Discovery of Mwinilunga alphavirus: a novel alphavirus in *Culex* mosquitoes in Zambia

Shiho Torii[1], Yasuko Orba[1*], Bernard M. Hang'ombe[2,10], Aaron S. Mweene[3,9,10], Yuji Wada[1], Paulina D. Anindita[1], Wallaya Phongphaew[1], Yongjin Qiu[4], Masahiro Kajihara[5], Akina Mori-Kajihara[5], Yoshiki Eto[5], Hayato Harima[4], Michihito Sasaki[1], Michael Carr[6,7], William W. Hall[7,8,9,10], Yuki Eshita[4], Takashi Abe[11], Hirofumi Sawa[1,7,9,10*]

[1]Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan

[2]Department of Para-clinical Studies, School of Veterinary Medicine, University of Zambia, Lusaka, Zambia

[3]Department of Disease Control, School of Veterinary Medicine, University of Zambia, Lusaka, Zambia

[4]Hokudai Center for Zoonosis Control in Zambia, Lusaka, Zambia

[5]Division of Global Epidemiology, Research Center for Zoonosis Control, Hokkaido University, Sapporo, Japan

[6]National Virus Reference Laboratory, School of Medicine, University College Dublin, Dublin, Ireland

[7]Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, Sapporo, Japan

[8]Centre for Research in Infectious Diseases, University College Dublin, Dublin, Ireland

[9]Global Virus Network, Baltimore, Maryland, USA

[10]Africa Center of Excellence for Infectious Diseases of Humans and Animals, University of Zambia, Lusaka, Zambia

[11]Department of Computer Science, Graduate School of Science and Technology, Niigata

University, Niigata, Japan

*Corresponding author:

Yasuko Orba

Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan

Tel: +81-11-706-9508, Fax: +81-11-706-7370,

E-mail: orbay@czc.hokudai.ac.jp

Hirofumi Sawa, M.D., Ph.D.

Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan

Tel: +81-11-706-5185; Fax: +81-11-706-7370

E-mail: h-sawa@czc.hokudai.ac.jp

Abstract

Mosquito-borne alphaviruses are disseminated globally and cause febrile illness in humans and animals. Since the prevalence and diversity of alphaviruses has not been previously investigated in Zambia, reverse transcription PCR was employed as a broad-spectrum approach for the detection of alphaviruses in mosquitoes. From 552 mosquito pools, a novel alphavirus, tentatively named Mwinilunga alphavirus (MWAV), was discovered from a single *Culex quinquefasciatus* mosquito pool. The full genome of MWAV was subsequently determined, and pairwise comparisons demonstrated that MWAV represented a new alphavirus species. Phylogenetic analyses and a linear discriminant analysis based on the dinucleotide ratios in various virus sequences indicated that MWAV is related to a mosquito-specific alphavirus distinct from other known mosquito-borne alphaviruses due to its inability to replicate in vertebrate cell lines. Further analyses of these novel alphaviruses will help to facilitate a greater understanding of the molecular determinants of host range restriction and the evolutionary relationships of alphaviruses.

Keywords: alphavirus; mosquito; *Culex*; Zambia

Main Text

Alphaviruses, belonging to the genus *Alphavirus* in the family *Togaviridae*, can infect diverse vertebrate hosts, including mammals, birds, reptiles, amphibians and fish (Griffin, 2013). Most alphaviruses are mosquito-borne and are transmitted by mosquitoes from at least eight genera (Ortiz et al., 2005; Webb et al., 2008). Restricted interactions between viruses, their invertebrate vector species and vertebrate hosts usually occur in nature; however, some alphaviruses occasionally can cause widespread epizootics (Weaver and Reisen, 2010). Humans and/or animals infected by pathogenic alphaviruses exhibit febrile illnesses that may culminate either in encephalitis or arthritis, depending upon the viral etiology (Atkins, 2013).

In Africa, human diseases caused by the Old World alphaviruses, such as chikungunya virus (CHIKV), o'nyong-nyong virus and Sindbis virus (SINV), have been reported (Lwande et al., 2015). While the Republic of Zambia is located in southern Africa and vector mosquitoes of alphaviruses have been found widely in the country (Gaffigan et al.; Kent, 2006), human cases of chikungunya fever, o'nyong-nyong fever or Sindbis fever have not been documented in prior studies (Lwande et al., 2015; Rezza et al., 2017; Storm et al., 2014) (https://www.cdc.gov/chikungunya/index.html). Moreover, comprehensive detection of alphaviruses in mosquitoes has not been conducted in Zambia to date. Therefore, screening of vector mosquitoes in Zambia is crucial to better predict future outbreaks of alphaviral diseases.

The alphavirus genome consists of a single-stranded, positive-sense RNA, encoding nonstructural proteins (nsPs; nsP1-nsP4), which are important for transcription and replication of viral RNA (Lemm and Rice, 1993), and structural proteins (sPs; Capsid, E3, E2, 6k and E1), which are the main constituents of virions (Voss et al., 2010). Based on genetic and serological similarities, around 30 species of alphaviruses are classified into antigenic complexes (Powers et al., 2012; Powers et al., 2001). In addition, an

unclassified alphavirus, Eilat virus (EILV), has been discovered in *Anopheles coustani* (Nasar et al., 2012) and Taï Forest alphavirus (TALV) has been recently discovered in *Culex decens* (Hermanns et al., 2017).

EILV appears to infect only mosquitoes and differs from other known mosquito-borne alphaviruses which generally maintain their lifecycles between mosquitoes and vertebrate hosts, although EILV is phylogenetically grouped within the mosquito-borne clade (Nasar et al., 2014; Nasar et al., 2012). Studies of EILV have contributed to the identification of the molecular determinants responsible for host range restriction and provided new insights into the evolutionary history of alphaviruses. However, further studies on insect-only alphaviruses are required for a better understanding of the ecology and evolutionary relationships in the genus *Alphavirus* and the molecular determinants of host range restriction (Nasar et al., 2015a; Nasar et al., 2015b). In this study, we have focused on the detection of alphaviruses in mosquitoes collected in Zambia.

To analyze the prevalence of alphaviruses in mosquitoes in Zambia, a total of 9,699 mosquitoes were collected from various locations between 2014 and 2017 (Fig. S1, available in the online Supplementary Material). Mosquito collection was carried out in protected areas with the permission of the Zambia Wildlife Authority, now the Department of National Parks and Wildlife, Ministry of Tourism and Arts of the Republic of Zambia and the Excellence in Research Ethics and Science (ERES) Converge Ethics Committee IRB (No: 00005948). Species were at first identified morphologically using a stereomicroscope, and subsequently some species were genetically confirmed by PCR and sequencing of the cytochrome-oxidase subunit I gene (Folmer et al., 1994). Information on the collected mosquitoes is summarized in Table S1 (available in the online Supplementary Material). The largest numbers of mosquito species were within the genus *Culex*. After species identification, mosquito samples were divided into 552 pools containing 1 to 40 individual mosquitoes, belonging to the same species and

sampling location. Subsequently, the pools were homogenized with Minimum Essential Medium containing 2% fetal bovine serum (Sigma, St. Louis, MO) using the BioMasher (Nippi, Tokyo, Japan). Total RNAs were extracted from 100 µl of the homogenates using the Direct-zol kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol and the remaining homogenates were stored at −80 °C until use for virus detection.

