Molecular surveillance and phylogenetic analysis of Old World arenaviruses in Zambia

Akihiro Ishii,¹ ² Yuka Thomas,¹ ² Ladslav Moonga,² Ichiro Nakamura,¹ ² Aiko Ohnuma,¹ Bernard M. Hang’ombe,² Ayato Takada,¹ ²Aaron S. Mweene² and Hirofumi Sawa¹ ²

¹Research Center for Zoonosis Control, Hokkaido University, Sapporo, Hokkaido, Japan
²School of Veterinary Medicine, University of Zambia, Lusaka, Lusaka province, Zambia

In order to survey arenaviruses in the Republic of Zambia, we captured 335 rodents from three cities between 2010 and 2011. Eighteen Luna virus (LUNV) and one lymphocytic choriomeningitis virus (LCMV)-related virus RNAs were detected by one-step RT-PCR from Mastomys natalensis and Mus minutoides, respectively. Four LUNV strains and one LCMV-related virus were isolated, and the whole genome nucleotide sequence was determined by pyrosequencing. Phylogenetic analyses revealed that the LUNV clade consists of two branches that are distinguished by geographical location and that the LCMV-related virus belongs to the LCMV clade, but diverges from the typical LCMVs. Comparison of nucleoprotein amino acid sequences indicated that the LCMV-related virus could be designated a novel arenavirus, which was tentatively named as the Lunk virus. Amino acid sequences of the GP, NP, Z and L proteins showed poor similarity among the three Zambian arenavirus strains, i.e. Luna, Lunk and Lujo virus.

The family Arenaviridae is divided into two groups, the New World arenaviruses (NWAs) and the Old World arenaviruses (OWAs), based on their antigenicity, host animal species, geographical distribution and genetic diversity (Bowen et al., 1997; Charrel et al., 2008). Viruses that are categorized as OWAs include the lymphocytic choriomeningitis virus (LCMV), which is primarily found in the northern hemisphere, and the Lassa virus-related viruses (LSRVs), which are found in sub-Saharan Africa (Charrel et al., 2008; Emonet et al., 2006). Serological analysis has provided some evidence of the presence of LCMV in Egypt (el Karamany & Imam, 1991). In addition, the Kudoko virus was detected in specimens of Mus minutoides from Guinea and Tanzania (de Belloq et al., 2010; Lecompte et al., 2007). Prior to 2008, only the Lassa virus, which is found exclusively in West Africa, was known to cause haemorrhagic fever (Briese et al., 2009; Frame et al., 1970).

In September 2008, a patient in Zambia developed haemorrhagic fever (LUJV) (Briese et al., 2009). Four out of the five patients who contracted the virus died of the infection. In this incident, the haemorrhagic fever was caused by a novel arenavirus, named the Lujo virus (LUJV) (Briese et al., 2009; Paweska et al., 2009). The geographical location in which LUJV was found classifies it as an OWA; however, the viral genes are slightly divergent from other OWAs (Briese et al., 2009). Other OWAs that are not pathogenic to humans have been discovered in West Africa (Gbagroube and Menekre viruses), Central Africa (Ippy and Mobala viruses), East Africa (Morogoro and Mopeia viruses) and South Africa (Merino Walk virus) from animals of the subfamily Murinae (Bowen et al., 2000; Coulibaly-N’Golo et al., 2011; Günther et al., 2009; Palacios et al., 2010; Wulff et al., 1977). Of note, the host of LUJV has not been identified to date and in order to identify the host of OWAs in Zambia, we conducted a molecular surveillance of arenaviruses. We identified a novel virus in Mastomys natalensis, the Luna virus (LUNV), which is an LSRV (Ishii et al., 2011). To date, neither LCMV nor LUJV have been detected in Zambia; therefore, to better characterize the diversity of arenavirus species in Zambia we captured 335 rodents from Lusaka (east area: 15°26′6.12″S, 28°26′9.93″E; north area: 15°48′6.12″S, 28°14′8.33″E; south area: 15°34′6.88″S, 28°16′5.13″E), 22 rodents from Kafue (15°49′27.8″S, 28°14′7.83″E) and 78 rodents from Livingston (17°50′8.72″S, 25°43′59.19″E). The rodents were captured in August 2010, November 2010 and July 2011 by using Sherman and cage traps that were placed around maize fields and in storage houses. The surveillance of arenavirus in rodents in Zambia was carried out with permission from the Zambia Wildlife Authority.

