MOLECULAR CHARACTERIZATION OF RNA VIRUSES IN SANDFLIES FROM SELECT REGIONS IN KENYA

JANE WAMBUI THIIRU

MASTER OF SCIENCE

(Biotechnology)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2025

Molecular Characterization of RNA Viruses in Sandflies from Select Regions in Kenya

Jane Wambui Thiiru

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

2025

DECLARATION

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

Signature.....Date: …….

Jane Wambui Thiiru

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

Signature.....Date: ……...

Dr. Fredrick Eyase, PhD

WRAIR-A, Kenya

Signature.....Date: …….

Dr. Samoel Khamadi, PhD

KEMRI, Kenya

Signature.....Date: …….

Prof Justus M. Onguso, PhD

JKUAT, Kenya

DEDICATION

I dedicate this thesis to my loving husband, Mwangi Murikwa, and our precious children, Allen Kaburu and Zion Baraka. I also dedicate it to my dear parents, Rachel and David Kamotho, and my beloved siblings, Geoffrey Nganga and Racheal Wanjiru. Your unwavering support and love have been my anchor. Thank you.

ACKNOWLEDGMENT

First, I am profoundly thankful to God for granting me the strength, perseverance, and wisdom to undertake this academic endeavor.

I am deeply indebted to my supervisors: Dr. Fredrick Eyase, Dr. Samoel Khamadi, and Prof. Justus M. Onguso, for their unwavering guidance, insightful feedback, and supervision throughout my study. Special thanks to Dr. Fredrick Eyase for providing the opportunity to carry out this research; your dedication has inspired me to strive for excellence.

I appreciate the U.S. Armed Forces Health Surveillance Branch and the Global Emerging Infections Surveillance Section for funding this research. Thanks also to the Department of Entomology and Vector-Borne Infections at the Walter Reed Army Institute of Research-Africa especially David Aboum, Nicholas Odemba, Janet Ambale, and Francis Ngere. I am grateful to the KEMRI Viral Hemorrhagic Fever Labs team, particularly Solomon Langat, Hellen Koka, Samuel Owaka, and Francis Mulwa, as well as my fellow researchers Bryson Kimemia, Miriam Ngiavi, Titus Kitonyo, Mercy Kibathi, Tabitha Wanjiru, Suhailah Kulthom, Cynthia Chelangat, George Kisoi, Faith Cherop, and Jacob Ngugi for their support.

Finally, I wish to thank the JKUAT Institute for Biotechnology for their support and for granting me the opportunity to pursue my studies, and the KEMRI Centre for Virus Research for resources, mentorship, and for allowing me to conduct this research.

TABLE OF CONTENTS

DECLARATION ..... **ii**

DEDICATION ..... **iii**

ACKNOWLEDGMENT ..... **iv**

TABLE OF CONTENTS ..... **v**

LIST OF TABLES ..... **xii**

LIST OF FIGURES ..... **xiii**

LIST OF APPENDICES ..... **xv**

ACRONYMS AND ABBREVIATIONS ..... **xvi**

ABSTRACT ..... **xix**

CHAPTER ONE **1**

INTRODUCTION **1**

1.1 Background of the Study 1

1.2 Statement of the Problem 4

1.3 Justification 5

1.4 Objectives of the Study 6

1.4.1 General Objective 6

1.4.2 Specific Objectives 6

1.5 Hypothesis 7

1.5.1 Null Hypothesis 7

CHAPTER TWO 8

LITERATURE REVIEW 8

2.1 RNA Viruses in Insect Vectors 8

2.2 Phlebotomine Sandflies as Vectors of Arboviruses 8

2.2.1 Phlebotomine Sandflies Classification 9

2.2.2 Morphological Characteristics of Phlebotomine Sandflies 10

2.2.3 Ecology of Phlebotomine Sandflies 11

2.2.4 Phlebotomine Sandflies Feeding Behaviours 11

2.2.5 Life Cycle of Phlebotomine Sandflies 11

2.2.6 Phlebotomine Sandflies Diversity and Distribution 12

2.3 Identification of Phlebotomine Sandflies 13

2.3.1 Morphological Identification 13

2.3.2 DNA Barcoding 13

2.3.3 Metabarcoding 14

2.3.4 Phelebotomine Sandflies in Kenya 14

2.4 Viruses Detected in Phlebotomine Sandflies 15

2.4.1 *Phenuiviridae* Family 15

2.4.2 *Rhabdoviridae* Family 17

2.4.3 *Reoviridae* Family 17

2.4.4 *Flaviviridae* Family 18

2.4.5 Negeviruses 19

2.5 Viruses Detected in Phlebotomine Sandflies in Kenya 19

2.6 Phlebotomine Sandfly-Borne Febrile Illnesses 20

2.6.1 Febrile Illnesses in Kenya 20

2.7 Control Measures against Phlebotomine Sand-fly Borne Viruses 21

2.8 Strategies for Detection and Characterization of Phlebotomine Sandfly-Borne Viruses. 22

2.8.1 Virus Isolation 22

2.8.2 PCR Detection 23

2.8.3 Metagenomics 24

2.8.4 Whole Genome Characterization 25

CHAPTER THREE 26

METHODOLOGY 26

3.1 Ethics Statement 26

3.2 Study Areas 26

3.3 Study Design 28

3.4 Study Population 28

3.5 Sample Size Determination 28

3.6 Sandfly Sorting, Identification, and Pooling 30

3.7 Homogenization of Phlebotomine Sandflies 31

3.8 To Determine the Diversity of Insect Specific Viruses and Arboviruses 31

3.8.1 Super Pools Generation 31

3.8.2 RNA Extraction and Quantification 33

3.8.3 Illumina Sequencing Libraries Preparation and Sequencing 34

3.8.4 Metagenomic Data Taxonomic Classification using DRAGEN and Kraken 2 35

3.8.5 Metagenomics Manual Data Analysis 35

3.8.6 Resequencing of Arbovirus Positive Sample on MiSeq 36

3.8.7 Data Analysis of MiSeq Resequencing Data 37

3.8.8 Characterization of Metagenome-Assembled Viral Genome Sequences 37

3.8.9 Phylogenetic Analysis 37

3.9 To Determine the Genomic Characteristics of the Isolated Arboviruses 38

3.9.1 Virus Isolation in Cell Culture 38

3.9.2 Illumina Library Preparation and Sequencing 39

3.9.3 Sequencing Data Analysis 39

3.9.4 Genome Organization and ORF Prediction 40

3.9.5 Amino acid Variations and Computational Protein Modeling 40

3.9.6 Phylogenetic Analysis 40

3.9.7 Molecular Clock Analysis 41

3.9.8 Evolutionary Selection Pressure Analysis 41

3.9.9 Confirmation of the Presence of Isolated Viruses in Original Pools. 42

3.10 To Determine the Taxonomic Identity of Phlebotomine Sandflies Species in Arbovirus Positive Pools. 44

3.10.1 Phlebotomine Sandflies DNA Extraction and Amplification of Mitochondrial Cytochrome Oxidase I (CO1) Gene 44

3.10.2 MinION Library Preparation and Sequencing 45

3.10.3 CO1 Gene Sequences Analysis and Taxonomic Assignment 46

3.11 Statistical Analysis 47

CHAPTER FOUR 48

RESULTS 48

4.1 Phlebotomine Sandflies Distribution 48

4.2 Diversity of Insect Specific Viruses and Arboviruses using Metagenomics 49

4.2.1 Metagenomics Reads Analysis 49

4.2.2 Metagenomics Contigs Analysis 52

4.2.3 Unclassified Virus Within the Order Picornavirales 54

4.2.4 *Dicistroviridae* Family 56

4.2.5 *Nodaviridae* Family....58

4.2.6 *Negevirus*....60

4.2.7 *Iflaviridae* Family....61

4.2.8 *Chuviridae* Family....63

4.2.9 *Flaviviridae* Family....64

4.2.10 *Phenuiviridae* Family....65

4.2.11 *Rhabdoviridae* Family....65

4.2.12 Metagenomics Diversity Statistical Analysis....67

4.3 Genetic Characterization of the Isolated Arboviruses....67

4.3.1 Virus Isolation....67

4.3.2 Genetic Organization of the Isolated Viruses....68

4.3.3 Amino Acid Variation among WN-KOUTV Strains....71

4.3.4 Positive Selection Pressure at Codon Site 1772 of the WN-KOUTV Lineage....74

4.3.5 Phylogeny of the WN-KOUTV Isolate....74

4.3.6 The Most Common Recent Ancestor of the Isolated West Nile Koutango Virus Potentially Existed in Kenya in 1930....75

4.3.7 The Phylogeny of Isolated Bogoria Virus....77

4.3.8 Confirmation of Identified Arboviruses using RT-PCR....79

4.4 Taxonomic Classification of Phlebotomine Sandflies Species Associated with

the Isolated Arboviruses 79

4.4.1 Distribution of Phlebotomine Sandflies Species in the WN-KOUTV Positive Pool 79

4.4.2 Distribution of Phlebotomine Sandflies Species in the BOGV Positive Pool 80

4.5 Data Availability 81

CHAPTER FIVE 82

DISCUSSION, CONCLUSION AND RECOMMENDATIONS 82

5.1 Discussion 82

5.2 Limitations of the Study 89

5.3 Conclusion 90

5.4 Recommendations 91

REFERENCES 92

APPENDICES 114

LIST OF TABLES

Table 3.1: Distribution of Phlebotomine Sandflies Samples across the Sampling Areas. 29

Table 3.2: Geographic Distribution and Pooling of Female Phlebotomine Sandflies for Metagenomics. 33

Table 3.3: Primer and Probe Sequences for Bogoria Virus L Segment. 43

Table 3.4: Primer and Probe Sequences for WN-KOUTV. 43

Table 3.5: Multiplex RT-PCR Assay Set up for WN-KOUTV and BOGV. 44

Table 3.6: Sequences of Primers for PCR Amplification of Phlebotomine Sandfly's CO1 Gene. 45

Table 4.1: Distribution of Phlebotomine Sandflies from Different Geographical Regions in Kenya. 49

Table 4.2: Number of Reads Before and After Quality Filtering per Sample..... 50

Table 4.3: Putative Viral Sequences and their Closest Hits 53

Table 4.4: Percentage Identities of Nucleotide and Amino Acid Sequences Between BAR_S2_S008 WN-KOUTV and other WN-KOUTV Strains 70

Table 4.5: Percentage Identities of Nucleotide and Amino Acid Sequences Between BAR_S5_230 Bogoria virus and other Bogoria Viruses..... 71

Table 4.6: Selection Pressure Analysis Table Indicating Positively Selected Site 1772 (NS3:267) of the WN-KOUTV Lineage. 74

Table 4.7: RT-PCR Amplification Results Confirming Presence of WN-KOUTV and BOGV in Original Pools..... 79

LIST OF FIGURES

Figure 2.1: Taxonomic Classification of Phlebotomine Sandflies. 9

Figure 2.2: Phlebotomine Sandfly. 10

Figure 2.3: Life Cycle of Phlebotomine Sandflies 12

Figure 3.1: Map of Kenya Indicating the Sampling Sites. 27

Figure 4.1: Taxonomic Classification of Metagenomic Reads..... 50

Figure 4.2: Proportion of Virus Family Reads Detected at each Sample Collection Site. 51