We employed a one-step reverse transcription PCR (RT-PCR) assay for detection of a wide-range of alphaviruses in the extracted mosquito RNAs. Oligonucleotide primers for the RT-PCR assay were designed based on an alignment of the conserved region of nsP4 among defined alphavirus species. One-step RT-PCR was performed with the PrimeScript One Step RT-PCR Kit, Ver. 2 (Takara, Shiga, Japan) with 1 µl of extracted mosquito RNA and 1 µM (final concentration) of each of the primers, pan-Alpha-nsP4-6692F (5'-CAYACRYTRTTYGAYATGTCDGC-3') and pan-Alpha-nsP4-7152R (5'-GCRTCDATKATYTTBACYTCCAT-3') in 15 µl reaction volumes, with the following thermocycling protocol: 50 °C for 30 minutes; 94 °C for 2 minutes; 43 cycles of 94 °C for 30 seconds, 52 °C for 30 seconds, and 72 °C for 30 seconds; and 72 ˚C for 5 minutes. The PCR products (approximately 460 bp) were visualized on a 1% (w/v) agarose gel stained with ethidium bromide. We confirmed that CHIKV and SINV nsP4 genes were also detected by our methodology (data not shown).

From 552 pools, one pool, which contained 19 *Culex quinquefasciatus* mosquitoes collected in Mwinilunga (North-Western Province of Zambia) in 2016, was found to be positive by RT-PCR assay. The amplified products were sequenced using the ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) with the pan-Alpha-nsP4-6692F primer and pan-Alpha-nsP4-7152R primer, and the nucleotide sequences were analyzed using GENETYX, version 13 (GENETYX, Tokyo, Japan). Sequence comparison of the amplified fragment in the National Center for Biotechnology Information database employing BLASTN analysis showed 76% nucleotide identity to

TALV and 75% identity to EILV. The detected alphavirus was tentatively named Mwinilunga alphavirus (MWAV). Other known alphaviruses were not detected in the present study.

We sought to isolate MWAV from RT-PCR-positive mosquitoes using *Aedes albopictus* cells (C6/36), African green monkey kidney cells (Vero) and baby hamster kidney cells (BHK-21). Homogenized samples of mosquitoes were diluted five-fold with DMEM supplemented with 2% fetal bovine serum, penicillin, streptomycin, gentamycin and 2 mM L-glutamine (Sigma), filtered and used to inoculate onto cells. After inoculation of the mosquito homogenates, supernatants of the cells were cultured through three passages for three weeks. Viral production into the supernatant was checked by RT-PCR with extracted RNA from supernatants of cultured cells and the pan-Alpha-nsP4 primers after three blind passages; however, proliferation of MWAV was not detected in any of the three cell lines tested.

To determine the full-length MWAV genome sequence, RNA extracted from the mosquitoes was subjected to next-generation sequencing. A library was constructed by the Nextera XT DNA Library Prep kit (Illumina, San Diego, CA). Sequencing was performed with a MiSeq Reagent Kit v3 (600 cycles) and Illumina MiSeq System (Illumina) following the manufacturer's protocol. After trimming low-quality reads, the resulting reads were *de novo* assembled using a CLC Genomics Workbench 10 (CLC bio, Qiagen, Hilden, Germany) with default parameters. The sequences were confirmed by specific PCRs using MWAV-specific primer pairs. The 5'- and the 3'- terminal regions were amplified using the rapid amplification of cDNA ends (RACE) approach with the SMARTer RACE cDNA Amplification Kit (Takara) and MWAV-specific primers. Amplified products were directly sequenced in both directions by Sanger sequencing. The full-genome sequence of MWAV was determined and the amino acid sequences were predicted using GENETYX, version 13. The MWAV genome consisted of 11,547 nt,

coding two open reading frames (ORFs) (2,439 and 1,235 amino acids) in the positive sense (Fig. 1). The full-genome sequence of MWAV was deposited in the GenBank/EMBL/DDBJ database under accession number LC361437. The pairwise nucleotide identities of MWAV were 70% to EILV, 58% to TALV and 52% to SINV. BLAST analysis indicated that the deduced amino acid sequence of one ORF (2,439 amino acids) was related to that of nsPs from alphaviruses and the highest identity was 83% to EILV. The deduced amino acid sequence of the second ORF (1,235 amino acids) was related to that of sPs from alphaviruses and the highest identity was 76% to TALV. The pairwise amino acid identities of MWAV ranged between 64% (E2 protein) and 88% (nsP1) to EILV for the corresponding proteins, and between 68% (nsP3) and 87% (nsP4) to TALV for the corresponding proteins (Fig. 1). According to the species designation criteria of the International Committee on Taxonomy of Viruses (Powers et al., 2012), the genetic analysis of MWAV suggested that this represented a novel alphavirus species, closely related to EILV and TALV. Eilat-like viruses, including EILV, TALV and MWAV were detected from different mosquito species including *Anopheles coustani* (Nasar et al., 2012), *Culex decens* (Hermanns et al., 2017) and *Culex quinquefasciatus*, respectively. This supports the view that those viruses have mosquito-restricted host ranges (Nasar et al., 2014). In addition, the narrow mosquito vector range may potentially explain why MWAV could not be isolated using *Aedes*-derived and mammalian-derived cells.

Since conserved sequence elements (CSEs) have important roles in the lifecycle of alphaviruses (Fayzulin and Frolov, 2004; Hyde et al., 2015; Niesters and Strauss, 1990; Zhu et al., 2013), we investigated the sequence identities of CSEs with other defined alphaviruses. The CSEs of MWAV within the nsP1 51 nt, subgenomic promoter sequence and the 3'-UTR shared more than 90% nt sequence identities with EILV and TALV (Fig. S2, available in the online Supplementary Material). The identity of the CSE within the putative 5'-UTR was 80% to that of EILV, and mFold analysis suggested that the putative

5'-UTR of MWAV might form secondary stem-loop structures, similar to EILV (data not shown) (Zuker, 2003). In addition, the conserved amino acid motifs, such as the ribosomal binding site, conserved RNA-dependent RNA polymerase motifs of MWAV were similar to those of other alphaviruses (data not shown) and the putative protease cleavage sites resembled those of EILV (Fig. S3, available in the online Supplementary Material). These sequence similarities suggest that MWAV may possess similar biological properties to EILV, which also displays a narrow vector range in mosquitoes (Nasar et al., 2014).

