Metagenomic Sequencing Characterizes a Wide Diversity of Viruses in Field Mosquito Samples in Nigeria

Judith Uche Oguzie
Redeemer's University
Udoka Nwangwu
National Arbovirus and Vectors Research Centre
Paul Oluniyi
Redeemer's University
Testimony Olumade
Redeemer's University
Uwem George
Redeemer's University
Akano Kazeem
Redeemer's University
Bolajoko Bankole
Redeemer's University
Farida Brimmo
African Centre of Excellence for Genomics of Infectious Diseases
Chukwuemeka Asadu
National Arbovirus and Vectors Research Centre
Okechukwu Chukwuekezie
National Arbovirus and Vectors Research Centre
Josephine Ochu
National Arbovirus and Vectors Research Centre
Catherine Makwe
Médecins Sans Frontières
Festus Dogunro
National Arbovirus and Vectors Research Centre
Cosmas Onwude
National Arbovirus and Vectors Research Centre
William Nwachukwu
Nigeria Center for Disease Control
Ebuka Ezihe

National Arbovirus and Vectors Research Centre
Gilkenny Okonkwo
Médecins Sans Frontières
Ndubuisi Umazi
Médecins Sans Frontières
Jacob Maikere
Médecins Sans Frontières
Nneka Agashi
National Arbovirus and Vectors Research Centre
Emelda Eloy
National Arbovirus and Vectors Research Centre
Stephen Anokwu
National Arbovirus and Vectors Research Centre
Angela Okoronkwo
National Arbovirus and Vectors Research Centre
Ebuka Nwosu
National Arbovirus and Vectors Research Centre
Sandra Etiki
National Arbovirus and Vectors Research Centre
Ifeoma Ngwu
National Arbovirus and Vectors Research Centre
Chikwe Ihekweazu
Nigeria Center for Disease Control
Onikepe Folarin
Redeemer's University
Isaac Komolafe
Redeemer's University
Christian Happi (happic@run.edu.ng)
Redeemer's University

Research Article

Keywords: Yellow fever, mosquitoes, vectors, entomological surveillance, Metagenomics

Posted Date: November 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1014537/v1

Metagenomic sequencing characterizes a wide diversity of viruses in field mosquito samples in Nigeria

Judith U. Oguzie[1,2], Udoka C. Nwangwu[3], Paul E Oluniyi[1,2], Testimony J. Olumade[1,2], Uwem E. George[1,2], Akano Kazeem[1,2], Bolajoko E. Bankole[1,2], Farida O. Brimmo[1], Chukwuemeka C. Asadu[3], Okechukwu C. Chukwuekezie[3], Josephine C. Ochu[3], Catherine O. Makwe[4], Festus A. Dogunro[3], Cosmas O. Onwude[3], William E. Nwachukwu[5], Ebuka K. Ezihe[3], Gilkenny K. Okonkwo[4], Ndubuisi E. Umazi[4], Jacob Maikere[4], Nneka O. Agashi[3], Emelda I. Eloy[3], Stephen O. Anokwu[3], Angela I. Okoronkwo[3], Ebuka M. Nwosu[3], Sandra O. Etiki[3], Ifeoma M. Ngwu[3], Chikwe Ihekweazu[5], Onikepe A. Folarin[1,2], Isaac O.O Komolafe[1,2], Christian T. Happi[1,2*]

1. African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Osun State, Nigeria.
2. Department of Biological Sciences, Faculty of Natural Sciences, Redeemer's University, Ede, Osun State, Nigeria.
3. National Arbovirus and Vectors Research Centre (NAVRC), Enugu, Enugu State, Nigeria
4. Médecins Sans Frontières (MSF Belgium).
5. Nigeria Center for Disease Control, Abuja, Nigeria.

*Correspondences: Professor Christian T. Happi (email: happic@run.edu.ng; Tel: +234-8023383684)

Abstract

Mosquito vectors are a tremendous public health threat as one in six diseases worldwide is vector-borne transmitted mainly by mosquitoes. In the last couple of years, there have been active Yellow fever virus (YFV) outbreaks in many settings in Nigeria, and nationwide, entomological surveillance has been a significant effort geared towards understanding these outbreaks. In this study, we used a metagenomic sequencing approach to characterize viruses present in vector samples collected during various outbreaks of Yellow fever (YF) in Nigeria between 2017 and 2020. Mosquito samples were grouped into pools of one to fifty mosquitoes, each based on species, sex and location. Twenty-five pools of *Aedes spp* and one pool of *Anopheles spp* collected from nine states were sequenced and metagenomic analysis was carried out. We identified a wide diversity of viruses belonging to various families in this sample set. Seven different viruses detected included: Fako virus, Phasi Charoen-like virus, Verdadero virus, Chaq like-virus, Aedes aegypti totivirus, cell fusing agent virus and Tesano Aedes virus. Although there are no reports of these viruses being pathogenic, they are an understudied group in the same families and closely related to known pathogenic arboviruses. Our work shows the revolutionary power of genomics to help detect the presence of viruses in mosquito vectors and unravel their pathogenic potential. This can help develop diagnostics and vaccines to help prevent the diseases caused by these viruses before they cause outbreaks in human populations.

Key words: Yellow fever, mosquitoes, vectors, entomological surveillance, Metagenomics.

Introduction

One in six diseases worldwide is transmitted by mosquito vectors[1]. It is estimated that half of the world's population is in danger of viral illnesses spread by mosquitoes and causing death in millions of people annually. Rift Valley fever virus (RVF), Dengue virus (DENV), Zika virus (ZIKV), Chikungunya virus (CHIKV), Yellow fever virus (YFV), Japanese encephalitis virus (JEV), and Ross River virus (RRV) are just a few of the mosquito-borne viruses that have caused disease epidemics both in humans and animals. Additionally, an increasing number of viruses specific to arthropods, classified as insect-specific viruses (ISVs), have been identified in the last two decades in diverse mosquito populations around the world [2-8]. ISVs and pathogenic arboviruses evolutionary relationship remains uncertain[9]. However, evolution from arthropod-specific viruses has been assumed for the genus Flavivirus. Pathogenic viruses are thought to have evolved from insect-specific to dual host viruses [10-12].

Metagenomic sequencing has increased exponentially the number of mosquito-borne viruses isolated in the last couple of years and further provided fresh insight into the enormous complexity and variety of invertebrate RNA viruses [4,6,11]. Recently, in Nigeria, metagenomic sequencing was used to identify an ongoing yellow fever outbreak and its aetiology and inform real-time public health actions, resulting in accurate and timely disease management and control[13]. A greater understanding of the virome in mosquito species in Nigeria could allow for a more accurate assessment of mosquito-borne disease risk, vector competence and mosquito management.

In the course of various YFV outbreaks in Nigeria between 2017 and 2020, we collected vector samples (mostly *Aedes spp*) in sites where there were active YFV cases. Next-generation sequencing (NGS) was carried out on 26 pools of 1,300 mosquitoes (50 mosquitoes per pool) across nine (9) states in Nigeria using a metagenomic protocol as previously described[14]. In this paper we present our findings and discuss the implications.