The captured animals were euthanized with diethyl ether and their kidney tissues were harvested. These samples were transported in a portable freezer from the field and
kept in a −80 °C freezer until analysis. The kidney tissue was homogenized in TRIzol reagent (Invitrogen). After the addition of chloroform, the total RNA sample was extracted from the supernatant of the homogenate by using PureLink RNA Mini kit (Invitrogen). The results of identification of the rodent species and arenavirus screening are summarized in Table 1. The arenaviruses were screened by using one-step RT-PCR with PrimeScript One-Step RT-PCR kit Ver.2 (Takara Bio Inc.) for 30 min at 50 °C and 2 min at 94 °C. Then, the samples were exposed to 45 cycles each of 30 s at 94 °C, 30 s at 50 °C, 30 s at 72 °C and 10 min at 72 °C. Universal primers were used to detect the arenavirus L gene as described previously (Ishii et al., 2011). The species of the captured rodents were identified by analyzing the nucleotide sequence of the cytochrome b gene (cytb). The cytb fragment was amplified by one-step RT-PCR using universal primers for multiple rodents that were designed for this study (5′-TCATCAGTAACCACAT-AYTYGCGAG-3′ and 5′-TTGGCGGAATATTAGGCC-TCTGTTG-3′). PCRs for cytb amplification were conducted under the following conditions: 30 min at 50 °C, 2 min at 94 °C and 45 cycles each of 30 s at 94 °C, 30 s at 50 °C, 1 min at 72 °C and 10 min at 72 °C. The results revealed that 18 of the 218 M. natalensis rodents and one of the three M. minutoides rodents were positive for an arenavirus. The cytb nucleotide sequences for the positive samples were deposited in GenBank (sample ID: accession number; 10LivR34, 10LivR62, 11LivR09, 11LivR21, 11LR24, 11LR25 and 11LR 26: AB731571; 10LR29 and 10LR33: AB731572; 10LR30: AB731573; 10LivR33: AB731574; 10LR34: AB731575; 10LR34: AB731576; 11LR10: AB731577; 11LR11: AB731578; 11LR36: AB731579; 11LivR31 and 11LivR33: AB731580; 10LR34: AB731581). BLAST searches revealed that the sequences of the amplified viral fragments extracted from the M. natalensis had high similarity to nucleotide sequences of LUNV (data not shown). The single isolated viral fragment from M. minutoides (10LR34) was similar to LCMV sequences; however, they were not identical because the highest E-value to an LCMV strain calculated by BLAST analysis was 1.72E-23.

In order to isolate the detected arenaviruses, the kidney tissues were homogenized in Dulbecco’s minimal essential medium (DMEM) supplemented with 2 % FBS and inoculated into Vero E6 cells that were cultured in DMEM. After one passage of the culture supernatant to new Vero E6 cells, the supernatants of the cell cultures were harvested and stored as P1 samples. Then, 1 ml of P1 sample was used to inoculate Vero E6 cells for 2 h. These cells were cultured for 6 days, and 100 μL of the culture supernatant was harvested daily to detect the presence of viral RNA. Preparation of the RNA sample and detection of viral RNA have been described above. The propagation of the viruses was evident in four samples of M. natalensis from Lusaka in 2010 (10LR33) and 2011 (11LR10), from Livingstone in 2010 (10LivR33 and 10LivR74), and one sample of M. minutoides from Lusaka in 2010 (10LR34). This was shown by a gradual increase of RT-PCR products for viral RNA in culture supernatants (data not shown). Pyrosequencing was used to determine the genome sequences of the isolated viruses. A double-stranded cDNA sample was prepared from 200 ng of the P1 RNA sample by using a Takara CDNA synthesis kit with 9-mer random nucleotides (Takara Bio Inc.). The cDNA sample was processed for pyrosequencing following the manufacturer’s instructions.