The full-length nucleotide sequences encoding either the nsPs or sPs ORFs of MWAV and alphaviruses available from GenBank (Table S2, available in the online Supplementary Material), were aligned using the ClustalW protocol (Thompson et al., 1994). For the alignments, the C terminus of nsP3 and the N terminus of capsid sequences were excluded, since these regions have numerous insertions and deletions and extensive sequence divergence in previous studies (Forrester et al., 2012; Nasar et al., 2012). Phylogenetic trees were constructed by maximum-likelihood method using MEGA, version 6.0 (Tamura et al., 2013). The robustness of the nodes was tested by 1,000 bootstrap replications. In both phylogenies based on nucleotide sequences of nsPs and sPs, MWAV was classified into the same phylogroup with EILV and TALV (Fig. 2), suggesting that MWAV belongs to the same antigenic complex with these two viruses in agreement with the genome analysis (Fig. 1). In both the nsPs and sPs trees, Eilat-like viruses (EILV, TALV and MWAV) have the same ancestral virus within the Western equine encephalitis (WEE) complex, placed between Trocara virus (TROV) and Aura virus (AURAV), as has been reported previously (Hermanns et al., 2017; Nasar et al., 2012). Although Eilat-like viruses have been detected in the Old World, TROV and AURAV have also been isolated from mosquitoes in South American countries (Rümenapf et al., 1995; Travassos da Rosa et al., 2001). Similar to Eilat-like viruses, vertebrate hosts of both TROV and AURAV have not been identified (Forrester et al.,

2012).

To predict the host range of MWAV, a linear discriminant analysis was performed based on the dinucleotide ratios in various virus sequences, which has been reported previously to determine host specificity based on the frequencies of dinucleotide in flavivirus sequences (Colmant et al., 2017). Complete genome sequences from 53 mosquito-borne alphaviruses with a defined host (bird, horse, donkey, human, hamster and mosquito) and full length sequences of nsPs and sPs from MWAV, EILV, TALV, TROV and AURAV were used for the analysis (Table S3, available in the online Supplementary Material). The dinucleotide ratios were calculated using the formula $P_{XY} = f_{XY}/f_X f_Y$, in which f_X and f_Y denote the frequencies of mononucleotides X and Y, respectively, and f_{XY} denotes the frequency of dinucleotide XY (Karlin and Mrázek, 1997; Lobo et al., 2009). The linear discriminant analysis was performed in the R package (Colmant et al., 2017; Di Giallonardo et al., 2017) (https://www.R-project.org/). The analysis showed correlations of dinucleotide usage patterns between alphaviruses and their host ranges (Fig. 3). Mosquito-borne alphaviruses showed similar dinucleotide usage patterns regardless of their phylogenetic relationships and hosts. On the other hands, the dinucleotide usage pattern of Eilat-like viruses were also similar to each other, indicating that MWAV may be a mosquito-specific alphavirus. Interestingly, TROV and AURAV, who have no known vertebrate hosts showed a different pattern from those of mosquito-bone alphaviruses and Eilat-like viruses.

If Eilat-like viruses and phylogenetically related viruses have adapted to specific mosquito hosts and relied on vertical and/or venereal transmission, it is possible that these viruses have not been spread *via* vertebrate hosts to other mosquito species. Therefore, 1) Eilat-like viruses, and phylogenetically related viruses, may exist in different parts of the world and have lifecycles involving vertical and/or venereal transmission exclusively in invertebrate mosquito hosts, or 2) Eilat-like viruses might be sporadically transported

from other areas and based on a previous report, geographic introductions of alphaviruses may have repeatedly occurred (Forrester et al., 2012). It is possible that Eilat-like viruses and/or their ancestral viruses might have lost their ability to infect vertebrates after being spread to other mosquito species in the Old World. Analyses of the biological properties of MWAV and TALV after isolation will greatly help to clarify the factors that determine the host range and the evolutionary histories of the Eilat-like alphaviruses.

In summary, we have identified a novel alphavirus, tentatively named MWAV from *Culex quinquefasciatus* mosquitoes in Zambia. Genetic analyses revealed that MWAV is closely related to, but distinct from, Eilat virus, a mosquito-specific virus and key to the better understanding of the evolution of alphaviruses. Further epidemiological and functional studies investigating arthropod and animal host species will help to better characterize future alphaviral disease risk and also provide new insights into their ecology.

Funding information

This study was supported by grants for Scientific Research (C), (B) from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT) /Japan Society for the Promotion of Science (JSPS) KAKENHI (JP15K07715, JP17K00401, JP16H05805); and grants for Scientific Research on Innovative Areas and International Group from the MEXT/JSPS KAKENHI (JP16H06431, JP16H06429, JP16K21723), and the Japan Initiative for Global Research Network on Infectious Diseases, from the Japan Agency for Medical Research and Development.

Acknowledgements

We appreciate the cooperation of Veterinary officers of the Department of Veterinary Services, Ministry of Fisheries and Livestock of the Republic of Zambia in each district for mosquito sampling.

References

Atkins, G.J., 2013. The Pathogenesis of Alphaviruses. Vol. 2018. ISRN Virology.

Colmant, A.M.G., Hobson-Peters, J., Bielefeldt-Ohmann, H., van den Hurk, A.F., Hall-Mendelin, S., Chow, W.K., Johansen, C.A., Fros, J., Simmonds, P., Watterson, D., Cazier, C., Etebari, K., Asgari, S., Schulz, B.L., Beebe, N., Vet, L.J., Piyasena, T.B.H., Nguyen, H.D., Barnard, R.T., Hall, R.A., 2017. A New Clade of Insect-Specific Flaviviruses from Australian Anopheles Mosquitoes Displays Species-Specific Host Restriction. mSphere 2(4).

Di Giallonardo, F., Schlub, T.E., Shi, M., Holmes, E.C., 2017. Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species. J Virol 91(8).

Fayzulin, R., Frolov, I., 2004. Changes of the secondary structure of the 5' end of the Sindbis virus genome inhibit virus growth in mosquito cells and lead to accumulation of adaptive mutations. J Virol 78(10), 4953-4964.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3(5), 294-299.

Forrester, N.L., Palacios, G., Tesh, R.B., Savji, N., Guzman, H., Sherman, M., Weaver, S.C., Lipkin, W.I., 2012. Genome-scale phylogeny of the alphavirus genus suggests a marine origin. J Virol 86(5), 2729-2738.