Results

Metagenomics Analysis

Metagenomics analysis carried out on sequenced samples revealed the presence of a wide range of viruses (Figure 1), some of which are being reported for the first time in Nigeria and Africa. The mosquito species that were pooled after morphological identification included the following: *Aedes aegypti* (n = 12 pools), *Aedes albopictus* (n = 10 pools), *Aedes simpsoni complex* (n = 2) *Aedes luteocephalus* and *Anopheles coustani* (n = 1 pool each) with *Aedes aegypti* and *Aedes albopictu*s accounting for more than 80% of the mosquito pools analyzed. Of the 26 pools, irrespective of the mosquito species, 7, 1, 3, 13, 13, 12, 2, 1 and 7 were positive for *Aedes aegypti* totivirus (AaTV), cell fusing agent virus (CFAV), Chaq like-virus (CLV), Phasi Charoen like phasivirus (PCLV) (L segment), PCLV (M segment), PCLV (S segment), Verdadero virus (VdV), Fako virus (FKV), Tesano Aedes virus (TeAV), respectively (Figure 1). The prevalence of PCLV was significantly higher compared to other viruses ($P < 0.0001$). Table 1 shows the distribution of the virus in the mosquito pools. Aedes aegypti and Aedes albopictus were the most common mosquitoes in the study areas. They accounted for >80% of the pools analyzed in the study. In addition, no viral genome was assembled from Aedes simpsoni, Aedes simpsoni complex and Anopheles coustani pools. Excluding these minor groups of the mosquito pools, distributions of the viruses between the two major species were similar. In *Aedes aegypti*: Excluding CFAV from statistical comparison, the prevalence of PCLV was significantly higher compared with other viruses ($P = 0.006$; Table 1). In *Aedes albopictus*: Excluding FKV and TeAV from statistical comparison, the prevalence of the viruses was similar ($P = 0.41$; Table 1).

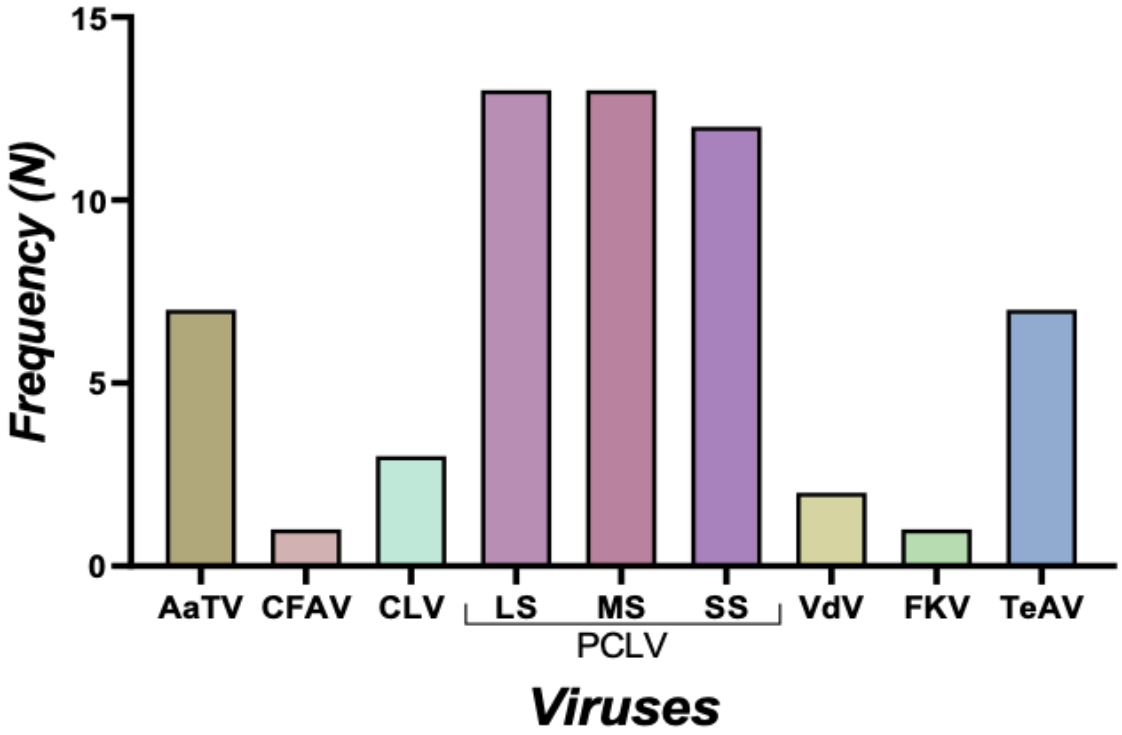


Figure 1: Frequency distribution of the viruses found in the mosquito pools

AaTV: *Aedes aegypti* totivirus, CFAV: cell fusing agent virus, CLV: Chaq-like virus, PCLV: Phasi Charoen-like phasivirus (L segment, M segment, and S segment), VdV: Verdadero virus, FKV: Fako virus, TeAV: Tesano Aedes virus.

Table 1: Distribution of the viruses based on the species of mosquitoes

Mosquito species	Viruses									P value
	AaTV	CFAV	CLV	PCLVLS	PCLVMS	PCLVSS	VdV	FKV	TeAV	
Aedes aegypti	4/12	0/12	2/12	9/12	9/12	9/12	0/12	2/12	1/12	0.003*
Aedes albopictus	3/10	1/10	1/10	4/10	4/10	3/10	2/10	0/10	0/10	0.41**
Aedes luteocephalus	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	-
Aedes simpsoni C.	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	-
Anopheles coustani	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	-

AaTV: *Aedes aegypti* totivirus, CFV: cell fusing agent virus, CLV: Chaq-like virus, PCLP: Phasi Charoen like phasivirus (L segment, M segment, and S segment), VdV: Verdadero virus, FV: Fako virus, TeAV: Tesano Aedes virus, C: complex

* = CFAV was excluded from this analysis

** = FKV and TeAV were excluded from this analysis

Genome Assembly and Phylogenetic Analysis

Genome assembly of viruses detected by our metagenomics pipeline was attempted to characterize the diversity of these viruses and their evolutionary relationship with previously reported viruses. A heatmap showing the distribution of viruses is shown in Supplementary Fig. S1.

Totiviridae

We assembled six (five full and one partial) genomes for *Aedes aegypti* totivirus (AaTV) out of the seven pools with reads for this virus. Phylogenetic analysis revealed that the *Aedes aegypti* totivirus sequences from this study clustered closely together and fell in the same major clade. Within the major clade, the sequence from Kwara state (BIS 100c) branched out on its own while the Ebonyi sequences (NB3 58, B2S 51, B2S 50 and B2S 83) clustered together, which could imply a localized/within-state spread of the virus. The short branch lengths of the sequences also show the limited diversity of the virus in Nigeria.

