

Three arenaviruses in three subspecific natal multimammate mouse taxa in Tanzania: same host specificity, but different spatial genetic structure?

Laura N. Cuypers,^{1,*}† Stuart J. E. Baird,² Alexandra Hánová,^{2,3} Tatjana Locus,¹ Abdul S. Katakweba,⁴ Sophie Gryseels,^{1,5,6,‡} Josef Bryja,^{2,3,§} Herwig Leirs,^{1,**} and Joëlle Goüy de Bellocq²

¹Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp, Belgium, ²Institute of Vertebrate Biology of the Czech Academy of Sciences, Brno, Czech Republic, ³Department of Botany and Zoology, Masaryk University, Brno, Czech Republic, ⁴Pest Management Centre, Sokoine University of Agriculture, Morogoro, Tanzania, ⁵Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA and ⁶Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium

*Corresponding author: E-mail: lauracuypers@outlook.be

†<https://orcid.org/0000-0003-4325-1304>

‡<https://orcid.org/0000-0002-4851-9384>

§<https://orcid.org/0000-0003-0516-7742>

**<https://orcid.org/0000-0002-7612-5024>

Abstract

Mastomys natalensis is widespread in sub-Saharan Africa and hosts several arenavirus species, including the pathogenic zoonotic Lassa virus in West Africa. Mitochondrial lineages sub-divide the range of *M. natalensis* and have been associated with cryptic structure within the species. To test specificity of arenaviruses to hosts carrying these lineages, we screened 1772 *M. natalensis* in a large area of Tanzania where three mitochondrial lineages meet. We detected fifty-two individuals that were positive for one of three arenaviruses: Gairo, Morogoro, and Luna virus. This is the first record of Luna virus in Tanzania. We confirmed the specificity of each arenavirus to a distinct host mitochondrial lineage except for three cases in one locality at the centre of a host hybrid zone. No arenaviruses were detected in a large part of the study area. Morogoro and Gairo virus showed differences in prevalence (Morogoro virus lower than Gairo virus) and in genetic structure (Morogoro virus more structured than Gairo virus). However, both viruses have genetic neighbourhood size estimates of the same order of magnitude as Lassa virus. While differences in arenavirus and/or host evolutionary and ecological dynamics may exist, Tanzanian arenaviruses could be suited to model Lassa virus dynamics in *M. natalensis*.

Key words: arenavirus host specificity; regional prevalence differences; spatial genetic structure; *Gairo mammarenavirus*; *Morogoro mammarenavirus*; *Luna mammarenavirus*.

1. Introduction

Viruses (and other pathogens and parasites) sometimes occur only in a restricted part of their host's geographic range. Several factors can play a role: a virus may be absent where the host density is below a persistence threshold (Patot et al. 2010), a vector involved in the transmission is absent (Cybinski, George, and Paull 1978), and/or environmental differences reduce viral survival outside the host (Linard et al. 2007). However, genetic structure of the host can also impact viral distribution (Saxenhofer et al. 2019). The impact of cryptic host structure is rarely assessed, as its detection requires multilocus genetics: the host of a virus may be a cryptic taxon below the species level (here referred to as subspecific taxon), rather than the species as a whole. One system where this has been hypothesized is the *Mastomys natalensis*-borne arenaviruses (Gryseels et al. 2017).

Mastomys natalensis (Smith, 1834), the Natal multimammate mouse, is one of the most widespread rodents in sub-Saharan Africa. Across its range, six divergent mitochondrial lineages with parapatric distribution can be distinguished: A-I in West Africa; A-II in Central Africa; A-III, B-IV and B-V in East Africa; and B-VI in South Africa (see Figure 1 in Colangelo et al. 2013). Different *M. natalensis*-borne arenaviruses were found in the ranges of these different mitochondrial lineages: Lassa virus in the A-I (Olayemi et al. 2016), a Mobala-like virus in the A-II (Olayemi et al. 2016), Gairo virus in the B-IV (Gryseels et al. 2015, 2017), Morogoro virus in the B-V (Günther et al. 2009; Gryseels et al. 2017), and both Mopeia (Wulff et al. 1977) and Luna virus (Ishii et al. 2011) in the B-VI range. These mitochondrial lineages may correspond to taxa that are distinct genome-wide and that meet in secondary contact zones, as multilocus genetics has shown for the B-IV and B-V mitochondrial lineages/taxa (Gryseels et al. 2017). In the [B-IV, B-V] contact zone in Tanzania, Gairo and Morogoro virus only co-occur at the centre of the narrow hybrid zone, otherwise they are only detected in localities dominated by individuals with either B-IV or B-V mitochondria, respectively (Gryseels et al. 2017). As the centre of the change in mitochondrial lineage over the hybrid zone coincides with the centre of the change in nuclear markers (Gryseels et al. 2017), B-IV mitochondrion dominated localities are on the B-IV taxon side of the hybrid zone and B-V mitochondrion dominated those on the B-V side. It thus appears that neither arenavirus has spread into the range of the other taxon despite hybridization of the taxa and in the absence of a clear environmental barrier (Gryseels et al. 2017).

Lassa virus, the arenavirus carried by *M. natalensis* in West Africa can cause haemorrhagic fever and is estimated to infect hundreds of thousands and kill thousands of people each year (CDC 2015). Recently, Olayemi et al. (2016) discovered Lassa virus in three *M. natalensis* individuals with A-II rather than A-I lineage mitochondria. They suggested that Lassa virus could potentially spread to *M. natalensis* lineages other than A-I and thus to the rest of sub-Saharan Africa. However, these individuals are from two A-I mitochondrion dominated localities in the [A-I, A-II] contact zone. Subspecific *M. natalensis* taxa can only be approximated by mitochondrial assays (Gryseels et al. 2017). Because mitochondrial copies from one taxon can cross a hybrid zone into the other, the three individuals may thus have swapped mitochondria rather than arenaviruses. In that case there would be no reason to suggest an immediate potential Lassa virus spread into the range of *M. natalensis* taxon A-II. On the other hand, Lassa virus has been found in other rodent species as well (Olayemi et al. 2016; Yadouleton et al. 2019), so the importance of *M. natalensis* intraspecific structure in restricting Lassa virus to West-Africa cannot yet be fully assessed.

Though many individuals were sampled, the [B-IV, B-V] contact zone study area in Gryseels et al. (2017) is limited in extent: a transect following a busy road between Tanzania's largest city, Dar es Salaam, and the capital, Dodoma. Along this transect, the B-IV subspecific taxon range was less sampled (only two B-IV versus nine B-V localities, one [B-IV, B-V] locality). Furthermore, the authors point out that a very recent (perhaps human-mediated) contact between the host taxa in this corridor could also explain the lack, or delay, of viral spread. However, it is unlikely that the *M. natalensis* taxa met that recently as the climate in Tanzania has been quite stable for at least 5,000 years (deMenocal et al. 2000; Garcin et al. 2012) and as B-IV mitochondrial copies can be found more than a few dozen kilometres from the hybrid zone centre in the B-V taxon range (Gryseels et al. 2017). As three *M. natalensis* mitochondrial lineage ranges meet at long contact fronts in Tanzania, the *M. natalensis*-borne arenaviruses in this region are an ideal system to investigate if pathogens/parasites only occur in a limited geographic area because they are specific to subspecific taxa.

The non-Lassa *M. natalensis*-borne arenaviruses have never been reported to cause human disease, but share many ecological features with Lassa virus (Demby et al. 2001; Borremans et al. 2011, 2015; Fichet-Calvet et al. 2014; Gryseels et al. 2015; Mariën et al. 2017) and may thus aid understanding of Lassa virus patterns. One feature that has not been compared yet among *M. natalensis*-borne arenaviruses is their spatial genetic structure. Because of the undersampling of the B-IV range, spatial genetic structure data on Tanzanian arenaviruses has been limited to that of Morogoro virus, and at a relatively small spatial scale within its geographic range. Gryseels et al. (2017) observed four Morogoro virus clades associated with one, two or three adjacent localities, and the distribution of these clades was not associated with genetic substructure of the host. Morogoro virus sequences from suburban Morogoro form a clade together with sequences from the adjacent, rural localities on both sides (Gryseels et al. 2017), while multilocus microsatellite markers indicate that the *M. natalensis* population in suburban Morogoro is different from those at nine surrounding rural localities (Gryseels et al. 2016). Furthermore, there are no rivers, roads or changes in land cover located between the different virus clades. The Morogoro virus spatial genetic structure rather shows a pattern of isolation by distance (Gryseels et al. 2017).

