METAGENOMIC ANALYSIS OF THE RNA METAVIROME OF *Aedes aegypti* MOSQUITOES COLLECTED FROM COASTAL KENYA

DONWILLIAMS OTIENO OMUOYO

MASTER OF SCIENCE

(Molecular Biology and Bioinformatics)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2024

Metagenomic Analysis of the RNA Metavirome of *Aedes aegypti* Mosquitoes Collected from Coastal Kenya

Donwilliams Otieno Omuoyo

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Molecular Biology and Bioinformatics of the Jomo Kenyatta University of Agriculture and Technology

2024

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature …….. Date…….

Donwilliams Otieno Omuoyo

This thesis has been submitted for examination with our approval as university supervisors.

Signature …….. Date…….

Prof. Johnson Kinyua, PhD

JKUAT, Kenya

Signature …….. Date…….

Prof. George Mbugua Warimwe, PhD

KWTRP, Kenya

DEDICATION

This work is dedicated to my parents, Mr. and Mrs. Ogutu, and uncle and mentor Mr. Sam Omondi Ogutu, for seeing me through the tough times over the years. Your support, inspiration and love will forever be engraved in my heart.

ACKNOWLEDGEMENT

I would like to express my utmost gratitude to God for the strength, energy, motivation, perseverance, and providence throughout the phases of this study and life in general. I wish to also extend my sincerest gratitude to my supervisors: Professor Johnson Kinyua of JKUAT and Professor George Warimwe of KWTRP. Your great stewardship, support and commitment are what led to the successful completion of this project.

Additionally, I would like to thank KWTRP's arbovirus, virus epidemiology and control (VEC), entomology and sequencing groups including George, James, Doris, Bernadette, John, Henry, Daisy, Charles, Evelyne, Mwangangi, Martin, Karisa, Chome and Zaydah. Your unwavering support, motivation, and positive critique throughout the course of this study will not be forgotten.

To my friends and JKUAT colleagues; Shaban, Arnold, Peter, Bungei and Mwita. *Aluta continua* is still our motto. Thank you for the many meaningful discussions and insights.

Lastly, I would like to thank JKUAT for giving me the opportunity to pursue this course.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATION AND ACRONYMS

cDNA	Complimentary Deoxyribonucleic Acid
CFAV	Cell Fusing Agent Virus
CHIKV	Chikungunya Virus
CxFV	*Culex* Flavivirus
DENV	Dengue Virus
dscDNA	Double Stranded Complimentary Deoxyribonucleic Acid
EILV	Eilat Virus
EIP	Extrinsic Incubation Period
HTV	Humaita-Tubiacanga Virus
ISV	Insect-Specific Viruses
JEV	Japanese Encephalitis Virus
KRV	Kamiti River Virus
MEV	Murray Encephalitis Virus
MSV	Mosquito Specific Viruses
NHUV	Nhumirim Virus
PCLV	Phasi Charoen-Like Phasivirus
RRV	Ross River Virus
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction

RT-qPCR	Reverse Transcriptase Real Time Polymerase Chain Reaction
RVFV	Rift Valley Fever Virus
SLEV	St Louis Encephalitis Virus
WNV	West Nile Virus
YFV	Yellow Fever Virus
ZKV	Zika Virus

ABSTRACT

The coastal region of Kenya has experienced repeated outbreaks of chikungunya (CHIKV) and dengue (DENV) viruses transmitted by local populations of the *Aedes aegypti* mosquito. In addition, this mosquito has been reported to harbor insect specific viruses (ISVs), some of which can either enhance or suppress arboviral transmission. Despite the arboviral burden in coastal Kenya, there has been no systematic molecular entomological surveillance in this region and the viral diversity of local *Aedes aegypti* populations remains largely unknown. To understand the *Aedes aegypti* virome from coastal Kenya sampled outside outbreaks, this study conducted a retrospective study of mosquitoes collected during a sectional survey to determine their urban ecology. Adult mosquitoes were lured using Biogent's sentinel traps at 16 different localities along the Kenyan coast between May to September 2017. Pools of 20 female *Aedes aegypti* mosquitoes were generated following grouping by morphological characteristics. Presence of arboviruses in the mosquito pools was determined using virus-specific and genera-specific primers on the real-time PCR. Metagenomic next generation sequencing (mNGS) on Illumina Miseq and analysis was used to characterize the virome. A total of 16,520 female *Aedes aegypti* grouped into 826 pools were analyzed. Flaviviruses were detected in 69/826 (8.4%) pools by real time PCR. Sequencing generated 8,459/971,754 (0.87%) clean reads that were taxonomically assigned to 16 and 28 viral families and species, respectively. The family *Phenuiviridae* represented by Phasi Charoen-like phasivirus (PCLV) species was the most abundant, detected in 64/73 (87%) mosquito pools. No pathogenic viruses were identified by mNGS. Phylogenetic analysis revealed local PCLV and Cell fusing agent virus (CFAV) were distinct from global sequences. Our data provides information about virus diversity and composition of the *Aedes aegypti* mosquitoes from coastal Kenya and contributes to the body knowledge of *Aedes aegypti* virome. To the best of our knowledge, this is the first study to provide this information from this region. Future studies should sample longitudinally across the country or region during and between arboviral outbreaks so as to gain understanding of the temporal viral diversity in local or regional *Aedes aegypti* populations.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

The *Aedes aegypti* mosquito transmits viruses of human importance such as chikungunya (CHIKV), dengue (DENV), yellow fever (YFV) and Zika (ZIKV) (Souza-Neto et al., 2019). The first documented DENV outbreak in Kenya (caused by DENV serotype 2) occurred in Malindi in 1982 and after a 30-year quiescence, emerged in Mandera and Mombasa between 2011 and 2014. Mombasa alone (in 2013) accounted for 58% (154/267) of the suspected hospital cases confirmed by RT-PCR (Ellis et al., 2015). In addition, a serosurvey from this period conducted in Nairobi, and in the northern and coastal regions showed 40% (345/868, either RT-PCR or IgM ELISA positive) of febrile patients were exposed to either DENV1, 2 or 3 (Konongoi et al., 2016). Subsequent outbreaks were reported in Mandera, Wajir, Kilifi, Malindi and Mombasa, respectively, in 2017 (Konongoi et al., 2016; WHO, 2017). The first notable CHIKV outbreak in Kenya occurred in 2004 in Lamu, resulting in an estimated 13,500 cases (Sergon et al., 2008). CHIKV then emerged over a decade later causing an outbreak in 2016 in Mandera, where 1,792 human cases were reported (WHO, 2016). Since these reports, CHIKV and DENV have continued to cause spontaneous outbreaks in different parts of the country.

The *Aedes aegypti* has a wide geographical distribution attributed to increased urbanization, globalization, and favorable climatic factors (Weaver, 2014). In Kenya, *Aedes aegypti* is abundant in the lowlands, with the coastal region reporting the presence of both the domestic and sylvatic forms (Lounibos & Kramer, 2016; Lutomiah et al., 2013; Mwangangi et al., 2012; Powell & Tabachnick, 2013). In recent times, following the distribution of *Aedes aegypti* and *Aedes albopictus* vectors, DENV and CHIKV have expanded their geographical territories and increased their abundance (Weaver, 2014). In Kenya, CHIKV and DENV have caused repeated outbreaks in the Northeastern and coastal regions (Sergon et al., 2008; Konongoi et al., 2016; WHO, 2017).

The successful transmission of arboviruses and other viruses in mosquitoes is affected or driven by extrinsic and intrinsic factors (Kramer & Ciota, 2015) that modulate vector competence - the ability of mosquitoes to allow infection, replication, and transmission of pathogen (Agboli et al., 2019; Agha et al., 2017; Aliota, et al., 2016; Chepkorir et al., 2014; Gloria-soria et al., 2017; Hobson-Peters et al., 2013). Rainfall for example, affects mosquito habitats, the likelihood of mosquito survival and adult mosquito densities by interfering with time-to-larval development (Souza-Neto et al., 2019). Temperature affects extrinsic incubation period (EIP) - time in days a pathogen takes to infect the mosquito midgut and appear in mosquito saliva (Souza-Neto et al., 2019). On the other hand, intrinsic factors such as the genetically regulated physiological barriers; the midgut infection barrier, which inhibits infection and replication of pathogens, and midgut escape barrier, which inhibits distribution of pathogens to other tissues, affect EIP (Bennett et al., 2002). The abundance of these barriers has been reported to vary in field *Aedes aegypti* populations and has particularly led to differential vector competence for DENV (Bennett et al., 2002).

A few studies have demonstrated the vector competence of *Aedes aegypti* mosquitoes collected from different regions in Kenya. Specifically, the impact of temperature and viral dose on competence have been investigated by comparing mosquitoes collected from inland regions of the country (Nairobi, Kisumu and Trans Nzoia) to those from the coastal region (Lamu, Mombasa and Kilifi), and found the latter were more competent to transmitting CHIKV and DENV (Agha et al., 2017; Chepkorir et al., 2014; Mbaika et al., 2016). These findings could explain why coastal Kenya experiences recurrent outbreaks of CHIKV and DENV viruses.

In addition to arboviruses, mosquitoes harbor other viruses that are incapable of replicating in vertebrate cells or tissues known as insect specific viruses (ISVs) (Bolling et al., 2012). The first generation of ISVs were discovered through culture-based methods. In recent times, however, the advent and application of metagenomics next generation sequencing (mNGS) in mosquitoes, has led to the discovery of novel ISVs (Agboli et al., 2019). Continuous detection of previously identified ISVs, in addition to novel ones, in mosquitoes has sparked research

interest with regards to how ISVs interact with arboviruses (Agboli et al., 2019). Several studies have demonstrated the ability of some ISVs to impair the transmission of specific arboviruses (Baidaliuk, et al., 2019; Bolling et al., 2012; Hobson-Peters et al., 2013; Kent et al., 2010; Koh et al., 2021; Talavera et al, 2018), although the mechanism through which ISVs modulate arboviruses is not clearly known. The ability of ISVs to prevent arboviral transmission has been shown to be arbovirus-specific (Baidaliuk et al., 2019; Kenney et al., 2014; Kent, et al., 2010; Koh et al., 2021; Nasar et al., 2015; Romo et al., 2018; Talavera et al., 2018). However, no single ISV capable of preventing the transmission of all the common arboviruses of veterinary and public importance was identified.

Despite a significant arbovirus burden in coastal Kenya, there has been no systematic molecular entomological surveillance and thus knowledge of the viral diversity, and the role of ISVs in modulating arboviruses in local *Aedes aegypti* populations remains largely unknown. This study leveraged on the unbiased and high throughput mNGS strategy to query the RNA virome of archived field *Aedes aegypti* mosquitoes collected from several localities within Kilifi, Kwale and Mombasa counties in coastal Kenya.

1.2 Statement of the Problem

The definitive global cumulative burden of arboviral diseases is unknow (LaBeaud et al., 2011). This is due to lack of arboviral risk awareness, unsystematic reporting of cases, self-limiting nature of infections and ineffective diagnosis in most endemic countries (Iwashita et al., 2018; Massengo et al., 2023). Nevertheless, it is estimated that arboviruses-related diseases claim 700,000 lives across the world every year (Ketkar et al., 2019). CHIKV has had 500,000 cases and 400 deaths globally (ECDC, 2024), DENV 390,000,000 cases and 36,000 deaths globally (Bhatt et al., 2013), YFV 200,000 cases and 30,000 deaths globally (Mutebi & Barrett, 2002) and ZIKV 900,000 cases and 51 deaths (Cardona-Ospina et al., 2019). While the overall global economic burden of arboviruses is unknown that of DENV and YFV is estimated at $2.1 and $57.3 billion with *Aedes* spp vector control strategies costing $5.62 to $73.5 million (Thompson, 2020). In the absence of effective licensed vaccines and

antiviral drugs, it is expected cases of exacerbated disease, disability-adjusted life years (DALY), lost productivity and mortality will increase further worsening the economic situation (Massengo et al., 2023). In Kenya, the Coastal and Northeastern regions have reported repeated outbreaks of CHIKV and DENV transmitted by competent *Aedes aegypti* mosquitoes. DENV emerged between 2011-2014 in Mandera and Mombasa, where 154 cases were confirmed by PCR in Mombasa (Ellis et al., 2015) and a contemporary serosurvey of these regions and Nairobi confirming 345 cases by laboratory tests (Konongoi et al., 2016). Following the biggest outbreak in Lamu with 13,500 cases and 75% seroprevalence in 2004, CHIKV occurred in Mandera affecting 1,792 people (Sergon et al., 2008; Sergon et al., 2007; WHO, 2016). Despite the disease and economic burden due to recurrent outbreaks of CHIKV and DENV, there is no regular vector surveillance in coastal Kenya, especially outside epidemics, to guide in the control and prevention of potential outbreaks.

1.3 Justification of the Study

Mosquito metagenomic surveillance offers an unbiased approach for viral detection and discovery. For example, metagenomic sequencing successfully detected pathogenic arboviruses including DENV, Japanese Encephalitis Virus (JEV), Ross River Virus (RRV) and ZIKV, in addition to a novel variant of ZIKV in mosquitoes (Batovska et al., 2017; He et al., 2021; Thannesberger et al., 2021; Xiao et al., 2018). These and other studies demonstrate the utility of metagenomic sequencing in viral surveillance and also underscore the need for continuous molecular surveillance of mosquitoes. In Kenya, however, metagenomic studies of mosquitoes of human importance are scarce. Hitherto, only one study has utilized metagenomics to characterize the virome of *Culex quinquefasciatus* and *Culex tritaeniorhynchus* mosquitoes virome from Kwale in coastal Kenya, where a rich viral diversity was observed (Atoni et al., 2018). No metagenomic studies of *Aedes aegypti* from coastal Kenya have been conducted. Thus, knowledge of the *Aedes aegypti* virome and diversity will increase our understanding of the circulating viruses and the potential for arboviral transmission. Additionally, knowledge of the core virome, including

ISVs that have previously been reported to modulate arboviruses, may inform further studies of interventions against medically important arboviruses.

1.4 Research Questions

1. What medically important arboviruses are present in *Aedes aegypti* populations from coastal Kenya sampled outside arboviral outbreaks?
2. What is the profile of viruses present in the *Aedes aegypti* populations from coastal Kenya sampled outside arboviral outbreaks?
3. What is the evolutionary relationship of viruses present in the *Aedes aegypti* populations from coastal Kenya sampled outside arboviral outbreaks?

1.5 Hypothesis

Null Hypothesis

The *Aedes aegypti* mosquitoes sampled outside arboviral outbreaks from the coastal counties of Kilifi, Kwale and Mombasa do not have high viral diversity.

1.6 Objectives

1.6.1 General Objective

To conduct metagenomic analysis of the RNA metavirome of *Aedes aegypti* mosquitoes collected from coastal Kenya.

1.6.2 Specific Objectives

1 To determine the abundance of mosquito-borne viruses in *Aedes aegypti* mosquito pools sampled outside arboviral outbreaks from Kilifi, Kwale and Mombasa counties in coastal Kenya.

2 To identify the viruses circulating in *Aedes aegypti* mosquito pools sampled outside arboviral outbreaks from Kilifi, Kwale and Mombasa counties in coastal Kenya.

3 To determine the evolutionary relationship of viruses in *Aedes aegypti* mosquito pools sampled outside arboviral outbreaks from Kilifi, Kwale and Mombasa counties in coastal Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 *Aedes aegypti* Taxonomy, Ecology and Life Cycle

The *Aedes aegypti* mosquito (*Ae. aegypti*) family *Culicidae* genus *Aedes* was identified and named by Linnaes in 1762 (Powell & Tabachnick, 2013). It is of global public health importance because it is the primary vector of human pathogenic viruses including CHIKV, DENV, YFV and ZIKV (Souza-Neto et al., 2019). Two subspecies exist in Africa: *Ae*. *aegypti aegypti* (*Aaa*) and *Ae*. *aegypti formosus* (*Aaf)* with the former being paler than the latter (Powell & Tabachnick, 2013).

Aedes aegypti is an urban mosquito that survives in humid and warm climate, preferentially feeds on humans during the day, lays multiple eggs in batches after a blood meal and hatches its eggs during floods and high temperature seasons (Lwande et al., 2020). The larvae take approximately 10 days to molt into pupae (males molt faster than females) and 2 more days to become adults after which they live for 2-4 weeks depending on conditions. The survival of this mosquito species has partly been attributed to its ecological plasticity including egg diapause, opportunistic biting habits and flexibility in breeding in both natural and artificial spots (Shi et al., 2019).

2.2 *Aedes aegypti* Distribution

The *Aedes aegypti* is native to Africa and its global distribution has been attributed to increased urbanization, international trading and travel and favourable climatic conditions (Moore et al., 2013). Despite their different historical backgrounds, the *Aedes aegypti* together with its closest relative *Aedes albopictus* are widespread within the tropical, sub-tropical and temperate regions of the world (Houé et al., 2019). In Kenya, *Aedes aegypti* is widely dispersed with a high abundance in lowland regions (Lutomiah et al., 2013). Coastal Kenya's tropical climate provides an ideal environment for its survival and breeding. Both the sylvatic and domestic forms of *Aedes aegypti* have been found to co-exist in Rabai, with the domestic form

being found both indoors and outdoors in Malindi and Mombasa, respectively (Lounibos & Kramer, 2016; Mwangangi, et al., 2012; Powell & Tabachnick, 2013).

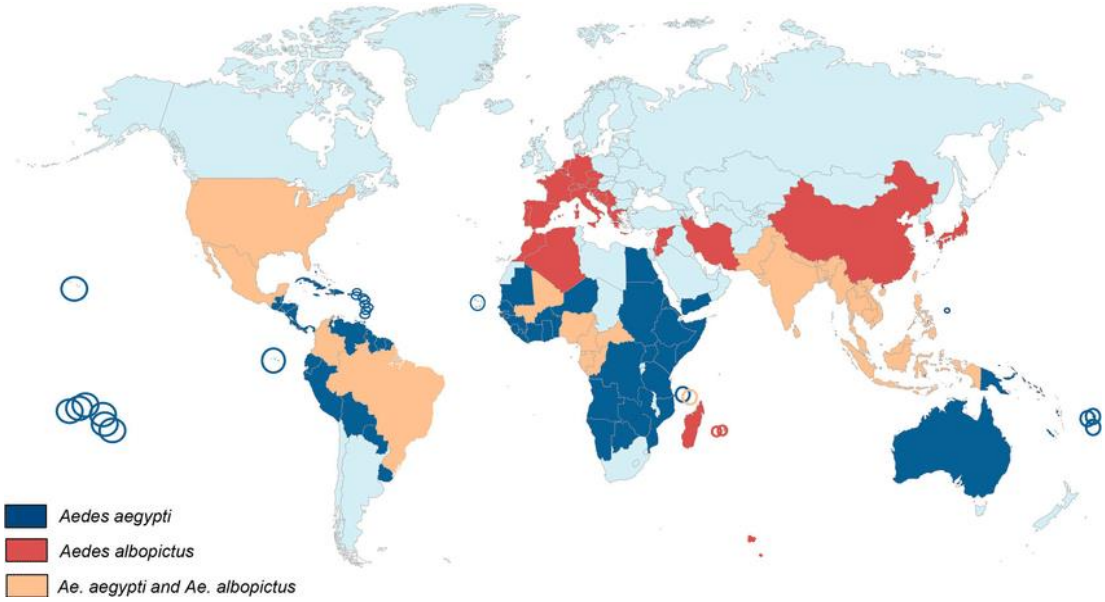

Figure 2.1: The Global distribution of *Aedes aegypti and Aedes albopictus.* Countries reporting mono-circulation of *Aedes aegypti* and *Aedes albopictus* are colored in blue and red, respectively, those reporting co-circulation in cream while those without these two mosquito species in turquoise (Houé et al., 2019).

2.3 Vector Competence of *Aedes aegypti* from Kenya

The *Aedes aegypti* and *Aedes albopictus* are competent vectors of CHIKV, DENV, YFV and ZIKV. Recently, DENV and CHIKV viruses in particular have expanded their geographical boundaries and increased their abundance (Weaver, 2014). The dispersal of these viruses likely follows the natural distribution of *Aedes aegypti* and *Aedes albopictus* vector populations. Successful transmission of arboviruses by mosquitoes depends on both intrinsic and extrinsic factors affecting the vector and virus. Temperature, an extrinsic factor, affects survival of mosquitoes, adult population densities by controlling the time-to-larval stage development and EIP (Souza-Neto et al., 2019). In Kenya, studies have investigated the impact of temperature and viral dose on vector competency of *Aedes aegypti*. Mosquitoes from

Nairobi and Kilifi reared at 30°C had higher midgut infection rates of DENV-2 compared to when reared at 26°C (Chepkorir et al., 2014). Similarly, *Aedes aegypti* from Lamu had higher mid gut infection and dissemination rates of CHIKV than those from Trans Nzoia at both 26°C and 32°C, with Trans Nzoia mosquitoes showing increased infection and dissemination rates when reared at 32°C (Mbaika et al., 2016). Viral dose has been found to be positively correlated to rate of infection (Souza-Neto et al., 2019). Increased doses of CHIKV in blood meal led to higher infection, dissemination, and transmission rates in *Aedes aegypti* from Nairobi, Kisumu and Mombasa compared to lower doses (Agha et al., 2017). Intrinsic factors like vector genetics likely influence competence. It is known that *Aedes aegypti* populations from different geographical backgrounds are genetically diverse, increasing the likelihood they will have varying vectorial competencies and susceptibilities to arboviruses (Gloria-soria et al., 2017). However, how the genetic diversity that affects vector competency varies among different mosquito populations remains to be elucidated.

Noteworthy, is the role of host microbiota on vector competence. The entomopathogenic fungi *Beauveria bassiana* and *Metarhizium anisopliae* kill larvae and adults of *Anopheline*, *Culex* and *Aedes spp* mosquitoes (Farenhorst et al., 2009). The wMEL strain of *Wolbachia* causes cytoplasmic incompatibility (CI) in *Aedes aegypti*, further limiting infection and transmission of DENV, CHIKV and ZIKV (Aliota et al., 2016; Aliota et al., 2016; Frentiu et al., 2014). On the contrary, *Serratia marcescens* produces the protein enhancin, which degrade membrane-bound mucins, enabling infection of mosquito midgut by DENV (Miesen & van Rij, 2019). Taken together, vector competence is a complicated and evolving phenotype, modulated by the tri-partite interaction between the vector, pathogen and vector-resident microbiota.

