One Health Outlook

RESEARCH

Open Access

Identification of insect-specific viruses in mosquitoes collected from wildlife and rural areas in north-eastern parts of South Africa using a metagenomic RNA sequencing approach

Sascha Kacnik[1], Caitlin MacIntyre[2], Milehna Guarido[1] and Marietjie Venter[1,2*]

Abstract

Background Next generation sequencing (NGS) has expanded virus detection capabilities beyond the limitations of sequence-specific methods. While mosquito collections in South Africa (SA) have been investigated for the major arbovirus genera (*Alpha-*, *Orthobunya-, and Orthoflavivirus*), research on insect-specific viruses (ISVs) is limited.

Methods Here we used an RNA-sequencing viral-metagenomics approach to investigate arboviruses and ISVs in mosquitoes collected in SA. Ten archived mosquito pools, representing up to 50 mosquitoes per pool, previously tested for the major arbovirus genera were investigated for the presence of additional viruses using NGS. These pools included known arbovirus mosquito vectors from the *Aedes*, *Anopheles*, *Culex*, and *Mansonia* genera, collected from wildlife areas in the Kruger National Park (Mpumalanga/Limpopo) and rural areas of Jozini (KwaZulu-Natal). Extracted RNA was DNA-depleted, subjected to Sequence Independent Single Primer Amplification (SISPA), and sequenced on the Illumina MiSeq platform.

Results Although no mammalian arboviruses were detected, bioinformatic analysis detected several ISVs in pools of *Aedes aegypti*, *Ae.ochraceus*, and *Anopheles squamosus* mosquitoes. These ISVs included the cell fusing agent-, Tesano aedes-, Fako-, formosus-, verdadero-, Aedes partiti-like virus 1 and the Kwale mosquito virus and several novel ISVs within the *Iflavivirus* genus as well as some unclassified viruses.

Conclusion This study confirms the value of sequence independent metagenomics approaches to detect novel viruses in mosquito pools. Such studies provide insight into viral diversity and mosquito-virus interactions, supporting the future evaluation of ISVs as biological control agents. This in turn can contribute to protecting the health of animals and humans within a One Health approach.

Keywords South Africa, Mosquitoes, Viral-metagenomics, Insect-specific viruses

*Correspondence:
Marietjie Venter
marietjie.venter1@wits.ac.za
[1]Centre for Emerging and Reemerging Arbo and Respiratory Virus Research (CEARV), Department of Medical Virology, University of Pretoria, Pretoria, South Africa
[2]Emerging Viral Threats, One Health Surveillance and Vaccines (EViTOH) Division, Infectious Disease and Oncology Research Institute (IDORI), University of Witwatersrand, Johannesburg, South Africa

Background

With one in six diseases being transmitted by mosquitoes, these insects are often considered the deadliest animals on the planet, indirectly causing the largest number of human deaths annually [1–3]. Mosquitoes are vectors for arboviruses of public health significance including dengue (DENV), Zika (ZIKV), and yellow fever virus (YFV). Transmission occurs when female mosquitoes take blood meals, and since many mosquito species feed on a variety of hosts, they provide opportunities for zoonotic spillover. The increase in emerging and re-emerging infectious diseases in recent decades has highlighted the importance of monitoring zoonotic pathogens [4, 5].

Knowing which viruses are circulating in mosquitoes is critical for preparing against potential outbreaks that could otherwise devastate vulnerable populations. Arbovirus detection relies on molecular techniques such as PCR, virus isolation, or antigen detection methods; however, these techniques require previous knowledge of target viruses and may miss unknown viruses [6, 7]. In contrast, next generation sequencing (NGS) offers the potential of a more comprehensive virome analysis. The increased utilization of metagenomic sequencing has significantly increased the number of mosquito-borne viruses identified from mosquito vectors without prior knowledge of specific viruses or mosquito vectors [3, 7]. Additionally, it has also provided insight into the virome of several mosquito hosts which has revealed a diverse range of invertebrate RNA viruses [3].

Since the characterization of mosquito viromes, several, novel insect-specific viruses (ISVs) have also been detected the first being, cell fusing agent virus (CFAV) [8, 9]. Although closely related to arboviruses, these viruses are not known to be pathogenic in vertebrates and are unable to infect vertebrate cells [10]. The role of many of these ISVs is uncertain, however several studies have demonstrated the effect that CFAV has on certain arboviruses, such as DENV in which CFAV has been shown to decrease it's dissemination in *Aedes aegypti* mosquitoes [11, 12]. These interactions between CFAV and DENV highlights the potential of ISVs to be used in biocontrol in influencing vector competence, emphasizing the role they could have in reducing the spread of medically-important arboviruses [11].

Researchers in the Centre for Emerging and Re-Emerging Arbo- and Respiratory virus (CEARV), University of Pretoria, previously conducted vector surveillance in the north-eastern parts of South Africa (SA) to describe the mosquito density and distribution [13, 14] across areas where outbreaks of arboviruses in the *Alpha-*, *Orthobunya-, and Orthoflavivirus* genera occurred in animals [15–20] and humans [21–24]. Using genus-specific PCRs, we described potential mosquito vectors for alphaviruses (Sindbis [SINV] and Middelburg virus [MIDV]) [25], orthobunyaviruses (Shuni virus [SHUV]) [26], and orthoflaviviruses (West Nile virus [WNV] and Banzi virus [BANV]) [27]. We also identified insect specific alpha [28] and orthoflaviviruses [29].

In order to determine if metagenomic approaches could detect additional viruses in the same mosquito pools, not identified through conventional genus-specific PCR screening, we selected known arbovirus mosquito vector species collected in wildlife and rural areas in the north-eastern parts of SA with high mosquito species richness. In this study, we tested ten mosquito pools of different mosquito genera (*Aedes*, *Anopheles, Culex*, and *Mansonia*) using Sequence Independent Single Primer Amplification (SISPA), a non-specific metagenomic approach, on DNA-depleted RNA followed by Illumina sequencing. Through this approach we identified several partial and full genomes of ISVs, not previously reported in SA as well as novel viruses. These viruses should be investigated further for their interaction with known arboviruses, mosquito vectors, and potential to infect mammalian cells and control programs.

Methods

Study design

Between 2014 and 2018, the CEARV maintained a mosquito surveillance program for arboviruses at several sites in the north-eastern region of SA, as previously described [13, 25–27, 30, 31]. These collections have previously been screened for the presence of alpha-, orthobunya (Simbu serogroup), and orthoflaviviruses, as described [13, 25–27, 29–31]. To explore if we could detect additional viruses through metagenomic sequencing approaches, we selected ten archived mosquito pools from the *Aedes, Anopheles, Culex*, and *Mansonia* genera, collected in the arboviral season (January to June) in wildlife areas in the Kruger National Park (KNP) in the Mpumalanga/Limpopo province, and rural areas in Jozini, in northern KwaZulu-Natal in 2017. One of these pools previously tested positive for CFAV, while the remaining nine tested negative for the presence of alpha-, orthobunya-, and orthoflaviviruses (Table 1).

Sample collection

Mosquito trapping was performed as previously described [13, 26]. In short, mosquito trapping was done between 16:00 pm to 8:00 am for one to three consecutive nights per month for each site using several different carbon dioxide (CO_2)-baited traps including net traps, Centres for Disease Control and Prevention (CDC) miniature light traps (Centres for Disease Control and Prevention, United States of America (USA)), and BioGents (BG)-sentinel traps (BioGents Corporation, Germany). Mosquitoes which were collected from the traps were euthanized by placing them on dry ice and subsequently

Table 1 Mosquito pools selected for metagenomic investigations collected in SA during the arbovirus season in 2017

Sample	location	Site	# mosquitoes in pool	Month of collection	Morphological identification of mosquitoes	Previous RT-PCR result
SHI17MP525	Shingwedzi rest camp	KNP	50	February	*Ae. aegypti*	Cell fusing agent virus
KNP7MP624	Malelane rest camp	KNP	33	March	*Ae. aegypti*	Negative
SHI17MP541	Shingwedzi rest camp	KNP	29	February	*Ae. ochraceus*	Negative
KNP17MP631	Malelane rest camp	KNP	27	March	*Ae. mcintoshi*	Negative
KNP17MP632	Malelane rest camp	KNP	31	March	*Ae. mcintoshi*	Negative
KNP17MP70	Satara rest camp	KNP	29	March	*An. squamosus*	Negative
KNP17MP688	Punda Maria rest camp	KNP	25	March	*Cx. theileri*	Negative
KZN17MP131	Mamfene	KZN	50	March	*Ae. durbanensis*	Negative
KZN17MP132	Mamfene	KZN	50	March	*Ae. durbanensis*	Negative
KZN17MP197	Mamfene	KZN	50	April	*Ma. uniformis*	Negative

SHI = Shingwedzi, KNP = Kruger National Park, KZN = Jozini (KwaZulu Natal), *Ae* = *Aegypti*, *An* = *Anopheles*, *Cx* = *Culex*, *Ma* = *Mansonia*

stored at −80 °C. Mosquitoes had previously been identified by entomologists in the CEARV using a stereomicroscope (Nikon SMZ 74ST, Nikon, SA) based on morphological characteristics using the dichotomous keys as a guide provided by Jupp (1996) [32], Huang (2001) [33], and Huang et al. (2014) [34] and were pooled according to species and sex at a maximum of 50 mosquitoes per pool.

RNA extraction

Mosquito pools were homogenized as previously described [13, 26]. Total RNA extraction was performed on the resulting homogenate using the Direct-zol RNA Miniprep kit (Zymo Research, Irvine, USA), including the DNase step as per the manufacturer's instructions. The RNA was purified using the RNA Clean and Concentrator^{TM}-5 kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions.

SISPA and library preparation

The methods used here were adapted from those previously described [35, 36]. The master mix for strand denaturation consisted of 2.00 µL of primer FR-26-RV-N (10 µM) (5'-GCCGGAGCTCTGCAGATATCNNNNNN-3), and 0.50 µL of Dimethyl sulfoxide (DMSO) (50%) (Sigma-Aldrich, St Louis, USA). The master mix was aliquoted, 2.50 µL per sample, and 7.50 µL of the sample RNA was added obtaining a reaction volume of 10.00 µL. The samples were incubated (MiniAmp Thermal cycler, Applied Biosystems, Thermo Fisher Scientific, Waltham, USA) at 65 °C for 5 minutes and 4 °C for 2 minutes and placed on ice. The second master mix for first strand cDNA synthesis using the SuperScript III First Strand Synthesis kit (Invitrogen, Thermo Fischer Scientific, Waltham, USA) was made according to the manufacturer's instructions. The reactions were incubated for 10 minutes at 25 °C, 50 minutes at 50 °C, 10 minutes at 85 °C and held at 4 °C.

For Klenow second strand synthesis the master mix was made as follows: 1.00 µL of RNAse H (Invitrogen, Thermo Fischer Scientific, Waltham, USA) and 1.00 µL of Klenow 3'-5'exo DNA polymerase (Invitrogen, Thermo Fischer Scientific, Waltham, USA). A volume of 2.00 µL was added to each reaction and incubated at 37 °C for 60 minutes, 75 °C for 15 minutes, and held at 4 °C.

The Qiagen MinElute PCR purification kit (Qiagen, Hilden, Germany) was used for cDNA purification, per the manufacturer's instructions.

The master mix for random fragment amplification was made as follows: 25.00 µL of MyTaq red 2X mix (Meridian Bioscience, Cincinnati, USA), 2.00 µL of primer FR-20-RV (10 µM) (5'- GCCGGAGCTCTGCAGATATC-3') and 13.00 µL of nuclease-free water (Sigma-Aldrich, St Louis, USA). The master mix was aliquoted, 40.00 µL per reaction together with 10.00 µL of cDNA template. The samples were incubated at 98 °C for 30 seconds, (98 °C for 10 seconds, 54 °C for 20 seconds, 72 °C for 45 seconds) x40, 72 °C for 5 minutes, and hold at 4 °C.