### Table 1. Species of the captured rodents and results of arenavirus screening

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acomys subspinosus</td>
<td>0/2*</td>
<td>0/1</td>
<td>0/10</td>
<td>0/6</td>
<td></td>
<td>0/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aethomys chrysophilus</td>
<td>0/5</td>
<td>0/5</td>
<td>0/10</td>
<td>0/6</td>
<td></td>
<td>0/26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arvicanthis sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crocidura hirta</td>
<td>0/1</td>
<td></td>
<td>0/1/1</td>
<td></td>
<td></td>
<td>0/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crocidura sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerbilliscus leucogaster</td>
<td>0/5</td>
<td>0/12</td>
<td>0/5</td>
<td>0/2</td>
<td></td>
<td>0/24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammomys sp.</td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
<td>0/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphiurus sp.</td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylomyscus sp.</td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastomys natalensis</td>
<td>4/24</td>
<td>7/20</td>
<td>0/20</td>
<td>1/85</td>
<td>0/5</td>
<td>6/42</td>
<td>18/218</td>
<td></td>
</tr>
<tr>
<td>Mus minutoides</td>
<td>1/2</td>
<td>0/1</td>
<td>0/1</td>
<td>0/7</td>
<td>0/1</td>
<td>0/35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otomys sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Praomys sp.</td>
<td>0/1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rattus rattus</td>
<td>0/1</td>
<td>0/2</td>
<td>0/7</td>
<td>0/1</td>
<td></td>
<td>0/35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccostomus sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0/1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Number of arenavirus-positive individuals/number of rodents captured. The positive number is in boldface.
and applied to the GS Junior sequencer (Roche). The obtained fragments were analysed by repetitive de novo assembly and reference mapping to the most similar viral genome by using CLC Genomics Workbench 4.9 software (CLC bio). We obtained 16,235, 6922, 496 and 16,654 reads for the 10LR33, 11LR10, 10LivR33 and 10LR34 viruses, respectively. The full-length sequence of the 10LR33 and the 10LR34 genomes was determined. The genomes of the remaining samples were determined except for the UTR in the termini of their genomes. The 10LR33, 11LR10 and 10LivR33 viruses were identified as LUNV variants genetically and were designated to LUNV LSK-2 (AB693148 and AB693149), LSK-3 (AB702940 and AB702941) and LVN-1 (AB697691 and AB697692) strains, respectively. The single virus from 10LR34 was tentatively named as the Lunk virus (LNKV) NKS-1 strain; it was named after the sampling location where it was discovered (Lusaka-New Kasama area) in order to distinguish it from LCMV. The virus had a typical bi-segmented genome structure as the arenavirus S (AB693150) and L segments (AB693151), with lengths of 3419 and 7196 bases, respectively. The intergenic regions of the LNKV genomes have the distinctive stem–loop structure that is present in other arenaviruses. The RNA secondary structure was predicted using the MFOLD program (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form). The calculated free energies of the stem–loops in the S and L segments were −53.2 and −55.7 kcal mol\(^{-1}\), respectively.