Gaffigan, T.V., Wilkerson, R.C., Pecor, J.E., Stoffer, J.A., Anderson, T., Systematic Catalog of Culicidae. Vol. 2018. The Walter Reed Biosystematics Unit

Griffin, D.E., 2013. Alphaviruses. In: David M. Knipe, P.M.H., Jeffrey I. Cohen, Diane E. Griffin, Robert A. Lamb, Malcolm A. Martin, Vincent R. Racaniello, Bernard Roizman (Ed.), Fields Virology 6th edition, 6th ed. Vol. 1. Wolters

Kluwer/Lippincott Williams & Wilkins, Philadelphia, United States, pp. 651-686.

Hermanns, K., Zirkel, F., Kopp, A., Marklewitz, M., Rwego, I.B., Estrada, A., Gillespie, T.R., Drosten, C., Junglen, S., 2017. Discovery of a novel alphavirus related to Eilat virus. J Gen Virol 98(1), 43-49.

Hyde, J.L., Chen, R., Trobaugh, D.W., Diamond, M.S., Weaver, S.C., Klimstra, W.B., Wilusz, J., 2015. The 5' and 3' ends of alphavirus RNAs--Non-coding is not non-functional. Virus Res 206, 99-107.

Karlin, S., Mrázek, J., 1997. Compositional differences within and between eukaryotic genomes. Proc Natl Acad Sci U S A 94(19), 10227-10232.

Kent, R.J., 2006. The Mosquitoes of Macha, Zambia. Vol. 2018.

Lemm, J.A., Rice, C.M., 1993. Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. J Virol 67(4), 1916-1926.

Lobo, F.P., Mota, B.E., Pena, S.D., Azevedo, V., Macedo, A.M., Tauch, A., Machado, C.R., Franco, G.R., 2009. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. PLoS One 4(7), e6282.

Lwande, O.W., Obanda, V., Bucht, G., Mosomtai, G., Otieno, V., Ahlm, C., Evander, M., 2015. Global emergence of Alphaviruses that cause arthritis in humans. Infect Ecol Epidemiol 5, 29853.

Nasar, F., Erasmus, J.H., Haddow, A.D., Tesh, R.B., Weaver, S.C., 2015a. Eilat virus induces both homologous and heterologous interference. Virology 484, 51-58.

Nasar, F., Gorchakov, R.V., Tesh, R.B., Weaver, S.C., 2015b. Eilat virus host range restriction is present at multiple levels of the virus life cycle. J Virol 89(2), 1404-1418.

Nasar, F., Haddow, A.D., Tesh, R.B., Weaver, S.C., 2014. Eilat virus displays a narrow mosquito vector range. Parasit Vectors 7, 595.

Nasar, F., Palacios, G., Gorchakov, R.V., Guzman, H., Da Rosa, A.P., Savji, N., Popov, V.L., Sherman, M.B., Lipkin, W.I., Tesh, R.B., Weaver, S.C., 2012. Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. Proc Natl Acad Sci U S A 109(36), 14622-14627.

Niesters, H.G., Strauss, J.H., 1990. Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. J Virol 64(4), 1639-1647.

Ortiz, D.I., Anishchenko, M., Weaver, S.C., 2005. Susceptibility of Psorophora confinnis (Diptera: Culicidae) to infection with epizootic (subtype IC) and enzootic (subtype ID) Venezuelan Equine encephalitis viruses. J Med Entomol 42(5), 857-863.

Powers, A., Huang, H., Roehrig, J., Strauss, E., Weaver, S., 2012. Family Togaviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E. (Eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Oxford, UK, pp. 1103-1110.

Powers, A.M., Brault, A.C., Shirako, Y., Strauss, E.G., Kang, W., Strauss, J.H., Weaver, S.C., 2001. Evolutionary relationships and systematics of the alphaviruses. J Virol 75(21), 10118-10131.

Rezza, G., Chen, R., Weaver, S.C., 2017. O'nyong-nyong fever: a neglected mosquito-borne viral disease. Pathog Glob Health 111(6), 271-275.

Rümenapf, T., Strauss, E.G., Strauss, J.H., 1995. Aura virus is a New World representative of Sindbis-like viruses. Virology 208(2), 621-633.

Storm, N., Weyer, J., Markotter, W., Kemp, A., Leman, P.A., Dermaux-Msimang, V., Nel, L.H., Paweska, J.T., 2014. Human cases of Sindbis fever in South Africa, 2006-2010. Epidemiol Infect 142(2), 234-238.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12), 2725-2729.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22), 4673-4680.

Travassos da Rosa, A.P., Turell, M.J., Watts, D.M., Powers, A.M., Vasconcelos, P.F., Jones, J.W., Klein, T.A., Dohm, D.J., Shope, R.E., Degallier, N., Popov, V.L., Russell, K.L., Weaver, S.C., Guzman, H., Calampa, C., Brault, A.C., Lemon, A.P., Tesh, R.B., 2001. Trocara virus: a newly recognized Alphavirus (Togaviridae) isolated from mosquitoes in the Amazon Basin. Am J Trop Med Hyg 64(1-2), 93-97.

Voss, J.E., Vaney, M.C., Duquerroy, S., Vonrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G., Rey, F.A., 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature 468(7324), 709-712.

Weaver, S.C., Reisen, W.K., 2010. Present and future arboviral threats. Antiviral Res 85(2), 328-345.

Webb, C.E., Doggett, S.L., Ritchie, S.A., Russell, R.C., 2008. Vector competence of three Australian mosquitoes, Verrallina carmenti, Verraullina lineata, and Mansonia septempunctata (Diptera: Culicidae), for Ross River virus. J Med Entomol 45(4), 737-740.

Zhu, W.Y., Zhao, T., Liang, G.D., 2013. Specific nucleotide changes in the subgenomic promoter region influence infectivity of the sindbis virus. Intervirology 56(5), 337-342.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13), 3406-3415.

Chikungunya virus. Centers for Disease Control and Prevention, Atlanta, GA, United

States. https://www.cdc.gov/chikungunya/index.html (accessed January 2018).

The R project for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed April 2018).

Figure Legends

Fig. 1. Genome organization of Mwinilunga alphavirus

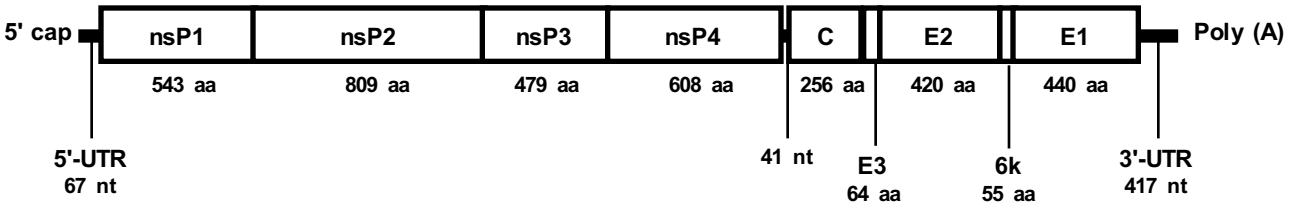
The nucleotide lengths of the UTRs and amino acid sizes of each protein are displayed under the genome. Amino acid sequence identities of each protein between Mwinilunga alphavirus and closely related alphaviruses are indicated.