The second amplification master mix was made as follows: 12.50 µL of MyTaq red 2X mix (Meridian Bioscience, Cincinnati, USA), 1.00 µL of primer FR-20-RV (10 µM), and 13.50 µL of nuclease-free water (Sigma-Aldrich, St Louis, USA). The master mix was aliquoted, 27.00 µL to each sample. The samples were incubated at 98 °C for 30 seconds, (98 °C for 10 seconds, 54 °C for 20 seconds, 72 °C for 45 seconds)x5 and held at 4 °C. The Qiagen MinElute PCR purification kit was used for cDNA purification as described above. Each sample was eluted with 12.00 µL of elution buffer.

DNA concentration and purity was determined using the Qubit 1X dsDNA HS Assay Kit on a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions. A concentration of > 4.00 ng/µL was needed to proceed with library preparation.

Sequencing libraries were prepared using the Illumina DNA Prep kit (Illumina, San Diego, USA) as per the manufacturer's instructions. A final concentration of > 4 ng/µl was required. The prepared libraries were sequenced on the Illumina MiSeq platform (Illumina, San Diego, USA).

Bioinformatics and quality control

The Illumina MiSeq system demultiplexed sequencing reads and removed adaptor sequences. Paired-end sequencing reads were trimmed using the default settings and quality control for sequencing reads was done on CLC Genomics workbench (CLC Bio, Aarhus, Denmark). The Q30 score, which represents the number of reads which had a base call accuracy of at least 99.90%, was recorded as a percentage. Galaxy (available at: https://usegalaxy.org/) was used for further downstream analysis. Host sequences were removed using Bowtie2 [37, 38]. If available, Galaxy's built-in host genome indexes were used, alternatively host sequences were obtained for corresponding mosquito species from Ensembl Metazoa (available at: https://metazoa.ensembl.org/index.html) [39]. Unaligned paired reads were assembled into contigs using MEGAHIT [40]. Subsequently, CD-HIT was used to cluster contigs and remove any duplicate sequences based on a 90.00% similarity [41]. The resulting contigs were submitted to the National Centre for Biotechnology Information (NCBI) nucleotide Basic Local Alignment Search Tool (BLASTn) Virus database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). BLAST hits with an E value of ≤ $10e^{-5}$ were considered significant [6]. Bacterial and eukaryotic sequences were eliminated, only viral sequences were analysed.

Phylogenetics

After BLAST analysis, the returned sequences were aligned using the Multiple Alignment using Fast Fourier Transform (MAFFT) program (available at: https://mafft.cbrc.jp/alignment/software/) [42] using the default parameters and edited using BioEdit v7.0.5.3 [43]. Reference sequences were downloaded from GenBank (available at: https://www.ncbi.nlm.nih.gov/genbank/) [44]. Upon alignment and editing, the sequences underwent phylogenetic analysis via the Molecular Evolutionary Genetics Analysis (MEGA X) program [45]. The best fit model test analysis was performed using the default parameters on MEGA X and was applied to the appropriate dataset. Subsequently, a maximum likelihood tree with a bootstrap analysis of 1000 repetitions was generated. Pairwise-distance indicating the percent nucleotide and amino acid similarity was generated using MEGA X.

Molecular barcoding of positive mosquito pools

Molecular barcoding was performed as previously described [29]. Mitochondrial DNA was extracted from positive mosquito homogenates using the Qiagen DNeasy blood and tissue kit (Qiagen, Germany) according to the manufacturer's instructions, specifically for insects. The molecular barcoding PCR was performed as previously described [28] using primers targeting the mitochondrial Cytochrome c oxidase subunit I (COX1) gene of mosquitoes [46].

Results

A total of ten mosquito pools were selected for NGS sequencing previously collected from the KNP and Jozini during the arbovirus season in 2017. These pools included *Ae.aegypti* (2 pools), *Ae.ochraceus* (1 pool), *Ae. mcintoshi* (2), *Ae.durbanensis* (2 pools), *An.squamosus* (1 pool), *Cx.theileri* (1 pool), and *Ma.uniformis* (1 pool) (Table 1).

All samples met quality control requirements for library preparation (Table 2). Following Illumina sequencing, all the samples had a Q30 score over 80.00%, passing quality control (Table 2). Viral reads were detected in 40.00% (4/10) of the pools using BLASTn. All viruses detected were ISVs and belonged to the *Flavi-*, *Iflavi- Spinareo-*, *Rhabdo-*, and *Partitiviridae* families (Table 2). Several unclassified viruses were also detected. Phylogenetic analysis of each detected virus is described below. Accession numbers can be seen in the supplementary material (Supplementary material 1, Table 1).

Molecular barcoding was performed for pools in which viruses were detected. CFAV was previously detected in SHI17MP525, therefore molecular barcoding had previously been done, confirming the morphological identification as *Ae.aegypti* (GenBank accession number: http://MW077855) [29]. Unfortunately, molecular barcoding was unsuccessful for the remaining three pools (KNP17MP624, KNP17MP70, and SHI17MP541). This could potentially be due to the degradation or low concentration of the respective target regions.

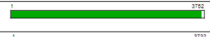
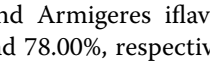
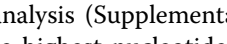
Flaviviridae

CFAV was detected in one pool of *Ae.aegypti* mosquitoes (SHI17MP525). Phylogenetic analysis (Fig. 1) showed that the CFAV identified here (designated ZA(Ar)SHI17MP525) clustered most closely with a strain isolated from *Ae.aegypti* mosquitoes from Zambia (LC770212.1) (bootstrap 96.00%). P-distance analysis (Supplementary material 1, Table 2) indicated 98.63% nucleotide identity and 96.71% amino acid identity to the Zambian strain. Due to the significant number of ambiguous bases ("N") across the partial genome detected here, it was not submitted to GenBank.

Iflaviviridae

Three Iflaviviruses were identified in the two *Ae.aegypti* pools collected in the KNP: two detections in pool

Table 2 Insect-specific viruses identified in mosquito samples through metagenomic sequencing

Mosquito pool	Species	Concentration following NGS library preparation (ng/µl)	# of trimmed reads	Q30 score (%)	Closest viral hit (Family, Genus)		Depth of coverage (X)	Total Sequence length obtained (nt)	Genome coverage (%)	Visual representation of the genome coverage (%)
SHI17MP525	*Ae. aegypti*	28.80	782416	84.83	Cell fusing agent virus (*Flaviviridae, Orthoflavivirus*)		1.19	4677	44.00 %	
					Armigeres iflavivirus (*Iflaviviridae, Iflavivirus*)		14.11	7660	83.90 %	
					Tesano aedes virus (*Iflaviviridae*, unclassified)		68.18	9502	99.00 %	
					Fako virus (*Spinareoviridae, Dinovernavirus*)	Segment 1	9.40	1702	63.00 %	
						Segment 2	1.72	1049	50.00 %	
						Segment 3	2.73	1365	65.00 %	
						Segment 4	7.01	1753	65.00 %	
						Segment 5	3.37	1993	74.00 %	
						Segment 6	4.77	1288	81.00 %	
					Formosus virus (*Rhabdoviridae*, unclassified)		28.50	12164	100.00 %	
					Kwale mosquito virus (unclassified)	Segment 1	178.37	3228	100.00 %	
						Segment 2	22.36	1054	84.00 %	
KNP17MP624	*Ae. aegypti*	30.00	984532	89.00	Tesano aedes virus (*Iflaviviridae*, unclassified)		29.53	9440	99.00 %	
					Fako virus (*Spinareoviridae, Dinovernavirus*)	Segment 1	13.43.	2569	84.00 %	
						Segment 2	18.42	3769	99.00 %	
						Segment 3	6.41	3102	92.00 %	
						Segment 4	7.40	3144	89.00 %	
						Segment 5	6.26	3417	99.00 %	
						Segment 6	7.87	1623	90.00 %	
						Segment 7	0.67	614	45.00 %	
						Segment 8	5.51	970	95.00 %	
						Segment 9	3.55	773	85.00 %	
					Formosus virus (*Rhabdoviridae*, unclassified)		3.83	5988	49.00 %	
					Kwale mosquito virus (Riboviria realm)	Segment 1	12.23	2775	93.00 %	
						Segment 2	1.51	633	40.00 %	
					Verdadero virus (*Partitiviridae*, unclassified)	Segment 1	1.15	736	45.00 %	
						Segment 2	24.73	1494	100.00 %	
					Aedes partiti-like virus 1 (*Partitiviridae*, unclassified)	Segment 1	62.36	1477	99.00 %	
						Segment 2	94.67	1273	97.00 %	
SHI17MP541		48.60	606432	95.09		Segment 1	92.39	1673	61.00 %	
	Ae. ochraceus				Kwale-/Guadeloupe mosquito virus (Riboviria realm)	Segment 2	0.46	455	29.00 %	
KNP17MP70	*An. squamosus*	55.20	989048	88.46	Hubei mosquito virus 1 (Riboviria realm)		68.90	9075	94.40%	
KNP17MP631	*Ae. mcintoshi*	185.00	703126	97.21	No significant viral reads					
KNP17MP632	*Ae. mcintoshi*	134.00	613060	97.43	No significant viral reads					
KNP17MP688	*Cx.theileri*	44.10	759396	97.07	No significant viral reads					
KZN17MP131	*Ae. durbanensis*	28.50	608218	94.70	No significant viral reads					
KZN17MP132	*Ae. durbanensis*	34.40	522700	95.84	No significant viral reads					
KZN17MP197	*Ma. uniformis*	53.00	518820	97.95	No significant viral reads					

SHI17MP525, referred to as ZA(Ar)SHI17MP525.1/.2 and one detection in pool KNP17MP624, referred to as ZA(Ar)KNP17MP624. The phylogenetic tree in Figure 2 indicates that ZA(Ar)SHI17MP525.1 groups with the Sassandra and Armigeres iflavivirus, with a bootstrap of 71.00% and 78.00%, respectively, but in its own clade. P-distance analysis (Supplementary material 1, Table 3) indicated the highest nucleotide identity of 70.25% and