Phylogenetic analysis of the Zambian arenaviruses was performed based on the L gene fragments that were obtained during this study and the corresponding fragments of other OWAs. The nucleotide sequences were aligned using CLUSTAL W, and the evolutionary history was inferred by using the maximum-likelihood method based on the Tamura–Nei model on the MEGA5 software (Tamura & Nei, 1993; Tamura et al., 2011). The analysis showed that all viral RNA fragments from \(M. natalensis\) in this study could be grouped into one clade of LUNVs (Fig. 1a). Three sub-branches contained LUNV strains from the rodents captured around Lusaka (LR), including the previously reported Namwala strain (NMW-1). LUNV strains from Livingstone rodents (LivR) slightly diverged from the Lusaka-Namwala strains. The LNKV strain from 10LR34 was related to LCMVs, Kodoko and the Tanzanian virus fragments, and it was clearly distinct from LSRVs, including LUNVs. Partial NP gene nucleotide sequences were used to analyse the diversity of LCMVs, including Kodoko virus fragments and the LNKV (Fig. 1b). From the analysis, four lineages could be identified. Lineage 1 consists of typical LCMVs such as the Armstrong strain. Lineage 2 is slightly diverged from lineage 1; however, viruses from these lineages have a similar geographical distribution and host animal species. Virus strains from wild mice in Spain with unknown pathogenicity comprise lineage 3; whereas, lineage 4 is an African lineage that includes the Kodoko virus, the Tanzanian virus and the LNKV strains.

In order to classify the OWAs, we used a divergence value of at least 12 % for the complete NP amino acid sequence, geographical distribution, host animal species and genetic character, as described in previous studies (Bowen et al., 2000; Charrel et al., 2008; Emonet et al., 2006). In order to identify newly isolated viruses, complete NP amino acid divergence values were calculated on CLC Genomics Workbench software version 5 (CLC bio). Divergence was less than 7.9 % among the LUNVs (LSK-1–3, NMW-1 and LVN-1 strains) and less than 5.6 % among the LSK strains. The newly isolated LUNV LSK-2, LSK-3 from \(M. natalensis\) had diversities less than 5.6 % among the LSK strains and the LVN-1 strain ranged from 6.32 to 7.88 % among the LUNV strains. The new isolates were verified to be LUNV strains based on these values of divergence.

Divergence among the LCMV strains ranged from 5.91 to 9.32 % (Armstrong strain in lineage 1, Dandenong strain in lineage 2 and CABN strain in lineage 3); however, the divergence between LNKV and the LCMVs ranged from 17.7 to 18.5 %. These observations indicate that this LNKV strain cannot be related to LUNV, LUJV or LCMVs regardless of geographical and phylogenetic relationships. Worldwide, excluding Africa, a number of LCMV strains have been detected and isolated from house mice and primates including humans, and phylogenetic analysis indicates that these strains had three or four lineages (Albariño et al., 2010). Our analysis, which was based on the NP amino acid sequences, also suggested that there are four lineages of LCMVs. Under the above-mentioned criterion, lineages 1, 2 and 3 consist of the same species of arenaviruses because, previously, viruses from lineages 1 and 2 have been isolated from similar hosts (mice and primates, including humans) since their discovery in 1933. Thus, owing to the aforementioned factors, viruses comprising lineages 1 and 2 are considered to be the same species. Only lineage 3 strains, which have been isolated from wild mice (\(Apodemus sylvaticus\)), can be distinguished from the typical LCMVs based on the host animal species. The African strains, Kodoko and the newly isolated Zambian strain, in lineage 4 were distinguished from LCMVs by the host animal (\(M. minutoides\)) and geographical distribution (sub-Saharan) in addition to the divergence of NP. Thus, we suggest that the Zambian strain is a novel LCMV-related arenavirus (LCRV), which we designated LNKV. Interestingly, the L gene indicates that LNKV and LCMV had a common ancestor; however, the NP gene indicates that LNKV and the Kodoko virus had a common ancestor (Fig. 1a and b). This discrepancy suggested that a recombination of the genes or reassortment of a genome segment among typical LCMV lineage and African LCRVs had occurred. In order to elucidate the evolution of the LCRV lineage, we need more genome sequence information for the African strains.