Bunyavirales

Phylogenetic analysis revealed similar clustering patterns across the three segments of the virus in Nigeria. According to states, there is no observed clustering pattern showing the virus's limited diversity in the country.

Other viruses (Tesano Aedes Virus**,** Chaq-like virus, Fako virus and Verdadero virus) were assembled from our sequenced data. We did not construct phylogenetic trees due to the lack of substantial genomes on NCBI for proper comparison. In which cases, only one to three genomes are available in the database.

The percentage similarity for these viruses assembled from this study is detailed in tables 2 and 4.

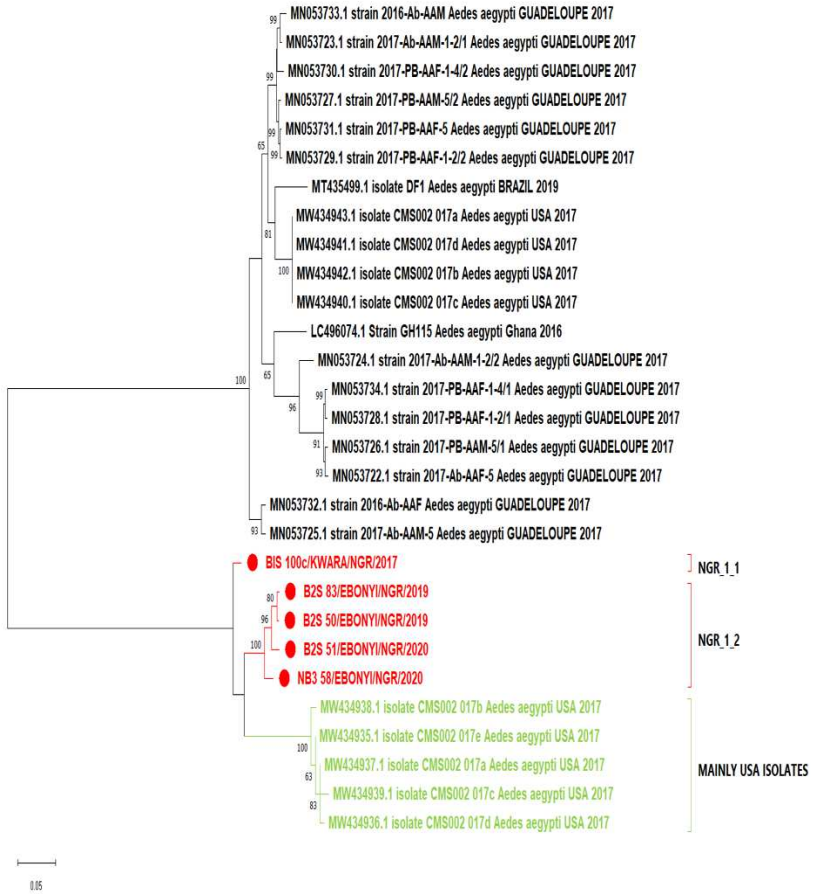


Figure 2: Phylogenetic analysis of orf1 nucleotide sequences of *Aedes aegypti* totivirus (AaTV). The evolutionary history was inferred using the Maximum Likelihood method and Tamura-Nei model with 1000 bootstrap replicates. The numbers at branch nodes indicate the bootstrap values ≥50%. All the reference strains are identified by name and GenBank accession number. Virus strains characterized in this study are highlighted in red.

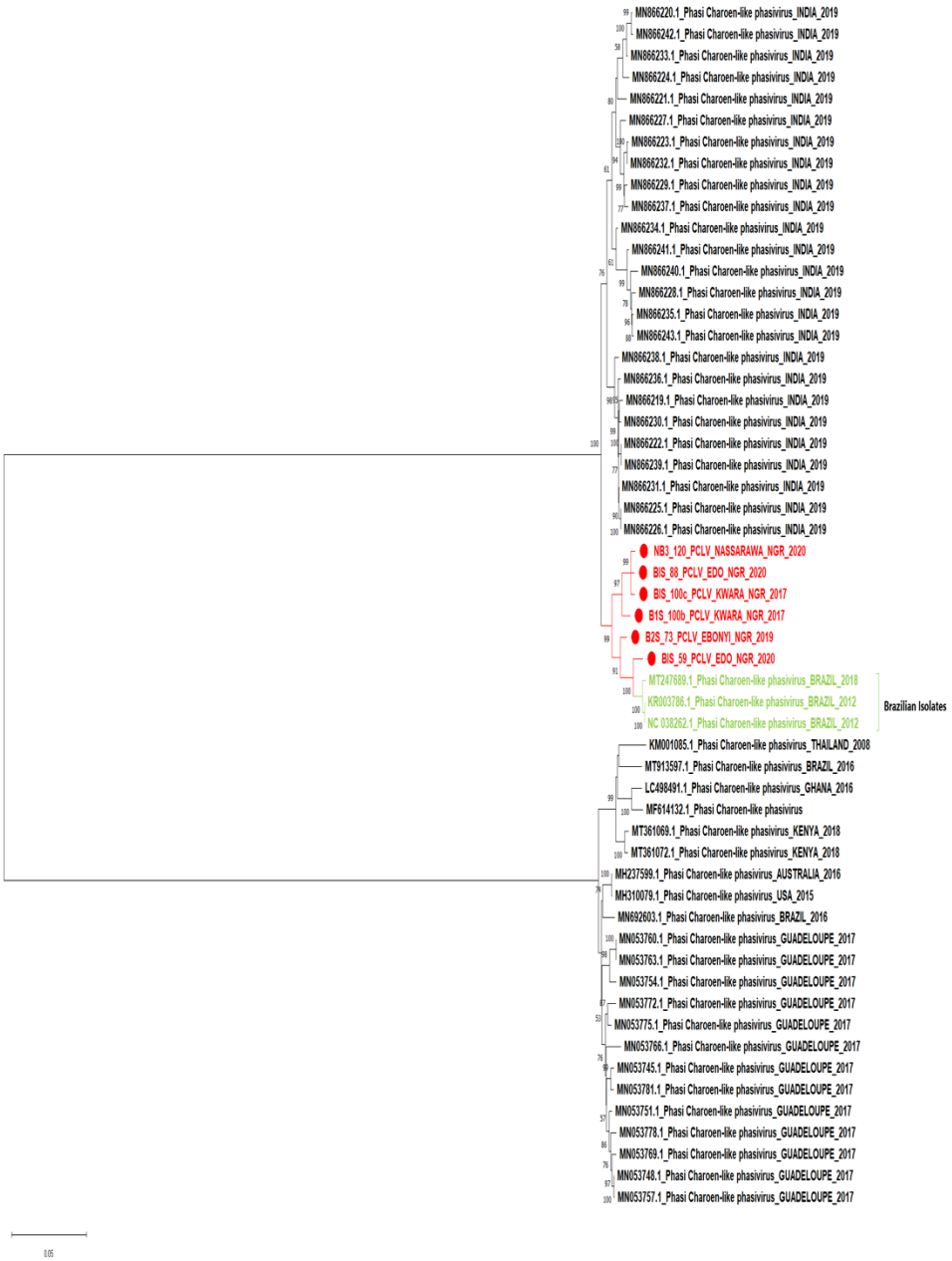


Figure 3: Maximum likelihood phylogenetic tree showing evolutionary relationship between the L segment (RdRp) of Phasi Charoen like phasivirus sequences from this study and those obtained from the NCBI database. The evolutionary history was inferred using the Maximum Likelihood method and Tamura-Nei model with 1000 bootstrap replicates. The numbers at branch nodes indicate the bootstrap values ≥50%.