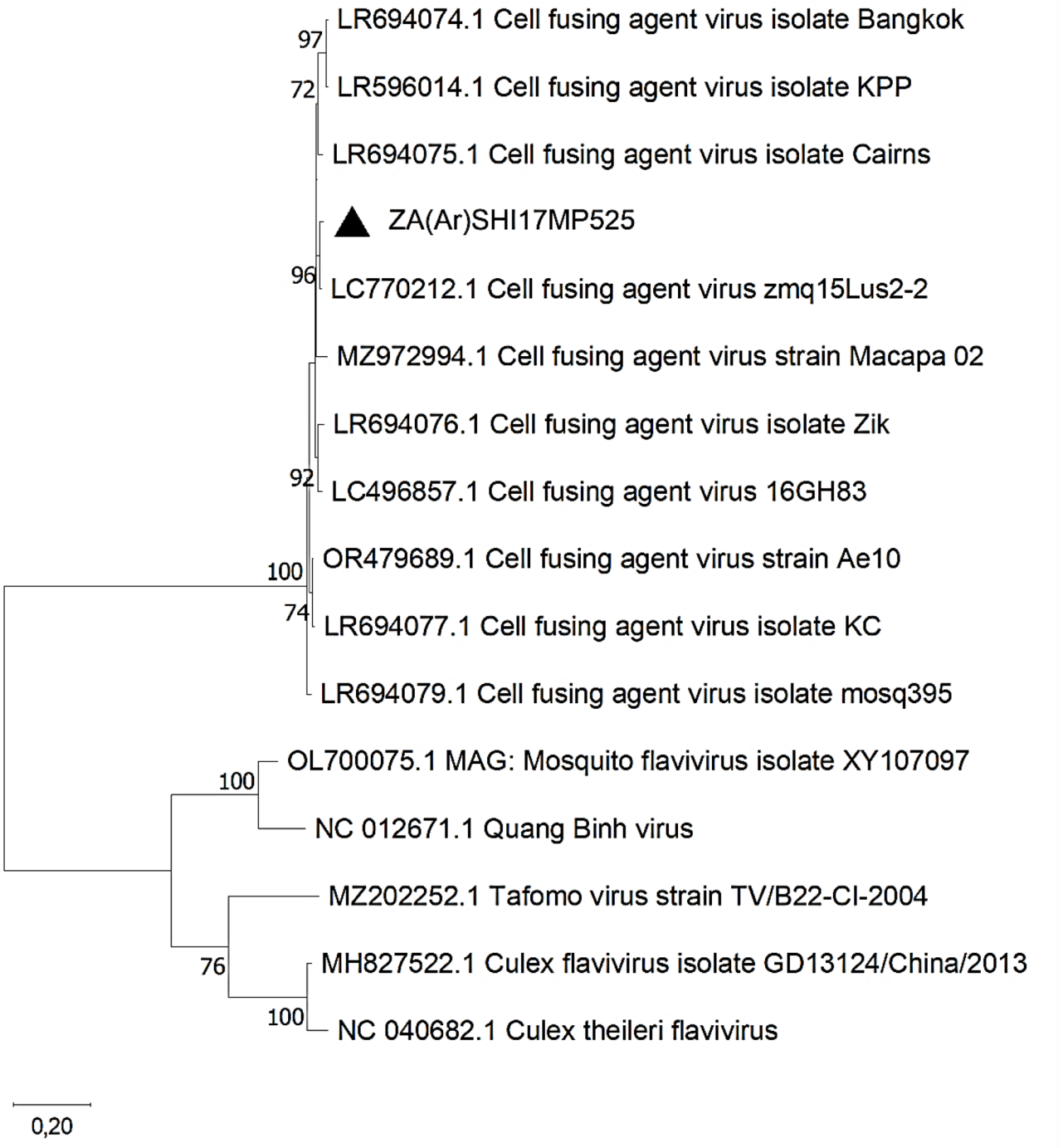


Fig. 1 Maximum likelihood phylogenetic analysis of cell fusing agent virus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model [47]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support > 70.00% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 1439 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle

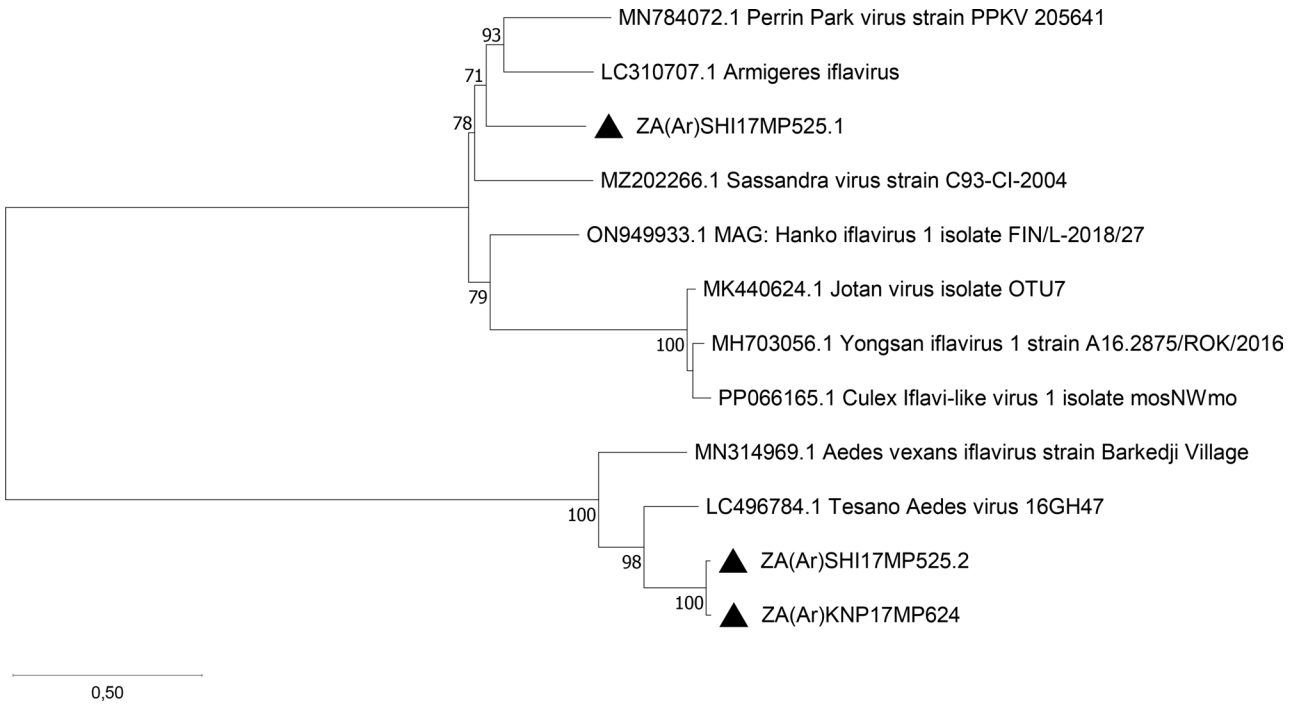


Fig. 2 Maximum likelihood phylogenetic tree for iflavivirus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model [48]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the values above 70% are shown next to the branches. Phylogenetic groupings with nodal support > 70% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 8896 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses: PV776881 (ZA(Ar)SHI17MP525.1), PV658503 (ZA(Ar)SHI17MP525.2), and PV658504 (ZA(Ar)KNP17MP624)

amino acid identity of 72.84% with Armigeres iflavivirus. The two other isolates, ZA(Ar)SHI17MP525.2 and ZA(Ar)KNP17MP624 grouped with the Tesano aedes virus (TeAV) with a bootstrap of 98%. These two samples shared a nucleotide identity of 78.40% and an amino acid identity of 94.00% with TeAV.

Spinareoviridae

The Fako virus (FAKV) was identified in the two *Ae. aegypti* pools, referred to as ZA(Ar)SHI17MP525 and ZA(Ar)KNP17MP624. The FAKV has a segmented genome consisting of nine genome segments. Nine segments were identified in pool KNP17MP624, however, only six were identified in SHI17MP525. Individual phylogenetic trees were drawn for each segment. Segments 1–6, with ZA(Ar)SHI17MP525 and ZA(Ar)KNP17MP624, are displayed in Fig. 3. Segments 7–9 are displayed in Fig. 4. Similar clustering patterns are observed across the segments in which the two isolates identified here group with the FAKV isolate KLF FAKV 21, which was originally isolated from *Ae.aegypti* mosquitoes in Kenya [49]. P-distance analysis (Supplementary material 1, Table 4A to 4I) supported the trees displayed in Figs. 3 and 4. The highest identities were between the samples or with FAKV isolate KLF FAKV 21.

Rhabdoviridae

The formosus virus (FORMV) was detected in both *Ae. aegypti* pools, SHI17MP525 and KNP17MP624, referred to as ZA(Ar)SHI17MP525 and ZA(Ar)KNP17MP624. As seen in Fig. 5, isolates ZA(Ar)SHI17MP525 and ZA(Ar)KNP17MP624 group together with FORMV with a bootstrap of 100.00%. The FORMV was identified in laboratory colonies established from wild *Ae. aegypti* mosquitoes from Bundibugyo, Uganda [51]. Between the samples and FORMV, the nucleotide identity ranged between 92.10% and 92.62%. P-distance analysis for amino acid identity for each protein encoded can be seen in the Supplementary material (Supplementary material 1, Tables 5A to 5F). The amino acid identity between the samples and representative FORMV sequences were all above 90.00% (91.04%-99.41%).

Partitiviridae

Two partitiviruses were identified in KNP17MP624 made up of *Ae.aegypti* mosquitoes, referred to as ZA(Ar)KNP17MP624.1 and ZA(Ar)KNP17MP624.2. Phylogenetic analysis was possible for Segment 1, however, due to a lack of sequences for segment 2 available on GenBank, phylogenetic analysis was not possible. As seen in Fig. 6, KNP17MP624.1 groups with the verdadero virus

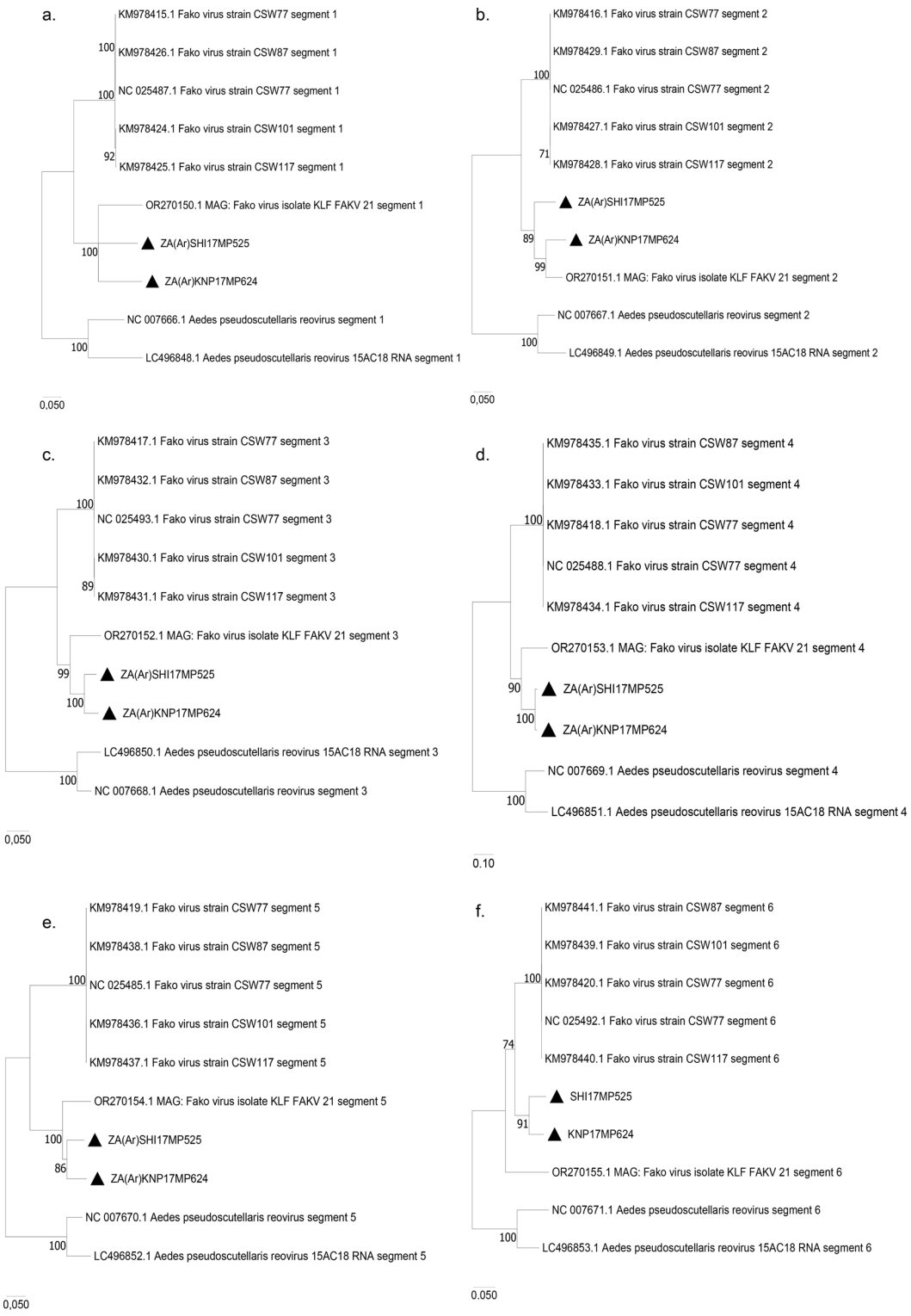


Fig. 3 Maximum likelihood phylogenetic trees for Fako virus segments 1 to 6. The phylogenetic trees for genome segments 1–6 are represented by Fig. 3**a**–**f**, respectively. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model [50]. The trees with the highest log likelihood are shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support >70.00% are considered statistically significant. The trees were drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 3828, 3711, 3685, 1758, 3192, and 1735 positions in the final dataset for **a**–**f**, respectively. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses (in order of segment 1 to 6): ZA(Ar)SHI17MP525 (PV730236, PV730253, PV730235, PV730239, PV730237, PV730238), and ZA(Ar)KNP17MP624 (PV730240- PV730245)

