Master thesis submitted to obtain the degree Master of Biology

Evolution and Behaviour Biology

Evolutionary diversity and distribution of arenaviruses in Tanzania

Laura Cuypers

Promotor

Prof. Dr. Herwig Leirs

Co-promotor

Dr. Joëlle Goüy de Bellocq

Faculty of Science

Department of Biology

Academic year 2016-2017

University of Antwerp

INSTITUTE OF VERTEBRATE BIOLOGY ACADEMY OF SCIENCES CR

SOKOINE UNIVERSITY OF AGRICULTURE *1984* ARDHI NI HAZINA

Table of contents

List of abbreviations

bp	base pairs
BI	Bayesian inference
cyt b	*cytochrome b* gene
GAIV	Gairo virus
GPC	glycoprotein
GSF	Genetic Service Facility
GTR	General Time Reversible
IFA	indirect immunofluorescence assay
IVB	Institute of Vertebrate Biology
LASV	Lassa virus
LUAV	Luna virus
ML	Maximum likelihood
MOPV	Mopeia virus
MORV	Morogoro virus
nt	nucleotides
PBS	phosphate-buffered saline
PCR	Polymerase Chain Reaction
RT	Reverse Transcription
SSP	stable signal peptide
SPMC	Sokoine University of Agriculture Pest Management Centre
VIB	Vlaams Instituut voor Biotechnologie

1. Summaries

1.1. Abstract

Different arenaviruses occur in different *Mastomys natalensis* and *Mus minutoides* mitochondrial lineages in distinct geographic regions throughout sub-Saharan Africa. In West Africa, *M. natalensis* is the reservoir of Lassa virus, which can spill over to humans and can cause fatal haemorrhagic fever. In East Africa, *M. natalensis* carries closely related arenaviruses which appear to exhibit similar dynamics, but which are not known to be pathogenic to humans.

The main objective of this Master thesis was to investigate the distribution, the genetic structure and the specificity of *M. natalensis* arenaviruses to three *M. natalensis* mitochondrial lineages (B-IV, B-V and B-VI). I screened 1155 *M. natalensis* individuals for arenavirus RNA, trapped in southwestern Tanzania, in the [B-IV, B-V] contact zone and in the putative three-way [B-IV, B-V, B-VI] contact zone. Additionally, I screened 21 *Mus minutoides* individuals for arenavirus RNA to explore the specificity of their arenaviruses to certain *M. minutoides* mitochondrial lineages.

I detected Gairo virus in B-IV, Morogoro virus in B-V and Luna virus in B-VI individuals, supporting the hypothesis that *M. natalensis* arenaviruses are constrained to a certain geographic region due to their specificity to a certain *M. natalensis* lineage. An arenavirus isolated from a *M. minutoides* individual is most likely Ngerengere virus, which was previously found in three individuals of the same mitochondrial lineage. It is possible that *M. minutoides* viruses are constrained by host lineages rather than by species as well.

No arenaviruses were detected in central Tanzania and in the north east, Morogoro virus prevalence was significantly lower than Gairo virus prevalence. Morogoro virus sequences also exhibited more spatial genetic structure than Gairo virus sequences. Both observations indicate that there might be a difference in virus and/or host dynamics. This could imply that both viruses are not equally suited as a model for Lassa virus.

1.2. Samenvatting

Verschillende arenavirussen komen voor in verschillende *Mastomys natalensis* en *Mus minutoides* mitochondriale lijnen in andere geografische regio's in sub-Sahara Afrika. In West-Afrika is *M. natalensis* het reservoir van Lassavirus. Dit virus kan overgedragen worden op mensen en fatale hemorragische koorts veroorzaken. In Oost-Afrika draagt *M. natalensis* nauwverwante

arenavirussen. Deze lijken een gelijkaardige dynamiek te hebben, maar niet pathogeen voor mensen te zijn.

Het doel van deze Masterproef was om de verspreiding, de genetische structuur en de specificiteit van *M. natalensis* arenavirussen voor drie *M. natalensis* mitochondriale lijnen (B-IV, B-V en B-VI) te onderzoeken. Ik heb 1155 *M. natalensis* individuen gescreend voor arenavirus RNA. Deze waren gevangen in het zuidwesten van Tanzania, in de [B-IV, B-V] contactzone en in de veronderstelde [B-IV, B-V, B-VI] contactzone. Daarnaast heb ik 21 *Mus minutoides* individuen gescreend voor arenavirus RNA om de specificiteit van hun arenavirussen voor bepaalde *M. minutoides* mitochondriale lijnen te onderzoeken.

Ik heb Gairovirus in B-IV, Morogorovirus in B-V en Lunavirus in B-VI individuen aangetroffen. Dit resultaat ondersteunt de hypothese dat *M. natalensis* arenavirussen begrensd zijn tot een bepaalde geografische regio door hun specificiteit voor een bepaalde *M. natalensis* genetische lijn. Een arenavirus geïsoleerd uit een *M. minutoides* individu is waarschijnlijk een Ngerengerevirus. Dit virus werd eerder gevonden in drie individuen van dezelfde mitochondriale lijn. Het zou kunnen dat ook *M. minutoides* virussen begrensd zouden worden door gastheerlijnen eerder dan door de gastheersoort.

Er werden geen arenavirussen gedetecteerd in centraal Tanzania en in het noordoosten was de Morogoro-virusprevalentie significant lager dan de Gairo-virusprevalentie. Morogoro-virussequenties vertoonden ook meer ruimtelijke genetische structuur dan Gairo-virussequenties. Beide observaties geven aan dat er mogelijk een verschil is in virus- en/of gastheerdynamiek. Dit zou kunnen impliceren dat beide virussen niet even geschikt zijn als model voor Lassavirus.

1.3. Lay summary

Multimammate Mice occur throughout Africa south of the Sahara desert. In different African regions they carry different arenaviruses, probably because these viruses are not specific to this mouse species, but to its genetic subdivisions (“lineages”). This idea was supported by my results. I screened 1155 Multimammate Mice from a strip of approximately 800 km from southwestern to northeastern Tanzania. In this strip three mouse lineages meet and indeed each lineage carried its own arenavirus. Two viruses, Gairo and Morogoro virus, had previously been described in northeastern Tanzania. The third, Luna virus, detected in the south west of Tanzania, had previously been found in Zambian mice of that lineage, but had never been detected outside of Zambia.

I also screened 21 African Pygmy Mice and found one arenavirus that is most likely Ngerengere virus. Ngerengere virus was previously found in three individuals from the same genetic lineage and two other arenaviruses have been detected in two other genetic lineages in Africa. Perhaps African Pygmy Mice arenaviruses are specific to certain genetic lineages of this mouse species as well.

In a strip of about 350 km from southwestern to central Tanzania I found no arenaviruses in Multimammate Mice and in the north east I found fewer Morogoro than Gairo viruses. Perhaps there are some slight differences in virus and/or host dynamics. That would also explain why Morogoro viruses found closer to each other resemble each other more than Gairo viruses found close to each other.

Researching how similar arenaviruses behave and what keeps them confined to certain regions is important for human health. The Multimammate Mouse arenaviruses in East Africa are not pathogenic, but in West Africa this mouse carries a closely related virus, Lassa virus, which can be transmitted to humans and is estimated to kill thousands of people each year.

2. Introduction

This Master thesis explores the diversity of arenaviruses in two rodent species in Tanzania. Some arenaviruses cause serious haemorrhagic fever or meningitis in humans (Armstrong and Wooley 1935; Rapp and Buckley 1962; Johnson et al. 1966; Frame et al. 1970; Milazzo et al. 2011), others appear to be non-pathogenic in humans (Wulff et al. 1977; Günther et al. 2009; Ishii et al. 2012; Gryseels 2015). The latter ones can be used to study ecological and epidemiological patterns that can be relevant for the management of their pathogenic relatives. They also provide an interesting model for the study of host-pathogen evolutionary relationships. In East Africa, different arenaviruses occur in different rodent species and even in different clades of the Natal Multimammate Mouse, *Mastomys natalensis*, and the African Pygmy Mouse, *Mus minutoides* (Walker et al. 1975; Goüy de Bellocq et al. 2010; Ishii et al. 2012; Gryseels et al. 2017)

2.1. Arenaviruses

For over forty years, the *Arenaviridae* family consisted of a single genus, *Arenavirus* (Radoshitzky et al. 2015). Based on phylogenetic differences, antigenic properties and geographical distribution, this genus was further divided into two clades, the Old World arenaviruses and the New World

arenaviruses (Radoshitzky et al. 2015). The Old World arenaviruses occur in rodents of the subfamily *Murinae* in Africa and Eurasia, except for Lymphocytic choriomeningitis virus which occurs worldwide due to the distribution of its host, the house mouse. The New World arenaviruses occur in rodents of the *Cricetidae* family in the Americas, except for Tacaribe virus which was isolated from bats. However, Stenglein et al. (2012) discovered that arenaviruses also infect snakes. A new genus, *Reptarenavirus,* was therefore established to accommodate these viruses and the old genus *Arenavirus* was renamed *Mammarenavirus* (Radoshitzky et al. 2015). As most of the recent literature still refers to mammarenaviruses by the more general term 'arenaviruses', I will do the same unless specifically mentioned.

Arenavirus genomes are made up of two single-stranded RNA segments: the large (L) segment (± 7.2 kb) and the small (S) segment (± 3.5 kb) (Charrel and de Lamballerie 2002). Each segment comprises two genes in non-overlapping reading frames that are read in opposite orientation (Charrel and de Lamballerie 2002) (Figure 1). The L segment contains the Z and *L* genes, which are translated by the host into the Z and L proteins, respectively. The S segment contains the *NP* and *GPC* genes, which encode the nucleoprotein (NP) and the glycoprotein precursor (pre-GPC), respectively. The Z protein or zinc-binding protein is the smallest arenavirus protein. It regulates viral RNA synthesis, viral assembly and budding, interacts with host cell proteins and inhibits host interferon activity (Fehling et al. 2012). The L protein is the largest arenavirus protein. It is an RNA-dependent RNA polymerase that catalyses viral transcription and replication in a ribonucleoprotein complex (Singh et al. 1987; Kranzusch and Whelan 2012). This ribonucleoprotein complex or nucleocapsid is formed through association with nucleoproteins. The complex encloses the genome segments and is enveloped by lipids with glycoprotein spikes (Eichler et al. 2003; Perez et al. 2003; Pinschewer et al. 2003). The glycoprotein precursor is post-translationally cleaved into the stable signal peptide (SSP) and the glycoprotein (GPC) (Eichler et al. 2003). The glycoprotein (GPC) is then further cleaved into two subunits: GP1, which binds to host transmembrane proteins, and GP2, which fuses the viral envelope with host cell membranes (Eichler et al. 2003; Günther and Lenz 2004). The SSP is also involved in fusion with host cell membranes and promotes GP1-GP2 cleavage (Messina et al. 2012).

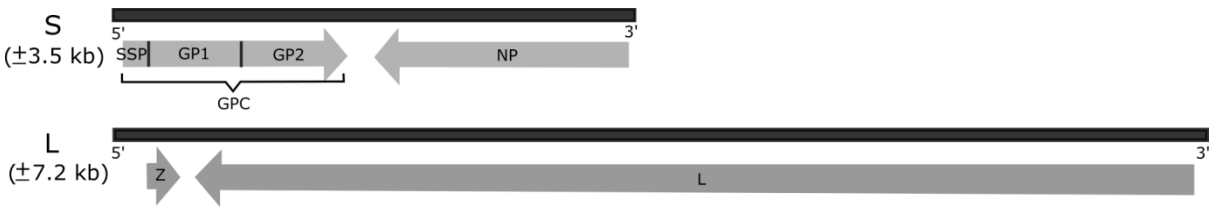


Figure 1: The bisegmented mammarenavirus genome structure.

The diversity of arenaviruses is large and we are only just beginning to unveil it. So far at least 55 mammarenaviruses and 5 reptarenaviruses have been discovered, more than half of these only in the last decade (Gryseels 2015). At the root of their diversity lies their high frequency of transcription errors (Zapata and Salvato 2013), because RNA dependent RNA polymerases of RNA viruses do not proofread and therefore incorporate approximately one mistake per genome replication (Holmes 2003). Reassortment (exchange of genomic segments) and recombination within segments could have contributed further to the current diversity (Zapata and Salvato 2013). Its traces have been observed among natural arenavirus infections with Lassa virus (Andersen et al. 2015) and it might even have given rise to a certain clade of New World arenaviruses (Charrel et al. 2002, however, also see Cajimat et al. 2011). Furthermore, reassortment and recombination of reptarenaviruses are widespread in captive snakes (Stenglein et al. 2015).

Arenaviruses are usually specific to a single rodent host species. First of all, most arenaviruses have only been detected in one species (Salazar-Bravo et al. 2002; Gryseels 2015) and if a virus is found in several species, one species is likely the true reservoir, while other murid species and humans are infected through spill-over (Fulhorst et al. 1999; Fulhorst et al. 2002; Mills et al. 1994). Secondly, distinct arenaviruses are carried by sympatric species (Fulhorst et al. 1999; Goüy de Bellocq et al. 2010; Ishii et al. 2012; Gryseels 2015). In other words, they occur alongside each other in other murids and thus appear to have the opportunity to switch hosts, but have not been observed to do so. Thirdly, closely related arenaviruses are often carried by closely related host species. As a result phylogenetic trees of arenaviruses match the phylogenetic trees of their hosts quite well, apart from a number of past host switches (Gryseels 2015; Irwin et al. 2012).

This match of virus and host trees could either be explained by co-speciation or by preferential host switching (Bowen et al. 1997; Irwin et al. 2012; Gryseels 2015). In the first case ancestral mammarenaviruses co-diverged with their hosts several million years ago. In the latter case ancestral mammarenaviruses spread more recently by jumping hosts and more easily so to hosts that were closely related. However, it is currently not possible to settle which is the case due to a lack of reliable estimates of early divergence times (Gryseels 2015). The problem is that divergence times for the deeper sections of RNA virus phylogenetic trees are likely underestimated by commonly used substitution models because small RNA genomes are subjected to strong purifying selection and are quickly saturated with neutral mutations (Duchêne et al. 2014; Gryseels 2015).

2.2. *Mastomys natalensis*-borne arenaviruses

Mastomys natalensis (Smith, 1834), the Natal Multimammate Mouse, is found throughout sub-Saharan Africa except in deserts, dense forests and very high mountainous areas (Coetzee 1975; Colangelo et al. 2013). Based on mitochondrial *cytochrome b* (*cyt b*) data, the species can be divided into six distinct clades: A-I in West Africa from Senegal to Nigeria; A-II in West and Central Africa from Niger to the Democratic Republic of Congo; A-III in East Africa, Kenya; B-IV in East Africa in Kenya, Tanzania and Rwanda; B-V in East Africa, Tanzania and B-VI in East and Southern Africa from Tanzania to South Africa (Figure 2 Top) (Colangelo et al. 2013). These lineages likely diverged from each other in isolated refugia when forests spread during climate fluctuations around one million years ago (Colangelo et al. 2013).

Several Old World arenaviruses have been detected in *M. natalensis*. Five of these viruses are not known to be pathogenic to humans, but a sixth, Lassa virus, is estimated to cause between 100,000 and 300,000 fever cases each year, resulting in about 5,000 deaths (CDC 2015). In contrast to the wide distribution range of *M. natalensis*, Lassa fever cases only occur in a few countries in West Africa, corresponding more or less to the distribution range of the *M. natalensis* A-I mitochondrial lineage. Moreover, the non-pathogenic arenaviruses also appear restricted to certain *M. natalensis* mitochondrial lineages: a Mobala-like virus in the A-II (Olayemi et al. 2016a), 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 Luna (Ishii et al. 2011, 2012) and Mopeia virus (Wulff et al. 1977) in the B-VI lineage (Figure 2 Top). *M. natalensis* arenaviruses are therefore likely specific to intraspecific lineages rather than to the species as a whole (Gryseels et al. 2017).

Gryseels et al. (2017) tested this hypothesis across a transect in Tanzania where the B-IV and B-V mitochondrial lineages come into contact. Using a mitochondrial *cyt b* marker, a single nucleotide polymorphism (SNP) on the Y chromosome and nuclear microsatellite markers, they showed that the B-IV and B-V mitochondrial lineages correspond to taxa that are distinct genome-wide and that these meet in a narrow hybrid zone which does not coincide with a river, road or a change in land cover. Both Gairo and Morogoro virus occur at the centre of the hybrid zone at the locality Berega (locality C on Figure 2 Bottom), but Gairo virus and Morogoro virus are only detected in B-IV and B-V individuals, respectively. Neither thus appears to have spread to the other taxon despite the close physical contact (Figure 2 Bottom). The taxa and/or their viruses could have only recently arrived at this boundary, so that the arenaviruses have not had time to spread into the other host range. Human migration might have facilitated recent contact as the transect is situated along a busy road between Tanzania's largest city, Dar es Salaam, and the capital, Dodoma. However, the climate has

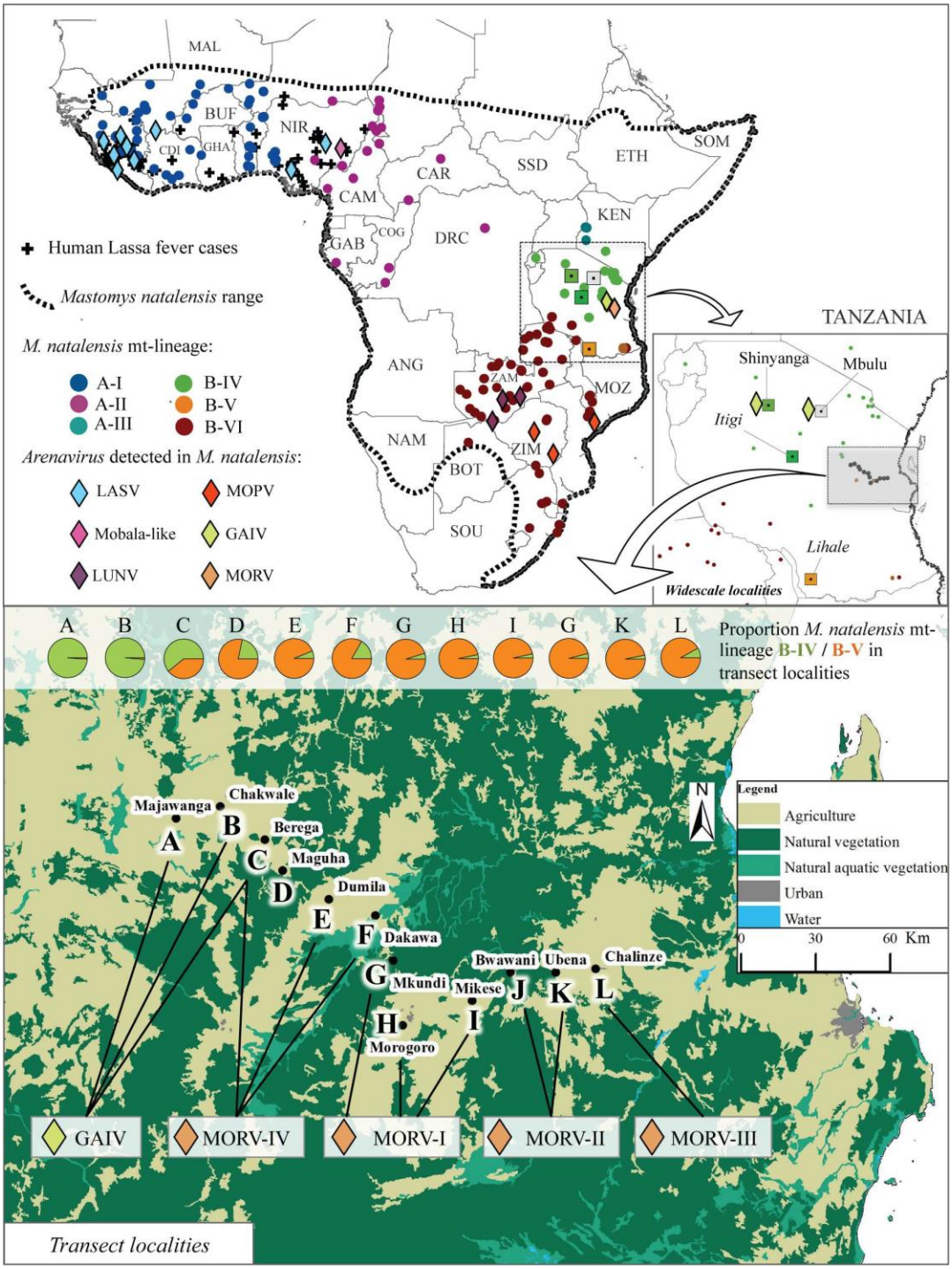


Figure 2: Distribution of *Mastomys natalensis* mitochondrial lineages and their arenaviruses. Top: Occurrence of *M. natalensis* mitochondrial lineages (circles), reported occurrences of *M. natalensis*-borne arenaviruses (diamonds; LASV = Lassa virus, LUNV = Luna virus, MOPV = Mopeia virus, GAIV = Gairo virus, MORV = Morogoro virus) and georeferenced human Lassa fever cases (crosses). Inset: Zoom-in on Tanzania with localities sampled in Gryseels et al. (2017). Bottom: Zoom-in on the transect across the [B-IV, B-V] hybrid zone. Pie charts represent the proportion of B-IV and B-V individuals at each locality. Presence of Morogoro and Gairo virus is visualised with diamonds and Roman numerals indicate different Morogoro virus clades (Figure from Gryseels et al. 2017).

been quite stable for at least 5,500 years, making it unlikely that contact between the *M. natalensis* taxa is that recent due to human activity (Gryseels et al. 2017). The distribution of Gairo and Morogoro virus is thus most likely mediated by the host genotype.