Figure 4: Maximum likelihood phylogenetic tree showing evolutionary relationship between the M segment (glycoprotein) of Phasi Charoen like phasivirus sequences from this study and those obtained from the NCBI database. The evolutionary history was inferred using the Maximum Likelihood method and Tamura-Nei model with 1000 bootstrap replicates. The numbers at branch nodes indicate the bootstrap values ≥50%.

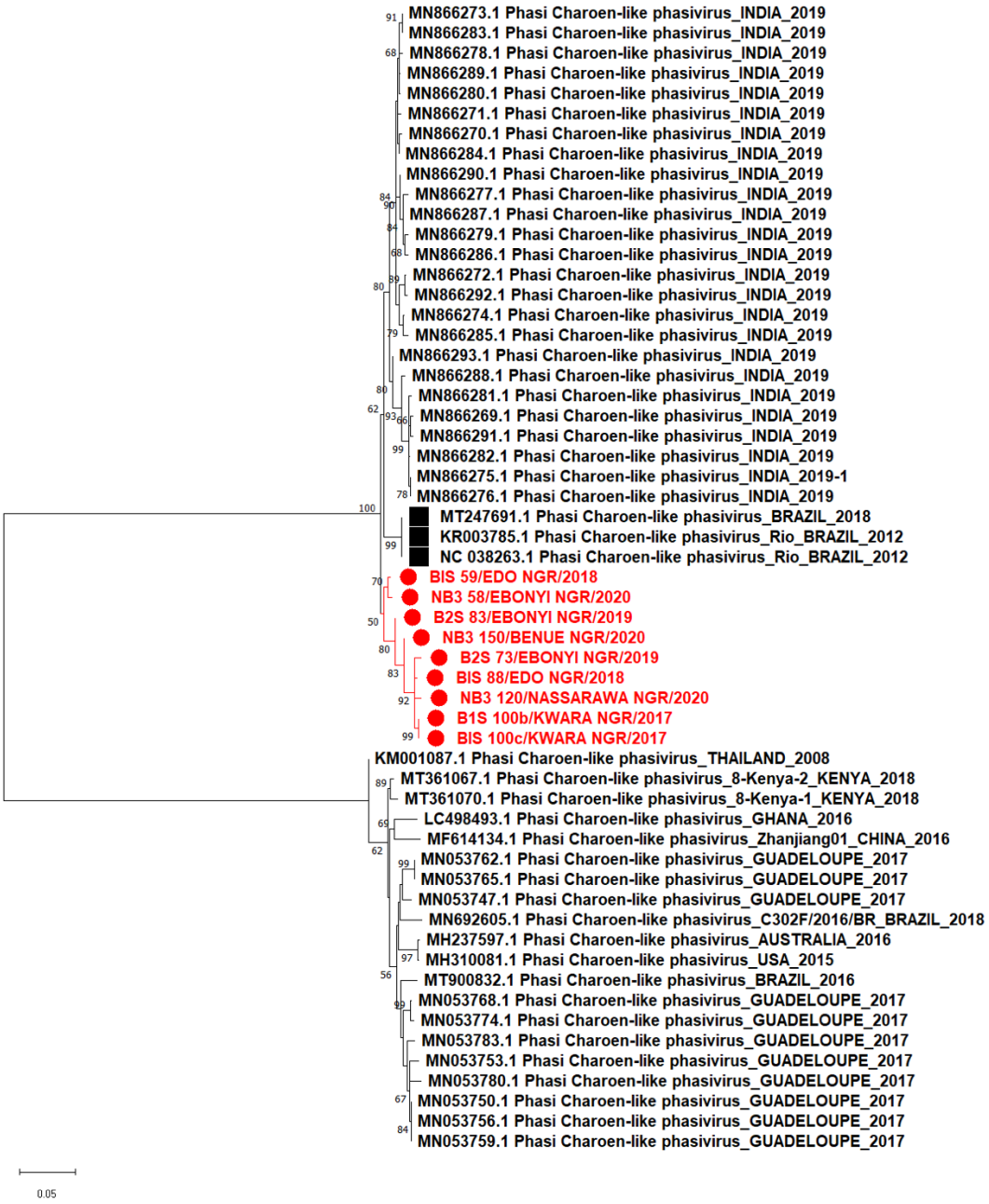
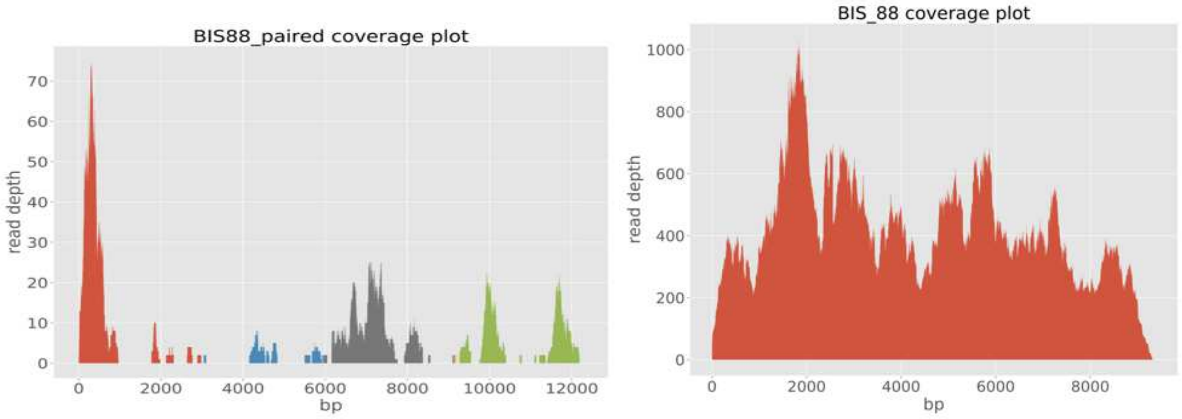
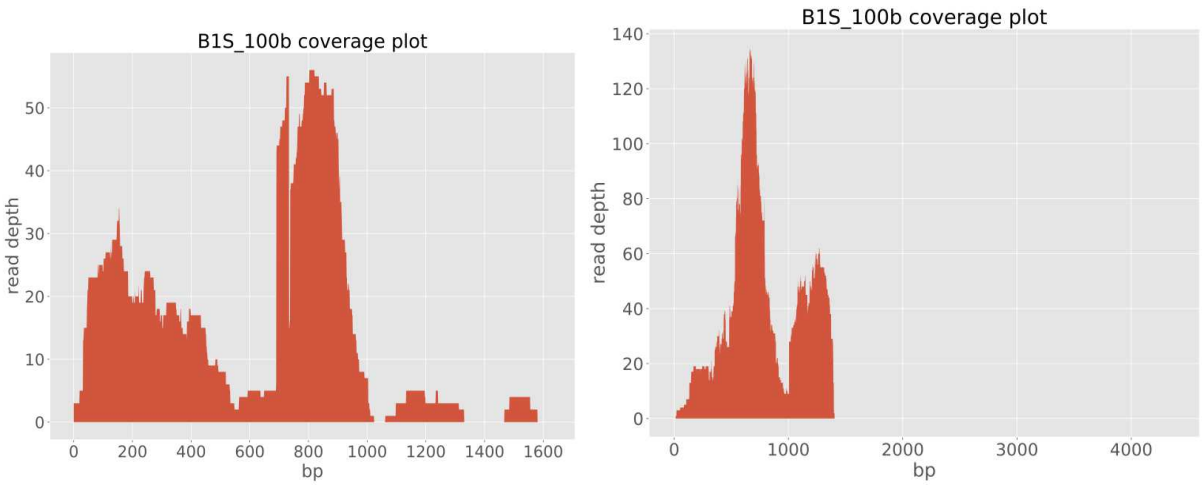