In this study we sampled 1772 *M. natalensis* individuals from localities spanning a strip of ~800 km from the south west to the north east of Tanzania, typed their mitochondria and screened them for arenaviruses in order to (1), gauge if the barrier observed for Morogoro and Gairo virus in Gryseels et al. (2017) is geographically restricted or holds throughout the length of the [B-IV, B-V] contact zone; (2), investigate whether Tanzanian B-VI *M. natalensis* (in contact with both Gairo-carrying B-IV and Morogoro-carrying B-V multimammate mice) carry Luna or Mopeia virus as B-VI individuals in Zambia (Ishii et al. 2011), Mozambique (Wulff et al. 1977), and Zimbabwe (Johnson et al. 1981), rather than Gairo or Morogoro virus; and (3), test if Morogoro and Gairo virus show similar spatial genetic structure and prevalence.

2. Materials and methods

2.1 Trapping and sampling

Mastomys natalensis were trapped in Sherman live traps (H.B. Sherman Traps Inc., Tallahassee, USA) and snap traps baited with a mixture of peanut butter, maize flour and canned fish in

June–August 2015 and 2016 (Supplementary Fig. S1). Mice caught in live traps were euthanized by cervical dislocation or an overdose of Isoflurane prior to dissection (Directive 2010/63/EU). Kidneys for arenavirus RNA screening were preserved in RNAlater and stored at -20°C and -80°C for short- and long-term storage, respectively. For anti-arenavirus antibody screening, blood from the heart was preserved on pre-punched Serobuvar filter papers that absorb $\sim 10 - 15\ \mu\text{l}$ per spot (LDA22 Zoopole, Ploufragan, France). Field work was approved by the University of Antwerp Ethical Committee for Animal Experimentation (2014-98) and complied with regulations of the Research Policy of Sokoine University of Agriculture as stipulated in the Code of Conduct for Research Ethics.

2.2 Arenavirus RNA screening and amplification

RNA was extracted from 1772 *M. natalensis* kidneys using the Nucleospin RNA Isolation kit (Macherey-Nagel, Düren, Germany). Samples were initially pooled by two or three (Supplementary Fig. S1), then positive pools were resolved by screening their constituting kidney samples. RNA extractions were screened with the SuperScript One-Step RT-PCR System kit (Invitrogen, Carlsbad, USA). Initial screening targeted a 340 nt fragment of the *L* gene using the following primers: MoroL3359-forward, LVL3359D-plus, LVL3359G-plus, MoroL3753-reverse, LVL3754A-minus, and LVL3754D-minus (Günther et al. 2009; Vieth et al. 2007). Additional screening assays were performed on samples from antibody positive localities in a strip from southwestern to central Tanzania where no *L* RNA was detected. This screening simultaneously targeted a 513–516 nt NP gene fragment and a 906–912 nt GPC gene fragment using the following primers: OWS0001-fwd and OWS1000-rev; and OWS2805-fwd, OWS2810-fwd, OWS3400-rev, and OWS3400A-rev (Ehichioya et al. 2011). The same primers were used to amplify NP and GPC gene fragments for kidney samples that tested positive for arenavirus *L* gene RNA. Amplicons from positive samples were Sanger-sequenced in both directions at the Genetic Service Facility of the Vlaams Instituut voor Biotechnologie (Antwerp, Belgium). The corresponding sequences were deposited in GenBank under accession numbers MH460970–MH461096.

2.3 Anti-arenavirus antibody screening

As no *L* gene RNA was detected at any locality within a strip of about 350 km from southwestern to central Tanzania despite a combined sample size of 551 individuals, nor at 2 localities in east Tanzania each with over 100 individuals screened, arenavirus presence in these regions was further assessed by screening 794 dried blood samples for IgG mouse antibodies specific for Old World arenaviruses. Furthermore, 50 dried blood samples from an arenavirus RNA positive locality in the south west (Ibohora) were screened to improve the statistical power for antibody analyses (see below). Dried blood spots were screened for antibodies using an indirect immunofluorescence assay as previously described (Günther et al. 2009; Borremans 2014; Gryseels et al. 2015). Dried blood spots were initially pooled by two, then blood samples from positive pooled samples were tested separately.

2.4 *Mastomys natalensis* mitochondrial genotyping

DNA was extracted from 96 per cent ethanol-preserved spleen, heart, or tail samples from a subset of individuals, including all arenavirus positive individuals, using the Exgene Tissue SV plus

mini (GeneAll Biotechnology, Seoul, Korea). Next, a 1,140 bp cytochrome *b* fragment was targeted in a PCR with L14723 and H15915 primers (Lecompte et al. 2002). One microlitre of DNA was added to 5 μl of Multiplex PCR Master Mix (Multiplex PCR Plus Kit, Qiagen, Hilden, Germany) and 0.3 μl of each primer in a total volume of 10 μl , followed by initial denaturation for 15 min at 95°C ; 35 cycles of denaturation for 30 s at 95°C , annealing for 30 s at 50°C and extension for 1 min at 72°C ; and final extension for 5 min at 72°C . PCR products were purified with an Exo-CIP PCR clean-up protocol and Sanger sequenced at GATC Biogen (Köln, Germany). The corresponding sequences were deposited in GenBank under accession numbers MK453920–MK454532.

They were aligned with sequences from Colangelo et al. (2013) and unambiguously assigned to one of three lineages (B-IV, B-V, B-VI) based on their position in a maximum likelihood (ML) phylogenetic tree. This tree was constructed in RAXML (Stamatakis 2014) in the CIPRES Science Gateway (Miller, Pfeiffer, and Schwartz 2010) with a general time reversible (GTR) substitution model and branch support assessed by 1000 bootstrap resampling. To explore substructure within mitochondrial lineages, a second ML tree was run with unique haplotypes instead of all sequences. Haplotypes were generated in R 3.4.3 (R Development Core Team 2017) using an in-house script (Supplementary Fig. S2) and the Ape Package (Paradis and Schliep 2019). The sequences were first aligned with sequences from Colangelo et al. (2013) and Gryseels et al. (2016, 2017). The alignment was then trimmed to the longest stretch of nucleotides with no more than 25 per cent missing data at any position, resulting in an alignment of 705 bp. After trimming, sequences shorter than 700 bp were excluded and the remaining sequences were condensed to haplotypes.

2.5 Analyses of regional differences in arenavirus detection

In order to investigate differences in observed arenavirus RNA prevalence among a northern region (including all Gairo virus localities), eastern region (including all Morogoro virus localities) and southwestern region (including all Luna virus localities): (1) the field area was split into three regions using straight-line boundaries, (2) viral positive/negatives falling in the regions were used to fill 3×2 contingency tables, and (3) non-random distributions of positives in the contingency tables were detected by G-tests (Woolf 1957). This process was repeated to allow for uncertainty in the region membership of those localities with no viral positives, by permuting how region boundaries split them. For this purpose, screening data were supplemented with data on 1077 dried blood samples from 15 localities from Gryseels et al. (2017). One locality from Gryseels et al. (2017), Berega, was not included in the test because both Gairo and Morogoro virus were detected there. For those localities where no arenavirus RNA was detected, we cannot be sure which arenavirus would be endemic there and therefore in which of the three regions they would be most sensibly classified. We therefore varied the classification of the remaining localities 15 times by drawing different sets of straight lines between them as geographic boundaries.

A second set of G-tests was performed on antibody data available from the same localities as those from the RNA G-tests (published in Gryseels et al. 2015; Mariën et al. 2017). The test was repeated 15 times with the same classifications as used above, but it only tested the difference in prevalence between the eastern and southwestern region. The northern region was not included because the available antibody data originated

from only two to three localities (depending on the classification). The G-tests were performed using the RVAideMemoire package (Hervé 2019) in R 3.3.2. For the RNA G-tests, G-test permutations with a ($P < 0.05$) significant outcome were further examined with pairwise G-tests from the same package. These pairwise tests used a Bonferroni correction for multiple testing.