Fig. 2. Phylogenetic analysis of Mwinilunga alphavirus with previously described alphaviruses

Phylogenetic trees were constructed using the maximum-likelihood method in MEGA, version 6.0 with 1,000 bootstrap replicates based on multiple alignment of nucleotide sequences of the (a) structural proteins and (b) nonstructural proteins. Bootstrap values greater than 50 are shown near the branch nodes and the scale bar indicates the number of substitutions per site.

Fig. 3. Score plot of the linear discriminant analysis

The figure shows a scatterplot of the two discriminant scores that explain the largest amount of the components from the linear discriminant analysis (72.3% and 16.2% for LD1 and LD2, respectively). Markers represent values for individual sequences as follows: complete genome sequences from mosquito-borne alphaviruses (crosses), full length of nsPs and sPs sequences from Mwinilunga alphavirus (black squares), Eilat-like viruses (triangles) or alphaviruses with no known vertebrate hosts (circles).



	Amino acid sequence identity to Mwinilunga alphavirus (%)								
Virus	**nsP1**	**nsP2**	**nsP3**	**nsP4**	**Capsid**	**E3**	**E2**	**6k**	**E1**
Eilat virus	88	83	69	87	75	73	64	83	77
Taï Forest alphavirus	83	82	68	87	75	73	71	83	79
Sindbis virus	70	64	41	78	55	51	43	47	50

Fig. 1.