Figure 5: Maximum likelihood phylogenetic tree showing evolutionary relationship between the S segment (Nucleocapsid) of Phasi Charoen like phasivirus sequences from this study and those obtained from the NCBI database. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model with 1000 bootstrap replicates. The numbers at branch nodes indicate the bootstrap values ≥50%.



A **B**



C **D**

Figure 6: Coverage plot for A) Fako virus segments, B) Tesano Aedes virus C.) Chaq like virus D.) Verdadero virus.

215 Table 2. Genome characterization of Fako virus genome described in this study

Segment (Accession No)	Segment Length (bp)	ORF Length (aa)	Protein	Encoded Gene	Closest strain in GenBank	Identity (%)
2 (XXXXXXX)	2398	132	VP2	RNA-dependent RNA polymerase	KM978429.1	94.29
3 (XXXXXXX)	2925	213	VP3	major capsid protein	KM978432.1	98.22
4 (XXXXXXX)	3176	528	VP4	nonstructural protein	KM978433.1	97.95
5 (XXXXXXX)	2983	286	VP5	turret protein	KM978436.1	97.00

216

217

218 Table 3. Genome characterization of Phasi Charoen-like phasivirus genome described in this study

Segment	Strain name	Segment (Accession_No)	Segment Length (bp)	ORF Length (aa)	Encoded Gene	Closest strain in GenBank	Identity (%)
L	B1S_100b_PCLP	(XXXXXXX)	6781	2217	RNA-dependent RNA polymerase	MT361069.1	97.50
	B2S_48_PCLP	(XXXXXXX)	6156	279	RNA-dependent RNA polymerase	MN053766	98.44
	B2S_73_PCLP	(XXXXXXX)	6471	351	RNA-dependent RNA polymerase	MT361069.1	97.62

	B2S_83_PCLP	(XXXXXXX)	5906	316	RNA-dependent RNA polymerase	MT361069.1	95.93
	B1S_59_PCLP	(XXXXXXX)	6731	2217	RNA-dependent RNA polymerase	MH237599.1	98.60
	B1S_88_PCLP	(XXXXXXX)	6738	2217	RNA-dependent RNA polymerase	MT361069.1	97.58
	B1S_100c_PCLP	(XXXXXXX)	6514	837	RNA-dependent RNA polymerase	MT361069.1	92.00
M	B1S_100b_PCLP	(XXXXXXX)	3819	1237	Glycoprotein	MH237598.1	97.07
	B2S_48_PCLP	(XXXXXXX)	3564	501	Glycoprotein	MN053776.1	97.41
	B2S_50_PCLP	(XXXXXXX)	3346	192	Glycoprotein	MN053776.1	96.84
	B2S_73_PCLP	(XXXXXXX)	3521	1492	Glycoprotein	MN053776.1	97.13
	B2S_83_PCLP	(XXXXXXX)	3600	620	Glycoprotein	MN053782.1	95.72
	B1S_59_PCLP	(XXXXXXX)	3768	1237	Glycoprotein	MH237598.1	96.97
	B1S_88_PCLP	(XXXXXXX)	3833	1237	Glycoprotein	MH237598.1	94.04
	B1S_100c_PCLP	(XXXXXXX)	3593	558	Glycoprotein	MH237598.1	97.84
	NB3_58_PCLP	(XXXXXXX)	3354	217	Glycoprotein	MN053776.1	98.20
	NB3_120_PCLP	(XXXXXXX)	3768	1237	Glycoprotein	MH237598.1	97.21

	NB3_126_PCLP	(XXXXXXX)	3614	552	Glycoprotein	MN053776.1	96.77
	NB3_150_PCLP	(XXXXXXX)	3620	350	Glycoprotein	MN053776.1	97.57
S	B1S_100b_PCLP	(XXXXXXX)	1331	268	Nucleocapsid	MN866293.1	96.31
	B2S_73_PCLP	(XXXXXXX)	769	250	Nucleocapsid	MN866293.1	95.84
	B2S_83_PCLP	(XXXXXXX)	728	109	Nucleocapsid	MN866293.1	97.67
	B1S_59_PCLP	(XXXXXXX)	1326	268	Nucleocapsid	MN866293.1	96.98
	B1S_88_PCLP	(XXXXXXX)	1331	268	Nucleocapsid	MN866293.1	96.53
	B1S_100c_PCLP	(XXXXXXX)	861	265	Nucleocapsid	MN866293.1	96.28
	NB3_58_PCLP	(XXXXXXX)	1290	199	Nucleocapsid	MT361067.1	95.06
	NB3_120_PCLP	(XXXXXXX)	1325	268	Nucleocapsid	MN866293.1	96.00
	NB3_126_PCLP	(XXXXXXX)	665	189	Nucleocapsid	MT361067.1	98.05
	NB3_150_PCLP	(XXXXXXX)	1024	117	Nucleocapsid	MN866293.1	96.62

Table 4. Assembly and Genome characterization of members of *Iflaviridae*, *Partitiviridae*, *Flaviviridae* and *Totiviridae* described in this study