2.6 Arenavirus genetic analyses

Sequences of coding regions were aligned per gene in Geneious R8.1 (Biomatters, New Zealand 2015) using translation alignment with the Geneious alignment algorithm. These alignments included the new sequences from this study; all sequences of Luna, Morogoro, and Gairo virus deposited in GenBank; some representatives of Old World arenaviruses (Lassa, Mopeia, Mobala, and Ippy virus); and Lymphocytic choriomeningitis virus as the outgroup.

Phylogenetic trees were inferred separately for the three genes so clear cases of re-assortment or recombination could be detected. Bayesian inference (BI) and ML were used as implemented in MrBayes v3.2.6 (Ronquist et al. 2012) and RaxML v8 (Stamatakis 2014), respectively, in the CIPRES Science Gateway (Miller, Pfeiffer, and Schwartz 2010). For both BI and ML analyses, sequences were partitioned according to codon position and the GTR nucleotide substitution model with four gamma rate variation (+G) categories was used as advised by jModelTest v2.0 (Darriba et al. 2012). We did not include a proportion of invariable sites (+I), as Stamatakis (2016) does not recommend to include both +G and +I. In ML analyses, 1000 bootstrap samples were simulated to determine branch support. During BI, Metropolis coupling with three heated and one cold chain was applied. In two independent runs the chains ran for 15 million generations. The cold chain was sampled every 500 generations after discarding the first 25 per cent as burn-in. The effective sample size (at least 200) and the trace pattern of the substitution model parameters were checked in Tracer v1.6 (Rambaut et al. 2014). Output trees were visualized in FigTree (Rambaut 2012).

Spatial autocorrelation in Morogoro and Gairo virus *L* and *NP* sequences from this study and in Lassa virus *NP* sequences from Fichet-Calvet et al. (2016) was tested and visualized for distance classes of 10 km with the non-parametric heterogeneity test of Smouse et al. (2008) in GenALEx 6.5 (Peakall and Smouse 2006, 2012). The first step of this analysis is a separate spatial autocorrelation analysis of the different groups using a genetic distance (for non-simple sequence repeats of haploid organisms, i.e. the number of mismatches) matrix and geographic distance matrix (Peakall and Smouse 2006). In this step autocorrelation coefficients (Smouse and Peakall 1999) are calculated and visualized per distance class in a spatial autocorrelation plot. These coefficients can range from -1 to 1 and express how similar sequences are to sequences found within a certain geographic distance range from them in comparison to random sequences from the dataset. As the largest geographic distance between Lassa virus sequences was 58 km, the Morogoro and Gairo virus datasets were truncated to sequences found no more than 60 km from each other and the analyses were thus limited to six distance classes. Next, the pooled autocorrelation across the different groups is estimated and the distribution of random departures from this null average is evaluated by bootstrapping 999 times (Peakall and Smouse 2006). Then, 'T2', a squared paired-sample t-test statistic, is calculated for each

distance class to assess the upper tail probability that the observed T2 value is larger than expected under the null hypothesis (no difference among the groups). Lastly, from the *P* values for the T2 statistics, the 'Omegas' over all distance classes are calculated and the probability that an observed Omega is larger than expected under the null hypothesis is evaluated. For both T2 and Omega, significance was set at 0.01 instead of 0.05, as recommended (Banks and Peakall 2012).

Finally, we used the weak-differentiation approximation to Rousset's isolation-by-distance analysis (Rousset 2000) developed for the case of a haploid organism (Rousset 2017). Genetic distances (\hat{a} -like statistics) and geographic distances were calculated with the R package Genepop (Rousset 2008). Next, ANCOVAs of genetic distance on ln-transformed geographic distance with virus species as a fixed factor and an interaction term between ln-transformed geographic distance and virus species were performed separately for the *L* and *NP* gene to test for differences among the viruses. The inverse of the slope of pairwise genetic distances on ln-transformed geographic distances is an estimator for the genetic neighbourhood size. This is the effective population size in a continuous population that is isolated by distance (Wright 1946), estimated as $4\pi D_e \sigma^2$ with D_e the effective population density and σ^2 the mean dispersal scale per generation between parent and offspring (Rousset 2000).

3. Results

3.1 Arenavirus detection and host genetic structure

A total of 52 arenavirus positives were detected in 1772 kidney samples: 38 Gairo positives with 30 unique *L* gene sequences, 10 Morogoro positives (all unique *L* gene sequences), and 4 Luna positives (all unique *L* gene sequences) (Fig. 1a and b; Supplementary Fig. S1). Identical sequences were only found at the same locality in all but one case (localities 29 km apart). For this case, we performed an independent RNA extraction, screening and sequencing which confirmed the result. *NP* and *GPC* gene sequences were obtained for 51 and 43 out of 52 *L* gene positive samples, respectively. Samples with identical *L* sequences also had identical *NP* sequences, but those for which the *GPC* sequencing was successful (four pairs of identical *L* and *NP* sequences) did not have identical *GPC* sequences.

The distribution of the three *M. natalensis* mitochondrial clades in the study area is shown in Fig. 2. While these clades were highly supported in the phylogeny, there was no clear structure within mitochondrial lineages (Supplementary Fig. S3 and Supplementary Table S4). All arenaviruses were found in correspondence with their expected *M. natalensis* mitochondrial clades (i.e. Gairo virus in B-IV, Morogoro virus in B-V and Luna virus in B-VI individuals), except for three Gairo viruses found in individuals with B-V mitochondria at locality Kimamba (Fig. 2b). No *L* gene RNA was detected at two localities in east Tanzania despite over 100 individuals screened in each (Selous and Ikwiriri) (Fig. 1a), nor at any locality of a 350 km strip from southwestern to central Tanzania (Fig. 1a). Therefore, dried blood samples from these localities were screened for anti-arenavirus antibodies. No antibodies were detected in the two localities in east Tanzania, but antibodies were detected in fourteen samples from four localities in the strip from southwestern to central Tanzania (Fig. 1c; Supplementary Fig. S1). During an additional *GPC-NP* gene screening, no sample from these antibody positive localities tested positive.