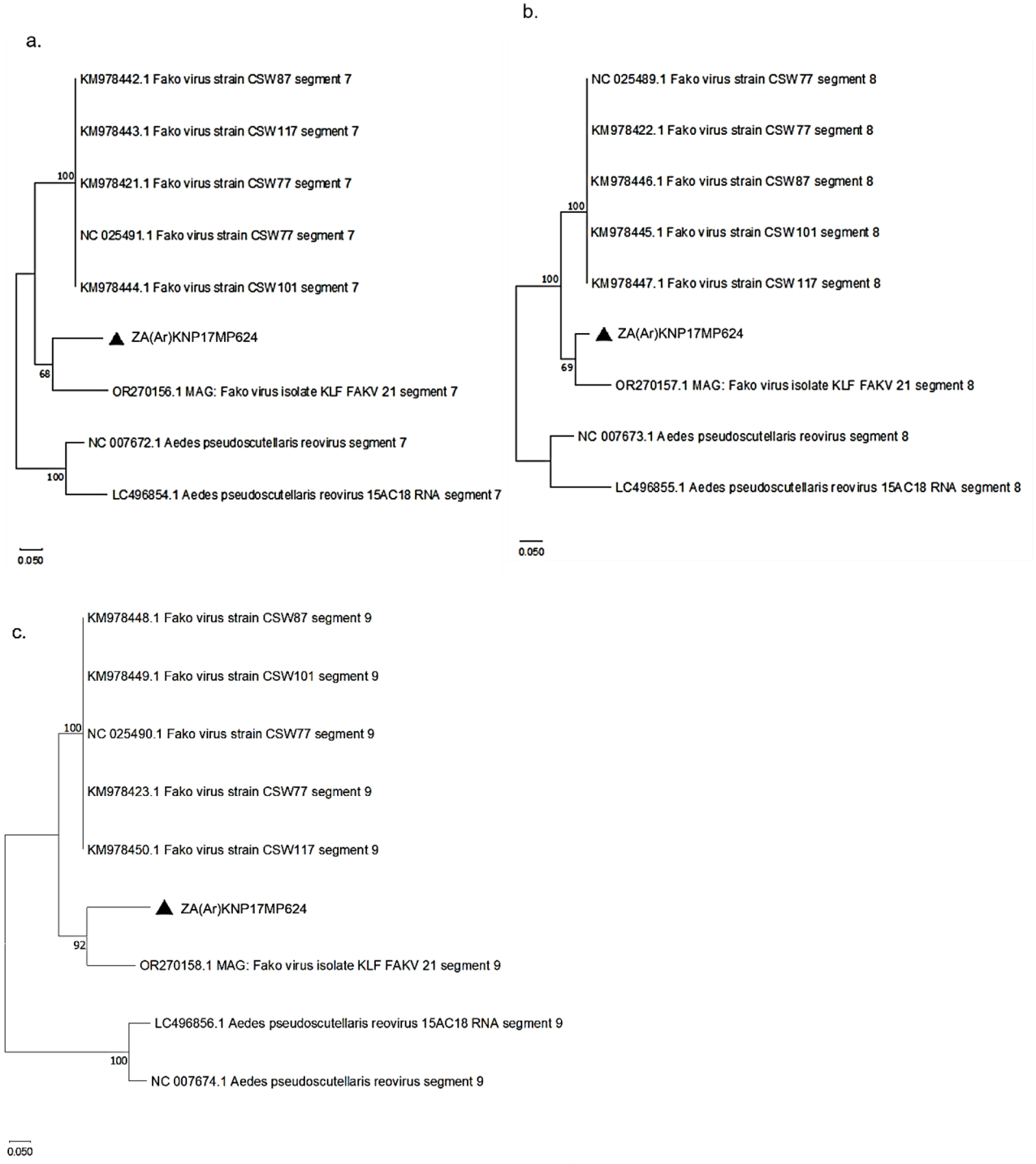


Fig. 4 Maximum likelihood phylogenetic trees for Fako virus segments 7 to 9. The phylogenetic trees for genome segments 7–9 are represented by Fig. 4**a**–**c**, respectively. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model [50]. The trees with the highest log likelihood are shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support > 70.00% are considered statistically significant. The trees were drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 432, 573, and 636 positions in the final dataset for **a**–**c**, respectively. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses (in order of segment 7 to 9): ZA(Ar)KNP17MP624 (PV730252, PV730246, PV730247)

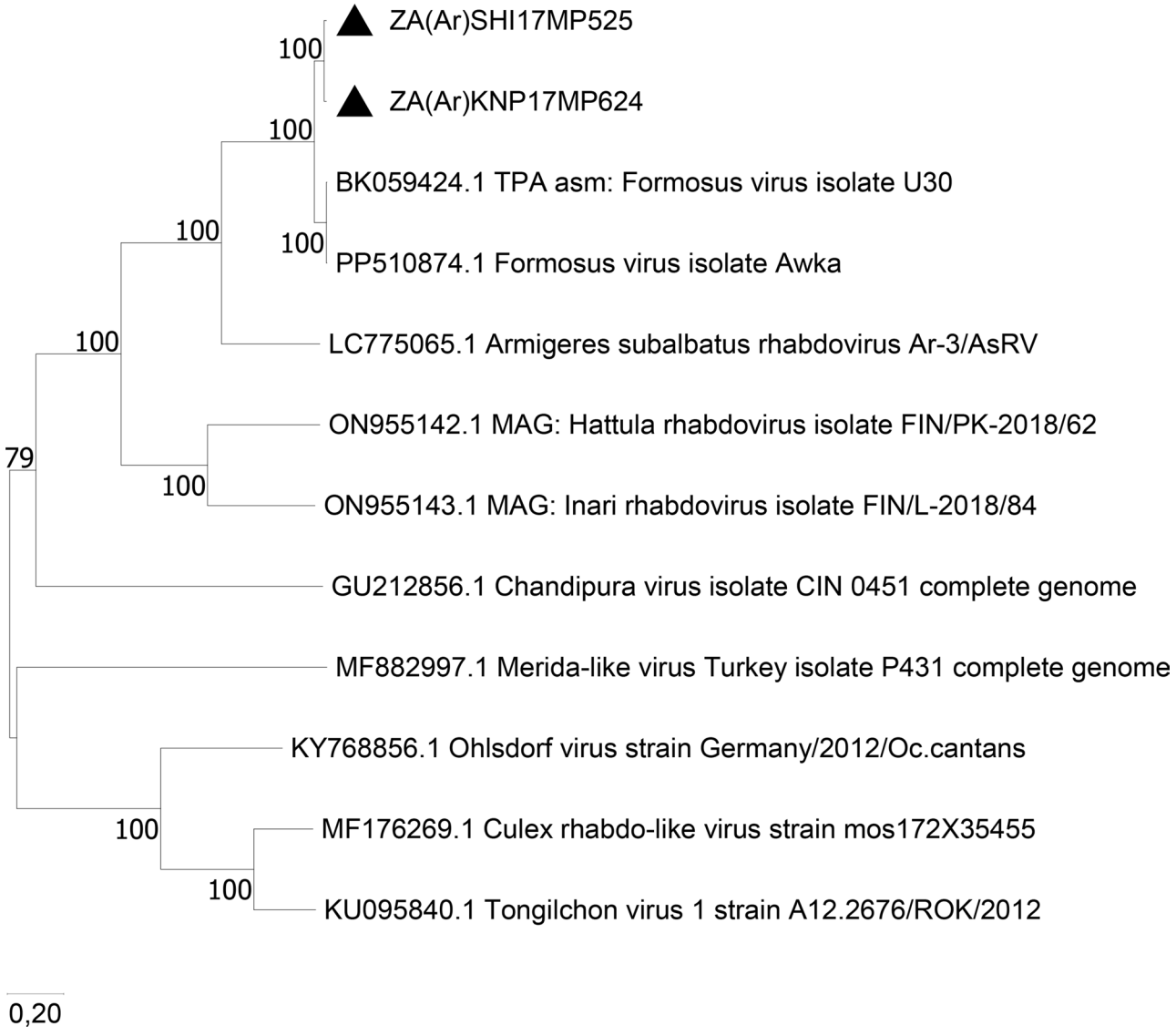


Fig. 5 Maximum likelihood phylogenetic tree for rhabdovirus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model [48]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support > 70.00% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site there were a total of 14614 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses: PV730233 (ZA(Ar)SHI17MP525) and PV730234 (ZA(Ar) KNP17MP624)

(VdV) (bootstrap of 99.00%), whereas KNP17MP624.2 groups with Aedes partiti-like virus 1 (APLV-1) (bootstrap < 70.00%). The VdV was originally isolated from a colony of *Ae.aegypti* mosquitoes in Mexico [52]. The same study that identified the FORMV, identified the APLV-1 in *Ae.aegypti* colonies located worldwide [51]. Segment 1 P-distance analysis (Supplementary material 1, Table 6A) supported the tree in Fig. 6. For ZA(Ar) KNP17MP624.1, P-distance analysis indicated the highest nucleotide identity of 97.88% and amino acid identity of 97.83% with VdV. For KNP17MP624.2, P-distance analysis indicated the highest nucleotide identity of 97.49% to 99.37% and amino acid identity of 100.00% with APLV-1.

P-distance analysis (Supplementary material 1, Table 6B) for segment 2 indicated that ZA(Ar)KNP17MP624.1 (GenBank: PV730256) had a nucleotide identity of 98.63% and amino acid identity of 100% with VdV. KNP17MP624.2 (GenBank: PV730258) had a nucleotide identity of 96.72% to 97.79% and amino acid identity of 97.03% to 98.29% with APLV-1.