Figure 4.3: Schematic Representation of the Genome Organization of the Identified RNA Viruses..... 55

Figure 4.4: Phylogenetic Relationship of Identified Genomes Related to Dicistroviruses and Reference Viruses. 57

Figure 4.5: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Nodaviridae* RdRp Amino Acid Sequences. 59

Figure 4.6: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative Negeviruses. 60

Figure 4.7: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Iflaviridae* Viruses. 62

Figure 4.8: Mid-Point Phylogenetic Relationship of Phlebotomine Chuvirus with other Viruses. 64

Figure 4.9: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Rhabdoviridae* Viruses. 66

Figure 4.10: Cytopathic Effect Observed in a Cell Culture Monolayer. 68

Figure 4.11: Genome Organization of the WN-KOUTV Strain BAR_S3_S008... 69

Figure 4.12: Genome Organization of the BAR_S5_230 Bogoria Virus. 70

Figure 4.13: Amino Acids Variation among the WN-KOUTV Strains 72

Figure 4.14: Amino Acid Variation at Position 653 within NS5 Protein 73

Figure 4.15: Maximum-Likelihood Phylogenetic Trees Showing the Relationship of the WN-KOUTV Strains BAR_S3_S008, with Select WNV Strains. 75

Figure 4.16: Time-calibrated Phylogeny of Different WNV Lineages from Different Countries 76

Figure 4.17: Phylogenetic Relationship of BAR_S5_230 with Select Phleboviruses. 78

Figure 4.18: Distribution of Phlebotomine Sandflies Species in Pool BAR_S3_S008..... 80

Figure 4.19: Distribution of Phlebotomine Sandflies Species in Pool BAR_S5_230. 81

LIST OF APPENDICES

Appendix I: SERU Ethical Approval for the Study 114

Appendix II: SERU Ethical Approval for the Study 115

Appendix III: Plot Showing the Depth of Coverage of WN-KOUTV Strain BAR_S3_S008 115

Appendix IV: Plots Showing the Depth of Coverage of Bogoria Virus BAR_S5_230 117

Appendix V: Gel Photo of the Vector COI Amplicons 118

Appendix VI: Publication 1 119

Appendix VII: Publication 2 120

ACRONYMS AND ABBREVIATIONS

BEAST	Bayesian Evolutionary Analysis Sampling Trees
BLAST	Basic Local Alignment Search Tool
BOGV	Bogoria Virus
Bp	Base pairs
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
CDC	Centers for Disease Control and Prevention
cDNA	Complementary Deoxyribonucleic Acid
CGLV	Changuiola Virus
CO1	Cytochrome c oxidase subunit I
Ct	Cycle Threshold
CPE	Cytopathic Effect
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide
EPEV	Ecuador Paraiso Escondido Virus
EMRV	Embossos Virus
ESS	Effective Sample Size
FEL	Fixed Effects Likelihood
FUBAR	Fast, Unconstrained Bayesian Approximation

HPD	Highest Posterior Density
HTS	High-Throughput Sequencing
ICTV	International Committee on Taxonomy of Viruses
IQ-TREE	Iterative Quick Tree
JKUAT	Jomo Kenyatta University of Agriculture and Technology
KBGV	Kiborgoch Virus
KEMRI	Kenya Medical Research Institute
KUV	Kunji Virus
MAFFT	Multiple Alignment using Fast Fourier Transform
MCC	Maximum Clade Credibility
MEGA	Molecular Evolutionary Genetics Analysis
MEM	Minimum Essential Medium
MEME	Mixed Effects Model of Evolution
MIDORI	Molecular Identification Database of Ribosomal RNA
MCMC	Markov Chain Monte Carlo
NACOSTI	National Commission for Science, Technology and Innovation
NCBI	National Center for Biotechnology Information
NGL	Next Generation Graphics Library
NGS	Next-Generation Sequencing

PCR	Polymerase Chain Reaction
PERV	Perkerra Virus
PYMOL	Python Molecular Graphics Tool
RDP	Ribosomal Database Project
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RT-RNA	Reverse Transcription RNA
SABV	Saboya Virus
SINV	Sindis Virus
SLAC	Single-Likelihood Ancestor Counting
SRA	Sequence Read Archive
tMRCA	Time to Most Recent Common Ancestor
USAMRD-A	United States Army Medical Research Directorate-Africa
VHF	Viral Hemorrhagic Fever
WN-KOUTV	West Nile Virus Koutango Lineage
WNV	West Nile Virus
WRAIR-A	Walter Reed Army Institute of Research-Africa

ABSTRACT

Phlebotomine sandflies are vectors of arboviruses from genera Phlebovirus, Vesiculovirus, and Orbivirus that cause febrile and neurological illnesses in humans and animals. However, the diversity and public health significance of Phelebotomine sandfly-associated viruses in Sub-Saharan Africa remain poorly understood, representing a critical surveillance gap. This retrospective study used metagenomic, virus isolation, next-generation sequencing, and metabarcoding to address this gap by characterizing RNA viruses and their sandfly vectors in Kenya, where unexplained febrile illnesses persist. Female sandflies, previously collected between September 2020 and July 2022 from Baringo, West Pokot, Nakuru, Kisumu, Kilifi, and Kwale counties, had been stored at –80°C to preserve RNA integrity prior to pooling and homogenization. Virus isolation was performed in Vero cells, with cytopathic effect (CPE)-positive samples subjected to next-generation sequencing. Supernatant super-pools were screened for RNA viruses using metagenomics, while genomic DNA from homogenate pellets of arbovirus-positive pools enabled sandfly species identification via metabarcoding. Metagenomic analysis revealed a diverse virome, including, West Nile virus Koutango lineage (WN-KOUTV), Bogoria virus (BOGV), Phlebotomine rhabdovirus and several insect-specific viruses from five families: Dicistroviridae, Chuviridae, Iflaviridae, Nodaviridae, and Negevirus. Genomic characterization of isolated WN-KOUTV and BOGV provided new insights into their phylogeny and potential public health relevance. Additionally, metabarcoding identified Sergentomyia species, including Sergentomyia schwetzi, Sergentomyia inermis, Sergentomyia sp. SERG.UNK. F. 1- 8, Sergentomyia squamipleuris, and other Phlebotominae species within each pool. The results highlight the ongoing circulation of medically important and insect-specific viruses in Kenyan sandflies, emphasizing the need for continued surveillance and vector control strategies to address undiagnosed febrile illnesses. These findings also demonstrate the utility of combining metagenomics and metabarcoding with traditional virological methods for comprehensive arbovirus surveillance.

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

RNA viruses constitute most viruses responsible for human and animal diseases worldwide (Carrasco-Hernandez et al., 2017). Characterized by ribonucleic acid as their genetic material, RNA viruses are highly diverse and include many significant pathogens (Payne, 2017). Among RNA viruses found in blood-feeding insects, two key categories are recognized: insect-specific viruses (ISVs), which replicate only in insects and do not infect vertebrate hosts (Bolling et al., 2015; Vasilakis & Tesh, 2015), and arthropod-borne viruses (arboviruses), which encompass a varied group transmitted by blood-feeding arthropods such as ticks, mosquitoes, Phlebotomine sandflies, and biting midges (Carrasco-Hernandez et al., 2017; Vasilakis & Gubler, 2016).

Phlebotomine sandflies are small hematophagous insects with a body length seldom exceeding 1.5–3 mm and a hairy appearance (Alkan et al., 2013). The current widely used classification groups *Phlebotominae* into five genera*:* two genera from the Old World *(Phlebotomus, Sergentomyia)*, and three from the New World *(Lutzomyia, Brumptomyia, and Warileya)* (Galati & Rodrigues, 2023). 1,060 Phlebotomine sandfly species have been described worldwide (Galati & Rodrigues, 2023), with 100 species implicated in transmitting various pathogens that infect humans and domestic animals (Benallal et al., 2022). In Kenya, forty-eight sandfly species have been identified, belonging to the genera *Phlebotomus* and *Sergentomyia* (Anjili et al., 2011).

Phlebotomine sandflies are important vectors of parasites, Bartonella bacteria, and viruses that threaten human and domestic animal health (Ayhan & Charrel, 2017; Jancarova et al., 2023). They are notoriously known for transmitting Leishmaniasis disease caused by parasitic protozoans of the genus *Leishmania* (*Kinetoplastida: Trypanosomatidae*), which infect approximately two million people annually (Dvorak et al., 2020). In addition to parasites, sandflies transmit a diverse group of

RNA arboviruses that include clinically important genera such as *Phlebovirus*, *Vesiculovirus*, and *Orbivirus*. Arboviruses are acquired by Phlebotomine sandflies from an infected host during blood meals and then replicate within the vector (Vasilakis & Gubler, 2016). These viruses cause a range of diseases, from acute febrile illnesses to severe neurological syndromes, impacting both humans and animals worldwide (Vasilakis & Gubler, 2016). Notable sandfly-borne arboviruses include Sandfly fever Sicilian, Naples, and Toscana viruses of the genus *Phlebovirus (*family *Phenuiviridae),* Chandipura and Vesicular Stomatitis viruses of the genus *Vesiculovirus (*family *Rhabdoviridae),* and Changuinola virus of genus *Orbivirus (*family *Reoviridae)* (Ayhan & Charrel, 2018; Maroli et al., 2013).

Phleboviruses are the most significant group as they are associated with severely incapacitating febrile illnesses and are widely distributed. These febrile illnesses are caused by Naples and Sicilian viruses serocomplexes and are clinically indistinguishable. They are known as "sandfly fever" or "papatacci fever". Sandfly fever is characterized by abrupt illness with fever, malaise, headache photophobia, myalgia, and retro-orbital pain usually lasting 2-3 days (Ayhan & Charrel, 2017). Vesicular Stomatitis viruses and Chandipura virus, both of which infect vertebrates, including humans, primarily represent the *Vesiculovirus* genus. Vesicular Stomatitis viruses cause vesicular stomatitis in humans and domestic animals, while Chandipura virus leads to severe encephalitis in humans (Tandale et al., 2008). The genus *orbivirus* is represented by Changuiola virus (CGLV) associated with Phlebotomine sandflies of the New World (Jaafar et al., 2014). To date, only one reported case of human illness associated with CGLV infection (Phan et al., 2020a).

The public health burden of sandfly arbovirus diseases is significant both globally and in Africa. Sandfly fever viruses are endemic in southern Europe, the Mediterranean, the Middle East, North Africa, and parts of Asia (Alkan et al., 2013), with outbreaks and serological evidence reported across Africa, including Sudan, Djibouti, Ethiopia, Egypt, Morocco, Algeria, Tunisia, Uganda, and Somalia (Izri et al., 2008; Tesh et al., 1976; Watts et al., 1994; Woyessa et al., 2014). Chandipura virus has been responsible for major outbreaks of acute encephalitis in children, with

high fatality rates in India (Tandale et al., 2008). In addition to India, the virus has also been detected in Senegal, Nigeria, and Kenya (Koskei et al., 2024; Sudeep et al., 2016).

