



Dhati Welel virus, the missing mammarenavirus of the widespread *Mastomys natalensis*

Authors: De Bellocq, Joëlle Goüy, Bryjová, Anna, Martynov, Aleksey A., and Lavrenchenko, Leonid A.

Source: Journal of Vertebrate Biology, 69(2)

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: <https://doi.org/10.25225/jvb.20018>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Dhati Welel virus, the missing mammarenavirus of the widespread *Mastomys natalensis*

Joëlle GOÛY DE BELLOCQ^{1*}, Anna BRYJOVÁ¹, Aleksey A. MARTYNOV² and Leonid A. LAVRENCHENKO²

¹ Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, 603 65 Brno, Czech Republic;
e-mail: joellegouy@gmail.com, bryjova@iob.cz

² A. N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 119071 Moscow, Russia;
e-mail: Aleksey.martynov@mail.ru, llavrenchenko@gmail.com

► Received 20 February 2020; Accepted 17 April 2020; Published online 19 June 2020

Abstract. The Natal multimammate mouse, *Mastomys natalensis*, occurs throughout sub-Saharan Africa. Mitochondrial phylogenetics indicate this species was fragmented during the Pleistocene, forming six matrilineage phylogroups: A-I, A-II, A-III, B-IV, B-V, B-VI with distinct ranges. All except the A-III lineage are identified as natural reservoirs of mammarenaviruses. *M. natalensis* A-III is found in western Ethiopia and is the only lineage reported in the country. While screening 203 small mammal samples from Dhati Welel National Park for mammarenaviruses, we detected mammarenavirus RNA in nine samples, eight from *M. natalensis* and one from *M. awashensis*. A sequence similarity search and phylogenetic analysis confirmed the *M. natalensis* mitochondrial DNA belongs to the A-III lineage. We characterised the complete virus genome, which showed typical mammarenavirus organisation. Phylogenetic analysis indicated it clusters with Gairo virus found in *M. natalensis* B-IV in Tanzania, while showing sufficient divergence from other mammarenaviruses to be considered as a new species, for which we proposed the name Dhati Welel. Additional sampling in the *M. natalensis* A-III phylogeographic range should help determine whether the detection of the virus in *M. awashensis* represents a local spill-over or if the virus circulates in both *Mastomys* species.

Key words: arenavirus, Natal multimammate mouse, zoonotic disease, Ethiopia, phylogeny

Introduction

The Natal multimammate mouse, *Mastomys natalensis*, is one of the most widespread rodent species in sub-Saharan Africa (Denys et al. 2017). It has a wide ecological range, occurring in natural savannahs, agricultural fields and houses but absent from deserts, alpine regions and rainforests (Denys et al. 2017). A study conducted across its entire range and based on a mitochondrial (Mt) marker, cytochrome *b* (*cytb*), showed six main matrilineages, forming two monophyletic clades A and B, each subdivided into three sub-clades (A-I,

A-II, A-III, B-IV, B-V, B-VI) (Colangelo et al. 2013). A-I is found in western sub-Saharan Africa; A-II from eastern Nigeria to the Democratic Republic of Congo; A-III from western Ethiopia to western Kenya, B-IV in southern Kenya, much of Tanzania and Rwanda; B-V in south-eastern Tanzania and northern Mozambique while B-VI covers a large part of southern Africa from south-west Tanzania through Zambia, Mozambique, Malawi, Botswana, Zimbabwe, to South Africa (Colangelo et al. 2013, Gryseels et al. 2017, Martynov et al. 2020) (Fig. 1). These Mt matrilineages likely correspond to taxa that are distinct genome-wide: this has

* Corresponding Author

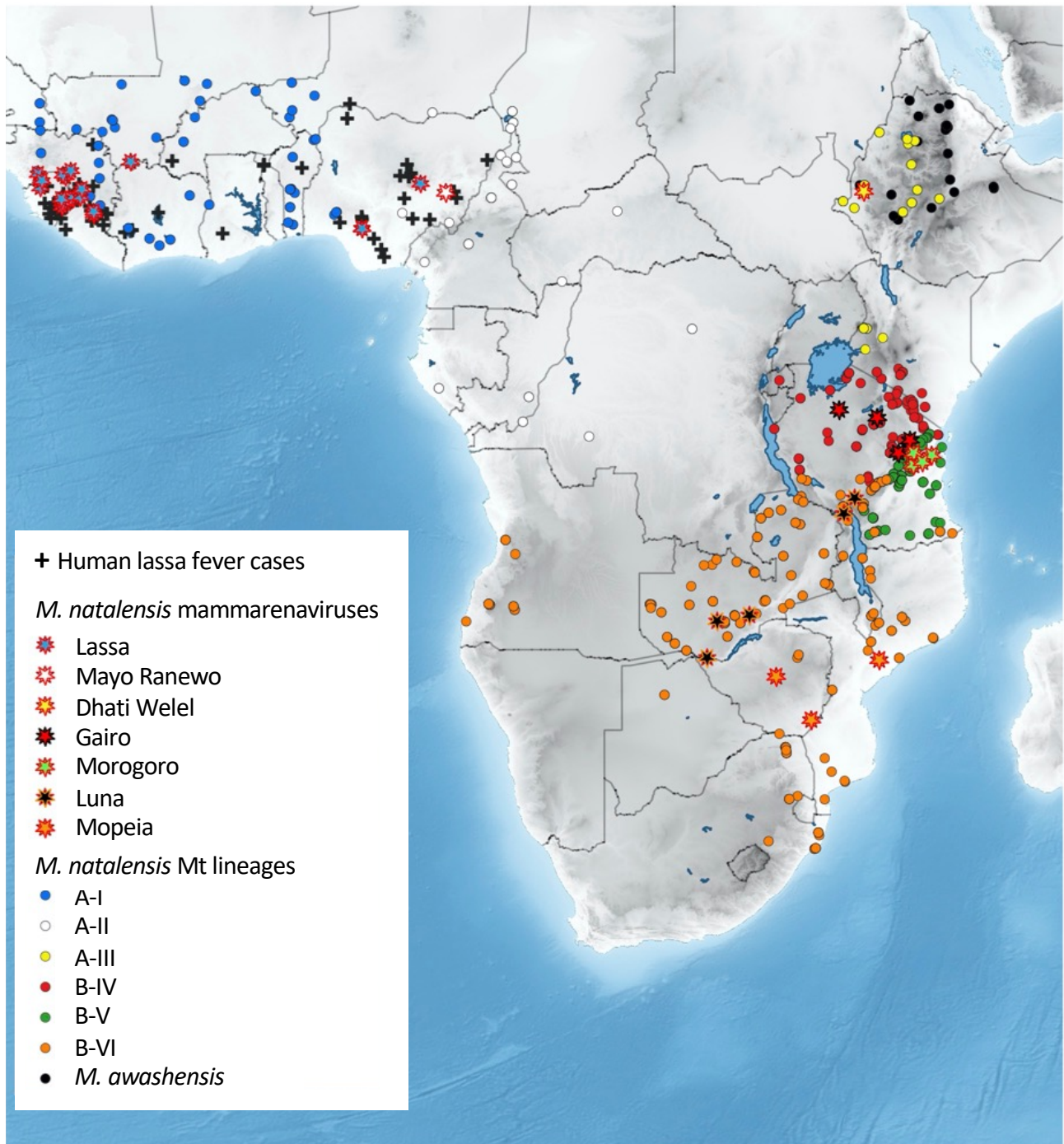


Fig. 1. Map of Africa showing the geographic distribution of the six mitochondrial lineages of *M. natalensis* (coloured circles), the geographic distribution of *M. awashensis* (black circles), human Lassa fever cases (black +) and the different mammarenaviruses (coloured stars) found in *M. natalensis* mitochondrial lineages. Data compiled from Fichet-Calvet & Rogers (2009), Colangelo et al. (2013), Olayemi et al. (2016), Gryseels et al. (2017), Cuypers et al. (2020), Martynov et al. (2020) and Hánová et al. (unpublished data).

been shown by Gryseels et al. (2017) using Mt *cytb*, Y chromosome and nuclear markers across Mt matrilineages B-IV and B-V in Tanzania. *M. natalensis* is also considered an important rodent pest in Africa affecting wheat, maize and other cereals and acting as a vector and reservoir of zoonotic diseases (Leirs 1994) including the agent of Lassa fever in western Africa, the Lassa mammarenavirus (Monath et al. 1974).

Mammarenaviruses are single-stranded RNA viruses from the family Arenaviridae. They typically infect murid rodents, New World (NW) mammarenaviruses infecting subfamilies Sigmodontinae and Neotominae in the Americas (with one exception, Tacaribe virus, isolated from bats) and Old World (OW) mammarenaviruses infecting the subfamily Murinae in Eurasia and Africa (with the exception of the Lymphocytic



choriomeningitis virus, infecting the house mouse and so having a worldwide distribution) (Radoshitzky et al. 2019). They have a bi-segmented genome with ambisense organization: the large segment (L) encodes a RNA-dependent RNA polymerase (L) and a zinc finger matrix protein (Z); the small segment (S) encodes a glycoprotein precursor (GPC) and a nucleocapsid protein (NP) (Radoshitzky et al. 2019). Mammarenaviruses do not generally infect mammals other than their natural hosts, in which they induce asymptomatic infections. However, some members can occasionally spill-over to humans and can be highly pathogenic causing haemorrhagic fevers such as the Lujó and Lassa mammarenaviruses in Africa (Frame et al. 1970, Briese et al. 2009), the latter being responsible for several thousand deaths annually (Richmond & Baglolle 2003).

So far, five out of the six matrilineages of *M. natalensis* have been found to carry at least one mammarenavirus: Lassa in A-I (Lecompte et al. 2006, Olayemi et al. 2016), a Mobala-like virus in A-II (strain Mayo Ranewo) (Olayemi et al. 2016), Gairo virus in B-IV (Gryseels et al. 2015, 2017, Cuypers et al. 2020), Morogoro virus in B-V (Günther et al. 2009, Cuypers et al. 2020) and Luna and Mopeia virus in B-VI (Wulff et al. 1977, Johnson et al. 1981, Ishii et al. 2011, Cuypers et al. 2020) (Fig. 1). Lassa virus has also been reported in three *M. natalensis* A-II Mt lineage individuals, but in the area of secondary contact between A-I and A-II lineages (Olayemi et al. 2016). It seems likely these individuals were A-I genome wide (they are located on the A-I side of the presumed barrier between the two sub taxa, the River Niger) but have introgressed A-II Mt, a common phenomenon in contact zones (see Gryseels et al. 2017 for details). To date only the *M. natalensis* A-III matrilineage lacked an associated mammarenavirus. This may partly due to the restricted geographic distribution of this lineage: it is one of the smallest, only reported in the western side of the Rift Valley in Ethiopia (Martynov et al. 2020) and in the vicinity of Lake Victoria in Kenya (Colangelo et al. 2013) (Fig. 1). In Ethiopia, it is the only *M. natalensis* matrilineage present, though the genus *Mastomys* is widespread in Ethiopia with four distinct species represented (*kollmannspergeri*, *awashensis*, *erythroleucus* and *natalensis*, see Martynov et al. 2020 this issue for details).

We use the opportunity of availability of samples from the Dhati Welel National Park in Ethiopia, which encompasses the range of *M.*

natalensis A-III lineage (Martynov et al. 2020), to search for the presence of mammarenaviruses. Ethiopia is renowned for its high diversity and endemism of plant and animal species and this also applies to rodent-borne-viruses: so far two mammarenaviruses have been reported in *Stenocephalemys albipes* and *M. awashensis* and an orthohantavirus, Tigray, in *S. albipes* and *Stenocephalemys* sp. A (this latter species is formally described in Mizerovská et al. 2020) in northern Ethiopia (Meheretu et al. 2012, 2019, Goüy de Bellocq et al. 2016). The samples from this study were previously screened for the presence of orthohantaviruses and one Cape pipistrelle bat was found positive for a strain of Mouyassué virus (Těšíková et al. 2017).

Material and Methods

Small mammal trapping and sample collection

218 small mammals (rodents, shrews and bats) were trapped during February 02-22, 2014 in the Dhati Welel National Park located in western lowlands of Ethiopia (09°13'33" N 34°52'37" E, elevation 1427 m) (Table S1). Samples were collected in four habitats: (1) gallery semi-evergreen forest; (2) wetlands with swamp grass vegetation with *Cipirus* sp. along the main riverbed of the River Dabus and its tributaries; (3) ecotonal area along the border of the wetland with savannah type vegetation; (4) tall grass savannah with sparse *Combretum-Terminalia* wood vegetation. Small terrestrial mammals (rodents and shrews) were captured using Sherman live traps (23 × 9.5 × 8 cm) set up each night and baited with a mixture of sliced carrot and sunflower oil or wheat with peanut butter. Root-rats were caught using a range of handmade traps. Bats were captured with the use of nylon mist-nets (size 10 × 4.5 m), put in the bat foraging sites and across flying paths, and also by mobile flap-trap (Borissenko 1999). Blood was collected on pre-punched filter papers. Preliminary species identification was performed in the field based on morphological characteristics and later confirmed by Mt *cytb* sequencing following Lecompte et al. (2002). GenBank accession numbers (AN) of representatives of some rodent species and individuals positive for mammarenaviruses are given in Table S1.

Mammarenavirus screening and prevalence analysis

203 dried blood samples for which enough blood was collected were screened for the presence of



mammarenavirus RNA. RNA was extracted from samples pooled by two or three following Goüy de Bellocq et al. (2010) i.e. using QIAamp Viral RNA Mini Kit reagents (Qiagen) with Zymo-Spin IIC columns (Zymo Research). cDNA was obtained with Maxima Reverse Transcriptase (Thermo Scientific) using random hexamers and then screened for mammarenaviruses by targeting a 340 nucleotide portion of the L gene (Vieth et al. 2007) using Phusion Hot Start High-Fidelity DNA Polymerase (Thermo Scientific). Positive pools were resolved by performing RNA extraction on individual dried blood samples, followed by a one-step RT-PCR assay using the Invitrogen SuperScript IV One-Step RT-PCR System (Thermo Scientific). An additional one-Step RT-PCR was performed on positive resolved samples to attempt sequencing a part of the GPC gene using primers OWS0001-fwd and OWS1000-rev (Ehichioya et al. 2011). Amplicons with bands of the expected size were purified with an Exo-CIP PCR clean-up protocol and Sanger sequenced at GATC Biogen (Köln, Germany) and sequences deposited in GenBank (AN: MT078820-MT078831). We used Quantitative Parasitology, version 3.0 (Rózsa et al. 2000) to estimate mammarenavirus prevalence with 95% confidence intervals (95% CI) estimated with Sterne's exact method (Reiczigel 2003). We used the Fisher exact test to determine whether the prevalence of infection differed between the positive host species.

Whole genome sequencing and assembly of Dhati Welel virus

We selected sample LAV2586, which showed a very bright band on the agarose gel after the L screening, for whole genome sequencing. The Ovation RNA-Seq system V2 (NuGEN) was used for cDNA synthesis before library preparation with the KAPA HyperPrep kit (Roche). Dual indexed library was sequenced with 150 paired-end (PE) reads on Illumina HiSeq X platform at BGI Genomics (Hong-Kong) together with 31 other samples from a wider RNA virome study. After read de-multiplexing we obtained a total of 37,828,616 PE raw reads. After quality filtering and trimming, we used 28,461,354 PE reads for a de-novo assembly using SPAdes (Bankevich et al. 2012). Assembled contigs were identified using Blobtools (Laetsch & Blaxter 2017). We identified five contigs corresponding to large parts of the four genes of the virus: almost complete NP (95 nucleotides missing at the 5' end), almost complete GPC (28 nucleotides missing at the 3' end), two

contigs covering almost all of the L gene (485 nucleotides missing at the 5' end) and one contig covering the full Z and part of the intergenic region of the L segment. Missing parts of the genome were completed by Sanger sequencing (primers in Table S2). Geneious mapper in Geneious 8.0.5 was used to finalise the assembly and estimate the read coverage after removal of duplicated reads. Potential N-glycosylation sites were searched for on the NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). From the SPAdes and Blobtools analysis of the same sample (LAV2586) we also obtained the almost complete (88%) Mt genome of the host rodent which we completed using Geneious mapper and submitted to GenBank (AN: MT093212).

Phylogenetic analysis

Mt sequences from *Mastomys* positive individuals were aligned with representative Mt sequences of this rodent genus available on GenBank. Mammarenavirus nucleotide sequences obtained from the molecular screening and/or from the *de novo* assembly were aligned with the sequence coding parts of the NP, GPC and L genes of OW mammarenavirus representatives for which the full genomes are available (see Table S3). For the L gene, we also added shorter sequences of viruses relevant for this study (i.e. mammarenaviruses found in Ethiopia or in other *Mastomys* species) (see Table S3). The nucleotide alignment was performed based on amino-acid sequences in Geneious using MAFFT (Katoh et al. 2002). We used the Bayesian Information Criterion (BIC) in JModelTest (Posada 2008) to evaluate 40 nested models of nucleotide substitution. The substitution model best fitting the data for *Mastomys* Mt sequences was the HKI + I + Γ and GTR + I + Γ for the mammarenavirus L, GPC and NP genes. Phylogenetic analyses were performed using Bayesian inference implemented in MrBayes 3.2.2 (Ronquist et al. 2012). We used the default priors for all parameters and two independent runs were conducted with 10,000,000 generations per run; trees and parameters were sampled every 500 generations. Runs were initiated from random trees, and three hot chains plus one cold chain were used in all analyses. Convergence was assessed by examining the average standard deviation of split frequencies and the potential scale reduction factor. For each run, the first 25% of trees sampled were discarded as burn-in. Bayesian posterior probabilities (PP) were used to assess branch support. *M. coucha* and Lujo virus were used as outgroups for the host and

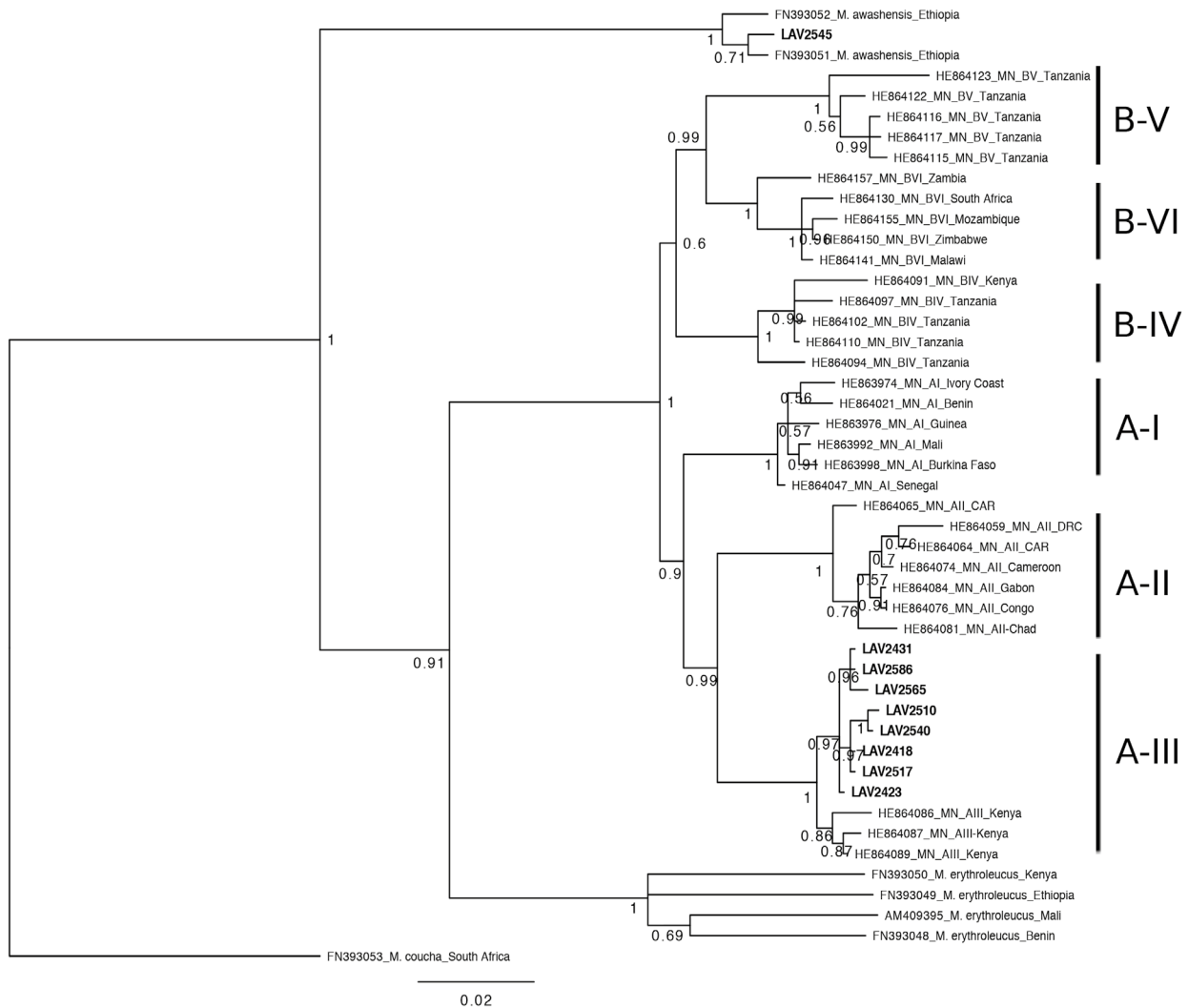


Fig. 2. Phylogenetic tree estimated from the Bayesian analysis of *cytb* sequences of *Mastomys* individuals positive for the mammarenavirus (in bold) and of representative sequences of the six *M. natalensis* mitochondrial lineages (sensu Colangelo et al. 2013) and other *Mastomys* species using the HKI + I + Γ model of evolution. Numbers below branches are the posterior probabilities. The small scale bar represents the number of nucleotide substitutions per site. GenBank AN and country of origin of *Mastomys* representative sequences are given (DRC: Democratic Republic of Congo; CAR: Central African Republic).

virus tree, respectively. Trees were visualized and annotated in FigTree, version 1.4.1. (<http://tree.bio.ed.ac.uk/software/figtree/>). Nucleotide and amino acid genetic p distances were estimated in MEGA7 (Kumar et al. 2016).

Results and Discussion

Virus screening

In total, 218 small mammals were trapped: 64 bats from 17 different species, 10 shrews from three different species and 144 rodents from 14 different species (see Table S1). We screened 54, 139 and 10 samples from bats, rodents and shrews, respectively, for mammarenaviruses, targeting a small part of the L gene. Six pools were positive, which, after depooling, corresponded to nine samples of four unique sequences. BLAST and a

preliminary phylogenetic analysis showed that the sequences corresponded to different strains of a unique mammarenavirus. For four out of the nine samples, we also successfully obtained part of the GPC gene corresponding to three unique sequences. The positive samples were all from the *Mastomys* genus. Small mammal *cytb* genotyping and BLAST analysis confirmed that eight samples were *M. natalensis* and one sample was *M. awashensis*. In the Bayesian phylogenetic analysis, all eight *M. natalensis* Mt *cytb* clustered with a posterior probability of one to the A-III Mt lineage (Fig. 2). The mammarenavirus L sequence found in *M. awashensis* was identical to the sequence found in two *M. natalensis* individuals.

The prevalence of the virus was 12.5% (95% CI: 0.6-50%) in *M. awashensis* and 36.4% (95% CI: 17.2-

59.4%) in *M. natalensis*. The limited sample size for the two species (N = 22 and N = 8 for *M. natalensis* and *M. awashensis*, respectively) did not allow us to distinguish if the virus is specific to *M. natalensis*, the single case in *M. awashensis* representing a spill-over, or if the virus circulates freely within the two *Mastomys* species: although the prevalence was higher in *M. natalensis* compared to *M. awashensis*, this was not statistically significant (Fisher exact test $p = 0.374$). *M. awashensis* has been shown to carry another mammarenavirus in northern Ethiopia (Meheretu et al. 2012). The nucleotide sequence identity of the short L fragment between the mammarenaviruses from *M. awashensis* from northern Ethiopia and the virus found in *M.*

awashensis in Dhati Welel was 78.9% supporting that the virus from this study is different than the one from northern Ethiopia. Eighteen *Stenocephalemys albipes* individuals were also screened during this study but none of them were found positive though mammarenaviruses have been previously reported in this rodent species in northern Ethiopia (Meheretu et al. 2012).

Characterisation of the full genome

Using a combination of high throughput and Sanger sequencing we successfully characterised the genome of the virus apart from 20 non-coding nucleotides at the 3' end of the S segment and 20 non-coding nucleotides at the 5' and 3' ends of the

Table 1. Amino-acid and nucleotide sequence identities of Dhati Welel virus with Old World representatives of the genus *Mammarenavirus*.

Virus species	Amino-acid sequence identity with Dhati Welel LAV2586				Nucleotide sequence identity with Dhati Welel LAV2586			
	NP	GPC	L	Z	NP	GPC	L	Z
Gairo	86.8	87.4	74.2	70.1	75.2	75.4	69.5	65.0
Mobala	84.2	85.5	71.9	66.0	74.7	74.8	68.2	65.6
Mayo Ranewo*	-	82.2	-	-	-	72.9	-	-
Luna	78.2	81.3	63.0	67.7	69.6	70.3	62.6	65.3
Mopeia (ZW)	77.0	80.9	62.7	64.6	69.5	70.8	62.8	63.0
Mopeia (MZ)	78.2	80.2	62.9	62.6	69.7	70.0	63.1	60.7
Gbagroube	72.8	79.8	-	-	67.0	67.4	-	-
Morogoro	78.6	79.2	63.1	59.6	71.5	71.1	61.9	61.3
Lassa Mad39	74.3	79.0	58.5	62.8	66.8	68.8	60.2	62.1
Lassa Josiah	74.3	79.0	59.0	63.8	67.0	68.8	59.9	62.5
Lassa Nig08-A41	72.7	78.6	58.9	64.9	66.1	70.3	60.0	60.0
Menekre	71.9	75.7	-	-	66.4	67.3	-	-
Mariental	71.9	74.0	-	-	65.2	67.6	-	-
Solwezi	69.6	73.4	54.5	57.7	64.0	68.5	57.9	57.2
Ippy	70.1	73.0	55.2	63.6	64.2	67.3	57.7	60.0
Kitale	68.7	71.5	55.1	57.9	64.0	66.9	57.8	60.5
Wenzhou	72.8	70.9	54.9	61.4	66.2	65.1	58.1	60.3
Loie River	71.5	70.7	54.7	63.0	65.1	66.8	57.7	53.2
Okahandja	66.9	67.4	52.6	56.0	63.5	65.1	56.2	60.9
Merino walk	67.0	67.0	53.3	51.7	63.7	64.1	56.8	55.6
LCMV Makokou	63.9	59.8	47.3	-	62.4	60.4	53.8	-
LCMV Arm.53b	63.9	59.6	47.6	55.6	61.1	60.4	53.8	53.5
Lunk	64.1	59.6	47.0	52.7	61.1	59.6	52.9	52.8
Souris	64.0	58.8	51.3	57.0	60.5	61.0	55.6	59.6
Ryuku	65.4	58.6	45.9	52.2	61.2	60.5	53.2	53.8
Lujo	58.6	44.4	43.9	45.1	60.3	51.4	51.2	48.2

Identities are shown only for the complete ORF at the exception of * where 60% of the GPC ORF is available. GPC: Glycoprotein; NP: Nucleoprotein; L: RNA-dependent-RNA polymerase; Z: Zinc finger protein. LCMV: Lymphocytic choriomeningitis virus.

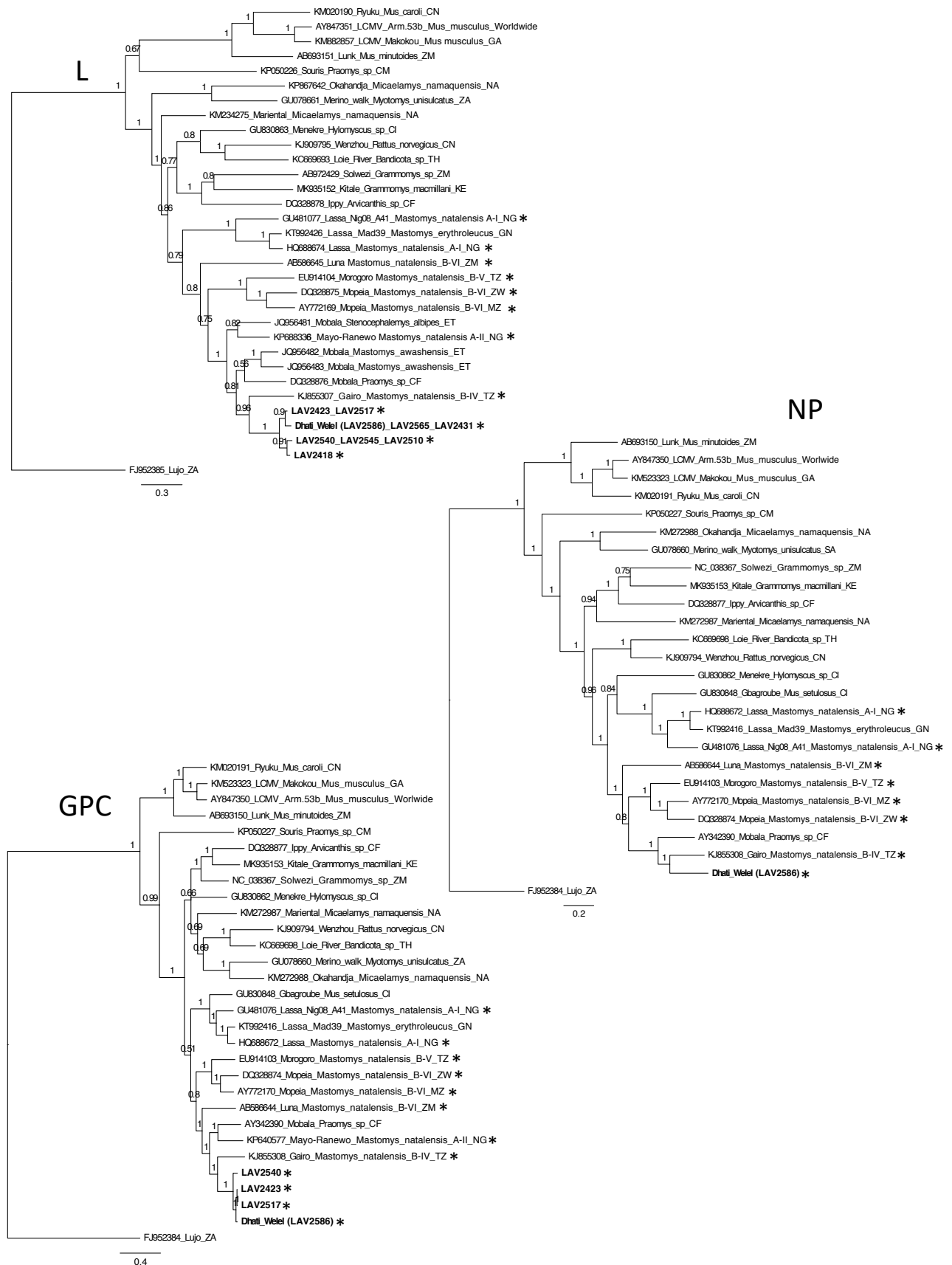


Fig. 3. Phylogenetic trees of Old World mammarenavirus representatives estimated from Bayesian analysis using the GTR + I + Γ model of evolution for the L, NP and GPC genes. Numbers above branches are the posterior probabilities. The small scale bar represents the number of nucleotide substitutions per site. Names of viruses are given in full with the GenBank AN, the natural reservoir hosts and the ISO code of the country (Cameroon CM, Central African Republic CF, China CN, Ethiopia ET, Gabon GA, Guinea GN, Kenya KE, Ivory Coast CI, Mozambique MZ, Namibia NA, Nigeria NG, South Africa ZA, Tanzania TZ, Thailand TH, Zambia ZM, Zimbabwe ZW). * indicates mammarenaviruses from *Mastomys natalensis*. Sequences from this study in bold. Note that for Lassa sequences, strain Josiah and Nig08-A41, *M. natalensis* A-I is indicated as the natural reservoir host though these sequences were isolated from human samples.



L segment (corresponding to the conserved end parts of mammarenaviruses and used as Sanger sequencing primers). High throughput read coverage was 40.5 ± 31 (SD) for the L segment and 25.2 ± 18.6 (SD) for the S segment. Each of the two segments of the virus showed mammarenavirus typical open reading frames (ORFs) separated by the typical stem-loop structures. The complete L segment was 7278 nucleotides long and contained two ORFs: the Z ORF of 300 nucleotides encodes a 99 amino-acid long zinc finger protein and the L ORF of 6660 nucleotides encodes a 2219 amino-acid long RNA-dependent RNA polymerase. In the L protein, the canonical polymerase domains (pre-A, A, B, C, D and E) and the key active site residues of the endonuclease domain NL1 (Morin et al. 2010) were well conserved. The motifs of the late domains of the Z protein, PSAP and PPPY, were identical to most of the African mammarenavirus genomes. The complete S segment was 3367 nucleotides long and contained two ORFs: the GPC ORF of 1470 nucleotides encodes a 489 amino-acid long glycoprotein precursor and the NP ORF of 1713 nucleotides encodes a 570 amino-acid long nucleoprotein. The motif at the GP1/GP2 cleavage site of the GPC protein was RLL as in most of the African mammarenavirus genomes. The DEDDh motif of the NP protein 3'-5' exonuclease domain found in other mammarenaviruses was also present. Four potential N-glycosylation sites on GP2 and seven potential N-glycosylation sites on GP1 could be detected in positions analogous to other African mammarenaviruses (Bonhomme et al. 2011). The sequences were deposited in GenBank (AN: MT078838-MT078839).

Molecular divergence and taxonomic considerations

The comparison of nucleotide and amino acid sequence identities of the new virus with other OW mammarenavirus representatives showed Gairo virus, found in *M. natalensis* B-IV in Tanzania is the most similar (Table 1). Nucleotide sequence identities between the two viruses were 75% (NP), 75% (GPC), 70% (L) and 65% (Z). Identities at amino acid sequences were 87% (NP), 87% (GPC), 74% (L) and 70% (Z). The International Committee on Taxonomy of Viruses recommends that a new species should share less than 80% and 76% nucleotide sequence identity in the S and L segments respectively, and less than 88% NP amino acid sequence identity with previously recognised mammarenavirus species. The new virus fits these criteria and we propose the name Dhati Welel

virus (DHWV), the National Park where the virus was found.

Evolutionary history of Dhati Welel and other *Mastomys natalensis* mammarenaviruses

We focus on the position of Dhati Welel virus in the OW mammarenavirus phylogeny and the phylogenetic relationships among *M. natalensis*-borne viruses. For all three genes, the new virus clustered with Gairo virus described from *M. natalensis* B-IV in Tanzania (PP = 1 for GPC and NP but PP = 0.96 for L) (Fig. 3). Mayo Ranewo virus found in *M. natalensis* A-II in Nigeria is basal to the clade (Gairo + Dhati Welel) with high support (PP = 1 in GPC and L; no sequence data from this virus for the NP gene). This is the sister group to a clade grouping Mopeia from *M. natalensis* B-VI and Morogoro from *M. natalensis* B-V in the NP and L trees. Luna is basal to this clade, with high support in the NP (PP = 1) but not the L tree (PP = 0.8). Finally, in all gene trees, Lassa virus found in *M. natalensis* A-I in Western Africa forms the sister group of the clade containing the viruses of all other *M. natalensis* Mt lineages (A-II, A-III, B-IV, B-V, B-IV) but with variable support depending on the gene (PP = 0.51 in GPC, 0.79 in L and 1 in NP). The branching of (Mopeia + Morogoro) and Luna is switched in the GPC tree with Luna basal to (Mayo Ranewo + (Gairo + Dhati Welel)) showing high support (PP = 1). The clade containing all *M. natalensis* viruses also contains mammarenaviruses found in other rodent species which do not always show similar branching depending on the gene: e.g. Mobala virus found in *Praomys* sp. in Central African Republic clusters with high support (PP = 1) with Mayo Ranewo virus from *M. natalensis* A-II in the GPC tree but forms a trifurcation with Mobala-like viruses found in *M. avashensis* in north Ethiopia and (Gairo + Dhati Welel) clade in the L tree. Finally, the position of the Mobala-like virus in *S. albipes* from Ethiopia clusters with Mayo Ranewo virus from *M. natalensis* A-II but with weak support (PP = 0.82 in the L tree). Although the L tree has the largest number of sequences, several of them are of limited size (340 nucleotides) and this explains in part the lack of resolution in this part of the tree. The lack of matching between the Mt phylogeny of *M. natalensis* lineages and its mammarenaviruses, the switch in the branching of some of the viruses within the clade containing all viruses from *M. natalensis* for some of the genes and the presence of related viruses from other rodent species such as *Praomys* sp., *S. albipes* or *M. avashensis* illustrate the dynamic evolutionary

history of this clade of mammarenaviruses involving likely several host jumps and maybe some recombinations or re-assortments among some of the viruses. Characterisation of the full genomes of the mammarenaviruses from *M. awashensis* and *S. albipes* from Ethiopia and of a few more viruses found in eastern Africa is currently underway (Gryseels & Goüy de Bellocq, unpublished) and will facilitate our understanding of the complex history of Eastern Africa mammarenaviruses.

Conclusion

We detected and characterised the almost complete genome (99.4%) of a novel mammarenavirus, Dhati Welel, which was found in several individuals of the *M. natalensis* A-III Mt lineage, hitherto not known to carry any mammarenavirus. This finding highlights the importance of the rodent species *M. natalensis* in the distribution and evolutionary history of mammarenaviruses throughout Africa. The virus was also found in one *M. awashensis* individual. Because we found four different strains of Dhati Welel in *M. natalensis* and because the strain found in *M. awashensis* individual was identical to one strain found in two *M. natalensis* individuals, it seems more likely that the reservoir of the virus is *M. natalensis* and the presence of the virus in *M. awashensis* species represents a spillover. We however cannot rule out that the virus circulates between both rodent species. Additional sampling in the areas where both rodent species co-occur in Ethiopia (Martynov et al. 2020) would help define the specificity of Dhati Welel virus for *M. natalensis* A-III.

Another African rodent species of widespread distribution similar to that of *M. natalensis* in Africa, is *Mus minutoides* (Bryja et al. 2014, Denys

et al. 2017) and is already known to harbour three different mammarenaviruses (Lecompte et al. 2007, Goüy de Bellocq et al. 2010, Ishii et al. 2012) in three out of the 11 described haplogroups (Bryja et al. 2014). These rodent species, *M. natalensis* and *M. minutoides* are clearly important to our understanding of the diversity and evolutionary history of this group of viruses on the continent.

Acknowledgements

Permission for sampling was provided by the Oromia Forest and Wildlife Enterprise (permission no. OFWE 20/01/2014). This work was supported by the Czech Science Foundation (GAČR grant no. 18-19629S) and the Russian Foundation for Basic Research (project no. 18-04-00563-a). We are grateful to Dr. A. Darkov (Joint Ethio-Russian Biological Expedition, Fourth Phase – JERBE IV) and Dr. S. Keskes (Ethiopian Ministry of Innovation and Technology) for management of the expedition in the field and in Addis Ababa. For help during the field work we acknowledge S.V. Kruskop, D. Yu. Alexandrov, K.A. Rogovin, M. Kasso, A.A. Warshavsky and M. Jemal. We thank E. Holánová for her help on mammarenavirus screening and D. Čížková for advice and assistance with high throughput sequencing and analysis. Computational resources were supplied by the project “e-Infrastruktura CZ” (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures. Author contributions: J. Goüy de Bellocq and L.A. Lavrenchenko designed the study; A.A. Martynov and L.A. Lavrenchenko collected the samples; J. Goüy de Bellocq and A. Bryjová performed the molecular work for viruses; A.A. Martynov and L.A. Lavrenchenko genotyped the small mammals; J. Goüy de Bellocq analysed the results and wrote the manuscript. All authors provided editorial advice and approved the final manuscript.



Literature

- Bankevich A., Nurk S., Antipov D. et al. 2012: SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19: 455–477.
- Bonhomme C.J., Capul A.A., Lauron E.J. et al. 2011: Glycosylation modulates arenavirus glycoprotein expression and function. *Virology* 409: 223–233.
- Borissenko A.V. 1999: A mobile trap for bats. *Plecotus et al.* 2: 10–19. (in Russian with English summary)
- Briese T., Paweska J.T., McMullan L.K. et al. 2009: Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from Southern Africa. *PLOS Pathog.* 5: e1000455.
- Bryja J., Mikula O., Šumbera R. et al. 2014: Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evol. Biol.* 14: 256.
- Colangelo P., Verheyen E., Leirs H. et al. 2013: A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biol. J. Linn. Soc.* 108: 901–916.
- Cuypers L.N., Baird S.J.E., Hánová A. et al. 2020: Three arenaviruses in three subspecific Natal multimammate mouse taxa in Tanzania: same host specificity, but different spatial genetic structure? *Virus Evol.* doi: 10.1093/ve/veaa039.
- Denys C., Taylor P.J. & Aplin K.P. 2017: Family Muridae (true mice and rats, gerbils and relatives). In: Wilson D.E., Lacher T.E. & Mittermeier R.A. (eds.), *Handbook of the mammals of the world, vol. 7. Rodents II. Lynx Edicions, Barcelona*: 536–884.
- Ehichioya D.U., Hass M., Becker-Ziaja B. et al. 2011: Current molecular epidemiology of Lassa virus in Nigeria. *J. Clin. Microbiol.* 49: 1157–1161.
- Fichet-Calvet E. & Rogers D.J. 2009: Risk maps of Lassa fever in West Africa. *PLOS Negl. Trop. Dis.* 3: e388.
- Frame J.D., Baldwin J.M., Jr., Gocke D.J. & Troup J.M. 1970: Lassa fever, a new virus disease of man from West Africa. *Am. J. Trop. Med. Hyg.* 19: 670–676.
- Goüy de Bellocq J., Borremans B., Katakweba A. et al. 2010: Sympatric occurrence of 3 arenaviruses, Tanzania. *Emerg. Infect. Dis.* 16: 692–695.
- Goüy de Bellocq J., Těšíková J., Meheretu Y. et al. 2016: Complete genome characterisation and phylogenetic position of Tigray hantavirus from the Ethiopian white-footed mouse, *Stenocephalemys albipes*. *Infect. Genet. Evol.* 45: 242–245.
- Gryseels S., Baird S.J.E., Borremans B. et al. 2017: When viruses do not go viral: the importance of host phylogeographic structure in the spatial spread of arenaviruses. *PLOS Pathog.* 13: e1006073.
- Gryseels S., Rieger T., Oestereich L. et al. 2015: Gairo virus, a novel arenavirus of the widespread *Mastomys natalensis*: genetically divergent, but ecologically similar to Lassa and Morogoro viruses. *Virology* 476: 249–256.
- Günther S., Hoofd G., Charrel R. et al. 2009: Mopeia virus-related arenavirus in Natal multimammate mice, Morogoro, Tanzania. *Emerg. Infect. Dis.* 15: 2008–2012.
- Ishii A., Thomas Y., Moonga L. et al. 2011: Novel arenavirus, Zambia. *Emerg. Infect. Dis.* 17: 1921–1924.
- Ishii A., Thomas Y., Moonga L. et al. 2012: Molecular surveillance and phylogenetic analysis of Old World arenaviruses in Zambia. *J. Gen. Virol.* 93: 2247–2251.
- Johnson K.M., Taylor P., Elliott L.H. & Tomori O. 1981: Recovery of a Lassa-related arenavirus in Zimbabwe. *Am. J. Trop. Med. Hyg.* 30: 1291–1293.
- Katoh K., Misawa K., Kuma K.I. & Miyata T. 2002: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30: 3059–3066.
- Kumar S., Stecher G. & Tamura K. 2016: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870–1874.
- Laetsch D. & Blaxter M. 2017: BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research* 6: 1287.
- Lecompte E., Fichet-Calvet E., Daffis S. et al. 2006: *Mastomys natalensis* and Lassa fever, West Africa. *Emerg. Infect. Dis.* 12: 1971–1974.
- Lecompte E., Granjon L., Peterhans J.K. & Denys C. 2002: Cytochrome *b*-based phylogeny of the *Praomys* group (Rodentia, Murinae): a new African radiation? *C. R. Biol.* 325: 827–840.
- Lecompte E., ter Meulen J., Emonet S. et al. 2007: Genetic identification of Kodoko virus, a novel arenavirus of the African pigmy mouse (*Mus Nannomys minutoides*) in West Africa. *Virology* 364: 178–183.



- Leirs H. 1994: Population ecology of *Mastomys natalensis* (Smith 1834): implications for rodent control in Africa. *Belgian Administration for Development Cooperation, Brussels*.
- Martynov A.A., Bryja J., Meheretu Y. & Lavrenchenko L.A. 2020: Multimammate mice of the genus *Mastomys* (Rodentia: Muridae) in Ethiopia – diversity and distribution assessed by genetic approaches and environmental niche modelling. *J. Vertebr. Biol.* 69: 20006. <https://doi.org/10.25225/jvb.20006>.
- Meheretu Y., Čížková D., Těšíková J. et al. 2012: High diversity of RNA viruses in rodents, Ethiopia. *Emerg. Infect. Dis.* 18: 2047–2050.
- Meheretu Y., Stanley W.T., Craig E.W. et al. 2019: Tigray orthohantavirus infects two related rodent species adapted to different elevations in Ethiopia. *Vector Borne Zoonotic Dis.* 19: 950–953.
- Mizerovská D., Mikula O., Meheretu Y. et al. 2020: Integrative taxonomic revision of the Ethiopian endemic rodent genus *Stenocephalemys* (Muridae: Murinae: Praomyini) with the description of two new species. *J. Vertebr. Biol.* 69: 20031. <https://doi.org/10.25225/jvb.20031>.
- Monath T.P., Newhouse V.F., Kemp G.E. et al. 1974: Lassa virus isolation from *Mastomys natalensis* rodents during an epidemic in Sierra Leone. *Science* 185: 263–265.
- Morin B., Coutard B., Lelke M. et al. 2010: The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLOS Pathog.* 6: e1001038.
- Olayemi A., Obadare A., Oyeyiola A. et al. 2016: Arenavirus diversity and phylogeography of *Mastomys natalensis* rodents, Nigeria. *Emerg. Infect. Dis.* 22: 694–697.
- Posada D. 2008: jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25: 1253–1256.
- Radoshitzky S.R., Buchmeier M.J., Charrel R.N. et al. 2019: ICTV virus taxonomy profile: arenaviridae. *J. Gen. Virol.* 100: 1200–1201.
- Reiczigel J. 2003: Confidence intervals for the binomial parameter: some new considerations. *Stat. Med.* 22: 611–621.
- Richmond J.K. & Baglole D.J. 2003: Lassa fever: epidemiology, clinical features, and social consequences. *BMJ* 327: 1271–1275.
- Ronquist F., Teslenko M., van der Mark P. et al. 2012: MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61: 539–542.
- Rózsa L., Reiczigel J. & Majoros G. 2000: Quantifying parasites in samples of hosts. *J. Parasitol.* 86: 228–232.
- Těšíková J., Bryjová A., Bryja J. et al. 2017: Hantavirus strains in East Africa related to Western African hantaviruses. *Vector Borne Zoonotic Dis.* 17: 278–280.
- Vieth S., Drosten C., Lenz O. et al. 2007: RT-PCR assay for detection of Lassa virus and related Old World arenaviruses targeting the L gene. *Trans. R. Soc. Trop. Med. Hyg.* 101: 1253–1264.
- Wulff H., McIntosh B.M., Hamner D.B. & Johnson K.M. 1977: Isolation of an arenavirus closely related to Lassa virus from *Mastomys natalensis* in south-east Africa. *Bull. World Health Organ.* 55: 441.

Supplementary online material

Table S1. List of the samples captured in Dharti Welel National Park in 2014. The samples screened for mammarenavirus are indicated with an X. The date of capture, sex, GPS and the GenBank AN of the *cytb* sequences of some samples are also mentioned.

Table S2. Primers used to complete the missing parts of the virus genome.

Table S3. GenBank accession number of the mammarenavirus sequences used for the analyses. The natural reservoir host and the country are indicated.

(<https://www.ivb.cz/wp-content/uploads/JVB-vol.-69-2-2020-Gouy-de-Bellocq-et-al.-Tables-S1-S3.xlsx>)