Family	Genus (Genome type)	Species	Isolate name	GenBank Accession number	Genome Length (bp)	Closest strain in GenBank	Identity (%)
Iflaviridae	Unclassified ssRNA (+)	Tesano virus	B1S_88_T	XXXXXXX	9310	LC496784.1 (16GH47)	93.30
			B1S_100c_T	XXXXXXX	9339	LC496784.1 (16GH47)	97.17
			NB3_120_T	XXXXXXX	9193	LC496784.1 (16GH47)	92.66
			NB3_150_T	XXXXXXX	9313	LC496784.1 (16GH47)	98.15
Partitiviridae	Unclassified Partitivirus (dsRNA)	Chaq-like virus	B1S_100b_Cv	XXXXXXX	1377	MT742176.1	98.77
		Verdadero virus	B1S_100b_V	XXXXXXX	1313	MT742174.1	95.83
Flaviviridae	Unclassisfied Flavivirus ssRNA(+)	Cell fusing agent virus	B2S_50_CFAV	XXXXXXX	4784	LR694076.1	97.79
Totiviridae	dsRNA	Aedes aegypti totivirus	B2S_50_AaT	XXXXXXX	7974	MN053727.1	91.83
			B2S_51_AaT	XXXXXXX	7557	MN053725.1	91.51
			NB3_58_AaT	XXXXXXX	7735	MN053727.1	91.73
			B2S_83_AaT	XXXXXXX	7677	MN053723.1	92.29
			B2S_100c_AaT	XXXXXXX	7953	MN053723.1	92.81

Discussion

Our study reports, for the first time, a metagenomics analysis of mosquitoes in Nigeria. We detected the presence of Fako virus, *Aedes aegypti* totivirus, Cell fusing agent, Tesano Aedes virus and Phasi Charoen-like phasivirus in Nigeria and Chaq-like and Verdadero viruses in Africa. Not only did we detect the presence of these viruses, but we were also able to obtain genome sequences from all. Our study highlights the application of next generation sequencing in identifying ISVs and granting insight into the mosquito microbiome in Nigeria. Interestingly, most mosquito pools showed more than one of the detected viruses, indicating the unbiased application of our protocol. *Aedes aegypti* and *Aedes albopictu*s were the most common mosquitoes in the study areas. They accounted for >80% of the pools analyzed in the study. Excluding the minor groups of the mosquito pools, distributions of the viruses between the two major species were similar. In *Aedes aegypti*: excluding CFAV and VdV from statistical comparison, TeAV was significantly less common than other viruses (Table 1). In *Aedes albopictus*: excluding FKV and TeAV from statistical comparison, the prevalence of the viruses was similar (Table 1).

We assembled four (segments 2,3,4 and 5) out of the nine segments of the Fako virus; reovirus of the genus Dinovernavirus. The first report of the virus is in mosquito pools from Cameroun[15]. This virus is maintained via mosquito to mosquito transmission and might have evolved from its initial ancestor through loss of function activities. There is no report of human infection with FKV. Our genome was isolated from one *Aedes aegypt*i pool from Edo State. Across all segments, our sequences were 96.7% similar to the first sequence from Cameroun (table 2).

There are seven genomes (four full and three partial) for Tesano Aedes virus (TeAV) for mosquito pools. TeAV is a member of the Iflaviridae family and was first isolated from Mosquito samples in Ghana[16]. The virus has shown evidence of vertical transmission[16]. It increases the growth of Dengue virus 1 (DENV-1) in a concentration-dependent manner under laboratory conditions[16]. Members of this family are known to infect insects with no reports of human infection. On average, our sequences were 96% similar to the only sequence for the TeAV available on NCBI (Table 4).

We had five complete genomes and one partial genome assembly for AaTV. This virus is a member of the *Totiviridae* family and a group of dsRNA viruses that infect fungi, protozoa, or invertebrates[17]. AaTV was highly distributed and present in mosquito pools from Kwara and Ebonyi States. Phylogenetic analysis of the virus genomes revealed that the Nigerian Strains clustered together in the same major clade, independent of the American/Asian/European lineages, implying continuous evolution and diversity of the virus (Figure 2). Sequences obtained from the same state (Ebonyi State) clustered closely in a sub-clade on the tree, resulting from the localized spread of the virus among the mosquito vectors in this community.

In addition, we had one partial genome assembly of a flavivirus; Cell fusing agent virus. Many medically important arboviruses belong to the *Flaviviridae* family. Our sequence shares 97.8% identity with the sequence from Uganda. CFAV is the first ISV reported and named after its characteristic CPE of fusion of cells[18]. Two decades later, researchers sequenced CFAV in 1992[19]. The virus has been isolated from *Aedes* species in dengue endemic areas [6, 20-24]. There is a close phylogenetic relationship between insect-specific (ISF) and medically important flaviviruses. This information could be valuable in understanding how ISFs enable/inhibit transmission of arboviruses in nature and their possible use as agents of biological control of vectors[2]. A study by Baidaliuk et al., 2019 evaluated how CFAV affects ZIKV and DENV-1 in *vitro* and *vivo* [25]. Their findings showed a negative correlation both in-*vitro* and *in-vivo,* indicating a decrease in transmission in both viruses due to the presence of CFAV.

Furthermore, we assembled 13 genomes for the L and M segments, as well as 12 for the S segment of Phasi Charoen-like-phasivirus (PCLV)- a bunyavirus first isolated from the Phasi Charoen district of Thailand from wild-caught *Aedes aegypti* larvae[26]. The phylogenetic tree based on the S segment (Nucleocapsid) and M segment (glycoprotein) displayed these segments as being in a separate cluster independent of previously detected lineages (figure 3 and 4). However, RdRp sequences encoded by the L segment displayed the PCLV from our study as being in the same clade as RdRp of previously detected PCLV from *Aedes aegypt*i from Brazil in 2012 (Figure 5). We have characterized our assemblies for the three segments of the virus (Table 3). There may be a link between PCLV and the transmission of arboviruses, e.g., Ae. *albopictus* cell line Aa23, persistently infected with CFAV, inhibited ZIKV replication and transmission[27]. At the same time, another study isolated PCLV from *Ae. aegypti* naturally infected with the Chikungunya virus

(CHIKV)[28] during an arbovirus surveillance program. The relationship between PCLV and CHIKV transmission is still unknown and needs further investigation.

Partitiviruses are known to infect a vast host, including plants, fungi with some members of this genus recently discovered in arthropods [29-31]. In this study, three (3) Chaq-like viruses were found from *Aedes spp* pools. The BLASTn search results of the three sequences from our study showed high nucleotide sequence identity (98.7%) to strain CLv.PozaRica20 (MT742176.1) was isolated from *Aedes aegypti* in Mexico in 2020. Showing the virus may be widely disseminated in the mosquito vector. Only three sequences are available on NCBI for this virus as of 23rd of October 2021, with limited information.

Another Partivirus isolated from this study is the Verdadero virus. Two pools of *Aedes aegypti* had reads for this virus from which we assembled one genome (Table 4). The coverage plots for some assembled viruses are presented in figure 6.

Generally, reports on the effects of ISVs on pathogenic arbovirus transmissions are controversial and conflicting from various studies and in vivo/in vitro conditions.