In addition to these medically important arboviruses, phlebotomine sandflies have been associated with several ISVs (Alkan et al., 2015; Moureau et al., 2010). Unlike arboviruses that infect vertebrates, insect-specific viruses replicate only in insects and do not cause disease in humans or animals (Bolling et al., 2015). However, recent research suggests that these viruses can modulate the replication of medically important arboviruses in co-infected vector cells and may influence vector competence (Patterson et al., 2020). This interaction offers potential avenues for novel vector control strategies and highlights the need to characterize the full spectrum of viruses present in sandfly populations.

Despite this, sandfly viruses remain under-researched in Sub-Saharan Africa compared to other vector-borne diseases (Alkan et al., 2013; Marklewitz et al., 2020; Tchouassi et al., 2019). Surveillance data on the incidence, genetic diversity, distribution, and evolutionary dynamics of these viruses are limited, hindering effective disease prevention and control. Moreover, the potential role of phlebotomine sandflies in the transmission of febrile illnesses of unknown origin has been largely overlooked, with emerging evidence suggesting that novel sandfly-borne phleboviruses may contribute to the burden of undiagnosed febrile diseases in Kenya (Tchouassi et al., 2019). The discovery and characterization of sandfly-borne viruses are further complicated by the high diversity of both vectors and viruses (Cholleti et al., 2018), as well as the limitations of traditional surveillance methods, which rely on manual identification, virus isolation in cell culture, and nucleic acid amplification (Cholleti et al., 2018; Edwards & Rohwer, 2005; Huemer et al., 2017). These methods are labor-intensive, biased toward known pathogens, and may miss novel or cryptic species (Huemer et al., 2017).

Recent advances in molecular approaches, such as next-generation sequencing (NGS) technologies, including metagenomics and metabarcoding, offer powerful alternatives. Metagenomics enables the unbiased detection and genomic

characterization of both known and novel viruses directly from vector samples, while metabarcoding allows for rapid identification of multiple sandfly species in bulk collections (Nooij et al., 2018; Roux et al., 2021). Genomic characterization of arboviruses through sequencing and phylogenetic analysis not only enables the identification of novel and emerging strains but also provides insight into their evolution, epidemiology, and the mechanisms underlying their pathogenicity and transmission dynamics (Patiño et al., 2024).

This study applied next generation sequencing for the detection and genomic characterization of RNA viruses isolated from sandflies, alongside metabarcoding of the cytochrome oxidase I (COI) gene to accurately identify sandfly species associated with these viruses; these integrated molecular approaches provide new insights into the diversity and transmission dynamics of sandfly-borne viruses in Kenya and offer valuable tools for improving arbovirus surveillance and vector control strategies.

1.2 Statement of the Problem

Febrile illnesses represent the leading cause of clinical attention in tropical and subtropical regions (Asuquo *et al.*, 2024), with Kenya reporting over 10 million annual episodes of acute febrile illness among children under five years (O'Meara et al., 2015). Despite this high burden, 37.1 % of febrile cases remain undiagnosed after standard testing(O'Meara et al., 2015). Phlebotomine sandflies mainly transmit pathogens responsible for these illnesses but are notably neglected as vectors of arboviruses, with research programs primarily focusing on mosquitoes and ticks (Lwande et al., 2013; Philbert et al., 2022; Tchouassi et al., 2019). Phlebotomine sandflies, of which there are over 1,000 species globally, but fewer than 10% have been studied for pathogens (Benallal et al., 2022). Also, their role in arbovirus transmission has been overshadowed by the focus on Leishmaniasis (Ayhan & Charrel, 2018). Therefore, the contribution of sandfly arboviruses to febrile illnesses is unknown. In addition, the diversity of insect-specific viruses within Phlebotomine sandfly populations, some of which may have potential applications in arbovirus transmission dynamics and vector control, remains unknown. In Kenya, 7 novel viruses, namely Ntepes virus (NPV), Bogoria virus (BOGV), Embossos virus

(EMRV), Kiborgoch virus (KBGV), and Perkerra virus (PERV), West Nile virus Koutango lineage (WN-KOUTV), Chandipura virus (CHPV), and Sindbis virus (SINV) were recently described in sandflies in the Rift valley region (Koskei *et al.,* 2024; Marklewitz *et al.,* 2020; Tchouassi *et al.,* 2019). Notably, specific neutralizing antibodies against NPV were detected in 13.9% of human serum samples taken at the site of isolation as well as disparate sites (Tchouassi *et al.,* 2019). However, the pathogenic potential, ecological cycles, mechanisms of maintenance, transmission, or presence in vertebrate reservoirs of these viruses remain limited. This suggests that the Kenyan population is at risk of febrile illnesses associated with sandfly-borne phlebovirus infections. In addition, with the increasing international travel (1.4 billion travelers annually) (UNWTO, 2024), globalization, and climate change which, through rising temperatures and altered rainfall patterns, expands the geographic range and extends the transmission season of arbovirus vectors, thereby increasing the risk of arboviral infections (Mordecai et al., 2020; Sant'Anna et al., 2025). Globally, vaccines against arboviruses are either not available or have limited use, and treatment is usually palliative (Goh et al., 2020). Traditional methods for identifying sandflies and detecting viruses are labor-intensive, require specialized expertise, and often miss novel or cryptic species (Huemer et al., 2017). Virus isolation and PCR-based techniques tend to be biased toward known pathogens, limiting the discovery of new viruses (Cholleti et al., 2018; Edwards & Rohwer, 2005). This study employed metagenomics, whole genome sequencing, and metabarcoding to overcome these challenges by enabling unbiased detection and characterization of viruses and accurate identification of sandfly vectors.

1.3 Justification

The recent identification of novel Phlebotomine sandfly arboviruses and serological evidence in humans necessitates the expansion of arbovirus surveillance to extensive and varied geographical regions in Kenya. This enhances the quantification of the risk posed by these arboviruses, especially in areas with historical febrile illnesses of unknown origin, such as Baringo, West Pokot, Nakuru, Kisumu, Kwale, and Kilifi counties.

Traditional Phlebotomine sandflies surveillance and diagnostic methods are limited by their labor-intensive nature, requirement for prior knowledge, and inability to detect novel or unexpected viruses (Huemer et al., 2017). Metagenomic sequencing and metabarcoding provide unbiased, comprehensive detection of both known and novel viruses and enable accurate identification of vector species directly from field samples, as successfully demonstrated in studies on sandflies (De Carvalho et al., 2018; Kocher et al., 2017; J. Wang et al., 2022).

Although Phlebotomine sandfly arboviruses have been studied previously in some geographical regions in Kenya, metagenomics has not been employed to define the overall diversity of RNA viruses that they harbor. With one exception, arboviruses are RNA viruses (Vasilakis *et al.,* 2013). The integration of virus isolation in cell culture and metagenomics enhances accuracy and provides an overview of all viruses circulating in the sandflies, as well as narrows down to presumptive pathogenic arboviruses.

Characterization of arboviruses in this study provides a critical early indication of circulating arboviruses in sandflies, essential for epidemic preparedness. Additionally, this information sheds light on sandflies as potential vectors of pathogenic arboviruses in Kenya. These findings would be useful in allowing evidence-based refinement of guidelines on Phlebotomine sandfly-borne viruses' surveillance and control activities.

1.4 Objectives of the Study

1.4.1 General Objective

To detect and characterize RNA viruses circulating in sandflies from select regions in Kenya.

1.4.2 Specific Objectives

I. To determine the diversity of insect-specific viruses and arboviruses circulating in sandflies from select regions of Kenya using metagenomics.

II. To determine the genomic characteristics of the isolated arboviruses of medical importance.
III. To determine the taxonomic identity of sandflies species associated with the isolated arboviruses using metabarcoding.

1.5 Hypothesis

1.5.1 Null Hypothesis

I. Insect-specific viruses and arboviruses harbored in sandflies collected from select regions of Kenya cannot be determined using metagenomics technique.
II. Arboviruses harbored in sandflies collected from select regions of Kenya cannot be isolated and genomically characterized through cell culture and sequencing.
III. Sandfly species associated with isolated arboviruses cannot be determined using metabarcoding technique.

CHAPTER TWO

LITERATURE REVIEW

2.1 RNA Viruses in Insect Vectors

RNA viruses constitute a major and diverse group of pathogens responsible for significant human and animal diseases worldwide (Carrasco-Hernandez et al., 2017). Among these, arboviruses are transmitted by blood-feeding arthropods such as mosquitoes, ticks, biting midges, and Phlebotomine sandflies (Vasilakis & Gubler, 2016). Arboviruses typically have complex life cycles involving replication in both invertebrate vectors and vertebrate hosts, which contributes to their epidemiological success and potential to cause outbreaks and emerging infectious diseases (Girard et al., 2020; Vasilakis & Gubler, 2016).

In addition to arboviruses, insect-specific viruses (ISVs) have been increasingly recognized as a widespread and genetically diverse group of RNA viruses that infect insects but do not replicate in vertebrates (Bolling et al., 2015; Vasilakis & Tesh, 2015). ISVs have been identified in a wide range of hematophagous insects, including mosquitoes and Phlebotomine sandflies, and are thought to influence arbovirus transmission dynamics by modulating vector competence or interfering with arbovirus replication (R. A. Hall et al., 2025; Vasilakis & Tesh, 2015). Furthermore, some ISVs have been harnessed as platforms for vaccine development against arboviral diseases. For example, chimeric vaccines based on insect-specific alphaviruses and flaviviruses have demonstrated safety and efficacy in preclinical models, offering promising tools for controlling arbovirus outbreaks (R. A. Hall et al., 2025).

2.2 Phlebotomine Sandflies as Vectors of Arboviruses

Phlebotomine sandflies are tiny insects mainly found in rural and peri-urban environments, often close to humans and domestic animals. They exist worldwide in tropical and subtropical, arid and semi-arid areas, and temperate zones (Munstermann, 2019). Approximately 100 sandfly species transmit various

pathogens, infecting humans and domestic animals (Benallal et al., 2022). They are the vectors of parasitic (leishmaniosis), bacterial (bartonellosis), and viral diseases by phleboviruses, orbiviruses, and vesiculoviruses (Maroli et al., 2013).

2.2.1 Phlebotomine Sandflies Classification

Phlebotomine sandflies belong to the order *Diptera*, which encompasses all true flies. They fall under the suborder *Nematocera,* known for their long-segmented antennae (Alkan et al., 2013). Within *Nematocera,* Phlebotomine sandflies belong to the family *Psychodidae*, which includes moths, flies, and sandflies (Alexander & Maroli, 2003; Alkan et al., 2013). Genus's level, they are classified into two main geographic regions. In the Old World, they included the genera *Phlebotomus* and *Sergentomyia.* In the New World, the genera are *Lutzomyia*, *Warileya,* and *Brumptomyia* (Galati & Rodrigues, 2023).