Within the distribution range of *M. natalensis* mitochondrial lineages, arenavirus distribution is likely further mediated by environmental variables, probably acting through host density regulation. Lassa virus prevalence can vary considerably among neighbouring villages (Demby et al. 2001), regions (Safronetz et al. 2013), and countries (Mylne et al. 2015). Lassa virus was long thought to be endemic only to Nigeria and Guinea/Sierra Leone/Liberia (Safronetz et al. 2010; Sogoba et al. 2012; Mylne et al. 2015). However, over the last decade several human and rodent Lassa virus infections have been reported in the countries in between (Safronetz et al. 2010, 2013; Dzotsi et al. 2012; Kouadio et al. 2015; Sogoba et al. 2016), including a Lassa fever outbreak in Benin (N'koué Sambiéni et al. 2015). Lassa virus is thus not absent from these countries, though it still appears to be less widespread there than in Nigeria, Guinea, Sierra Leone and Liberia (Coulibaly-N'Golo et al. 2011; Safronetz et al. 2013; Kronmann et al. 2013). However, studies investigating the distribution of Lassa virus in West Africa are almost exclusively based on or biased towards reported human Lassa fever cases and few studies sample rodents in more than a handful of villages at a large spatial scale. It is therefore unclear whether the lower prevalence is caused by a lower Lassa virus prevalence in *M. natalensis*, a lower *M. natalensis* density, a lower transmission to humans (e.g. due to lower virulence, environmental factors, a lower contact rate with rodents…) or even misdiagnosis and underreporting of human cases.

Based on (the limited) available data, Mylne et al. (2015) modelled Lassa virus distribution in West Africa. Their model indicates that vegetation, night temperature, elevation and modelled *M. natalensis* habitat suitability might predict environmental suitability for Lassa virus. However, they did not take into account that sampling intensity is unequally distributed throughout West Africa, nor that while positive sampling results can be interpreted as virus presence, negative sampling results do not necessarily mean the virus is truly absent. However, both could greatly influence the model's results (Peterson et al. 2014).

Environmental suitability for *M. natalensis* and its arenaviruses could also be expected to influence genetic structure of *M. natalensis* and its arenaviruses within the distribution range of *M. natalensis* lineages. Environmental barriers could constrain *M. natalensis* dispersal, and thus gene flow, resulting in spatial genetic structure within a lineage. As arenaviruses depend on their hosts for dispersal, and thus gene flow, and as they appear to be tightly associated with *M. natalensis* between-lineage genetic structure, intraspecific arenavirus genetic structure might match *M.*

natalensis within-lineage spatial genetic structure to some extent. Gryseels et al. (2016, 2017) and a Master student from last year (Locus 2016) explored spatial clustering of Morogoro virus and *M. natalensis* B-V sequences from ten localities across a transect in central Tanzania. They observed four Morogoro virus clades linked to one, two or three adjacent localities (Figure 2 Bottom). They looked for potential environmental barriers relating to land use, vegetation, soil, elevation, precipitation, rivers and roads, but were not able to identify any that could be responsible for the spatial genetic structure (Locus 2016; Gryseels et al. 2017). Furthermore, Morogoro virus spatial genetic structure does not match *M. natalensis* B-V spatial genetic structure (Gryseels et al. 2017). Microsatellite markers indicate that suburban Morogoro city samples are different from samples from the nine surrounding rural localities (Gryseels et al. 2016), while Morogoro virus sequences from Morogoro form a clade together with sequences from the adjacent localities on both sides. Rather than host genetic substructure or investigated environmental elements, the Morogoro virus spatial genetic structure appears to show a pattern of isolation by distance (Gryseels et al. 2017).

2.3. *Mus minutoides*-borne arenaviruses

Mus (Nannomys) minutoides Smith, 1834, the African Pygmy Mouse, is widespread throughout sub-Saharan Africa, except in deserts and continuous forest areas in the Congo Basin (Bryja et al. 2014). It is probably the rodent with the largest distribution range in Africa, surpassing even *Mastomys natalensis* (Bryja et al. 2014). Across its distribution range 11 distinct mitochondrial clades can be distinguished (Figure 3). These clades appear to have diverged about a million years ago during the same climate fluctuations as the *M. natalensis* mitochondrial clades, resulting in similar genetic distances among the clades of both species (Bryja et al. 2014).

As for *M. natalensis*, there is evidence that different mitochondrial clades carry distinct arenaviruses. In Guinea, Kodoko virus has been detected in two *M. minutoides* individuals (Lecompte et al. 2007) from the 'West African clade' (W, sensu Bryja et al. 2014), which occurs from Guinea to Ghana. In Zambia, Lunk virus has been isolated (Ishii et al. 2012) from a 'Zambia and surrounding clade' individual (ZA, sensu Bryja et al. 2014). This clade is situated in Zambia, Botswana, the Democratic Republic of Congo and South Africa. A third virus, Ngerengere virus, has been detected in Tanzania (Goüy de Bellocq et al. 2010; Gryseels 2015) in three individuals from the 'South East Africa clade' (SE, sensu Bryja et al. 2014), which stretches from south Kenya to eastern South Africa. However, the genetic information available for Ngerengere virus is too limited to determine if it is a distinct virus species from Lunk virus.

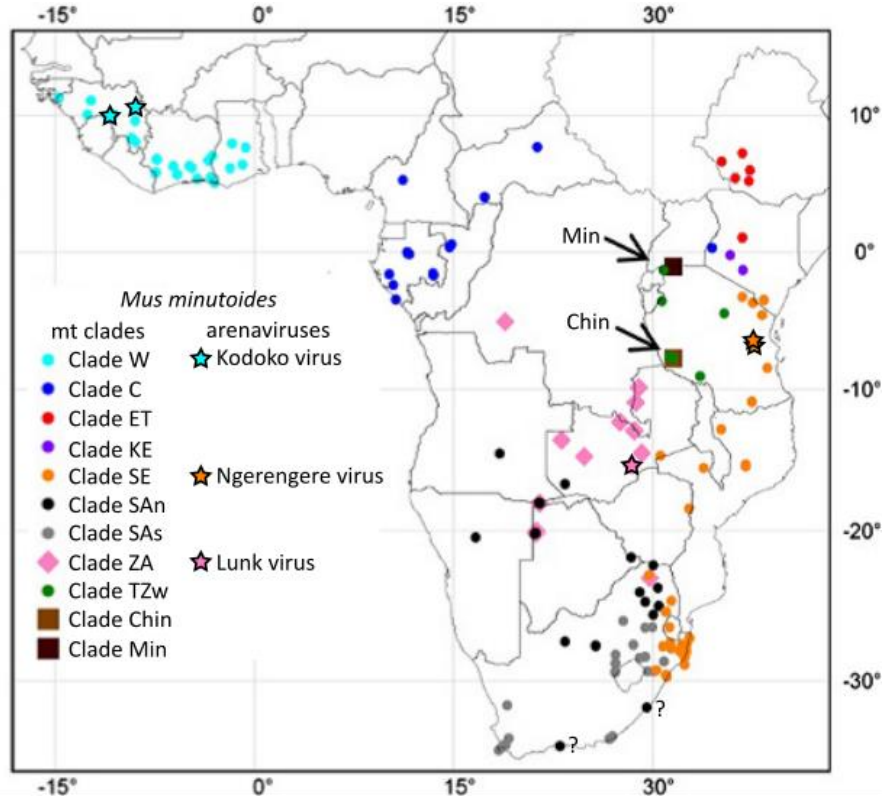


Figure 3: Distribution of *Mus minutoides* mitochondrial lineages and arenaviruses. Doubtful records based on genotyping of old museum samples are indicated by question marks. (Figure adapted from Bryja et al. 2014).

2.4. Aims and objectives

The main objective of this Master thesis was to investigate the distribution, specificity and genetic structure of Tanzanian *Mastomys natalensis* arenaviruses at a larger scale than the transect of Gryseels et al. (2017). For this purpose, samples from localities spanning a strip of approximately 800 km from the south west to the north east of Tanzania were screened for arenaviruses. These localities are situated in and around the contact zone between the [B-IV, B-V] and between the [B-IV, B-V, B-VI] lineages.

If the B-IV and B-V lineages or Gairo and Morogoro virus only arrived recently at their boundary in Gryseels et al. (2017) and the pattern of arenavirus specificity was observed because they lacked the time to spread into the range of the other, then the strict specificity in Gryseels et al. (2017) (Figure 2 Bottom) would be a special case and this specificity or even arenaviruses will not be observed close to the [B-IV, B-V] contact zone at a larger geographic scale. Conversely, if the B-IV and B-V lineages and their arenaviruses came into contact some time ago and if arenaviruses are strictly specific to *M. natalensis* host lineages, then strict specificity will be found across the entire host lineage range. Moreover, if arenaviruses are specific to *M. natalensis* host lineages, an arenavirus observed in the

B-VI lineage should also show a strict association with that lineage. This arenavirus could be Luna and/or Mopeia virus, as both have been detected in B-VI individuals.

Furthermore, I expected that Gairo virus prevalence and genetic structure would be similar to those reported for Morogoro virus in Gryseels et al. (2017). Similar prevalence and genetic structure might indicate similar dynamics. Indeed, so far Morogoro, Gairo and Lassa virus appear to share many ecological features. Reported RNA and antibody prevalence are similar for Gairo, Morogoro, Lassa and Luna virus (Fichet-Calvet et al. 2007; Borremans et al. 2011; Ishii et al. 2011, 2012; Gryseels et al. 2015). Antibody prevalence tends to be relatively high for very young individuals, decline steeply as they grow and then increase linearly with age (Demby et al. 2001; Borremans et al. 2011; Fichet-Calvet et al. 2014; Gryseels et al. 2015). In agreement with this pattern, it is assumed that these viruses are mainly transmitted horizontally and that maternal antibodies are likely passed on. Furthermore, Morogoro and Lassa virus inoculations appear to cause acute infections in adult *M. natalensis*, but chronic infections in neonates (Walker et al. 1975; Borremans et al. 2015) and natural Morogoro, Gairo and Lassa virus infections do not seem to negatively affect *M. natalensis* (Mariën et al. 2017). Apart from its ability to spill over, Lassa virus thus appears to exhibit similar dynamics as non-pathogenic *M. natalensis* arenaviruses. Research on their distribution, prevalence and genetic structure might therefore help to understand the current and future distribution range of Lassa virus.

Additionally, a small number of *Mus minutoides* were screened for arenaviruses. Like *Mastomys natalensis*, *Mus minutoides* has a pan-sub-Saharan-African distribution range, a strong geographic intraspecific structure and multiple arenaviruses. It would therefore be interesting to compare the evolution of this system to the evolution of *Mastomys natalensis* and its associated arenaviruses. If *Mus minutoides* arenaviruses are also specific to certain mitochondrial host lineages, then different arenaviruses will be carried by distinct lineages as well.

3. Materials and methods

3.1. Trapping and sampling

Mastomys natalensis individuals were trapped during the summers of 2015 and 2016 by the University of Antwerp, the Institute of Vertebrate Biology of the Czech Academy of Sciences (IVB; Studenec, Czech Republic) and the Pest Management Centre of the Sokoine University of Agriculture (SPMC; Morogoro, Tanzania). They were trapped in the south west of Tanzania, in the putative

three-way [B-IV, B-V, B-VI] contact zone and in the [B-IV, B-V] contact zone. In the [B-IV, B-V] contact zone extra localities on the B-IV side of the transect from Gryseels et al. (2017), a new transect about 100 km to the north of this transect and several localities around these transects were sampled. As *M. natalensis* is mostly active during the night (Delany 1964; Coetzee 1975), Sherman live traps (H.B. Sherman Traps Inc., Tallahassee, USA) and snap traps baited with a mixture of peanut butter, maize flour and dried fish were set out in the late afternoon and collected during the morning.

A total of 1445 *M. natalensis* were caught at 50 localities spanning a strip of about 800 km from the south west to the north east of Tanzania (see Supplementary Table 1). 546 *M. natalensis* were caught in June 2015 and June to early July 2016 by Czech-Tanzanian research teams. In mid-July 2016 I assisted a Czech-Tanzanian research team in trapping 322 *M. natalensis* at 10 localities. Subsequently I led a Tanzanian research team, catching 577 *M. natalensis* at one previously trapped locality and five new localities from late July to August 2016. In addition to *M. natalensis* several other rodents and shrews were caught, including 21 *Mus minutoides* individuals spread over 11 localities (see Supplementary Table 2). Species were identified in the field based on morphological characteristics. When in doubt, the Mammals of Tanzania Rodentia Skin key (The Field Museum, Chicago, USA 2016) was consulted. To confirm the species identification in the field, A. Hánová from the IVB sequenced a 1140 bp fragment of the *cyt b* gene for a subset of individuals including all arenavirus-positive *M. natalensis* and *M. minutoides*.

Mice caught in live traps were euthanized by cervical dislocation prior to dissection (Directive 2010/63/EU). Spleens for host mitochondrial genotyping were preserved in 96% ethanol and stored at -20 °C, while 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 Serobuvard filter papers that absorb about 10 to 15 μL per spot (LDA22 Zoopole, Ploufragan, France).

3.2. RNA Extraction

RNA was extracted from kidneys from up to 50 individuals per locality using the Nucleospin RNA Isolation kit (Macherey-Nagel, Düren, Germany). As viral RNA screening with specific primers does not require DNA digestion, the membrane desalting and DNA digestion steps were omitted. Otherwise, the manufacturer's protocol was followed. If more than 50 kidney samples were available, a random selection was made, except for choosing kidneys from individuals caught alive in Sherman live traps over those from individuals caught dead in snap traps. In order to reduce the

time and cost of screening, *M. natalensis* kidney samples were pooled by three. In case of a positive band after gel electrophoresis (see 3.3), the three kidneys making up a pooled sample were extracted and screened separately to determine which kidney(s) contained arenavirus RNA. Due to their smaller kidney size, *M. minutoides* kidneys were not pooled, but extracted separately from the start. A total of 1155 *M. natalensis* and 21 *M. minutoides* kidney samples were extracted in this way and their extractions were stored at -20 °C and -80 °C for short and long term storage, respectively.

3.3. Arenavirus *L* gene RNA screening

RNA extractions were screened with a Reverse Transcription Polymerase Chain Reaction (RT-PCR) using the SuperScript One-Step RT-PCR System kit (Invitrogen, Carlsbad, USA). 4.5 µl of template RNA was added to 15.5 µl master mix consisting of 10 µl 2X Reaction Mix, 0.3 µl magnesium sulfate, 0.4 µl Superscript II RT/Platinum Taq Mix, 0.05 µl of RNase free water and 0.8 µl of each of the following primers: MoroL3359-forward, LVL3359D-plus, LVL3359G-plus, MoroL3753-reverse, LVL3754A-minus and LVL3754D-minus (see Table 1). These primers target a 340 nt fragment of the *L* gene. The LVL primers were designed by Vieth et al. (2007) to bind to regions that are very conserved among the Old World arenaviruses that had been discovered up to that time (including Mopeia virus). They have been shown to detect other Old World arenaviruses discovered since then including Morogoro virus (Günther et al. 2009), Gairo virus (Gryseels et al. 2015) and Luna virus (Gryseels 2015). Morogoro virus, however, is detected with low sensitivity by these primers. Therefore MoroL primers specific to Morogoro virus were designed by Günther et al. (2009). The reactions were run with the same temperature profile and number of cycles as in Vieth et al. (2007): reverse transcription for 30 min at 50 °C; initial denaturation and Platinum Taq activation for 2 min at 95 °C; 45 cycles of denaturation for 20 s at 95 °C, annealing for 30 s at 55 °C and extension for 1 min at 72 °C; and final extension for 10 min at 72 °C. In each RT-PCR a positive control (a known positive sample) and a negative control (RNase free water instead of template RNA) were included to validate the assay.

Next, RT-PCR products were verified by gel electrophoresis. DNA was visualised on a 1.4% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium, Hayward, USA). In case of a band between 300 and 400 bp positive, pooled samples were depooled and positive single samples were sent for purification and Sanger sequencing in both directions at the Genetic Service Facility (GSF) of the Vlaams Instituut voor Biotechnologie (VIB, Antwerp, Belgium).

Table 1: Overview of the primers used in PCRs to amplify a portion of arenavirus *L*, *NP* and *GPC* genes.

Primer	Sequence	Target gene	Target fragment length in nt (of which coding)	Reference
LVL3359D-plus	AGAATCAGTGAAAGGGAAAGCAAYTC	*L*	340 (340)	Vieth et al. 2007
LVL3359G-plus	AGAATTAGTGAAAGGGAGAGTAAYTC			
LVL3754A-minus	CACATCATTGGTCCCCATTTACTATGRTC			
LVL3754D-minus	CACATCATTGGTCCCCATTTACTGTGRTC			
MoroL3359-forward	AGGATTAGTGAGAGAGAGAGTAATTC			Günther et al. 2009
MoroL3753-reverse	GACCATAGTAAGTGGGGCCCAATGATGT			
OWS2805-fwd	GTCAGGCTTGGCATTGTCCCAAACTGRTTRTT	*NP*	531-536 (513 or 516)	Ehichioya et al. 2011
OWS2810-fwd	CTTGGCATTGTCCCAAACTGRTTRTT			
OWS3400-rev	CGCACAGTGGATCCTAGGCTATTKGATTGCGC			
OWS3400A-rev	GCGCACAGTGGATCCTAGGC			
OWS0001-fwd	GCGCACCGGGGATCCTAGGC	*GPC*	953-972 (906 or 912)	Ehichioya et al. 2011
OWS1000-rev	AGCATGTCACAGAAYTCYTCATCATG			

3.4. Antibody screening

As no *L* gene RNA was detected at any locality within a strip of about 350 km from south west to central Tanzania (see Results), arenavirus presence in this region was further assessed by screening dried blood samples for IgG mouse antibodies specific for Old World arenaviruses. Up to 50 dried blood samples per locality were screened, adding up to 540 samples from 17 localities. If more than 50 dried blood samples were available, a random selection was made except for choosing dried blood samples from individuals caught alive in Sherman traps over those from individuals caught dead in snap traps. This selection was independent from the kidney sample selection. For efficiency, dried blood samples were pooled by two. Then blood samples from positive pooled samples were tested separately.

Anti-arenavirus antibody presence was tested with an indirect immunofluorescence assay (IFA) as in previous studies (Günther et al. 2009; Gryseels et al. 2015). Dried blood spots were eluted overnight at 4 °C in 200 µL or 100 µL of phosphate-buffered saline (PBS) for pooled and single samples, respectively. A few dried blood samples did not elute well in the PBS, as indicated by their transparency instead of a yellow to brown colour. This can happen due to suboptimal sampling,

transportation or storage conditions, but can be remedied by adding 0.2% ammonium (1.6 μL or 0.8 μL of 25% NH_3 for pooled or single samples, respectively) as advised by Borremans (2014). After 5 hours these samples had eluted enough to resume the protocol. 10 μL of each elution was pipetted on wells of slides coated with Vero cells infected with Morogoro virus (Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany). A positive control (a known positive elution sample) was added on every slide and a negative control (PBS only) on every five slides. After an incubation step of one hour at 37 °C, slides were washed thrice with PBS for 5 min. When the slides had dried, 10 μL of 1:100 rabbit anti-mouse IgG antibodies was added to each well. These secondary antibodies were conjugated with fluorescein isothiocyanate (FITC) for visualisation under a fluorescence microscope. Next, the slides were incubated again for one hour at 37 °C and washed thrice with PBS for 5 min. When the slides had dried, 3 μL of glycerol with DABCO was added to each well to delay fading of secondary antibodies. Lastly, wells were verified for fluorescent antibodies under a fluorescent microscope with blue LED light (480 ± 30 nm) at 10 x 40 magnification. In case of doubt, the well was checked by Joachim Mariën, who is experienced in this assay.

3.5. Additional *L* gene screening and *GPC* and *NP* gene screening

In case anti-arenavirus antibodies were detected in localities where no *L* gene RNA was detected, additional kidney samples were extracted if available and screened for the *L* gene in the same way as described in 3.2. and 3.3. Furthermore, all pooled kidney samples from that locality were screened with primers targeting the *GPC* and *NP* gene to detect virus strains that might not have annealed with the primers used in the *L* gene screening. The reaction was performed similarly to the *L* gene screening above in 3.3., but with 0.8 μl of each of the following primers: OWS0001-fwd and OWS1000-rev; and OWS2805-fwd, OWS2810-fwd, OWS3400-rev and OWS3400A-rev (see Table 1). The former primer pair targets the first part of the *GPC* gene; the latter pairs target a fragment of the *NP* gene.

3.6. *GPC* and *NP* amplification

For all kidneys positive for arenavirus *L* gene RNA, parts of the *GPC* and *NP* genes were amplified as well. These genes were amplified in separate PCRs with the OWS primers mentioned in 3.5. and Table 1 and conditions set as in 3.3. As for the *L* gene, *GPC* and *NP* gene amplicons were purified and Sanger sequenced in both directions at the GSF of the VIB.

3.7. Arenavirus genetic analyses

Raw sequence data was imported into Geneious R8.1 (Biomatters, New Zealand 2015). Forward and reverse sequences were aligned, manually edited and the primer regions were cut. The resulting consensus sequences were 340 nt long for the *L* gene, 531-536 nt long for the *NP* gene and 953-972 nt long for the *GPC* gene (see Table 1). Subsequently, the sequences were aligned with annotated sequences of the same virus species from GenBank using the Geneious alignment algorithm. The non-coding regions were cut (for the *NP* and *GPC* sequences). As a result the *NP* and *GPC* gene sequences were 513 or 516 nt and 906 or 912 nt long, respectively (Table 1). Next, these coding sequences were aligned with other Old World arenavirus sequences using the translation alignment option with the Geneious alignment algorithm for protein alignment and the BLOSUM62 substitution matrix. These sequences included a sequence of each published African Old World arenavirus species (a full segment sequence if available); all partial sequences of Luna, Morogoro and Gairo virus deposited in GenBank; unpublished Morogoro virus sequences (Locus 2016) and unpublished Luna virus and Ngerengere virus sequences (Gryseels 2015).

Phylogenetic trees were inferred separately for the three genes using Bayesian inference (BI) and Maximum likelihood (ML) as implemented in MrBayes v3.2.6 (Ronquist et al. 2003) and RaxML v8 (Stamatakis 2014), respectively, in the CIPRES web portal (Miller et al. 2010). As the glycoprotein precursor (pre-GPC) is post-translationally cleaved into three different peptides with different functions, the *GPC* gene sequences were partitioned into these three parts in both tree building methods. Moreover, sequences of the three genes were partitioned according to codon position because mutations at a different codon position do not have the same effect on the corresponding amino acid translation. For example, mutations of the third codon position are often synonymous, resulting in the same amino acid, while mutations of the first codon position are not.