Although ISVs cannot grow in mammalian cell lines, a study in Brazil isolated a novel insect-specific virus, Guapiaçu virus (GUAPV), from the plasma sample of a febrile person[9]. Furthermore, the discovery of ISVs within the families of pathogenic viruses provided insights into the evolution and adaptation of these groups of viruses[32]. For example, ISVs belonging to *Flaviviridae* and *Bunyaviridae* families are thought to be ancient viruses with distinct lineages that have evolved at the same time and diversified with their vector hosts[10, 33-34]. Studies establishing vertical transmission[20, 35] and evidence of ISVs genomic sequence integration in the genome of insect vectors[36] have supported the hypothesis. Against this background, many pathogenic arboviruses probably gained their dual-host range by an adaptive evolution process that conferred the ability to infect vertebrates to ISVs. There is limited information on the pathogenicity of ISVs to their insect host. Furthermore, ISVs were considered possible biological control agents for vectors and arboviruses of public health importance due to their characteristic lack of replication in mammalian cell lines. Given all these possible applications of ISVs, they are an exciting group of viruses for further investigations.

The identification of these diverse groups of viruses from our study is a first step in applying local genomics capacity within the country for a holistic approach to disease outbreaks. Further

investigation of the pathogenic potential of these viruses, how they enhance/inhibit transmission of circulating arboviruses in Nigeria needs to be carried out.

Methods

Sample Description

During active yellow fever virus outbreaks, mosquito samples were trapped across nine (9) states in Nigeria between 2017 and 2020 for metagenomic sequencing. Generally, three trapping methods were used in this study; egg, larval and adult collections. At each sampling point, the coordinates were taken using a global positioning system (GPS) gadget as described by [37]. Live adult mosquitoes collected in the field were immobilized using Ethyl Acetate. All adult mosquitoes collected were morphologically identified to species level using keys of[38-40] and then pooled based on species and sex into 50 mosquitoes per pool. They were then introduced into well-labelled Eppendorf tubes containing RNAlater. The tubes were stored in freezers for the duration of the surveillance to keep the samples genetically intact. Immature stages (pupae and larvae) are reared to adults before pooling.

A cohort of samples making 26 pools of 1,300 mosquitoes were sequenced. Twenty-five (25) pools were *Aedes spp* (*Aedes aegypti*, *Aedes albopictus*, *Aedes luteocephalus*, *Aedes simpsoni*, *Aedes simpsoni complex*) and one pool of *Anopheles coustani* (Supplementary Table S1). Mosquitoes were trapped by the National Arbovirus and Vectors Research Centre (NAVRC), Enugu, Enugu State, Nigeria.

RNA Extraction and Metagenomics Sequencing

A total of 1,300 mosquitoes made into twenty-six (26) mosquito pools were sequenced based on the established unbiased protocol[14]. Briefly, Vector pools were initially homogenized in 1 ml of cooled Dulbecco's Modified Eagle Medium (DMEM) (composition- 500 ml DMEM High Glucose (4.5 g/I) with L-Glutamine), 1 ml Penicillin-Streptomycin, 15 ml Fetal Calf Serum (FCS) 3% and 5ml Amphotericin B) and 500ml of Zirconia beads (Firma Biospec: 2.0 mm, Cat. No 1107912). The contents were macerated for 10 minutes on the Qiagen Tissuelyser LT followed by centrifuging at 4,500 xg for 15 minutes. According to the manufacturer's instructions, the supernatant was further used for RNA extraction using the QIAamp Viral RNA extraction kit

(Qiagen, Hilden, Germany). Extracted RNA was turbo Dnased to remove contaminating DNA and cDNA synthesis was carried out according to the published protocol [13-14]. Sequencing libraries were made using the Illumina Nextera XT kit. Next generation sequencing was carried out on Illumina Miseq at the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Nigeria.

Bioinformatics Analysis

Bioinformatics analysis was carried out on the sequence data generated to identify viruses present in the samples. First, we demultiplexed individual libraries and removed reads mapping to the human genome or other known technical contaminants (e.g., sequencing adapters). Raw reads from the sequencing machine were passed through Microsoft's Premonition metagenomics pipeline (https://innovation.microsoft.com/en-us/premonition). We also carried out genome assembly of individual viruses detected using our publicly available viral-ngs [41-42] and VGEA pipelines[43].

Assembled genomes were annotated using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) with a 300-nt minimum length and the genome structure was annotated using NCBI Conserved Domain Database version 3.19 (expected value threshold of 1 x 10-2) with the NCBI viral genome database as references.

Multiple alignments of nucleotide sequences and deduced amino acids with those of reference strains from GenBank were analyzed using the MAFFT online version[44]. To determine the phylogenetic relationship of viruses detected from the mosquito pools, reference nucleotide sequences representing *Aedes aegypti* totivirus and Phasi Charoen-like phasivirus (strains with all three segments) were obtained from GenBank. The evolutionary history was inferred using the Maximum Likelihood method and Tamura-Nei model[45] with 1000 bootstrap replicates. Bootstrap support is indicated by values on branches. Evolutionary analysis was conducted in MEGA X version 10.1.8[46].

Data Availability

All the sequences from this study are available on github: https://github.com/acegid/Vector_Genomics_Study and will be released on NCBI once the paper is published.

Reference

1. World Health Organization. *A global brief on vector-borne diseases*. (2014).

2. Blitvich, B. J. & Firth, A. E. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* **7**, 1927–1959 (2015).

3. Vasilakis, N. & Tesh, R. B. Insect-specific viruses and their potential impact on arbovirus transmission. *Curr. Opin. Virol.* **15**, 69–74 (2015).

4. Cholleti, H. *et al.* Discovery of Novel Viruses in Mosquitoes from the Zambezi Valley of Mozambique. *PLoS One* **11**, e0162751 (2016).

5. Guzman, H. *et al.* Characterization of Three New Insect-Specific Flaviviruses: Their Relationship to the Mosquito-Borne Flavivirus Pathogens. *Am. J. Trop. Med. Hyg.* **98**, 410–419 (2018).

6. Zakrzewski, M. *et al.* Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok. *Sci. Rep.* **8**, 4690 (2018).

7. Cross, S. T. *et al.* Partitiviruses Infecting Drosophila melanogaster and Aedes aegypti Exhibit Efficient Biparental Vertical Transmission. *J. Virol.* **94**, (2020).

8. Faizah, A. N. *et al.* Deciphering the Virome of Culex vishnui Subgroup Mosquitoes, the Major Vectors of Japanese Encephalitis, in Japan. *Viruses* **12**, (2020).

9. de Oliveira Ribeiro, G. *et al.* Guapiaçu virus, a new insect-specific flavivirus isolated from two species of Aedes mosquitoes from Brazil. *Sci. Rep.* **11**, 4674 (2021).

10. Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C. & Junglen, S. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7536–7541 (2015).

11. Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* **540**, 539–543 (2016).

12. Halbach, R., Junglen, S. & van Rij, R. P. Mosquito-specific and mosquito-borne viruses: evolution, infection, and host defense. *Curr Opin Insect Sci* **22**, 16–27 (2017).

13. Ajogbasile, F. V. *et al.* Real-time Metagenomic Analysis of Undiagnosed Fever Cases Unveils a Yellow Fever Outbreak in Edo State, Nigeria. *Sci. Rep.* **10**, 3180 (2020).

