health England, National Infection Service, Porton Down, UK.

2. National Institute of Health Research (NIHR), Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, 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, CA, USA.

6. German Centre for Infection Research (DZIF), partner site Hamburg, 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. Faculty of Infectious and Tropical Diseases, Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, UK.

12. Faculty of Clinical Sciences and International Public Health, Liverpool School of Tropical Medicine, Liverpool, UK.

13. Singapore Immunology Network, Agency for Science, Technology and Research (A\*STAR), Singapore.

14. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK.

15. Nigeria Centre for Disease Control, Abuja, Nigeria.

16. Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. guenther@bni.unihamburg.de.

# Contributed equally

**Abstract:** The 2018 Nigerian Lassa fever season saw the largest 5 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 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 rats 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 25 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 outbreak - has been the largest on record, reaching 1,495 suspect and 376 confirmed cases, spread over 18 states, by the 18th March. This notably exceed the total 269 confirmed cases reported over the three previous years combined (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 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, Oxford, UK). This instigated a major upscale in sequencing efforts, leading to sequencing of 120 samples.

Nanopore sequencing is an emerging technology with significant potential, in particular the MinION, a small and robust sequencing device perfectly 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, for which even PCR-based laboratory diagnosis poses a significant challenge, due to inter-strain nucleic acid sequence variation of up to 32% and 25% for the L and S segments respectively (9). Designing targeted whole-genome sequencing approaches, such as PCR amplicons or bait/capture probes, without prior knowledge of the targeted LASV lineage is therefore impractical. 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 Ct value and location from the 334 cases reported by ISTH between 1st January and 16th March 2018. To reflect the geographic case distribution of the outbreak, the majority of samples originated from Edo state followed by Ondo and Ebonyi. Samples selected covered the wide range of clinical viral loads observed, including several samples testing negative in one of the two qRT-PCR assays used (Figure S1). Up to six samples were run in multiplex per MinION flow cell, along with a negative control. In order to produce high confidence consensus sequences for phylogenetic inference we replicated the reference mapping/variant calling approach using Nanopolish variants, as developed for the West African Ebola virus disease outbreak (12, 13). Due 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 with a maximum of 42.9% (on average: 4.26%) allowing for sufficient genomic sequence (>70%) for phylogenetic comparison of at least one segment in 90 of the samples tested (detailed in Supplementary methods, Figures S1-4). Metagenomic classification using Centrifuge (14) identified 0.10% of reads from sample ISTH\_18\_110 as originating from Hepatitis A virus; providing 74% genome coverage at 20x 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 of RNA viruses, including those present as co-infections. In all other samples tested, LASV was the sole pathogen identified.

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 released to assist the outbreak response and publicly available sequences. We use this as a frame of reference to document how the genomic data generated in real-time 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 (Figure 2). This phylogenetic pattern is mimicked by the L segment reconstruction (Figure S5). This is not surprising since only 7 viruses in the entire complete genome data set were identified as clustering significantly differently in the L and S segment (cfr. 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.

The fact that the 2018 outbreak was fuelled by the circulating LASV diversity and not by transmission of a new or divergent LASV lineage was already evident from the first 7 genomes (generated on 10th March, Figure 1). 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 (Figure 2), further supporting the evidence of spill-over from the extensive LASV diversity in rodents. However, the collection of 36 samples also revealed sequences clustering together as a pair in the phylogeny (n = 6 pairs) suggesting some degree of epidemiological linkage between those 5 and possibly even human-to-human transmission. The pair with the highest sequence similarity (137 and 138, Figure S6) differed by only a single substitution in the S segment. However, retrospective tracing of the origin of the respective samples for these sequences revealed that they were in fact drawn from the same patient. In order to assess the possibility of human-to-human transmission for the remaining pairs, as well as for all other sequences from the 2018 outbreak, we conducted a Bayesian inference of time-measured trees to estimate the time to the most recent common ancestor (TMRCA) of each sequence with its most closely related sequence in the S segment data set (Figure 3). Our analysis estimates an S segment substitution rate (1.06x10-3

subst./site/year, 95% credible interval = [8.01x10-4,1.32x10-3]) similar to previous estimates (9) and a genotype II TMRCA of 236 [160,325] years. For three pairs of sequences, we estimate an average TMRCA close to one year (Figure 3), but with credible intervals that do not necessarily exclude a recent date that could be compatible human-to-human transmission. The number of substitutions between those pairs in the S segment range from 5 to 7. All other 2018 sequences were estimated to share a common ancestor with another sequence too distantly in the past to implicate possible human-to-human transmission. Similar results are obtained for the time measured analysis of the L segment data (Figure S7). Sequences generated by Illumina resequencing of 14 SISPA preparations consistently clustered with their Nanopore counterpart with little to no divergence between them, confirming the accuracy of the Nanopore approach (Figure S10).

A request for information on circulating strains was made on the 28th February at the height of the outbreak, within 10 days our pilot research study was expedited and the initial analysis completed. The conclusions drawn from that first set of genome sequences immediately removed fears of the emergence of a novel strain and allowed public health resources to be allocated appropriately. Further research is needed to evaluate if better surveillance and/or ecological factors promoting transmission could be responsible for the unprecedented rise in Lassa fever cases observed. Portable metagenomic sequencing of genetically diverse RNA viruses on the MinION, direct from patient sample, with no pathogen specific enrichment, is shown here for the first time to be a feasible methodology enabling a real-time characterization of outbreaks in the field.

References:

1. J. D. Frame, J. M. Baldwin Jr, D. J. Gocke, J. M. Troup, 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).

2. D. A. Asogun 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).

3. WHO | Lassa Fever – Nigeria (2018) (available at [http://www.who.int/csr/don/23-march- 2018-lassa-fever-nigeria/en/](http://www.who.int/csr/don/23-march-%202018-lassa-fever-nigeria/en/)).

4. M. Jain, H. E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 17, 239 (2016).

5. J. Quick et al., Real-time, portable genome sequencing for Ebola surveillance. Nature. 530, 228–232 (2016).

6. J. Quick et al., Multiplex PCR method for MinION and Illumina sequencing of Zika and 15 other virus genomes directly from clinical samples. Nat. Protoc. 12, 1261–1276 (2017).

7. N. R. Faria et al., Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 546, 406–410 (2017).

8. N. R. Faria et al., Genomic and epidemiological monitoring of yellow fever virus transmission potential. bioRxiv (2018), p. 299842.

9. K. G. Andersen et al., Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. Cell. 162, 738–750 (2015).

10. A. L. Greninger et al., Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7, 99 (2015).

11. L. E. Kafetzopoulou 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 (2018), , doi:10.1101/355560.

12. S. Koren et al., Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. Genome Res. 27, 722–736 (2017).

13. J. Quick et al., Real-time, portable genome sequencing for Ebola surveillance. Nature. 530, 228–232 (2016).

14. D. Kim, L. Song, F. P. Breitwieser, S. L. Salzberg, Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26, 1721–1729 (2016).

15. Nigeria Centre for Disease Control, (available at https://ncdc.gov.ng/news/121/early-resultsSubmitted of-lassa-virus-sequencing-%26-implications-for-current-outbreak-response-in nigeria).

16. S. Nikisins et al., International external quality assessment study for molecular detection of Lassa virus. PLoS Negl. Trop. Dis. 9, e0003793 (2015).

17. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013), (available 5 at http://arxiv.org/abs/1303.3997).

18. N. J. Loman, J. Quick, J. T. Simpson, A complete bacterial genome assembled de novo using only nanopore sequencing data (2015), , doi:10.1101/015552.

19. A. R. Penedos, R. Myers, B. Hadef, F. Aladin, K. E. Brown, Assessment of the Utility of Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks. PLoS One. 10, e0143081 (2015).

20. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5, 113 (2004).

21. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003 (2015).

22. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312–1313 (2014).

23. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2, vew007 (2016).

24. N. S. Trovão et al., Host ecology determines the dispersal patterns of a plant virus. Virus Evol. 1, vev016 (2015).

25. M. A. Suchard et al., Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, vey016 (2018).

26. D. Edo-Matas 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).

27. M. S. Gill et al., Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724 (2013).

28. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88 (2006).

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