During BI the General Time Reversible (GTR) nucleotide substitution model was used as selected for the data by jModelTest v2.0 (Guindon and Gascuel 2003; Darriba et al. 2012). In this model a separate rate is estimated for each type of interchange between bases (Tavare 1986). The model test further recommended to implement models with a proportion of invariable sites and with a gamma distributed variation in substitution rates among sites (Yang 1993) to account for site-dependent variation. This gamma distributed variation was implemented over four categories. The branch lengths were not constrained (i.e. there were no molecular clock priors), allowing different branches of the tree to evolve at different rates. In order to improve mixing and thus speed up Markov Chain Monte Carlo convergence, Metropolis coupling with three heated and one cold chain was applied. In two independent runs the chains ran for 15 or 20 million generations for the *L* and

NP gene and for the *GPC* gene analysis, respectively. The cold chain was sampled every 500 generations after discarding the first 25% as burn-in. The effective sample size (ESS) and the trace pattern of the substitution model parameters were checked in Tracer v1.6 (Rambaut et al. 2014). The ESS of a given parameter estimates how many independent samples the output of the analysis represents. These numbers should therefore be sufficiently high (as a rule of thumb at least 200) to assess if the posterior probability distribution was sampled adequately. Adequate sampling was further assessed by checking trace patterns for normal mixing behaviour.

As in the BI analyses, the GTR substitution model was used in the ML analyses. However, no gamma distributed variation with a proportion of invariable sites was implemented because it is not recommended to do so in RAxML (Stamatakis 2016). In order to determine branch support, 1000 bootstrap samples were simulated. Output trees were visualised in FigTree (Rambaut 2012) with Lujo virus as outgroup because it is basal to other Old World arenaviruses (Briese et al. 2009).

3.8. *Mastomys natalensis* and *Mus minutoides genetic* analyses

Cyt b sequences from arenavirus-positive mice were obtained from A. Hánová (IVB) and imported into Geneious R8.1. They were aligned with a sequence from each *Mastomys natalensis* and *Mus minutoides* lineage from Colangelo et al. (2013) and Bryja et al. (2014), respectively, and were assigned to one of these lineages based on their position in a Maximum likelihood phylogenetic tree. This tree was constructed in RAxML in the CIPRES web portal with a GTR substitution model and 1000 bootstrap trees.

3.9. Analyses of regional differences in *Mastomys natalensis* arenavirus detection

A G-test (Woolf 1957) was carried out to investigate differences in arenavirus RNA detection level between the south west and the north east of Tanzania and between different arenavirus species. For this purpose screening data were supplemented with Morogoro and Gairo virus data from Gryseels et al. (2017) and from Locus (2016). From Gryseels et al. (2017) 1077 dried blood samples from 15 localities were analysed. They were initially pooled by two and screened for *L* gene RNA in two PCRs, one with the LVL and one with the MoroL primers described in Table 1. Locus (2016) screened 619 kidney samples from 5 localities. These samples were also initially pooled by two, but were only screened with the MoroL and not with the LVL primers. Northeastern localities were split into a Gairo virus and a Morogoro virus group, except for one locality from Gryseels et al. (2017) (Berega, locality C in Figure 2 Bottom) which was not included in the test because both viruses were

detected here. Southwestern and central localities, however, could not be split according to virus species, because viral RNA was not detected at most localities. The G-test thus compared prevalence among northern Gairo localities, eastern Morogoro localities, and southwestern and central localities. Not all localities could be assigned to a single fixed group, because no arenavirus RNA was detected. Therefore, the G-test was repeated 15 times with varying classification of these localities by drawing different straight lines between them as geographic boundaries.

A second G-test was performed on the antibody data that was available from the same localities as those from the RNA G-test. It was also repeated 15 times with the same classifications, but it only tested the difference in prevalence between the Morogoro virus and the southwestern and central group. The Gairo virus group was not included because the available antibody data originated from only two to four localities (depending on the classification). The used data consisted of 540 dried blood samples from 17 localities in this study, 306-444 dried blood samples from 2-3 localities from Locus (2016) and 710-732 dried blood samples from 8-9 localities from Gryseels et al. (2017) which were not published in this study, but in Gryseels et al. (2015) and Mariën et al. (2017). As IFA-positive dried blood samples from an antibody-positive locality were not depooled in Locus (2016), the number of positive single samples was estimated from the number of positive pooled samples with the following formula:

$p + n = (p + n)^2 = p^2 + 2\,pn + n^2 = 1$

with p = proportion of positive single samples

n = proportion of negative single samples

$p^2 + 2\,pn$ = proportion of positive pooled samples

(a pooled sample is positive if at least one of its constituting samples is positive)

n^2 = proportion of negative pooled samples

(a pooled sample is negative if both of its constituting samples are negative)

The proportion of negative single samples ‘n’ can then easily be calculated by taking the square root of the proportion of negative pooled samples ‘n^2’ and the proportion of positive single samples is simply equal to one minus this proportion. In this way it was estimated that the 16 IFA-positive pooled samples in Locus (2016) likely correspond to 17 IFA-positive single samples.

The G-tests were performed using the RVAideMemoire package (Hervé 2017) in R 3.3.2 (R Development Core Team 2017). For the RNA G-test, G-test repetitions with a significant outcome, set at $P < 0.05$, were further examined with pairwise G-tests from the same package. These pairwise tests used a Bonferroni correction for multiple testing.

4. Results

4.1 Arenavirus RNA and anti-arenavirus antibody detection

Out of 21 *Mus minutoides* kidneys, one sample from Ngana tested positive for arenavirus *L* gene RNA (Supplementary Table 2). The nucleotide sequence and corresponding amino acid translation were compared to Ngerengere and Lunk virus sequences and corresponding translations available from Goüy de Bellocq et al. (2010), Ishii et al. (2012) and Gryseels (2015). The new sequence differed from Ngerengere virus in only one or two and from Lunk virus in four out of 98 amino acids. Nucleotide pairwise comparisons are summarized in Table 2.

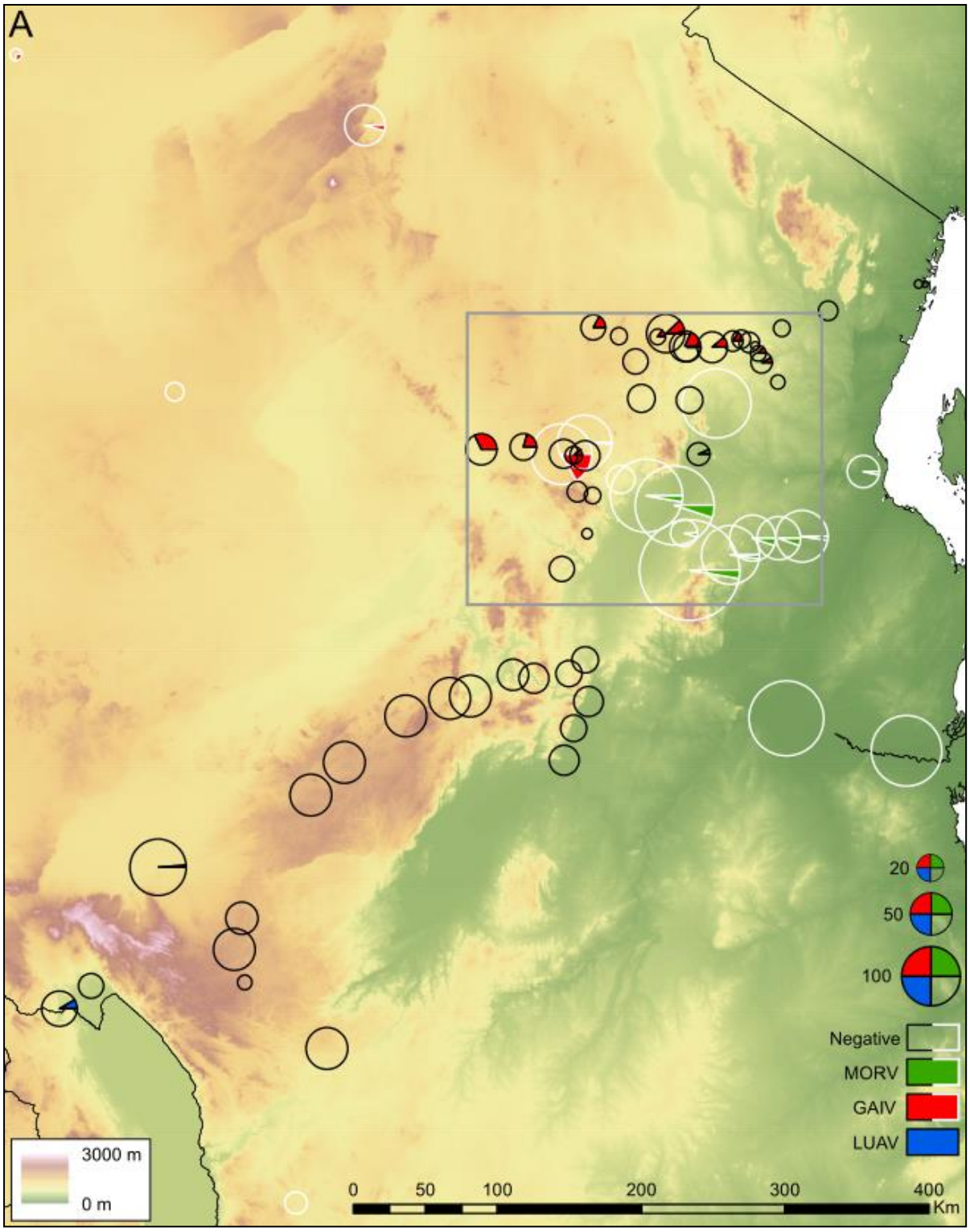
Table 2: Nucleotide pairwise identities for the *Mus minutoides* virus *L* gene sequence and available *L* Ngerengere and Lunk virus sequences. Numbers between brackets after the horizontal header indicate sequence length in nt.

	M. minutoides virus TZ28088 (294) (this study)	NGEV TZ22285 (340) (Goüy de Bellocq et al. 2010)	NGEV TZ23131 (340) (Gryseels 2015)	LNKV (6,246) (Ishii et al. 2012)
NGEV TZ22285	81%			
NGEV TZ23131	82%	91%		
LNKV	80%	78%	79%	

A total of 43 arenaviruses were detected in 1155 *M. natalensis* kidney samples: 38 Gairo viruses, 4 Luna viruses and 1 Morogoro virus (Figure 4A-B, Supplementary Table 1). All Luna and Morogoro virus *L* gene sequences, but only 27 out 38 Gairo virus sequences were unique. Identical sequences were mostly found at the same locality, but in one case 2 km apart and in another case 29 km apart. They were sometimes found in different batches of extractions and/or PCRs, the negative control was never positive, and re-extractions were performed for suspected contaminations, so there is no indication for contamination in the lab. *NP* and *GPC* gene sequences were obtained for 42 and 32 out of 43 *L* gene positive samples, respectively. Samples with identical Gairo *L* sequences also had identical *NP* sequences, but not always identical *GPC* sequences.

As no *L* gene RNA was detected at any *M. natalensis* locality of a 350 km strip from south west to central Tanzania (Figure 4A), 540 dried blood samples spread over 17 localities were screened for anti-arenavirus antibodies. Antibodies were detected in 14 of these samples originating from five localities (Supplementary Table 1, Figure 4C). For one antibody-positive locality more than 50 kidney samples were available, so the remaining 42 kidney samples were screened for arenavirus RNA as well, resulting in the detection of an additional Luna virus (already included in the count above). Like

the *L* gene screening, the *GPC* and *NP* gene screening only detected arenavirus RNA from this sample, but not from any other pooled kidney sample from the five antibody-positive localities.



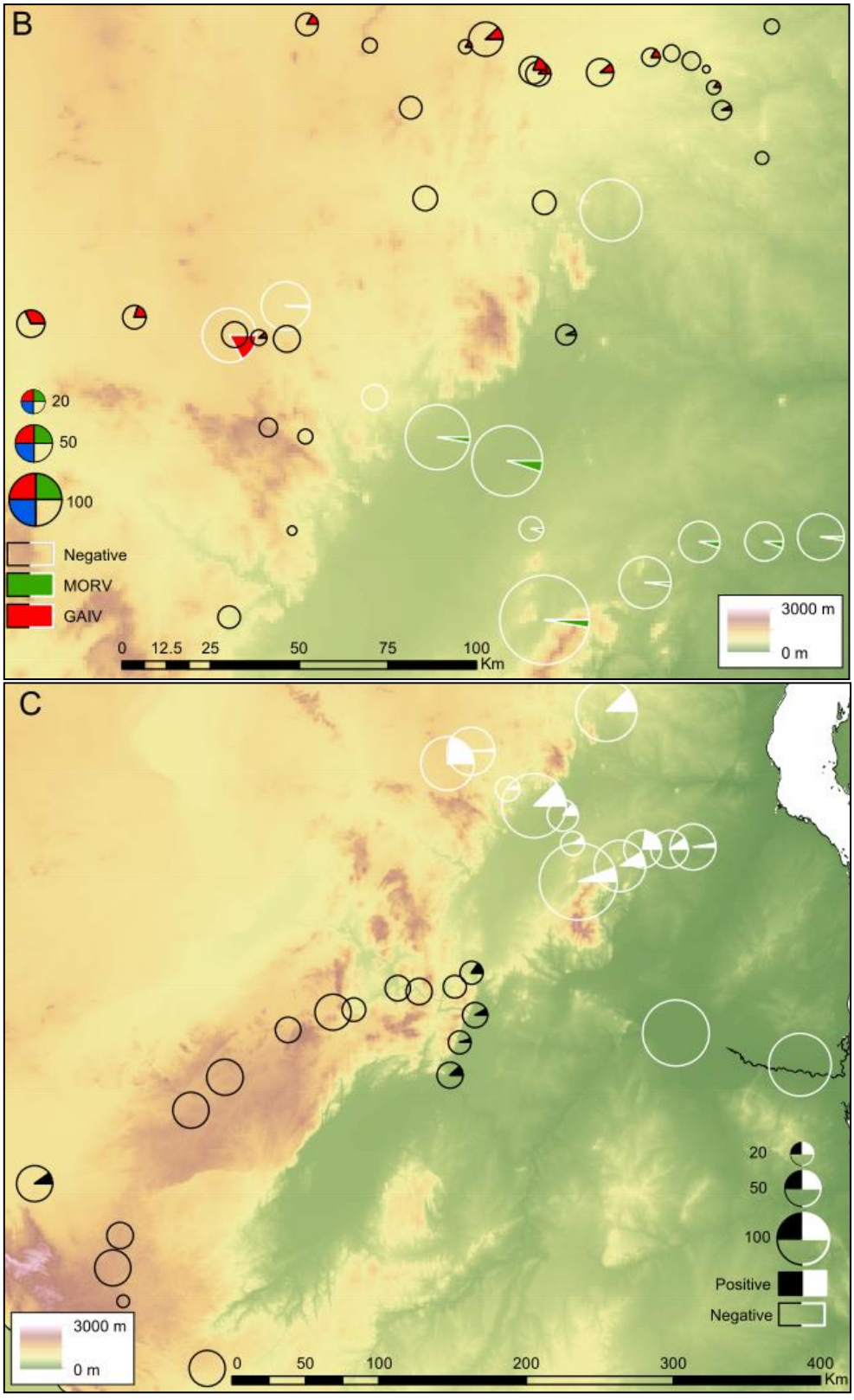


Figure 4: Arenavirus *L* gene RNA (A-B) and anti-arenavirus antibodies (C) detected in *Mastomys natalensis* in Tanzania. Figure B is an enlargement of the area in the grey rectangle in A. Pie chart areas are scaled to the number of individuals screened. Black pie charts represent localities screened in this study. White pie charts represent localities screened in Locus (2016) and in Gryseels et al. (2017) (RNA)/ Gryseels et al. (2015) and Mariën et al. (2017) (antibodies). Elevation data was made available by the U.S. Geological Survey's Center for Earth Resources Observation and Science.

4.2 Arenavirus genetic analyses

In the phylogenetic analyses based on a short portion of the *L* gene, the *Mus minutoides* virus clusters together with a pair of Ngerengere virus sequences with limited support (0.78 for BI/ 78 for ML analysis) (Figure 5). All Gairo, Morogoro and Luna *L*, *NP* and *GPC* sequences cluster together with sequences from their respective virus species with high support (1 for BI/ 98 - 100 for ML analyses), so no re-assortment or recombination is detectable among the three virus species (Figure 5, 6 and Supplementary Figure 1).

Four Morogoro virus clades have been described in Locus (2016) and Gryseels et al. (2017): sequences from Mkundi, Morogoro and Mikese (MORV-I); sequences from Bwawani and Ubena (MORV-II); sequences from Chalinze and Matipwili (MORV-III); and sequences from Berega, Dumila and Dakawa (MORV-IV). MORV-I, MORV-II and MORV-IV form monophyletic clades with a posterior probability between 0.83 and 1 in BI analyses for all three genes (Figure 7 and Supplementary Figure 2). Some of these clades are also supported in ML analyses, but always with lower support than in BI analyses (Figure 7 and Supplementary Figure 2). MORV-III is supported in BI *NP* and in BI and ML *GPC* analyses, but not in BI and ML *L* and in ML *NP* analyses (Figure 7 and Supplementary Figure 2). In the BI *L* tree, these sequences from Chalinze and Matipwili do not form a monophyletic clade, but are basal to all other Morogoro virus sequences (Figure 7). The new Morogoro virus sequence from Kunke does not cluster consistently across the gene trees, being a sister clade to MORV-II in the *L* trees (support of 0.90 in BI/ 70 in ML analysis), a sister clade to all other Morogoro virus sequences in the *NP* trees (support of 1 in BI/ 98 in ML analysis) and a sister clade to a clade consisting of both MORV-I and MORV-III in the *GPC* trees (support of 0.90 in BI/ 51 in ML analysis) (Figure 7 and Supplementary Figure 2).

Gairo virus was previously detected in three localities from the Gryseels et al. (2017) transect along the road from Dar es Salaam to Dodoma (Majawanga, Chakwale and Berega) and in two more distant localities in that study (Shinyanga-Lubaga and Mbulu). In this study Gairo virus was detected in three more localities supplementing that transect (Mbande, Mtanana and Ibuti), and in nine localities forming a new transect (from Meriongima to Magamba) along a less busy paved road (Figure 8 Bottom). Gairo virus sequences from neighbouring localities on a transect do not cluster all together, nor do they cluster together per transect. For example, some sequences from Makasini and from Majawanga cluster together with sequences from relatively distant localities on the other transect rather than with other sequences from the same locality or from neighbouring localities (Figure 8 and Supplementary Figure 3). Two medium-sized clades do show genetic spatial (and temporal) clustering (Figure 8 and Supplementary Figure 3). The first clade comprises the sequences

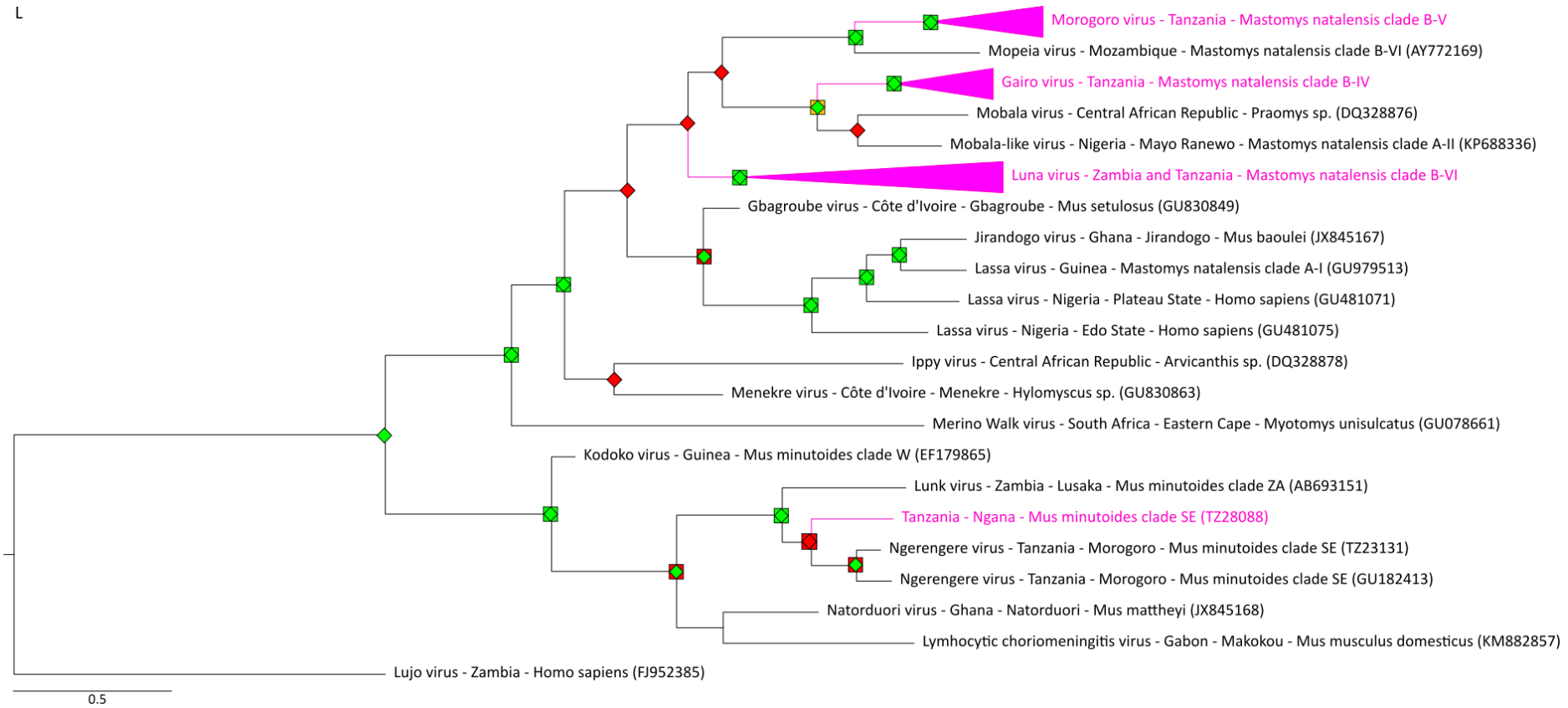


Figure 5: *L gene* Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/ 70 (Maximum likelihood), red for supports of 0.70/ 70 to 0.90/ 90, yellow for supports of 0.90/ 90 to 0.95/ 95, and green for supports of 0.95/ 95 and above. Taxa are named as the virus species followed by the sampling country, the locality or region (if available), the host species and the accession number from GenBank or a sample code starting with 'TZ' between brackets. Gairo virus, Morogoro virus and Luna virus sequences are collapsed to triangles (see Figures 7, 8 and 9 for these branches). Taxa are coloured fuchsia if the taxon is or contains a sample screened in this study. The scale bar represents the number of nucleotide substitutions per site.

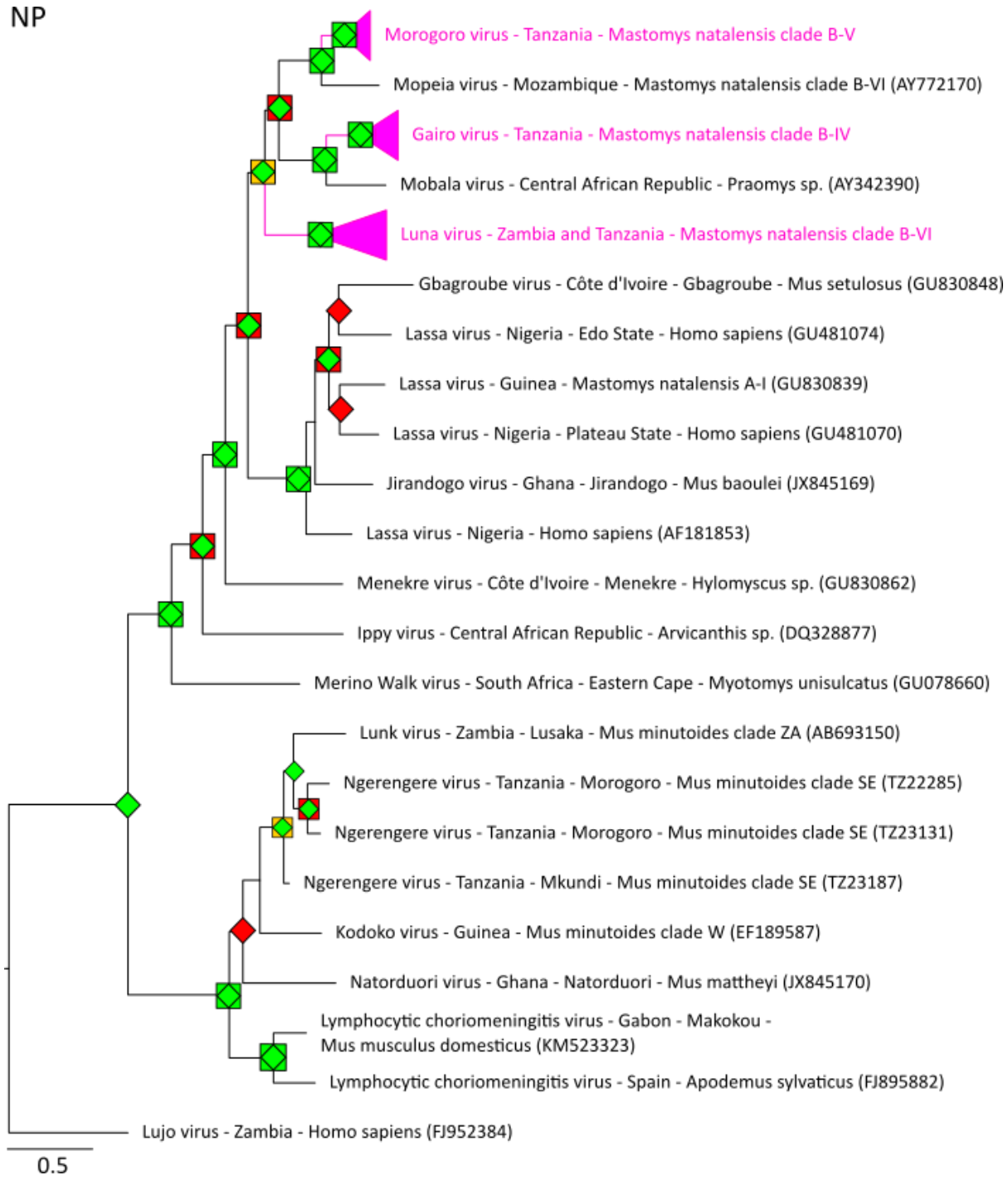


Figure 6: *NP* gene Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/ 70 (Maximum likelihood), red for supports of 0.70/ 70 to 0.90/ 90, yellow for supports of 0.90/ 90 to 0.95/ 95, and green for supports of 0.95/ 95 and above. Taxa are named as the virus species followed by the sampling country, the locality or region (if available), the host species and the accession number from GenBank or a sample code starting with 'TZ' between brackets. Gairo virus, Morogoro virus and Luna virus sequences are collapsed to triangles (see Figure 8 and Supplementary Figures 2 and 4 for these branches). Taxa are coloured fuchsia if the taxon is or contains a sample screened in this study. The scale bar represents the number of nucleotide substitutions per site.

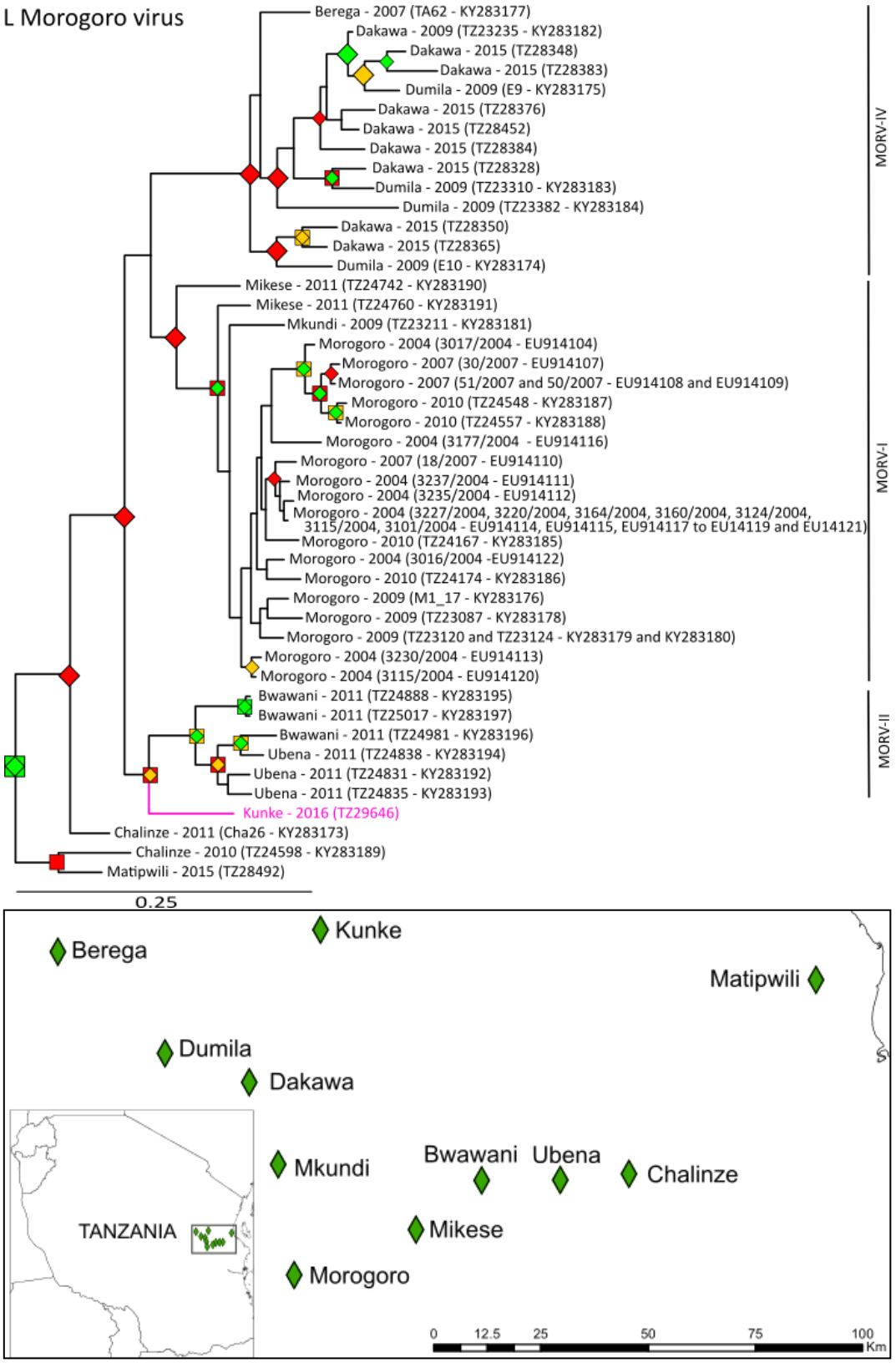


Figure 7: Top: Morogoro virus *L* gene Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/ 70 (Maximum likelihood), red for supports of 0.70/ 70 to 0.90/ 90, yellow for supports of 0.90/ 90 to 0.95/ 95, and green for supports of 0.95/ 95 and above. Sequences are named as the locality, the year and a sample code and accession number from GenBank between brackets. Clades with Roman numbers indicate clades described in Locus (2016) and Gryseels et al. (2017). The fuchsia sequence is new to this study. The scale bar represents the number of nucleotide substitutions per site. Bottom: Map of Morogoro virus localities.

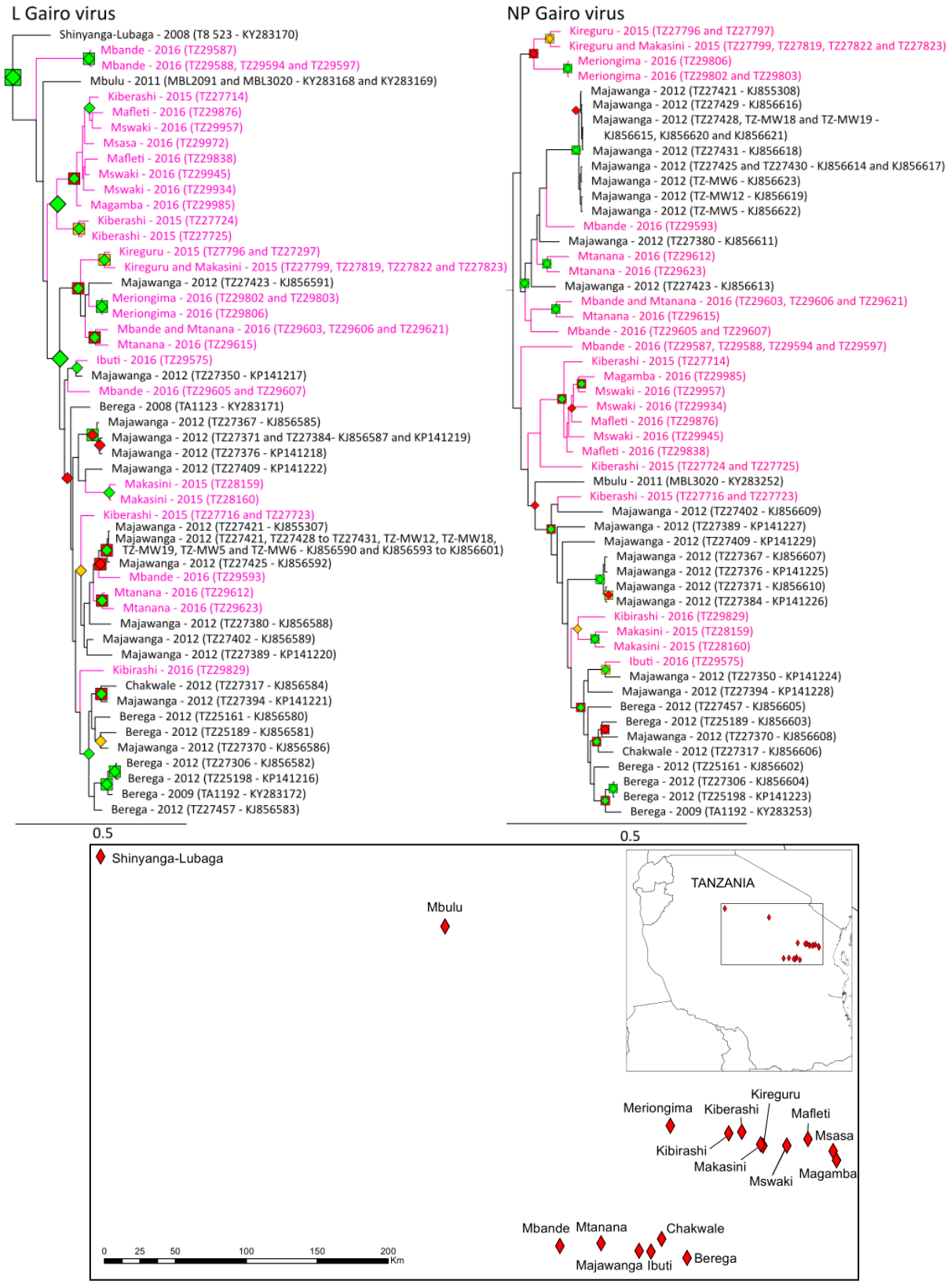


Figure 8: Top: Gairo virus *L* gene Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/ 70 (Maximum likelihood), red for supports of 0.70/ 70 to 0.90/ 90, yellow for supports of 0.90/ 90 to 0.95/ 95, and green for supports of 0.95/ 95 and above. Sequences are named as the locality, the year and a sample code and accession number from GenBank between brackets. Sequences are coloured fuchsia if they are new to this study. The scale bar represents the number of nucleotide substitutions per site. Bottom: Map of Gairo virus localities.

collected in 2015 and 2016 from Magamba, Msasa, Mafleti, Mswaki and some from Kiberashi (support of 0.87-1 for BI/ 68-89 for ML analyses). The other clade comprises sequences collected from 2009 to 2012 from Berega and Chakwale and some from Majawanga (support of 0.98-1 for BI/ 30-74 for ML analyses). In the *NP* gene ML tree an additional 2012 sample from Majawanga and a 2016 sample from Ibuti are also situated in this clade (Figure 8).

The three Luna virus sequences from Ngana form a monophyletic clade with high support (1 in BI/ 95-100 in ML analyses) for the three genes, but a Tanzanian monophyletic clade together with the Luna virus sequence from Ibohora is not supported (clade not found in the trees or with a support of 0.56 or 0.67 in BI/ 61 or 65 in ML analyses) (Figure 9 and Supplementary Figure 4).

4.3 *Mastomys natalensis* and *Mus minutoides genetic* analyses

The phylogenetic tree analysis of the *M. natalensis cyt b* sequences indicated that Gairo virus was detected in B-IV, Morogoro virus in B-V and Luna virus in B-VI individuals. All *M. natalensis* arenaviruses were thus found in correspondence with their respective host mitochondrial clades. The *M. minutoides* virus from Ngana (orange triangle near the border with Malawi in Figure 10B) was found in an individual of the SE clade, which carries Ngerengere virus in Morogoro and Mkundi (Goüy de Bellocq et al. 2010; Gryseels 2015). The distribution of the *M. natalensis* and *M. minutoides* mitochondrial clades in Tanzania can be seen in Figure 10.

4.4 Analyses of regional differences in *Mastomys natalensis* arenavirus detection

Because arenavirus RNA was not detected at all *M. natalensis* localities, the arenavirus RNA and anti-arenavirus antibody G-tests were repeated 15 times with varying classification of undetermined localities to either a Gairo virus group in north Tanzania, a Morogoro virus group in the east or an arenavirus group (including Luna virus) in the south (see Figure 11). All repetitions of the arenavirus RNA G-test revealed a significant non-random distribution of arenavirus positives over the three groups (G: 49.17-79.79, Df = 2, $P < 0.001$). Pairwise tests showed this result was due to a significantly higher prevalence of Gairo virus in north Tanzania compared to Morogoro virus in the east (P: < 0.001-0.001) and compared to arenaviruses in the southwest and centre (P: < 0.001-0.036). The result of the pairwise comparison between the Morogoro virus group and the southwestern and central group depended on the classification of the localities in central to south west Tanzania where no arenavirus RNA was detected (P: < 0.001-1). The more virus-free localities were assigned to the Morogoro virus group, the less clear the higher prevalence in the Morogoro virus group was (Figure

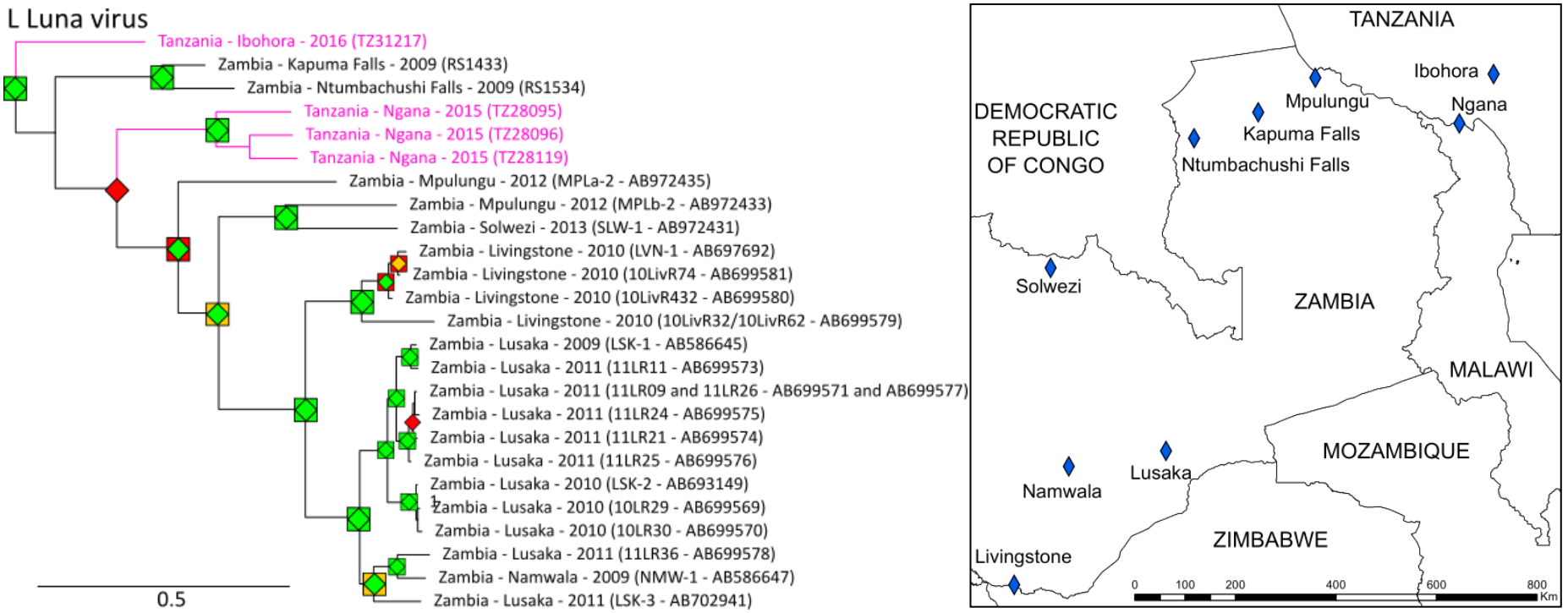


Figure 9: Left: Luna virus *L gene* Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/ 70 (Maximum likelihood), red for supports of 0.70/ 70 to 0.90/ 90, yellow for supports of 0.90/ 90 to 0.95/ 95, and green for supports of 0.95/ 95 and above. Sequences are named as the sampling country, the locality, the year and a sample code and accession number from GenBank between brackets. Sequences are coloured fuchsia if they are new to this study. The scale bar represents the number of nucleotide substitutions per site. Right: Map of Luna virus localities. Coordinates from the Solwezi and Mpulungu samples were not available on GenBank, but approximated by coordinates of the city/town centre.

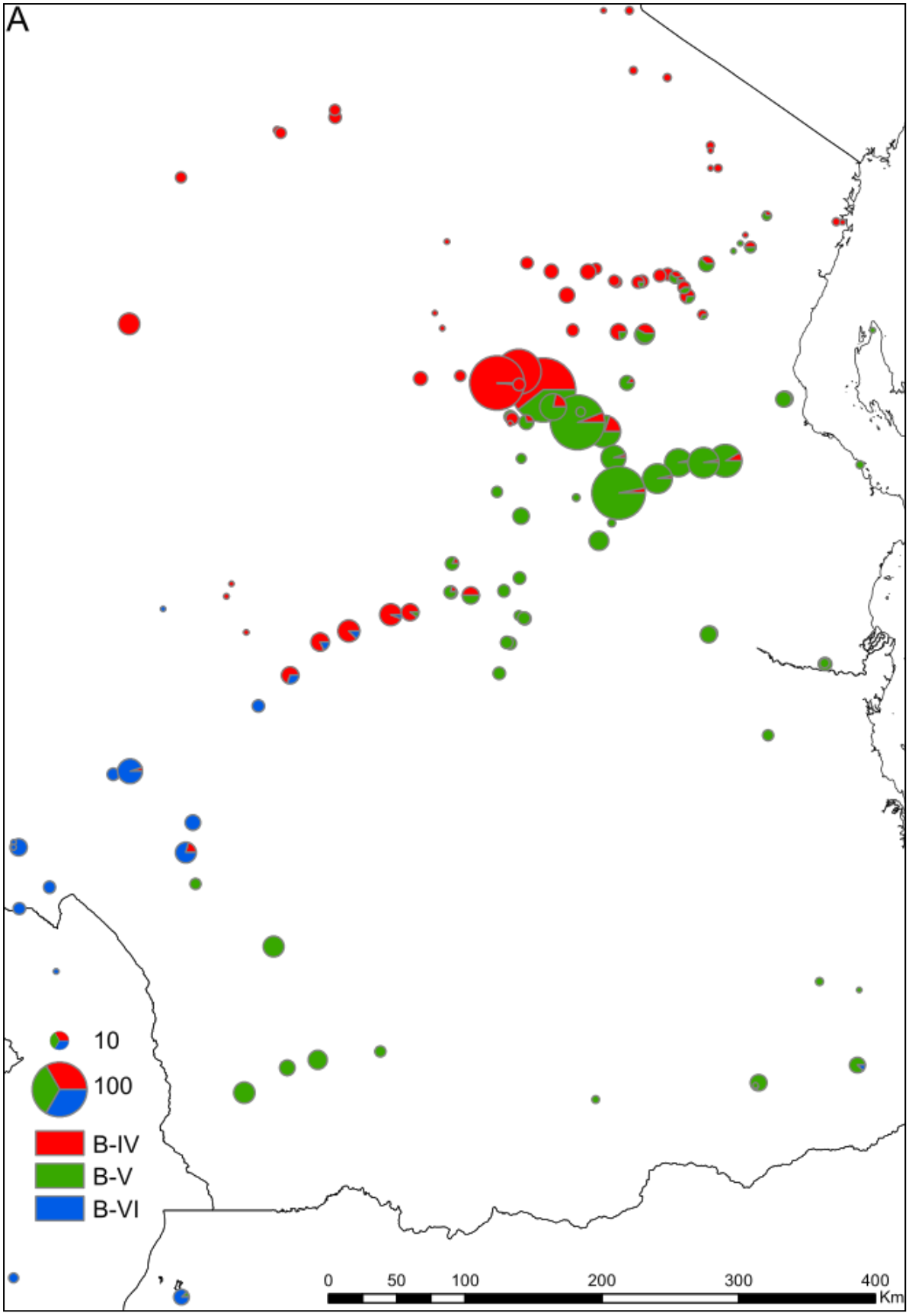


Figure 10A: Distribution of *Mastomys natalensis* mitochondrial lineages sensu Colangelo et al. (2013). Data from J. Bryja and A. Hánová from the IVB.