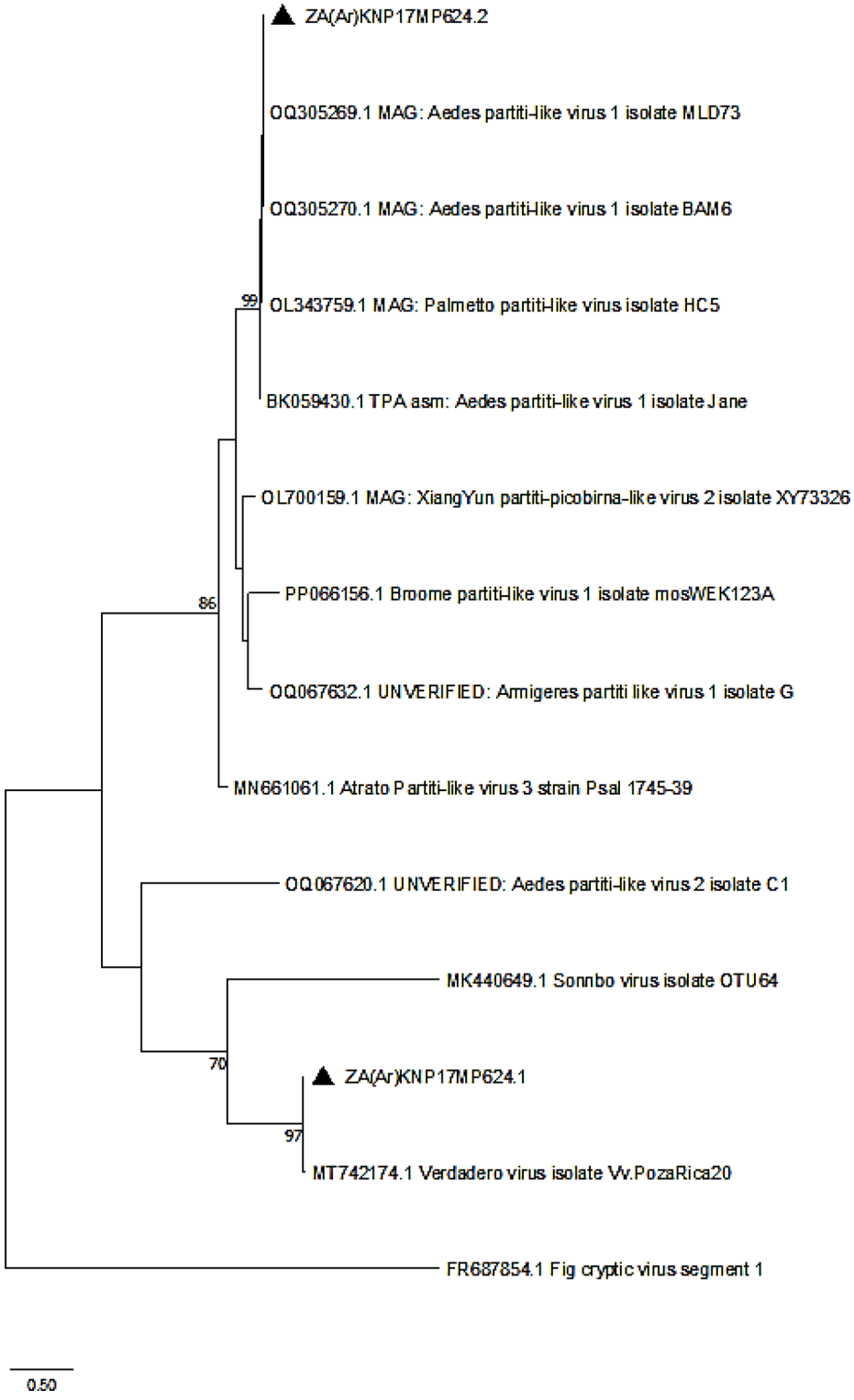


Fig. 6 Maximum likelihood phylogenetic tree for partitivirus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model [47]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support > 70.00% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 344 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses: PV730255 (ZA(Ar)SHI17MP525) and PV730257 (ZA(Ar) KNP17MP624)

Unclassified viruses

Kwale-/Guadeloupe mosquito-like viruses

The Kwale mosquito virus (KMV) was identified in sample SHI17MP525 (ZA(Ar)SHI17MP525) (*Ae.aegypti*) and KNP17MP624 (ZA(Ar)KNP17MP624) (*Ae.aegypti*), and a virus similar to the KMV and Guadeloupe mosquito virus (GMV) was detected in SHI17MP541 (ZA(Ar) SHI17MP541) (*Ae.ochraceus*). Both the KMV and GMV, are unclassified viruses in the *Riboviria* realm, and were isolated in *Ae.aegypti* mosquitoes from Kenya and Guadeloupe, respectively [53, 54]. Phylogenetic trees were drawn for segment 1 (Fig. 7) and segment 2 (Fig. 8). The trees in Figs. 7 and 8 display similar relationships, with ZA(Ar)SHI17MP525 and ZA(Ar)KNP17MP624 grouping most closely with KMV, and ZA(Ar)SHI17MP541 grouping with the KMV and GMV, but in its own separate clade. P-distance analysis for segment 1 (Supplementary material 1, Table 7A) indicated that ZA(Ar)SHI17MP525 had the highest nucleotide identity of 97.19% with KMV. ZA(Ar)KNP17MP624 had a nucleotide identity of 94.38% with the KMV. Nucleotide identities between ZA(Ar) SHI17MP541 and the KMV and GMV ranged between 82.48% and 83.23%, the highest with KMV. P-distance analysis for segment 1 (Supplementary material 1, Table 7B) indicated that ZA(Ar)SHI17MP525 and ZA(Ar) KNP17MP624 had the highest amino acid identity of 98.58% and 95.41% with the KMV for ORF 1, respectively. For ORF 2, both ZA(Ar)SHI17MP525 and ZA(Ar) KNP17MP624 shared the highest amino acid identity of 99.55% with the KMV. ZA(Ar)SHI17MP541 shared the highest amino acid identity of 74.18% with GMV isolate PB-AAF 1–3 for ORF 1, and 93.20% with the KMV and GMV isolates CMS002 017a and 018a for ORF 2.

P-distance analysis for segment 2 (Supplementary material 1, Table 7C) indicated that ZA(Ar)SHI17MP525 had the highest nucleotide identity of 97.37% with ZA(Ar)KNP17MP624, followed by 95.99% with the KMV. ZA(Ar)KNP17MP624 had a nucleotide identity of 93.46% with KMV. The nucleotide identity between ZA(Ar) SHI17MP541 and the KMV and GMV ranged between 79.56% and 81.76%, the highest with the KMV. P-distance analysis for segment 2 (Supplementary material 1, Table

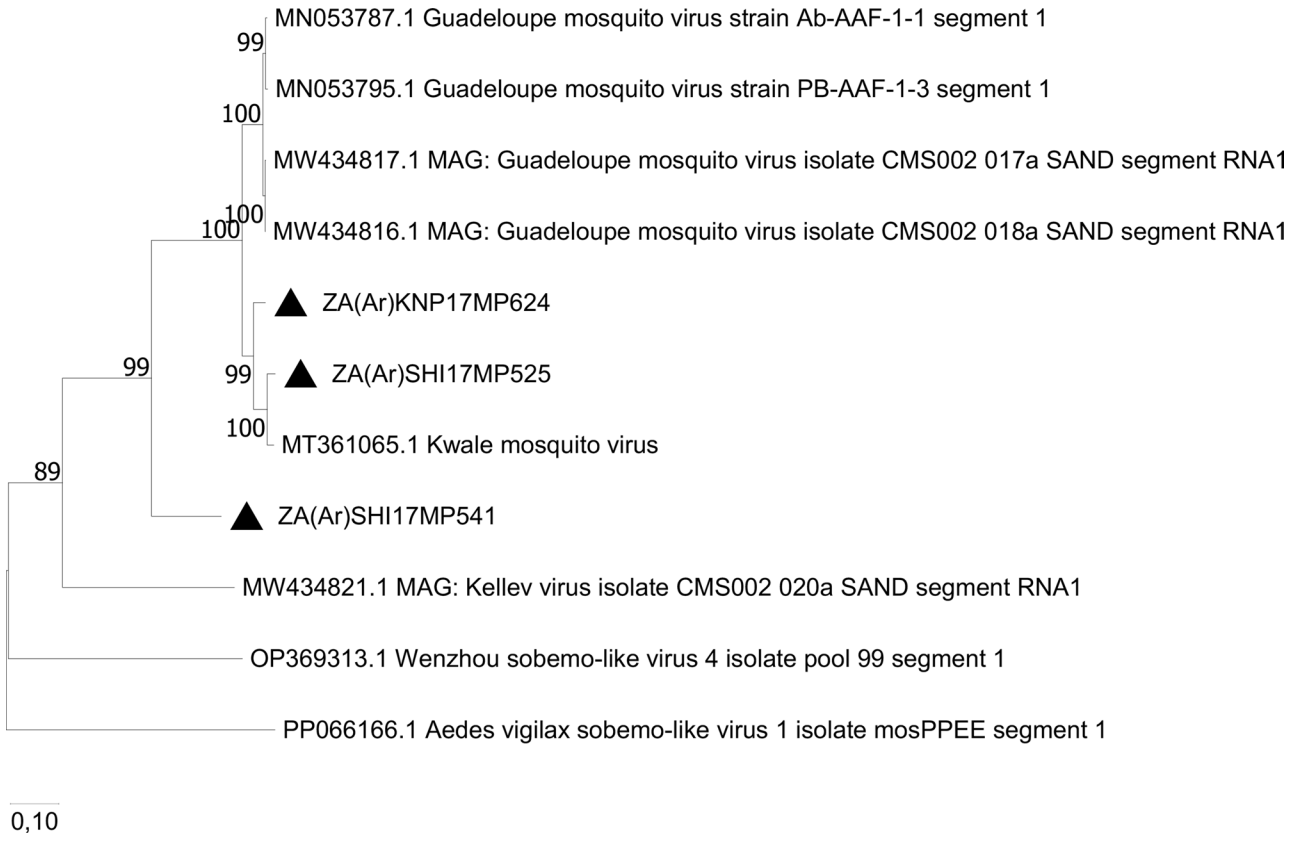


Fig. 7 Maximum likelihood phylogenetic tree for segment 1 of Kwale-/Guadeloupe mosquito like virus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model [50]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support >70.00% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 3156 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses: PV730248 (ZA(Ar) SHI17MP525), PV730250 (ZA(Ar)KNP17MP624), and PV730254 (ZA(Ar)SHI17MP541)

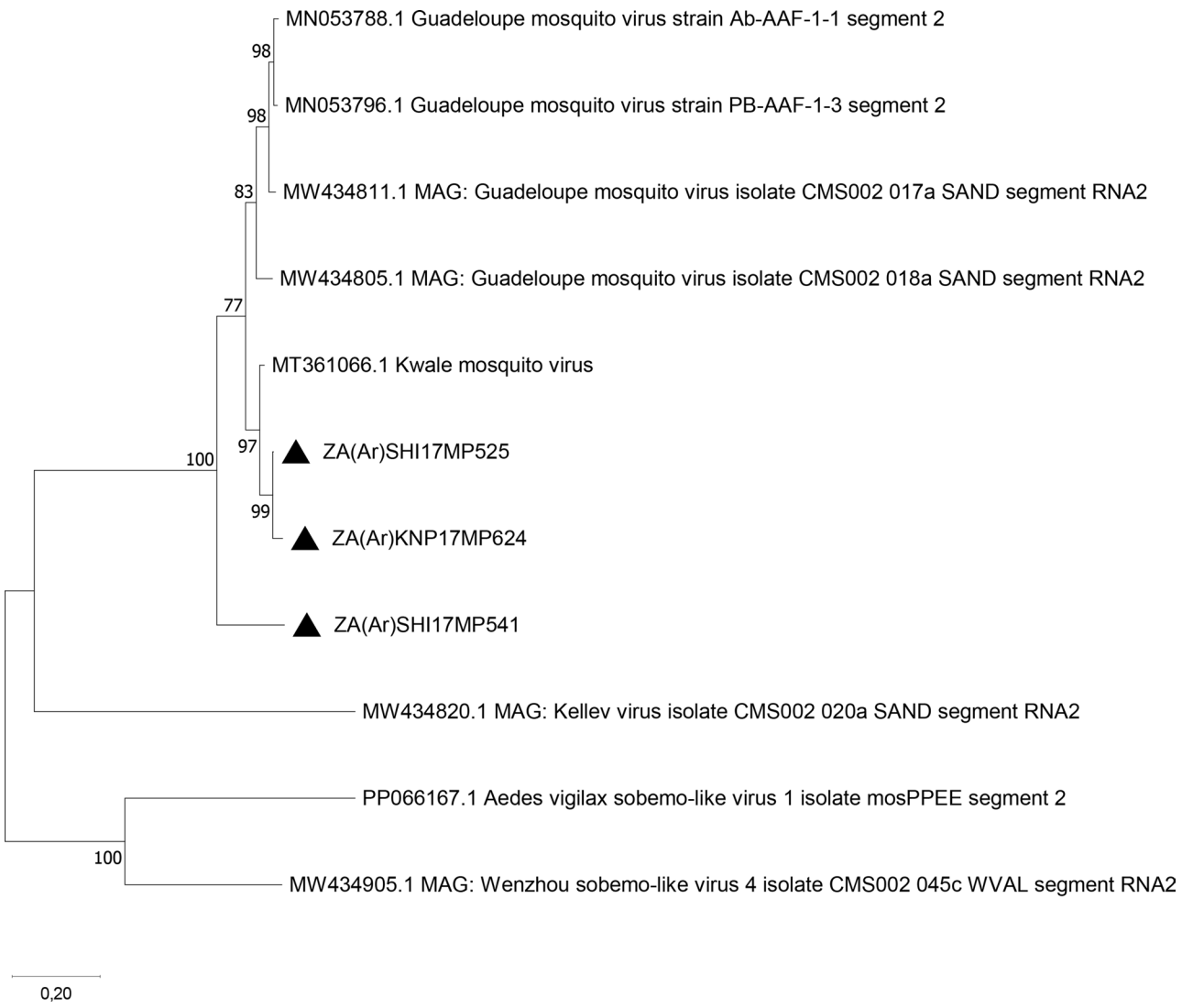


Fig. 8 Maximum likelihood phylogenetic tree for segment 2 of Kwale-/Guadeloupe mosquito like virus. Evolutionary analyses were conducted in MEGA X [42]. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model [50]. The tree with the highest log likelihood is shown. The trees were run under 1000 bootstrap replicates and the bootstrap values above 70.00% are shown next to the branches. Phylogenetic groupings with nodal support > 70.00% are considered statistically significant. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 3156 positions in the final dataset. GenBank accession numbers are indicated for reference sequences. Samples that were a part of this study are marked with a triangle. GenBank accession numbers for newly sequenced viruses: PV730249 (ZA(Ar)SHI17MP525), PV730251 (ZA(Ar)KNP17MP624), and PV776883 (ZA(Ar)SHI17MP541)