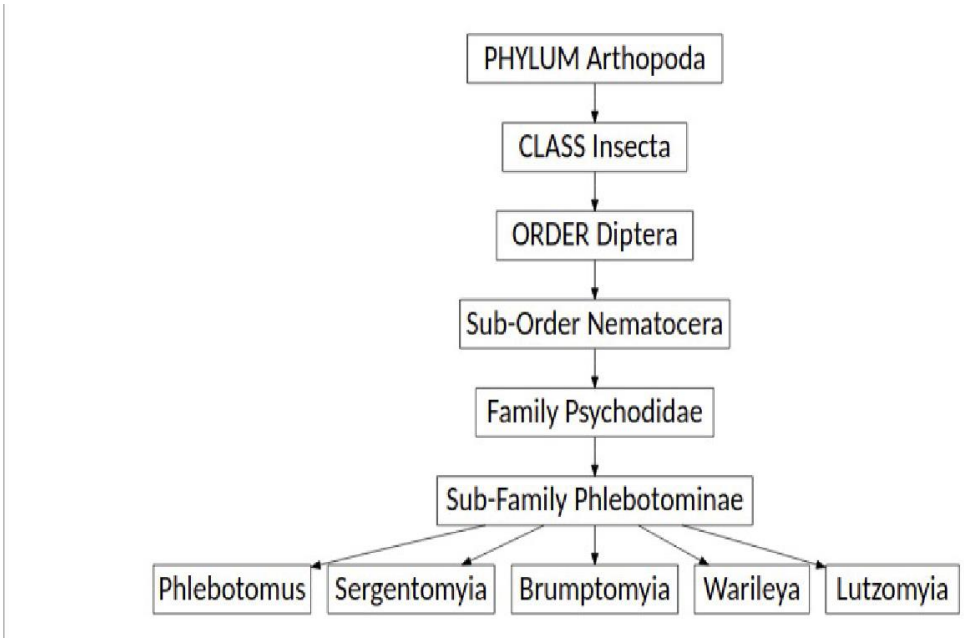


Figure 2.1: Taxonomic Classification of Phlebotomine Sandflies (Galati & Rodrigues, 2023)

2.2.2 Morphological Characteristics of Phlebotomine Sandflies

Morphologically, a delicate build, hairy bodies, and long legs (Alexander & Maroli, 2003) characterize Phlebotomine sandflies. They typically measure between 1.5 to 3.5 mm in length, and their color ranges from almost white to almost black (Alexander & Maroli, 2003; Alkan et al., 2013). The head of a Phlebotomine sandfly features prominent, large compound eyes and long, beaded antennae, which are essential sensory organs (Munstermann, 2019). Their mouthparts are adapted for piercing skin and sucking blood in females and sucking in males. Their wings are relatively large compared to their body size, held at 45-degree angles above the body when at rest, and are covered with dense, fine hairs that give them a characteristic appearance (Lawyer et al., 2017).

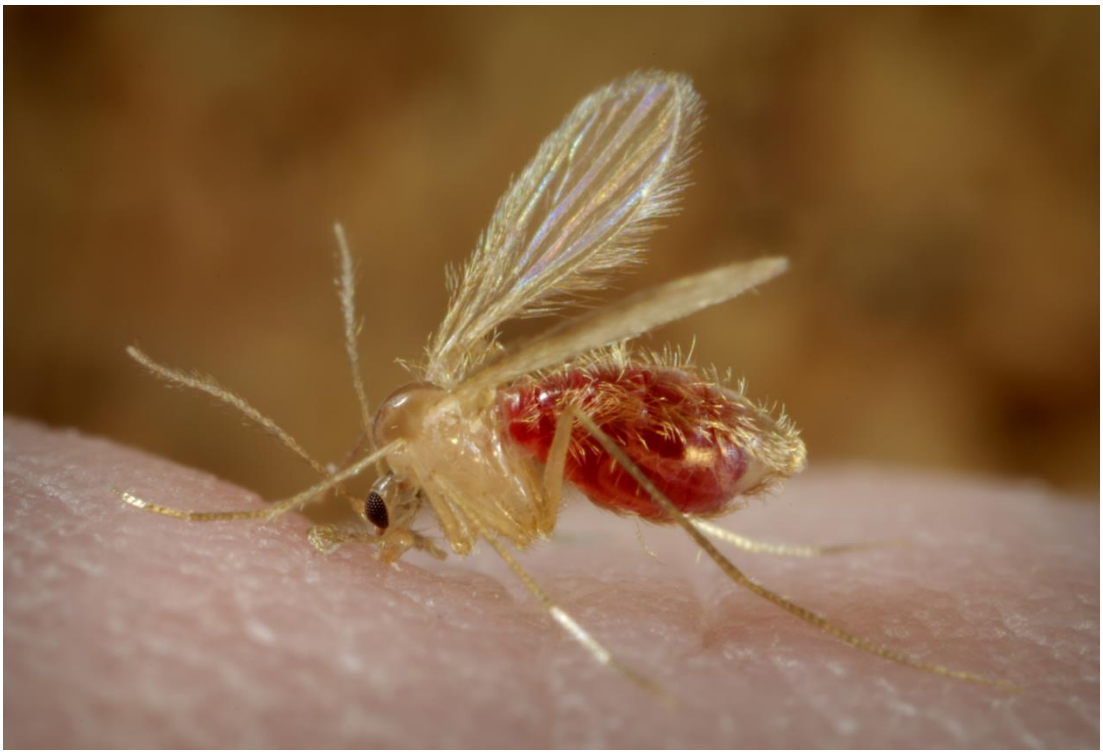


Figure 2.2: Phlebotomine Sandfly (CDC/Frank Collins, Public Health Image Library).

Legend

Note the delicate, hairy body, long legs, and characteristic wing position held at a 45-

degree angle above the body.

2.2.3 Ecology of Phlebotomine Sandflies

Ecologically, Phlebotomine sandflies are versatile and inhabit a wide range of environments from deserts to tropical forests. Adult sandflies shelter in humid and dark places such as cracks and crevices, caves, tree holes, rock walls, a stonewall, termite mounds, human habitations, animal sheds, and animal burrows during the day (Maroli et al., 2013; Philip, 1995). Their activity peaks during twilight and nighttime hours, making them primarily nocturnal, which influences their trapping and control strategies (Polseela et al., 2016).

2.2.4 Phlebotomine Sandflies Feeding Behaviours

Both females and males feed on sugar, but only females take a blood meal before laying eggs in organic-rich terrestrial microhabitats that nourish the larvae (Alkan et al., 2013). Depending on the sandfly species, they take blood from various animals, including mammals, birds, and cold-blooded vertebrates (Abbate et al., 2020).

2.2.5 Life Cycle of Phlebotomine Sandflies

The life cycle of Phlebotomine sandflies includes four development stages: egg, larva, pupa, and adult. The life cycle begins with the females taking a blood meal and laying their eggs in moist, dark environments rich in organic matter, where the eggs hatch within 4 to 20 days (Alkan et al., 2013). The resulting larvae go through four instar stages over 20 days to several months, depending on environmental conditions. The fourth-instar larvae metamorphose into pupal stage, then into adults (Volf & Volfova, 2011) within 7 to 14 days. After feeding, the females lay eggs, continuing the life cycle. The lifespan of adult sandflies typically lasts about two weeks.

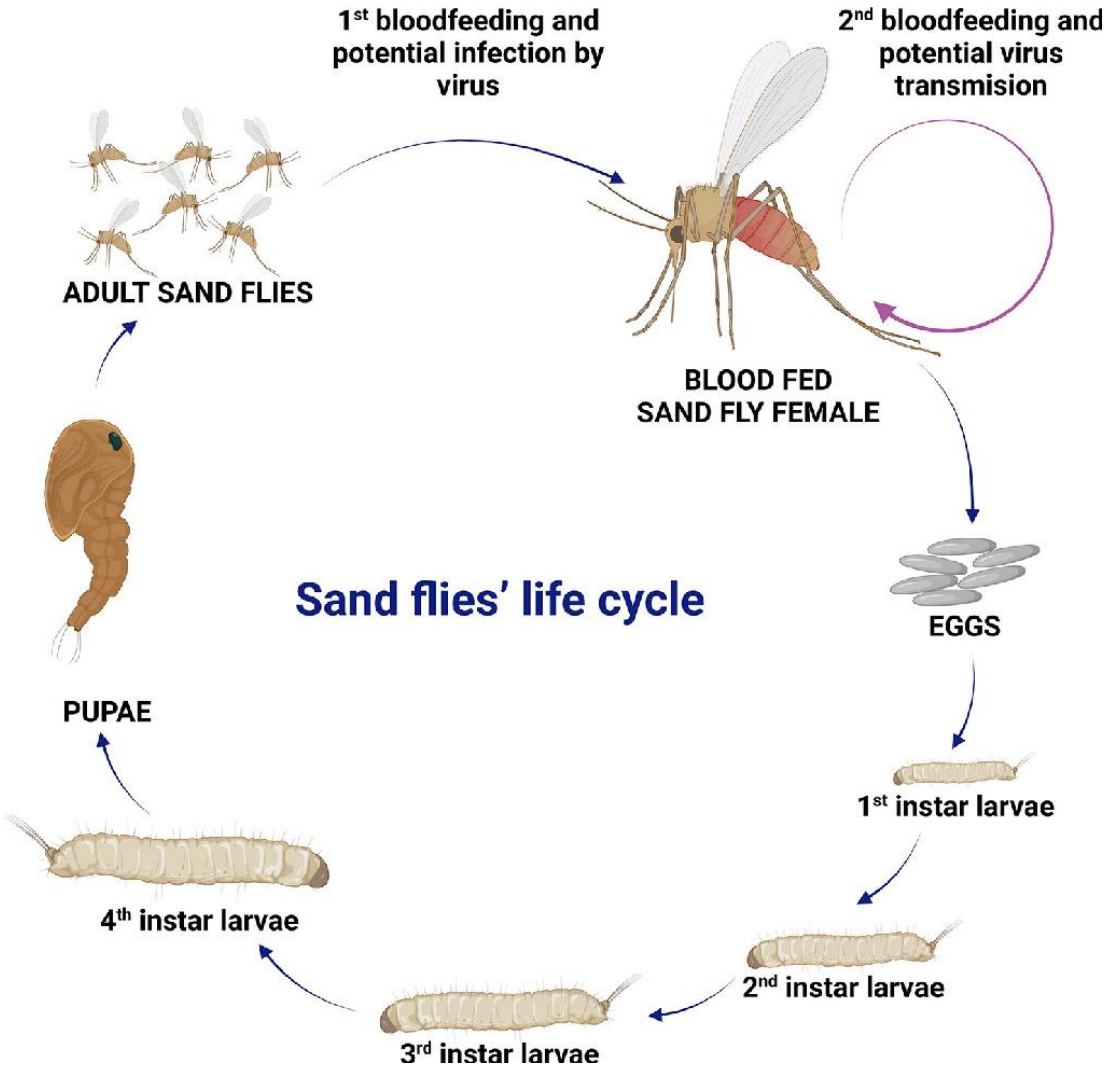


Figure 2.3: Life Cycle of Phlebotomine Sandflies (Jancarova et al., 2023)

2.2.6 Phlebotomine Sandflies Diversity and Distribution

The distribution of Phlebotomine sandflies varies widely, with species occupying diverse habitats from tropical forests to arid regions. Principally they are found in the warmer parts of the world, including southern Europe, Asia, Africa, Australia, and South America (Alten et al., 2015). To date, 1,060 Phlebotomine sandfly species have been described worldwide in tropical and subtropical regions, arid and semi-arid areas, and temperate zones (Galati & Rodrigues, 2023).

2.3 Identification of Phlebotomine Sandflies

Sandfly-borne pathogens are vectored by Phlebotomine sandflies belonging to diverse species with different ecological niches and geographic distributions. Therefore, entomological and genetic identification is crucial (Ayhan & Charrel, 2017).

2.3.1 Morphological Identification

Presently, sandfly species are identified based on diagnostic characteristics chiefly by internal structures like the ciborium, pharynx, spermatheca of females, and genitalia of males (Benallal et al., 2022). This method is the gold standard, but it is time-consuming, laborious, challenging, and requires a high level of expertise, especially for female specimens (Alten et al., 2015). In addition, extraordinary measures are required to prevent the degradation of pathogens. Furthermore, some sandfly species, isomorphic females, belonging to different species, damaged specimens, immature stages, and intra-, and inter-specific cryptic diversity may also not be reliably identified (Rodrigues et al., 2020). In addition, if traps are subsampled, there is a risk that rare species are missing. Therefore, in most cases taxonomy of Phlebotomine sandflies is never resolved, necessitating the use of protein such as MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) and nucleic-based methods such as DNA barcoding and Metabarcoding.

2.3.2 DNA Barcoding

DNA barcoding involves amplifying and sequencing a short standard DNA marker for the taxonomic assignment of a sample by looking for the closest match in a reference database (Hebert et al., 2003). The most used barcode is cytochrome oxidase subunit 1(CO1). In the last decade, it has emerged as the critical alternative for species identification in many taxa, including medically essential arthropods (Ondrejicka et al., 2014). This method has been successively used in several studies to identify Phlebotomine sandflies species, including immature stages and cryptic species (De Souza Pinto et al., 2015; Kumar et al., 2012; Nzelu et al., 2015; Polseela

et al., 2016). However, it is limited as it requires a single specimen per reaction and creates a hindrance in surveillance programs where large numbers of the target insects are involved.

2.3.3 Metabarcoding

Metabarcoding is an advanced molecular technique that allows for the simultaneous identification of multiple species within a mixed environmental sample by analyzing DNA sequences (Taberlet et al., 2012). This method combines DNA barcoding principles with high-throughput sequencing technologies, enabling researchers to assess biodiversity more efficiently and accurately, especially in ecosystems with complex species compositions (Ji et al., 2013; Taberlet et al., 2012). Illumina's sequencing technologies are the most used for metabarcoding. The requirement of short fragment length barcode sequences (less than 500 bp), however, limits the use of CO1 barcode (approximately 710 bp), whose library references are readily available (Krehenwinkel et al., 2019).

To address these limitations, third-generation sequencing platforms such as Pacific Biosciences and Oxford Nanopore Technologies that can sequence longer fragments offer a possible solution. Nanopore sequencing's underlying principle is based on passing single-stranded DNA through a nanopore and using the current fluctuations to reconstruct the DNA sequence (van Dijk et al., 2018). The MinION platform has the advantage of being readily accessible to researchers while at the same time yielding data quickly (Jain et al., 2016). It can also generate long DNA sequences with high phylogenetic and taxonomic resolution (Krehenwinkel et al., 2019).

2.3.4 Phelebotomine Sandflies in Kenya

In Kenya, the earliest reports of Phlebotomine sandflies date back to 1921 and later in 1930 and 1932 (Mukhwana et al., 2018). More than 48 species of sandflies have been identified in various habitats in Kenya, falling into the genera *Phlebotomus* (Rondani & Berte, 1840) and *Sergentomyia* (Franca & Parrot1920) (Anjili et al., 2011). Genus Phlebotomus is represented in five subgenera, namely Phlebotomus, *Larroussius*, *Synphlebotomus*, *Paraphlebotomus*, and *Anaphlebotomus*. Genus

Sergentomyia is the largest in four subgenera: *Sergentomyia*, *Sintonius*, *Grassomyia*, and *Parvidens* (Anjili et al., 2011).

Traditional morphological identification remains the mainstay of sandfly taxonomy in Kenya, as molecular approaches are not yet routinely applied. While DNA barcoding and other molecular methods have been increasingly used elsewhere to resolve cryptic species and support vector surveillance (Owino et al., 2021), their application to sandfly studies in Kenya remains limited. To date, no published studies have reported the use of high throughput metabarcoding for sandfly identification in Kenya. As a result, there is a significant gap in knowledge regarding the full spectrum of sandfly species involved in arbovirus transmission in the country.

2.4 Viruses Detected in Phlebotomine Sandflies

Phlebotomine sandflies are known to harbor a diverse array of viruses, some of which have significant implications for human and animal health. The viruses detected so far in Phlebotomine sandflies belong to five virus families, namely *Rhabdoviridae*, *Flaviviridae*, *Reoviridae*, *Peribunyaviridae*, *Phenuiviridae*, and recently established taxon *Negevirus* (Jancarova et al., 2023).

2.4.1 Phenuiviridae Family

The family *Phenuiviridae* (Bunyavirales), which harbors the genus *Phlebovirus* is the most significant and widely studied (Alkan et al., 2013; Ergunay et al., 2015; Marklewitz et al., 2020). Phleboviruses are the most medically important sandfly-borne viruses. They are distributed in the New World tropics and the Old World semi-arid and temperate regions, including the Mediterranean, North Africa, the Indian subcontinent, the Middle East, and Central Asia (Alkan et al., 2013).

Currently, the ICTV recognizes nine sandfly-borne phleboviral species serocomplex. These viruses are found in different geographic locations, in the new world (Bujaru virus, Candiru virus, Chilibre virus, Frijoles virus, and Punta Toro virus) and sandfly fever Naples virus, sandfly fever Sicilian virus, Salehabad virus, and Punique virus are reported around the Mediterranean basin, North Africa, central and western Asia

(Ayhan & Charrel, 2018). They are transmitted between vertebrates by Lutzomyia spp. in the New World and by Phlebotomus spp. in the Old World when they take a blood meal. Transovarial transmission (female to offspring) and venereal transmission play an essential role in the natural cycle (Tesh et al., 1976). In the recent past, five novel phleboviruses have been identified in Kenya. These include Ntepes virus, Bogoria virus, Embossos virus, Kiborgoch virus, and Perkerra virus (Koskei et al., 2024; Marklewitz et al., 2020; Tchouassi et al., 2019). Specific neutralizing antibodies against the Ntepes virus were detected in 13.9% of human serum samples taken at the site of isolation as well as a disparate site (Tchouassi et al., 2019).

Medically relevant sand-fly-borne phleboviruses belong to the sand-fly fever Naples serocomplex (comprising of Toscana virus, sandfly fever Naples virus, and related viruses), and sandfly fever Sicilian serocomplex (comprising of sandfly fever Sicilian virus and related viruses) (Alkan et al., 2013; Calisher & Calzolari, 2021).

Naples and Sicilian virus infections are clinically indistinguishable. They have an incubation period of 3-5 days followed by abrupt onset of severe fever, headache, malaise, photophobia, myalgia, and retro-orbital pain for 2-3 days (Tesh et al., 1976). However, even when mild, these infections have shown to be highly incapacitating for the time patients are affected (Depaquit et al., 2010). A substantial proportion of infections result in being asymptomatic. Toscana virus clinical infection starts with an incubation period ranging from a few days to 2 weeks. In some cases, the incubation period is followed by a mild febrile illness without the involvement of the central nervous system (Charrel et al., 2005). In other cases, the onset of the disease is intense with headache, fever, nausea, vomiting, myalgia, neck rigidity, kerning signs, unconsciousness, tremors, paresis, and nystagmus (Ergunay et al., 2015). The disease outcome is usually favorable, with a mean duration of 7 days. A small number of severe cases have been reported in the form of severe meningoencephalitis or meningitis without encephalitis (Charrel et al., 2005).

2.4.2 *Rhabdoviridae* Family

Members of four genera, namely *Vesiculovirus*, *Curiovirus*, *Sripuvirus*, and *Arurhavirus*, have been detected in Phlebotomine sandflies (Jancarova et al., 2023). Of this, only those of the genus Vesiculovirus are of medical and veterinary importance. Eleven Vesiculovirus species have been isolated from Phlebotomine sandflies, among which the Vesicular stomatitis complex and Chandipura virus are of high medical and veterinary importance (Jancarova et al., 2023). Vesicular stomatitis virus (VSV) is a significant pathogen within the Vesicular stomatitis complex, which includes various serotypes such as Indiana, New Jersey, and Alagoas. This complex is primarily of veterinary importance, affecting livestock and causing economic losses, particularly in the Americas (Rodrigues et al., 2020; Rozo-Lopez et al., 2018). Chandipura virus is primarily known for causing Chandipura viral encephalitis, a severe neurological disease that predominantly affects children in rural areas of India (Chadha et al., 2005; Mourya et al., 2019). Outside of India, CHPV was detected in West Africa (Senegal, Nigeria), Sri Lanka, (Jancarova et al., 2023) and Kenya (Koskei et al., 2024).

2.4.3 *Reoviridae* Family

Viruses transmitted by various arthropod vectors, including insects and ticks, characterize the genus *Orbivirus*, within the family Reoviridae. Changuinola virus (CGLV) is the only Orbivirus confirmed to be transmitted by sandflies (Jaafar et al., 2014). The virus was first isolated in Panama from sandfly pools, with subsequent isolation confirming its presence in various sandfly species, including *Lutzomyia* spp., *Sergentomyia khawi,* and *Phelebotomus papatasi* (Peralta et al., 1966; Travassos da Rosa et al., 1984). Recent findings in Thailand have also reported the presence of CGLV in sandflies, (Phumee et al., 2024). The proposed transmission cycle of CGLV involves sandflies and sloths, with sloths serving as a significant blood source and exhibiting long-term viremia (Medlin et al., 2016). Despite its presence in various vertebrate hosts and sandflies, Changuinola virus CGLV is not considered a major human pathogen, with only a single symptomatic human case reported (Phan et al., 2020b; Rosa et al., 1984).

2.4.4 *Flaviviridae* Family

Although Family *Flaviviridae* arboviruses are predominantly associated with mosquitoes and ticks, their presence in sandflies has also been documented, albeit rarely. In 1992, four strains of Saboya virus (SABV) were isolated from sandfly pools collected in Senegal. The appearance of SABV in sandflies in this area was later confirmed (Traoré-lamizana et al., 2001). SABV has also been detected in other arthropods, including Aedes mosquitoes and *Ixodidae* ticks, suggesting that its vector specificity remains unclear (Butenko, 1996). Recently it was also reported again in Phlebotomine sandflies sampled in Niger (Fall et al., 2021).

In the New World, Ecuador Paraiso Escondido virus (EPEV) was isolated from Phlebotomine sandflies. EPEV was discovered in a pool of non-engorged and non-gravid female *Lutzomyia (Psathyromyia) abonnenci* collected in Ecuador (Alkan et al., 2015). EPEV replicated in insect cells, as evidenced by cytopathic effects and RT-PCR, but it did not infect or replicate in vertebrate cells or suckling mice; therefore an insect-specific virus (Alkan et al., 2015).

Recently, West Nile virus Koutango lineage (WN-KOUTV) was isolated from Phlebotomine sandflies in Niger (Fall et al., 2021) and Kenya (Koskei et al., 2024). WN-KOUTV, predominantly found in Africa, was for many years reported as a distinct flavivirus. However, based on phylogenetic evidence, it is presently regarded as a distant variant of the West Nile Virus (WNV) (Charrel et al., 2003; Scherret et al., 2001). WN-KOUTV takes its name from the Koutango district of the Kaolack region in Senegal, where the virus was first recovered from the wild rodent *Tatera kempi* in 1968 (Coz et al., 1975). Subsequently, it was detected in M*astomys sp.* 1974 in Senegal and later detected in *Lemnyscomys sp.* in Central African Republic (Fall et al., 2021). In addition, it has been isolated from a range of arthropods, including mosquitoes and ticks. In animal models, investigations done on mice have consistently shown that WN-KOUTV is the most neurovirulent WNV lineage (Fall et al., 2017; Pérez-Ramírez et al., 2017; Prow et al., 2014). Potential risk for humans has been highlighted following an accidental infection of a Senegalese lab worker characterized by fever and rash (Shope, 2003). There is also documented serological

evidence in humans in Gabon (Fall et al., 2021) and Sierra Leone (De Araujo Lobo, 2012). *Culex spp.* primary competent vectors of WNV lineage 1 and 2 were found to be incompetent vectors of WN-KOUTV (Fall et al., 2014). However, transovarial transmission has been documented in *Aedes Aegypti* (Coz et al., 1976). Also, dissemination of WN-KOUTV by *Aedes aegypti* was demonstrated, but only at high titers (Lobo et al., 2014).

2.4.5 Negeviruses

Negeviruses are a newly discovered taxon of insect-specific viruses (Vasilakis et al., 2013). They have been isolated from various insects, including Phlebotomine sandflies, in geographically distant regions of America, Europe, Africa, Australia, and Asia (Jancarova et al., 2023). To date, there have only been two detections of Negevirus from sandflies, Loreto virus obtained from a pool of *Lutzomyia spp*. in Peru in 1977, and Piura virus obtained from *Lutzomyia evansi* sampled in 2013 in Florida (Vasilakis et al., 2013). Though non-pathogenic, several co-infection studies with arboviruses suggest that negeviruses may potentially be used for pathogen control, either by natural inhibition of other viruses or, in their modified form, to control a specific arbovirus in a vector (Patterson et al., 2020).

2.5 Viruses Detected in Phlebotomine Sandflies in Kenya

For a long time, arboviruses have not been associated with Phlebotomine sandflies in Kenya, though human infections with sandfly fever Naples, and Sicilian viruses have been reported in neighboring countries, including Uganda, Somalia, Djibouti, and Ethiopia (Rodhain et al., 1989; Tesh et al., 1976). In recent studies, several viruses have been detected in Phlebotomine sandfly populations in Kenya, particularly in the North Rift region, highlighting the significance of Phlebotomine sandflies as potential vectors for various arboviruses (Koskei et al., 2024; Marklewitz et al., 2020; Tchouassi et al., 2019).

Ntepes virus, isolated from Phlebotomine sandflies collected in Baringo County in 2014 and 2016, has been shown to infect humans. Serological evidence indicates the

presence of neutralizing antibodies in human serum samples from various locations across Kenya, suggesting a broader distribution of the virus (Tchouassi et al., 2019). In addition to the Ntepes virus, seven strains of four previously unknown phleboviruses were detected, including Bogoria virus, Embossos virus, Kiborgoch virus, and Perkerra virus in Baringo (Marklewitz et al., 2020). Several other viruses were identified in a study conducted on Phlebotomine sandflies sampled between 2015 and 2018. The study reported the isolation of four arboviruses: Sindbis virus (four isolates from Kacheliba and Baringo), Chandipura virus (four isolates from Turkana and Baringo), Koutango virus (one isolate from Baringo), and Bogoria virus (one isolate from Kacheliba) (Koskei et al., 2024). Given the diversity of sandfly-borne viruses now detected in Kenya, understanding their contribution to the burden of febrile illnesses is increasingly important.

2.6 Phlebotomine Sandfly-Borne Febrile Illnesses

Sandfly fever, also known as phlebotomus fever, papatasi fever, or three-day fever, is a viral illness transmitted by the bite of infected female Phlebotomine sandflies (Ayhan & Charrel, 2017). The disease is caused by phleboviruses and is endemic in parts of Europe, Asia, Africa, and Latin America, particularly in regions with hot, dry climates (Alkan et al., 2013).

The clinical presentation typically includes a sudden onset of high fever, headache, photophobia, malaise, myalgia, and retro-orbital pain (Ayhan & Charrel, 2017). The incubation period ranges from 4 to 8 days, and the illness usually resolves spontaneously within three to six days (Dehghani et al., 2021; Jarvis, 2024). Additional symptoms can include fatigue, chills, abdominal pain, dizziness, and, occasionally, facial flushing and tachycardia. Although symptoms can be severe, sandfly fever is rarely fatal (Depaquit et al., 2010).

2.6.1 Febrile Illnesses in Kenya

Febrile illnesses represent the leading cause of clinical attention in tropical and subtropical regions (Maze et al., 2018). In Kenya, more than 10 million episodes of

acute febrile illness are treated each year among children under five, highlighting the significant burden of these conditions in the pediatric population (O'Meara et al., 2015).

While malaria is often presumed to be the primary cause of fever in sub-Saharan Africa, recent surveillance in Kenya has shown that a broad range of pathogens contribute to Acute Febrile Illnesses. Plasmodium species remain the most common pathogen, detected in 35.8% of cases as a single infection and in 2.9% as a co-infection. Other notable causes include HIV (3.8% of undifferentiated fever cases), chikungunya virus (2.3%, detected primarily in Mombasa), and various respiratory viruses, which are responsible for a substantial proportion of febrile illnesses, especially in children (Verani et al., 2024). Rickettsial diseases, scrub typhus, and Q fever are increasingly recognized as important contributors to undifferentiated fevers, particularly in regions where animal husbandry is prevalent.

In addition to well-known pathogens, recent studies have detected novel sandfly-associated phleboviruses in Kenya, some closely related to viruses known to cause febrile illness such as sandfly fever Naples and Sicilian virus serogroups (Marklewitz et al., 2020; Tchouassi et al., 2019). The Ntepes virus (NTPV), a newly identified phlebovirus, has been isolated from Phlebotomine sandflies in Baringo County and evidence of human infection has been demonstrated through the detection of neutralizing antibodies in serum samples (Tchouassi et al., 2019).

2.7 Control Measures against Phlebotomine Sand-fly Borne Viruses

Sandfly fever viruses are critical public health concerns, especially given the absence of approved vaccines or specific antiviral therapies for these diseases (Alkan et al., 2013). Research has shown that a vaccine using recombinant Toscana virus proteins can provide complete protection in animal models, but the genetic diversity among sandfly-borne viruses poses challenges for creating a universally effective vaccine (Alkan et al., 2013; Gori Savellini et al., 2008). Insecticides and repellents are the primary tools for reducing the spread of sandfly-borne diseases. Indoor residual spraying has been effective in inhabited areas, particularly against anthroponotic

sandflies, but its efficacy diminishes against non-anthroponotic species (Balaska et al., 2021; Killick-Kendrick, 1999). Continuous spraying is necessary to maintain effectiveness, as sporadic campaigns are largely ineffective (Alexander & Maroli, 2003). Various insecticides, including organochlorines, organophosphates, and synthetic pyrethroids, have been used, but resistance to some of these chemicals has been documented in different regions (Alexander & Maroli, 2003; Balaska et al., 2021). Moreover, attempts to control sandflies through insecticide-impregnated bed nets and other innovative methods have yielded mixed results, indicating the need for further research and trials (Balaska et al., 2021).

In addition to insecticides, novel sustainable approaches, such as pheromone dispenser baits and the cultivation of noxious plants, are being explored as potential control measures against sandflies (Bray et al., 2009). These methods aim to reduce sandfly populations in a more environmentally friendly manner. However, comprehensive strategies that integrate these various control measures, along with community education and engagement, are essential for effectively managing the risks posed by sandfly-borne phleboviruses (Alexander & Maroli, 2003; Alkan et al., 2013).

2.8 Strategies for Detection and Characterization of Phlebotomine Sandfly-Borne Viruses.

2.8.1 Virus Isolation

Viruses are isolated through inoculation into newborn mice intracerebrally or by seed cell lines that are competent for the replication of the viruses (Huemer et al., 2017). The Naples virus, Sicilian virus, and Toscana virus can replicate in Vero, LLC-MK2, and BHK-21 cells (Karabatsos; & Taylor, 1985). Vero cells have been the most frequently used because of sand-fly-borne phleboviruses. In Kenya, the first successful isolation of a sandfly-borne virus was the Ntepes virus, which was isolated from Phlebotomine sandflies collected in Baringo County (Tchouassi et al., 2019). This provided the initial evidence of active sandfly-borne phlebovirus circulation in the country and enabled subsequent serological and genomic characterization.

Subsequent studies in Kenya have attempted to isolate additional novel phleboviruses, such as Bogoria, Embossos, Kiborgoch, and Perkerra viruses. However, Marklewitz et al. (2020) reported that while these viruses were detected by PCR and sequencing, isolation in cell culture was not successful for these viruses. More recently, Koskei et al. (2024) achieved successful isolation of several arboviruses from Kenyan sandflies, including West Nile virus Koutango lineage (WN-KOUTV), Chandipura virus, Sindbis virus, and notably, Bogoria virus. These viruses were isolated using Vero cell cultures, where cytopathic effects were observed and confirmed by molecular assays.

This demonstrates both the challenges and progress in isolating sandfly-borne viruses: while some viruses have only been detected by molecular methods, others have been successfully isolated, providing critical material for further study. Importantly, successful isolation in vertebrate cell lines suggests these viruses are not insect-specific and may have implications for animal or human health. (Elrefaey et al., 2020).

2.8.2 PCR Detection

Unlike bacteria, viruses do not share any common gene that could similarly qualify as a unified phylogenetic marker (Mokili et al., 2012). Molecular techniques based on Polymerase Chain Reaction (PCR) are more sensitive for the detection of viruses as compared to virus isolation (Cassedy et al., 2021). In Kenya, molecular detection of sandfly-borne viruses has relied extensively on PCR and sequencing, given the limitations of cell culture in detecting all virus types. Targeted PCR assays, often using primers designed for conserved regions of phlebovirus, rhabdovirus, and flavivirus genomes, have enabled the identification of several novel viruses from sandfly pools, including Ntepes, Bogoria, Embossos, Kiborgoch, Perkerra, Chandipura, and Koutango viruses (Koskei et al., 2024; Marklewitz et al., 2020; Tchouassi et al., 2019). These studies have used both conventional and nested PCR protocols, followed by Sanger sequencing for confirmation and genetic characterization. However, the natural genetic variability of sandfly-borne viruses and the scarcity of comprehensive genomic data have posed challenges, sometimes

limiting detection to known or closely related viruses (Ayhan & Charrel, 2018; Sánchez-Seco et al., 2003). Despite these constraints, PCR-based approaches remain the cornerstone of arbovirus surveillance in Kenyan sandfly populations, pending the adoption of broader, unbiased metagenomic techniques.

2.8.3 Metagenomics

Metagenomics studies the collective viral genomes from primary samples (Cholleti et al., 2018). Unlike PCR amplicon sequencing, which is limited to a few predefined targets, metagenomics takes advantage of high-throughput sequencing and is broadly non-specific to identifying any viruses present in a given sample (Mokili et al., 2012). It has been applied extensively in sequencing uncultivable microbes, characterization of the complete viral populations in a given sample and detecting novel viruses without isolation and cloning. Overcoming the challenges posed by the high genetic diversity observed across most virus families and the lack of available full-length sequences for comparison (Hang et al., 2012). Novel sandfly-borne viruses, including the Viola virus, have been discovered using this technique (De Carvalho et al., 2018). To date, the detection of sandfly-borne viruses in Kenya-including Ntepes, Bogoria, Embossos, Kiborgoch, and Perkerra viruses-has relied primarily on virus isolation targeted PCR and sequencing approaches, rather than unbiased metagenomic methods (Koskei et al., 2024; Marklewitz et al., 2020; Tchouassi et al., 2019). These molecular studies have significantly expanded our understanding of the diversity of sandfly-associated viruses in the North Rift region and beyond. However, no comprehensive metagenomic profiling of sandfly viromes has yet been conducted in Kenya. As a result, the full spectrum of insect-specific viruses and arboviruses circulating in sandfly populations across diverse ecological zones in Kenya remains largely unexplored. This highlights a critical gap in current knowledge and underscores the importance of applying metagenomic techniques to reveal the true diversity of viruses present in Kenyan sandflies.

2.8.4 Whole Genome Characterization

Whole-genome sequencing is a comprehensive method for analyzing the DNA sequence of an organism's genome at a single time (Kwong et al., 2015). It provides a base-by-base view of the genome and hence can provide large and small variants to facilitate comparative genomic studies such as evolutionary relationships and evolutionary selection pressure analysis.

Genomic characterization of sandfly-borne viruses in Kenya is progressing but remains limited in scope. The complete genomes of several novel phleboviruses-including Ntepes, Bogoria, Embossos, Kiborgoch, and Perkerra viruses-have been sequenced following their discovery in sandflies from the North Rift region (Marklewitz et al., 2020; Tchouassi et al., 2019). Comparative genomic analyses have revealed unique genetic features among these viruses, contributing valuable insights into their evolution, diversity, and potential pathogenicity. In addition, partial or full genome sequences have been obtained for other medically relevant viruses such as Chandipura and Koutango viruses recently detected in Kenyan sandflies (Koskei et al., 2024). While these studies provide foundational data for understanding sandfly-borne viruses in the Kenyan context, comprehensive genomic analyses encompassing the full diversity of circulating arboviruses across different ecological regions are still lacking.

CHAPTER THREE

METHODOLOGY

3.1 Ethics Statement

This study was retrospective and involved Phlebotomine sandfly samples collected between September 2020 and July 2022 under the Kenya Medical Research Institute (KEMRI) Scientific and Ethics Review Unit study CCR SERU/SSC No: 3948. The processing and analysis of these previously collected samples in this study was approved by the under-protocol number KEMRI/SERU/CVR/007-2022/4570. As the study did not involve human subjects, it was exempted from further approval according to Walter Reed Army Institute of Research (WRAIR) policy #25. Additionally, the study received approval from the National Commission for Science, Technology & Innovation (NACOSTI) under License No: NACOSTI/P/23/24603.

3.2 Study Areas

The study areas included sites across six geographically distinct administrative counties in Kenya: Baringo, Nakuru, Kwale, Kilifi, West Pokot, and Kisumu as shown in Figure 3.1. These counties were selected based on their history of febrile illness outbreaks, previous cases, or reports of human fevers of unknown origin (Chretien et al., 2007; Hooft et al., 2021; Nguku et al., 2010; Sanders et al., 1998; Sergon et al., 2008).

Kisumu (0°05'S and 34° 46' E) and Nakuru (0°30'S and 36° 0' E) represent counties of a medium potential agro-ecological zone (Recha, 2018), ideal for agricultural cultivation and the most resettled by a human. The annual rainfall ranges between 950 and 1500 mm.

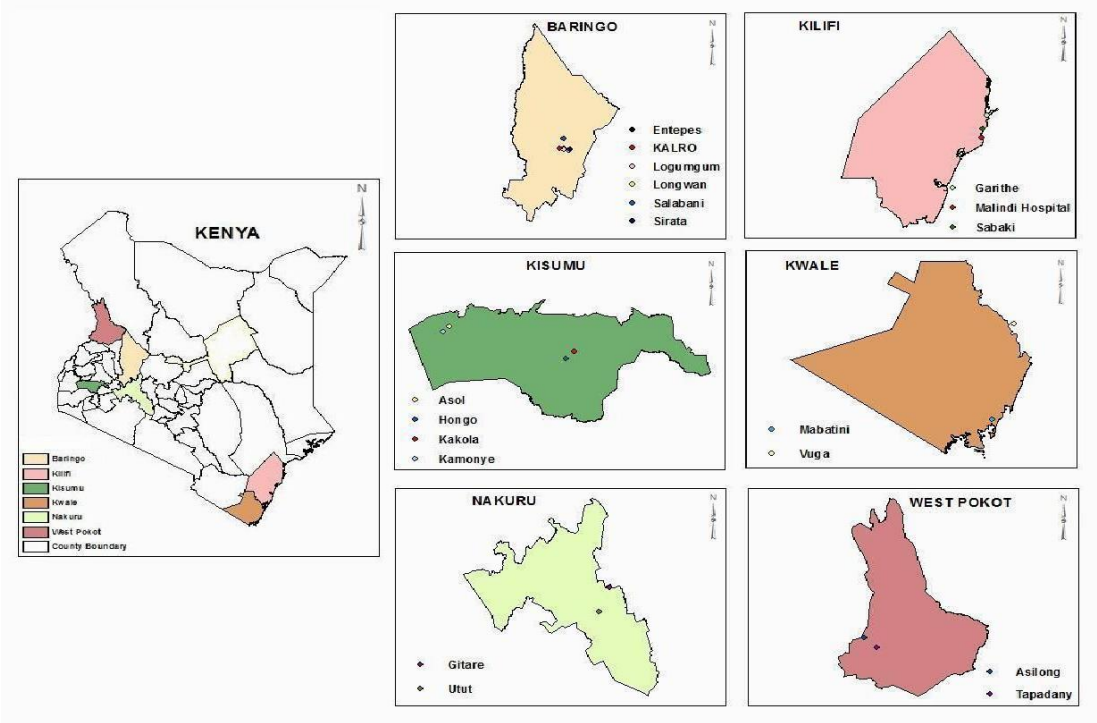


Figure 3.1: Map of Kenya Indicating the Sampling Sites

This map was developed using ArcGIS Software Version 10.2.2 (http://desktop.arcgis.com/en/arcmap) advanced license.

Baringo (0°40'N and 36° 00'E) and West Pokot (1°62'N and 35° 39'E) represent semi-arid to very arid Zones, which are predominantly inhabited by pastoralists and agro-pastoralists (Recha, 2018). Annual rainfall is 300-600mm, therefore, unsuitable for rain-fed cultivation due to physical limitations such as aridity and poor vegetation. Their main economic activity is livestock production.

Kilifi (3°67'S and 39° 75'E) and Kwale (4° 10'S and 39° 27'E) represent coastal lowlands agro-ecological zones (Recha, 2018). The main economic activities of Kilifi County are fishing and tourism due to its proximity to the Indian Ocean. Kwale is an inland county and engages in the agriculture of fruits, livestock, and mixed farming.

3.3 Study Design

This was a retrospective study aimed at determining the diversity of insect-specific and arboviruses circulating in sandflies from the selected counties, as well as conducting baseline characterization of the isolated arboviruses of medical importance and their associated sandflies species.

3.4 Study Population

The study populations were female sandflies samples collected from September 2020 to July 2022 under KEMRI study CCR SERU/SSC No: 3948.

3.5 Sample Size Determination

6,312 female Phlebotomine sandflies were collected between September 2020 and July 2022. The sandflies were grouped into pools for downstream analysis. Most of the pools (n = 689) contained ≤ 10 sandflies per pool, grouped according to genus, collection location, and trapping site. However, 11 pools from West Pokot were formed with larger sample sizes, each containing ≤ 50 sandflies. These larger pools were created in specific instances where higher numbers of sandflies were collected from the same site and genus, as reflected in the pooling records. Each pool was assigned a unique code (e.g., BAR/S3/001, KSM/S6/485, PKT/S2/700) reflecting its specific site and batch information, allowing for traceability and reproducibility as shown in Table 3.1.

Table 3.1: Distribution of Phlebotomine Sandflies Samples across the Sampling Areas

Study area	Collection Sites	Number of female sandflies	Number of pools	Pool numbers
Baringo	Salabani, Longumgum, Sirata, Entepes, Longwan, KALRO	3679	406	BAR/S3/001-BAR/S5/406
Nakuru (Gilgil)	Utut, Gitare	1787	156	GIL/S1/407-GIL/S1/456
Kilifi (Malindi)	Malindi Hospital, Vuga	350	60	MAL/S3/457-MAL/S2/475
Kwale	Mabatini	46	9	KWA/S6/476-KWA/S6/484
Kisumu	Kamonye, Asol	92	19	KSM/S6/485-KSM/S7/544
West Pokot	Asilong, Tapadany	358	50	PKT/S1/545-PKT/S2/700
Total		**6,312**	**700**	

Since no specific RNA virus was targeted in this study, the degree of variability of Phlebotomine sandflies population with RNA viruses is unknown, therefore, the Cochran formula (1977) was used to determine the minimum sample size.

$$n_0 = \frac{z^2 p(1-p)}{e^2}$$

Where: n_0= Sample size

z= Normal deviation at the desired confidence interval. At 95%, Z value at 95% is 1.96

P= Proportion of the population with the desired characteristic.

q = Proportion of the population without the desired characteristic.

e^2= Degree of precision; will be taken to be 5%.

$$n = \frac{1.96^2 \times 0.5(1-0.5)}{0.05^2}$$

$$n_0 = 384.16 = 385\ pools$$

Since the collected female sandflies population is known sample adjustment was made using the following formula,

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

Where:

n= the desired sample size for the finite population

n_0= the calculated sample size

N= the total population.

$$n = \frac{385}{1 + \frac{385 - 1}{700}}$$

$$n = 249\ pools$$

Therefore, the minimum sample size was 249 pools, however, total sampling was chosen because this approach maximized the results from the samples, which is important given the relatively small population and the potential rarity RNA of sandfly borne virus infections. Additionally, total sampling eliminates the risk of sampling bias. Since the available data was finite, analyzing all collected samples was more efficient and increased the statistical power of the study.

3.6 Sandfly Sorting, Identification, and Pooling

Sorting and morphological identification of collected sandflies was done by expert

taxonomists to the genus level using established taxonomic keys (Abonnenc, 1951; Kirk & Lewis, 1951). Identification was based on key morphometric features such as wing venation patterns, antennal segmentation, palpal segment proportions, and the presence or absence of cibarial and pharyngeal armature. To minimize specimen manipulation and preserve viral integrity for downstream virus isolation, slide-mounting and detailed species-level identification were not performed at this stage.. The female sandflies were pooled according to genus, location, trapping site, and day of capture in a 1.5 mL tube and stored at -80°c.