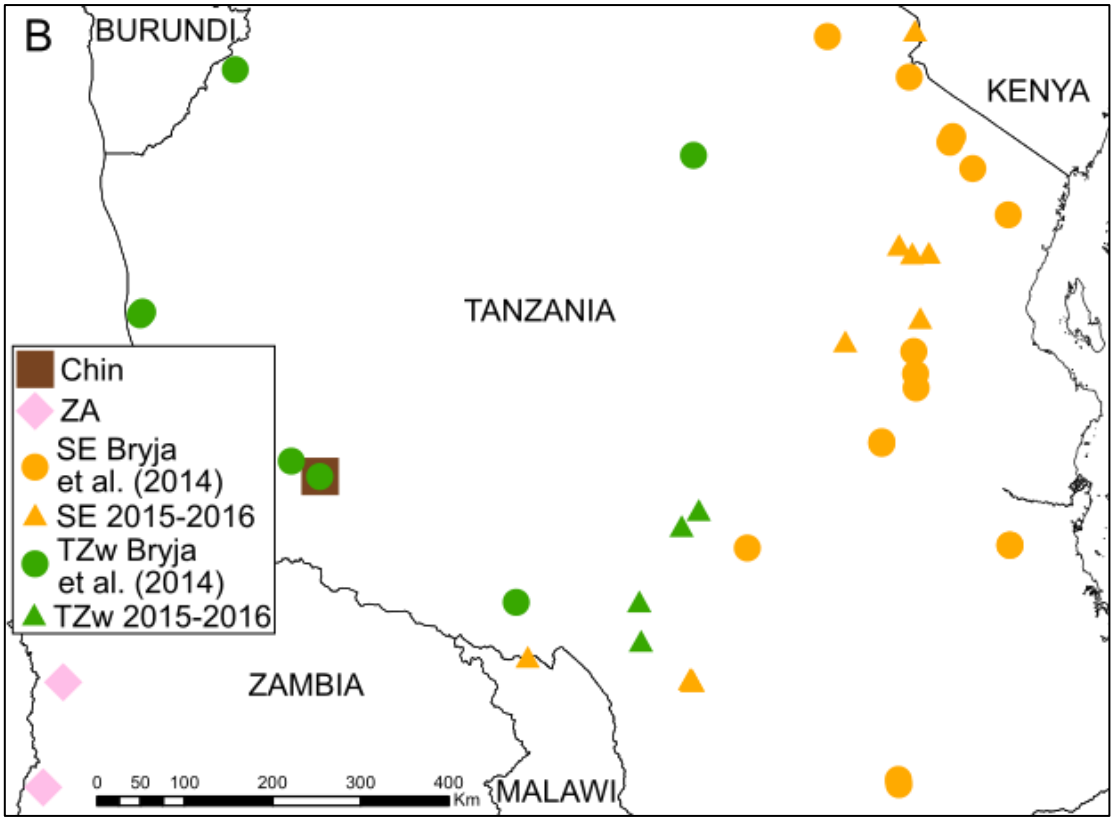


Figure 10B: Distribution of *Mus minutoides* mitochondrial lineages sensu Bryja et al. (2014). Data from J. Bryja and A. Hánová from the IVB.

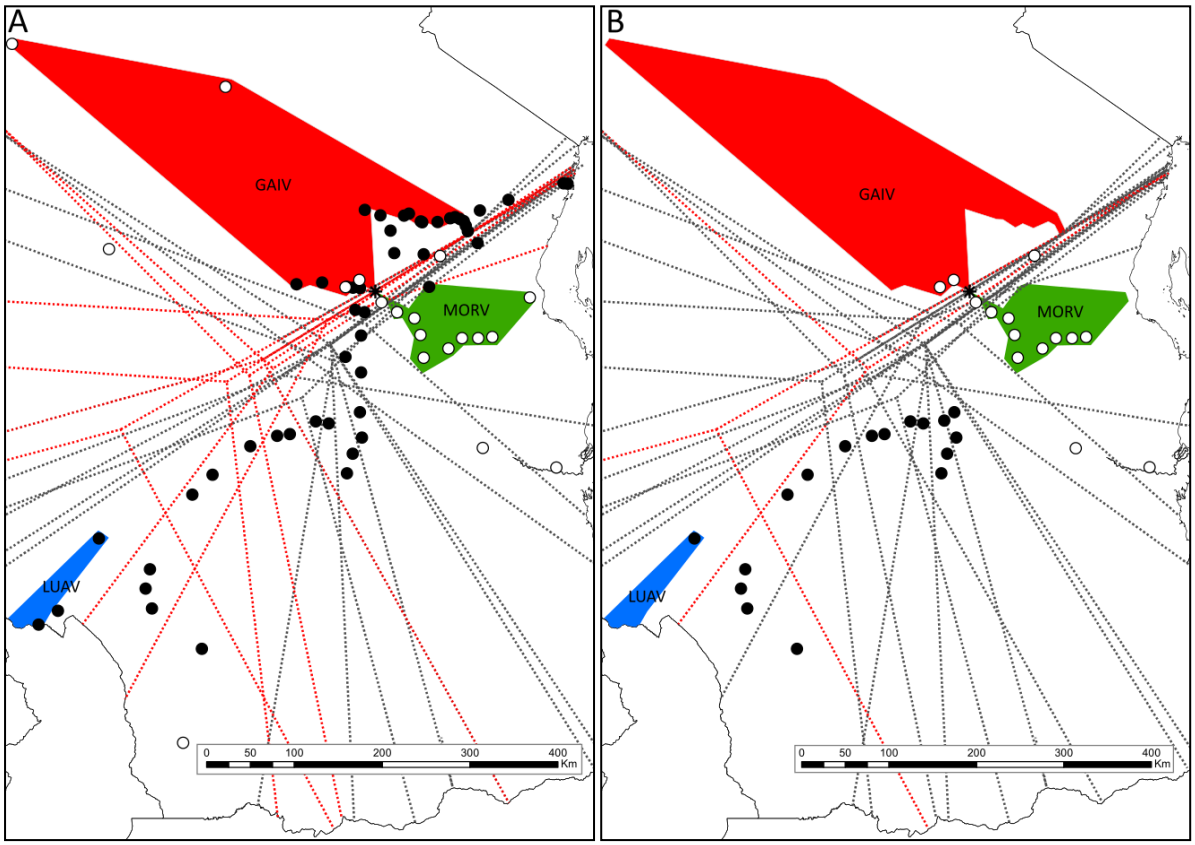


Figure 11: Variable classification of localities for the *Mastomys natalensis* arenavirus RNA G-test (A) and the anti-arenavirus antibody G-test (B). Filled circles represent localities screened in this study; open circles represent localities screened in Locus (2016) and in Gryseels et al. (2017) (RNA data)/ 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). Coloured polygons indicate 'core regions' connecting localities where a specific arenavirus species was found: Gairo virus (GAIV) in the north, Morogoro virus (MORV) in the east and Luna virus (LUAV) in the south west of Tanzania. Dotted lines divide the remaining localities into the Gairo virus group in the north (containing at least the Gairo virus core region), the Morogoro virus group in the east (containing at least the Morogoro virus core region) or a more general arenavirus group in the southwest and centre (containing at least the Luna virus core region). A first G-test revealed significant differences in arenavirus RNA prevalence between the three groups for all classifications. Significant differences between the Gairo virus and Morogoro virus group and between the Gairo virus and the southwestern and central group were also found for all classifications. However, significant differences between the Morogoro virus and southwestern and central group were only found for classifications based on the dark grey dotted lines, not for the red dotted lines (P values above 0.05). A second G-test revealed a significant difference in anti-arenavirus antibody prevalence between the Morogoro virus and southwestern and central group for classifications based on the dark grey dotted lines, but not for those based on the red dotted lines.

11A). The antibody G-test explored this Morogoro virus – southwest and centre pairwise comparison further: P values were lower (G: 2.65-47.79, Df = 1, P: < 0.001-0.104) than those for the RNA data, resulting in more classifications with a significant difference set at 0.05 (13 vs. 9 out of 15) (Figure 11B).

5. Discussion

5.1. Arenavirus specificity

Four Luna viruses were found at two localities in the south west of Tanzania and are the first Luna viruses to be detected outside of Zambia. The fact that these B-VI individuals carried Luna virus, like B-VI individuals in Zambia, and not Morogoro or Gairo virus like other Tanzanian individuals (of other lineages), supports Gryseels et al.'s hypothesis (Gryseels et al. 2017) that intraspecific *M. natalensis* lineages constrain the geographic ranges of their arenaviruses. This hypothesis is further supported by the association of Gairo virus with the B-IV and Morogoro virus with the B-V lineage found at a larger geographic scale than the transect of Gryseels et al. (2017), including at the lineage contact zone in a new transect (Figures 4B and 10). It is thus unlikely that the observed specificity in Gryseels et al. (2017) is a special case that came about because arenaviruses only recently met at that busy road. Nor can the absence of Morogoro virus in B-IV individuals in Gryseels et al. (2017)'s transect be explained by the limited number of B-IV dominated localities (only two). The present study adds 20

B-IV dominated localities, at 12 of which Gairo virus was detected and at none of which Morogoro virus was detected.

Mus minutoides also appears to carry different arenaviruses in distinct mitochondrial lineages in restricted geographical ranges. If the virus detected in the SE individual is Ngerengere virus, which has been detected in three SE individuals from Morogoro and Mkundi, this would support that *M. minutoides* arenaviruses might also be constrained by intraspecific *M. minutoides* lineages. The amino acid sequence of the *L* gene fragment indeed suggests that the virus is most similar to Ngerengere virus. The nucleotide sequence, however, is only slightly more similar to Ngerengere virus than it is to Lunk virus detected in a ZA individual from Zambia (Table 2). This is reflected in both Bayesian inference and Maximum likelihood trees as the sequence clusters with two other Ngerengere virus sequences rather than with the Lunk virus sequence, though only with a limited branch support (Figure 5). For the *NP* gene an extra Ngerengere virus sequence is available compared to the *L* gene. In this tree however, Ngerengere virus monophyly is not supported (Figure 6). In fact, it is not sure yet if Ngerengere virus truly represents a different species from Lunk virus. Further research including whole genome sequencing is needed to resolve this. The *M. minutoides* virus detected in this Master thesis should certainly be included in future pairwise comparisons because of the intermediate nature of its *L* gene fragment nucleotide sequence.

If Ngerengere and Lunk virus are not distinctly different from each other or if they are distinct, but a larger fragment of the new sample would indicate it is Lunk virus, then *M. minutoides* arenaviruses are most likely not constrained by mitochondrial lineages. In the first case they could still be constrained by larger mitochondrial clades, as the SE and TZw lineages appear slightly more related to each other than to other lineages (Bryja et al. 2014). In both cases, however, potential environmental barriers to *M. minutoides* arenavirus spread should be investigated.

In any case, East African *M. natalensis* arenaviruses have not been detected in *M. minutoides* individuals and vice versa, despite co-occurrence at at least four localities: Morogoro (Goüy de Bellocq et al. 2010), Mkundi (Gryseels et al. 2015) and Ngana (this study) in Tanzania and Lusaka in Zambia (Ishii et al. 2012). Furthermore, Ngerengere virus and Lunk virus are more related to each other and to Lymphocytic choriomeningitis virus in *Mus musculus* in Central Africa (N'Dilimabaka et al. 2015), than they are to *M. natalensis* arenaviruses. In West Africa, Kodoko virus in *Mus minutoides* and Natorduori virus in *Mus mattheyi* are closely related to these *Mus sp.* arenaviruses, while Jirandogo virus in *Mus baoulei* and Gbagroube virus in *Mus setulosus* are more closely related to Lassa virus in *M. natalensis* (Figures 5 and 6 and Supplementary Figure 1), indicating a past host switch of a Lassa(-like) virus. Furthermore, while Lassa virus is primarily born by *M. natalensis*, it

spills over to *Mastomys erythroleucus*, *Hylomyscus pamfi* and humans (Olayemi et al. 2016b). The pathogenic Lassa virus thus appears to spill over more easily than non-pathogenic East African arenaviruses, both in present and past times.

5.2. Spatial genetic structure of *Mastomys natalensis*-borne arenaviruses

The Morogoro virus trees show clear spatial genetic structure. As in Locus (2016) and Gryseels et al. (2017), four clades are present that contain all sequences from one, two or three adjacent localities. Three of these clades are supported by a posterior probability of at least 70% in BI trees for the three genes (Figure 7 and Supplementary Figure 2). The fourth is not supported in the *L* gene BI analysis (Figure 7), but this was also the case in the *L* (and *NP*) gene BI analyses with unconstrained branch lengths in Gryseels et al. (2017). However, the *L* gene tree is based on a very restricted fragment of only 340 nt. The new sequence from Kunke does not appear to belong to any of the previously described clades and could represent a new separate lineage (Figure 7 and Supplementary Figure 2).

Gairo virus spatial genetic structure is much more limited compared to that of Morogoro virus. Two clades contain all but a few sequences from neighbouring localities, but most sequences cluster together with sequences from another transect rather than with sequences from the same or an adjacent locality. Several factors could hypothetically contribute to a lower spatial genetic structure for Gairo virus compared to Morogoro virus. Gairo virus dynamics might be slightly different than that of Morogoro virus. For example, a longer infectious period, a longer latent period, a higher transmission efficiency (e.g. due to higher viral load), a slower mutation rate or a higher proportion of chronic compared to acute infections could have an impact. The latter could not only be caused by a difference in virus dynamics, but also by a difference in host population age structure (e.g. due to a difference in timing of reproduction). Host age likely matters as chronic Morogoro and Lassa virus infections only occur in laboratory conditions when *M. natalensis* are infected at a very young age (Walker et al. 1975; Borremans et al. 2015). Furthermore, Gairo virus hosts might have migrated more than Morogoro virus hosts, either due to environmental factors in the area or due to an intrinsic higher migration rate in B-IV compared to B-V individuals. However, the reduced spatial genetic structure mostly stems from the fact that many recent samples cluster together with 2012 samples from Majawanga and this might simply be the result of an outbreak of a very successful and mobile Gairo virus strain. Indeed, Gairo virus prevalence in Majawanga was 16%, much higher than the prevalence in Mbulu (4.3%) and Chakwale (1.2%) in 2011 and 2012, respectively (Gryseels et al.

2017) (Supplementary Table 1). Furthermore, Fichet-Calvet et al. (2016) also found evidence for multiple movements of Lassa virus strains between villages, though at a smaller spatial scale.

As there are fewer Luna virus sequences from much more distant localities than Morogoro and Gairo virus, it is not possible to comment much on Luna virus spatial genetic structure. However, an Ibohora-Ngana clade is not supported (Figure 9 and Supplementary Figure 4), even though Ngana is located much closer to Ibohora than any other locality (Figure 9 Right). Perhaps the mountain range in between them (Figure 4A) forms a strong environmental barrier.

5.3 Prevalence of *Mastomys natalensis*-borne arenaviruses

Significantly fewer arenaviruses were detected in south west to central Tanzania compared to the north east. In fact, initially no viruses were detected in 17 localities spanning a strip of about 350 km from the south west to the centre of Tanzania. Antibodies were detected in five of these localities, either at the eastern or at the western edge of the strip (Figure 4C). For the westernmost locality, extra samples were available and additional screening resulted in one Luna virus sample. The antibodies in this locality were thus most likely produced in response to Luna virus infections. For the four antibody-positive localities at the eastern edge of the strip, no extra samples were available and like the *L* gene RNA screening, the *NP* and *GPC* gene screening did not yield any positives. It is therefore not possible to determine in response to which arenavirus these antibodies were produced. However, as all *M. natalensis* individuals typed in these lineages belong to the B-V lineage, they were likely produced in response to Morogoro virus infections. Furthermore, RNA prevalence is generally lower than antibody prevalence (Mariën et al. 2017). As Morogoro virus RNA prevalence is usually as low as or lower than 5% (Gryseels et al. 2017), and as only 20 to 26 samples were available for each of these localities (Supplementary Table 1), the probability of a positive individual among them is very low.

In contrast, anti-arenavirus antibodies were detected in about 12% of 138 samples at a certain locality in north east Tanzania in Locus (2016), but not a single one of those 138 samples was positive for arenavirus RNA. With a sample size that large, we might expect a few RNA-positive samples. The current *M. natalensis* mitochondrial data indicates that two-thirds of the 24 individuals genotyped from this locality belong to the B-V lineage, while the other third belongs to the B-IV lineage. However, these samples were only screened with MoroL primers and MoroL primers are able to detect Gairo virus, but at a lower sensitivity than the LVL primers. It is therefore possible that

the detected antibodies were produced in response to Gairo virus infections and that Gairo virus was present in some samples, but was not picked up well by the MoroL primers.

In the north east, significantly fewer Morogoro viruses were detected compared to Gairo viruses. This difference has not been reported before, possibly because available data on Gairo virus was restricted to five localities, and at one of which it co-occurred with Morogoro virus (Gryseels et al. 2015, 2017). The differences in prevalence between Gairo virus in the north, Morogoro virus in the east and the arenavirus group in south west to central Tanzania could be related to the methodology, temporal variation, and differences in host and/or virus dynamics.

5.3.1 Methodology

The samples included in the G-test were screened for arenavirus RNA by three different people with minor differences in methodology. I screened all samples from southwestern and central localities, samples from all but four Gairo virus localities, and only a limited amount of samples from Morogoro virus localities. I initially pooled kidney samples by three and used both MoroL and LVL primers in a single PCR. S. Gryseels screened samples from four Gairo virus localities and most samples of the Morogoro virus localities. She initially pooled dried blood samples by two and used MoroL and LVL primers in two separate PCRs. T. Locus screened kidney samples from five localities, initially pooled by two and using only MoroL, not LVL primers. My screening might have been less sensitive than that of T. Locus and S. Gryseels because I pooled samples by three. Conversely, arenavirus RNA might remain in kidney tissue for a longer time than in blood, so T. Locus and I might have been able to detect more positives than S. Gryseels. However, I found both less (in the south west to the centre of Tanzania) and more arenaviruses (Gairo virus in the north) than S. Gryseels and T. Locus. Furthermore, the antibody G-test indicated an even stronger significant difference between the Morogoro virus group and the arenavirus group in south west to central Tanzania, i.e. more localities from the south west to the centre could be assigned to the Morogoro virus group before the P-value rose above 0.05. As the antibodies were screened for in the same way, a difference in methodology cannot explain this difference.

It cannot be excluded that some virus strains were not detected by our assays. Mutations in a PCR primer binding region could strongly affect the annealing of the primers to the template RNA (during the RT phase) and to the template DNA (during the PCR phase) and thus in a lower screening sensitivity. For examples, LVL3359-plus and LVL3754-minus primer pairs (which were used in combination with MoroL primers) were not able to detect a few Lassa virus positive samples from Sierra Leone in Leski et al. (2015). Likewise, Emmerich et al. (2008) showed that IFA slides coated

with cells infected with a certain strain of Lassa virus have a lower sensitivity for antibodies against divergent Lassa virus strains from other West African countries. Nonetheless, IFA slides coated with Lassa-infected cells were still able to detect antibodies against other arenaviruses from the Lassa virus complex such as Mopeia virus from Mozambique (Wulff et al. 1977) and Zimbabwe (Johnson et al. 1981), Mobala virus from the Central African Republic (Gonzalez et al. 1984) and Morogoro virus from Tanzania (Günther et al. 2009). Similarly, IFA slides coated with Morogoro-infected cells can detect antibodies against Gairo virus (Gryseels et al. 2015) and Luna virus from Tanzania (Figure 4C). A negative result due to a lower sensitivity to a more divergent strain is thus possible for any of our assays. However, the probability that such a strain in south west to central Tanzania is not picked up by the *L* gene, *NP* and *GPC* gene or antibody screening seems low if its prevalence and viral load are comparable to those of Gairo and Morogoro virus strains in the north east.

5.3.2 Temporal variation

The differences in arenavirus prevalence among the three groups could be temporal. Three localities in the extended dataset of the G-tests were sampled in two years or seasons, the others only in one (Supplementary Table 1). The observed prevalence in any given locality is thus just a snapshot, while prevalence likely fluctuates through time. However, each group was represented by multiple localities in the G-tests, which should have reduced effects of temporal stochasticity.

Non-random temporal variation might have had a more important impact. In Guinea, Lassa virus prevalence is two to three times higher during the rainy season compared to the dry season (Fichet-Calvet et al. 2007). As Morogoro virus localities were sampled both in the dry and rainy season, while Gairo virus and southwestern localities were only sampled in the dry season (Supplementary Table 1), a similar pattern for East African *M. natalensis* arenaviruses might explain why the Morogoro virus group had a higher prevalence than the southwestern group, but not why it had a lower prevalence than the Gairo virus group. Furthermore, while both southwestern and Gairo virus localities were mostly sampled in the dry season, southwestern localities were sampled more extensively in August and Gairo virus localities more extensively in June. However, while Lassa virus prevalence differed between the beginning and the end of the rainy season (no comparison was made between the beginning and the end of the dry season), it did so in opposite directions in two consecutive years (Fichet-Calvet et al. 2007). Differences throughout a season were thus not consistent. Sampling in different years could also have affected prevalence in a consistent way, but in 2016 many arenaviruses were detected in Gairo virus localities, while only one was detected in the southwest (Supplementary Table 1). In summary, there appear to be no differences in sampling

time that can explain all pairwise differences. Even though temporal variation might affect the differences in arenavirus prevalence, other factors at least appear to play an important role as well.