7D) indicated that ZA(Ar)SHI17MP525 had the highest amino acid identity of 100.00% and 98.20% with the KMV for ORF 1 and ORF 2, respectively. Unfortunately, the coding sequence for ORF 1 for ZA(Ar)KNP17MP624 was not obtained, however the highest amino acid identity for ORF 2 was 94.06% with ZA(Ar)SHI17MP525 followed by 92.38% with the KMV. ZA(Ar)SHI17MP541 shared the highest amino acid identity of 94.74% with the KMV for ORF 1, and 79.87% with the KMV and GMV isolates CMS002 017a and 018a for ORF 2.

Malelane mosquito virus 1, (novel Hubei mosquito-like virus 1 in Riboviria realm)

According to the BLAST analysis, a virus most similar to the Hubei mosquito virus-1 was identified in KNP17MP70 (referred to as ZA(Ar)KNP17MP70), in a pool of *An.squamosus* mosquitoes. The full length of this virus's genome was obtained, (>9000bp), but the BLAST search only returned 3 hits all of which were smaller areas of the Hubei mosquito-like virus 1. The virus detected here was named, Riboviria sp. isolate Malelane mosquito virus 1 isolate ZA(Ar)KNP17MP70 (GenBank: http://PV776882). The genome structure is indicated in the supplementary material (Supplementary material

1, Fig. 1). However, due to the lack of relevant genome sequences available on GenBank for this virus, a phylogenetic tree could not be drawn. The P-distance matrix (Supplementary material 1, Table 8) illustrates that the nucleotide identity with Hubei mosquito virus-1 was between 70.71% and 70.83%, whereas amino acid identity was between 79.80% and 79.85.

Discussion

Historically, studies looking at arbovirus surveillance in mosquitoes typically used a targeted PCR-based approach, however, several recent studies have utilized a metagenomics approach to identify viruses from mosquitoes in Africa [3, 9, 55, 56]. To explore if we could identify additional viruses in mosquito pools which were previously screened for alpha-, orthobunya (Simbu serogroup)-, and orthoflaviviruses we used a viral metagenomics approach. We selected ten pools of mosquitoes belonging to the *Ae.aegypti, Ae.ochraceus, Ae.mcintoshi, Ae.durbanensis, An.squamosus, Cx.theileri, and Ma.uniformis* species, collected in the KNP (Shingwedzi and Malelane rest camps) and Jozini, representing wildlife and rural sites in warmer and sub-tropical areas, respectively. These mosquito species are commonly found in SA and have previously shown to be arbovirus vectors [25–27, 57–60]. While no known arboviruses were identified, numerous different ISVs, which have not previously been reported in SA were identified, predominantly in *Aedes and Anopheles* species. Viruses were identified in 40% (4/10) of pools: SHI17MP525 (*Ae.aegypti*), KNP17MP624 (*Ae.aegypti*), SHI17MP541 (*Ae.ochraceus*), and KNP17MP70 (*An.squamosus*), all of which were collected from the KNP. Several of these pools had more than one ISV identified. As these pools have previously undergone cycles of freeze-thawing and have been stored for several years, we did not attempt virus isolation in this project, although this will be followed up in future research

Through metagenomic approaches, we were able to obtain 44.00% of the CFAV genome whereas previously, a small region of the NS5 gene was identified through a targeted PCR [29]. This shows that the SISPA method followed by Illumina NGS sequencing on the MiSeq could accurately identify this virus and allowed basic phylogenetic analysis despite missing areas of the genome. The CFAV belongs to the *Orthoflavivirus* genus which contains several medically important arboviruses such as WNV, DENV, and YFV [8, 11]. It was originally isolated from an *Ae.aegypti* cell line after it was found to cause cell fusion of *Ae.albopictus* cells and has since been found in wild mosquito populations, specifically *Ae. aegypti* mosquitoes [8, 11, 61]. CFAV is one of the most well-described ISV, especially its interactions with other medically important arboviruses such as DENV and ZIKV, however, contrasting results have been observed [11, 12]. In the presence of CFAV, replication of DENV was inhibited in *Ae.albopictus* C6/36 cells, whereas it's replication was enhanced in *Ae.aegypti* Aa20 cells [11]. It has been suggested that these contrasting results may be due to the use of different viral strains or cell types [11]. Furthermore, CFAV has also shown to hinder the dissemination of ZIKV and DENV-1 in *Ae.aegypti* mosquitoes [11]. Phylogenetic analysis of the CFAV strain (ZA(Ar) SHI17MP525) identified here in an *Ae.aegypti* mosquito pool, grouped closely with a CFAV strain isolated from *Ae.aegypti* mosquitoes in Zambia. The incidence of CFAV in *Ae. aegypti* in SA should be evaluated to determine if it may have an inhibiting action on virus infection in the local mosquito population.

Iflaviviruses (*Iflaviviridae* family) were detected in two pools collected in the KNP, two in a single pool of *Ae.aegypti* mosquitoes (SHI17MP525) and one in an *Ae.aegypti* pool (KNP17MP624). The *Iflavivirus* genus is made up of RNA viruses that only infect arthropods [62]. The TeAV was identified in both pools (TeAV isolate ZA(Ar)SHI17MP525/ZA(Ar)KNP17MP624). The TeAV was originally found in female and male *Ae. aegypti* and female *Culex* mosquitoes in Ghana and has since also been identified in *Ae.aegypti* mosquitoes from Nigeria [3, 9]. A previous study has shown that TeAV may increase the growth of DENV-1 in a concentration-dependent manner [3, 9]. The second iflavivirus detected in SHI17MP525 (ZA(Ar)SHI17MP525.1) grouped with the Perrin park-, Sassandra-, and Armigeres iflaviviruses, but in its own clade. The Sassandra virus, an unclassified iflavivirus, was initially detected in a pool of *Culex* mosquitoes collected in Cote d'Ivoire, whereas the Armigeres iflavivirus was isolated from *Armigeres* mosquitoes collected in the Philippines [63, 64]. Perrin park virus was identified in mosquito excreta in Australia [7]. This iflavivirus had the highest amino acid identity of 72.84% with the Armigeres iflavivirus and based on the guidelines provided by the International Committee on the Taxonomy of Viruses (ICTV) (amino acid identity >90% represents the same species) [65] based on the partial genome isolated (83.90%), this iflavivirus represents a novel iflavivirus.

The FAKV was originally isolated from *Aedes spp*. and *Eretmapodites spp*. mosquitoes during a surveillance study in the Fako region of Cameroon in 2010 [66]. Placed within the *Reovirales* order, this reovirus is made up of nine segments [66]. In the *Reovirales* order, members of the same genus typically have the same number of segments, therefore FAKV was tentatively placed within the *Dinovernavirus* genus alongside *Aedes pseudoscutellaris* reovirus, neither of which have previously been identified in SA [66, 67]. Here we identified six of the nine segments in SHI17MP525 (ZA(Ar)SHI17MP525) and all nine

segments in KNP17MP624 (ZA(Ar)KNP17MP624). Phylogenetic analysis across the different segments indicated similar clustering patterns in that the samples group with each other and isolate KLF FAKV 21 which was also isolated from *Ae.aegypti* mosquitoes in Kenya [49]. Nine segments were identified for KNP17MP624, placing it within the *Dinovernavirus* genus. Although only six segments were identified for ZA(Ar)SHI17MP525, no other genus within the *Spinareoviridae* family is made up of six segments of genomic material and taking into account the high nucleotide and amino acid identity between the respective segments of ZA(Ar)SHI17MP525 and ZA(Ar) KNP17MP624, we also propose that this virus is within the *Dinovernavirus* genus. Species demarcation according to the ICTV, is based on sequence similarity, serological comparisons, analysis of electropherotypes, and identification of conserved terminal regions [67]. However, based on the amino acid similarity of segment 2 (the RNA-dependent RNA polymerase) which is above 90.00% (91%-93%) between the isolates and FAKV reference strains, it's likely that these isolates belong to the FAKV species.

The FORMV (*Rhabdoviridae* family, *Merhavirus* genus) was identified in SHI17MP525 and KNP17MP624. The FORMV was identified in a study which analysed RNA sequencing libraries deposited in the Sequence Read Archive (SRA) hosted by NCBI [47]. FORMV was found in laboratory colonies established from wild *Ae. aegypti* mosquitoes from Bundibugyo, Uganda [51]. This is in line with our study where it was identified in *Ae. aegypti* mosquitoes collected from the KNP. To our current knowledge, these Merhaviruses have not previously been identified in SA.

Two partitiviruses were detected in KNP17MP624: the VdV and the APLV-1. Partitiviruses are double-segmented viruses which until recently were thought to only infect fungi, plant, and protozoan hosts [52]. These viruses have now also been identified in arthropod hosts [52, 68]. The VdV was originally isolated from a colony of *Ae.aegypti* mosquitoes in Mexico [52]. The same study that identified the FORMV, identified the APLV-1 in *Ae. aegypti* colonies located worldwide [51].

Three pools of *Aedes* mosquitoes SHI17MP525 (*Ae.aegypti*), KNP17MP624 (*Ae.aegypti*), and SHI17MP541(*Ae.ochraceus*) contained viruses that grouped most closely with the KMV and GMV. Both the KMV and GMV, were originally isolated in *Ae.aegypti* mosquitoes from Kenya and Guadeloupe, respectively [53, 54]. These double-segmented viruses have also been found in several *Ae.aegypti* and *Ae.albopictus* pools in different studies [53, 69]. These viruses are unclassified in the *Riboviria* realm, hence no species demarcation exists. However, if following the trend that species demarcation is typically based on a minimum amino acid sequence divergence of 95.00% and taking into account the phylogenetic relationships displayed, we can tentatively say that ZA(Ar)SHI17MP525 and ZA(Ar)KNNP17MP624 belong to the KMV species, whereas ZA(Ar)SHI17MP541 is a novel virus as the amino acid similarities were below 95.00% and it was placed separately from both the KMV and GMV.

Lastly, in a pool of *An.squamosus* mosquitoes (KNP17MP70) collected from the KNP, we identified a novel virus, which through BLAST analysis was most similar to Hubei mosquito virus 1 (70% nucleotide identity), an ISV isolated from a pool of mosquitoes in China [70]. Based on amino acid similarity (< 80.00%), it is likely that this is a novel virus, tentatively named Malelane mosquito virus 1 (unclassified in the *Riboviria* realm).

Several studies have identified the same ISVs in the same mosquito species from different geographical locations [3, 61, 71]. Interestingly, several of the viruses identified here are found in the same mosquito species i.e. *Ae.aegypti*, potentially indicating that these viruses are part of this species' innate virome. Furthermore, several studies have shown that ISVs are capable of vertical transmission, maintaining their presence in mosquitoes which also makes ISVs ideal as biological control agents in arbovirus transmission, in that they will be maintained in mosquito populations [10]. With the exception of the ISV identified in the *An.squamosus* mosquitoes, all the ISVs identified here were identified in *Aedes* mosquitoes, a trend observed in several studies with *Aedes* mosquitoes having diverse viromes. No ISVs were detected in the *Culex* or *Mansonia* mosquitoes although only one pool of each was included in this study and we can therefore not make a conclusion on the presence of ISVs in these species. Although not commonly detected in *Culex* and *Mansonia* mosquitoes, a few such as the Culex flavivirus and the Atrato partiti-like virus 5 have been found in these species, respectively [29, 68, 72]. While we were able to detect partial and full genomes of both known and novel ISVs in important arbovirus vector species in SA including A*e.aegypti, Ae.ochraceus*, and *An.squamosus*, future studies could benefit from high-capacity sequencing platforms (which will allow for more reads), capture probes for RNA enrichment or improved host RNA/DNA depletion techniques to maximize detection of viruses that may be present in the mosquitoes. Additionally, while we sequenced only ten mosquito pools, which may have lacked arboviruses, increasing sequencing depth might enhance detection of these pathogens. Despite this limitation, SISPA followed by Illumina NGS sequencing allowed us to identify several viruses not previously reported in SA. Considering that certain ISVs may have co-evolved with arboviruses, and that several ISVs are not necessarily found in arbovirus-containing families, looking into whether these seemingly non-pathogenic

viruses have potential in adapting to infect vertebrates could be of importance [72]. Therefore, the importance of these viruses not only lies in their potential to control the circulation of arboviruses without infecting mammalian hosts but by understanding how arboviruses evolve and adapt may allow for the implementation of new intervention strategies within a One Health approach [72].

Conclusion

Viral metagenomics was used to determine whether arboviruses not detected by PCR could be identified. Given that PCRs target specific regions of the genome, viral metagenomics was crucial for capturing a broader spectrum of viruses. While no arboviruses were identified, several known and novel ISVs were detected. Several ISVs have shown to influence the transmission of arboviruses, therefore understanding the virome of key mosquito species, such as *Aedes aegypti*, is crucial for assessing the risk of arboviruses becoming endemic. This knowledge may contribute to vector control strategies and prevention of arbovirus infection.

Abbreviations

Ae	Aedes
An	*Anopheles*
APLV-1	Aedes partiti-like virus 1
BANV	Banzi virus
BG	BioGents
BLASTn	nucleotide Basic Local Alignment Search Tool
CDC	Centres for Disease Control and Prevention
cDNA	Complementary DNA
CEARV	Centre for Emerging and Re-Emerging Arbo- and Respiratory virus
CFAV	Cell fusing virusagent
COXI	Cytochrome c oxidase subunit I
Cx	*Culex*
DENV	Dengue virus
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
FAKV	Fako virus
FORMV	Formosus virus
GMV	Guadeloupe mosquito virus
HS	High sensitivity
ICTV	International Committee on Taxonomy of Viruses
ISV	Insect-specific virus
KMV	Kwale mosquito virus
KNP	Kruger National Park
Ma	*Mansonia*
MAFFT	Multiple Alignment using Fast Fourier Transform
MEGA	Molecular Evolutionary Genetics Analysis
MIDV	Middelburg virus
NCBI	National Centre for Biotechnology Information
NGS	Next-Generation sequencing
PCR	Polymerase Chain Reaction
P-distance	Pairwise distance
RNA	Ribonucleic acid
SA	South Africa
SHUV	Shuni virus
SINV	Sindbis virus
SISPA	Sequence Independent Single Primer Amplification
SRA	Sequence Read Archive
TeAV	Tesano aedes virus
USA	United States of America
VdV	Verdadero virus
WNV	West Nile virus
YFV	Yellow fever virus
ZIKV	Zika virus

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42522-025-00185-1.

Supplementary Material 1

Acknowledgements

We thank the South African National Parks for logistical assistance and permission to collect mosquitoes under a research permit (SANPARKS permit reference number: VENTM1055); Dr Basil Brooke and Danny Govender for their contribution to the collection of the mosquitoes; all previous staff and students in the Centre for Emerging and Reemerging Arbo and Respiratory Virus Research program, Department of Medical Virology, University of Pretoria for their technical support.

Author contributions

MV conceived and designed the study. MG performed mosquito collections and morphological identifications. SK and CM conducted metagenomic sequencing. SK performed laboratory work, conducted bioinformatic analysis and interpreted the results. SK took the lead in developing the original manuscript. All authors contributed to the final draft of the paper. All the authors read and approved the final manuscript.

Funding

This work was funded by Prof Marietjie Venter's development fund, University of Pretoria.

Data availability

The datasets supporting the conclusions of this article are available in the NCBI Sequence Read Archive (SRA) under BioProject accession number http://PRJNA1267582. Individual sample data can be accessed via SRA accession numbers http://SRR33693534 (ZA(Ar)SHI17MP525), http://SRR33693536 (ZA(Ar)KNP17MP624), http://SRR33783618 (ZA(Ar)SHI17MP541) and http://SRR33906205 (ZA(Ar)KNP17MP70). Viral sequences have been submitted to GenBank under accession numbers http://PV658503, http://PV658504, http://PV776881, http://PV730233, http://PV730234, http://PV730235 http://PV730236, http://PV730237, http://PV730238, http://PV730239, http://PV730240, http://PV730241, http://PV730242, http://PV730243, http://PV730244, http://PV730245, http://PV730246, http://PV730247, http://PV730248, http://PV730249, http://PV730250, http://PV730251 http://PV730252, http://PV730253, http://PV730254, http://PV730255, http://PV730256, http://PV730257, http://PV730258, http://PV776882 and http://PV776883.

Declarations

Ethical approval

Mosquito collections and screenings were approved by the Department of Agriculture Land Reform and Rural Development (DALRRD) under section 20 (12/11/1/1) and the MSc research committee, University of Pretoria, South Africa.

Consent for publication

Not applicable.

Competing interests

MV is a member of the Editorial Board of One Health Outlook.

Received: 1 September 2025 / Accepted: 12 November 2025
Published online: 06 December 2025

References

1. Prudencio M. In fairness to mosquitoes. Trends Parasitol. 2020;36(11):876–77. https://doi.org/10.1016/j.pt.2020.08.003.
2. Centers for Diseaase Control and Prevention (CDC). [Internet]. Fighting the World's deadliest animal. [updated 2024 Aug 14; cited 2024 Oct 3]. Available from: https://www.cdc.gov/global-health/impact/fighting-the-worlds-deadliest-animal.html.
3. Oguzie JU, Nwangwu UC, Oluniyi PE, Olumade TJ, George UE, Kazeem A, et al. Metagenomic sequencing characterizes a wide diversity of viruses in field mosquito samples in Nigeria. Sci Rep. 2022;12(1):7616. https://doi.org/10.1038/s41598-022-11797-2.
4. Mukherjee S. Emerging infectious diseases: epidemiological perspective. Indian J Dermatol. 2017;62(5):459–67. https://doi.org/10.4103/ijd.IJD_379_17.
5. Chapman GE, Baylis M, Archer D, Daly JM. The challenges posed by equine arboviruses. Equine Vet J. 2018;50(4):436–45. https://doi.org/10.1111/evj.12829.
6. Xiao P, Li C, Zhang Y, Han J, Guo X, Xie L, et al. Metagenomic sequencing from mosquitoes in China reveals a variety of insect and human viruses. Front Cell Infect Microbiol. 2018;8. https://doi.org/10.3389/fcimb.2018.00364.
7. Ramírez AL, Colmant AMG, Warrilow D, Huang B, Pyke AT, McMahon JL, et al. Metagenomic analysis of the virome of mosquito excreta. mSphere. 2020;5(5). https://doi.org/10.1128/mSphere.00587-20.
8. Stollar V, Thomas VL. An agent in the *aedes aegypti* cell line (Peleg) which causes fusion of *aedes albopictus* cells. Virology. 1975;64(2):367–77. https://doi.org/10.1016/0042-6822(75)90113-0.
9. Amoa-Bosompem M, Kobayashi D, Murota K, Faizah AN, Itokawa K, Fujita R, et al. Entomological assessment of the status and risk of mosquito-borne arboviral transmission in Ghana. Viruses. 2020;12(2). https://doi.org/10.3390/v12020147.
10. Carvalho VL, Long MT. Insect-specific viruses: an overview and their relationship to arboviruses of concern to humans and animals. Virology. 2021;557:34–43. https://doi.org/10.1016/j.virol.2021.01.007.
11. Zhou N, Huang E, Guo X, Xiong Y, Xie J, Cai T, et al. Cell fusing agent virus isolated from Aag2 cells does not vertically transmit in *aedes aegypti* via artificial infection. Parasites Vectors. 2023;16(1):402. https://doi.org/10.1186/s13071-023-06033-3.
12. Baidaliuk A, Miot EF, Lequime S, Moltini-Conclois I, Delaigue F, Dabo S, et al. Cell-fusing agent virus reduces arbovirus dissemination in *aedes aegypti* mosquitoes in vivo. J Virol. 2019;93(18):10.1128/jvi.00705–19. https://doi.org/10.1128/jvi.00705-19.
13. Johnson T, Braack L, Guarido M, Venter M, Gouveia Almeida AP. Mosquito community composition and abundance at contrasting sites in northern South Africa, 2014-2017. J Vector Ecol. 2020;45(1):104–17. https://doi.org/10.1111/jvec.12378.
14. Guarido MM, Riddin MA, Johnson T, Braack LEO, Schrama M, Gorsich EE, et al. *Aedes* species (*diptera: Culicidae*) ecological and host feeding patterns in the north-eastern parts of South Africa, 2014–2018. Parasites Vectors. 2021;14(1):339. https://doi.org/10.1186/s13071-021-04845-9.
15. Steyn J, Botha E, Stivaktas B VI, Beechler P, Myburgh BR, JG, et al. West Nile virus in wildlife and nonequine domestic animals, South Africa, 2010-2018. Emerg. Infect. Dis. 2019;25(12):2290–94. https://doi.org/10.3201/eid2512.190572.
16. Fourie I, Snyman J, Williams J, Ismail A, Jansen van Vuren P, Venter M. Epidemiological and genomic characterisation of Middelburg and sindbis alphaviruses identified in horses with febrile and neurological infections. In: Viruses. Vol. 14(9) doi:South Africa; 2014-2018. 2022. https://doi.org/10.3390/v14092013.
17. van Niekerk S, Human S, Williams J, van Wilpe E, Pretorius M, Swanepoel R, et al. Sindbis and Middelburg old world alphaviruses associated with neurologic disease in horses, South Africa. Emerg. Infect. Dis. 2015;21(12):2225–29. https://doi.org/10.3201/eid2112.150132.
18. Steyn J, Fourie I, Steyl J, Williams J, Stivaktas V, Botha E, et al. Zoonotic alphaviruses in fatal and neurologic infections in wildlife and nonequine domestic animals, South Africa. Emerg. Infect. Dis. 2020;26(6):1182–91. https://doi.org/10.3201/eid2606.191179.
19. Motlou TP, Williams J, Venter M. Epidemiology of shuni virus in horses in South Africa. Viruses. 2021;13(5):937. https://doi.org/10.3390/v13050937.
20. Steyn J, Motlou P, van Eeden C, Pretorius M, Vi S, Williams J, et al. Shuni virus in wildlife and nonequine domestic animals, South Africa. Emerg. Infect. Dis. 2020;26(7):1521–25. https://doi.org/10.3201/eid2607.190770.
21. MacIntyre C, Lourens C, Mendes A, de Villiers M, Avenant T, du Plessis NM, et al. West Nile virus, an underdiagnosed cause of acute fever of unknown origin and neurological disease among hospitalized patients in South Africa. Viruses. 2023;15(11):2207.
22. Fourie I, Williams J, Ismail A, Jansen van Vuren P, Stoltz A, Venter M. Detection and genome characterization of Middelburg virus strains isolated from CSF and whole blood samples of humans with neurological manifestations in South Africa. PLoS Negl Trop Dis. 2022;16(1):e0010020. https://doi.org/10.1371/journal.pntd.0010020.
23. Meno K, Yah C, Mendes A, Venter M. Incidence of sindbis virus in hospitalized patients with acute fevers of unknown cause in South Africa, 2019-2020. Front Microbiol. 2021;12:798810. https://doi.org/10.3389/fmicb.2021.798810.
24. Motlou TP, Venter M. Shuni virus in cases of neurologic disease in humans, South Africa. Emerg. Infect. Dis. 2021;27(2):565–69. https://doi.org/10.3201/eid2702.191551.
25. Guarido MM, Fourie I, Meno K, Mendes A, Riddin MA, MacIntyre C, et al. Alphaviruses detected in mosquitoes in the north-eastern regions of South Africa, 2014 to 2018. Viruses. 2023;15(2):414.
26. Guarido MM, Motlou T, Riddin MA, MacIntyre C, Manyana SC, Johnson T, et al. Potential mosquito vectors for shuni virus, South Africa, 2014-2018. Emerg. Infect. Dis. 2021;27(12):3142–46. https://doi.org/10.3201/eid2712.203426.
27. MacIntyre C, Guarido MM, Riddin MA, Johnson T, Braack L, Schrama M, et al. Survey of West Nile and banzi viruses in mosquitoes, South Africa, 2011-2018. Emerg. Infect. Dis. 2023;29(1):164–69. https://doi.org/10.3201/eid2901.220036.
28. Guarido M Identification and phylogenetic analysis of Aedes species (Diptera: Culicidae) and arboviruses associated with them across tropical and temperate regions of South Africa (Pretoria. 2015-2018. University of Pretoria 2020.
29. Guarido MM, Govender K, Riddin MA, Schrama M, Gorsich EE, Brooke BD, et al. Detection of insect-specific flaviviruses in mosquitoes (*diptera: Culicidae*) in northeastern regions of South Africa. Viruses. 2021;13(11):2148.
30. Schrama M, Hunting ER, Beechler BR, Guarido MM, Govender D, Nijland W, et al. Human practices promote presence and abundance of disease-transmitting mosquito species. Sci Rep. 2020;10(1):13543. https://doi.org/10.1038/s41598-020-69858-3.
31. Gorsich EE, Beechler BR, van Bodegom PM, Govender D, Guarido MM, Venter M, et al. A comparative assessment of adult mosquito trapping methods to estimate spatial patterns of abundance and community composition in southern Africa. Parasites Vectors. 2019;12(1):462. https://doi.org/10.1186/s13071-019-3733-z.
32. Jupp P. Mosquitoes of southern Africa : *culicinae* and *toxorhynchitinae*. Ekogilde Publishers; 1996.
33. Yiau-Min H. A pictorial key for the identification of the subfamilies of *Culicidae*, genera of *Culicinae*, and subgenera of *aedes* mosquitoes of the afrotropical region. *Diptera*: *Culicidae*): Walter Reed Biosystematics Unit; 2001.
34. Huang YM, Rueda LM. A pictorial key to the species of *aedes* (*ochlerotatus* and *Coetzeemyia*) in the afrotropical region (*diptera: Culicidae*). Zootaxa. 2014;3754:592–600. https://doi.org/10.11646/zootaxa.3754.5.5.
35. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, et al. Viral genome sequencing by random priming methods. BMC Genomics. 2008;9(1):5. https://doi.org/10.1186/1471-2164-9-5.
36. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc Natl Acad Sci, India, Sect B Biol Sci. 2001;98(20):11609–14. https://doi.org/10.1073/pnas.211424698.
37. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. Genome Biol. 2009;10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25.
38. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–59. https://doi.org/10.1038/nmeth.1923.
39. Yates Andrew D, Allen J, Amode RM, Azov AG, Barba M, Becerra A, et al. Ensembl genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Res. 2021;50(D1):D996–1003. https://doi.org/10.1093/nar/gkab1007.
40. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. Bioinformatics. 2015;31(10):1674–76. https://doi.org/10.1093/bioinformatics/btv033.
41. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–52. https://doi.org/10.1093/bioinformatics/bts565.
42. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res. 2002;30(14):3059–66. https://doi.org/10.1093/nar/gkf436.

43. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. In: Hall TA, editor. Nucleic acids symposium series. Oxford; 1999.
44. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2005;33(suppl_1):D34–8. https://doi.org/10.1093/nar/gki063.
45. Kumar S, Stecher G, Li M, Knyaz C, Tamura KMX. Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–49. https://doi.org/10.1093/molbev/msy096.
46. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. Dna primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3(5):294–99.
47. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111–20. https://doi.org/10.1007/BF01731581.
48. Nei M, Kumar S. Molecular evolution and phylogenetics. Oxford University Press; 2000.
49. National Center for Biotechnology Information (NCBI) [Internet]. Mag: fako virus isolate KLF_FAKV_21 segment 1 genomic sequence. NCBI [updated 2023 Aug 13.; cited 2024 Nov 24]. Available from: https://www.ncbi.nlm.nih.gov/nuccore/OR270150.1.
50. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol Biol Evol. 1992;9(4):678–87. https://doi.org/10.1093/oxfordjournals.molbev.a040752.
51. Parry R, James ME, Asgari S. Uncovering the worldwide diversity and evolution of the virome of the mosquitoes *aedes aegypti* and *aedes albopictus*. Microorganisms. 2021;9(8). https://doi.org/10.3390/microorganisms9081653.
52. Cross ST, Maertens BL, Dunham TJ, Rodgers CP, Brehm AL, Miller MR, et al. Partitiviruses infecting *drosophila melanogaster* and *aedes aegypti* exhibit efficient biparental vertical transmission. J Virol. 2020;94(20). https://doi.org/10.1128/jvi.01070-20.
53. Calle-Tobón A, Pérez-Pérez J, Forero-Pineda N, Chávez OT, Rojas-Montoya W, Rúa-Uribe G, et al. Local-scale virome depiction in Medellín, Colombia, supports significant differences between *aedes aegypti* and *aedes albopictus*. PLoS One. 2022;17(7):e0263143. https://doi.org/10.1371/journal.pone.0263143.
54. Shi C, Beller L, Deboutte W, Yinda KC, Delang L, Vega-Rúa A, et al. Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. Microbiome. 2019;7(1):121. https://doi.org/10.1186/s40168-019-0734-2.
55. Omuoyo D, Nyamwaya D, Kamau E, Nyagwange J, Karanja H, Gitonga J, et al. Metagenomic analysis of coastal Kenya female aedes aegypti mosquito rna metaviromes reveal presence of diverse insect specific viruses [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res. 2023;8(136). https://doi.org/10.12688/wellcomeopenres.18868.2.
56. Maia LJ, Oliveira CH, Silva AB, Souza PAA, Müller NFD, Cardoso JDC, et al. Arbovirus surveillance in mosquitoes: historical methods, emerging technologies, and challenges ahead. Exp Biol Med (maywood). 2023;248(22):2072–82. https://doi.org/10.1177/15353702231209415.
57. van den Bergh C, Thompson PN, Swanepoel R, Almeida APG, Paweska JT, van Vuren P J, et al. Detection of rift Valley fever virus in *aedes* (*aedimorphus*. Durbanensis, South Africa. Pathog. 2022;11(2). https://doi.org/10.3390/pathogens11020125.
58. Sang R, Arum S, Chepkorir E, Mosomtai G, Tigoi C, Sigei F, et al. Distribution and abundance of key vectors of Rift Valley fever and other arboviruses in two ecologically distinct counties in Kenya. PLoS Negl Trop Dis. 2017;11(2):e0005341. https://doi.org/10.1371/journal.pntd.0005341.
59. Campbell LP, Reuman DC, Lutomiah J, Peterson AT, Linthicum KJ, Britch SC, et al. Predicting abundances of *aedes mcintoshi*, a primary Rift Valley fever virus mosquito vector. PLoS One. 2019;14(12):e0226617. https://doi.org/10.1371/journal.pone.0226617.
60. Cornel AJ, Lee Y, Almeida APG, Johnson T, Mouatcho J, Venter M, et al. Mosquito community composition in South Africa and some neighboring countries. Parasit Vectors. 2018;11(1):331. https://doi.org/10.1186/s13071-018-2824-6.
61. Martin E, Tang W, Briggs C, Hopson H, Juarez JG, Garcia-Luna SM, et al. Cell fusing agent virus (flavivirus) infection in *aedes aegypti* in Texas: seasonality, comparison by trap type, and individual viral loads. Arch Virol. 2020;165(8):1769–76. https://doi.org/10.1007/s00705-020-04652-0.
62. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB, Estrada A, et al. Discovery of a novel alphavirus related to Eilat virus. J Gen Virol. 2017;98(1):43–49. https://doi.org/10.1099/jgv.0.000694.
63. Hermanns K, Marklewitz M, Zirkel F, Kopp A, Kramer-Schadt S, Junglen S. Mosquito community composition shapes virus prevalence patterns along anthropogenic disturbance gradients. Elife. 2023;12. https://doi.org/10.7554/eLife.66550.
64. Kobayashi D, Isawa H, Fujita R, Murota K, Itokawa K, Higa Y, et al. Isolation and characterization of a new iflavirus from *armigeres spp.* mosquitoes in the Philippines. J Gen Virol. 2017;98(11):2876–81. https://doi.org/10.1099/jgv.0.000929.
65. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, Herrero S, et al. ICTV virus taxonomy profile. Iflaviridae. J Gen Virol. 2017;98(4):527–28. https://doi.org/10.1099/jgv.0.000757.
66. Auguste AJ, Kaelber JT, Fokam EB, Guzman H, Carrington CV, Erasmus JH, et al. A newly isolated reovirus has the simplest genomic and structural organization of any reovirus. J Virol. 2015;89(1):676–87. https://doi.org/10.1128/jvi.02264-14.
67. Matthijnssens J, Attoui H, Bányai K, Brussaard CPD, Danthi P, Del Vas M, et al. ICTV virus taxonomy profile: *spinareoviridae* 2022. J Gen Virol. 2022;103(11). https://doi.org/10.1099/jgv.0.001781.
68. Li C, Liu S, Zhou H, Zhu W, Cui M, Li J, et al. Metatranscriptomic sequencing reveals host species as an important factor shaping the mosquito virome. Microbiol Spectr. 2023;11(2):e0465522. https://doi.org/10.1128/spectrum.04655-22.
69. Rangel MEO, Oliveira LPR, Cabral AD, Gois KC, Lima MVM, Reis B, et al. Dengue-2 and Guadeloupe mosquito virus rna detected in *aedes* (*Stegomyia*) *spp.* collected in a vehicle impound yard in Santo André, sp, Brazil. Insects. 2021;12(3). https://doi.org/10.3390/insects12030248.
70. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate rna virosphere. Nature. 2016;540(7634):539–43. https://doi.org/10.1038/nature20167.
71. Langat SK, Kerich G, Cinkovich S, Johnson J, Ambale J, Yalwala S, et al. Genome sequences of Phasi Charoen-like phasivirus and fako virus from *aedes aegypti* mosquitoes collected in coastal Kenya. Microbiol Resour Announc. 2023;12(11):e00678–23. https://doi.org/10.1128/MRA.00678-23.
72. Öhlund P, Lundén H, Blomström A-L. Insect-specific virus evolution and potential effects on vector competence. Virus Genes. 2019;55(2):127–37. https://doi.org/10.1007/s11262-018-01629-9.

Publisher's Note