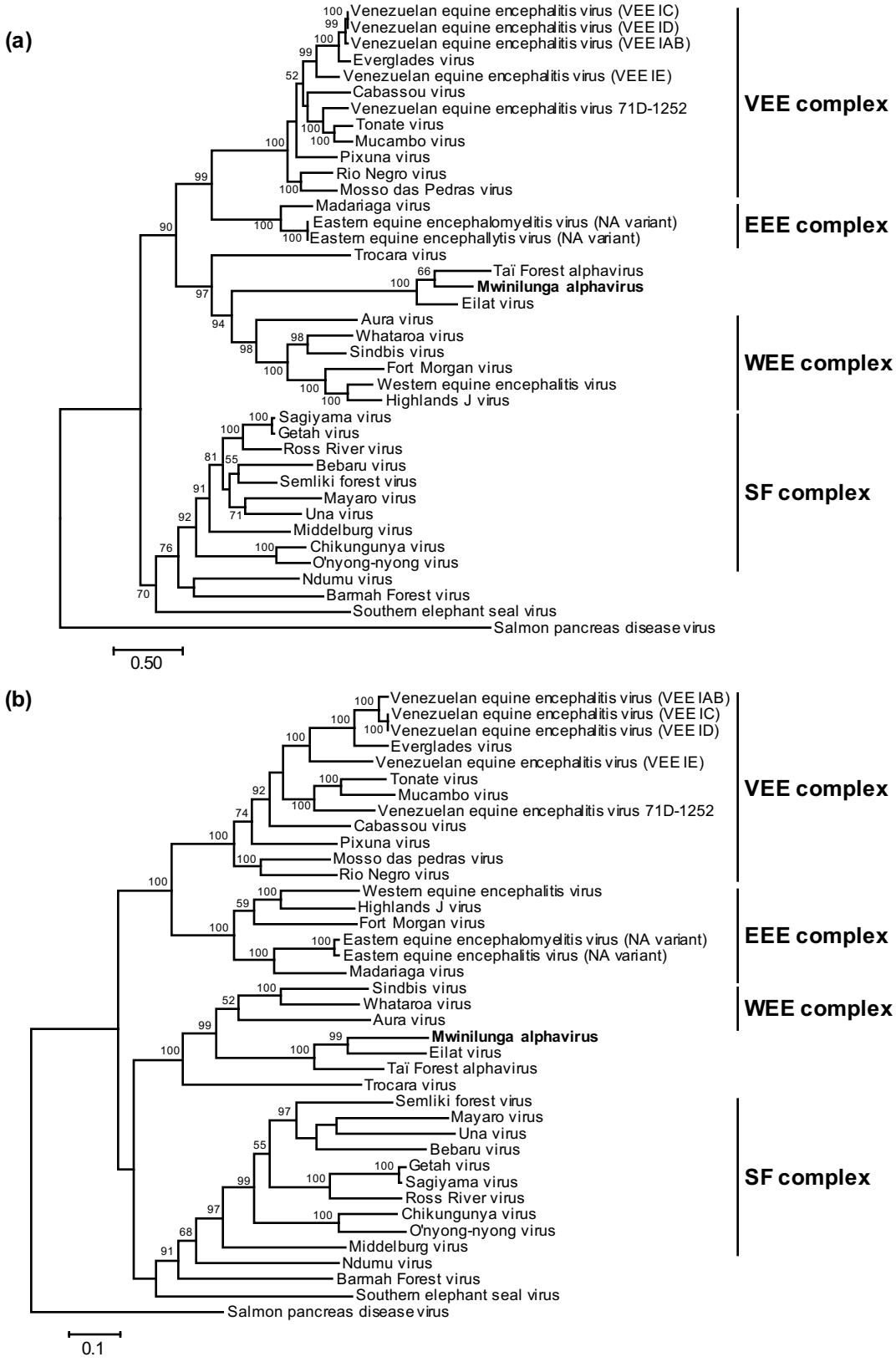

(a)
Venezuelan equine encephalitis virus (VEE IC)
Venezuelan equine encephalitis virus (VEE ID)
Venezuelan equine encephalitis virus (VEE IAB)
Everglades virus
Venezuelan equine encephalitis virus (VEE IE)
Cabassou virus
Venezuelan equine encephalitis virus 71D-1252
Tonate virus
Mucambo virus
Pixuna virus
Rio Negro virus
Mosso das Pedras virus
Madariaga virus
Eastern equine encephalomyelitis virus (NA variant)
Eastern equine encephallytis virus (NA variant)
Trocara virus
Taï Forest alphavirus
Mwinilunga alphavirus
Eilat virus
Aura virus
Whataroa virus
Sindbis virus
Fort Morgan virus
Western equine encephalitis virus
Highlands J virus
Sagiyama virus
Getah virus
Ross River virus
Bebaru virus
Semliki forest virus
Mayaro virus
Una virus
Middelburg virus
Chikungunya virus
O'nyong-nyong virus
Ndumu virus
Barmah Forest virus
Southern elephant seal virus
Salmon pancreas disease virus
VEE complex
EEE complex
WEE complex
SF complex
0.50
(b)
Venezuelan equine encephalitis virus (VEE IAB)
Venezuelan equine encephalitis virus (VEE IC)
Venezuelan equine encephalitis virus (VEE ID)
Everglades virus
Venezuelan equine encephalitis virus (VEE IE)
Tonate virus
Mucambo virus
Venezuelan equine encephalitis virus 71D-1252
Cabassou virus
Pixuna virus
Mosso das pedras virus
Rio Negro virus
Western equine encephalitis virus
Highlands J virus
Fort Morgan virus
Eastern equine encephalomyelitis virus (NA variant)
Eastern equine encephalitis virus (NA variant)
Madariaga virus
Sindbis virus
Whataroa virus
Aura virus
Mwinilunga alphavirus
Eilat virus
Taï Forest alphavirus
Trocara virus
Semliki forest virus
Mayaro virus
Una virus
Bebaru virus
Getah virus
Sagiyama virus
Ross River virus
Chikungunya virus
O'nyong-nyong virus
Middelburg virus
Ndumu virus
Barmah Forest virus
Southern elephant seal virus
Salmon pancreas disease virus
VEE complex
EEE complex
WEE complex
SF complex
0.1

Fig. 2.

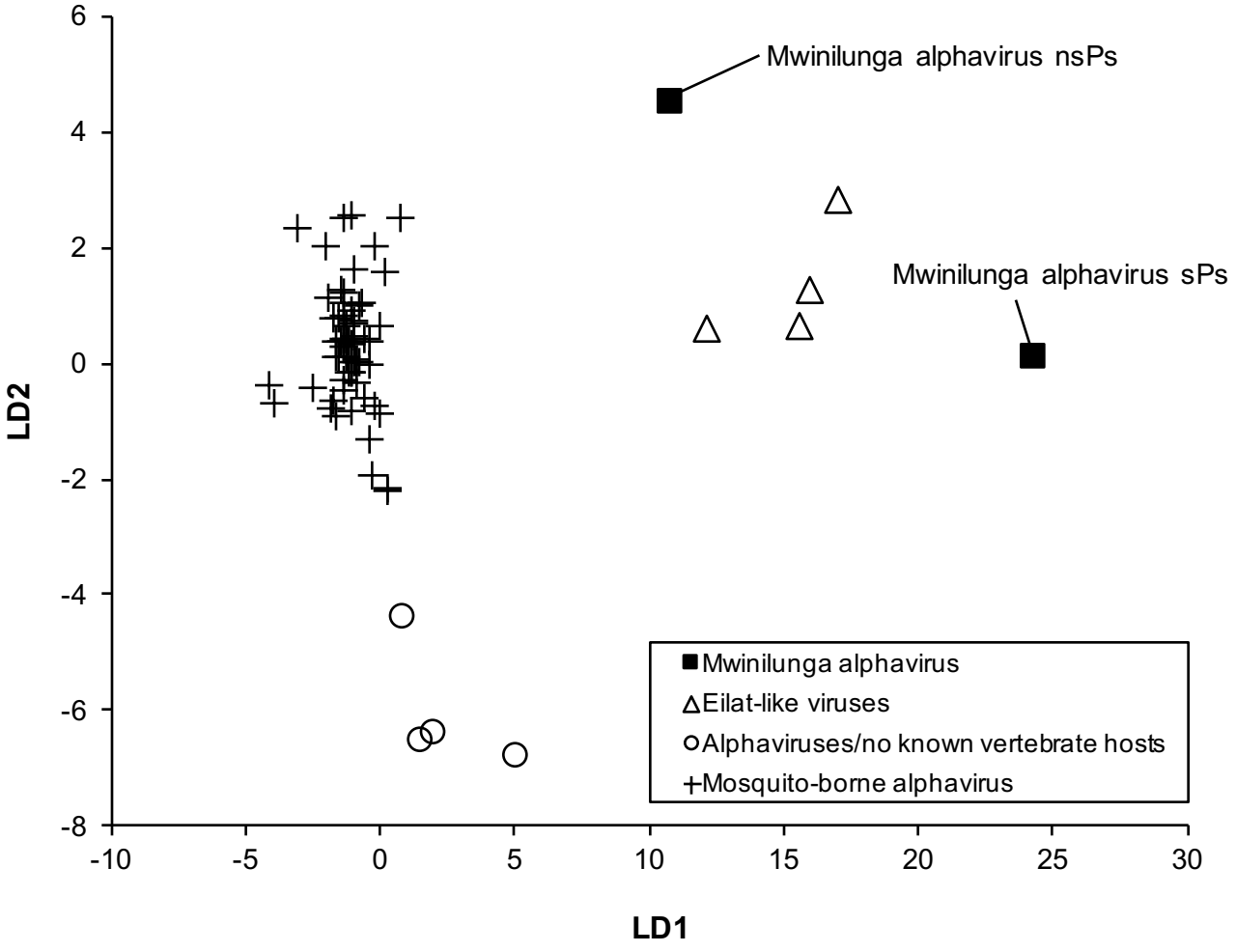

Mwinilunga alphavirus nsPs
Mwinilunga alphavirus sPs
Mwinilunga alphavirus
Eilat-like viruses
Alphaviruses/no known vertebrate hosts
Mosquito-borne alphavirus
LD1
LD2
6
4
2
0
-2
-4
-6
-8
-10
-5
0
5
10
15
20
25
30

Fig. 3.

Table S1. Information of collected mosquitoes

Year	Place	Mosquito genus	Pool no.	Mosquito no.
2014	Lusaka	*Culex*	10	258
	Mongu	*Anopheles*	2	19
		Coquillettidia	6	19
		Culex	52	1399
		Mansonia	6	26
2015	Chipata	*Aedes*	1	1
		Culex	16	306
	Kazungula	*Aedeomya*	1	1
		Aedes	3	5
		Anopheles	9	38
		Culex	5	45
		Mansonia	1	2
	Livingstone	*Aedes*	3	12
		Anopheles	3	4
		Culex	27	774
		Mansonia	1	8
	Lusaka	*Aedes*	3	20
		Anopheles	1	1
		Culex	54	1574
		Mansonia	2	2
2016	Kitwe	*Culex*	10	275
	Mongu	*Aedes*	3	4
		Anopheles	25	366
		Coquillettidia	10	153
		Culex	23	413
		Mansonia	5	99
		Uranotaenia	1	2
	Mwinilunga	*Aedes*	4	4
		Anopheles	2	3
		Culex	6	40
		Uranotaenia	1	2
	Ndola	*Culex*	2	19
	Siavonga	*Aedes*	22	141
		Anopheles	3	3
		Culex	19	385
2017	Livingstone	*Aedes*	8	12
		Anopheles	16	73
		Culex	46	812
		Mansonia	2	2
	Mongu	*Aedeomya*	2	10
		Aedes	4	11
		Anopheles	66	1451
		Coquillettidia	9	62
		Culex	49	710
		Mansonia	7	132
		Uranotaenia	1	1
		Total	**552**	**9699**