5.3.3 Host dynamics

Differences in host population dynamics could result in year-round differences in arenavirus prevalence inherent to the three regions. For example, *M. natalensis* density, migration rate and age population structure might vary throughout the study area. Age population structure could for instance vary between sampled localities if there is some variability in timing of reproduction (Leirs et al. 1993; Makundi et al. 2005, 2007). *M. natalensis* age could be an important factor because Gairo and Morogoro virus RNA are detected more in younger individuals (Borremans et al. 2011; Gryseels et al. 2015) and because Morogoro and Lassa virus inoculations only appear to result in chronic infections in very young individuals (Walker et al. 1975; Borremans et al. 2015). The extent of *M. natalensis* migration likely affects arenavirus persistence in a locality or region, and thus prevalence, and could, for example, depend on topography and vegetation cover (Russo et al. 2016). *M. natalensis* density influences their contact rate (Borremans et al. 2013), but of course also the number of susceptible individuals. However, given the strict specificity of arenaviruses to certain *M. natalensis* lineages, effective host density in or close to *M. natalensis* hybrid zones could be lower than the *M. natalensis* density. Perhaps this in itself could result in a lower arenavirus prevalence in the three-way contact zone from the south-west to the centre of Tanzania.

Differences in host population dynamics throughout the study area could arise due to variation in interactions with other species and habitat suitability. It is striking that the virus-free strip from south west to central Tanzania and two east Tanzanian virus-free localities with a large sample size from Locus (2016), correspond more or less to regions which are predicted to have a low suitability for *M. natalensis* and Lassa virus in Mylne et al. (2015) (Figure 12). With the current data it is not possible to investigate if these regions are indeed less suitable for *M. natalensis* and/or their arenaviruses. Trapping success in these localities at least does not appear to be lower than in other localities, but these numbers are just a proxy of *M. natalensis* density at patches of suitable field habitat, which could be surrounded by less suitable habitat. Furthermore, they are only a proxy at the time of capture, while population sizes can fluctuate strongly throughout and between years (Leirs et al. 1993). However, the prediction maps from Mylne et al. (2015) should be looked at with caution for Lassa virus predictions in West-Africa (see Introduction), and even more so for predictions about East-African arenaviruses. Nonetheless, they do suggest that there might be a specific set of environmental conditions present in these regions. For the strip from south west to central Tanzania these conditions might be linked to high elevation (Figure 4), but environmental

conditions that could affect arenavirus prevalence in the two localities from Locus (2016) are less clear.

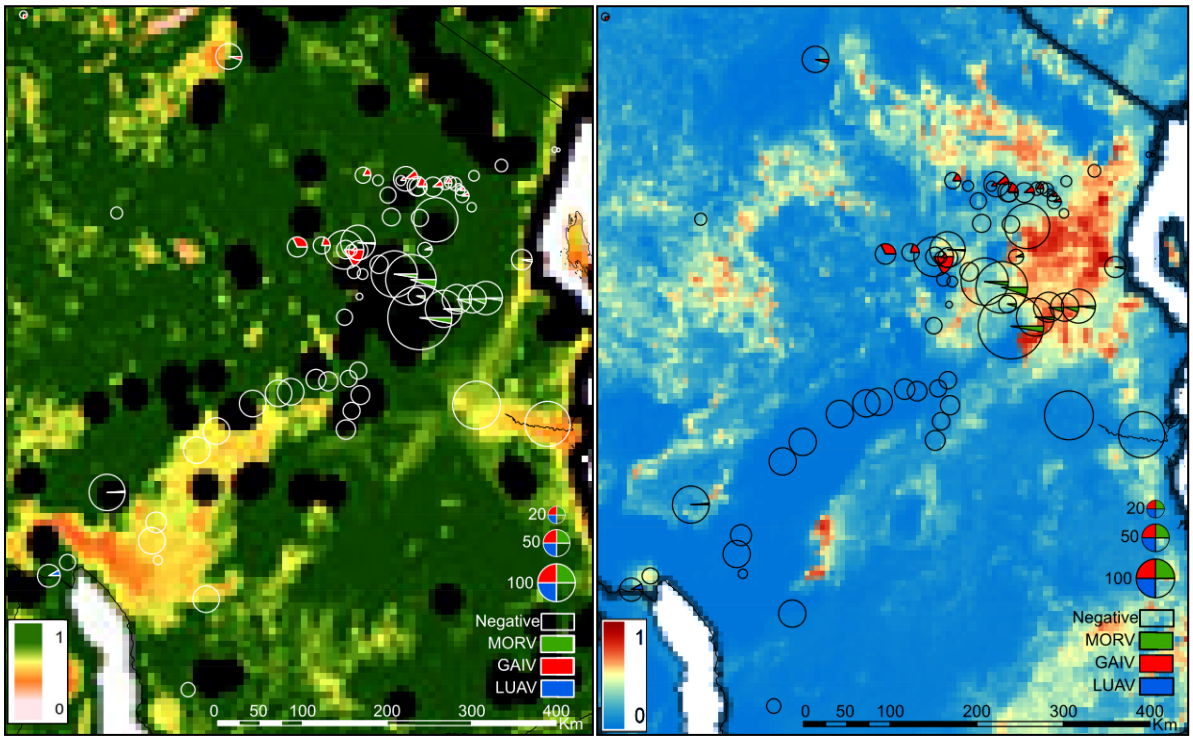


Figure 12: Arenavirus *L* gene RNA prevalence in *Mastomys natalensis* plotted on a predicted distribution of *Mastomys natalensis* (A) and a predicted distribution of Lassa virus (B) from Mylne et al. (2015). Pie charts are scaled to the number of individuals screened. The colour scale reflects environmental suitability with areas closer to 1 (green in A/ red in B) predicted to be more suitable and areas closer to 0 (pink in A/ blue in B) predicted to be less suitable. Black spots in A are *M. natalensis* trapping locations which were used to construct the model.

5.3.4 Arenavirus dynamics

Differences in arenavirus prevalence in the three regions could be caused by differences in arenavirus dynamics. As for host population dynamics, these differences could be linked to variation in environmental conditions, but they could also be inherent to the viruses or the strains themselves. A few viral properties that could influence arenavirus persistence and prevalence are length of the infectious period, transmission efficiency and ability to chronically infect host individuals (Goyens et al. 2013; Borremans 2015). Moreover, a higher viral load may not only affect transmission efficiency (Gray et al. 2001) and thus actual prevalence, but also the probability of detection and thus observed prevalence.

That fact that most southwestern localities are located in a three-way hybrid zone also makes it difficult to predict which virus(es) might be present in this region where no arenaviruses were detected. Perhaps Luna virus occurs here, like in two other southwestern localities, and has an

intrinsically lower prevalence. The Luna virus prevalence in Zambia, however, does not appear to be especially low compared to Morogoro and Gairo virus (Ishii et al. 2011, 2012).

6. Conclusion

Luna virus, previously only known from Zambian *Mastomys natalensis* individuals, was detected for the first time at two localities in the south west of Tanzania. It was found in individuals belonging to *M. natalensis* lineage B-VI, Morogoro virus in B-V and Gairo virus in B-IV. All *M. natalensis* arenaviruses were thus only found in combination with their corresponding *M. natalensis* mitochondrial lineage. This observation supports the hypothesis that *M. natalensis* arenaviruses are restricted to certain geographic regions due to their specificity to certain host lineages. Furthermore, Gairo virus was again detected at the contact zone with the B-V lineage in a new transect and sequences from this transect clustered together with sequences from the transect in Gryseels et al. (2017), indicating that the Gairo and Morogoro virus did not meet only recently at the latter transect along a busy road. The *M. natalensis* arenaviruses boundaries thus appear to be stable in Tanzania. Further research is needed to assess if this is also the case for *Mus minutoides* arenaviruses.

Further research is also needed to clarify why Morogoro virus in the east of Tanzania was detected less than Gairo virus in the north, but more than arenaviruses in the centre and south west and why Morogoro virus sequences show more spatial genetic structure than Gairo virus sequences. Such differences have not been reported before and could be caused by differences in virus and/or host dynamics, which could possibly but not necessarily relate to environmental factors. For example, a relatively recent spread of a successful and highly mobile Gairo virus strain could explain both the higher prevalence and the lower degree of spatial genetic structure of Gairo virus compared to Morogoro virus in this study. However, if such differences exist, it may imply that one or the other virus may be more suitable as a model for Lassa virus.

7. Acknowledgements

I would like to express my gratitude to everyone who contributed to this Master thesis. I would first like to thank my promotor, Prof. Dr. Herwig Leirs, who gave me the wonderful opportunity to investigate arenaviruses in Tanzania, supervised me and put me in touch with so many helpful people.

I am extremely grateful to my co-promotor Dr. Joëlle Goüy de Bellocq from the IVB. She supervised me almost every step of the way and was always ready to help me despite the long distance.

I thank VLIR-UOS for their financial support to perform fieldwork in Tanzania and Prof. Dr. Apia Massawe and the rest of the SPMC for receiving me there. Dr. Adam Konečný and Dr. Ondřej Mikula from the IVB took me under their wing and showed me how to catch and dissect mice in the field. Dr. Abdul Katakweba from the SPMC looked after me during my fieldwork and arranged for everything I needed. He even managed for us to trap again in some fields where one week earlier local people had thought we were laying bombs instead of small aluminium live traps. Khalid Kibwana from the SPMC always knew the best places to set traps and provided excellent assistance in the field. I was also lucky to be able to screen samples collected during other field expeditions by Tatiana Aghová, Dr. Josef Bryja, Alexandra Hánová, Jarmila Krásová, Vladímir Mazoch Dr. Ondřej Mikula and Jana Vrbová Komárková from the IVB, and Dr. Abdul Katakweba and Dr. Christopher Sabuni from the SPMC.

I also received a warm welcome at the IVB by Josef Bryja and many others. I thank Josef Bryja and Alexandra Hánová for sharing their host data and Stuart J.E. Baird for his sound statistical advice. Anna Bryjová and Natalie Van Houtte helped me during my lab work at the IVB and at the University of Antwerp, respectively. Joachim Mariën from the University of Antwerp taught me how to check IFA slides and checked them when I was unsure about the result. I also thank Sophie Gryseels. It was a joy building upon her PhD data and results.

This study was supported by a project of the FWO (G0A4815N) and of the Czech Science Foundation (15-20229S).

8. References

Andersen K.G. et al. (2015). Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. *Cell.* 162:738-750. DOI: 10.1016/j.cell.2015.07.020.

Armstrong C. and Wooley J.G. (1935). Studies on the Origin of a Newly Discovered Virus Which causes Lymphocytic Choriomeningitis in Experimental Animals. *Public Health Reports.* 50:537-541.

Borremans B. (2015). Transmission ecology of Morogoro arenavirus in the Multimammate Mouse *Mastomys natalensis* in Tanzania. PhD Thesis. University of Antwerp, Belgium. 223 pp.

Borremans B. (2014). Ammonium improves elution of fixed dried blood spots without affecting immunofluorescence assay quality. *Tropical Medicine and International Health.* 19:413-416. DOI: 10.1111/tmi.12259.

Borremans B., Hughes N.K., Reijniers J., Sluydts V., Katakweba A.A.S., Mulungu L.S., Sabuni C.A., Makundi R.H. and Leirs H. (2013). Happily together forever: temporal variation in spatial patterns and complete lack of territoriality in a promiscuous rodent. *Population Ecology.* 56:109-118. DOI: 10.1007/s10144-013-0393-2.

Borremans B., Leirs H., Gryseels S., Günther S., Makundi R. and Goüy de Bellocq, J. (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.* 11:1125-1131. DOI: 10.1089/vbz.2010.0010.

Borremans B., Vossen R., Becker-Ziaja B., Gryseels S., Hughes N., Van Gestel M., Van Houtte N., Günther S. and Leirs H. (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. DOI: 10.1038/srep10445.

Bowen M.D., Peters C.J. and Nichol S.T. (1997). Phylogenetic Analysis of the *Arenaviridae*: Patterns of Virus Evolution and Evidence for Cospeciation between Arenaviruses and Their Rodent Hosts. *Molecular Phylogenetics and Evolution.* 8: 301-316. DOI: 10.1006/mpev.1997.0436.

Briese T. et al. (2009). Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever-Associated Arenavirus from Southern Africa. *PLoS Pathogens.* 5:e1000455. DOI: 10.1371/journal.ppat.1000455.

Bryja J. et al. (2014). Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology.* 14:256. DOI: 10.1186/s12862-014-0256-2.

Cajimat M.N.B., Milazzo M.L., Haynie M.L., Hanson J.D., Bradley R.D. and Fulhorst C.F. (2011). Diversity and phylogenetic relationships among the North American Tacaribe serocomplex viruses (Family *Arenaviridae*). *Virology.* 421:87-95. DOI: 10.1016/j.virol.2011.09.013.

CDC. (2015). Lassa Fever. Retrieved from https://www.cdc.gov/vhf/lassa/index.html.

Charrel R.N. and de Lamballerie X. (2002). Chapter 16: Molecular Epidemiology of Arenaviruses. In: *The Molecular Epidemiology of Human Viruses* (ed.: Leitner T.). New York, USA: Springer Science+Business Media. Pp. 385-404.

Coetzee C.G. (1975). The Biology, Behaviour, and Ecology of *Mastomys Natalensis* in Southern Africa. *Bulletin of the World Health Organization.* 52:637-644.

Colangelo P., Verheyen E., Leirs H., Tatard C., Denys C., Dobigny G., Duplantier J.M., Brouat C., Granjon L. and Lecompte E. (2013). A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys Natalensis*. *Biological Journal of the Linnean Society.* 108:901-916. DOI: 10.1111/bij.12013.

Coulibaly-N'Golo D. et al. (2011). Novel Arenavirus Sequences in *Hylomyscus* Sp. and *Mus (Nannomys) setulosus* from Côte d'Ivoire: Implications for Evolution of Arenaviruses in Africa. *PLoS ONE.* 6:e20893. DOI: 10.1371/journal.pone.0020893.

Darriba D., Taboada G.L., Doallo R. and Posada D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods.* 9:772. DOI: 10.1038/nmeth.2109.

Delany M.J. (1964). A study of the ecology and breeding of small mammals in Uganda. *Journal of Zoology.* 142:347-370. DOI: 10.1111/j.1469-7998.1964.tb04627.x.

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.* 1:283-297. DOI: 10.1089/15303660160025912.

Duchêne S., Holmes E.C. and Ho S.Y.W. (2014). Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proceedings of the Royal Society of London B: Biological Sciences.* 281:20140732. DOI: 10.1098/rspb.2014.0732.

Dzotsi E.K. et al. (2012). The first cases of Lassa Fever in Ghana. *Ghana Medical Journal.* 46:166-170.

Ehichioya D.U. et al. (2011). Current Molecular Epidemiology of Lassa Virus in Nigeria. *Journal of Clinical Microbiology.* 49:1157-1161. DOI: 10.1128/JCM.01891-10.

Eichler R., Lenz O., Strecker T., Eickmann M., Klenk H.-D. and Garten W. (2003). Identification of Lassa virus glycoprotein signal peptide as a trans-acting maturation factor. *EMBO Reports.* 4:1084-1088. DOI: 10.1038/sj.embor.7400002.

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.* 42:40-44. DOI: 10.1016/j.jcv.2007.11.019.

Fehling S.K., Lennartz F. and Strecker T. (2012). Multifunctional nature of the arenavirus RING finger protein Z. *Viruses.* 4:2973-3011. DOI: 10.3390/v4112973.

Fichet-Calvet E., Becker-Ziaja B., Koivogui L. and Günther S. (2014). Lassa Serology in Natural Populations of Rodents and Horizontal Transmission. *Vector-Borne and Zoonotic Diseases.* 14:665-674. DOI: 10.1089/vbz.2013.1484.

Fichet-Calvet E., Lecompte E., Koivogui L., Soropogui B., Doré A., Kourouma F., Sylla O., Daffis S., Koulémou K. and Ter Meulen J. (2007). Fluctuation of abundance and Lassa virus prevalence in *Mastomys natalensis* in Guinea, West Africa. *Vector Borne* a*nd Zoonotic Disease*s. 7:119-128. DOI: 10.1089/vbz.2006.0520.

Fichet-calvet E., Ölschläger S., Strecker T., Koivogui L., Becker-Ziaja B., Bongo Camara A., Soropogui B. and Magassouba N.F. (2016). Spatial and temporal evolution of Lassa virus in the natural host population in Upper Guinea. *Scientific Reports.* 6:21977. DOI: 10.1038/srep21977.

Frame J.D., Baldwin J.M.Jr., Gocke D.J. and Troup J.M. (1970). Lassa Fever, a new virus disease of man from West Africa. *American Journal of Tropical Medicine and Hygiene.* 19:670-676.

Fulhorst C.F. et al. (1999). Natural rodent host associations of Guanarito and Pirital viruses (family *Arenaviridae*) in central Venezuela. *American Journal of Tropical Medicine and Hygiene.* 61:325-330.

Fulhorst C.F., Milazzo M.L., Carroll D.S., Charrel R.N. and Bradley R.D. (2002). Natural host relationships and genetic diversity of Whitewater Arroyo virus in southern Texas. *American Journal of Tropical Medicine and Hygiene.* 67:114-118.

Gonzalez J.P., McCormick J.B., Georges A.J. and Kiley M.P. (1984). Mobala virus: Biological and physicochemical properties of a new arenavirus isolated in the Central African Republic. *Annual Review of Virology*. 135:145-158.

Goüy de Bellocq J., Borremans B., Katakweba A., Makundi R., Baird S.J.E., Becker-Ziaja B., Günther S., and Leirs H. (2010). Sympatric Occurrence of 3 Arenaviruses, Tanzania. *Emerging Infectious Diseases.* 16:692-695. DOI: 10.3201/eid1604.091721.

Goyens J., Reijniers J., Borremans B. and Leirs H. (2013). Density thresholds for Mopeia virus invasion and persistence in its host *Mastomys natalensis. Journal of Theoretical Biology.* 317: 55-61. DOI: 10.1016/j.jtbi.2012.09.039.

Gray R.H. et al. (2001). Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet*. 357: 1149-1153. DOI: 10.1016/S0140-6736(00)04331-2.

Gryseels S. (2015). Evolutionary relationships between arenaviruses and their rodent hosts. PhD Thesis. University of Antwerp, Belgium. 205 pp.

Gryseels S., Goüy de Bellocq J., Makundi R., Vanmechelen K., Broeckhove J., Mazoch V., Šumbera R., Zima J., Leirs H. and Baird S.J.E. (2016). Genetic distinction between contiguous urban and rural multimammate mice in Tanzania despite gene flow. *Journal of Evolutionary Biology.* 29:1952-1967. DOI: 10.1111/jeb.12919.

Gryseels S., Baird S.J.E., Borremans B., Makundi R., Leirs H. and Goüy de Bellocq J. (2017). When Viruses Don't Go Viral: The Importance of Host Phylogeographic Structure in the Spatial Spread of Arenaviruses. *PLOS Pathogens*. 13:e1006073. DOI: 10.1371/journal.ppat.1006073.

Gryseels S., Rieger T., Oestereich L., Cuypers B., Borremans B., Makundi, R., Leirs H., Günther S. and Goüy de Bellocq J. (2015). Gairo virus, a novel arenavirus of the widespread *Mastomys natalensis*: Genetically divergent, but ecologically similar to Lassa and Morogoro viruses. *Virology.* 476:249-256. DOI: 10.1016/j.virol.2014.12.011.

Guindon S. and Gascuel O. (2003). A simple, fast and accurate method to estimate large phylogenies by Maximum-Likelihood. *Systematic Biology.* 52:696-704.

Günther S. et al. (2009). Mopeia Virus-related Arenavirus in Natal Multimammate Mice, Morogoro, Tanzania. *Emerging Infectious Diseases.* 15:2008-2012. DOI: 10.3201/e*id1512.090864.*

Günther S. and Lenz O. (2004). Lassa Virus. *Critical Reviews in Clinical Laboratory Sciences.* 41:339-390. *DOI: 10.1080/10408360490497456.*

Hervé M. (2017). Package 'RVAideMemoire' v0.9-65. Retrieved from https://cran.r-project.org/web/packages/RVAideMemoire/RVAideMemoire.pdf.

Holmes E.C. (2003). Molecular Clocks and the Puzzle of RNA Virus Origins. *Journal of Virology.* 77:3893-3897. DOI: 10.1128/JVI.77.7.3893-3897.2003.

Irwin N.R., Bayerlová M., Missa O. and Martínková N. (2012). Complex patterns of host switching in New World arenaviruses. *Molecular Ecology.* 21:4137-4150. DOI: 10.1111/j.1365-294X.2012.05663.x.

Ishii A., Thomas Y., Moonga L., Nakamura I., Ohnuma A., Hang'ombe B.M., Takada A., Mweene A.S. and Sawa H. (2012). Molecular surveillance and phylogenetic analysis of Old World Arenaviruses in Zambia. *Journal of General Virology*. 93:2247-2251. DOI: 10.1099/vir.0.044099-0.

Ishii A., Thomas Y., Moonga L., Nakamura I., Ohnuma A., Hang'ombe B., Takada A., Mweene A. and Sawa H. (2011). Novel Arenavirus, Zambia. *Emerging Infectious Diseases.* 17:1921-1924. DOI: 10.3201/eid1710.10452.

Johnson K.M., Kuns M.L., Mackenzie R.B., Webb P.A. and Yunker C.E. (1966). Isolation of Machupo Virus from Wild Rodent *Calomys Callosus*. *American Journal of Tropical Medicine and Hygiene.* 15:103-106. DOI: 10.4269/ajtmh.1966.15.103.

Johnson K.M., Taylor P., Elliott L.H. and Tomori O. (1981). Recovery of a Lassa-related arenavirus in Zimbabwe. *American Journal of Tropical Medicine and Hygiene.* 30: 1291-1293. DOI: 10.4269/ajtmh.1981.30.1291.

Kouadio L., Nowak K., Akoua-Koffi C., Weiss S., Allali B.K., Witkowski P.T., Krüger D.H., Couacy-Hymann E., Calvignac-Spencer S. and Leendertz F.H. (2015). Lassa Virus in Multimammate Rats, Côte d'Ivoire, 2013. *Emerging Infectious Diseases.* 21:1481-1483. DOI: 10.3201/eid2108.150312.

Kranzusch P.J. and Whelan S.P.J. (2012). Architecture and regulation of negative-strand viral enzymatic machinery. *RNA Biology.* 9: 941-948. DOI: 10.4161/rna.20345.

Kronmann K.C., Nimo-Paintsil S., Obiri-danso K., Ampofo W. and Fichet-Calvet E. (2013). Arenaviruses Detected in Pygmy Mice, Ghana. *Emerging Infectiou*s Diseases. 19:1832-1835. DOI: 10.3201/eid1911.121491.

Lecompte E., ter Meulen J., Emonet S., Daffis S. and Charrel R. N. (2007). Genetic identification of Kodoko virus, a novel arenavirus of the African Pigmy Mouse (*Mus Nannomys minutoides*) in West Africa. *Virology.* 364:178-183. DOI: 10.1016/j.virol.2007.02.008.

Leirs H., Verhagen R. and Verheyen W. (1993). Productivity of different generations in a population o*f Mastomys natalensis* rats in Tanzania. O*ikos.* 68:53-60. DOI: 10.2307/3545308.

Leski T.A., Stockelman M.G., Moses L.M., Park M., Stenger D.A., Ansumana R., Bausch D.G. and Lin B. (2015). Sequence Variability and Geographic Distribution of Lassa Virus, Sierra Leone. *Emerging Infectious Diseases.* 21:609-618. DOI: http://dx.doi.org/10.3201/eid2104.141469.

Locus T. (2016). Landscape genetics of *Mastomys natalensis*-borne arenaviruses in Tanzania. Master thesis. University of Antwerp, Belgium. 53 pp.

Makundi R.H., Massawe A.W. and Mulungu L.S. (2005). Rodent population fluctuations in three ecologically heterogeneous locations in northeast, central and southwest Tanzania. *Belgian Journal of Zoology.* 135:159-165.

Makundi R.H., Massawe A.W. and Mulungu L.S. (2007). Reproduction and population dynamics of *Mastomys natalensis* Smith, 1834 in an agricultural landscape in the Western Usambara Mountains, Tanzania. *Integrative Zoology.* 2:233-238. DOI: 10.1111/j.1749-4877.2007.00063.x.

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. DOI: 10.1186/s13071-017-2146-0.

Messina E.L., York J. and Nunberg J.H. (2012). Dissection of the Role of the Stable Signal Peptide of the Arenavirus Envelope Glycoprotein in Membrane Fusion. *Journal of Virology.* 86:6138-6145. DOI: 10.1128/JVI.07241-11.

Milazzo M.L., Campbell G.L. and Fulhorst C.F. 2011. Novel arenavirus infection in humans, United States. *Emerging Infectious Diseases.* 17:1417-1420. DOI: 10.3201/eid1708.110285.

Miller M.A., Pfeiffer W. and Schwartz T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*. New Orleans, USA. pp. 1-8.

Mills J.N., Ellis B.A., Childs J.E., McKee K.T.Jr., Maiztegui J.I., Peters C.J., Ksiazek T.G. and Jahrling P.B. (1994). Prevalence of infection with Junin virus in rodent populations in the epidemic area of Argentine hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene.* 51:554-562.

Mylne A.Q.N., Pigott D.M., Longbottom J., Shearer F., Duda K.A., Messina J.P., Weiss D.J., Moyes C.L., Golding N. and Hay S.I. (2015). Mapping the zoonotic niche of Lassa Fever in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 109:483-492. DOI: 10.1093/trstmh/trv047.

N'koué Sambiéni E., Danko N. and Ridde V. (2015). La Fièvre Hémorragique à Virus Lassa Au Bénin en 2014 en Contexte d'Ebola: une épidémie révélatrice de la faiblesse du système sanitaire. *Anthropologie & Santé [En ligne].* 11. DOI: 10.4000/anthropologiesante.1772.

N'Dilimabaka N. et al. (2015) Evidence of Lymphocytic Choriomeningitis Virus (LCMV) in Domestic Mice in Gabon: Risk of Emergence of LCMV Encephalitis in Central Africa. *Journal of Virology.* 89:1456-1460. DOI: 10.1128/JVI.01009-14.

Olayemi A. et al. (2016a). Arenavirus Diversity and Phylogeography of *Mastomys natalensis* Rodents, Nigeria. *Emerging Infectious Diseases.* 22:694-697. DOI: 10.3201/eid2204.150155.

Olayemi A. et al. (2016b). New Hosts of The Lassa Virus. *Scientific Reports.* 6:25280. DOI: 10.1038/srep25280.

Perez M. and de la Torre J.C. (2003). Characterization of the Genomic Promoter of the Prototypic Arenavirus Lymphocytic Choriomeningitis Virus. *Journal of Virology.* 77:1184-1194. DOI: 10.1128/JVI.77.2.1184.

Peterson A.T., Moses L.M. and Bausch D.G. (2014). Mapping Transmission Risk of Lassa Fever in West Africa: The Importance of Quality Control, Sampling Bias, and Error Weighting. *PloS One.* 9: e100711. DOI: 10.1371/journal.pone.0100711.

Pinschewer D.D., Perez M. and de la Torre J.C. (2003). Role of the Virus Nucleoprotein in the Regulation of Lymphocytic Choriomeningitis Virus Transcription and RNA Replication. *Journal of Virology*. 77:3882-3887. DOI: 10.1128/JVI.77.6.3882.

R Development Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org.

Radoshitzky S.R. et al. (2015). Past, present, and future of arenavirus taxonomy. *Archives of Virology*. 160:1851-1874. DOI: 10.1007/s00705-015-2418-y.

Rambaut A. (2012). FigTree v1.4. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A., Suchard M.A., Xie D. and Drummond A.J. (2014). Tracer v1.6. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology. Retrieved from http://beast.bio.ed.ac.uk/Tracer.

Rapp F. and Buckley S.M. (1962). Studies with the etiologic agent of Argentinian epidemic hemorrhagic fever (Junín virus). *American Journal of Pathology.* 40:63-75.

Ronquist F. and Huelsenbeck J.P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572-1574.

Russo I.-R.M., Sole C.L., Barbato M., von Bramann U. and Bruford M.W. (2016). Landscape determinants of fine-scale genetic structure of a small rodent in a heterogeneous landscape (Hluhluwe-iMfolozi Park, South Africa). *Scientific Reports.* 6:29168. DOI: 10.1038/srep29168.

Safronetz D. et al. (2010). Detection of Lassa Virus, Mali. *Emerging Infectious Diseases.* 16:1123-1126. DOI: 10.3201/eid1607.100146.

Safronetz D. et al. (2013). Geographic Distribution and Genetic Characterization of Lassa Virus in Sub-Saharan Mali. *PLoS Neglected Tropical Diseases*. 7:4-12. DOI: 10.1371/journal.pntd.0002582.

Salazar-Bravo J., Ruedas L.A. and Yates T.L. (2002). Mammalian reservoirs of arenaviruses. In: *Arenaviruses. I. The Epidemiology, Molecular and Cell Biology of Arenaviruses*. (ed.: Oldstone M.B.A.). Current Topics in Microbiology and Immunology. Heidelberg, Germany: Springer-Verlag. 262:25–64.

Salvato M., Shimomaye E. and Oldstone M.B. (1989). The Primary Structure of the Lymphocytic Choriomeningitis Virus L Gene Encodes a Putative RNA Polymerase. *Virology.* 169:377-384.

Singh M.K., Fuller-Pace F.V., Buchmeier M.J. and Southern P.J. (1987). Analysis of the Genomic L RNA Segment from Lymphocytic Choriomeningitis Virus. *Virology.* 161:448-456. DOI: 10.1016/0042-6822(87)90138-3.

Sogoba N., Feldmann H. and Safronetz D. (2012). Lassa Fever in West Africa: Evidence for an Expanded Region of Endemicity. *Zoonoses and Public Health.* 59:43-47. DOI: 10.1111/j.1863-2378.2012.01469.x.

Sogoba N. et al. (2016) Lassa Virus Seroprevalence in Sibirilia Commune, Bougouni District, Southern Mali. *Emerging Infectious Diseases.* 22:657-663. DOI: 10.3201/eid2204.151814.

Stamatakis A. (2014). RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics. 30*:1312-1313. DOI: 10.1093/bioinformatics/btu033.

Stamatakis A. (2016). The RAxML v8.2.X Manual. Retrieved from https://sco.h-its.org/exelixis/resource/download/NewManual.pdf.

Stenglein M.D. et al. (2015). Widespread Recombination, Reassortment, and Transmission of Unbalanced Compound Viral Genotypes in Natural Arenavirus Infections. *PLoS Pathogens*. 11:e1004900. DOI: 10.1371/journal.ppat.1004900.

Stenglein M.D., Sanders C., Kistler A.L., Ruby J.G., Franco J.Y., Reavill D.R., Dunker F. and DeRisi J.L. (2012). Identification, Characterization, and *In Vitro* Culture of Highly Divergent Arenaviruses from Boa Constrictors and Annulated Tree Boas: Candidate Etiological Agents for Snake Inclusion Body Disease. *mBio.* 3:e00180-12: DOI: 10.1128/mBio.00180-12.

Tavare S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. In: *Lectures on Mathematics in the Life Sciences* (ed.: Miura R.M.). Providence: American Mathematical Society. 17:57-86.

The Field Museum, Chicago, USA. (2016). Mammals of Tanzania: Rodentia Skin Key. Retrieved from http://archive.fieldmuseum.org/tanzania/SkinKey.asp?ID=336.

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.* 101:1253-1264. DOI: 10.1016/j.trstmh.2005.03.018.

Walker D.H., Wulff H., Lange J.V, and Murphy F.A. (1975). Comparative pathology of Lassa virus infection in monkeys, guinea-pigs, and *Mastomys natalensis*. *Bulletin of the World Health Organization*. 52:523-534.

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. DOI: 10.1111/j.1469-1809.1972.tb00293.x.

Wulff H., McIntosh B.M., Hamner D.B. and Johnson. K.M. (1977). Isolation of an arenavirus closely related to Lassa virus from *Mastomys natalensis* in south-east Africa. *Bulletin of the World Health Organization.* 55:441-444.

Yang Z. (1993). Maximum-Likelihood Estimation of Phylogeny from DNA Sequences When Substitution Rates Differ over Sites. *Molecular Biology and Evolution.* 10:1396-1401.

Zapata J.C. and Salvato M.S. (2013). Arenavirus variations due to host-specific adaptation. *Viruses.* 5:241-278. DOI: 10.3390/v5010241.

9. Supplementary material

Supplementary Table 1: Coordinates, elevation, sampling time and summary of arenavirus prevalence in *Mastomys natalensis* per locality. Unless mentioned otherwise I screened kidney samples and dried blood samples for arenavirus *L gene* RNA and anti-arenavirus antibodies, respectively. For RNA screening kidneys samples were initially pooled by two instead of three by T. Locus. Dried blood samples (initially pooled by two), not kidney samples, were screened in Gryseels et al. (2017). For anti-arenavirus antibody screening T. Locus did not depool the positive dried blood samples from Kimamba. The number of IFA-positive single samples at this locality was therefore estimated using the equations $p^2 + 2pn + n^2 = 1$ and $p + n = 1$ with p the proportion of positive single samples and n the proportion of negative single samples and given the proportion of negative pooled samples (n^2). * indicates that the sampled mice originated from different fields with different GPS coordinates and elevation data and that an average weighted for the number of individuals is given; ** indicates that GPS elevation data was not available and that instead elevation was estimated from a Digital Elevation Model ArcGIS layer with a resolution of 1 km from the U.S. Geological Survey's Center for Earth Resources Observation and Science.

Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled together with a Tanzanian team from the Pest Management Centre of the Sokoine University of Agriculture (SPMC)											
Ibohora	-8.70, 34.31*	1060*	August 2016	0 / 92	(0.0)	0 / 92	(0.0)	1 / 92	(1.09)	6 / 50	(12.0)
Ikokoto	-7.65, 36.13*	1241*	July 2016	0 / 42	(0.0)	0 / 42	(0.0)	0 / 42	(0.0)	0 / 44	(0.0)
Ifunda	-8.04, 35.48*	1732*	July 2016	0 / 25	(0.0)	0 / 25	(0.0)	0 / 25	(0.0)	0 / 34	(0.0)
	-8.09, 35.44*	1707*	July 2016	0 / 25	(0.0)	0 / 25	(0.0)	0 / 25	(0.0)	0 /15	(0.0)
Kibena	-9.22, 34.78*	1874*	August 2016	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)
Lilondo	-9.82, 35.35*	1443*	August 2016	0 / 7	(0.0)	0 / 7	(0.0)	0 / 7	(0.0)	0 / 7	(0.0)
	-9.84, 35.37*	1375*	August 2016	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)	0 / 19	(0.0)
	-9.85, 35.36	1243	August 2016	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)
	-9.83, 35.35*	1333*	August 2016	0 / 3	(0.0)	0 / 3	(0.0)	0 / 3	(0.0)	0 / 3	(0.0)
Mafinga	-8.25, 35.26*	1802*	August 2016	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)

Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled together with a Czech-Tanzanian team from the Institute of Vertebrate Biology (IVB) and the SPMC											
Ikokoto	-7.65, 36.13	1250	July 2016	0 / 8	(0.0)	0 / 8	(0.0)	0 / 8	(0.0)	0 / 6	(0.0)
Kidatu	-7.66, 36.97	297	July 2016	0 /3	(0.0)	0 /3	(0.0)	0 /3	(0.0)	0 / 3	(0.3)
	-7.68, 37.01	282	July 2016	0 / 13	(0.0)	0 / 13	(0.0)	0 / 13	(0.0)	0 / 12	(0.0)
	-7.68, 37.01	280	July 2016	0 / 9	(0.0)	0 / 9	(0.0)	0 / 9	(0.0)	2 / 9	(22.2)
Kidayi 'A'	-7.53, 36.66	517	July 2016	0 / 26	(0.0)	0 / 26	(0.0)	0 / 26	(0.0)	0 / 26	(0.0)
Lugalo	-7.76, 35.85	1583	July 2016	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)	0 / 25	(0.0)
Mahenge	-7.64, 36.26	669	July 2016	0 / 50	(0.0)	0 / 50	(0.0)	0 / 50	(0.0)	0 / 21	(0.0)
Mang'ula	-7.84, 36.92	292	July 2016	0 /15	(0.0)	0 /15	(0.0)	0 /15	(0.0)	1 / 15	(6.7)
	-7.84, 36.89	310	July 2016	0 / 5	(0.0)	0 / 5	(0.0)	0 / 5	(0.0)	0 / 5	(0.0)
Mbuyuni	-7.51, 36.53	529	July 2016	0 / 16	(0.0)	0 / 16	(0.0)	0 / 16	(0.0)	0 / 13	(0.0)
	-7.50, 36.52	525	July 2017	0 / 12	(0.0)	0 / 12	(0.0)	0 / 12	(0.0)	0 / 12	(0.0)
Mikumi	-7.41, 36.98	521	July 2016	0 /20	(0.0)	0 /20	(0.0)	0 /20	(0.0)	3 / 20	(15.0)
Msimba	-7.01, 36.99	726	July 2016	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)
Signal	-8.04, 36.84	265	July 2016	0 /26	(0.0)	0 /26	(0.0)	0 /26	(0.0)	3 / 26	(11.5)

Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled by a Czech-Tanzanian team from the IVB and SPMC											
Amboni caves	-5.07, 39.06	22	June 2015	0 / 2	(0.0)	0 / 2	(0.0)	0 / 2	(0.0)	-	-
Chabima	-6.85, 36.83	701	June 2015	0 / 18	(0.0)	0 / 18	(0.0)	0 / 18	(0.0)	-	-
Gairo	-6.13, 36.84	1279	June 2016	0 / 24	(0.0)	0 / 24	(0.0)	0 / 24	(0.0)	-	-
Handeni	-5.45, 38.04	703	June 2015	0 / 2	(0.0)	0 / 2	(0.0)	0 / 2	(0.0)	-	-
Ibuti	-6.14, 36.90	1323	June 2016	1 / 9	(11.1)	0 / 9	(0.0)	0 / 9	(0.0)	-	-
Ilunda	-9.02, 34.83	1827	June 2015	0 / 30	(0.0)	0 / 30	(0.0)	0 / 30	(0.0)	0 / 27	(0.0)
Kanga	-9.45, 33.89	521	June 2015	0 / 18	(0.0)	0 / 18	(0.0)	0 / 18	(0.0)	-	-
Kiberashi	-5.38, 37.48	1034	June 2015	5 / 42	(11.9)	0 / 42	(0.0)	0 / 42	(0.0)	-	-
Kibirashi	-5.40, 37.43	1228*	July 2016	1 / 7	(14.3)	0 / 7	(0.0)	0 / 7	(0.0)	-	-
Kijungu	-5.39, 37.19	1343	June 2016	0 / 8	(0.0)	0 / 8	(0.0)	0 / 8	(0.0)	-	-
Kimbe	-5.79, 37.63	705*	June 2016	0 / 20	(0.0)	0 / 20	(0.0)	0 / 20	(0.0)	-	-
Kireguru	-5.47, 37.61	848	June 2015	3 / 22	(13.6)	0 / 22	(0.0)	0 / 22	(0.0)	-	-
Kisongwe	-6.63, 36.99	878	June 2015	0 / 3	(0.0)	0 / 3	(0.0)	0 / 3	(0.0)	-	-
Komtema	-5.41, 37.95	697*	July 2016	0 / 10	(0.0)	0 / 10	(0.0)	0 / 10	(0.0)	-	-
Korogwe	-5.24, 38.50	313*	June 2015	0 / 11	(0.0)	0 / 11	(0.0)	0 / 11	(0.0)	-	-
Kunke	-6.13, 37.68	372*	June 2016	0 / 15	(0.0)	1 / 15	(6.7)	0 / 15	(0.0)	-	-
Kwekivu	-5.78, 37.33	827*	June 2016	0 / 22	(0.0)	0 / 22	(0.0)	0 / 22	(0.0)	-	-
Mafleti	-5.42, 37.90	692	July 2016	2 / 12	(16.7)	0 / 12	(0.0)	0 / 12	(0.0)	-	-

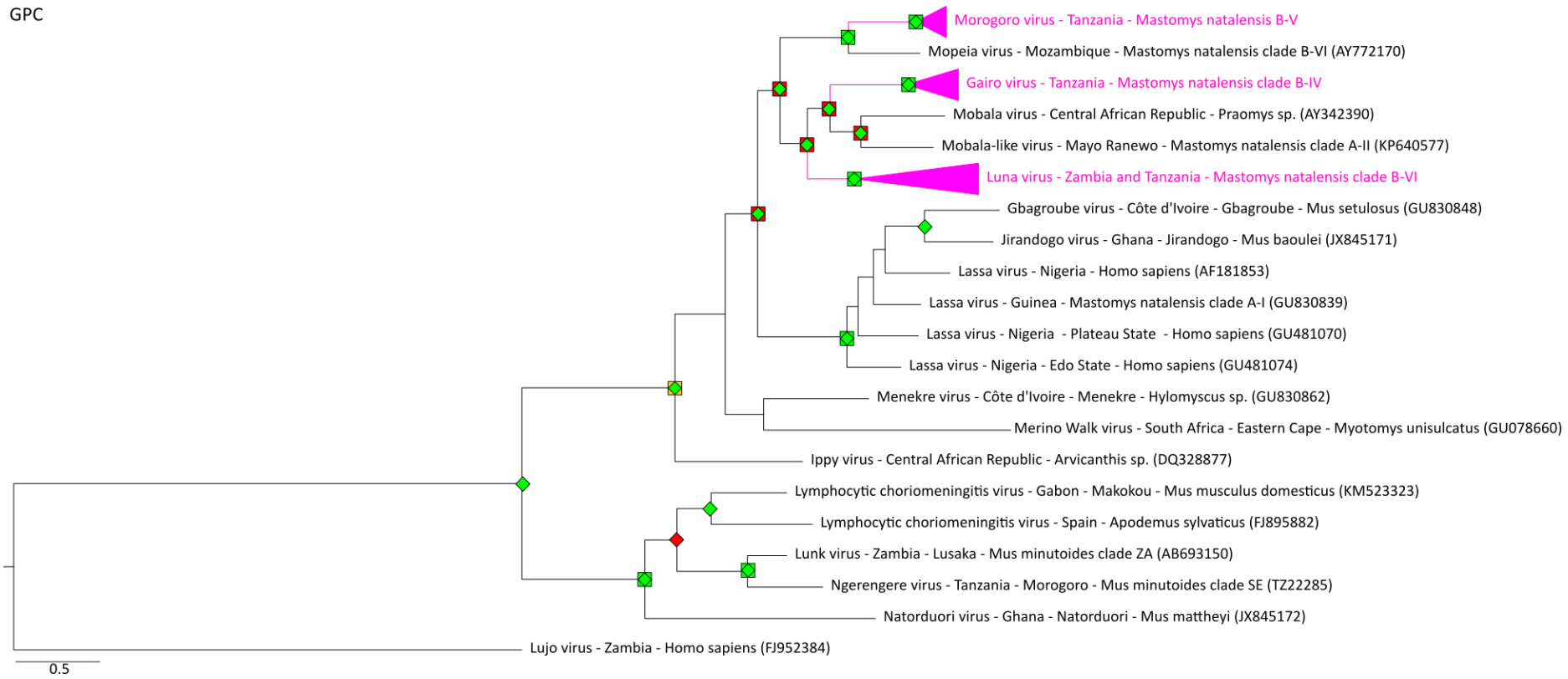
Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled by a Czech-Tanzanian team from the IVB and SPMC											
Magamba	-5.56, 38.08	535*	July 2016	1 / 13	(7.7)	0 / 13	(0.0)	0 / 13	(0.0)	-	-
Makasini	-5.46, 37.60	878	June 2015	5 / 27	(18.5)	0 / 27	(0.0)	0 / 27	(0.0)	-	-
Masenge	-6.36, 36.93	1780*	June 2015	0 / 12	(0.0)	0 / 12	(0.0)	0 / 12	(0.0)	-	-
Mbande	-6.10, 36.33	975*	June 2016	9 / 28	(32.1)	0 / 28	(0.0)	0 / 28	(0.0)	-	-
Meriongima	-5.34, 37.03	1271	June 2016	3 / 18	(16.7)	0 / 18	(0.0)	0 / 18	(0.0)	-	-
Msasa	-5.50, 38.06	598	July 2016	1 / 7	(14.3)	0 / 7	(0.0)	0 / 7	(0.0)	-	-
Mswaki	-5.46, 37.78	715	July 2016	1 /14	(7.1)	0 /14	(0.0)	0 /14	(0.0)	-	-
	-5.47, 37.76	725	July 2016	2 / 13	(15.4)	0 / 13	(0.0)	0 / 13	(0.0)	-	-
Mtanana	-6.08, 36.59	1151	June 2016	4 / 18	(22.2)	0 / 20	(0.0)	0 / 20	(0.0)	-	-
Mtumbatu	-6.14, 36.98	1205	June 2016	0 / 26	(0.0)	0 / 26	(0.0)	0 / 26	(0.0)	-	-
Musilibuha	-6.39, 37.02	867	June 2015	0 / 8	(0.0)	0 / 8	(0.0)	0 / 8	(0.0)	-	-
Ndelema	-5.43, 38.00	624	July 2016	0 / 12	(0.0)	0 / 12	(0.0)	0 / 12	(0.0)	-	-
Ngana	-9.59, 33.69	542	June 2015	0 / 36	(0.0)	0 / 36	(0.0)	3 / 36	(8.3)	-	-
Nundu	-9.42, 34.85	2004	June 2015	0 / 6	(0.0)	0 / 6	(0.0)	0 / 6	(0.0)	0 / 6	(0.0)
Picha Ya Ndege	-5.68, 38.18	493	July 2016	0 / 6	(0.0)	0 / 6	(0.0)	0 / 6	(0.0)	-	-
Sindeni	-5.35, 38.21	500	July 2016	0 / 8	(0.0)	0 / 8	(0.0)	0 / 8	(0.0)	-	-
Songe	-5.55, 37.29	1227	June 2016	0 / 18	(0.0)	0 / 18	(0.0)	0 / 18	(0.0)	-	-
Tanga	-5.07, 39.10	22	June 2015	0 / 1	(0.0)	0 / 1	(0.0)	0 / 1	(0.0)	-	-

Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled and screened in Gryseels et al. (2017)											
Bwawani	-6.66, 38.03	285*	January 2011	0 / 61	(0.0)	3 / 61	(4.9)	0 / 61	(0.0)	12 / 53	(22.6)
Chakwale	-6.05, 36.96	1080**	August 2012	1 / 85	(1.2)	0 / 85	(0.0)	0 / 85	(0.0)	1 / 85	(1.2)
Chalinze	-6.66, 38.36	202*	December 2010 - January 2011	0 / 78	(0.0)	2 / 78	(2.6)	0 / 78	(0.0)	2 / 77	(2.6)
Dumila	-6.38, 37.36	426*,**	December 2009	0 / 152	(0.0)	4 / 152	(2.6)	0 / 152	(0.0)	29 / 152	(12.5)
Majawanga	-6.11, 36.82	1272**	August 2012	17 / 106	(16.0)	0 / 106	(0.0)	0 / 106	(0.0)	24 / 106	(22.6)
Dakawa	-6.45, 37.54	359*	December 2009	0 / 35	(0.0)	1 / 35	(2.9)	0 / 35	(0.0)	5 / 35	(14.3)
Itigi	-5.74, 34.41	1307	July - August 2010	0 / 10	(0.0)	0 / 10	(0.0)	0 / 10	(0.0)	-	-
Lihale	-10.80, 35.17	913	July - August 2008	0 / 14	(0.0)	0 / 14	(0.0)	0 / 14	(0.0)	-	-
Maguha	-6.29, 37.19	793*	December 2009	0 / 23	(0.0)	0 / 23	(0.0)	0 / 23	(0.0)	1 /22	(4.5)
Mbulu	-4.08, 35.60	1355**	January 2011 & November 2011	2 / 47	(4.3)	0 / 47	(0.0)	0 / 47	(0.0)	-	-
Mikese	-6.77, 37.86	424*	January 2011	0 / 98	(0.0)	2 / 98	(2.0)	0 / 98	(0.0)	9 / 98	(9.2)
Mkundi	-6.62, 37.60	453*	December 2009	0 / 21	(0.0)	1 / 21	(4.8)	0 / 21	(0.0)	2 / 21	(9.5)
Morogoro	-6.85, 37.65	494*	December 2009	0 / 133	(0.0)	4 / 133	(3.0)	0 / 133	(0.0)	5 / 133	(3.8)
		509*	December 2010	0 / 157	(0.0)	4 / 157	(2.5)	0 / 157	(0.0)	8 / 89	(9.0)
Shinyanga-Lubaga	-3.64, 33.42	1127	July - August 2009	1 / 4	(25.0)	0 / 4	(0.0)	0 / 4	(0.0)	-	-
Ubena	-6.64, 38.19	239*	January 2011	0 / 54	(0.0)	3 / 54	(5.6)	0 / 54	(0.0)	5 / 52	(9.6)

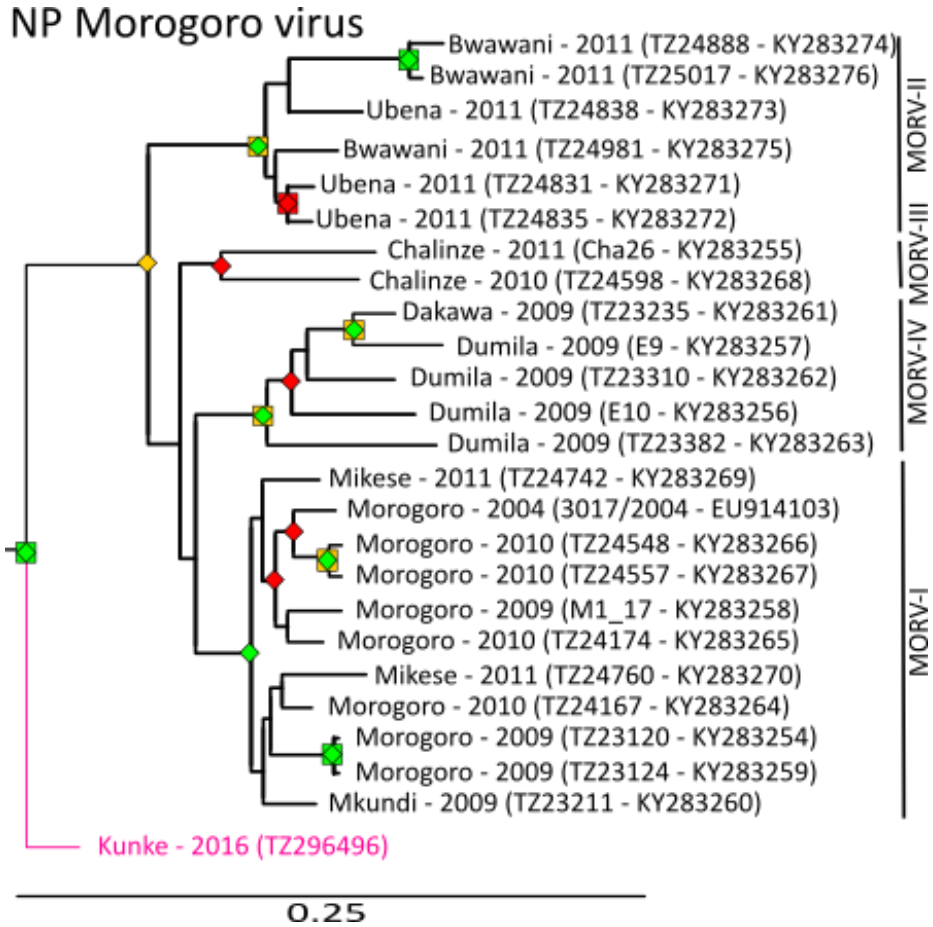
Locality	Coordinates	Elevation (m)	Sampling time	GAIV RNA positive / no. tested (prevalence in %)		MORV RNA positive / no. tested (prevalence in %)		LUAV RNA positive / no. tested (prevalence in %)		Antibody positive / no. tested (prevalence in %)	
Sampled by T. Locus, J. Favaits and a Tanzanian team from the SPMC, screened by T. Locus											
Dakawa	-6.45, 37.54*	349*	August 2015	0 / 145	(0.0)	8 / 145	(5.5)	0 / 145	(0.0)	-	-
Ikwiriri	7.98, 38.99*	12*	September 2015	0 / 144	(0.0)	0 / 144	(0.0)	0 / 144	(0.0)	0 / 144	(0.0)
Kimamba	-5.81, 37.80*	494*	August 2015	0 / 138	(0.0)	0 / 138	(0.0)	0 / 138	(0.0)	17 / 138	(12.4)
Matipwili	-6.24, 38.71*	5*	August 2015	0 / 30	(0.0)	1 / 30	(3.3)	0 / 30	(0.0)	-	-
Selous	-7.78, 38.23*	50*	August 2015	0 / 162	(0.0)	0 / 162	(0.0)	0 / 162	(0.0)	0 / 162	(0.0)

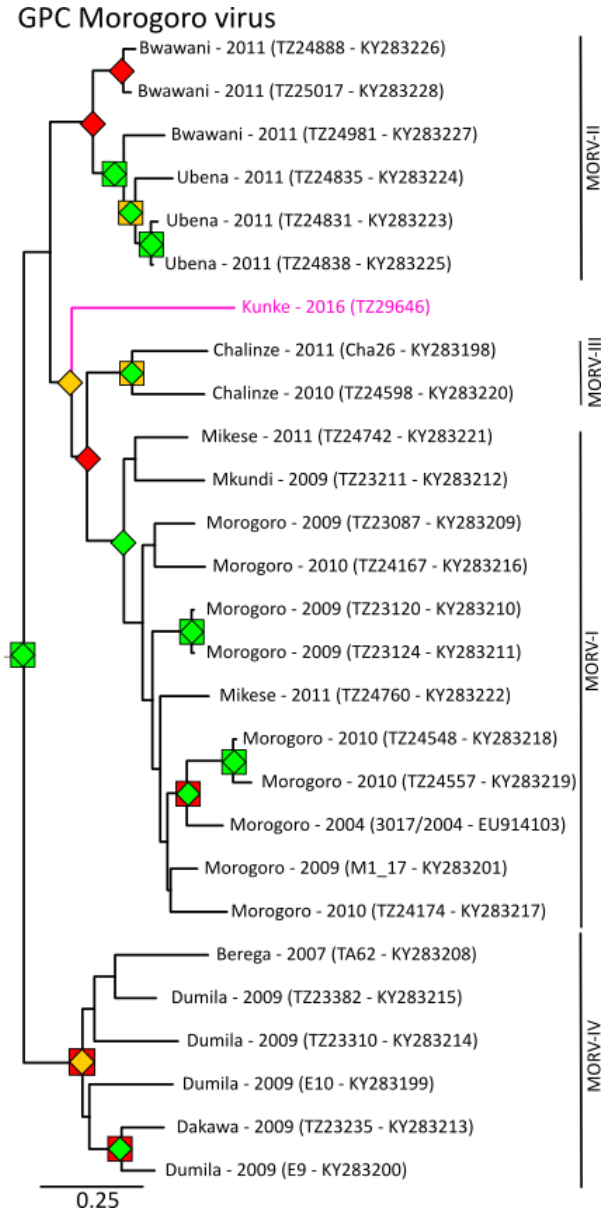
Supplementary Table 2: *Mus minutoides* kidney samples screened for *L gene* arenavirus RNA. Six were sampled together with a Tanzanian research team from the Pest Management Centre of the Sokoine University of Agriculture (SPMC; Morogoro, Tanzania); 15 were sampled by a Czech-Tanzanian research team from the Institute of Vertebrate Biology (IVB; Studenec, Czech Republic) and the SPMC. A fragment of the *cytochrome b gene* was amplified to determine the mitochondrial (Mt) lineage (sensu Bryja et al. (2014)) by A. Hánová from the IVB.

Locality	Coordinates	Mt lineage	Arenavirus RNA positive / no. tested (prevalence in %)
Sampled together with a Tanzanian team from the SPMC			
Ifunda	-8.09, 35.44	TZw	0 / 1 (0.0)
Lilondo	-9.82, 35.35	SE	0/ 3 (0.0)
	-9.84, 35.37	SE	0 / 2 (0.0)
Sampled by a Czech-Tanzanian team from the IVB and SPMC			
Rombo	-3.19, 37.64	SE	0 / 1 (0.0)
Kiberashi	-5.38, 37.48	SE	0 / 2 (0.0)
Kireguru	-5.47, 37.61	SE	0 / 2 (0.0)
Masenge	-6.36, 36.93	SE	0 / 1 (0.0)
Ngana	-9.59, 33.69	SE	1 / 2 (50.0)
Nundu	-9.42, 34.85	TZw	0 / 1 (0.0)
Ilunda	-9.02, 34.83	TZw	0 / 4 (0.0)
Kunke	-6.12, 37.70	SE	0 / 1 (0.0)
Mswaki	-5.46, 37.78	SE	0 / 1 (0.0)

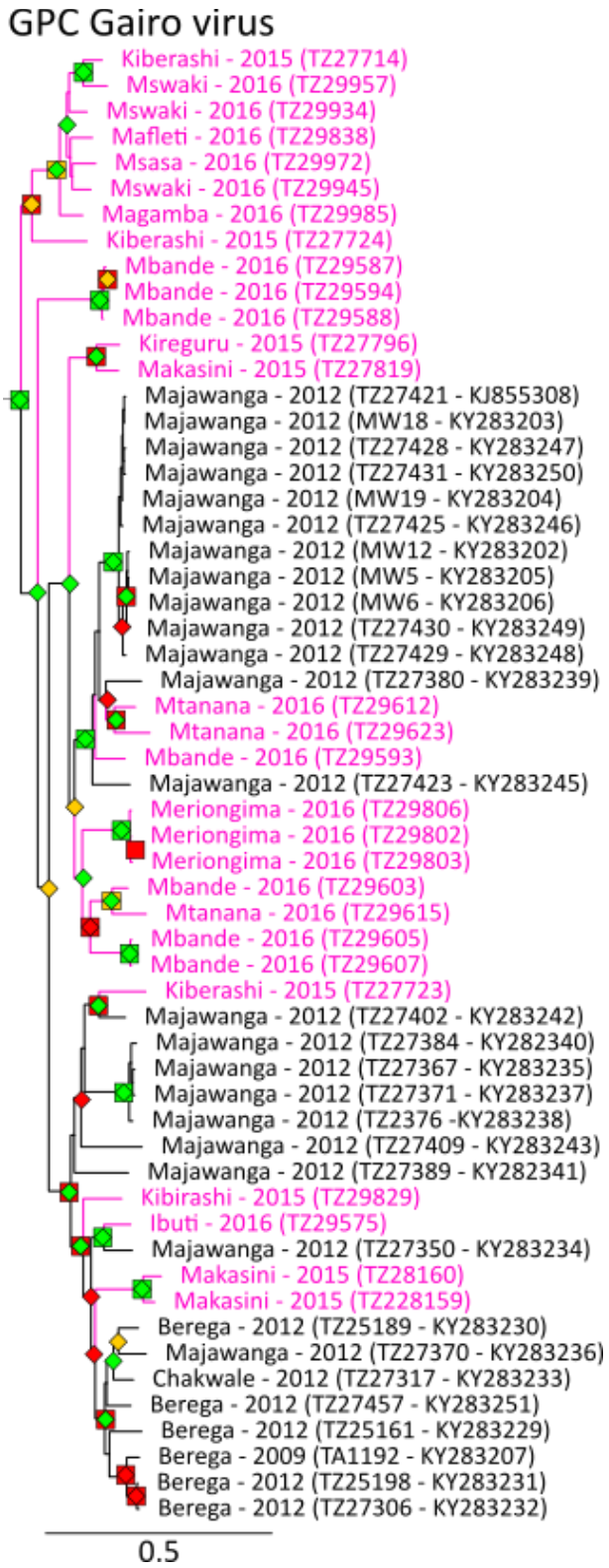


Supplementary Figure 1: *GPC gene* Bayesian inference tree. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/70 (Maximum likelihood), red for supports between 0.70/70 and 0.90/ 90, yellow for supports between 0.90/90 and 0.95/95, and green for supports above 0.95/95. Taxa are named as the virus species followed by the sampling country, the locality or region (if available), the host species extracted from and the accession number from GenBank or a sample code starting with 'TZ' between brackets. Gairo virus, Morogoro virus and Luna virus sequences are collapsed to triangles (see Supplementary Figures 2, 3 and 4 for these branches). Taxa are coloured fuchsia if the taxon is or contains a sample screened in this study. The scale bar represents the number of nucleotide substitutions per site.

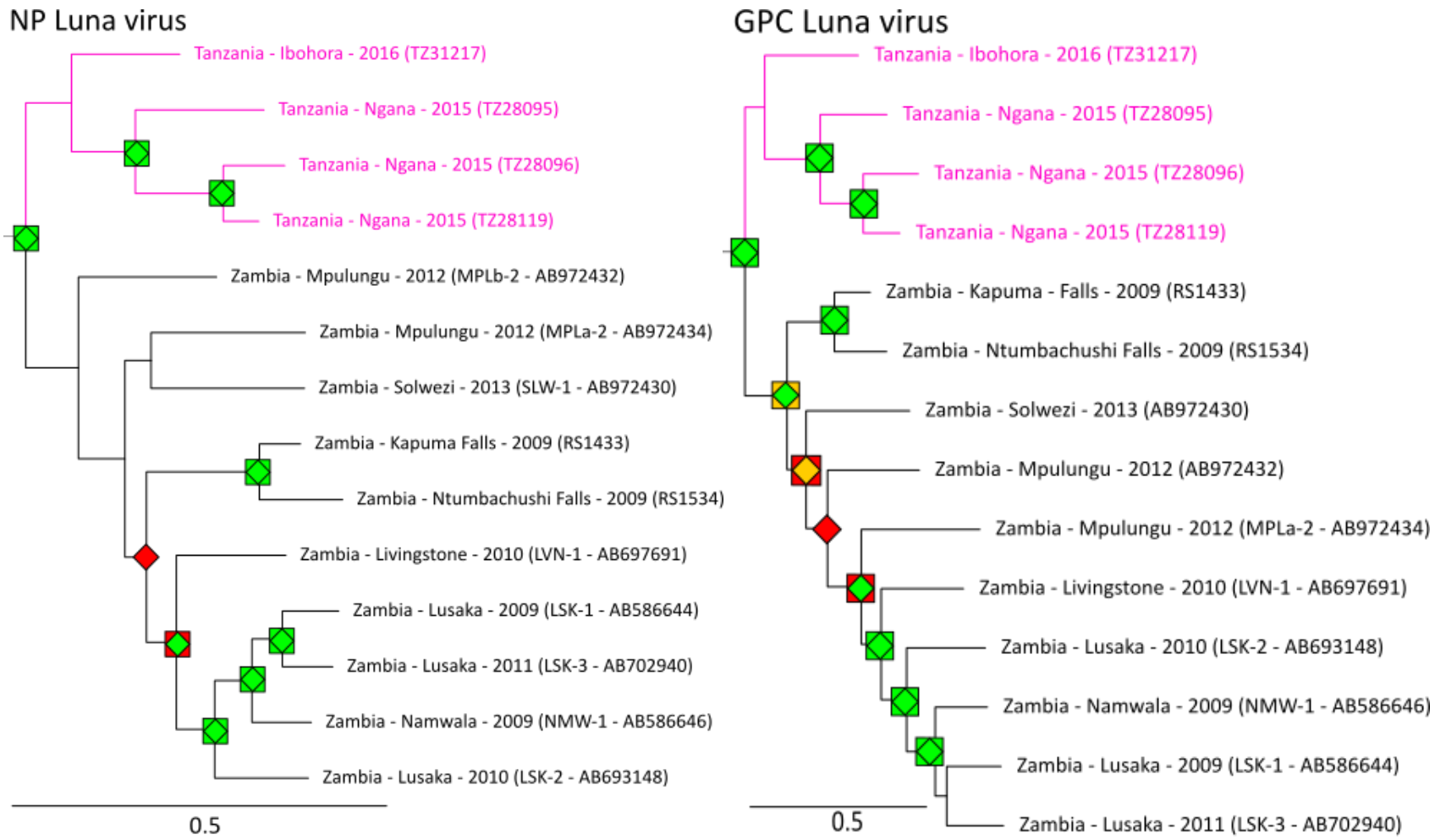




Supplementary Figure 2: Morogoro virus *NP* and *GPC gene* Bayesian inference trees. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/70 (Maximum likelihood), red for supports of 0.70/70 to 0.90/90, yellow for supports of 0.90/90 to 0.95/95, and green for supports of 0.95/95 and above. Sequences are named as the locality, the year and a sample code and accession number from GenBank between brackets. Clades with Roman numbers indicate clades described in Gryseels et al. (2017). The fuchsia sequence is new to this study. The scale bar represents the number of nucleotide substitutions per site.



Supplementary Figure 3: Gairo virus *NP* and *GPC gene* Bayesian inference trees. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/70 (Maximum likelihood), red for supports of 0.70/70 to 0.90/90, yellow for supports of 0.90/90 to 0.95/95, and green for supports of 0.95/95 and above. Sequences are named as the locality, the year and a sample code and accession number from GenBank between brackets. Sequences are coloured fuchsia if they are new to this study. The scale bar represents the number of nucleotide substitutions per site.



Supplementary Figure 4: Luna virus *NP* and *GPC gene* Bayesian inference trees. Diamonds and squares indicate node support for Bayesian inference and Maximum likelihood analyses, respectively. Node support categories are as follows: no symbol for supports under 0.70 (Bayesian inference)/70 (Maximum likelihood), red for supports of 0.70/70 to 0.90/90, yellow for supports of 0.90/90 to 0.95/95, and green for supports of 0.95/95 and above. Sequences are named as the sampling country, the locality, the year and a sample code and accession number from GenBank between brackets. Sequences are coloured fuchsia if they are new to this study. The scale bar represents the number of nucleotide substitutions per site.