2.4 Prevalent Arboviruses Transmitted by *Aedes aegypti* in Kenya

Given the demonstrable vector competence of *Aedes aegypti*, Kenya has concomitantly experienced recurrent outbreaks of CHIKV and DENV viruses, with most cases reported in the Northeastern and Coastal regions. CHIKV (family

Togaviridae genus *Alphavirus*) is a positive sense single stranded RNA virus that is 12 kb long, which causes Chikungunya disease (CHIKVD) characterized by fever, rash and severe arthralgia, which is the hallmark of infection (Volk et al., 2010). Native to East Africa in Tanzania where it was initially isolated in the early 1950s, CHIKV spread and caused random outbreaks mediated by *Aedes aegypti* in Africa and Asia in the 1960s (Houé et al., 2019). The biggest outbreak in East Africa emerged in Lamu Kenya in 2004 then spread into the Indian ocean islands (Comoros, Reunion, Madagascar and Seychelles) and India between 2004-2007 where it was estimated that hundreds of thousands and 60% of the islands' inhabitants were affected and exposed to the virus, respectively (Sergon et al., 2008; Njenga et al., 2008; Sergon et al., 2007). In India an estimated 1.3 million people in 13 states were infected (Arankalle et al., 2007). Since 2006, CHIKV has been introduced to Europe and the Americas by viremic travelers from India and continues to spread (Volk et al., 2010).

DENV (family *Flaviviridae* genus *Flavivirus*) is a positive sense single stranded RNA virus that is 12 kb long (Pollett et al., 2019). It exists as four serotypes (1-4), which cause dengue fever (DF) and dengue hemorrhagic fever (DHF) that affects hundreds of millions of people resulting in thousands of deaths across the world (Bhatt et al., 2013). In Africa, DENV-1 and 2, and 3 were first isolated in Nigeria and Mozambique in 1968 and 1984, respectively (Shah et al., 2020). Since the first laboratory-confirmed DENV case in Kenya in 1982 in the coastal region, several outbreaks of DENV-1,2 and 3 have been reported in the northern and coastal parts of the country between 2011-2014 and 2017, infecting several hundred people (Ellis et al., 2015; Konongoi et al., 2016; Masika et al., 2020; Obonyo et al., 2018; Pollett et al., 2019; Shah et al., 2020).

Since these reports, these arboviruses have continued to cause spontaneous outbreaks in several parts of the country.

2.5 Metagenomics and Mosquito Virome Studies

Besides arboviruses, mosquitoes naturally harbor insect specific viruses (ISVs). ISVs have sparked attention in recent times with regards to how they interact with

pathogenic viruses in co-infected mosquitoes. The first generation of ISVs were discovered through culture-based methods and Cell-fusing agent virus (CFAV) is the first ISV (family *Flaviviridae*, genus *Flavivirus)* to be isolated and characterized in *Aedes aegypti* cell line in 1975 followed by later detection in *Aedes aegypti* natural populations (Agboli et al., 2019). The second ISV to be discovered using culture methods was Kamiti River Virus (KRV) in 1999 from field *Aedes macintoshi* mosquitoes collected from central Kenya (R. C. Sang et al., 2003). However, the last decade has seen an increase in the discovery of ISVs through the molecular-based metagenomic next generation sequencing (mNGS) method.

Metagenomic sequencing is an indispensable tool in vector surveillance of arboviruses, understanding viral diversity and virus discovery (Agboli et al., 2019). In particular, recent mNGS studies of *Aedes aegypti* and *Aedes albopictus* viromes sampled from Africa, Asia, Australia, the Americas and Europe have revealed presence of diverse ISVs (Gomez, 2022). A majority of ISVs have been classified into various families based on their characteristics and evolutionary relationships including *Flaviviridae, Togaviridae, Bunyaviridae and Mesoniviridae* (Agboli et al., 2019). Although the viral diversity in *Aedes aegypti* varies by geographical location, some viruses are core. For example, field *Aedes aegypti* from Australia, Brazil, China, India, Grenada, Thailand and Nigeria have a high abundance of Phasi Charoen-like phasivirus (PCLV, family *Phenuiviridae*) (Duarte et al., 2022; Gangopadhayya et al., 2024; He et al., 2021; Oguzie et al., 2022; Ramos-Nino et al., 2020; Zakrzewski et al., 2018). Phylogenetic studies of *Aedes aegypti* mitochondrial genes show that this mosquito originated from either of two maternal lineages in Africa (Ramos-Nino et al., 2020). The prevalence of PCLV in mosquitoes from different geographical regions outside Africa suggests that this virus likely infected the ancestral mosquito species. Other abundant ISVs include Humaita-Tubiacanga (HTV) and CFAV (Gomez, 2022). It is likely that the core viruses of *Aedes aegypti* are vertically transmitted (Shi et al., 2019, 2020b, 2020a).

2.6 Insect Specific Viruses and their Interactions with Arboviruses

The abilities of some ISVs to suppress the replication of arboviruses has been demonstrated. While initial infection of *Culex pipiens* mosquitoes with Culex flavivirus (CxFV) followed by West Nile Virus (WNV) suppressed WNV replication, no impact was observed when the experiment was replicated in *Culex quinquefasciatus* mosquitoes and when CxFV-infected *Culex pipiens* were infected with Rift Valley fever virus (RVFV) (Bolling et al., 2012; Kent et al., 2010; Talavera et al., 2018). Prior infection of *Aedes aegypti* with wild type CFAV followed by DENV-1 and ZIKV reduced dissemination of these viruses in the head tissues, both in-vitro and in-vivo (Baidaliuk et al., 2019). Infection of *Aedes albopictus* C6/36 cells first with Palm Creek Virus (PCV) then with either West Nile virus (WNV) or Murray Encephalitis virus (MEV) suppressed the replication of these arboviruses but did not affect Ross River virus (RRV) (Hobson-Peters et al., 2013). Interestingly, infection of *Aedes aegypti* and *Aedes albopictus* mosquitoes with PCV followed by infection with CHIKV and ZIKV did not hamper infection (Koh et al., 2021). Prior infection of *Aedes albopictus* C6/36 cells with Nhumirim virus (NHUV) then with either WNV, St Louis Encephalitis virus (SLEV), Japanese Encephalitis virus (JEV), ZIKV or DENV-2 reduced viral titers but did not affect CHIKV (Kenney et al., 2014; Romo et al., 2018). Similarly, initial infection of *Aedes albopictus* C7/10 cells and *Aedes aegypti mosquitoes* with Eilat virus (EILV) reduced viral load of CHIKV and delayed distribution of CHIKV, respectively (Nasar et al., 2015).

Despite a high abundance of *Aedes aegypti* in coastal Kenya, no metagenomic studies of this mosquito have been conducted. Whether or not these mosquitoes contain the abundant ISVs or those implicated in the suppression of arboviruses remains to be investigated.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Area and Site Selection

Mosquitoes were sampled from 16 different localities spanning Kilifi (3.5107° S; 39.9093° E), Kwale (4.1816° S; 39.4606° E) and Mombasa (4.0435° S; 39.6682°E) counties along the Kenyan coast between May 2017 to September 2017. These counties were chosen because of their history of arbovirus outbreaks (Chepkorir et al., 2014; Ellis et al., 2015; Eyase et al., 2020; Iwashita et al., 2018; Pollett et al., 2019; Villinger et al., 2017).

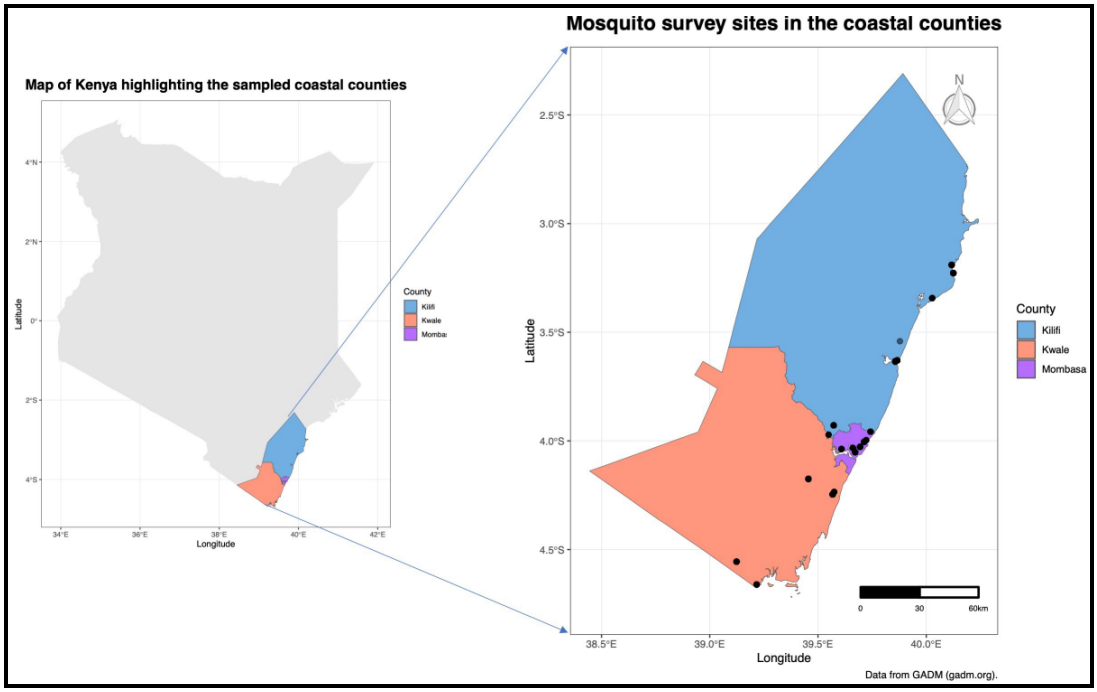

Figure 3.1: Map of *Aedes aegypti* collection sites. The map of Kenya is shown on the left with the coastal counties under study highlighted. A zoom in of the coastal counties is shown on the right with the sampling localities within each county indicated with black dots.

The ecologies of Kilifi, Mombasa and Kwale counties along the coastal region of Kenya have been described previously (Atoni et al., 2018; Karisa et al., 2021). Kilifi County sites included Kilifi town, Malindi town, Mazeras, Rabai and Watamu. Sites

in Mombasa County included Bamburi, Kisauni, Mombasa town, Portreiz, Shanzu, Tononoka and Tudor. Kwale County was comprised of Kwale town, Lunga-Lunga, Ukunda and Vanga sites. All study sites were mapped by ascertaining their respective geographical position system (GPS) coordinates using the Garmin Etrex 10 GPS gadget (Garmin Ltd, USA) fitted on Biogents sentinel (BG) traps (Biogents AG, Germany) used to trap *Aedes aegypti* mosquitoes. The national and regional boundaries for Kenya were downloaded from the open-source online platform GADM, data version 3.6 accessible at (https://gadm.org/download_country36.html) and used to generate a map of the sampling sites highlighted in Figure 3.1.

3.2 Mosquito Collection and Identification

This study was a retrospective cross-sectional survey of adult *Aedes aegypti* mosquitoes collected from their natural habitat in Kilifi, Kwale and Mombasa counties. Ethical approval was provided by the Kenya Medical Research Institute Scientific and Ethics Review Unit (KEMRI SERU #3296, SSC Protocol 2675). Given the difficulty in ascertaining adult mosquito population densities due to the impact of ecological and human factors, larvae densities from a previous entomological survey of the same urban sites in Kilifi, Malindi and Mombasa (Karisa et al., 2021) were used as proxy for predicting the potential adult densities. Number of traps per sampling sites (D) were calculated based on the formula $D = 1+(m-1)\rho$, where m is average number of positive mosquito traps per site and ρ is the coefficient of correlation between sites. An $m = 30$ and $\rho = 0.9$ were used and 27 traps were set per site and sampled successively. Major urban areas within Kilifi, Kwale and Mombasa counties were purposely selected based on potential mosquito breeding sites.

A total of 16,520 adult *Aedes aegypti* mosquitoes were sampled using BG trap lured with carbon dioxide in the form of dry ice between 0600 hrs to 1700 hrs as previously described (Karisa et al., 2021). Briefly, the traps were randomly set outdoors approximately 100 m from each other, lured with ~3 kg of dry ice as the source of carbon dioxide and allowed to operate from 0600 hrs to 1700 hrs. In the evening, the collected samples were retrieved from the traps, placed in cool boxes

with ice packs and transported to the KEMRI-Wellcome Trust Research laboratories, Kilifi for further analysis. In the laboratory, the mosquitoes were knocked down by freezing at -20°C for two minutes, sorted, identified morphologically using mosquito identification keys (Gillies & De Meillon, 1968; Iv Edwards, 1941) and then pooled on a chill table (maximum of 20 mosquitoes per pool) by trap, species, sex, collection date and sites to generate 826 pools. All the samples were preserved in 1.5 ml cryogenic vials at -80°C awaiting further analysis.

3.3 Total Ribonucleic Acid (RNA) Extraction

Each of the 826 cryopreserved female *Aedes aegypti* pools were thawed on ice and then suspended in 200 µL of Trizol reagent (ThermoFisher, USA, catalogue number 19738). Homogenization was done by hand using micro pestle (Thomas Scientific, USA), Trizol topped up to 1 ml, and the homogenate pulse vortexed and left to lyse at room temperature for 10 minutes. Then 200 µL of chloroform (Carlo Erba, Italy, catalogue number 438603) was added to the homogenate and vigorously vortexed for 15 seconds. The homogenate was incubated for 2-3 minutes and then centrifuged at 12,000 g for 15 minutes at 4ºC. The aqueous phase was then transferred to a fresh sterile 1.5 ml Eppendorf tube, 1 µL of glycol blue co-precipitant (ThermoFisher, USA, catalogue number AM9515) added and the RNA precipitated by adding 500 µL of isopropanol (Finar, India, catalogue number 713090LC250) followed by incubation for 10 minutes at room temperature. The mixture was then centrifuged at 12,000 g at 4ºC for 10 minutes, supernatant discarded, and pellets washed with 1 ml of 75% ethanol (Carlo Erba, Italy, catalogue number 4146082) followed by centrifugation at 7,500 g for 10 minutes at 4ºC. Supernatant was discarded and the pellet air-dried. The RNA pellet was suspended in 50 µL of nuclease-free water to solubilize it (ThermoFisher, USA, catalogue number 10977-035) at room temperature by pipetting up and down, heated at 56ºC for 15 minutes and stored on ice or frozen at −80°C. RNA purity and concentration was checked using NanoDrop 1000 v3.8 (ThermoFisher, USA).

3.4 Arbovirus Detection and Identification

3.4.1 Reverse Transcription Real Time PCR (RT-qPCR)

A total of 826 RNA samples were subjected to RT-qPCR using the QuantiFast RT-PCR +R Kit (Qiagen, catalogue number 204956) for the detection of Rift Valley fever virus (RVFV) and members of the genera *Alphavirus* and *Flavivirus* using published primers and probes sets.

The RVFV assay was virus-specific using a primer-probe set targeting the L segment (Mwaengo et al., 2012). The PCR master mix was comprised of 6.25 µL of a 2X master mix, 1.625 µL of RNase-free water, 0.625 µL of 10 µM forward primer (5'-TTCTTTGCTTCTGATACCCTCTGT-3') and 10 µM reverse primer (5'-GTTCCACTTCCTTGCATCATCTG-3'), 0.5 µL of 5 µM probe (FAM-TTGCACAAGTCCACACAGGCCCCT - TAMRA), 0.25 µL of Rox dye, 0.125 µL of reverse transcriptase and 2.5 µL of the RNA template. The primers and probe were used at a final concentration of 0.5 µM and 0.2 µM, respectively. PCR conditions were 50°C for 20 minutes, 95 °C for 30 seconds, followed by 45 cycles of 95°C for 15 seconds and 60°C for 60 seconds.

The Pan-Alphavirus assay targeted the non-structural protein 4 (NSP4) (Giry et al., 2017) of genus *Alphavirus*. The PCR master mix components were: 6.25 µL of the 2X reaction mix, 0.25 µL of RNase-free water, 0.625 µL of 10 µM forward primer (F2A, 5'- ATGATGAARTCIGGIATGTTYYT-3'), 0.625µL of each of three reverse primers at 5 µM concentration (rev2A (R2A 5'-ATYTTIACTTCCATGTTCATCCA-3'), rev3A (R3A 5'-ATYTTIACTTCCATRTTCARCCA-3) and rev 4A (R4A 5'-ATYTTIACTTCCATGTTGACCCA-3')), 0.625 µL of 5 µM probe (5'- ATTO425- - AT+GTT+GTC+GT+CIC+CIAT- BHQ1/LNA-3'), 0.25 µL of Rox dye, 0.125 µL of reverse transcriptase and 2.5 µL RNA template. The forward primer, three reverse primers and probe were used at a final concentration of 0.5 µM, 0.25 µM and 0.25 µM, respectively. PCR cycling conditions were 45°C for 5 minutes, 98°C for 20 seconds and 45 cycles comprising 2 steps of 98°C for 3 seconds and 58°C for 45 seconds.

The Pan-Flavivirus assay targeted the non-structural 5 region (NS5) (Patel et al., 2013) of the genus *Flavivirus*. The PCR master mix was comprised of: 6.25 µL of 2X master mix, 2.175 µL of RNase free water, 0.2 µL of each primer at 50 µM; Flavi allS (Fwd 5'-TACAACATgATggggAARAgAgARAA-3'), Flavi all AS2 (Rev 5'-gTgTCCCAgCCNgCKgTgTCATCWgC-3'), Flavi all AS4 (Rev 5'-gTgTCCCAGCCNgCKgTRTCRTC-3'), Flavi all probe 3 mix at 10 µM each (0.5 µL of $3P_i$ (FAM-Tg+gTWYATgT+ggYTNg+gRgC-NFQ-MGB), 0.05 µL of $3P_{ii}$ (FAM-CCgTgCCATATggTATATgTggCTgggAgC-NFQ-MGB) and $3P_{iii}$ (FAM-TTTCTggAATTTgAAgCCCTgggTTT-NFQ-MGB)), 0.25 µL of Rox dye, 0.125 µL of reverse transcriptase mix and 2.5 µL RNA template. The forward primer, two reverse primers and three probes were used at final concentrations of 0.5 µM, 0.25 µM and 0.25 µM, respectively. PCR cycling conditions were 55°C for 20 minutes, 95°C for 30 seconds and 45 cycles comprising 2 steps 95°C for 10 seconds and 60°C for 25 seconds.

The RNA extracted from RVFV, CHIKV and DENV culture supernatants were used as positive controls in the RVFV, Pan-Alphavirus and Pan-Flavivirus assays, respectively. A no template control (NTC) comprised of the RT-PCR master mix only and nuclease free water was used in all assays as negative control. Virus amplification was performed with the ABI 7500 real time PCR machine running software v2.1 (Applied Biosystems, USA, catalogue number 4345241). The cycle threshold (Ct) cut-off for RT-qPCR positivity was set at 35 based on standard curve analysis of RVF, CHIKV and DENV RNA of known viral loads.

3.4.2 Sanger Sequencing

3.4.2.1 Conventional RT-PCR

All the 64 RT-qPCR positive mosquito pools RNA were reverse-transcribed and amplified using the SuperScript™ III One-Step RT-PCR system with platinum™Taq kit (Invitrogen, catalogue number 12374035). The PCR master mix was a 25 µL reaction comprised of: 12.5 µL of 2X reaction mix, 0.5 µL of 5 µM of the primers used in RT-qPCR, 0.5 µL of enzyme mix, 5.5 µL of RNase free water and 5.0 µL of RNA template. PCR conditions were 1 cycle of 50˚C for 30 minutes and 94 ˚C for 2

minutes, 40 cycles of 94˚C for 15 seconds, 60˚C for 1 minute and 68˚C for 1 minute, followed by 68˚C for 5 minutes. All the assays were performed with the Veriti 96 well Fast thermal cycler (Applied Biosystems, USA, catalogue number 4375305). Agarose gels (1.5%) were prepared by dissolving 1.5 grams of agarose powder (Bioline, UK, catalogue number B1041026) in 100 ml of 0.5X Tris-acetate-EDTA (TAE) buffer (ThermoFisher, USA, catalogue number 15558042) and staining with RedSafe DNA dye (iNtRON Biotech, catalogue number 21141). Amplicons and DNA ladder (Promega, catalogue number G2101) were individually mixed with 6X loading dye (Promega, catalogue number G2101), loaded into gels placed in gel tanks (Biorad, catalogue number 1704469) containing 0.5X TAE, and a current of 90 volts applied for 45 min using the PowerPac basic supply (Biorad, catalogue number 1645050). Bands were visualized under the ChemiDoc XRS + imaging system (BioRad, catalogue number 1708265).

3.4.2.2 Sanger Library Preparation (Cycle Sequencing)

Amplicons from 64 samples were purified using the QIAmp MinElute PCR purification kit (Qiagen, catalogue number 28004) and then subjected to Big Dye Terminator Cycle sequencing (Applied Biosystems, catalogue number 4336917) protocol using the primers used in RT-PCR. Briefly, 10 µL reaction mix for individual primers, consisting of 0.5 µL BigDye terminator 3.1 ready mix, 1.75 µL 5X BigDye sequencing buffer, 4.75 µL of deionized water, 1 µL of 5 µM concentration sequencing primer and 2 µl of purified amplicons (template) was prepared. Cycle sequencing was then performed with the following cycling conditions; initial denaturation at 95˚C for 10 seconds, amplification of 25 cycles (96˚C for 10 seconds, annealing at 60˚C for 5 seconds, elongation at 60˚C for 4 minutes) and final extension of 15˚C for 10 minutes. The sequencing reaction was then purified using ethanol/EDTA precipitation.

3.4.2.3 Capillary Sequencing and Data Analysis

The sequence for both the forward and reverse strands of the 64 flavivirus positive amplicons were determined using the 3730 DNA analyzer (ThermoFisher, USA, catalogue number A41046). This instrument was outsourced, and so samples (cycle

sequencing products) were shipped to the International Livestock Research Institute (ILRI) labs in Nairobi, Kenya. The BioEdit sequence alignment editor program version v7.0.5.3 (Hall, 1999) was used to generate contigs from the reads (.ab1 format) and the Basic Local Alignment Search Tool (BLASTn) v2.11.0 (Camacho et al., 2009a) software used to identify the viruses by searching against the GenBank nucleotide database.

3.5 Metagenomic Sequencing

3.5.1 Complementary DNA (cDNA) Synthesis

Total RNA from 73 (26 RT-qPCR positive and 44 RT-qPCR negative) mosquito pools were treated with 2 U/µL TURBO DNAse kit (ThermoFisher, catalogue number AM1907) at 37°C for 30 minutes. DNAse-treated RNA was purified by extraction with Acid-Phenol: Choloform: Isoamyl alcohol pH 4.5 (ThermoFisher, catalogue number AM9722) and the RNA eluted in 50 µL nuclease-free water and stored on ice for immediate use or frozen at −80°C. A 5 µL volume of DNAse-treated RNA was used as a template for cDNA synthesis using Superscript III reverse transcriptase kit (Invitrogen, catalogue number 18080051). For selective amplification of viral genomes, cDNA synthesis was primed with a set of 96 hexamers (Endoh et al., 2005), with each oligonucleotide pooled at equal amount at a concentration of 100 µM. An initial mixture containing RNA, hexamers and 1 µL of 10 mM dNTPs was incubated at 65°C for 5 minutes, and then immediately placed on ice for 1 minute. A reaction mix containing 200 U of SuperScript III RT, 40 U RNaseOUT, 0.1 M DTT and 1X RT buffer was added to the initial mix, followed by incubation at 25°C for 10 minutes, 50°C for 50 minutes and 85°C for 10 minutes. 5 U of Klenow fragment 3'->5' exo- (New England Biolabs, catalogue number M0212L), 10 U of Ribonuclease H (Invitrogen, catalogue number 18080051) and 1µL of 10mM dNTPs (Invitrogen, catalogue number 18080051) were added into the cDNA reaction mix, and incubated at 37°C for 1.5 hours, for double stranded cDNA synthesis (dsDNA). dsDNA products were purified using QIAquick PCR purification kit (QIAgen, catalogue number 28104).

3.5.2 Library Preparation and Sequencing

The quantity of the 73 purified DNA products was measured using Qubit dsDNA HS kit (Invitrogen, catalogue number Q32854) with the Qubit 3 fluorometer device (ThermoFischer, catalogue number 15387293). 100 ng of purified DNA was used to prepare indexed, paired-end libraries using Nextera XT DNA library Prep kit (Illumina, catalogue number 15032354 and 15032355). dsCDNA was tagmented and then indexed using Illumina Tru-Seq RNA UD indexes (Integrated DNA Technologies, USA, catalogue number 20022371). Size distributions of prepared libraries were assessed using an Agilent 2100 Bioanalyzer DNA 1000 HS assay (Agilent Technologies, USA, catalogue number 50674625). The average library was 400bp. Small DNA fragments were removed from the libraries using Agencourt's AmpPure XP beads (Beckman Coulter, catalogue number A63881). DNA quantity of the cleaned libraries was measured using Qubit dsDNA HS kit. Indexed libraries were then pooled at equimolar concentration and then diluted to 8 pM for sequencing on the MiSeq platform (Illumina, USA). Paired end sequencing (500 cycles) was done using Miseq V2 reagent kit (Illumina, USA, catalogue number MS-102-2002).

3.6 Bioinformatics Analysis

3.6.1 Quality Control of NGS Sequencing Reads

The fastq files obtained from the sequencing machine containing paired-end reads were subjected to quality control using the fastp software v0.23.1 (Chen et al., 2018) for the removal of Illumina adapter sequences, duplicate reads, PhiX reads, low complexity reads, and low-quality bases (PHRED quality score < 20). Quality-checked reads were collated and visualized using MultiQC v1.0 (Ewels et al., 2016).

3.6.2 Taxonomic Assignment

The clean reads were then subjected to decontamination to remove *Aedes aegypti* and human reads using the BBsplit algorithm in BBmap suite v38.96 (Bushnell, 2014) , by aligning to mosquito and human genome references, respectively. The clean reads were then subjected to taxonomic assignment using Kraken 2 v2.0.9 (Wood et al., 2019) and the BioSankey v0.13 (Platzer et al., 2018) diagram visualized in Pavian.

3.6.3 Mapping of Sequencing Reads and Genome Assembly

Viral species reads with counts > 50 from taxonomic assignment were mapped on to their respective reference genomes downloaded from the National Centre for Biotechnology Information's (NCBI) GenBank using BWA mem algorithm v0.7.17 (Li & Durbin, 2009). The number of mapped reads, and percent genome coverage was determined using Samtools v1.7 depth, mpileup and bcftools v1.6 (Li et al., 2009). Genome coverage plots were generated using R v4.2.0 software (R Core Team, 2022). Reads with a genome coverage > 70 % were used for assembly. De novo assembly was performed with the SPAdes assembler v3.13.0 (Bankevich et al., 2012). Resulting contigs longer than 100 nucleotides (nt) and encoding for viral genes were identified using BLASTn v2.11.0 (Camacho et al., 2009b). Contigs matching to viral reference sequences with at least 70% alignment identity and e-value 0.001 were marked as significant.

3.6.4 Phylogenetic analysis

Only genomes with coverage > 70% were used for phylogenetic analysis. Complete genomes of viruses similar to the newly assembled viruses were downloaded from GenBank, the sequences collated and aligned using MAFFT v7.220 (Katoh & Standley, 2013). The alignment was visualized and edited to only include the protein coding regions using AliView software v1.289 (Larsson, 2014). The best nucleotide substitution model was selected using jModelTest in IQ-TREE v1.6.9 (Nguyen et al., 2015). Maximum likelihood phylogenetic trees were inferred in IQ-TREE using the General Time Reversible substitution model with invariable sites and 4 gamma categories (GTR+I+G4), and branch support assessed by 1000 bootstrap iterations(Nguyen et al., 2015). Trees were visualized with either FigTree v1.4.4.7 (Rambaut, 2010) and ggtree v3.15 (Yu et al., 2017) in R v4.2.0, and mid-point rooted.

CHAPTER FOUR

RESULTS

4.1 Mosquito Collections and RT- PCR Screening

Adult female *Aedes aegypti* mosquitoes were sampled from Kilifi, Kwale and Mombasa counties of coastal Kenya (Figure 3.1). A total of 16,520 individual mosquitoes grouped into 826 pools of 20 mosquitoes each were analyzed (Table 4.1). Ukunda had the highest mosquito pool collection with 212/826 (25.7%) followed by Malindi town with 196/826 (23.7%) and Kilifi town with 149/826 (18%). Rabai had the lowest collection with 0.2% (2/826). Molecular surveillance by RT-PCR for RVFV and viruses belonging to *Alphavirus* and *Flavivirus* genera found 69/826 (8.4%) of the pools were positive for *Flavivirus* only (Table 4.1). Kilifi town had the highest number of *Flavivirus* positive pools accounting for 21.7% (15/69) followed by Bamburi with 18.8% (13/69) and Malindi town with 15.9% (11/69). Mombasa town, Mazeras, Watamu and Vanga sites did not record any positive pools. The RT-PCR cycle threshold (Ct) ranged between 26 and 34 (Table 4.2).

Table 4.1: Summary of *Aedes aegypti* Pools Collection and Genus *Flavivirus* RT-PCR Results

Region	Sampling site	No. of pools generated	No. of *Flavivirus* positive pools
Kilifi county	Kilifi Town	149	15
	Rabai	2	2
	Malindi Town	196	11
	Watamu	10	0
	Mazeras	4	0
Mombasa county	Mombasa Town	5	0
	Bamburi	33	13
	Shanzu	5	3
	Tudor	8	2
	Portreiz	23	2
	Kisauni	55	7
	Tononoka	21	2
Kwale county	Lungalunga	34	2
	Kwale Town	36	1
	Vanga	33	0
	Ukunda	212	9
Total		826	69

Table 4.2: Summary of PCR-Positive Mosquito Pools

Region	Sampling site	Mosquito pool ID	*Flavivirus* RT-qPCR Ct	mNGS dcDNA concentration (ng/µL)
Kilifi county	Kilifi Town	KLF8	30.77	0.45
		KLF31	28.18	3.14
		KLF33	32.41	0.74
		KLF36	32.37	0.58
		KLF43	28.38	7.13
		KLF44	33	1.20
		KLF45	30	3.94
		KLF47	31	0.21
		KLF57	33.6	0.96
		KLF61	28.35	1.76
		KLF66	26.99	3.13
		KLF67	26.74	2.67
		KLF69	27.09	3.07
		KLF77	29.88	1.89
	Rabai	RB08	30.31	3.28
		RB20	31.58	1.20
	Malindi Town	MAL5	32.24	2.35
		MAL9	27.24	3.85
		MLD20	32.24	3.40
		MLD22	29.22	0.90
		MLD73	31.67	3.57
		MLD95	31.6	1.02
		MLD154	28.89	2.20
		MLD164	30.54	2.47
		MLD167	30.81	1.09
		MLD177	27.63	2.56
		MLD182	29.84	0.85
Mombasa county	Bamburi	BAM2	29	1.34
		BAM6	30	2.18
		BAM7	26	2.31
		BAM10	27	7.52
		BAM14	29	3.38
		BAM16	29	0.74
		BAM19	26	2.31
		BAM20	29	3.25
		BAM22	27	1.91
		BAM25	29	3.06
		BAM27	28	1.75
		BAM28	28	3.38
		BAMX	27	3.30
	Shanzu	SH01	29	1.08
		SH03	30	1.98
		SH06	24	3.92
	Tudor	TD03	30	1.57

		Mosquito pool ID	*Flavivirus* RT-qPCR Ct	mNGS dcDNA concentration (ng/µL)
		TD14	29	1.98
	Portreiz	PR13	31	2.78
		PR22	30.36	2.67
	Kisauni	KS11	30	7.22
		KS12	26	2.20
		KS17	27	1.27
		KS31	32.13	2.35
		KS31	32.13	2.35
		KS44	28	3.21
		KS48	31	2.27
	Tononoka	TN26	33	2.93
		TN66	27	0.31
Kwale county	Lungalunga	LN1	30.97	1.19
		LN5	32.65	0.55
	Kwale Town	KW37	30.64	2.03
	Ukunda	UK22	30.79	0.81
		UK37	33.24	5.57
		UK65	33.22	0.68
		UK68	33.87	0.35
		UK80	30.24	3.92
		UK85	31.94	2.29
		UK86	31.61	0.11
		UK88	34	2.55
		UK121	31.89	6.25
Total			69	

4.2 Sanger Sequencing and Nucleotide Analysis

Sanger sequencing was done on the 69 *Flavivirus* positive mosquito pools. BLASTn analysis of the 260bp amplicons of the *Flavivirus* positive pools identified several insect specific viruses (Figure 4.1). A total of 36/69 (52.2%) mosquito pools had matches to the Portugal mosquito flavivirus, 2/69 (2.9%) to Kamiti river virus (2.9%) and 6/69 (8.7%) to Phlebotomus flavivirus. Five pools (7.2%) showed high similarity to sections of *Aedes aegypti* genome while 19 pools (27.5%) did not have a similarity hit. Notably, one pool from Kisauni matched the pathogenic West Nile virus (1.4%).

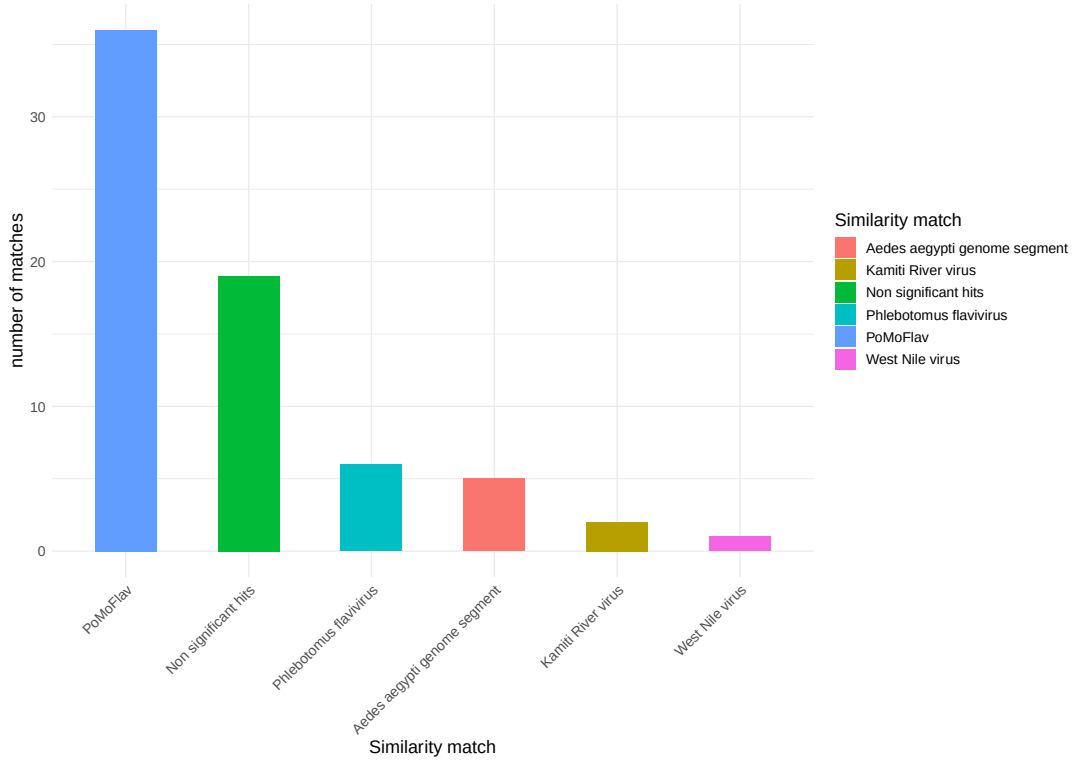

Figure 4.1: Summary of Sanger Sequencing Contigs Analysis. NCBI nucleotide database matches of genus *Flavivirus* positive *Aedes aegypti* pools' sequences are color coded as blue, red, green, purple and orange. The X-axis indicates the database matches while the Y-axis indicates the number of individual mosquito pools that had database matches. The analyzed contigs had an e-value of 0.001 and nucleotide similarity of 70%.

4.3 Metagenomic Analysis of *Aedes aegypti* Pools

A total of seventy-three *Aedes aegypti* pools (29 RT-PCR positive and 44 RT-PCR negative) passed the library preparation quality control steps and were sequenced using mNGS (Table 4.3 and 4.4). A total of 64 million reads were generated. After quality control to remove low quality, duplicated, adapter and host-specific reads, 971,754 (1.5%) reads were retained. Taxonomic assignment of reads showed a majority of mosquito pools had higher proportion of unclassified reads compared to classified reads (Figure 4.2). Amongst the classified reads, viruses comprised 0.87% (8469/971,754) of the retained reads and formed a relatively small proportion compared to eukaryota and bacteria (Figure 4.3). Further removal of eukaryotic reads

enabled the identification of the most abundant bacteria and viruses (Figure 4.4). Viral reads were taxonomically assigned to 16 viral families (Table 4.3) including *Phenuiviridae* which constituted 78.6% of the reads, *Flaviviridae* (9%), *Reoviridae* (6.2%), *Baculoviridae* (2.7%), *Closteroviridae* (1.1%), *Iridoviridae* (0.6%), *Orthomyxoviridae* (0.5%), *Microviridae* (0.5%), *Leviviridae* (0.5%), *Siphoviridae* (0.09%), *Picornaviridae* (0.06%), *Iflaviviridae* (0.05%), *Podoviridae* (0.02%), *Peribunyaviridae* (0.01%), *Pneumoviridae* (0.01%) and *Phycodnaviridae* (0.01%). The *Phenuiviridae* family was the most abundant detected in 64 of the 73 samples (Table 4.3). Viral diversity was identified in twenty *Aedes aegypti* pools from Bamburi, Kilifi town, Kisauni, Lunga Lunga, Malindi, Shanzu, Tononoka and Ukunda, which had at least 4 viral families (Figure 4.5). A total of 28 viral species were identified with Phasi Charoen-like phasivirus (PCLV) being the most abundant (Figure 4.6). No pathogenic viruses were identified. In general, RT-PCR positive mosquito pools had more viral reads than the RT-PCR negative pools (Figure 4.7).

Table 4.3: Summary of PCR Positive *Aedes aegypti* Pools Selected for Sequencing after Passing Library Preparation Quality Control

Region	Sampling site	Mosquito pool ID	*Flavivirus* RT-PCR Ct	Sequenced using metagenomics
Kilifi county	Kilifi Town	KLF8	30.77	yes
		KLF31	28.18	yes
		KLF33	32.41	yes
		KLF36	32.37	
		KLF43	28.38	yes
		KLF44	33	
		KLF45	30	
		KLF47	31	yes
		KLF57	33.6	
		KLF61	28.35	yes
		KLF66	26.99	
		KLF67	26.74	
		KLF69	27.09	
		KLF77	29.88	
	Rabai	RB08	30.31	
		RB20	31.58	
	Malindi Town	MAL5	32.24	yes
		MAL9	27.24	
		MLD20	32.24	
		MLD22	29.22	
		MLD73	31.67	yes
		MLD95	31.6	
		MLD154	28.89	yes
		MLD164	30.54	
		MLD167	30.81	
		MLD177	27.63	
		MLD182	29.84	
Mombasa county	Bamburi	BAM2	29	
		BAM6	30	yes
		BAM7	26	yes
		BAM10	27	
		BAM14	29	yes
		BAM16	29	
		BAM19	26	yes
		BAM20	29	yes
		BAM22	27	yes

		Mosquito pool ID	*Flavivirus* RT-PCR Ct	Sequenced using metagenomics
		BAM25	29	
		BAM27	28	
		BAM28	28	
		BAMX	27	
	Shanzu	SH01	29	yes
		SH03	30	
		SH06	24	yes
	Tudor	TD03	30	yes
		TD14	29	
	Portreiz	PR13	31	
		PR22	30.36	
	Kisauni	KS11	30	yes
		KS12	26	
		KS17	27	
		KS31	32.13	yes
		KS31	32.13	
		KS44	28	yes
		KS48	31	
	Tononoka	TN26	33	
		TN66	27	yes
Kwale county	Lungalunga	LN1	30.97	yes
		LN5	32.65	
	Kwale Town	KW37	30.64	yes
	Ukunda	UK22	30.79	yes
		UK37	33.24	yes
		UK65	33.22	
		UK68	33.87	
		UK80	30.24	yes
		UK85	31.94	yes
		UK86	31.61	
		UK88	34	
		UK121	31.89	yes
Total			69	29

Table 4.4: Summary of PCR Negative *Aedes aegypti* Pools Selected for Sequencing after Passing Library Preparation Quality Control

Region	Sampling site	Mosquito pool ID
Kilifi county	Kilifi town	KLF3
		KLF11
		KLF26
		KLF41
		KLF76
		KLF81
	Rabai	RB31
	Malindi town	MLD3
		MLD11
		MLD14
		MLD18
		MLD118
	Watamu	W5
Mombasa county	Bamburi	BAM1
		BAM4
		BAM8
		BAM24
		BAM31
	Tudor	TD04
	Portreiz	PR18
	Kisauni	KS07
		KS08
		KS26
		KS37
		KS46
	Tononoka	TN03
Kwale county	Lungalunga	LN7
		LN10
		LN25
		LN27
		LN31
	Kwale	KW06
		KW07
		KW19
		KW25
		KW36
	Ukunda	UK8
		UK46
		UK49
		UK71
		UK132
	Vanga	VG13
	Mazeras	MZ11
		MZ12
Total		44

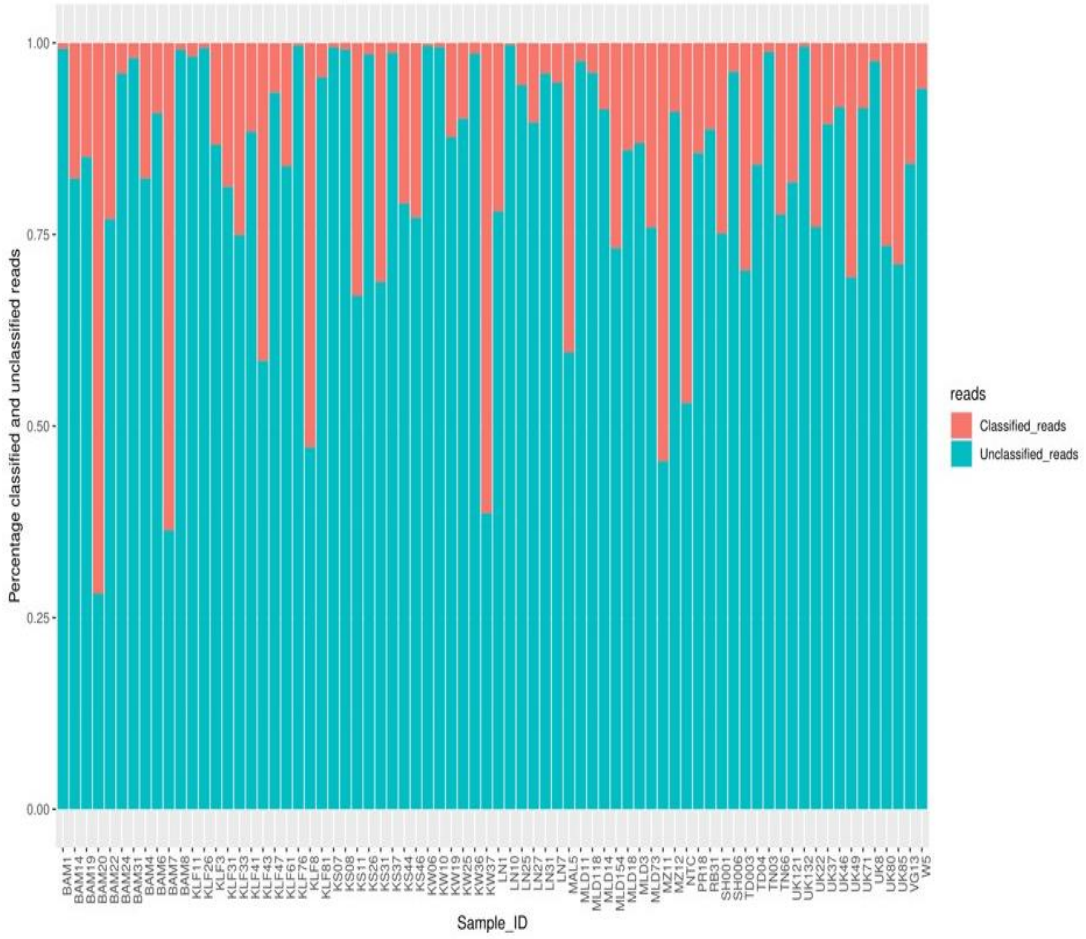

Figure 4.2: Proportion of Classified and Unclassified Total Reads across *Aedes aegypti* Pools (n=73)

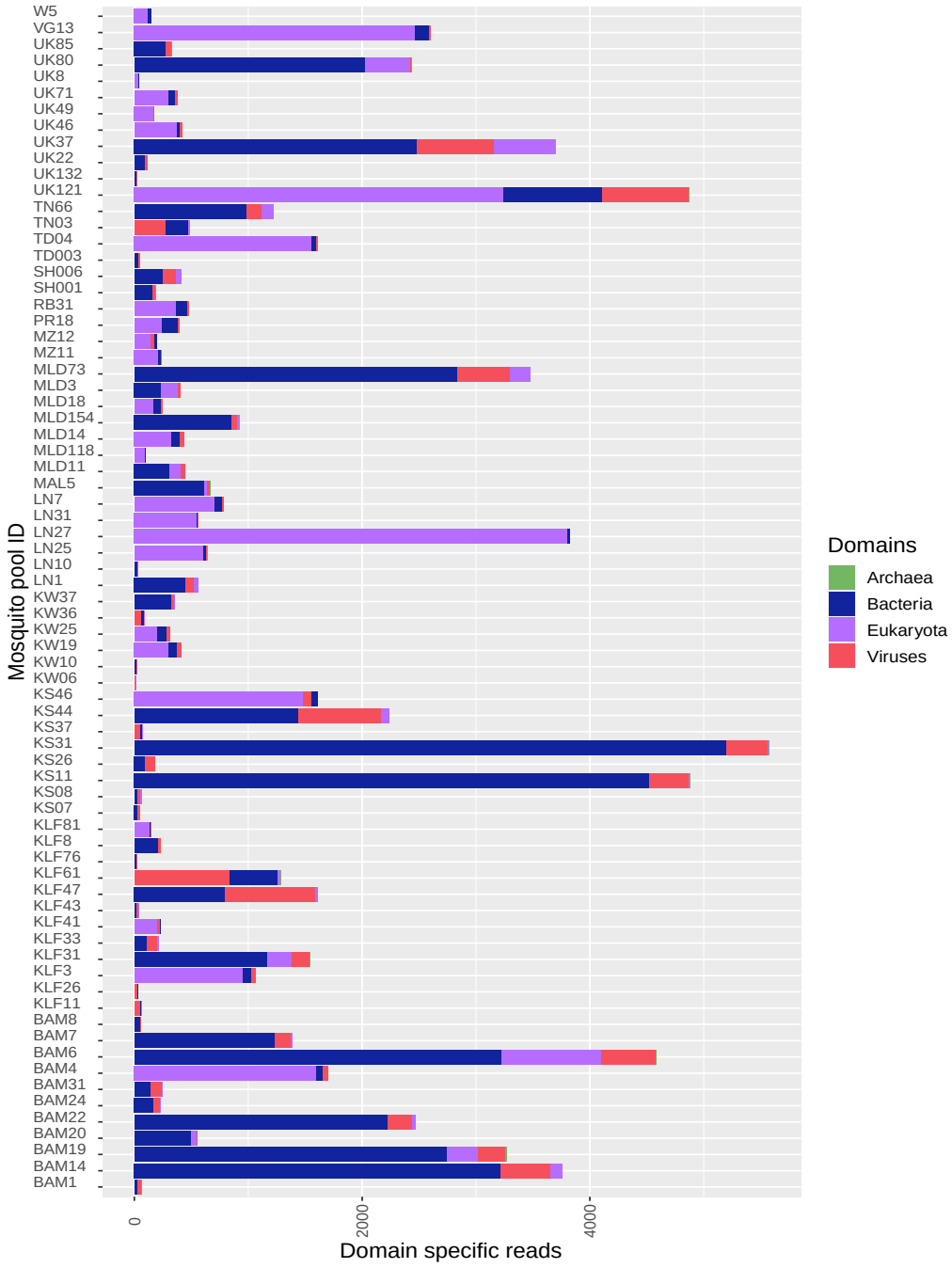

Figure 4.3: Domain-Specific Reads Distribution across *Aedes aegypti* Pools. Taxonomic classification of reads was done using Kraken 2 software using the Minikaren database that has reference sequences for archaea, bacteria, eukaryota and viruses. Proportion of reads in the mosquito pools belonging to archaea, bacteria, eukaryota and viruses are differentially colored.

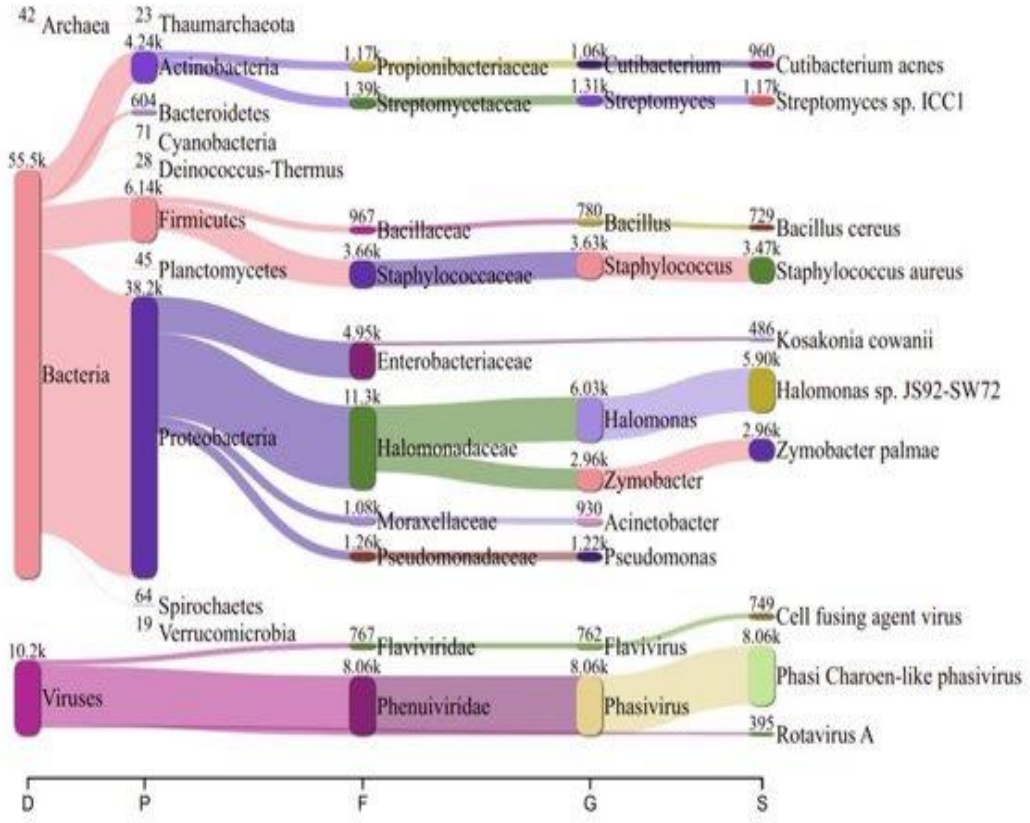

Figure 4.4: Microbial Domain-Specific Abundant Reads Distribution across *Aedes aegypti* Pools. Taxonomic classification of reads was done using Kraken 2 software using the Minikaren database that has reference sequences for archaea, bacteria, eukaryota and viruses.

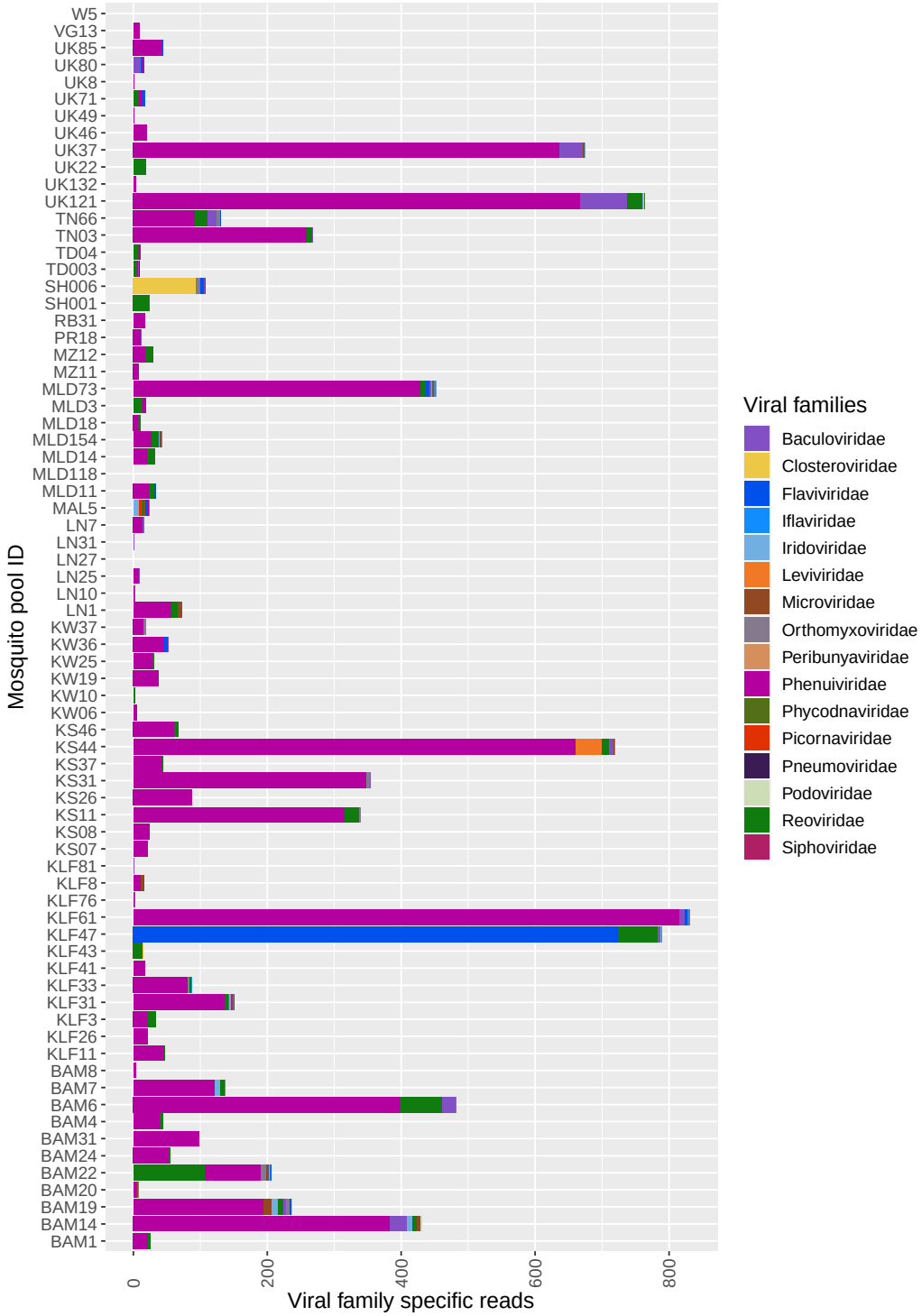

Figure 4.5: Distribution of Viral Family Reads across *Aedes aegypti* Pools. Viral families are assigned different colors as indicated on the legend on the right of the graph.

Table 4.5: Proportion of *Aedes aegypti* Pools with Viral Family Reads Following Metagenomics Sequencing (n=73)

Viral family	Total reads	No. of mosquito pools present	Proportion of positive mosquito pools
Phenuiviridae	6653	64	87.7%
Baculoviridae	226	29	39.7 %
Iridoviridae	49	14	19.2%
Siphoviridae	8	5	6.8%
Podoviridae	2	1	1.4%
Reoviridae	522	41	56.2%
Microviridae	42	16	21.9%
Orthomyxoviridae	46	20	27.4%
Flaviviridae	764	18	24.7%
Closteroviridae	96	4	5.5%
Iflaviridae	4	4	5.5%
Leviviridae	39	1	1.4%
Picornaviridae	5	1	1.4%
Peribunyaviridae	1	1	1.4%
Pneumoviridae	1	1	1.4%
Phycodnaviridae	1	1	1.4%
Total viral reads	8469		

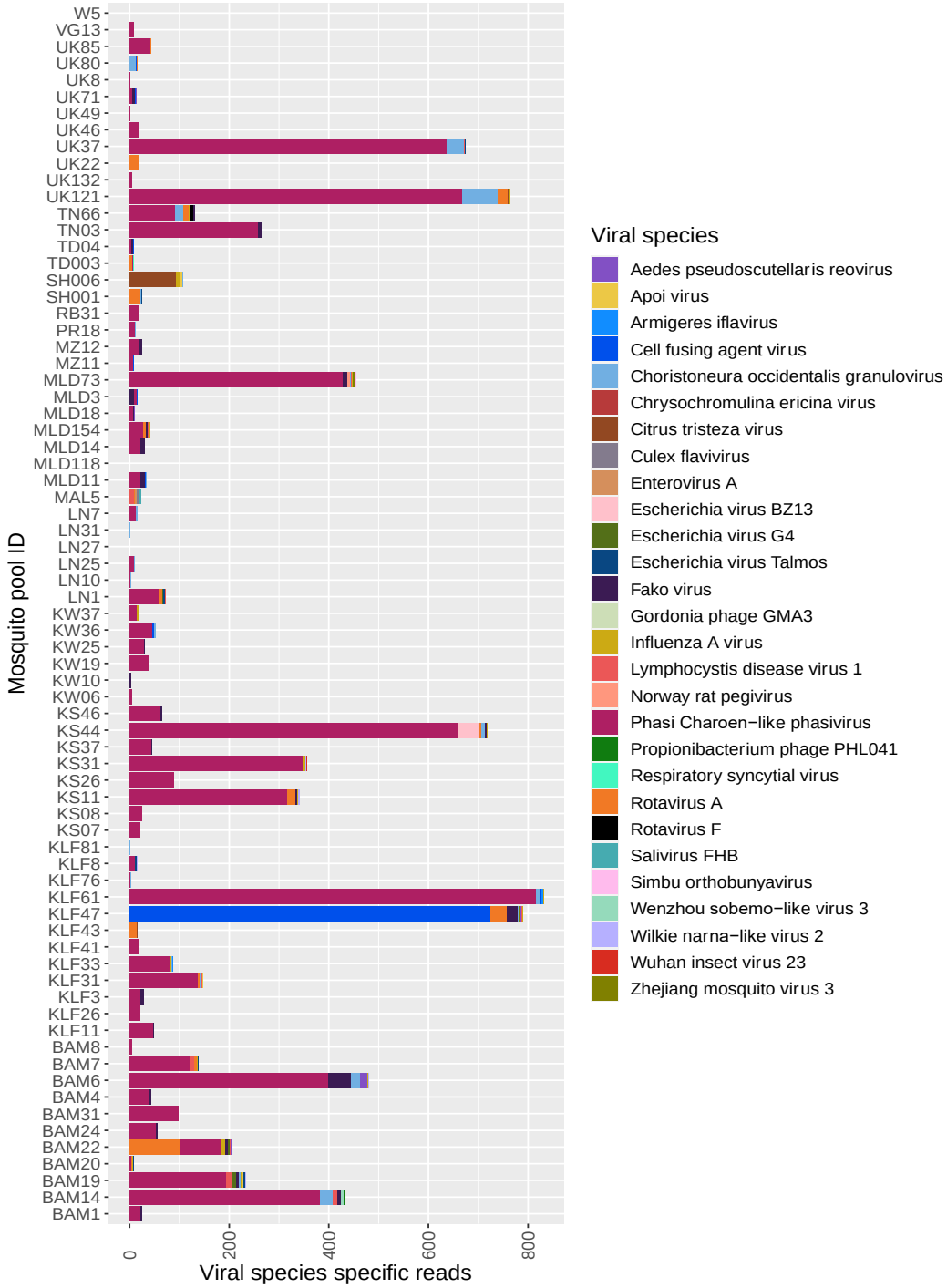

Figure 4.6: Distribution of Virus Species-Specific Reads across the *Aedes aegypti* Pools. The different viral species are color coded as highlighted in the legend on the right.

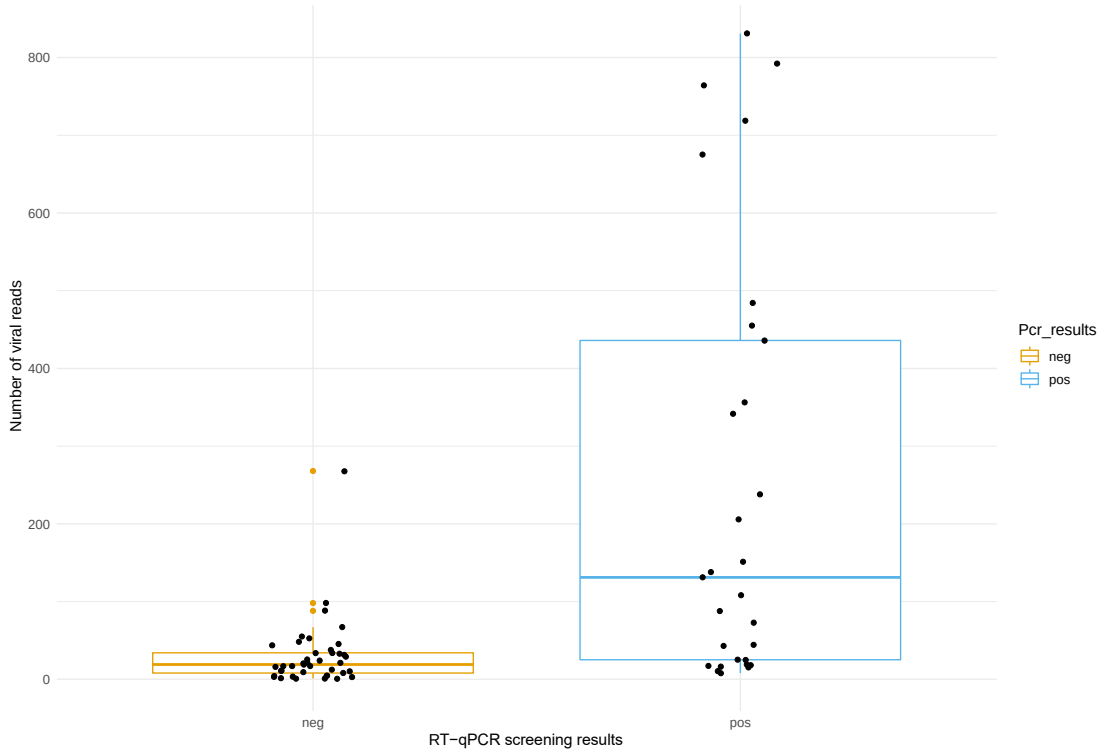

Figure 4.7: Distribution of Viral Reads from RT-PCR Positive and Negative *Aedes aegypti* Pools. Box plot of RT-PCR negative pools (n=44) is colored orange while that of RT-PCR positives pools (n=29) is colored blue. RT-PCR positive mosquito pools had a higher read counts with higher dispersal than RT-PCR negative pools.

4.4 Genome Assembly of Identified Viruses

A total of 57/73 (78%) *Aedes aegypti* pools had viral contigs longer than 100bp after assembly. BLASTn analysis of the contigs identified a total of nine viral species including Aedes partiti-like virus 1, cell fusing agent virus (CFAV), Fako virus, Humaita-Tubiacanga virus (HTV), PCLV, rotavirus A, Tesano virus, totivirus and Verdadero virus (Figure 4.8). The most abundant contigs belonged to totivirus and PCLV**.** Assembly did not generate full genomes from totivirus, FAKV, Tesano virus, rotavirus A and Verdadero virus. Mapping of reads to a complete genome of Tesano virus retrieved from GenBank (accession number LC496784.1) showed there were a lot of gaps (Figure 4.9).

A near-complete CFAV genome was generated in one pool from Kilifi township (KLF47). The sole CFAV sequence from Kilifi shared 96.95% whole genome nucleotide identity with CFAV identified in *Aedes aegypti* collected in 2014 in Cairns, Australia (GenBank accession number LR694075.1) (Table 4.6). The Kilifi CFAV had a good genome coverage (100%) with relatively fewer reads towards the tail end of the genome, a region which codes for structural genes (Figure 4.10).

The PCLV has three genomic segments L (large), M (medium) and S (small). Near-complete L and M segments from eleven mosquito pools, and partial S segment from six out of these eleven mosquito pools were generated (Figure 4.11). The PCLV segments shared 97.94–99.14%, 95.1–98.94% and 92.78–98.95% nucleotide similarities with the L (accession no MT361069.1), M (accession no MT361068.1) and S (accession no MT361067.1) segments, respectively, of PCLV detected in *Aedes aegypti* mosquitoes collected in Kisumu district in Kenya in 2018 (Table 4.6).

Aedes partiti-like virus 1 and HTV virus genomes are unclassified viruses comprised of two open reading frames encoding the capsid (*C*) and RNA dependent RNA polymerase (*RdRp*) genes, respectively. Near-complete genome coverage (>97 %) for both *C* and *RdRp* genes of HTV (in twelve mosquito pools for each gene) and *Aedes* partiti-like virus 1 (in seven and eight mosquito pools for *C* and *RdRp* gene, respectively) were obtained (Figure 4.12). The *C* and *RdRp* genes for HTV and *Aedes* partiti-like virus 1 shared at least 91% nucleotide sequence identity with sequences deposited in GenBank (Table 4.6).

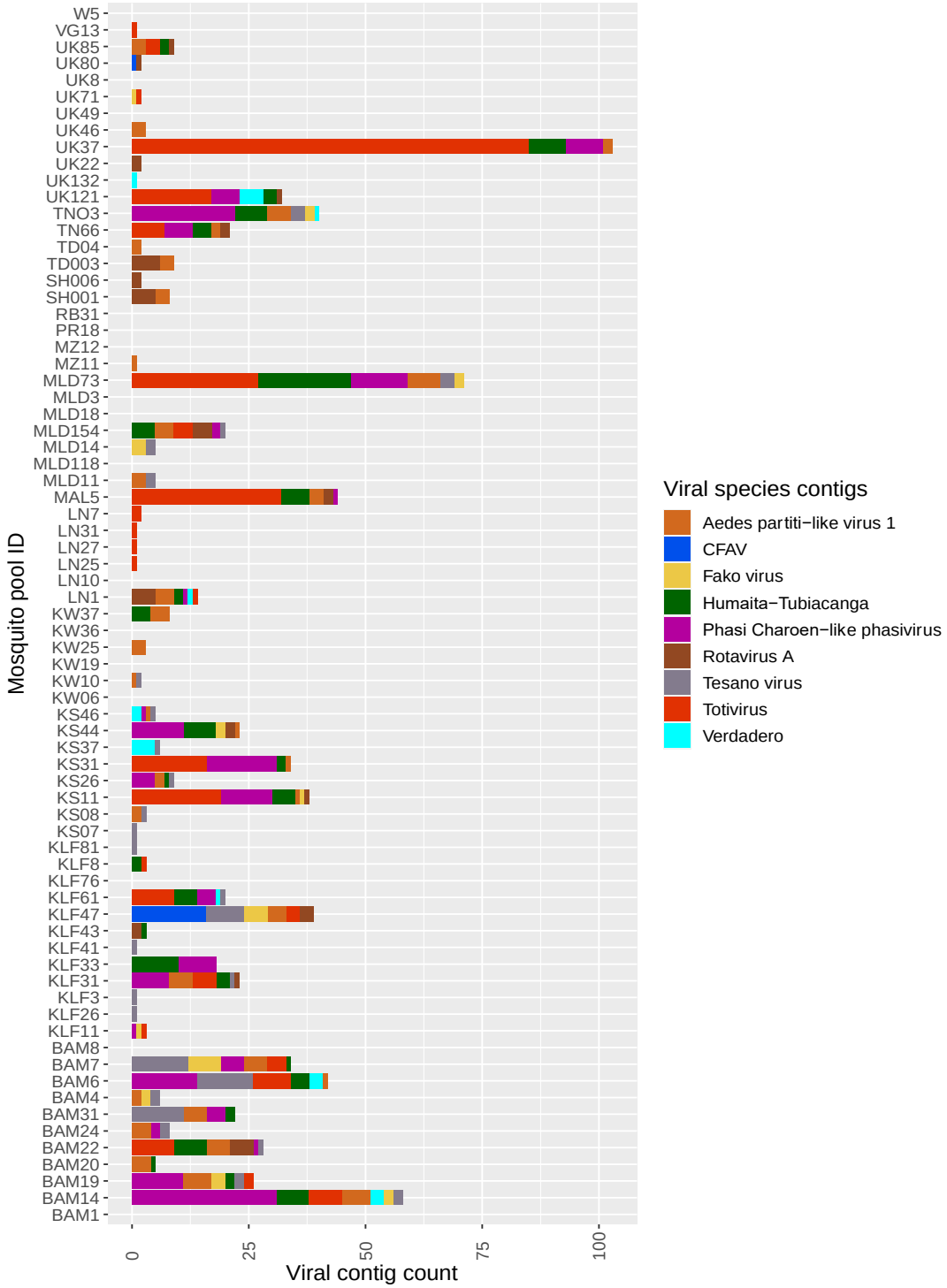

Figure 4.8: Distribution of Assembled Viral Species-Specific Contigs Longer than 100bp Across *Aedes aegypti* Pools. Contigs were generated from reads using metaSPades software. The identity of the contigs was determined using BLASTn against the NR database in GenBank.

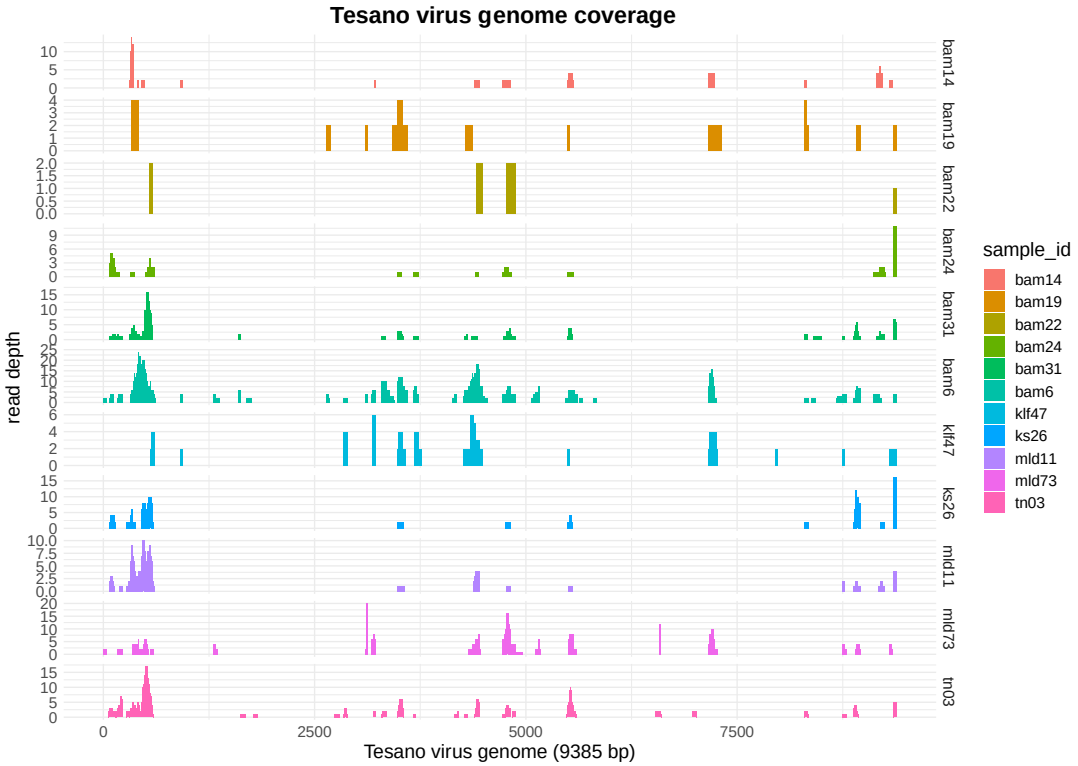

Figure 4.9: Genome Coverage Plot of Tesano Virus. Viral reads from eleven *Aedes aegypti* pools were mapped on to a Tesano virus complete genome (LC496784.1) retrieved from GenBank. The coverage plots of the respective mosquito pools are differentially colored as highlighted on the legend on the right.

Table 4.6: Summary of Homology Search (BLASTn) of Viruses with Near Coding Complete Genomes

Virus name	Virus segment	Mosquito pool ID	Sequence length (nt)	Closest related sequence	% Identity	No. of mapped reads
Aedes partiti-like virus 1	Capsid	BAM6	1301	BK059431.1	98.93	2525
	Capsid	MLD73	1301	BK059431.1	99.01	132
	Capsid	KS26	1301	BK059431.1	99.09	241
	Capsid	TN03	1301	BK059431.1	99.13	2445
	Capsid	BAM31	1301	BK059431.1	99.29	526
	Capsid	BAM14	1301	BK059431.1	99.25	443
	Capsid	BAM24	1301	BK059431.1	98.73	303
	RdRp	BAM6	1494	BK059430.1	98.32	4619
	RdRp	MLD73	1494	BK059430.1	98.44	108
	RdRp	UK37	1494	BK059430.1	98.35	320
	RdRp	TN66	1494	BK059430.1	96.8	26
	RdRp	BAM14	1494	BK059430.1	98.46	621
	RdRp	KS44	1494	BK059430.1	98.58	2226
	RdRp	KS31	1494	BK059430.1	98.49	189
	RdRp	KS11	1494	BK059430.1	98.29	128
Cell fusing agent virus	Whole genome	KLF47	10684	LR694075.1	96.95	2627
Humaita-Tubiacanga virus	Capsid	UK121	1609	MN053812.1	97.28	64
	Capsid	UK85	1609	MN053812.1	97.97	143
	Capsid	UK37	1609	KR003802.1	97.91	9123
	Capsid	TN66	1609	MN053814.1	97.84	180
	Capsid	TN03	1609	MN053814.1	96.9	156
	Capsid	KS44	1609	MN053814.1	97.86	1604
	Capsid	KS31	1609	MN053814.1	97.52	224
	Capsid	KS26	1609	MN053824.1	93.72	68
	Capsid	KS11	1609	MN053812.1	97.59	86
	Capsid	KLF31	1609	MN053814.1	97.25	444
	Capsid	BAM31	1609	MN053812.1	96.87	136
	Capsid	BAM6	1609	MN053810.1	97.59	838
	RdRp	UK121	2794	KR003801.1	97.24	1049
	RdRp	UK85	2794	KR003801.1	97.44	308
	RdRp	UK37	2794	KR003801.1	97.28	1173
	RdRp	TN66	2794	KR003801.1	97.36	658
	RdRp	TN03	2794	KR003801.1	97.05	269
	RdRp	KS44	2794	KR003801.1	97.48	604
	RdRp	KS31	2794	KR003801.1	97.32	1259
	RdRp	KS26	2794	MN053811.1	91.67	129
	RdRp	KS11	2794	KR003801.1	96.96	493
	RdRp	KLF31	2794	KR003801.1	96.49	146

Virus name	Virus segment	Mosquito pool ID	Sequence length (nt)	Closest related sequence	% Identity	No. of mapped reads
	RdRp	BAM31	2794	KR003801.1	95.45	180
	RdRp	BAM6	2794	KR003801.1	97.59	1245
Phasi	L	KLF33	6474	MN866234.1	98.18	144
Charoen-like	L	KLF31	6474	MT361069.1	98.28	216
Phasivirus	L	BAM19	6474	MT361069.1	99.14	325
	L	BAM7	6474	MT361069.1	98.49	308
	L	BAM14	6474	MT361069.1	98.58	541
	L	KS11	6474	MT361069.1	98.11	360
	L	KS31	6474	MT361069.1	98.24	584
	L	KLF61	6474	MT361069.1	98.99	1041
	L	BAM6	6474	MT361069.1	99.03	715
	L	KS44	6474	MT361069.1	97.94	1115
	L	UK121	6474	MT361069.1	98.97	968
	L	UK37	6474	MT361069.1	98.88	1246
	L	MLD73	6474	MT361069.1	98.11	1160
	M	KLF33	3724	MT361068.1	95.35	75
	M	BAM7	3725	MT361068.1	95.1	125
	M	KLF31	3725	MT361068.1	96.16	156
	M	MLD73	3725	MT361068.1	98.77	239
	M	BAM19	3725	MT361068.1	97.43	214
	M	UK37	3725	MT361068.1	98.58	317
	M	KS31	3725	MT361068.1	97.52	341
	M	KS11	3725	MT361068.1	98.7	349
	M	KLF61	3725	MT361068.1	97.57	972
	M	UK121	3725	MT361068.1	98.94	665
	M	KS44	3725	MT361068.1	97.95	691
	M	BAM6	3725	MT361068.1	98.03	431
	M	BAM14	3725	MT361068.1	98.56	566
	S	BAM7	1208	MN866293.1	92.78	15
	S	MLD73	1208	MT361070.1	98.66	19
	S	KLF33	1208	MT361067.1	97.26	32
	S	KLF31	1208	MT361067.1	97.9	46
	S	KS31	1208	MT361067.1	98.6	125
	S	BAM14	1208	MT361067.1	98.95	142
	S	KS11	1208	MT361067.1	98.85	44
	S	UK37	1208	MT361067.1	98.08	38
	S	KLF61	1208	MT361067.1	98.61	65
	S	KS44	1208	MT361067.1	98.93	138
	S	UK121	1208	MT361067.1	98.54	114
	S	BAM6	1208	MT361067.1	98.75	172

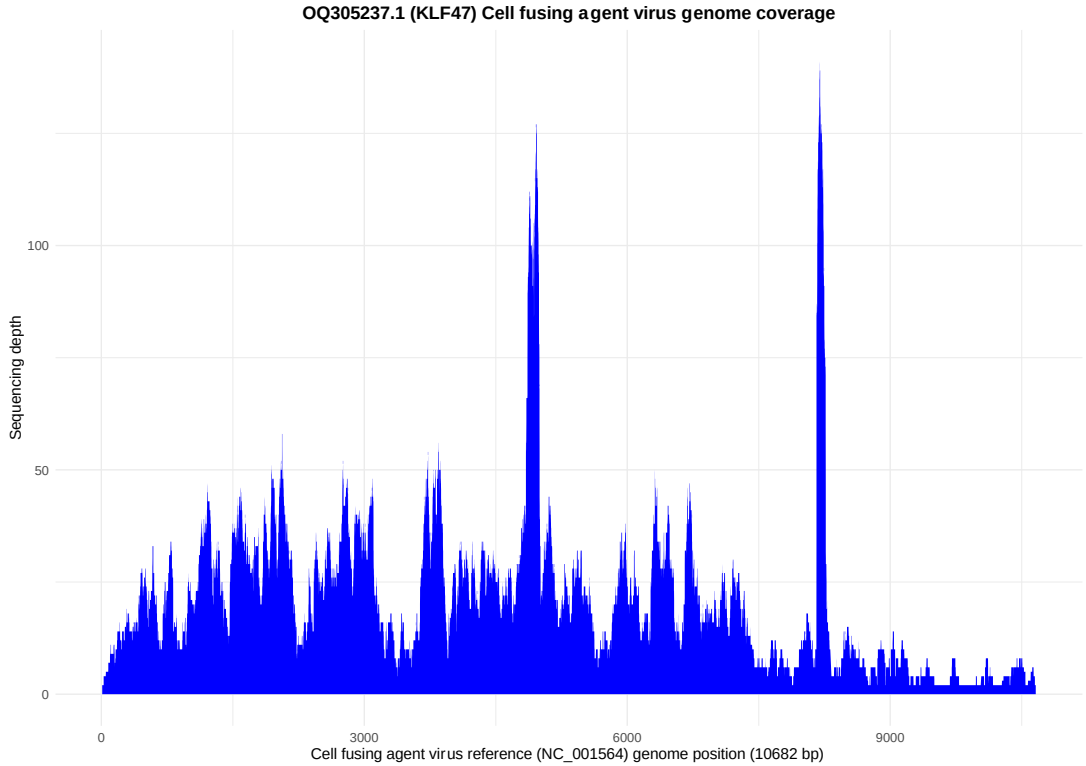

Figure 4.10: Cell Fusing Agent Virus (CFAV) Genome Coverage Plot. *Aedes aegypti* reads (blue) from a pool from Kilifi township (KLF47, accession OQ305237.1) are mapped on to CFAV reference genome (accession no NC_001564.2) retrieved from GenBank. The X-axis indicates the reference genome position while the Y-axis indicates the sequencing depth.

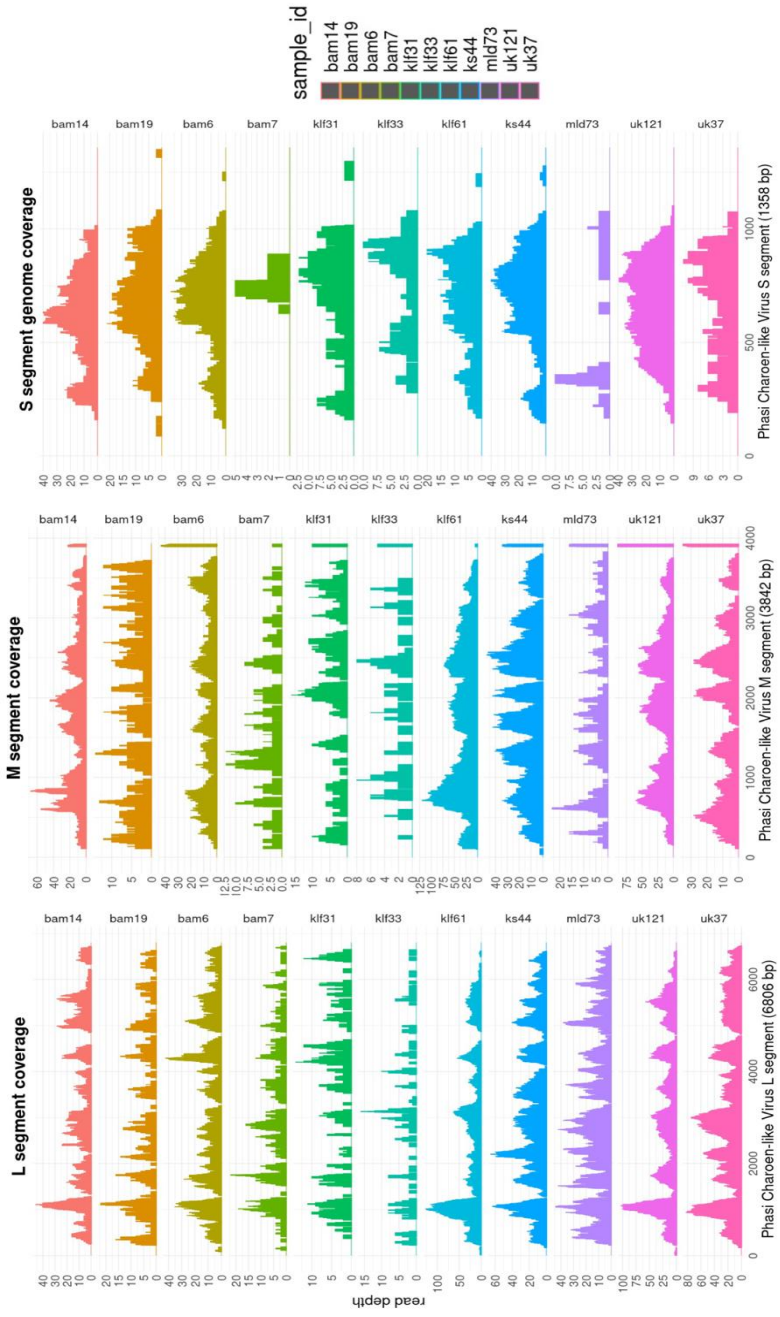

Figure 4.11: Phasi Charoen-like Phasivirus (PCLV) Genome Coverage Plots. Sequencing reads from *Aedes aegypti* pools are mapped on to the reference sequences for L (large, accession no NC_038262.1), M (medium, accession no NC_038261.1) and S (small, NC_038263.1) gene segments, respectively, retrieved from GenBank. The X-axis indicates sequencing depth, Y axis indicates genome segments and Z-axis reference genome lengths. Mosquito pool accession numbers and collection sites are indicated on the right of each plot.

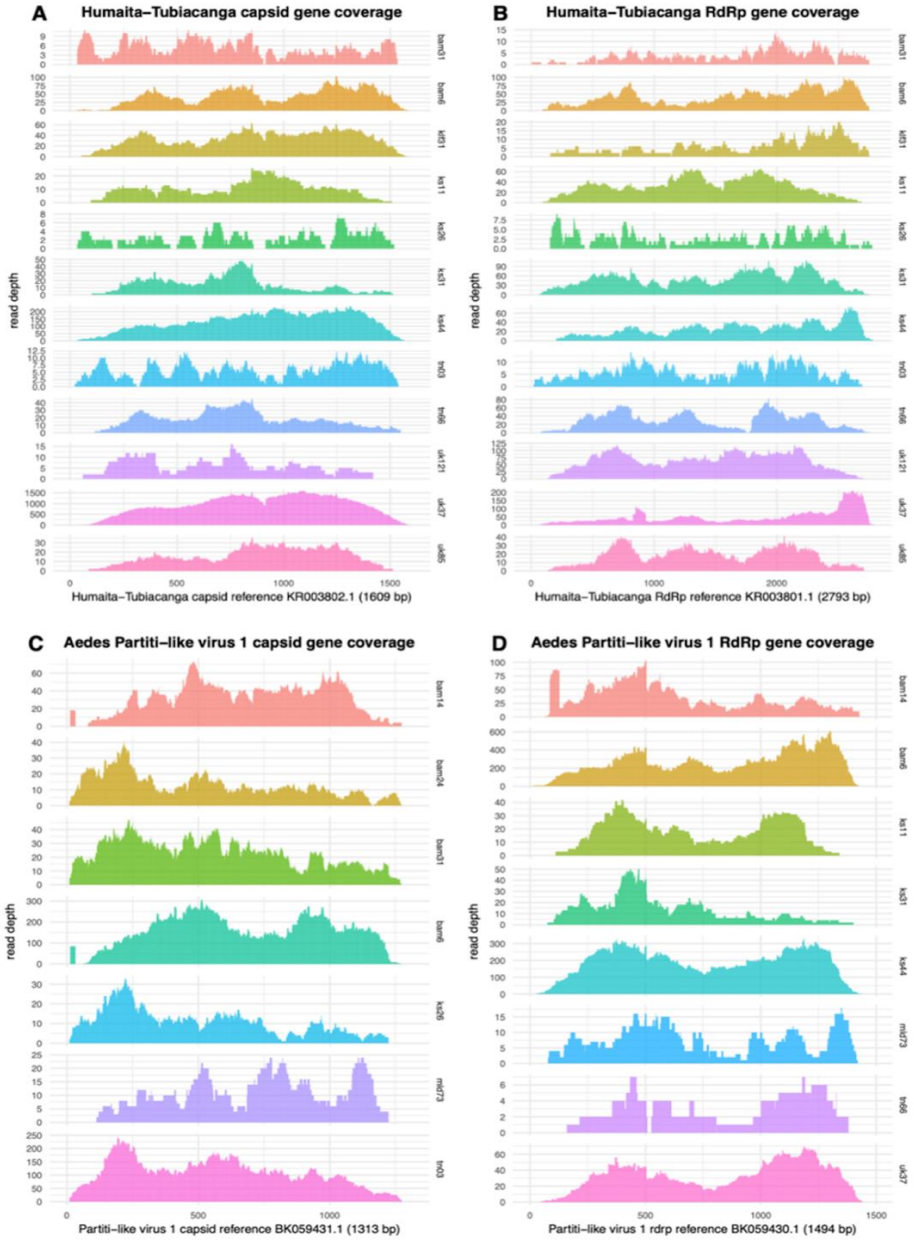

Figure 4.12: Genome Coverage Plots for Humaita-Tubiacanga Virus (HTV) and Aedes Partiti-Like Virus 1. Sequencing reads from *Aedes aegypti* pools are mapped on to HTV's capsid (A) and RdRp (B), and Aedes partiti-like virus 1's capsid (C) and RdRp (D) genes, respectively. The accession numbers for the respective GenBank reference genes are indicated below each plot. The mosquito pool accession numbers and sampling site are indicated on the right of each plot.

4.5 Phylogenetic Analysis of Selected Viruses

Two viruses, CFAV and PCLV, had good genome coverage (>90%) and a substantial number of sequences in NCBI amenable to evolutionary analysis. GTR+I+G4 was the best nucleotide substitution model for both CFAV and PCLV analyses. Whole genome phylogenetic analysis of the sole CFAV sequence from this study (KLF47) showed it was basal in a major clade that had global sequences from outside Africa (Figure 4.13). However, only a few whole genome sequences were available for this analysis (n=22) with low representation from Africa (n=2). On the other hand, there were more CFAV non-structural 5 (*NS5*) gene sequences in the NCBI database. The analysis based on *NS5* gene sequences (n=96) revealed that all Kenyan sequences belonged to four different lineages. (Figure 4.14).

Phylogenetic analysis of PCLV's L, M and S segments from this study also showed unique clustering. All Kenyan sequences belonged to clades made up exclusively of Kenyan sequences, highlighted in the middle of the respective phylogenetic trees (Figures 4.15 to 4.17). These clades were basal to the Asian clade predominated by sequences from India. The short branches within the Kenyan clades are indicative of the unique diversity of the Kenyan sequences. Notably, sequences from other African countries were distant from the Kenyan sequences.

Although we obtained near-complete *C* and *RdRp* gene coverages for *Aedes* partiti-like virus 1 and Humaita-Tubiacanga virus, these viruses had very few sequences available (n <10 for each genome segment) in the public nucleotide databases for phylogenetic comparison.

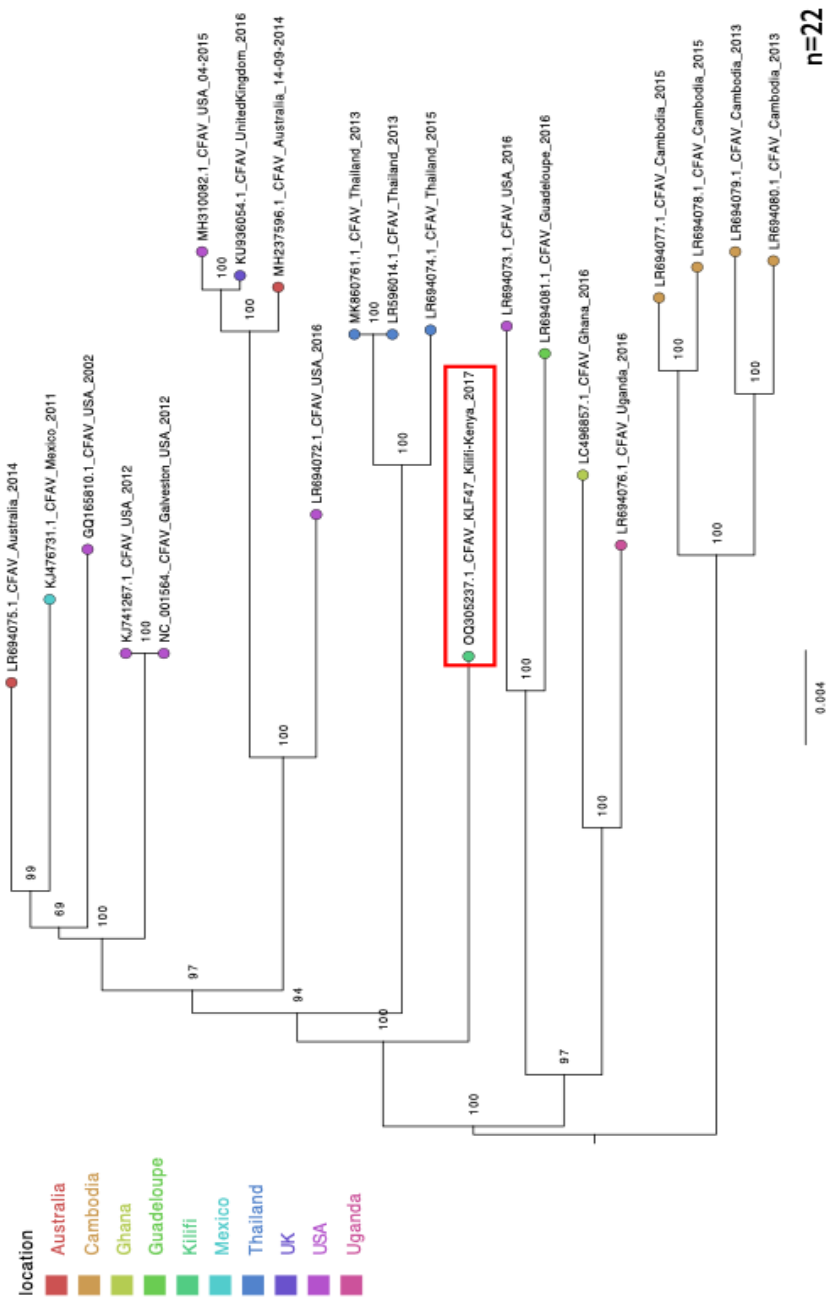

Figure 4.13: Maximum Likelihood Phylogenetic Tree of Cell Fusing Agent Virus (CFAV) Based on Whole Genomes. The tree indicates the evolutionary relationship of CFAV genome from this study (encircled in red) and global sequences retrieved from GenBank (n=21). The origin of each sequence is indicated at the branch tips and color-coded based on year of collection.

Figure 4.14: Maximum Likelihood Phylogenetic Tree of Cell Fusing Agent Virus (CFAV) Based on a 740 bp Region of the NS5 Gene (n=96). Branches are labelled with bootstrap values after 1,000 iterations. External nodes are labelled with countries of origin and differentially colored based on year of collection. A clade comprised of sequences exclusively from Thailand has been collapsed to aid visibility (colored in blue at top of the tree). The sole sequence from Kilifi in this study (Kilifi-Kenya) is pointed by red arrow near bottom of the tree.

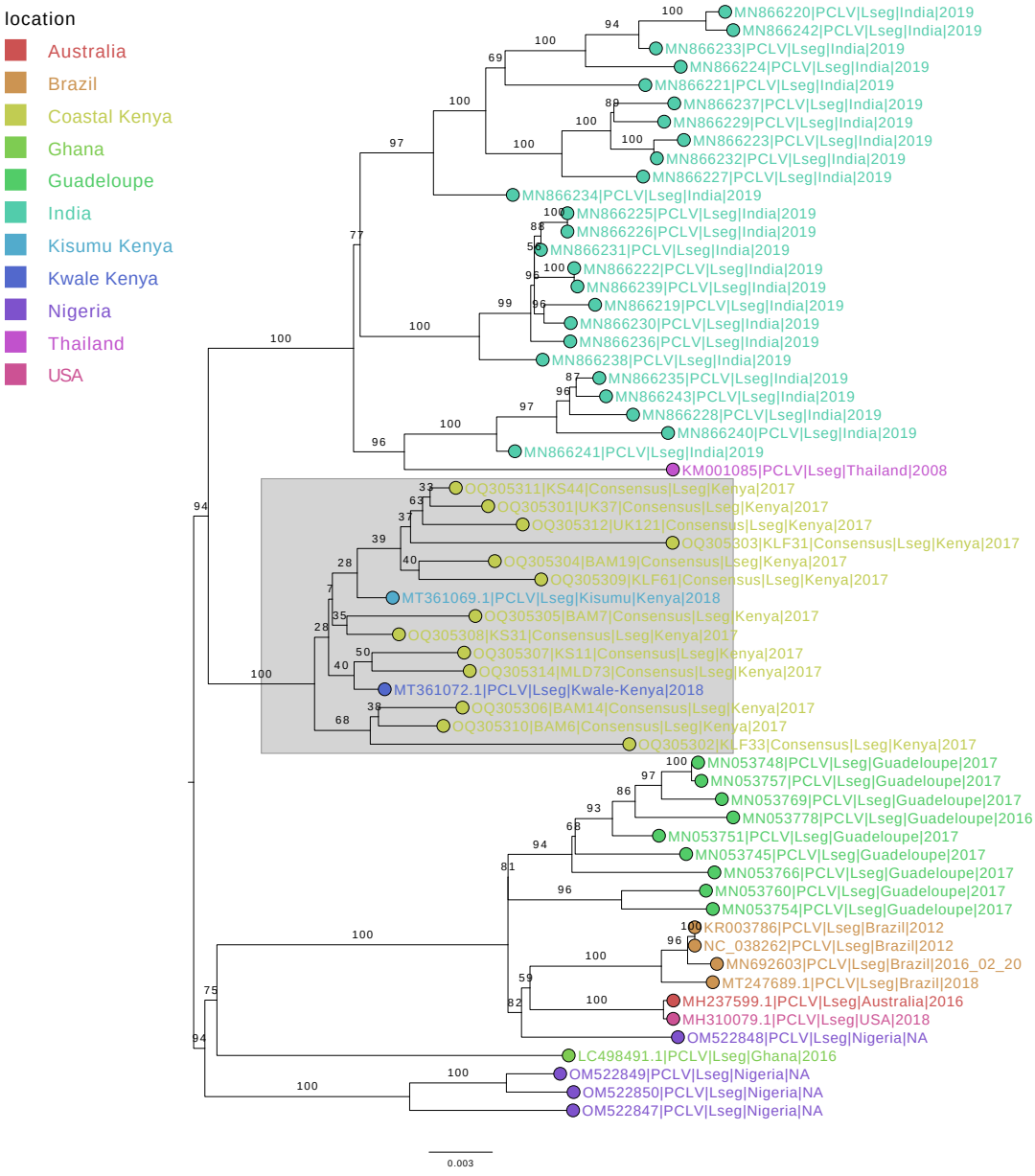

Figure 4.15: Maximum Likelihood Phylogenetic Tree of Phasi Charoen-like Phasivirus (PCLV) Large Segment (Lseg). The tree indicates the evolutionary relationship of PCLV's Lseg from this study (highlighted) and global sequences retrieved from GenBank. Origin of the sequences are indicated at the branch tips as differentially colored circles and texts. Branch supports are indicated as bootstrap values following 1,000 iterations.

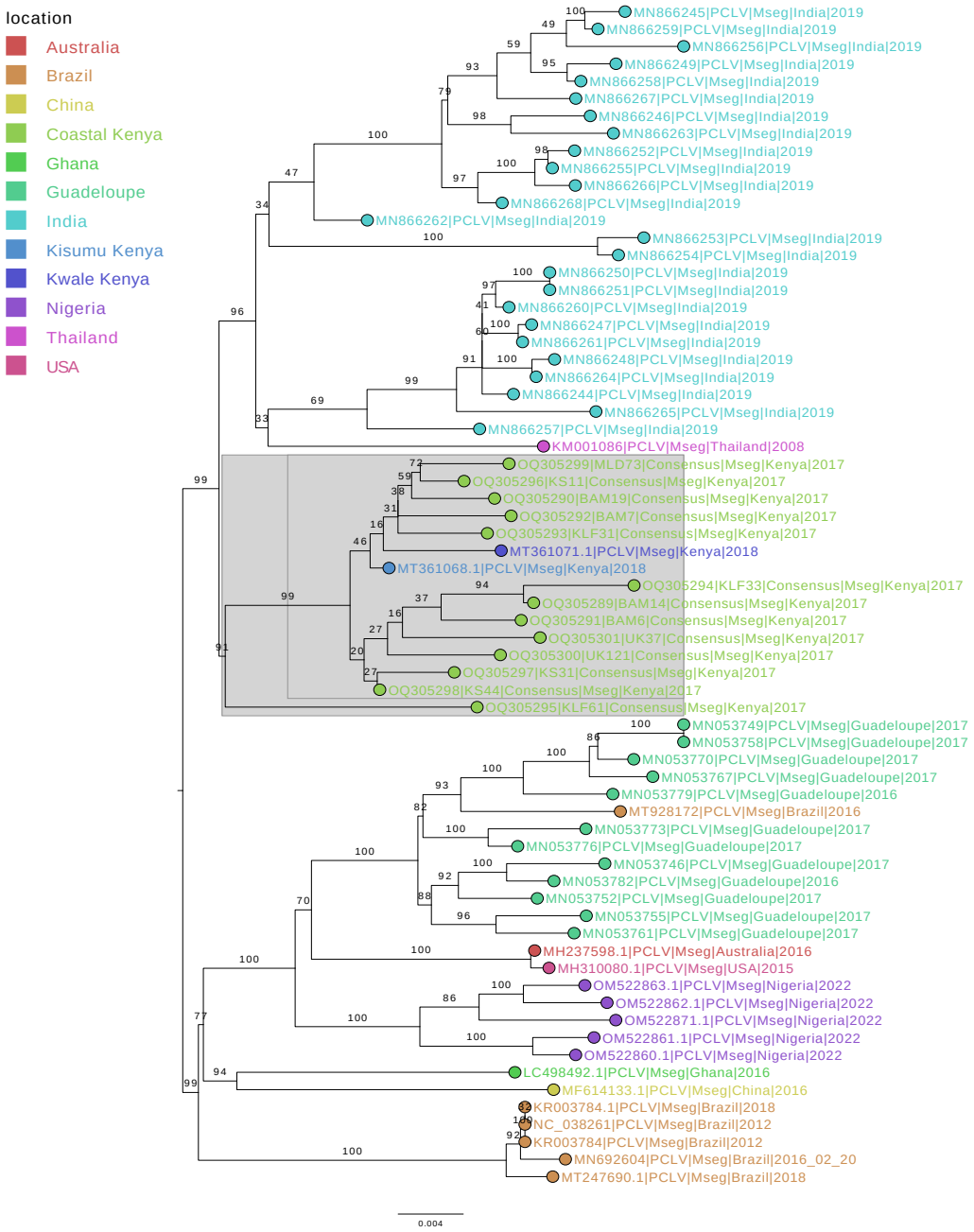

Figure 4.16: Maximum Likelihood Phylogenetic Tree of Phasi Charoen-like Phasivirus (PCLV) Medium Segment (Mseg). The tree indicates the evolutionary relationship of PCLV's Mseg from this study (highlighted) and global sequences retrieved from GenBank. Origin of the sequences are indicated at the branch tips as differentially colored circles and texts. Branch supports are indicated as bootstrap values following 1,000 iterations.

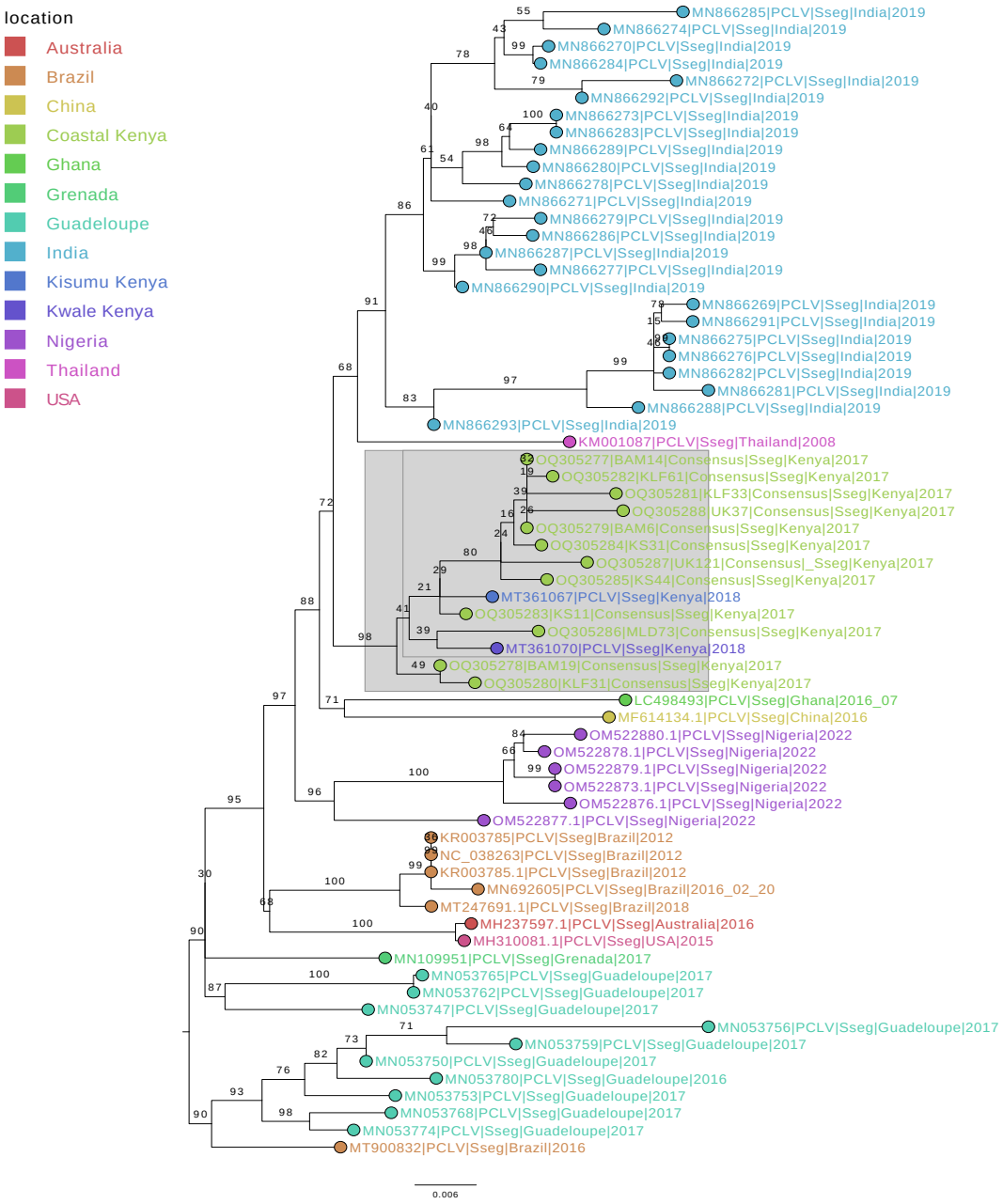

Figure 4.17: Maximum Likelihood Phylogenetic Tree of Phasi Charoen-like Phasivirus (PCLV) Small Segment (Sseg). The tree indicates the evolutionary relationship of PCLV's Sseg from this study (highlighted) and global sequences retrieved from GenBank. Origin of the sequences are indicated at the branch tips as differentially colored circles and texts. Branch supports are indicated as bootstrap values following 1,000 iterations.

CHAPTER FIVE

DISCUSSION

5.1 Discussion

This study identified diverse ISVs from 16 families in *Aedes aegypti* using mNGS. Mosquitoes' studies are important as they are vectors for pathogenic arboviruses in humans but also harbor non-pathogenic insect specific viruses (Agboli et al., 2019; Ferreira-De-Lima & Lima-Camara, 2018; Romo et al., 2018). Metagenomic studies of different mosquito species from different regions in Africa have provided insights into the viral diversity in mosquitoes (Amoa-Bosompem et al., 2020; Atoni et al., 2018; Auguste et al., 2015a; Fauver et al., 2019). Despite the arbovirus infection burden in coastal Kenya being primarily mediated by *Aedes aegypti*, no metagenomic analysis has been conducted on the local populations of this mosquito. The only mosquito metagenomics study in the region was on *Culex* species in 2018 (Atoni et al., 2018). This study utilized the metagenomics approach to elucidate on the extant viral diversity of local *Aedes aegypti* populations.

This study focused on female *Aedes aegypti* as they carry active infections and are responsible for human-to-human and animal host-to-human transmission of the common arboviruses (Gubler, 1998; Gupta et al., 2012). The presence of arboviruses in naturally infected *Aedes aegypti* mosquitoes is generally considered to be low due to low vector densities and focal nature of human-mosquito interactions (Gu & Novak, 2004; Stoddard et al., 2013). Molecular detection of arboviruses thus requires a large number of mosquito collections. To reduce overall cost of arbovirus surveillance and increase chances of arbovirus detection, many studies have pooled mosquitoes in the range of 15–50 per pool (Atoni et al., 2018; Fauver et al., 2016; Oguzie et al., 2022; Ramos-Nino et al., 2020; Shi et al., 2019). Pools comprised of 20 mosquitoes after collections were generated. The number of pools varied per site. Differences in mosquito collections could partly be attributed to availability of breeding habitats (Ferede et al., 2018). Destruction of these breeding habitats leads to reductions in mosquito larva, adult mosquito densities and arbovirus transmission.

This study identified ISV families that have previously been identified in *Ae. aegypti*, a confirmation that these viruses are ubiquitous and naturally infect this mosquito species (Ashok et al., 2021; Auguste et al., 2015b, 2015a; Baidaliuk et al., 2020; Cunha et al., 2020; Oguzie et al., 2022; Sang et al., 2003; Zakrzewski et al., 2018). Since mosquitoes harbor a diverse microbiota, bacteriophages of mosquito endosymbionts were also detected. Additionally, given female mosquitoes feed on both blood and nectar, and breed in different environments, some viral sequences with environmental or dietary origin were identified. Viruses with good genome coverage (>70%) including CFAV, PCLV, Aedes partiti-like 1 virus and HTV were focused on. Further, the evolutionary history of CFAV and PCLV were studies as only these had substantial sequences for comparison.

The CFAV belongs to family *Flaviviridae* and was the first ISV to be isolated in *Aedes aegypti* mosquitoes (Stollar & Thomas, 1975). The first flavivirus to be detected in Kenya belonging to same family as CFAV is KRV. KRV was isolated in 1999 from adult *Aedes macintoshi* mosquitoes whose larvae and pupae were collected from dambos in Central province and reared to adulthood (Sang et al., 2003). However, since its discovery, KRV has not been detected in *Aedes* spp mosquitoes in the country. On the other hand, mosquito molecular surveillance studies have detected partial gene (*NS5*) of CFAV in field *Aedes aegypti* mosquitoes from the western region of Kenya (Ajamma et al., 2018; Chiuya et al., 2021; Iwashita et al., 2018). Whole genome phylogenetic analysis revealed the Kilifi sequence clustered within a major clade comprised of global sequences, with the closest relatives being sequences from Thailand. On the contrary, analysis based on the *NS5* gene previously collected in western Kenya showed Kenyan sequences belonged to four lineages, and the Kilifi sequence was distant from Thailand sequences. In the first lineage, one sequence from Homa Bay Kenya (MG372051.1, collected in 2012) clustered with sequences from Africa (Uganda and Ghana), which belonged to a global clade predominated by non-African sequences. In the second lineage, Kenyan sequences collected in 2018 and 2019 formed a unique cluster that shared a common ancestor with a clade comprised of global sequences. In the third lineage, the sequence from Kilifi belonged to a cluster comprised of a few sequences collected in 2012 in western and coastal regions of Kenya. The fourth lineage was

comprised of sequences from Homa Bay in western Kenya collected in 2012. Differential clustering of CFAV sequences from the same geographical location has been observed before. CFAV phylogeny based on both full genome and partial sequences showed sequences from Thailand and Cambodia belonged to separate clades (Baidaliuk et al., 2020). The same study showed the *Aedes aegypti* population from both countries were distinct based on analysis of mitochondrial genes, and given ISVs are vertically transmitted, it is probable host genetics plays a role in the evolution of CFAV(Baidaliuk et al., 2020). Taken together, these results are indicative of an ongoing evolution of CFAV probably driven by the host, localized dissemination, and sustained transmission of the virus over the years. However, more CFAV sequences from Kenya and Africa are needed to support this hypothesis. The interaction between CFAV and arboviruses has been investigated and one study showed CFAV isolated from wild *Aedes aegypti* reduced the infection and transmission of DENV-1 and ZIKV (Baidaliuk et al., 2019). Since CFAV reads were detected in only one mosquito pool it is hard to conclude with certainty that the presence of this virus led to the absence of arboviruses.

First isolated using C6/36 cells from the larvae of wild-caught *Aedes aegypti* mosquitoes in the Phasi Charoen district in Thailand, PCLV has subsequently been detected in wild *Aedes aegypti* mosquitoes collected from other geographical locations (Atoni et al., 2018; Oguzie et al., 2022; Ramos-Nino et al., 2020; Shi et al., 2019, 2020b). The PCLV segments in this study shared 94-97% nucleotide similarities (BLASTn) with PCLV segments detected in *Aedes aegypti* mosquitoes from Kisumu and Kwale districts in Kenya in 2018 (Shi et al., 2020a). Phylogenetic analysis showed Kenyan sequences belonged to a unique cluster within a major clade comprised of sequences mainly from Asia. The short branch within the Kenyan cluster is indicative of the limited diversity of PCLV within the country. A similar study in Nigeria also reported unique clustering of PCLV segments collected from different sites within the country (Oguzie et al., 2022). The unique clustering of sequences from Africa is suggestive of within-country circulation and diversity, probably also influenced by host genetic diversity in *Aedes aegypti* from different populations within the same geographical location (Baidaliuk et al., 2020). However, more sequences and studies of *Aedes aegypti* genetic structure from populations

across Africa are needed to ascertain this hypothesis. The role of PCLV in the transmission of arboviruses is not well known. Previous studies of arboviral transmission inhibition by ISVs observed that inhibition was more successful when both viruses belonged to the same family (Baidaliuk et al., 2019; Kenney et al., 2014; Kent et al., 2010; Koh et al., 2021; Romo et al., 2018). Wild *Aedes aegypti* mosquitoes collected in arbovirus-endemic regions of Brazil and India have been found to be coinfected with PCLV, CHIKV and DENV 1 and 3(Ashok et al., 2021; Cunha et al., 2020) . It is probable that PCLV does not inhibit transmission of CHIKV and DENV, which belong to distant families from PCLV. Therefore, further studies are needed to elucidate the interaction between PCLV and RVFV, both members of family *Phenuiviridae,* and also the interaction with other arboviruses in field collected *Aedes aegypti* mosquitoes.

Partitiviruses, family *Partitiviridae,* are segmented dsDNA viruses that infect a wide host range including plants, fungi, and arthropods. In Africa, different species of partitiviruses have been detected in wild *Aedes aegypti* and *Anopheles gambiae* from West Africa (Fauver et al., 2016; Oguzie et al., 2022). This study identified *Aedes* partiti-like virus 1. BLASTn analysis showed these contigs had >98% nucleotide identity to *Aedes* partiti-like virus 1 isolate Jane capsid (accession no BK059431.1) and *RdRp* genes (accession no BK059430.1). Partitiviruses have not been shown to infect humans and their interaction with arboviruses in mosquitoes remain unknown.

Humaita-Tubiacanga (HTV) is an unclassified virus that was first identified in *Aedes aegypti* from Brazil (Aguiar et al., 2015). It has subsequently been detected in wild mosquitoes from Bangkok-Thailand, Cairns-Australia, Guadeloupe and Nigeria (Oguzie et al., 2022; Shi et al., 2019; Zakrzewski et al., 2018). This study assembled contigs for both *C* and *RdRp* genes. BLASTn analysis showed these sequences shared > 97% and > 97.56% nucleotide identity to *RdRp* (accession no KR003801.1) and *C* segment sequences assembled from wild *Aedes aegypti* in Guadeloupe (accession no MN053812.1), respectively. The interaction of HTV and arboviruses is also not yet known.

Fako virus, family *Spinareoviridae*, is a multi-segmented dsRNA virus that was first isolated from wild *Aedes albopictus* mosquitoes in Cameroon (Auguste et al., 2015a). Subsequently, it was detected in *Aedes aegypti* mosquitoes from Nigeria (Oguzie et al., 2022). Although this study did not assemble full length sequences for any of the 9 segments that constitute the genome, contigs that matched to segments 2,3,5,6,7 and 8 were generated, with the longest contigs being 649bp, 566bp, 398bp, 436bp,212bp and 514bp, respectively. Fako virus, like partitiviruses, has not been reported to infect humans.

Although this study only recovered small fractions of the genome, the sequences were identical to totivirus. Totiviruses, family *Totiviridae*, have a linear dsRNA genome comprised of two overlapping open reading frames (ORFs) encoding *C* and *RdRp* genes. Previously, totiviruses were known to only infect plants and protists. However, owing to mNGS, totiviruses have increasingly been described in arthropods including mosquito, ants, ticks, fruit fly, shrimp and horseshoe crab, and even piscine species in recent times (Sandlund et al., 2021). A study of West African *Anopheles gambiae* mosquitoes showed the identified totiviruses phylogenetically clustered with totiviruses detected in other arthropods (Fauver et al., 2016). Nonetheless, totiviruses are yet to be accepted as ISVs. Even though it is likely the totivirus identified in this study could have an environmental or dietary (nectar) origin, detection in *Aedes aegypti* contributes to increasing evidence that this virus infects arthropods.

This study also generated rotavirus A contigs. Rotavirus is the leading cause of diarrhea in children below the age of five years but has not been detected before in mosquitoes. The rotavirus genome segments were not present across all the samples and were missing in the negative control. In addition, the identified contigs shared 99-100% nucleotide identity with rotavirus from India for non-structural protein (NSP) genes 2, 3, 4, and 6 based on BLASTn analysis. Moreover, the VP4 region was related to sequences from either Kilifi, Nairobi or Java while NSP5 was related to sequences from Kiambu-Kenya or the USA. Since our lab also does rotavirus sequencing, it is expected that carry over contamination from sequencing would have resulted in a majority of the contigs sharing high sequence similarity with rotavirus

from Kilifi deposited in GenBank. In this regard, it is likely the rotavirus genome segments detected in this study are laboratory contamination or, given whole mosquito lysates were used, these viral segments were picked by the mosquitoes from their environment for example sewage water.

5.2 Limitations of the Study

One limitation of this study is that metagenomics did not detect any arboviruses (even though sanger sequencing detected WNV in one pool), probably due to low sensitivity of mNGS or low viral load in the vectors. This is not unique as mNGS did not detect YFV in mosquitoes sampled during the 2017-2020 YFV outbreak in Nigeria (Oguzie et al., 2022). Additionally, sampling was done outside an outbreak period, which probably contributed to lack of detection of arboviruses. With regards to the difference in the detected viral profiles between mNGS (unbiased) and sanger (amplicon-based), this can be attributed to difference in methods and low sensitivity of mNGS. Another shortcoming of the study is that our analysis focused on known mosquito viruses. Considering a significant proportion of reads were unclassified, future works should focus on viral discovery and characterization. Knowledge of the newly identified viruses will shed light on their genome organization, evolution, genetic diversity and potential to cause diseases in humans, plants and animals. Lastly, this study had a small sample size and only probed *Aedes aegypti* species. Future studies should have larger sample sizes comprised of an array of different mosquito species in order to establish the core virome of mosquitoes from coastal Kenya.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Overall, this study has shown the virome of female *Aedes aegypti* populations from coastal Kenya did not have medically important arboviruses but had a high abundance of insect specific viruses (ISVs). Further, it has shown that family *Phenuiviridae* was the most abundant ISV, and viral species CFAV (family *Flaviviridae*) and PCLV (family *Phenuiviridae*) present in these mosquitoes were phylogenetically distinct from those circulating in Africa and the rest of the world. *In vitro* studies have shown that ISVs can either enhance or suppress the replication of medically important viruses. In the absence of widespread epidemiological surveillance, the potential for the identified ISVs to demonstrate their ability to suppress or enhance arbovirus replication in nature remains unknown. Future studies are needed to study phenotypic and genetic characteristics, as well as the possibility of human or animal infection by the viruses identified in this study.

6.2 Recommendations

This study characterized known mosquito viruses. Since a significant proportion of reads were of unknown origin (unclassified), future works should also attempt virus discovery and characterization. Knowledge of the newly identified viruses will shed light on their genome organization, evolution, genetic diversity, and zoonotic potential. Additionally, this study was based on a small sample size of adult female Aedes *aegypti* pools from coastal Kenya. Future studies should be longitudinal in design and have larger sample sizes comprised of *Aedes* spp pools collected during epidemic and inter-epidemic period across the country in order to establish the national temporal core virome.

REFERENCES

Agboli, E., Leggewie, M., Altinli, M., & Schnettler, E. (2019). Mosquito-specific viruses transmission and interaction. *Viruses*, *11*(9), 1–26. https://doi.org/10.3390/v11090873

Agha, S., Chepkorir, E., Mulwa, F., Tigoi, C.,, Arum, S., Guarido, M., … & Sang, R. (2017). Vector competence of populations of Aedes aegypti from three distinct cities in Kenya for chikungunya virus. *PLoS Negl Trop Dis*, *11*(8). https://doi.org/https://doi.org/10.1371/journal.pntd.0005860

Aguiar, E. R. G. R., Olmo, R. P., Paro, S., Ferreira, F. V., de Faria, I. J. D. S., Todjro, Y. M. H., … & Marques, J. T. (2015). Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. *Nucleic Acids Research*, *43*(13), 6191–6206. https://doi.org/10.1093/nar/gkv587

Ajamma, Y. U., Onchuru, T. O., Ouso, D. O., Omondi, D., Masiga, D. K., & Villinger, J. (2018). Vertical transmission of naturally occurring Bunyamwera and insect-specific flavivirus infections in mosquitoes from islands and mainland shores of Lakes Victoria and Baringo in Kenya. *PLoS Neglected Tropical Diseases*, *12*(11). https://doi.org/10.1371/journal.pntd.0006949

Aliota, M. T., Peinado, S. A., Velez, I. D., & Osorio, J. E. (2016). The wMel strain of Wolbachia Reduces Transmission of Zika virus by Aedes aegypti. *Scientific Reports*, *6*(July), 1–7. https://doi.org/10.1038/srep28792

Aliota, M. T., Walker, E. C., Uribe Yepes, A., Dario Velez, I., Christensen, B. M., & Osorio, J. E. (2016). The wMel Strain of Wolbachia Reduces Transmission of Chikungunya Virus in Aedes aegypti. *PLoS Neglected Tropical Diseases*, *10*(4), 1–13. https://doi.org/10.1371/journal.pntd.0004677

Amoa-Bosompem, M., Kobayashi, D., Murota, K., Faizah, A. N., Itokawa, K., Fujita, R., … & Iwanaga, S. (2020). Entomological assessment of the status and risk of mosquito-borne arboviral transmission in Ghana. *Viruses*, *12*(2). https://doi.org/10.3390/v12020147

Arankalle, V. A., Shrivastava, S., Cherian, S., Gunjikar, R. S., Walimbe, A. M., Jadhav, S. M., Sudeep, A. B., & Mishra, A. C. (2007). Genetic divergence of Chikungunya viruses in India (1963-2006) with special reference to the 2005-2006 explosive epidemic. *Journal of General Virology*, *88*(7), 1967–1976. https://doi.org/10.1099/vir.0.82714-0

Ashok, M. Dimpal, N., Pragya, Y., Manjushree, R., Savita, P., Triparna, M., … & Manjunath, J. (2021). Identification of Phasi Charoen-Like Phasivirus in Field Collected Aedes aegypti from Karnataka State, India. *Vector Borne Zoonotic Dis* , *21*(11), 900–909.

Atoni, E., Wang, Y., Karungu, S., Waruhiu, C., Zohaib, A., Obanda, V., … & Yuan, Z. (2018). Metagenomic virome analysis of Culex mosquitoes from Kenya and China. *Viruses*, *10*(1), 1–15. https://doi.org/10.3390/v10010030

Auguste, A. J., Kaelber, J. T., Fokam, E. B., Guzman, H., Carrington, C. V. F., Erasmus, J. H., … & Weaver, S. C. (2015a). A Newly Isolated Reovirus Has the Simplest Genomic and Structural Organization of Any Reovirus. *Journal of Virology*, *89*(1), 676–687. https://doi.org/10.1128/jvi.02264-14

Auguste, A. J., Kaelber, J. T., Fokam, E. B., Guzman, H., Carrington, C. V. F., Erasmus, J. H., … & Weaver, S. C. (2015b). A Newly Isolated Reovirus Has the Simplest Genomic and Structural Organization of Any Reovirus. *Journal of Virology*, *89*(1), 676–687. https://doi.org/10.1128/jvi.02264-14

Baidaliuk, A., Lequime, S., Moltini-Conclois, I., Dabo, S., Dickson, L. B., Prot, M., … & Lambrechts, L. (2020). Novel genome sequences of cell-fusing agent virus allow comparison of virus phylogeny with the genetic

structure of Aedes aegypti populations. *Virus Evolution*, *6*(1). https://doi.org/10.1093/VE/VEAA018

Baidaliuk, A., Miot, E.F., Lequime, S., Moltini-Conclois, I., Delaigue, F., Dabo, S., … & Lambrechts, L. (2019). Cell-Fusing Agent Virus Reduces Arbovirus Dissemination in Aedes aegypti Mosquitoes In Vivo. *Journal of Virology*, *93*(18), e00705-19.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Batovska, J., Lynch, S. E., Rodoni, B. C., Sawbridge, T. I., & Cogan, N. O. (2017). Metagenomic arbovirus detection using MinION nanopore sequencing. *Journal of Virological Methods*, *249*, 79–84. https://doi.org/10.1016/j.jviromet.2017.08.019

Bennett, K. E., Olson, K. E., De, M., Mun˜oz, L., Mun˜oz, M., Fernandez-Salas, I., … & Beaty, B. J. (2002). Variation in vector competence for dengue 2 virus among 24 collections of aedes aegypti from mexico and the united states. *American Journal of Tropical Medicine and Hygiene*, *67*(1), 85-92.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., … & Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, *496*(7446), 504–507. https://doi.org/10.1038/nature12060

Bolling, B. G., Olea-Popelka, F. J., Eisen, L., Moore, C. G., & Blair, C. D. (2012). Transmission dynamics of an insect-specific flavivirus in a naturally infected Culex pipiens laboratory colony and effects of co-infection on vector competence for West Nile virus. *Virology*, *427*(2), 90–97. https://doi.org/10.1016/j.virol.2012.02.016

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7055E.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009a). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*. https://doi.org/10.1186/1471-2105-10-421

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009b). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*. https://doi.org/10.1186/1471-2105-10-421

Cardona-Ospina, J. A., Henao-SanMartin, V., Acevedo-Mendoza, W. F., Nasner-Posso, K. M., Martínez-Pulgarín, D. F., Restrepo-López, A., … & Rodriguez-Morales, A. J. (2019). Fatal Zika virus infection in the Americas: A systematic review. *International Journal of Infectious Diseases*, *88*, 49–59. https://doi.org/10.1016/j.ijid.2019.08.033

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chepkorir, E., Lutomiah, J., Mutisya, J., Mulwa, F., Limbaso, K., Orindi, B., Ng'Ang'a, Z., & Sang, R. (2014). Vector competence of Aedes aegypti populations from Kilifi and Nairobi for dengue 2 virus and the influence of temperature. *Parasites and Vectors*, *7*(1), 1–8.

Chiuya, T., Masiga, D. K., Falzon, L. C., Bastos, A. D. S., Fèvre, E. M., & Villinger, J. (2021). A survey of mosquito-borne and insect-specific viruses in hospitals and livestock markets in western Kenya. In *PLoS one*, *16*(5). Public Library of Science. https://doi.org/10.1371/journal.pone.0252369

Cunha, M. dos P., Ioshino, R. S., Costa-da-Silva, A. L., Petersen, V., Capurro, M. L., & Zanotto, P. M. de A. (2020). A Metagenomic Approach Identified a Novel Phasi Charoen-Like Virus Coinfecting a Chikungunya Virus-Infected Aedes aegypti Mosquito in Brazil. *Microbiology Resource Announcements*, *9*(31). https://doi.org/10.1128/mra.01572-19

Duarte, M. A., Campos, F. S., Araújo Neto, O. F., Silva, L. A., Silva, A. B., Aguiar, T. C., … & Ribeiro, B. M. (2022). Identification of potential new mosquito-associated viruses of adult Aedes aegypti mosquitoes from Tocantins state, Brazil. *Brazilian Journal of Microbiology*, *53*(1), 51–62. https://doi.org/10.1007/s42770-021-00632-x

Ellis, E. M., Neatherlin, J. C., Delorey, M., Ochieng, M., Mohamed, A. H., Mogeni, D. O., … & Tomashek, K. M. (2015). A Household Serosurvey to Estimate the Magnitude of a Dengue Outbreak in Mombasa, Kenya, 2013. *PLoS Neglected Tropical Diseases*, *9*(4). https://doi.org/10.1371/journal.pntd.0003733

Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., & Hayashi, M. (2005). Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Research*, *33*(6), 1–11. https://doi.org/10.1093/nar/gni064

European Centre for Disease Prevention and Control. (2023). *Chikungunya virus disease cases January - December 2023*. Retrieved from https://www.ecdc.europa.eu/en/publications-data/chikungunya-virus-disease-cases-january-december-2023#:~:text=In%202023% 20and%20 as%20of, deaths%20had%20been%20reported%20worldwide.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/ bioinformatics/ btw354

Eyase, F., Langat, S., Berry, I. M., Mulwa, F., Nyunja, A., Mutisya, J., … & Sang, R. (2020). Emergence of a novel chikungunya virus strain bearing the E1:V80A substitution, out of the Mombasa, Kenya 2017-2018 outbreak. *PLoS one*, *15*(11), 1–14. https://doi.org/10.1371/journal.pone.0241754

Farenhorst, M., Mouatcho, J. C., Kikankie, C. K., Brooke, B. D., Hunt, R. H., Thomas, M. B., … & Coetzee, M. (2009). Fungal infection counters insecticide resistance in African malaria mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17443–17447. https://doi.org/10.1073/pnas.0908530106

Fauver, J. R., Akter, S., Morales, A. I. O., Black, W. C., Rodriguez, A. D., Stenglein, M. D., … & Weger-Lucarelli, J. (2019). A reverse-transcription/RNase H based protocol for depletion of mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and pathogen discovery. *Virology*, *528*(December 2018), 181–197. https://doi.org/10.1016/j.virol.2018.12.020

Fauver, J. R., Grubaugh, N. D., Krajacich, B. J., Weger-Lucarelli, J., Lakin, S. M., Fakoli, L. S., … & Stenglein, M. D. (2016). West African Anopheles gambiae mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses. *Virology*, *498*, 288–299. https://doi.org/10.1016/j.virol.2016.07.031

Ferede, G., Tiruneh, M., Abate, E., Kassa, W. J., Wondimeneh, Y., Damtie, D., & Tessema, B. (2018). Distribution and larval breeding habitats of Aedes mosquito species in residential areas of northwest Ethiopia. *Epidemiology and Health*, *40*, e2018015. https://doi.org/10.4178/epih.e2018015

Ferreira-De-Lima, V. H., & Lima-Camara, T. N. (2018). Natural vertical transmission of dengue virus in Aedes aegypti and Aedes albopictus: A systematic review. In *Parasites and Vectors*, *11*(1). BioMed Central. https://doi.org/10.1186/s13071-018-2643-9

Frentiu, F. D., Zakir, T., Walker, T., Popovici, J., Pyke, A. T., van den Hurk, A., … & O'Neill, S. L. (2014). Limited Dengue Virus Replication in Field-Collected Aedes aegypti Mosquitoes Infected with Wolbachia. *PLoS*

Neglected Tropical Diseases, *8*(2), 1–10. https://doi.org/10.1371/journal.pntd.0002688

Gangopadhayya, A., Lole, K., Ghuge, O., Ramdasi, A., Kamble, A., Roy, D., Thakar, S., Nath, A., Sudeep, A. B., & Cherian, S. (2024). Metagenomic Analysis of Viromes of Aedes Mosquitoes across India. *Viruses*, *16*(1). https://doi.org/10.3390/v16010109

Gillies, M. T., & De Meillon, B. (1968). *The Anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region). The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region)* (Vol. 54). Johannesburg.

Giry, C., Roquebert, B., Li-Pat-Yuen, G., Gasque, P., & Jaffar-Bandjee, M. C. (2017). Improved detection of genus-specific Alphavirus using a generic TaqMan® assay. *BMC Microbiology*, *17*(1). https://doi.org/10.1186/s12866-017-1080-9

Gloria-soria, A., Ayala, D., Bheecarry, A., Calderon-arguedas, O., Chadee, D. D., Chiappero, M., … & Health, G. (2017). *Global Genetic Diversity of Aedes aegypti*. *25*(21), 5377–5395. https://doi.org/10.1111/mec.13866.Global

Gomez, M. M. D. M. M. R. J. D. (2022). Aedes aegypti and Ae. albopictus microbiome/virome: new strategies for controlling arboviral transmission? *Parasites Vectors*, *15*(287).

Gu, W., & Novak, R. J. (2004). Short Report: Detection Probability Of Arbovirus Infection In Mosquito Populations. In *J. Trop. Med. Hyg*, *71*(5).

Gubler, D. J. (1998). Dengue and Dengue Hemorrhagic Fever. In *Clinical Microbiology Reviews*, *11*(3).

Gupta, N., Srivastava, S., Jain, A., & Chaturvedi, U. C. (2012). Dengue in India. In *Indian J Med Res*, *136*.

Hall, T. A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids*, *41*, 95–98.

He, X., Yin, Q., Zhou, L., Meng, L., Hu, W., Li, F., … & Wang, H. (2021). Metagenomic sequencing reveals viral abundance and diversity in mosquitoes from the Shaanxi-Gansu-Ningxia region, China. *PLoS Neglected Tropical Diseases*, *15*(4). https://doi.org/10.1371/journal.pntd.0009381

Hobson-Peters, J., Yam, A. W. Y., Lu, J. W. F., Setoh, Y. X., May, F. J., Kurucz, N., … & Hall, R. A. (2013). A New Insect-Specific Flavivirus from Northern Australia Suppresses Replication of West Nile Virus and Murray Valley Encephalitis Virus in Co-infected Mosquito Cells. *PLoS One*, *8*(2), 1–12. https://doi.org/10.1371/journal.pone.0056534

Houé, V., Bonizzoni, M., & Failloux, A. B. (2019). Endogenous non-retroviral elements in genomes of Aedes mosquitoes and vector competence. In *Emerging Microbes and Infections*, *8*(1), 542–555). Taylor and Francis Ltd. https://doi.org/10.1080/22221751.2019.1599302

Iv Edwards, B. F. (1941). *Mosquitoes of the ethiopian region iii.-culicine adults and pupae london : printed by order of the trustees sold at the rritish museum (natc'ral history), cromwell road, S.iV 1941 (A 11 rights reserved).*

Iwashita, H., Higa, Y., Futami, K., Lutiali, P. A., Njenga, S. M., Nabeshima, T., & Minakawa, N. (2018). Mosquito arbovirus survey in selected areas of Kenya: Detection of insect-specific virus. *Tropical Medicine and Health*, *46*(1). https://doi.org/10.1186/s41182-018-0095-8

Karisa, J., Muriu, S., Omuoyo, D., Karia, B., Ngari, M., Nyamwaya, D., … & Mbogo, C. M. (2021). Urban ecology of arboviral mosquito vectors along the kenyan coast. *Journal of Medical Entomology*, *58*(1), 428–438. https://doi.org/10.1093/jme/tjaa136

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kenney, J. L., Solberg, O. D., Langevin, S. A., & Brault, A. C. (2014). Characterization of a novel insect-specific flavivirus from Brazil: Potential for inhibition of infection of arthropod cells with medically important flaviviruses. *Journal of General Virology*, *95*(May), 2796–2808. https://doi.org/10.1099/vir.0.068031-0

Kent, R.J.; Crabtree, M.B.; Miller, B. R. (2010). Transmission of West Nile Virus by Culex Quinquefasciatus Say Infected with Culex Flavivirus Izabal. *PLoS Negl. Trop. Dis.*, *4*, e671.

Ketkar, H., Herman, D., & Wang, P. (2019). Genetic determinants of the re-emergence of arboviral diseases. In *Viruses*, *11*(2). MDPI AG. https://doi.org/10.3390/v11020150

Sergon, K., Njuguna, C., Kalani, R., Ofula, V., Onyango, C., Konongoi, L. S., ... & Breiman, R. F. (2008). Seroprevalence of chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. *The American journal of tropical medicine and hygiene*, *78*(2), 333-337.

Koh, C., Henrion-Lacritick, A., Frangeul, L., & Saleh, M. C. (2021). Interactions of the insect-specific palm creek virus with zika and chikungunya viruses in aedes mosquitoes. *Microorganisms*, *9*(8), 14–18. https://doi.org/10.3390/microorganisms9081652

Konongoi, L., Ofula, V., Nyunja, A., Owaka, S., Koka, H., Makio, A., … & Sang, R. (2016). Detection of dengue virus serotypes 1, 2 and 3 in selected regions of Kenya: 2011-2014. *Virology Journal*, *13*(1), 1–11. https://doi.org/10.1186/s12985-016-0641-0

Kramer, L. D., & Ciota, A. T. (2015). Dissecting vectorial capacity for mosquito-borne viruses. In *Current Opinion in Virology, 15*, 112–118. Elsevier B.V. https://doi.org/10.1016/j.coviro.2015.10.003

LaBeaud, A. D., Bashir, F., & King, C. H. (2011). Measuring the burden of arboviral diseases: The spectrum of morbidity and mortality from four prevalent infections. *Population Health Metrics*, *9*. https://doi.org/10.1186/1478-7954-9-1

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/ 10.1093/ bioinformatics/btu531

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/ 10.1093/ bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/ 10.1093/bioinformatics/btp352

Lounibos, L. P., & Kramer, L. D. (2016). Invasiveness of aedes aegypti and aedes albopictus and vectorial capacity for chikungunya virus. *Journal of Infectious Diseases*, *214*(Suppl 5), S453–S458. https://doi.org/10. 1093/infdis/jiw285

Lutomiah, J., Bast, J., Clark, J., Richardson, J., Yalwala, S., Oullo, D., … & Sang, R. (2013). Abundance, diversity, and distribution of mosquito vectors in selected ecological regions of Kenya: Public health implications. *Journal of Vector Ecology*, *38*(1), 134–142. https://doi.org/10.1111/j.1948-7134.2013.12019.x

Lwande, O. W., Obanda, V., Lindström, A., Ahlm, C., Evander, M., Näslund, J., & Bucht, G. (2020). Globe-Trotting Aedes aegypti and Aedes albopictus:

Risk Factors for Arbovirus Pandemics. In *Vector-Borne and Zoonotic Diseases* (Vol. 20, Issue 2, pp. 71–81). Mary Ann Liebert Inc. https://doi.org/10.1089/vbz.2019.2486

Masika, M. M., Korhonen, E. M., Smura, T., Uusitalo, R., Vapalahti, K., Mwaengo, D., … & Huhtamo, E. (2020). Detection of dengue virus type 2 of Indian origin in acute febrile patients in rural Kenya. *PLoS Neglected Tropical Diseases*, *14*(3), 1–17. https://doi.org/10.1371/journal.pntd.0008099

Massengo, N. R. B., Tinto, B., & Simonin, Y. (2023). One Health Approach to Arbovirus Control in Africa: Interests, Challenges, and Difficulties. *Microorganisms*, *11*(6). https://doi.org/10.3390/microorganisms11061496

Miesen, P., & van Rij, R. P. (2019). Crossing the Mucosal Barrier: A Commensal Bacterium Gives Dengue Virus a Leg-Up in the Mosquito Midgut. *Cell Host and Microbe*, *25*(1), 1–2. https://doi.org/10.1016/j.chom.2018.12.009

Moore, M, Sylla, M, Goss, L, Burugu, MW, Sang, R, Kamau, LW, Kenya, E, Bosio, C, Munoz, M, Sharakova, M, B. W. (2013). Dual African Origins of Global Aedes aegypti s.l. Populations Revealed by Mitochondrial DNA. *PLoS Negl Trop Dis*, *7*(4), e2175. https://doi.org/https://doi.org/10.1371/journal.pntd.0002175

Mutebi, J.-P., & Barrett, A. D. T. (2002). *The epidemiology of yellow fever in Africa*. Retrieved from www.elsevier.com/locate/micinf

Mwaengo, D., Lorenzo, G., Iglesias, J., Warigia, M., Sang, R., Bishop, R. P., & Brun, A. (2012). Detection and identification of Rift Valley fever virus in mosquito vectors by quantitative real-time PCR. *Virus Research*, *169*(1), 137–143. https://doi.org/10.1016/j.virusres.2012.07.019

Mwangangi, J.M., Midega, J., Kahindi, S., Njoroge, L., Nzovu, J., Githure, J., Mbogo, C.M. & Beier, J. C. (2012). Mosquito species abundance and diversity in Malindi, Kenya and their potential implication in pathogen

transmission. *Parasitol Res*, *10*, 61–71. https://doi.org/https://doi.org/10.1007/s00436-011-2449-6

Nasar, F., Erasmus, J. H., Haddow, A. D., Tesh, R. B., & Weaver, S. C. (2015). Eilat virus induces both homologous and heterologous interference. *Virology*, *484*, 51–58. https://doi.org/10.1016/j.virol.2015.05.009

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-Tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Njenga, M. K., Nderitu, L., Ledermann, J. P., Ndirangu, A., Logue, C. H., Kelly, C. H. L., … & Powers, A. M. (2008). Tracking epidemic Chikungunya virus into the Indian Ocean from East Africa. *Journal of General Virology*, *89*(11), 2754–2760. https://doi.org/10.1099/vir.0.2008/005413-0

Obonyo, M., Fidhow, A., & Ofula, V. (2018). Investigation of laboratory confirmed dengue outbreak in north-eastern Kenya, 2011. *PLoS One*, *13*(6), 1–11. https://doi.org/10.1371/journal.pone.0198556

Oguzie, J. U., Nwangwu, U. C., Oluniyi, P. E., Olumade, T. J., George, U. E., Kazeem, A., … & Happi, C. T. (2022). Metagenomic sequencing characterizes a wide diversity of viruses in field mosquito samples in Nigeria. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-11797-2

Patel, P., Landt, O., Kaiser, M., Faye, O., Koppe, T., Lass, U., Sall, A. A., & Niedrig, M. (2013). Development of one-step quantitative reverse transcription PCR for the rapid detection of flaviviruses. *Virology Journal*, *10*. https://doi.org/10.1186/1743-422X-10-58

Platzer, A., Polzin, J., Rembart, K., Han, P. P., Rauer, D., & Nussbaumer, T. (2018). BioSankey: Visualization of Microbial Communities Over Time. *Journal*

of Integrative Bioinformatics, *15*(4). https://doi.org/10.1515/jib-2017-0063

Pollett, S., Gathii, K., Figueroa, K., Rutvisuttinunt, W., Srikanth, A., Nyataya, J. N., … & Waitumbi, J. N. (2019). The evolution of dengue-2 viruses in Malindi, Kenya and greater East Africa: Epidemiological and immunological implications. *BioRxiv*. https://doi.org/10.1101/850107

Powell, J. R., & Tabachnick, W. J. (2013). History of domestication and spread of Aedes aegypti--a review. *Memórias Do Instituto Oswaldo Cruz*, *108*(October), 11–17. https://doi.org/10.1590/0074-0276130395

R Core Team. (2022). *R: A language and environment for statistical computing* (4.2.0). R Foundation for Statistical Computing.

Rambaut, A. (2010). *FigTree* (1.3.1). Institute of Evolutionary Biology, University of Edinburgh.

Ramos-Nino, M. E., Fitzpatrick, D. M., Eckstrom, K. M., Tighe, S., Hattaway, L. M., Hsueh, A. N., … & Cheetham, S. (2020). Metagenomic analysis of Aedes aegypti and Culex quinquefasciatus mosquitoes from Grenada, West Indies. *PLoS one*, *15*(4), 1–16. https://doi.org/10.1371/journal.pone.0231047

Romo, H., Kenney, J. L., Blitvich, B. J., & Brault, A. C. (2018). Restriction of Zika virus infection and transmission in Aedes aegypti mediated by an insect-specific flavivirus. *Emerging Microbes and Infections*, *7*(1). https://doi.org/10. 1038/s41426-018-0180-4

Sandlund, L., Mor, S. K., Singh, V. K., Padhi, S. K., Phelps, N. B. D., Nylund, S., & Mikalsen, A. B. (2021). Comparative molecular characterization of novel and known piscine toti-like viruses. *Viruses*, *13*(6). https://doi.org/10.3390/v13061063

Sang, R. C., Gichogo, A., Gachoya, J., Dunster, M. D., Ofula, V., Hunt, A. R., … & Dunster, L. M. (2003). Isolation of a new flavivirus related to Cell fusing agent virus (CFAV) from field-collected flood-water Aedes mosquitoes sampled from a dambo in central Kenya. *Archives of Virology*, *148*(6), 1085–1093. https://doi.org/10.1007/s00705-003-0018-8

Sergon, K, Yahaya, A.A, Brown, J, Bedja, S.A, Mlindasse, M, Agata, N, … & Njenga MK, B. R. (2007). Seroprevalence of Chikungunya virus infection on Grande Comore Island, union of the Comoros, 2005. *Am J Trop Med Hyg*, *76*(6), 1189–1193.

Shah, M. M., Ndenga, B. A., Mutuku, F. M., Vu, D. M., Grossi-Soyster, E. N., Okuta, V., … & LaBeaud, A. D. (2020). High dengue burden and circulation of 4 virus serotypes among children with undifferentiated fever, Kenya, 2014–2017. *Emerging Infectious Diseases*, *26*(11), 2638–2650. https://doi.org/10.3201/eid2611.200960

Shi, C., Beller, L., Deboutte, W., Yinda, K. C., Delang, L., Vega-Rúa, A., Failloux, A. B., & Matthijnssens, J. (2019). Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. *Microbiome*, *7*(1), 1–20. https://doi.org/10.1186/s40168-019-0734-2

Shi, C., Zhao, L., Atoni, E., Zeng, W., Hu, X., Matthijnssens, J., Yuan, Z., & Xia, H. (2020a). Stability of the Virome in Lab- and Field-Collected Aedes albopictus Mosquitoes across Different Developmental Stages and Possible Core Viruses in the Publicly Available Virome Data of Aedes Mosquitoes . *MSystems*, *5*(5). https://doi.org/10.1128/msystems.00640-20

Shi, C., Zhao, L., Atoni, E., Zeng, W., Hu, X., Matthijnssens, J., Yuan, Z., & Xia, H. (2020b). The conservation of a core virome in Aedes mosquitoes across different developmental stages and continents. *MSystems*, 1–28. https://doi.org/10.1101/2020.04.23.058701

Sophiah Mbaika, S., Lutomiah, J., Chepkorir, E., Mulwa, F., Khayeka-Wandabwa, C., Tigoi, C., … & Sang, R. (2016). Vector competence of Aedes aegypti in transmitting Chikungunya virus: effects and implications of extrinsic incubation temperature on dissemination and infection rates. *Virology Journal*, *13*(114). https://doi.org/DOI 10.1186/s12985-016-0566-7

Souza-Neto, J. A., Powell, J. R., & Bonizzoni, M. (2019). Aedes aegypti vector competence studies: A review. *Infection, Genetics and Evolution*, *67*(November 2018), 191–209. https://doi.org/10.1016/j.meegid. 2018.11.009

Stoddard, S. T., Forshey, B. M., Morrison, A. C., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., … & Scott, T. W. (2013). House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 994–999. https://doi.org/10.1073/pnas.1213349110

Stollar, V., & Thomas, V. (1975). An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. *Virology*, *64*(2), 367–377.

Talavera, S.; Birnberg, L.; Nuñez, A.I.; Muñoz-Muñoz, F.; Vázquez, A.; Busquets, N. (2018). Culex Flavivirus Infection in a Culex pipiens Mosquito Colony and Its Effects on Vector Competence for Rift Valley Fever Phlebovirus. *Parasit. Vectors*, *11*(310).

Thannesberger, J., Rascovan, N., Eisenmann, A., Klymiuk, I., Zittra, C., Fuehrer, H. P., … & Steininger, C. (2021). Viral metagenomics reveals the presence of novel Zika virus variants in Aedes mosquitoes from Barbados. *Parasites and Vectors*, *14*(1). https://doi.org/10.1186/s13071-021-04840-0

Thompson, R. M. D. C. J.C. . (2020). A review of the economic evidence of Aedes-borne arboviruses and Aedes-borne arboviral disease prevention and control strategies. *Expert Review of Vaccines*, *19*(2), 143–162.

Villinger, J., Mbaya, M. K., Ouso, D., Kipanga, P. N., Lutomiah, J., & Masiga, D. K. (2017). Arbovirus and insect-specific virus discovery in Kenya by novel six genera multiplex high-resolution melting analysis. *Molecular Ecology Resources*, *17*(3), 466–480. https://doi.org/10.1111/1755-0998.12584

Volk, S. M., Chen, R., Tsetsarkin, K. A., Adams, A. P., Garcia, T. I., Sall, A. A., … & Weaver, S. C. (2010). Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates. *Journal of Virology*, *84*(13), 6497–6504. https://doi.org/10.1128/jvi.01603-09

Weaver, S. C. (2014). Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. *PLoS Neglected Tropical Diseases*, *8*(6), 6–9. https://doi.org/10.1371/journal.pntd.0002921

WHO. (2017). *Weekly Bulletin on Outbreaks and Other Emergencies: Week 28: 08-14 July 2017, World Health Organization*.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1891-0

World Health Organization (WHO). (2016). *Emergencies preparedness, response: Disease outbreak news, Chikungunya Kenya*. Retrieved from http://www.who.int/csr/don/09-august-2016-chikungunya-kenya/en/.

Xiao, P., Han, J., Zhang, Y., Li, C., Guo, X., Wen, S., … & Jin, N. (2018). Metagenomic Analysis of Flaviviridae in Mosquito Viromes Isolated From Yunnan Province in China Reveals Genes From Dengue and Zika Viruses. *Frontiers in Cellular and Infection Microbiology*, *8*. https://doi.org/10.3389/fcimb.2018.00359

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their

covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zakrzewski, M., Rašić, G., Darbro, J., Krause, L., Poo, Y. S., Filipović, I., … & Suhrbier, A. (2018). Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-22945-y

APPENDICES

Appendix I: Abstract of Peer Reviewed Research Article

Wellcome Open Research

Wellcome Open Research 2023, 8:136 Last updated: 30 AUG 2023

Check for updates

RESEARCH ARTICLE

REVISED

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]

Donwilliams O. Omuoyo[1], Doris K. Nyamwaya[1], Everlyn Kamau[1], James N. Nyagwange[1], Henry K. Karanja[1], John N. Gitonga[1], Zaydah de Laurent[1], Jonathan C. Ngala[2], Johnson Kinyua[3], Martin K. Rono[1], Joseph M. Mwangangi[1], Charles N. Agoti[1], George M. Warimwe[1,4]

[1]KEMRI-Wellcome Trust Research Programme, Center for Geographic Medicine Research, Coast, Kilifi, P.O Box 230-80108, Kenya
[2]Medical School, Mount Kenya University, Thika, P.O Box 342-01000, Kenya
[3]Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Nairobi, P.O Box 62000 – 00200, Kenya
[4]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Old Road Campus, NDM Research Building, Oxford OX3 7FZ, UK

v2 **First published:** 23 Mar 2023, **8**:136
https://doi.org/10.12688/wellcomeopenres.18868.1
Latest published: 16 Aug 2023, **8**:136
https://doi.org/10.12688/wellcomeopenres.18868.2

Abstract

Background: Coastal Kenya has experienced repeated outbreaks of Chikungunya (CHIKV) and Dengue (DENV) viruses mediated by competent *Aedes aegypti* mosquito populations. These mosquitoes harbor insect specific viruses (ISVs), some of which can prevent arboviral transmission. However, there has been no systematic molecular entomological surveillance in coastal Kenya and the diversity of viruses in local *Aedes aegypti* populations remains largely unknown.

Methods: To obtain a snapshot of the *Aedes aegypti* viromes from coastal Kenya, we took advantage of a cross-sectional survey of mosquitoes to determine the prevalence of Zika virus. We collected adult mosquitoes using lured Biogent's sentinel traps at 16 different localities along the Kenyan coast between May to September 2017. Pools of 20 female *Aedes aegypti* mosquitoes were generated following grouping by morphological characteristics. Presence of arboviruses in the mosquito pools was determined using virus-specific and genera-specific primers with real-time PCR. Metagenomic next generation sequencing (mNGS) on Illumina Miseq and analysis was used to characterize the virome.

Results: A total of 16,520 female *Aedes aegypti* grouped into 826 pools were analysed. Flaviviruses were detected in 69/826 (8.4%) pools by

Open Peer Review

Approval Status ? ✓

	1	2
version 2 (revision) 16 Aug 2023	? view	
version 1 23 Mar 2023	? view	✓ view

1. **Jandouwe Villinger**, International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya

2. **Lícia Natal Fernandes**, Universidade de Sao Paulo, São Paulo, Brazil

Any reports and responses or comments on the article can be found at the end of the article.

real time PCR. Sequencing generated 8,459/971,754 (0.87%) clean reads that were taxonomically assigned to 16 and 28 viral families and species, respectively. The family *Phenuiviridae* represented by *Phasi Charoen-like phasivirus* (PCLV) species was the most abundant, detected in 64/73 (87%) mosquito pools. No pathogenic viruses were identified by mNGS. Phylogenetic analysis revealed local PCLV and Cell fusing agent virus (CFAV) were distinct from global sequences.
Conclusions: Our data provides information about virus diversity and composition of the *Aedes aegypti* mosquitoes from coastal Kenya and contributes to the body of knowledge of the *Aedes aegypti* virome. To the best of our knowledge, this is the first study to provide this information from this region.

Keywords
Aedes aegypti, virome, metagenomics, phylogenetics

KEMRI wellcome This article is included in the KEMRI | Wellcome Trust gateway.

Corresponding author: Donwilliams O. Omuoyo (domuoyo@kemri-wellcome.org)

Author roles: Omuoyo DO: Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Nyamwaya DK**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Kamau E**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Nyagwange JN**: Formal Analysis, Investigation, Writing – Review & Editing; **Karanja HK**: Investigation, Methodology, Writing – Review & Editing; **Gitonga JN**: Investigation, Methodology, Writing – Review & Editing; **de Laurent Z**: Investigation, Methodology, Writing – Review & Editing; **Ngala JC**: Methodology, Resources, Writing – Review & Editing; **Kinyua J**: Resources, Supervision, Writing – Review & Editing; **Rono MK**: Methodology, Resources, Writing – Review & Editing; **Mwangangi JM**: Methodology, Resources, Writing – Review & Editing; **Agoti CN**: Formal Analysis, Methodology, Writing – Review & Editing; **Warimwe GM**: Conceptualization, Funding Acquisition, Investigation, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome [203077, https://doi.org/10.35802/203077] and an award from the Medical Research Council, Wellcome Trust, and Newton Fund (grant no. MC_PC_15092) that is part of the EDCTP2 programme supported by the European Union.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Omuoyo DO, Nyamwaya DK, Kamau E *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 Research 2023, **8**:136 https://doi.org/10.12688/wellcomeopenres.18868.2

First published: 23 Mar 2023, **8**:136 https://doi.org/10.12688/wellcomeopenres.18868.1

Appendix II: Ethics Review Approval from KEMRI/SERU

In Search of Better Health

KENYA MEDICAL RESEARCH INSTITUTE

OFFICE OF THE DIRECTOR RESEARCH & DEVELOPMENT

Tell: +254 020 2722541, 2713349,
0722 205 901, 0733 400 003

P.O. Box 54840-00200, Nairobi
Email: ddrt@kemri.go.ke
Website: www.kemri.go.ke

KEMRI/RES/7/3/1 **October 31, 2022**

TO: **PROF. GEORGE WARIMWE,**
PRINCIPAL INVESTIGATOR.

THROUGH: **THE DEPUTY DIRECTOR, CGMR-C,**
KILIFI.

Dear Sir,

RE: **PROTOCOL NO. SERU 3296 (REQUEST FOR ANNUAL RENEWAL AND PROTOCOL DEVIATION): ESTIMATING THE TRANSMISSION AND CASE BURDEN OF ZIKA VIRUS IN KENYA.**

Thank you for continuing review report for the periods **July 29, 2021** to **June 16, 2022** and **June 17, 2022** to **October 11, 2022**

The Committee noted that a protocol deviation report has been submitted as the request for annual renewal was done after the required date of submission. Measures taken to address the deviation were deemed adequate.

This is to inform you that the expedited review team of the KEMRI Scientific and Ethics Review Unit (SERU) was of the informed opinion that the progress made during the reported period is satisfactory. The study has therefore been granted **approval**.

This approval is valid from **October 31, 2022** for a period of **one (1) year**. Please note that authorization to conduct this study will automatically expire on **October 30, 2023**. If you plan to continue data collection or analysis beyond this date, please submit an application for continuation approval by **September 18, 2023.**

You are required to submit any amendments to this protocol and any other information pertinent to human participation in this study to the SERU for review prior to initiation.

You may continue with the study.

Yours faithfully,

ENOCK KEBENEI,
THE ACTING HEAD,
KEMRI SCIENTIFIC AND ETHICS REVIEW UNIT

Appendix III: Funding Statement

‘This work was supported by Wellcome [203077, https://doi.org/10.35802/203077] and an award from the Medical Research Council, Wellcome Trust, and Newton Fund (grant no. MC_PC_15092) that is part of the EDCTP2 programme supported by the European Union. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.’