Table S2. Reference sequences for phylogenetic analyses

Species	Acession Number
Aura virus	NC 003900
Barmah Forest virus	NC 001786
Bebaru virus	NC 016962
Cabassou virus	AF075259
Chikungunya virus	NC 004162
Eastern equine encephalitis virus (NA variant)	NC 003899
Eastern equine encephalomyelitis virus (NA variant)	U01034
Eilat virus	NC 018615
Everglades virus	AF075251
Fort Morgan virus	NC 013528
Getah virus	NC 006558
Highlands J virus	NC 012561
Madariaga virus	NC 023812
Mayaro virus	NC 003417
Middelburg virus	NC 024887
Mosso das Pedras virus	AF075257
Mucambo virus	AF075253
Ndumu virus	NC 016959
O'nyong-nyong virus	NC 001512
Pixuna virus	AF075256
Rio Negro virus	AF075258
Ross River virus	DQ226993
Sagiyama virus	AB032553
Salmon pancreas disease virus	NC 003930
Semliki forest virus	NC 003215
Sindbis virus	NC 001547
Southern elephant seal virus	NC 016960
Taï Forest alphavirus	NC 032681
Tonate virus	AF075254
Trocara virus	HM147991
Una virus	HM147992
Venezuelan equine encephalitis virus (VEE IAB)	KC344505
Venezuelan equine encephalitis virus (VEE IC)	KP282671
Venezuelan equine encephalitis virus (VEE ID)	NC 001449
Venezuelan equine encephalitis virus (VEE IE)	KC344501
Venezuelan equine encephalitis virus 71D-1252	AF075255
Western equine encephalitis virus	NC 003908
Whataroa virus	NC 016961

Table S3. Reference sequences for linear discriminant analysis

Accession number	Species	Isolation host
NC 018615	Eilat virus	Mosquito
NC 032681	Taï Forest alphavirus	Mosquito
NC 003900	Aura virus	Mosquito
HM147991	Trocara virus	Mosquito
GU001935	Eastern equine encephalitis virus	Bird
GU001913	Eastern equine encephalitis virus	Bird
GU001914	Eastern equine encephalitis virus	Bird
FJ827631	Highlands J virus	Bird
GU167952	Highlands J virus	Bird
KT429024	Highlands J virus	Bird
KJ409555	Highlands J virus	Bird
KT429026	Highlands J virus	Bird
GU001917	Eastern equine encephalitis virus	Horse
GU001915	Eastern equine encephalitis virus	Horse
GU001920	Eastern equine encephalitis virus	Horse
U01557	Eastern equine encephalomyelitis virus	Horse
KT429021	Highlands J virus	Horse
KT429023	Highlands J virus	Horse
AY973944	Venezuelan equine encephalitis virus	Horse
AY986475	Venezuelan equine encephalitis virus	Horse
L01442	Venezuelan equine encephalitis virus	Donkey
KJ554965	Western equine encephalitis virus	Horse
GQ287645	Western equine encephalomyelitis virus	Horse
GQ287643	Western equine encephalomyelitis virus	Horse
KX702402	Chikungunya virus	Human
KX496989	Chikungunya virus	Human
KY703990	Chikungunya virus	Human
KY704955	Chikungunya virus	Human
KX000164	Eastern equine encephalitis virus	Human
NC 023812	Madariaga virus	Human
NC 003417	Mayaro virus	Human
KY302801	Ross River virus	Human
JQ771794	Sindbis virus	Human
JQ771797	Sindbis virus	Human
JQ771798	Sindbis virus	Human
KC344505	Venezuelan equine encephalitis virus	Human
KP282671	Venezuelan equine encephalitis virus	Human
GQ287640	Western equine encephalomyelitis virus	Human
KC344491	Venezuelan equine encephalitis virus	Hamster
KC344515	Venezuelan equine encephalitis virus	Hamster
GU908223	Chikungunya virus	Mosquito
AY726732	Chikungunya virus	Mosquito
HM045805	Chikungunya virus	Mosquito
HM045818	Chikungunya virus	Mosquito
HM045820	Chikungunya virus	Mosquito
HM045815	Chikungunya virus	Mosquito
HM045819	Chikungunya virus	Mosquito
HM045784	Chikungunya virus	Mosquito
HM045785	Chikungunya virus	Mosquito
GU361116	Sindbis virus	Mosquito
JQ771793	Sindbis virus	Mosquito
JQ771799	Sindbis virus	Mosquito
KC344525	Venezuelan equine encephalitis virus	Mosquito
KC344512	Venezuelan equine encephalitis virus	Mosquito
KC344500	Venezuelan equine encephalitis virus	Mosquito
KJ554983	Western equine encephalitis virus	Mosquito
KJ554989	Western equine encephalitis virus	Mosquito

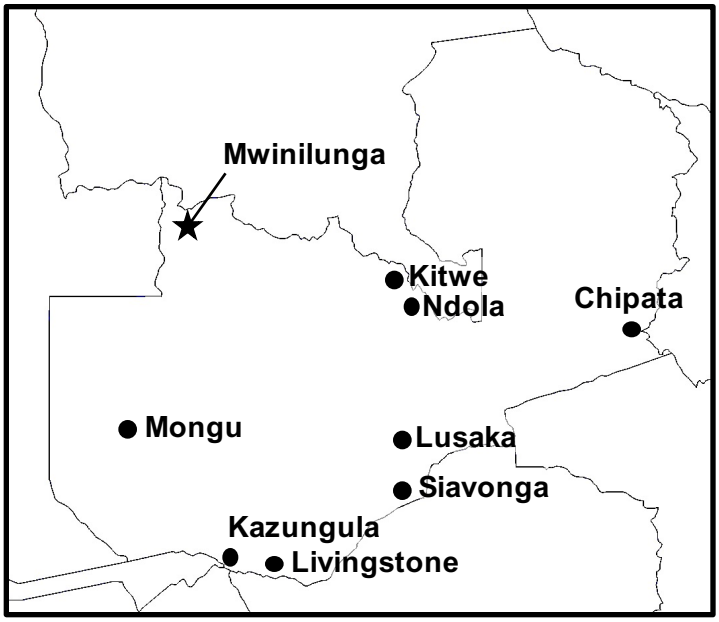

Mwinilunga
Kitwe
Ndola
Chipata
Mongu
Lusaka
Siavonga
Kazungula
Livingstone

Fig. S1.

(a)

Virus	5'-UTR
Mwinilunga alphavirus	ACAUGGGGAUAGGCUAUAUAACACACAAU-UAAACCCAGUACCAAAUAGCCUCCCACUUUCAUCGAAU
Eilat virus	--------.....GC.C.UU.....U.AG..U.----GUA.CG.....U.....C
Taï Forest alphavirus	--------.....CG.U.....C.U.....U.GU.GAU.....U.....

(b)

Virus	nsP1 51 nt CSE
Mwinilunga alphavirus	ACAGGUCACUCCGAAUGACCACGCCAAUGCGAGGGCGUUCUCGCAUUGCGC
Eilat virus	CAU.....C.....A..C.....C.....
Taï Forest alphavirus	C.U.....C.....U..A..U.....
Trocara virus	G.....G.C.....U..U.....C..A.....CUG..
Aura virus	G.....U..U.....C..A..U..U.....CUG..
Western equine encephalitis virus	G.....GAC.....U.....C..A.....U.....GUG..
Sindbis virus	G.....A.....U..U.....C..A..A..U.....CUG..
Eastern equine encephalitis virus	G.....GAC.....U..U.....U.....U.....CCUA..
Venezuelan equine encephalitis virus	G.....GAU.....U..U.....C..A.....U.....CUG..
Ross River virus	G.....A..C.....U..U.....C..A..U..U.....CUG..
Semliki forest virus	G.....A..A.....U..A.....C..A..A..U.....CCUG..
Una virus	G.....A.....U..G..C.....U.....CCU...
Chikungunya virus	G.....A.....U..U.....U..A.....CUA..
Barmah Forest virus	G...AC...A..A..C..U..U..AC.C.....U.....CCUU..
Southern elephant seal virus	G.....A..U.....U.....C..A..U..U.....CUG..
Salmon pancreas disease virus	CA.UAGGU.GU.U..C....U...GCC..C..A..U....C..C.UG..

(c)

Virus	Subgenomic promoter
Mwinilunga alphavirus	CCCUCUACAACUGACCUAAAUAGU
Eilat virus	.....A.....
Taï Forest alphavirus	A...G..GGUG.....
Trocara virus	A.....G.A...U..C..
Aura virus	A.....GGUG.U.....A
Western equine encephalitis virus	.....GG.....G
Sindbis virus	AU.....GGUG.U.....
Eastern equine encephalitis virus	.....GG..... .....G
Venezuelan equine encephalitis virus	.U.....GG..A...G...G.A
Ross River virus	A.....GG.G.U.....A
Semliki forest virus	A.....GG.G.U....G..U.G
Chikungunya virus	.UU.G..GG.G.U.....G
Barmah Forest virus	AU.....GGUG.U.....
Southern elephant seal virus	.G.....GG...U.....A
Salmon pancreas disease virus	.....GU..A....U....U.

(d)

Virus	3'-UTR
Mwinilunga alphavirus	AUUUUGUUUUUAAUAUUUC
Eilat virus	.A.....C
Taï Forest alphavirus	.....
Western equine encephalitis virus	.....A.....
Sindbis virus	.....C.....
Eastern equine encephalitis virus	.....
Venezuelan equine encephalitis virus	.....
Ross River virus	U..G.U.....UAC
Semliki forest virus	.A..G.....C
Una virus	...GGU.....C
Barmah Forest virus	....GU.....UAC
Salmon pancreas disease virus	C.A...G.....A....UCAAUAC

Fig. S2.

Virus	nsP1/nsP2	nsP2/nsP3	nsP3/nsP4	Capsid/E3	E3/E2
Mwinilunga alphavirus	DVGG/ALVE	EIGA/APSY	GAGG/YIFS	TVEW/SAIV	RSKR/SAPG
Eilat virus	.I../....	GV../....	..../....	..../....	.AR./AVAP
Taï Forest alphavirus	..../....	GV../....	..../....	..../....	.IR./.TQP
Trocara virus	.I.A/...D	G..C/....	.V../....	..K./..AT	.P../.TEL
Aura virus	.A.A/....	GS../....	.V../....	..../.RAI	.HV./.T.T
Western equine encephalitis virus	EA.A/GS..	EA.R/..A.	E..A/....	SES./.LVT	.Q../.ITD
Sindbis virus	.I.A/....	GV../....	.V../....	.E../..AP	..../.VID
Eastern equine encephalitis virus	EA.A/GS..	EA.R/..A.	E..A/....	SEP./.LAT	.TR./DLDT
Venezuelan equine encephalitis virus	EA.A/GS..	EA.C/....	D..A/....	CEQ./.LVT	.KR./.TEE
Ross River virus	RA.A/GV..	TA.C/....	R..A/....	.E../..AL	.HR./.VTE
Semliki forest virus	HA.A/GV..	TA.C/....	R..A/....	SE../..PL	.HR./.VS.
Una virus	RA.A/GV..	TA.C/..A.	R.../....	..../..PL	.HR./.VT.
Chikungunya virus	RA.A/GII.	RA.C/....	R.../....	AE../.LAI	.QR./.IKD
Barmah Forest virus	RA.E/GV..	PA.S/..A.	R.../....	S.../..AA	.PK./.VAH
Southern elephant seal virus	RA.A/GV..	PA.T/..N.	R..A/....	..../..LT	.GK./.VAS
Salmon pancreas disease virus	GA.A/TIID	MV../..G.	.L../....	AIP./TRAP	.RK./AVST

Fig. S3.

Figure Legends

Fig. S1. The map of study sites in Zambia

Geographic areas where mosquitoes were collected from 2014 to 2017 (black circles) are shown. The district where the alphavirus-positive mosquito pool was collected in North-Western Province is indicated by a black star.

Fig. S2. Comparison of deduced conserved sequence elements (CSEs) of alphaviruses

Alignment of the deduced nucleotide sequences of CSEs were analyzed with the Mwinilunga alphavirus and selected alphavirus sequences using MEGA, version 6.0. Alignment of CSE (a) in the 5'-UTR region, (b) in nsP1 51 nt ORF, (c) in subgenomic promoter and (d) in the 3'-UTR region. Nucleotides identical to Mwinilunga alphavirus sequence are shown with black dots.

Fig. S3. Comparison of the protease cleavage sites of alphaviruses

Alignment of the amino acid sequences of protease cleavage sites were analyzed with Mwinilunga alphavirus and selected alphavirus sequences using MEGA, version 6.0. Conserved residues are indicated with black dots.