The OWAs consist of two major clades: the LSRVs and the LCMVs. The novel LNKV belongs to the LCMV clade, and together they form the LCRV clade. Another Zambian strain, LUNV, belongs to the LSRV clade. However, although LUJV is geographically related to these strains (as it is present in Zambia), it has less identity and does not
Fig. 1. Phylogenetic analysis of OWAs. The percentage (<90) of trees in which the associated taxa clustered together is shown next to the branches. Initial tree for the heuristic search were obtained automatically as follows: when the number of common sites was <100 or less than one-fourth of the total number of sites, the maximum-parsimony method was used; otherwise the BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Codon positions included were 1st + 2nd + 3rd + non-coding. All positions containing gaps and missing data were eliminated. (a) The phylogenetic tree of the nucleotide sequences of the partial L gene fragment with the highest log-likelihood (−5532.8371) is shown. The analysis involved 36 nt sequences. There were a total of 335 positions in the final dataset. Tacaribe virus was considered to be an outgroup. (b) The phylogenetic tree of the nucleotide sequences of the partial NP gene fragment with the highest log-likelihood (−4006.9075) is shown. The analysis involved 21 nt sequences. There were a total of 370 positions in the final dataset. Lassa virus was considered to be an outgroup. Strains that determined the nucleotide sequences of their genomes in this study are in boldface. Viruses and the accession numbers used in this analysis are listed in Table S1 (available in JGV Online). GBAV, Gbagroube virus; KDKV, Kodoko virus; LASV, Lassa virus; LUJV, Lujo virus; LUNV, Luna virus; MENV, Menekre virus; MVV, Merino Walk virus; MOBV, Mobala virus; MOPV, Mopeia virus; MORV, Morogoro virus; TCRV, Tacaribe virus.
belong to either of the clades. Divergences of amino acid sequences of GP, NP, Z and L among Zambian OWAs (LUNV LSK-1, LNKV and LUJV) ranged from 42.9 to 60.6, 36.8 to 41.6, 47.5 to 57.6 and 52.5 to 57.3 %, respectively. Comparisons of amino acid sequence identities indicated that LUJV was less closely related to LUNV and LNKV than LUNV was to LNKV. A previous study reported the differences in the NP amino acid sequences in the OWA and NWA strains (Palacios et al., 2010). The calculated mean differences among OWAs were demonstrated to be less than 32 % in the same clade, 36.5 ± 1.4 % between different clades. Among NWA clades, the mean difference was 40.6 ± 4.8 %. Between the OWAs and NWAs, the mean difference was 49.6 ± 1.7 %. Following these values, LUJV was designated to be an OWA, although not in the same clade as LSRVs and LCMVs because the mean LUJV NP sequence difference was 42.6 and 52.7 % as compared with the OWAs and NWAs, respectively. In our study, comparison of amino acid sequence among Zambian strains also indicated that LUJV was 41.3 and 41.6 % different from LUNV and LNKV (an extended LCMV clade), respectively. In addition, mean percentage differences in GP, Z and L were greater than 57 %. In particular, the GP amino acid sequence of LUJV was highly divergent. The mean GP sequence difference between LUJV and the two other strains was 60.2 and 42.9 % between LUNV and LNKV, respectively. This high divergence suggests that LUJV evolved in a different location in Zambia, in a different host than LUNV (in *M. natalensis*) and LNKV (in *M. minutoides*), and/or experienced potent selective pressure. In order to identify the ancestral strain of LUJV, it might be necessary to survey geographical locations outside of Zambia and/or survey host animals other than those belonging to the subfamily Murinae.

**Acknowledgements**

The authors would like to thank the Zambia Wildlife Authority and Kwatala Safaris for assistance with sample collection in Zambia. This study was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology; the Ministry of Health, Labour and Welfare, Japan; the Global COE program; and the Japan Initiative for Global Research Network on Infectious Diseases, MEXT Japan.

**References**