14. Matranga, C. B. *et al.* Unbiased Deep Sequencing of RNA Viruses from Clinical Samples. *J. Vis. Exp.* (2016) doi:10.3791/54117.

15. Auguste, A. J. *et al.* A newly isolated reovirus has the simplest genomic and structural organization of any reovirus. *J. Virol.* **89**, 676–687 (2015).

16. Amoa-Bosompem, M. *et al.* Entomological Assessment of the Status and Risk of Mosquito-borne Arboviral Transmission in Ghana. *Viruses* **12**, (2020).

17. Ghabrial, S. A. & Nibert, M. L. Victorivirus, a new genus of fungal viruses in the family Totiviridae. *Arch. Virol.* **154**, 373–379 (2009).

18. Stollar, V. & Thomas, V. L. An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. *Virology* **64**, 367–377 (1975).

19. Cammisa-Parks, H., Cisar, L. A., Kane, A. & Stollar, V. The complete nucleotide sequence of cell fusing agent (CFA): homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. *Virology* **189**, 511–524 (1992).

20. Cook, S. *et al.* Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. *J. Gen. Virol.* **87**, 735–748 (2006).

21. Espinoza-Gómez, F. *et al.* Detection of sequences from a potentially novel strain of cell fusing agent virus in Mexican Stegomyia (Aedes) aegypti mosquitoes. *Arch. Virol.* **156**, 1263–1267 (2011).

22. Hoshino, K., Isawa, H., Tsuda, Y., Sawabe, K. & Kobayashi, M. Isolation and characterization of a new insect flavivirus from Aedes albopictus and Aedes flavopictus mosquitoes in Japan. *Virology* **391**, 119–129 (2009).

23. Kihara, Y. *et al.* Rapid determination of viral RNA sequences in mosquitoes collected in the field. *J. Virol. Methods* **146**, 372–374 (2007).

24. Yamanaka, A., Thongrungkiat, S., Ramasoota, P. & Konishi, E. Genetic and evolutionary analysis of cell-fusing agent virus based on Thai strains isolated in 2008 and 2012. *Infect. Genet. Evol.* **19**, 188–194 (2013).

25. Baidaliuk, A. *et al.* Cell-Fusing Agent Virus Reduces Arbovirus Dissemination in Aedes aegypti Mosquitoes In Vivo. *J. Virol.* **93**, (2019).

26. Yamao, T. *et al.* Novel virus discovery in field-collected mosquito larvae using an improved system for rapid determination of viral RNA sequences (RDV ver4.0). *Arch. Virol.* **154**, 153–158 (2009).

27. Schultz, M. J., Frydman, H. M. & Connor, J. H. Dual Insect specific virus infection limits Arbovirus replication in Aedes mosquito cells. *Virology* **518**, 406–413 (2018).

28. Cunha, M. D. P. *et al.* A Metagenomic Approach Identified a Novel Phasi Charoen-Like Virus Coinfecting a Chikungunya Virus-Infected Aedes aegypti Mosquito in Brazil. *Microbiol Resour Announc* **9**, (2020).

29. Nibert, M. L. *et al.* Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. *Virus Res.* **188**, 128–141 (2014).

30. Pettersson, J. H.-O. *et al.* Characterizing the virome of Ixodes ricinus ticks from northern Europe. *Sci. Rep.* **7**, 10870 (2017).

31. Faizah, A. N. *et al.* Deciphering the Virome of Culex vishnui Subgroup Mosquitoes, the Major Vectors of Japanese Encephalitis, in Japan. *Viruses* **12**, (2020).

32. Öhlund, P., Lundén, H. & Blomström, A.-L. Insect-specific virus evolution and potential effects on vector competence. *Virus Genes* **55**, 127–137 (2019).

33. Cook, S. *et al.* Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* **8**, e80720 (2013).

34. Li, C.-X. *et al.* Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* **4**, (2015).

35. Bolling, B. G., Eisen, L., Moore, C. G. & Blair, C. D. Insect-specific flaviviruses from Culex mosquitoes in Colorado, with evidence of vertical transmission. *Am. J. Trop. Med. Hyg.* **85**, 169–177 (2011).

36. Roiz, D., Vázquez, A., Seco, M. P. S., Tenorio, A. & Rizzoli, A. Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. *Virol. J.* **6**, 93 (2009).

37. Booman, M. *et al.* Using a geographical information system to plan a malaria control programme in South Africa. *Bull. World Health Organ.* **78**, 1438–1444 (2000).

38. Edwards, Wallace, F. & British Museum (Natural History). *Mosquitoes of the Ethiopian region. Vol. III, Vol. III,*. (British Museum, 1941).

39. Gillett, D., J., Smith & G., J. *Common African mosquitos and their medical importance*. (Heinemann Medical, 1972).

40. Gillies, T., M., Coetzee & Maureen. *A supplement to the anophelinae of Africa south of the Sahara (Afrotropical Region)*. (South African Institute for Medical Research, 1987).

41. Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161**, 1516–1526 (2015).

42. Park, D. et al. broadinstitute/viral-ngs: v1.19.2. Zenodo https://zenodo.org/record/1167849 (2018).

43. Oluniyi, P. E. *et al.* VGEA: an RNA viral assembly toolkit. *PeerJ* **9**, e12129 (2021).

44. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).

45. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).

46. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

Acknowledgements

This work is made possible by support from ACEGID lab and a cohort of generous donors through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott, the Skoll Foundation, and Open Philanthropy. This work was supported by grants from the National Institute of Allergy and Infectious Diseases (https://www.niaid.nih.gov), NIH-H3Africa (https://h3africa.org) (U01HG007480 and U54HG007480 to Christian T Happi), the World Bank grant (worldbank.org) (project ACE019 to Christian T Happi). We would like to appreciate Clementina Chekwube Anioke, Francis Uzoma Orizu, Esther Ngozi Onuora, Linda Chioma Ikechukwu, and Oscar Nnanna Nwaogo for their assistance during sample collections.

Author Contributions

J.U.O.-performed NGS on samples, initial data review and bioinformatics analysis and wrote the manuscript, P.E.O. and U.E.G.- carried out bioinformatics analysis and wrote the manuscript. T.J.O., B.E.B. and F.O.B.- performed NGS and contributed to the first draft of the manuscript. A.K.- performed the statistical analysis. U.C.N., C.C.A. F.A.D., C.O.O., E.K.E., N.O.A., E.I.E., S.O.A., A.I.O., E.M.N., S.O.E., G.K.O, N.E.U, J.M and I.M.N. collected, identified and preserved the mosquito samples; C.O.M., C.I., O.C.C., C.C.A. and J.C.O supervised sample collection, identification and preservation of the mosquito samples. O.A.F., and I.O.O.K reviewed the manuscript. C.T.H.- Conceived the study, supervised the NGS experiment, data analysis and discussions and reviewed the manuscript. All authors contributed to the final version of the manuscript.

Competing Interests: The authors declare no competing interests

Additional Information

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx