Co-infection of mammarenaviruses in a wild mouse, Tanzania

Laura N. Cuypers1,7 Dagmar Čízková,2,4 and Joëlle Goüy de Bellocq2,4

1Evolutionary Ecology Group, Department of Biology, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium and 2Institute of Vertebrate Biology of the Czech Academy of Sciences, Kvetna 8, 603 65 Brno, Czech Republic

Abstract

Mammarenaviruses are bi-segmented RNA viruses. They encompass viruses responsible for several severe diseases in humans. While performing a de novo assembly of a new virus found in a wild single-striped grass mouse in Tanzania, we found a single S but two divergent L segments. Natural co-infections, common within reptarenaviruses in captivity, were never reported for mammarenaviruses and never in a wild sample. This finding can have implications for virus evolution as co-infection could trigger viral recombination/reassortment in natural reservoirs.

Key words: co-infection; Mafiga mammarenavirus; Lemniscomys rosalia; zoonotic infections; recombination; BlobTools.

Some members of the arenavirus family are among the leading etiological agents of human hemorrhagic fever. This family encompasses four genera: antennaviruses infect fish, hartmanniviruses and leptorenaviruses infect snakes and mammarenaviruses infect mammals and rodents in particular. Nine mammarenaviruses, including Lassa, are known to spill over to humans and can cause mild symptoms of fatal disease, while reptarenaviruses, including Lassa, are known to spill over to humans and can cause boid inclusion body disease in captive snakes. In Reptarenavirus and Mammarenavirus genera, the RNA genome is composed of an S (∼3.5 kb) and an L (∼7 kb) segment that each encode two proteins in ambisense orientation: a glycoprotein precursor (GPC) and a nucleoprotein (NP) for the S and a Z protein and an RNA-dependent RNA polymerase for the L segment (Radoshitzky et al. 2019).

In segmented viruses, three evolutionary mechanisms generate genetic diversity: mutation, recombination, and reassortment. The last two mechanisms require the occurrence of co-infecting viruses within the same host cell and have important implications in the evolution of viruses because they may influence the virus capacity to jump hosts and evade the immune system (Vijaykrishna, Mukerji, and Smith 2015). The two mechanisms have been observed for mammarenaviruses in laboratory conditions (Fernandes et al. 2018). Moreover, evidence for Lassa virus reassortment in human clinical samples has been reported in West Africa (Andersen et al. 2015). Recently, reptarenavirus studies showed that captive snakes were prone to co-infections with multiple S and L segments, with an unbalanced ratio in favor of the L segment and with ongoing recombination and reassortment events (Hepojoki et al. 2015; Stenglein et al. 2015). This unique phenomenon may have been induced by captivity, but in a recent preprint, two reptarenavirus L segments but only one S segment were detected in a wild Boa constrictor (Alfaro-Alarcón et al. 2022). How often co-infections and to what extent recombination/reassortment are happening in natural conditions is unknown.

Here, we report evidence of a mammarenavirus co-infection in a wild single-striped grass mouse, Lemniscomys rosalia, in Tanzania. The first detection of a mammarenavirus in this mouse species referred to as Lemniscomys arenavirus F4-8/TZA/2008 (Goüy de Bellocq et al. 2010), motivated the molecular screening of twelve L. rosalia individuals captured in 2016 for mammarenavirus RNA presence. This work complemented a recent screening of many Mastomys natalensis individuals to describe their arenavirus genetic diversity in Tanzania (Cuypers et al. 2020). Following the methods of Cuypers et al. (2020), we found one mammarenavirus-positive sample (individual T229841), and Sanger sequenced a 340 and 540 nt-long sequence of the L and NP genes, respectively. BLASTn analysis confirmed that the L sequence was 82 per cent identical to Lemniscomys arenavirus F4-8/TZA/2008. The sample was subject to metagenomic sequencing to determine its complete viral genome sequences. A small piece of kidney preserved in RNAlater was extracted with RNeasy Mini Kit (Qiagen, Hilden, Germany). Ribosomal RNA (rRNA) was depleted with
RiboCop rRNA Depletion Kit V1.3 (Lexogen, Vienna, Austria), and a library with dual indexes prepared with Swift RNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA). The library was sequenced with 150 paired-end (PE) reads on Illumina HiSeq X at Novogene (UK) together with twenty-three other samples from a wider RNA virome study. This batch included RNA samples from a bat, a rabbit, and different rodent species (Mus, Mastomys, Stenocephalemys, Praomys, Saccostomus, and Clethrionomys) to attempt sequencing of mammarenaviruses, orthohantaviruses, and lagoviruses. After read de-multiplexing, we obtained a total of 11,667,572 PE raw reads (deposited under European Nucleotide Archive project no. PRJEB52502). After quality filtering and trimming with Skewer (Jiang et al. 2014), 11,652,508 PE reads remained.

Metagenomic Illumina reads were assembled into a complete viral genome using two approaches: (1) We used Geneious Read mapper with the low sensitivity option in Geneious 11.1.5 (Biomatters, Auckland, New Zealand) to perform iterative mapping with 15 per cent of the trimmed read data and the L and NP Sanger-sequenced fragments as starting seeds. A 3,364 nt-long S segment and a 7,252 nt-long L segment were assembled after twenty-five and thirty-five iterations, respectively, with a 75 per cent consensus sequence threshold. (2) We performed a de novo assembly using metaSPAdes (Nurk et al. 2017). Assembled contigs were classified using BlobTools 1.1.1 (Laetsch and Blaxter 2017) after similarity searches of the contigs against NCBI and UniProt databases. Among thirty-six contigs larger than 250 nt identified as viruses, metaSPAdes assembly retrieved three large and highly (k-mer) covered contigs of 3,401, 7,389, and 7,315 nt (Fig. 1), the other thirty-three contigs being weakly covered and much shorter (< 608 nt). These three contigs were trimmed to conserved terminal nucleotides common to mammarenaviruses in Geneious to 3,364 (S segment, k-mer coverage 1,499), 7,254 (L1 segment; k-mer coverage 178), and 7,252 nt (L2 segment; k-mer coverage 555). S and L2 segments were identical to the segments retrieved with the first approach, while segment L1 showed 87 per cent identity with segment L2. Upon reexamination of the iterative mapping of the first approach, we found that those L1 reads that met the low sensitivity criterion (allowing 10 per cent divergence to the reference) were mapped, but the L1-specific variants were not included in the final consensus sequence. This was most likely because the consensus threshold for each iteration was set at 75 per cent. The mapped L1 variants were below this threshold in the initial extension from the L2 seed and once the iteratively extending consensus sequence was biased toward L2, it preferentially attracted L2 reads in the next iterations, exacerbating the omission of L1 variants.

**Figure 1.** BlobPlot of sample TZ29841 by BlobTools 1.1.1 (Nurk et al. 2017). BlobTools collects the information from hit files resulting from sequence similarity searches of the assembly against NCBI and UniProt sequence databases using BLASTn and Diamond (Bao, Chetvernin, Tatusova 2014) and sums up bit-scores by the taxonomic group at each taxonomic rank. The taxonomic groups in the legend are followed by their contig number, total span, and N50. Sequences are represented by circles (blobs) with diameters proportional to sequence length and colored by superkingdom taxonomic affiliation: y-axis: k-mer coverage in log scale as provided by metaSPAdes; x-axis: GC contents. The three larger red blobs correspond to the two L and one S mammarenavirus segments with 0.42 GC proportion for both Ls and 0.45 for S. The larger blue blob at 56% GC with coverage above 10 is a 10,359 nt-long mitochondrial contig showing 97% identity with L. rosalia (GenBank accession no. NC_053799) in BLAST, confirming the field identification of the host. The assembly of the mitochondrial genome was completed using the Geneious Read mapper, and the full genome has been deposited in GenBank (Accession no. ON380917). Other examples of blobplots for mammarenaviruses assemblies are provided in S1 for comparative purposes.
Figure 2. L, NP, and GPC gene phylogenetic trees. The trees were inferred with MrBayes v3.2.7a with codon partitioning and a GTR+G nucleotide substitution model. In two independent runs, four chains ran for 10,000,000 generations with a burn-in of 25%. Average standard deviations of split frequencies were examined for convergence, as were parameter effective sample sizes and trace patterns in Tracer (https://beast.community/tracer). Trees were visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Asterisks indicate sequences that were slightly modified from the corresponding GenBank sequence to improve translation alignment (see Supplementary Materials S4–S6). The scale bars indicate 0.5 nt substitutions per site. Numbers at the nodes represent the posterior probability.
Competitive mapping using de novo assembled L1, L2, and S segments as references resulted in read coverages of 1,023 ± 393 (SD) for L1, 3,311 ± 1,317 (SD) for L2, and 9,059 ± 3,538 (SD) for the S segment (after deduplication in Geneious). Higher coverage of the S relatively to the L segment seems to be relatively common in mammarenaviruses (e.g. Lassa (Andersen et al. 2015), but not always true (Těšíková et al. 2021; see Supplementary Material S1). Each segment showed a typical mammarenavirus genomic organization with two ambisense open reading frames separated by noncoding intergenic regions (GenBank accession nos. ON381477–ON381479).

To check the phylogenetic relationships of the L. rosalia-borne virus, MrBayes (Ronquist et al. 2012) phylogenetic trees were produced for the L, GPC, and NP genes with Old World mammarenavirus representatives in the CIPRES Science Gateway V3.3 (Miller, Pfeiffer, and Swartz 2010). The L sequences first cluster together with a posterior probability (PP) of 0.85, then with the partial sequence of arenavirus Lemniscomys/F4-8/TZA/2008 found 162 km further (PP: 1, Fig. 2). The new sequences consistently form a clade with Solwezi, Ippy, and Kitale virus (PP: 1) found in Grammomys sp., Arvicanthis sp., and Grammomys macmillani, respectively (all Arvicanthini tribe members, including L. rosalia). Its closest relative in the L and NP tree and in a Pairswise Sequence Comparison (PASC) analysis (https://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi; Bao, Chetvernin, Tatusova 2014) was Solwezi virus (S segment: 70 per cent, L segments: 61 per cent, see Supplementary Material S2 for PASC figures). Following the recommendations of the International Committee on Taxonomy of Viruses, these divergence percentages confirm that the L. rosalia-borne mammarenavirus represents a new virus species, which we name Mafiga (MAFV) on the basis of the location where the virus was originally found in 2008 (Gouy de Bellocq et al. 2010).

The two L sequences appear to be the result of a co-infection and not of a laboratory contamination as (1) the two L segments cluster together before any other known Old World mammarenavirus; (2) a single L. rosalia sample and no other rodent from the Arvicanthini tribe were included in the library preparation batch; and (3) the library preparation was performed in a different laboratory and country from the one that screened the TZ29841 and F4-8 L. rosalia samples (which were the only L. rosalia samples to test positive for mammarenavirus RNA). Finally, we reexamined the initial short L Sanger sequence to check if double peaks could be distinguished on the chromatogram that might have been overlooked. We did not detect any, which could be explained by the coverage difference and a preferential annealing of the primers for the L2 over the L1 segment, which could be expected as four out of six primers used for the screening have higher sequence similarity to the L2 segment in the primer interval. However, the timing of the co-infection is unknown. As we were only able to detect a single S segment with few, low-frequency SNPs, we likely did not sample during an active co-infection. The co-infection could have taken place before sampling or could even have taken place in a different mouse that transmitted the three segments to the mouse we sampled (Stenglein et al. 2015).

**Supplementary data**

Supplementary data are available at Virus Evolution online.

**Acknowledgements**

This collaborative work was also part of the Czech Republic-Flanders, Belgium Mobility projects 2020 (Research Foundation – Flanders no. VS07521N and CAS no. FWO-21-02). Permission to collect specimens and handle animals was provided by the Sokoline University of Agriculture with the assistance of the Pest Management Centre, Morogoro, Tanzania. All animal-related procedures were carried out in accordance with the International Guiding Principles for Biomedical Research Involving Animals. We thank Ondřej Mikula, Adam Konečný, Alexandra Hánová, and Abdul A. S. Katakweba for field sampling and J Bryja for access to the sample. Computational resources were supplied by the project ‘e-Infrastruktura CZ’ (e-INFRA CZ LM20181401) supported by the Ministry of Education, Youth and Sports of the Czech Republic. L.N.C. was a Ph.D fellow with the FWO (1171921N).

**Funding**

This research was funded by the Czech Science Foundation (GAČR grant no. 18-19629S).

**Conflict of interest**: None declared.

**Data availability**

Raw metagenomic data are available under ENA project no. PRJEBS2502. Assembled mammarenavirus segment and Lemniscomys rosalia mitochondrial sequences are available under GenBank accession nos. ON381477–ON381479 and ON380917, respectively. Other data are available in the Supplementary Material.

**References**

Alfaro-Alarcón, A. et al. (2022) ‘Boid Inclusion Body Disease (BIBD) Is also a Disease of Wild Boa Constrictors’, BioRxiv [Preprint] <https://doi.org/10.1101/2022.04.25.489483> accessed 1 Jun 2022.

Andersen, K. G., et al. (2015) ‘Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus’, Cell, 162: 738–50.

Bao, Y., Chetvernin, V., and Tatusova, T. (2014) ‘Improvements to Pairwise Sequence Comparison (PASC): A Genome-Based Web Tool for Virus Classification’, Archives of Virology, 159: 3293–304.

Buchfink, B., Xie, C., and Huson, D. H. (2015) ‘Fast and Sensitive Protein Alignment Using DIAMOND’, Nature Methods, 12: 59–60.

Cuypers, L. N. et al. (2020) ‘Three Arenaviruses in Three Sub-specific Natal Multimammate Mouse Taxa in Tanzania: Same Host Specificity. But Different Spatial Genetic Structure?’, Virus Evolution, 6: veaa039.

**Conclusions**

Mafiga is a new virus and the first report of co-infecting mammarenavirus segments in a wild host. The detection of the two L segments was dependent on the genomic characterization method, so we recommend de novo approaches over mapping methods using references or starting seed sequences that will miss very divergent or low-frequency co-infected segment(s). Alternatively, complex stepwise pipelines can also be efficient at recovering multiple segments (Stenglein et al. 2015). Although higher levels of co-infection have been described in reptarenaviruses, with up to four S and eleven L segments in a single captive snake (Stenglein et al. 2015), our results show that this phenomenon occurs in natural conditions. As co-infection is the prerequisite for recombination and reassortment, documenting such cases in nature can help to better understand the evolution of arenaviruses and their capacity for spillover or host jumps.
Fernandes, J. et al. (2018) ‘Xapuri Virus, a Novel Mammarenavirus: Natural Reassortment and Increased Diversity between New World Viruses’, Emerging Microbes & Infections, 7: 120.

Goüy de Bellocq, J. et al. (2010) ‘Sympatric Occurrence of 3 Arenaviruses, Tanzania’, Emerging Infectious Diseases, 16: 692–5.

Hepojoki, J. et al. (2015) ‘Arenavirus Coinfections Are Common in Snakes with Boid Inclusion Body Disease’, Journal of Virology, 89: 8657–60.

Jiang, H. et al. (2014) ‘Skewer: A Fast and Accurate Adapter Trimmer for Next-generation Sequencing Paired-End Reads’, BMC Bioinformatics, 15: 182.

Laetsch, D., and Blaxter, M. (2017) ‘BlobTools: Interrogation of Genome Assemblies’, F1000Research, 6: 1287.

Lukashevich, I. S. (1992) ‘Generation of Reassortants between African Arenaviruses’, Virology, 188: 600–5.

Miller, M. A., Pfeiffer, W., and Swartz, T. (2010) ‘Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees’, in Proceedings of the Gateway Computing Environments Workshop (GCE), Nov 14; New Orleans, LA, pp. 1–8.

Nurk, S. et al. (2017) ‘metaSPAdes: A New Versatile Metagenomic Assembler’, Genome Research, 27: 824–34.

Radoshitzky, S. R. et al. (2019) ‘ICTV Virus Taxonomy Profile: Arenaviridae’, The Journal of General Virology, 100: 1200–1.

Ronquist, F. et al. (2012) ‘MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space’, Systematic Biology, 61: 539–42.

Stenglein, M. D. et al. (2015) ‘Widespread Recombination, Reassortment, and Transmission of Unbalanced Compound Viral Genotypes in Natural Arenavirus Infections’, PLoS Pathogens, 11: e1004900.

Těšíková, J., Krásová, J., and Goüy de Bellocq, J. (2021) ‘Multiple Mammarenaviruses Circulating in Angolan Rodents’, Viruses, 13: 982.

Vijaykrishna, D., Mukerji, R., and Smith, G. J. D. (2015) ‘RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion’, PLoS Pathogens, 11: e1004902.