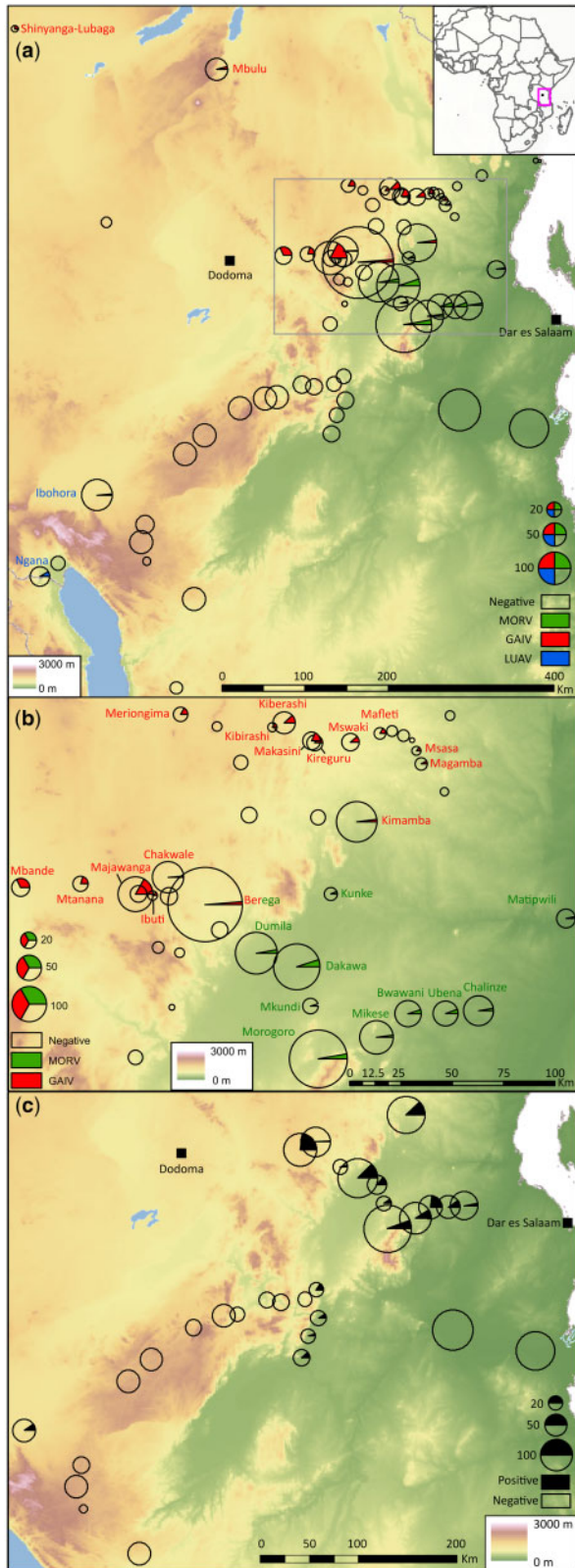


Figure 1. Proportion of arenavirus L gene RNA (a, b) and anti-arenavirus antibodies (c) detected in *Mastomys natalensis* in Tanzania in this study and in Gryseels et al. (2015) and Mariën et al. (2017). (b) An enlargement of the area in the grey rectangle in (a). Pie chart areas are scaled to the number of individuals screened. Locality names are coloured by virus species detected there. Elevation data from the U.S. Geological Survey’s Center for Earth Resources Observation and Science. Lake data from Sebastian (2014).

3.2 Analyses of regional differences in detection

A set of G-tests was carried out to compare the observed arenavirus RNA prevalence among three geographic regions: a northern region including all Gairo virus localities, an eastern region including all Morogoro virus localities, and a southwestern region including all Luna virus localities. Localities where no arenavirus RNA was detected were variably classified in each of the three regions by drawing lines between them as geographic boundaries (Fig. 3). All permutations with varying locality-classifications into the three regions show a significant non-

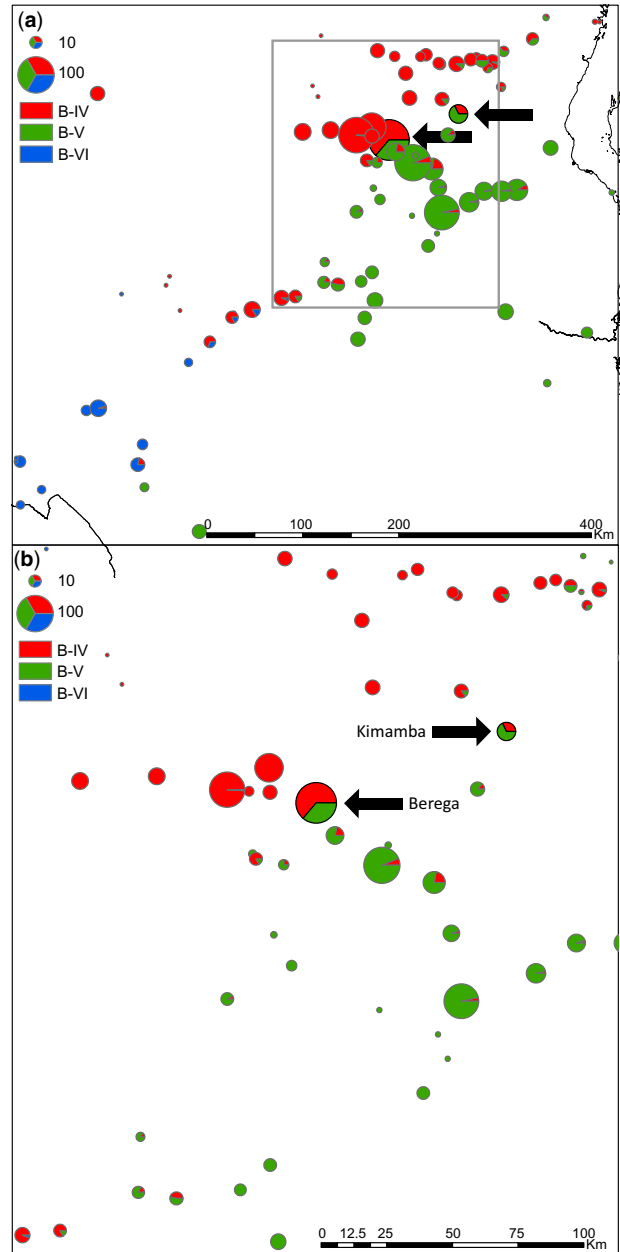


Figure 2. (a) Distribution of *Mastomys natalensis* mitochondrial lineages sensu Colangelo et al. (2013) in the study area. The mitochondrial lineages were determined from cytochrome b sequences from this study, GenBank and the African Mammalia database from the Belgian Biodiversity Platform (Van de Perre et al. 2019). (b) An enlargement of the area in the grey rectangle in (a). Pie charts are scaled to the number of individuals. The localities where Gairo virus was found in individuals with B-V mitochondria, Berega (screened in Gryseels et al. 2017) and Kimamba (screened in this study), are indicated with black arrows.

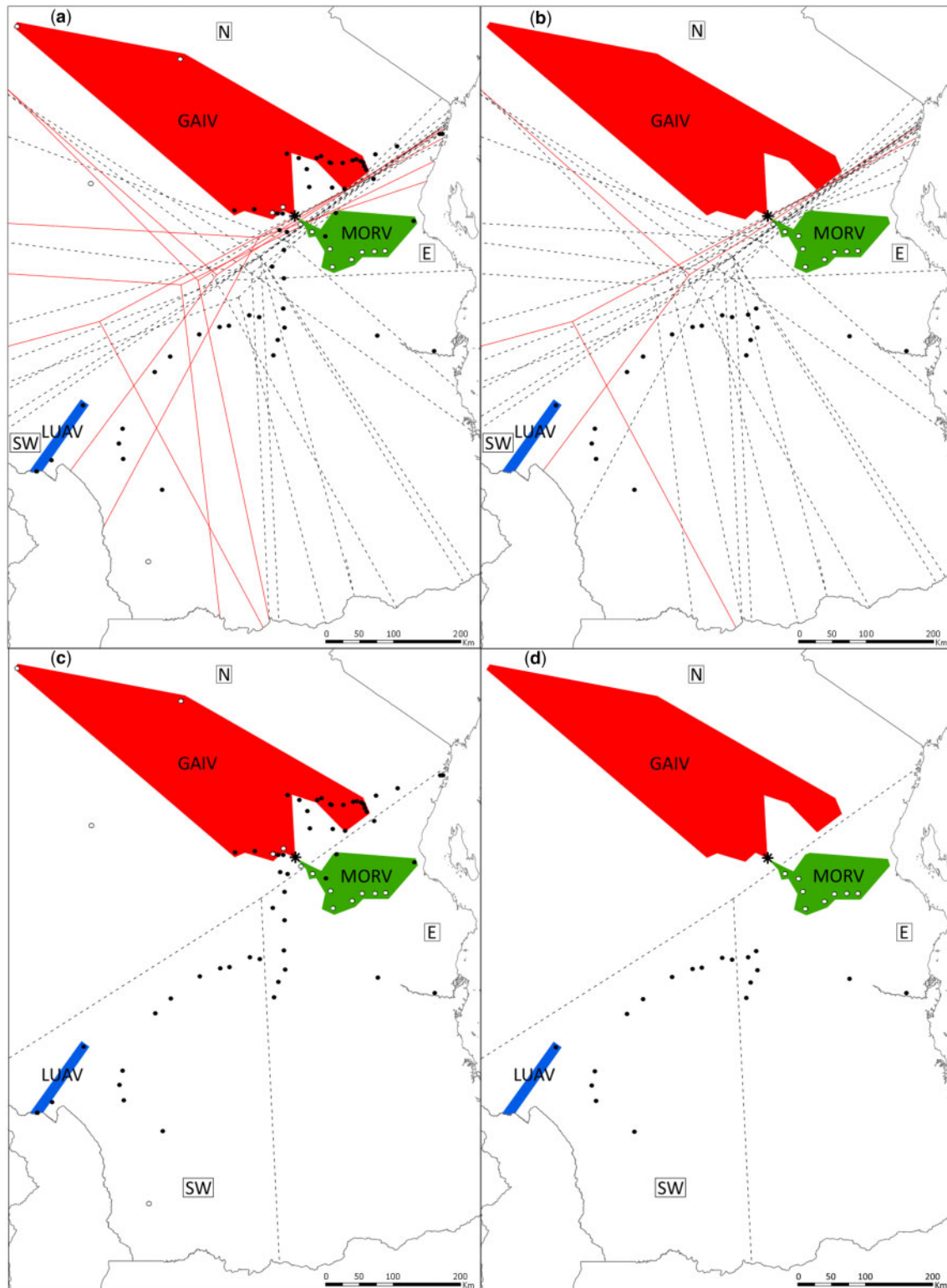


Figure 3. Variable classifications of localities into three geographic regions as used in the *Mastomys natalensis* arenavirus RNA G-tests (a, c) and the anti-arenavirus antibody G-tests (b, d). (a, b) All fifteen classifications overlaid, but for clarity only one of these is shown in (c, d). The other fourteen classifications are shown in [Supplementary Fig. S5](#). A set of red solid lines indicate no significant difference was found between the E and SW region in that classification. Filled circles represent localities screened in this study; open circles represent localities screened in [Gryseels et al. \(2017\)](#) (RNA data), or [Gryseels et al. \(2015\)](#) and [Mariën et al. \(2017\)](#) (antibody data). The asterisk represents a locality where both Morogoro and Gairo virus were detected in [Gryseels et al. \(2017\)](#) and was thus excluded from the G-tests.

random distribution of arenavirus positives over the three regions ($G: 47.02 - 73.47, Df = 2, P < 0.001$). Pairwise tests reveal this result is due to a significantly higher prevalence of

arenaviruses in north Tanzania compared with the east ($P < 0.001$) and the south west ($P < 0.001 - 0.031$). The result of the pairwise comparison between the eastern and

southwestern regions depends on the classification of the localities in southwestern to central Tanzania where no arenavirus RNA was detected ($P < 0.001 - 1$). If fewer arenavirus RNA-free localities are assigned to the southwestern region, the P -values are higher, and thus the higher prevalence in the eastern compared with the southwestern region becomes less clear (Fig. 3a).

A second set of G -tests was performed on the anti-arenavirus antibody data that were available from the same localities as those from the RNA G -tests. These tests explored the eastern–southwestern region pairwise comparison further: P values are lower (G : 1.71 – 47.79, $Df = 1$, $P < 0.001 - 0.191$) than those for the RNA data pairwise comparison, resulting in more locality-classifications with a significantly higher prevalence for eastern compared with southwestern Tanzania (13 vs. 10 out of 15 classifications) (Fig. 3b).

3.3 Arenavirus genetic analyses

All Gairo, Morogoro, and Luna L, NP, and GPC sequences cluster with sequences from their respective virus species (support of 0.98 – 1 for BI/88 – 100 for ML analyses), so no re-assortment or pronounced recombination is observed among the three virus species (Fig. 4; Supplementary Fig. S6).

Luna virus was detected in two localities in southwestern Tanzania (Fig. 4d). Sequences from the two Tanzanian localities

cluster together in NP and GPC analyses (0.92 – 0.94 for BI and 82 – 89 for ML, see Supplementary Fig. S6), but not in the L analysis (Fig. 4d). Morogoro virus sequences cluster with sequences from neighbouring localities (Fig. 4b; Supplementary Fig. S6). The support of these clades, however, depends on the analysis (BI or ML) and the gene. One clade comprises sequences from Morogoro, Mikese, and Mkundi (see Fig. 1b for locality position) and is highly supported in NP and GPC BI analyses (posterior probability (PP) of 0.98 and 0.99, respectively). Sequences from Bwawani and Ubena form a second clade with high support in BI analyses (PP: 0.97 – 0.99). The new sequence from Matipwili clusters with older samples from Chalinze, the locality it is closest to, though this is only well supported in GPC trees (PP 1 for BI and bootstrap value of 94 for ML analysis). The fourth clade contains sequences from Berega, Dumila, and Dakawa (both old and new). This clade has high PP for L and NP BI trees (0.98 and 1, respectively). The new sequence from Kunke cannot be assigned to any of the four clades as it does not cluster consistently with other clades across gene trees. Gairo virus was previously detected in three localities along the road from Dar es Salaam to Dodoma (Majawanga, Chakwale, and Berega) and in two more distant localities (Shinyanga-Lubaga and Mbulu) (Gryseels et al. 2017). In this study Gairo virus was detected in three more localities on the Dodoma side of that road (Mbande, Mtanana, and Ibuti), in nine localities (from



Figure 4. Arenavirus L gene Bayesian inference tree. Diamonds indicate node support: no symbol for supports <math>< 0.70</math>, red for 0.70 – 0.90, yellow for 0.90 – 0.95, and green for > 0.95. The scale bars represent the number of nucleotide substitutions per site. (b–d) Subtrees from the tree in (a), for which the subtrees for Morogoro, Gairo, and Luna virus, respectively, have been expanded. In (a) taxa are named as the virus species followed by the sampling country, the host species and the accession number from GenBank between brackets. In (b–d), sequences are named as the country (for d only), locality, the year, the sample code, and the accession number from GenBank between brackets. New sequences from this study are coloured.

Meriongima to Magamba) along a less busy paved road in the north, and in one locality in between these two roads (Kimamba) (Fig. 1a and b). Contrary to the pattern observed for Morogoro virus sequences, Gairo virus sequences from neighbouring localities on a road do not cluster all together, nor do they cluster together per road (Fig. 4c; Supplementary Fig. S6).

In order to further explore differences in spatial genetic structure between Gairo and Morogoro virus, virus-specific genetic neighbourhood sizes were estimated and spatial autocorrelograms were constructed. As in the phylogenetic trees, spatial genetic structure is less pronounced for Gairo compared with Morogoro virus in spatial autocorrelograms (Fig. 5; Supplementary Fig. S7). The variation in sequences is thus less well explained by isolation-by-distance. The correlograms differ significantly from each other in non-parametric heterogeneity

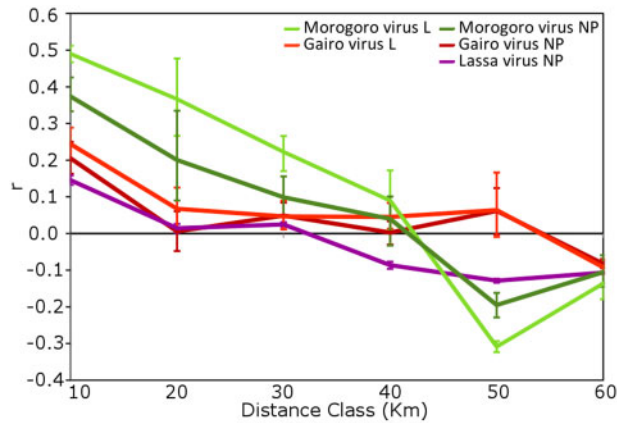


Figure 5. Spatial autocorrelograms for Morogoro, Gairo, and Lassa virus pairs found up to 60 km from each other. Each autocorrelation coefficient (*r*) is plotted at the end of a 10 km distance class with 95 per cent bootstrap confidence intervals.

tests due to a higher autocorrelation coefficient for Morogoro virus in the first (NP) or first three (L) distance classes and a lower autocorrelation coefficient in the fifth distance class (Supplementary Fig. S7). For Morogoro virus, but not for Gairo virus, there is a significant difference between L and NP correlograms. This difference is due to larger absolute correlation coefficient values for L compared with NP, significant at the first, third and fifth distance classes (the classes with the highest sample sizes). A correlogram was also constructed for NP Lassa virus sequences from Fichet-Calvet et al. (2016). This correlogram differs significantly from the Morogoro and Gairo virus NP correlograms due to a lower autocorrelation coefficient in the first two (Morogoro virus) and the fourth and fifth (Gairo virus) distance classes and a higher autocorrelation coefficient in the fifth distance class (Morogoro virus) (Supplementary Fig. S7).

Slopes of pairwise genetic distances on ln-transformed geographic distances between sequences are significantly different among different viruses for both L and NP sequences (Table 1; Fig. 6 with log-transformed instead of ln-transformed geographic distances for readability). The inverse of the slope is an estimator for the genetic neighbourhood size (N_b) (Table 1). Where the probability of coalescence in the previous generation in a panmictic population is $1/N_b$, in a spatially continuous population it is $1/N_b$. Although the neighbourhood size estimates are significantly different among Morogoro, Gairo, and Lassa virus, they are of the same order of magnitude for all viruses and genes (Table 1).

4. Discussion

The hypothesis of *M. natalensis*-borne arenavirus restriction to subspecific taxa predicts three pairings of host and virus in the Tanzanian field area. Three arenavirus species were detected in association with their expected *M. natalensis* mitochondrial clades over a large area in Tanzania, including the first

Table 1. ANCOVA virus species—ln geographic distance interaction term statistics and neighbourhood size estimates for L and NP gene sequences.

		L			NP			
ANCOVA statistics		Df	F	P	Df	F	t	P
	Three-way				2	17.79		<0.001
	Pairwise							
	GAIV vs. MORV	1	432.50	<0.001			5.25	<0.001
	GAIV vs. LASV						4.46	<0.001
	MORV vs. LASV						-3.60	<0.001
Neighbourhood sizes (individuals)								
	GAIV		12.71				18.13	
	MORV		6.15				9.71	
	LASV						13.87	

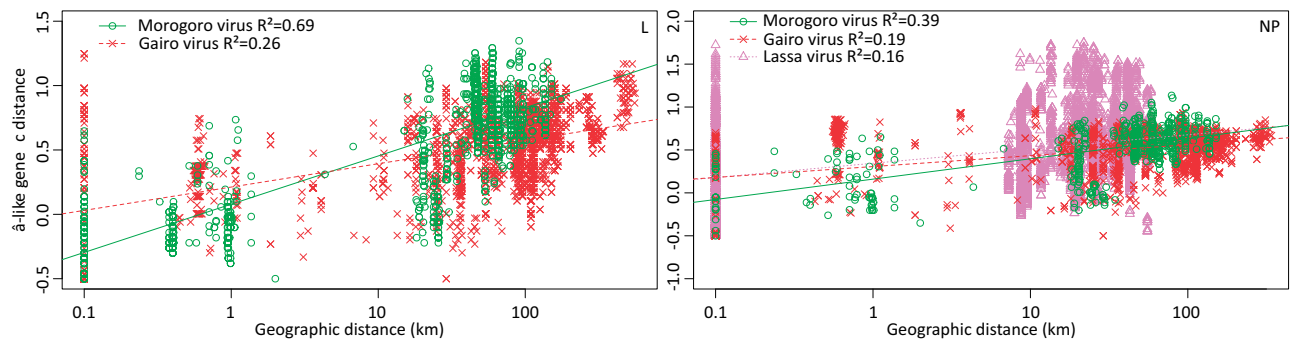


Figure 6. Linear regressions of genetic distances (\hat{a} -like statistics) on log-transformed geographic distances for the L (left) and NP gene (right).

detection of Luna virus outside of Zambia. No arenaviruses were detected in a large part of central Tanzania. The arenavirus with the highest prevalence, Gairo virus, shows less spatial genetic structure than Morogoro virus.

4.1 Arenavirus specificity

In the *cytochrome b* data, the host mitochondrial lineages could easily be distinguished, but there was no clear geographic structure within the lineages (Supplementary Fig. S3). In 49 out of 52 virus positive samples, arenavirus RNA was detected in combination with their corresponding *M. natalensis* mitochondrial lineage: Gairo virus in B-IV, Morogoro virus in B-V, and Luna virus in B-VI individuals. The four Luna virus positive mice were found at two localities in southwestern Tanzania and are the first Luna viruses to be detected outside of Zambia. Given the large geographic range of the *M. natalensis* B-VI mitochondrial lineage in Southern Africa (Colangelo et al. 2013), it might be expected that Luna virus will be found in other Southern African countries besides Zambia and Tanzania, such as Malawi and Angola. However, not only Luna virus, but also Mopeia virus occurs in association with the B-VI mitochondrial lineage (Wulff et al. 1977) and it is currently not clear if there is a geographic divide in the distribution of these two arenaviruses nor whether there is further cryptic structure within this large host range.

Given the widespread match between arenavirus species and *M. natalensis* mitochondrial clades in this study, we can confirm the specificity of the two virus-host pairs observed in Gryseels et al. (2017) is not a special geographically limited case, for example, due to human-mediated dispersal on a busy road between Dar es Salaam and Dodoma: sampling of that contact zone now spans over 100 km. Hence, our study supports the hypothesis that subspecific *M. natalensis* taxa constrain the geographic ranges of their arenaviruses.

However, of the fifty-two viral positive samples, there were three host–virus mismatches: three Gairo virus samples were found in host individuals carrying B-V instead of B-IV mitochondria. Kimamba, the locality where these individuals were caught, has eight out of the twenty-five typed hosts belonging to the B-IV mitochondrial lineage and seventeen to the B-V lineage, indicating that it lies close to the centre of the [B-IV, B-V] hybrid zone. Cline analysis based on nuclear diagnostic SNPs confirms this is the case (Cuypers and Baird, unpublished results). In Gryseels et al. (2017) dataset, there were also three Gairo virus positive individuals with B-V mitochondria, again from a locality with intermediate frequencies of both mitochondrial lineages, but also both alleles of a diagnostic Y chromosome SNP marker and where microsatellite markers indicated that most individuals were autosomal hybrids. One of these mismatches was a male individual which carried a B-V mitochondrion, but a B-IV Y chromosome (Gryseels et al. 2017). We suggest in both localities mitochondrial mismatches are hybrids carrying Gairo virus, rather than individuals of one taxon carrying the ‘wrong’ arenavirus typical of the neighbouring taxon. Such cases show transmission into hybrid zones is possible, making the restriction of viral transmission across hybrid zones all the more striking.

4.2 Regional differences in prevalence

The data in this study were collected for diverse purposes during different expeditions with diverse sampling strategies. For this reason, while they allow for genetic meta-analysis, they are not well suited for ecological hypothesis testing. We did,

however, test for heterogeneity in prevalence, (i) to obtain an approximate prevalence estimate of Gairo and Morogoro virus for our genetic structure analyses and (ii) to explore if the lack of arenaviruses in southwestern to central Tanzania was significant given the prevalence in other areas of the study region. There is a non-random distribution of arenavirus prevalence: the northern region, including all Gairo virus localities, has a significantly higher prevalence than the eastern region, including all Morogoro virus localities (Fig. 3a; Supplementary Fig. S5) and the southwestern region has significantly less arenaviruses than both the northern and eastern region (Fig. 3; Supplementary Fig. S5), though this distinction becomes less clear depending on how boundaries are defined.

Factors that could play a role at this regional level include (1), Cryptic viruses: it is possible the low prevalence region corresponds to the range of a viral strain undetected by our assays due to lower detection sensitivity (Emmerich, Günther, and Schmitz 2008; Leski et al. 2015). However, our screening methods can detect RNA from, and antibodies against, different arenaviruses in different countries (Wulff et al. 1977; Johnson et al. 1981; Gonzalez et al. 1984; Vieth et al. 2007; Günther et al. 2009; Yama et al. 2012; Kronmann et al. 2013; Witkowski et al. 2015; Gryseels et al. 2017; Mariën et al. 2017) (Fig. 1). The probability that a divergent strain in southwestern to central Tanzania is not picked up at all by the L, NP, and GPC gene primer pairs (not three, but six pairs in total) or antibody screening thus seems low. (2), RNA screening: there are slight variations in the arenavirus RNA screening of samples included in the G-tests (dried blood or kidney samples, pooled by two or three; see Supplementary Fig. S1). However, antibodies were screened in the same way for all samples, and those G-tests indicated an even stronger difference between the eastern and southwestern region, i.e. more arenavirus-free localities from central to southwestern Tanzania could be assigned to the eastern region instead of the southwestern region before P-values rose above 0.05. (3), Sampling season: most localities were only sampled once, and not all of them in the same year or season (Supplementary Fig. S1). While each prevalence region includes many localities, there is one axis of this temporal variation which could cause a consistent difference at the regional level: localities in the eastern region were sampled both in the dry and rainy season, while the other localities were only sampled in the dry season (Supplementary Fig. S1). In Guinea, Lassa virus rainy season prevalence is two to three times higher than dry season (Fichet-Calvet et al. 2007). A similar pattern for East African *M. natalensis*-borne arenaviruses might explain why the eastern region has higher observed prevalence than the southwestern region, but not why it has a lower prevalence than the northern region. (4), Virus intrinsic factors: each prevalence region corresponds to a different species of virus. Different virus species may exhibit different prevalences in similar hosts. (5), Host intrinsic factors: prevalence regions are strongly correlated with host mitochondrial clade: cryptic host taxa may exhibit different prevalences for similar viruses. (6), Host density: first, it is striking that the virus-free strip from southwestern to central Tanzania and two east Tanzanian virus-free localities with a large sample size, correspond more or less to regions which are predicted to have low suitability for *M. natalensis* and Lassa virus in Mylne et al. (2015). For the virus-free strip the adverse environmental condition could be higher elevation (see Fig. 1). While host density might fall below a viral persistence threshold in the low prevalence region, extrapolation about East African arenaviruses from West African arenavirus data should naturally be treated with caution. Secondly, the three-way M.

natalensis contact zone from southwestern to central Tanzania may contribute, through selection against hybrids, to lower host density in this low prevalence region. Estimating the relative contribution of such factors when total viral prevalence is only ~3 per cent, with negative binomial distribution, is beyond the scope of the current study.

4.3 Spatial genetic structure

Morogoro virus spatial genetic structure is much more pronounced compared with that of Gairo and Lassa virus. As in Gryseels et al. (2017), Morogoro virus sequences from up to three adjacent localities cluster together forming three to four main clades (though the support depends on the gene and analysis). The position of the new sequences here from Matipwili and Dakawa further emphasizes that spatial genetic structure. In Fichet-Calvet et al. (2016) some Lassa virus sequences form monophyletic clusters within a locality, others cluster with sequences from neighbouring villages and few cluster with sequences from villages further away. More Gairo virus sequences cluster with sequences from more far-off localities, especially striking are the high number of Majawanga sequences from 2012 which cluster with sequences from many other localities sampled both in Gryseels et al. (2017) and in this study (Fig. 4c; Supplementary Fig. S6). However, Extended Bayesian Skyline plots (not included because of low power) show no signal of a major recent outbreak. Perhaps B-IV mice can move more easily through their landscape than B-V mice in theirs, however, Gryseels et al. (2017) could not find any landscape barriers for B-V mice that might explain the stronger Morogoro virus spatial genetic structure.

The difference in spatial autocorrelation coefficients in the spatial autocorrelograms (Fig. 5; Supplementary Fig. S7) and the difference in neighbourhood size estimates (the Gairo virus neighbourhood size estimate is two times larger for *L* sequences and almost two times larger for *NP* sequences) corroborate the apparent difference in spatial genetic structure between Morogoro and Gairo virus. The difference in neighbourhood size is also in agreement with a higher prevalence for Gairo virus (2–5× depending on the locality-classification), as neighbourhood size is a product of dispersal scale and (effective) population density.

The spatial autocorrelation coefficients (Fig. 5; Supplementary Fig. S7) of Lassa virus sequences from Fichet-Calvet et al. (2016) and, though with considerable scattering, the slope of the linear regression between Lassa virus genetic distance and geographic distance (Fig. 6) seem more similar to those of Gairo virus sequences at a small spatial scale, but more similar to those of Morogoro virus sequences at a larger spatial scale. Moreover, the neighbourhood size estimate of Lassa virus sequences was intermediate between those of Morogoro and of Gairo virus sequences and all neighbourhood sizes were of the same order of magnitude.

5. Conclusion

In this study, we show that cryptic structure within the host species can play a role in the restricted distribution of viruses: Morogoro, Gairo, and Luna arenavirus distribution is linked tightly to the ranges of subspecific taxa of *M. natalensis* on a large geographic scale. Estimators of arenavirus prevalence and genetic structure vary between geographic regions corresponding to the three host–arenavirus pairs. Estimators for Lassa virus fall within this variation, suggesting East African arenaviruses are safe models for understanding the spatial ecology of the biosafety level 4 Lassa virus.

Data availability

Sequence data are available in GenBank under accession numbers MK453920–MK454532 (*Mastomys natalensis* cytochrome *b* sequences) and MH460970–MH461096 (arenavirus sequences). Other data are available in [Supplementary Material](#).

Supplementary data

[Supplementary data](#) are available at *Virus Evolution* online.

Acknowledgements

The authors thank Tatiana Aghová, Khalid Kibwana, Adam Konečný, Jarmila Krásová, Vladimír Mazoch, Ondřej Mikula, Jana Vrbová Komárková and Christopher Sabuni for their assistance in the field. We also acknowledge Anna Bryjová, Joachim Mariën and Natalie Van Houtte for their help during the laboratory work and Ondřej Mikula for sharing his in-house script for haplotype generation (Supplementary Fig. S2).

Funding

This study was supported by projects of the Research Foundation - Flanders (GOA4815N), the Czech Science Foundation (15-20229S and 18-19629S) and the German Research Foundation (LE3111/1 (SP1596)). L.N.C. was a PhD fellow with the University of Antwerp and FWO (1171921N) during part of the work.

Authors' contributions

L.N.C., S.J.E.B., J.B., H.L., and J.G.B. participated in conceptualization; J.B., H.L., and J.G.B. contributed in funding acquisition; L.N.C., A.H., T.L., A.S.K., and J.B. contributed in fieldwork; L.N.C. and T.L. contributed in arenavirus laboratory work; A.H. contributed in *Mastomys natalensis* laboratory work; L.N.C., S.J.E.B., and J.G.B. participated in formal analysis and the original draft; S.J.E.B., S.G., H.L., and J.G.B. contributed in supervision; L.N.C. contributed in visualization; L.N.C., S.J.E.B., A.H., T.L., A.S.K., S.G., J.B., H.L., and J.G.B. contributed in the final draft.

Conflict of interest: None declared.

References

- Banks, S. C., and Peakall, R. (2012) 'Genetic Spatial Autocorrelation can Readily Detect Sex-biased Dispersal', *Molecular Ecology*, 21: 2092–105.
- Borremans, B. (2014) 'Ammonium Improves Elution of Fixed Dried Blood Spots Without Affecting Immunofluorescence Assay Quality', *Tropical Medicine and International Health*, 19: 413–6.
- Borremans, B. et al. (2011) 'Presence of Mopeia Virus, an African Arenavirus, Related to Biotope and Individual Rodent Host Characteristics: Implications for Virus Transmission', *Vector Borne and Zoonotic Diseases*, 11: 1125–31.
- Borremans, B. et al. (2015) 'Shedding Dynamics of Morogoro Virus, an African Arenavirus Closely Related to Lassa Virus, in Its Natural Reservoir Host *Mastomys natalensis*', *Scientific Reports*, 5: 10445.
- CDC (2015) *Lassa Fever*. <<https://www.cdc.gov/vhf/lassa/index.html>> accessed 30 Apr 2020.

- Colangelo, P. et al. (2013) 'A Mitochondrial Phylogeographic Scenario for the Most Widespread African Rodent, *Mastomys natalensis*', *Biological Journal of the Linnean Society*, 108: 901–16.
- Cybinski, D. H., George, T. D. S., and Paull, N. I. (1978) 'Akabane Virus in Australia', *Australian Veterinary Journal*, 54: 1–3.
- Darriba, D. et al. (2012) 'jModelTest 2: More Models, New Heuristics and Parallel Computing', *Nature Methods*, 9: 772.
- Demby, A. H. et al. (2001) 'Lassa Fever in Guinea: II. Distribution and Prevalence of Lassa Virus Infection in Small Mammals', *Vector Borne and Zoonotic Diseases*, 1: 283–97.
- deMenocal, P. et al. (2000) 'Abrupt Onset and Termination of the African Humid Period: Rapid Climate Responses to Gradual Insolation Forcing', *Quaternary Science Reviews*, 19: 347–61.
- Ehichioya, D. U. et al. (2011) 'Current Molecular Epidemiology of Lassa Virus in Nigeria', *Journal of Clinical Microbiology*, 49: 1157–61.
- Emmerich, P., Günther, S., and Schmitz, H. (2008) 'Strain-specific Antibody Response to Lassa Virus in the Local Population of West Africa', *Journal of Clinical Virology*, 42: 40–4.
- Fichet-Calvet, E. et al. (2014) 'Lassa Serology in Natural Populations of Rodents and Horizontal Transmission', *Vector-Borne and Zoonotic Diseases*, 14: 665–74.
- Fichet-Calvet, E. et al. (2007) 'Fluctuation of Abundance and Lassa Virus Prevalence in *Mastomys natalensis* in Guinea, West Africa', *Vector borne and zoonotic diseases*, 7: 119–28.
- Fichet-Calvet, E. et al. (2016) 'Spatial and Temporal Evolution of Lassa Virus in the Natural Host Population in Upper Guinea', *Scientific Reports*, 6: 21977.
- Garcin, Y. et al. (2012) 'East African mid-Holocene Wet-dry Transition Recorded in Palaeo-Shorelines of Lake Turkana, Northern Kenya Rift', *Earth and Planetary Science Letters*, 331 – 332: 322–34.
- Gonzalez, J. P. et al. (1984) 'Mobala Virus: Biological and Physicochemical Properties of a New Arenavirus Isolated in the Central African Republic', *Annual Review of Virology*, 135: 145–58.
- Gryseels, S. et al. (2016) 'Genetic Distinction Between Contiguous Urban and Rural Multimammate Mice in Tanzania Despite Gene Flow', *Journal of Evolutionary Biology*, 29: 1952–67.
- Gryseels, S. et al. (2017) 'When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure in the Spatial Spread of Arenaviruses', *PLoS Pathogens*, 13: e1006073.
- Gryseels, S. et al. (2015) 'Gairo Virus, a Novel Arenavirus of the Widespread *Mastomys natalensis*: Genetically Divergent, but Ecologically Similar to Lassa and Morogoro Viruses', *Virology*, 476: 249–56.
- Günther, S. et al. (2009) 'Mopeia Virus-related Arenavirus in Natal Multimammate Mice, Morogoro, Tanzania', *Emerging Infectious Diseases*, 15: 2008–12.
- Hervé, M. (2019) *RVAideMemoire: Testing and Plotting Procedures for Biostatistics*. R Package Version 0.9-73. <https://CRAN.R-project.org/package=RVAideMemoire>.
- Ishii, A. et al. (2011) 'Novel Arenavirus, Zambia', *Emerging Infectious Diseases*, 17: 1921–4.
- Johnson, K. M. et al. (1981) 'Recovery of a Lassa-related Arenavirus in Zimbabwe', *American Journal of Tropical Medicine and Hygiene*, 30: 1291–3.
- Kronmann, K. C. et al. (2013) 'Two Novel Arenaviruses Detected in Pygmy Mice, Ghana', *Emerging Infectious Diseases*, 19: 1832–5.
- Lecompte, É. et al. (2002) 'Cytochrome b-based Phylogeny of the Praomys Group (Rodentia, Murinae): A New African Radiation?' *C. R. Biologies*, 325: 827–40.
- Leski, T. A. et al. (2015) 'Sequence Variability and Geographic Distribution of Lassa Virus, Sierra Leone', *Emerging Infectious Diseases*, 21: 609–18.
- Linard, C. et al. (2007) 'Environmental Conditions and Puumala Virus Transmission in Belgium', *International Journal of Health Geographics*, 6: 55.
- Mariën, J. et al. (2017) 'No Measurable Adverse Effects of Lassa, Morogoro and Gairo Arenaviruses on Their Rodent Reservoir Host in Natural Conditions', *Parasites & Vectors*, 10: 210.
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010) 'Creating the CIPRES Science Gateway for inference of large phylogenetic trees', *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, pp. 1–8.
- Mylne, A. Q. N. et al. (2015) 'Mapping the Zoonotic Niche of Lassa Fever in Africa', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 109: 483–92.
- Olayemi, A. et al. (2016) 'New Hosts of The Lassa Virus', *Scientific Reports*, 6: 25280.
- Olayemi, A. et al. (2016) 'Arenavirus Diversity and Phylogeography of *Mastomys natalensis* Rodents, Nigeria', *Emerging Infectious Diseases*, 22: 687–7.
- Paradis, E., and Schliep, K. (2019) 'Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.
- Patot, S. et al. (2010) 'Prevalence of a Virus Inducing Behavioural Manipulation Near Species Range Border', *Molecular Ecology*, 19: 2995–3007.
- Peakall, R., and Smouse, P. E. (2006) 'GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research', *Molecular Ecology Notes*, 6: 288–95.
- Peakall, R., and Smouse, P. E. (2012) 'GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update', *Bioinformatics*, 28: 2537–9.
- Van de Perre, F. et al. (2019) *African Mammalia*. <<http://projects.biodiversity.be/africanmammalia>> accessed 30 Apr 2020.
- R Development Core Team. (2017) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rambaut, A. (2012) *FigTree v1.4*. <<http://tree.bio.ed.ac.uk/software/figtree/>>> accessed 30 Apr 2020.
- Rambaut, A. et al. (2014) *Tracer v1.6*. <<http://beast.bio.ed.ac.uk/Tracer>> accessed 18 Mar 2018.
- 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.
- Rousset, F. (2000) 'Genetic Differentiation Between Individuals', *Journal of Evolutionary Biology*, 13: 58–62.
- Rousset, F. (2008) 'GENEPOP'007: A Complete Re-implementation of the GENEPOP Software for Windows and Linux', *Molecular Ecology Resources*, 8: 103–6.
- Rousset, F. (2017) 'Population Genetic Data Analysis Using Genepop'.
- Saxenhofer, M. et al. (2019) 'Secondary Contact Between Diverged Host Lineages Entails Ecological Speciation in a European Hantavirus', *PLoS Biology*, 17: e3000142.
- Sebastian, K. (ed.) (2014) *Atlas of African Agriculture Research & Development*. Washington, DC: International Food Policy Research Institute.
- Smouse, P. E., and Peakall, R. (1999) 'Spatial Autocorrelation Analysis of Individual Multiallele and Multilocus Genetic Structure', *Heredity*, 82: 561–73.
- Smouse, P. E., Peakall, R., and Gonzales, E. (2008) 'A Heterogeneity Test for Fine-scale Genetic Structure', *Molecular Ecology*, 17: 3389–400.
- Stamatakis, A. (2014) 'RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies', *Bioinformatics*, 30: 1312–3.

- Stamatakis, A. (2016) 'The RAxML v8.2.X Manual'.
- Vieth, S. et al. (2007) 'RT-PCR Assay for Detection of Lassa Virus and Related Old World Arenaviruses Targeting the L Gene', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101: 1253–64.
- Witkowski, P. T. et al. (2015) 'Novel Arenavirus Isolates from Namaqua Rock Mice, Namibia, Southern Africa', *Emerging Infectious Diseases*, 21: 1213–6.
- Woolf, B. (1957) 'The Log Likelihood Ratio Test (the G-test): Methods and Tables for Tests of Heterogeneity in Contingency Tables', *Annals of Human Genetics*, 21: 397–409.
- Wright, S. (1946) 'Isolation by Distance Under Diverse Mating Systems', *Genetics*, 31: 39–59.
- Wulff, H. et al. (1977) 'Isolation of an Arenavirus Closely Related to Lassa Virus From *Mastomys natalensis* in South-East Africa', *Bulletin of the World Health Organization*, 55: 4: 441–4.
- Yadouleton, A. et al. (2019) 'Lassa Virus in Pygmy Mice, Benin, 2016 – 2017', *Emerging Infectious Diseases*, 25: 1977–9.
- Yama, I. N. et al. (2012) 'Isolation and Characterization of a New Strain of Lymphocytic Choriomeningitis Virus from Rodents in Southwestern France', *Vector-Borne and Zoonotic Diseases*, 12: 893–903.