3.7 Homogenization of Phlebotomine Sandflies

The scope of this study started with the homogenization of the pooled female sandflies. Phlebotomine sandfly pools were first homogenized using 2.0 mm zirconium beads in 500 µL of homogenization media containing minimum essential medium with Earle's salts, supplemented with Gibco 15% heat-inactivated fetal bovine serum (FBS) (Gibco by Life Technologies, Grand Island, NY, USA), 2% Gibco antibiotic-antimycotic (100×), 2% Sigma-Aldrich L-glutamine (200 mM), and 7.5% $NaHCO_3$. Homogenization was performed using Mini-Beadbeater-16 (Biospec, Bartlesville, OK, USA) (Langat et al., 2021). The homogenates were clarified by centrifugation at 10,000 rpm for 10 minutes at 4ºC, and the supernatant was collected. The supernatant was transferred to a different tube for use in metagenomics and virus isolation, while the pellets were preserved for metabarcoding.

3.8 To Determine the Diversity of Insect Specific Viruses and Arboviruses

3.8.1 Super Pools Generation

Aliquots from the original pools' supernatants were combined to generate superpools of up to 500 sandflies, grouped according to their region of collection for metagenomic analysis. These superpools were mixed thoroughly by vortexing before viral RNA extraction. A total of 11 superpools, each containing ≤ 500 specimens and representing different study sites, were generated. For study sites with more than 500 specimens, a representative sample was selected using systematic sampling. Detailed

descriptions of each pool, including the number of specimens, collection site, and collection dates, are provided in Table 4.2.

Table 3.2: Geographic Distribution and Pooling of Female Phlebotomine Sandflies for Metagenomics

County	Site	Super Pool ID	Number of Specimens per Pool	Date of Collection
Baringo	Salabani	BARM1	399	March 2021
	Longumgum/Sirata	BARM3	170	March 2021
	Entepes	BARM4	496	March 2021
	Longwan	BARM5	461	March 2021
	KALRO	BARM6	500	March 2021
West Pokot	Asilong	PKTM1	454	Sept 2021
	Tapadany	PKTM4	469	Sept 2021
Kisumu	Kamonye/Asol	KSMM1	350	Sept 2020
Kwale	Mabatini	KWAM1	46	Oct-2021
Kilifi (Malindi)	Malindi Hospital/Vuga	MALM1	92	Oct-2021
Nakuru (Gilgil)	Utut/Gitare	GILM1	358	Jun-2022

3.8.2 RNA Extraction and Quantification

The supernatant from the superpools was syringe-filtered using 0.22 µm filters to remove host debris and bacteria while concentrating the viral particles. Viral RNA was extracted from sandfly homogenates using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) to detect both ISVs and arboviruses. Briefly, 140 µL of the filtered supernatant was added to microcentrifuge tubes containing 560 µL of Buffer AVL and carrier RNA, which helps stabilize the viral RNA and improve yield. The tubes were vortexed and incubated at room temperature for 10 minutes to allow the lysis of viral particles. Next, 560 µL of molecular-grade ethanol was added and pulse-vortexed for 15 seconds to precipitate the RNA, facilitating its binding to the QIAamp Mini column during subsequent centrifugation steps. The sample was then applied to the QIAamp Mini column, where viral RNA bound to the membrane while contaminants passed through. The column was washed twice through centrifugation, first with 500 µL of Buffer AW1 at 8000 rpm to remove any residual contaminants, and then with 500 µL of Buffer AW2 at 14,000 rpm to ensure the

purity of the RNA. Finally, viral RNA was eluted in 60 µL of Buffer AVE, ensuring the recovery of purified RNA.

The extracted nucleic acid was quantified using a Qubit RNA fluorometer (ThermoFisher, Waltham, MA, USA). Briefly, the Qubit RNA HS Reagent was diluted 1:200 with Qubit RNA HS Buffer to prepare the working solution. Standards and samples were prepared by adding 190 µL of the working solution to the assay tubes for the standards and 199µL of the working solution to the assay tubes for the samples, followed by the addition of 10 µL of RNA standard or extracted RNA sample. After gently mixing by vortexing, the tubes were incubated at room temperature for 2 minutes. Fluorescence was then measured using the Qubit RNA fluorometer, which calculates RNA concentration by comparing the sample fluorescence to that of the standards.

3.8.3 Illumina Sequencing Libraries Preparation and Sequencing

Sequencing libraries were prepared using the Illumina RNA Prep with Enrichment (L) Tagmentation Kit (Illumina, San Diego, California, USA) and IDT for Illumina DNA/RNA UD 96 Indexes Set A (New England Biolabs, Ipswich, MA, USA), with modifications to exclude hybridization and enrichment steps. Briefly, total RNA samples were denatured by combining 10–100 ng of RNA with 8.5 µL of Elute, Prime, Fragment High Concentration Mix and subjected to a thermal cycling program (65°C for 5 minutes, then held at 4°C) to disrupt RNA secondary structures.

First-strand complementary DNA synthesis was done by adding 8 µL of First Strand Synthesis Act D Mix and 1 µL of Reverse Transcriptase to the denatured RNA, with incubation conditions set at 25°C for 10 minutes, 42°C for 15 minutes, 70°C for 15 minutes, and then held at 4°C. This step converts RNA into complementary DNA, which is more stable and suitable for further processing.

The second strand of complementary DNA was synthesized using 25 µL of Second Strand Marking Master Mix and incubated at 16°C for 1 hour to generate double-stranded complementary DNA. The resulting complementary DNA was purified with

Agencourt AMPure XP beads at a 1.8× bead-to-sample ratio to remove excess reagents. Tagmentation was performed using 11.5 µL of Enrichment Bead-Linked Transposomes and 11.5 µL of Tagmentation Buffer, along with 14.5 µL of nuclease-free ultrapure water, which fragments the complementary DNA and incorporates adapter sequences in a single step. The tagmented complementary DNA was then amplified through polymerase chain reaction with 23 µL of Enhanced PCR Mix and indexed with 10 µL of unique dual-index adapters, followed by additional purification with AMPure XP beads at a 1.0× ratio to ensure the removal of residual reagents. The final libraries were quantified using the Qubit 2.0 Fluorometer with the Qubit™ dsDNA High Sensitivity Assay Kit (Invitrogen,) and diluted to 4 nM. Alongside experimental RNA samples, a negative control (no-template control) was included and processed through all steps of library preparation and sequencing to monitor potential cross-contamination. Four prepared superpools libraries and a negative control were pooled to equimolar concentrations, diluted to 120 pM, and 20 µL of this pool was loaded onto the Illumina Iseq cartridge for 2 × 150 bp paired-end sequencing.

3.8.4 Metagenomic Data Taxonomic Classification using DRAGEN and Kraken 2

The raw, unassembled metagenomic reads were uploaded to Illumina BaseSpace and analyzed using the DRAGEN Metagenomics pipeline with Kraken 2 for taxonomic classification (Wood & Salzberg, 2014). The classification is performed to the lowest common ancestor in the taxonomic tree, enabling robust identification and quantification of major taxonomic groups, including Eukaryota, Bacteria, Archaea, and Viruses (Wood & Salzberg, 2014).

3.8.5 Metagenomics Manual Data Analysis

The quality of generated sequences was analyzed with FastQC and processed using the Trimmomatic v0.39 using the following settings (ILLUMINACLIP option enabled, seed mismatch threshold = 2, palindrome clip threshold = 40, simple clip threshold of 15; SLIDING WINDOW option enabled, window size = 4, quality

threshold = 20; MINLEN =35; LEADING = 3; TRAILING = 3) (Bolger et al., 2014). This involved trimming of sequences with Phred score ≤ Q30 while removing, low complexity reads, and low-quality bases, minimum length 35bp. The resulting high-quality reads were mapped to their *Phelebotomus papatasi* M1 reference genome available at (https://vectorbase.org/vectorbase/app/record/dataset/TMPTX_ppapM1) for host reads removal using Bowtie2 (v. 2.3.4.3). The unmapped reads were assembled de novo using MEGAHIT version 1.2.9, employing a multi-k-mer strategy with k-mer sizes of 21, 29, 39, 59, 79, 99, 119, and 141 to optimize assembly across varying read lengths and coverage depths. This approach utilizes succinct de Bruijn graphs to efficiently construct contigs from metagenomic data, producing the final assembled contigs (Li et al., 2015). Viral contigs were identified by performing an online BLASTx (Altschul et al., 1990) against the NCBI databases (https://BLAST.ncbi.nlm.nih.gov/BLAST.cgi) with an E-value cut-off of 1×10^{-3}. Additional unassembled reads were mapped to built protein databases of various sandfly-borne arboviruses including Vesiculoviruses, Phleboviruses, and Orbiviruses using the BLASTX program built in DIAMOND v2.0.15 (Buchfink et al., 2015), with a cut-off E-value of $< 10^{-5}$. The final taxonomic classifications were assigned according to the contemporary taxonomic guidelines of the International Committee on Taxonomy of Viruses (ICTV).

3.8.6 Resequencing of Arbovirus Positive Sample on MiSeq

Arbovirus-positive superpools, for which primers could not be designed, were resequenced on the Illumina MiSeq platform after viral RNA enrichment for RNA-seq. The use of MiSeq was to increase the sequencing depth to get longer contigs that can be used to infer the phylogeny of viruses related to arboviruses, of which only short sequence reads were obtained using iSeq sequencing. Briefly, viral RNA was extracted from 140 µL of filtered superpool supernatant using the QIAamp Viral RNA® Kit (Qiagen, Hilden, Germany). The extracted nucleic acid was treated with TURBO™ DNase (Ambion) to digest DNA, followed by ribosomal RNA removal using the QIAseq FastSelect –rRNA HMR (Qiagen). The cleaned RNA was prepared for sequencing using the Illumina RNA Prep with Enrichment (L) Tagmentation Kit

(Illumina, San Diego, California, USA) as described in section 3.5.3 and sequenced on the Illumina MiSeq platform using 2 × 300 bp paired-end chemistry.

3.8.7 Data Analysis of MiSeq Resequencing Data

The data generated from the MiSeq resequencing run were analyzed using the same automated (DRAGEN & Kraken 2) and manual bioinformatics workflows as described for the initial Iseq sequencing (see section 3.8.4). This included quality control, host read removal, assembly, and taxonomic classification using BLASTx and DIAMOND against curated viral databases.

3.8.8 Characterization of Metagenome-Assembled Viral Genome Sequences

Open reading frames (ORFs) in the viral genomes were predicted using the NCBI ORFfinder tool (https://www.ncbi.nlm.nih.gov/orffinder/), which identifies potential protein-coding regions within nucleotide sequences.. Putative protein domains within the predicted ORFs were then annotated using the NCBI Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to identify functional motifs and conserved regions.

3.8.9 Phylogenetic Analysis

Phylogenetic inference was conducted using the RNA-dependent RNA polymerase (RdRp) coding regions, a conserved marker for viral classification, from the virus genomes identified in this study, along with reference sequences from GenBank. Multiple sequence alignment of the generated datasets was performed using MAFFT v7.511 (Katoh & Standley, 2013). Aligned viral sequences were trimmed using BioEdit (T. A. Hall, 1999) and sites containing more than 50% gaps were removed from alignments. IQ-TREE v2.0.7 software was employed to generate a maximum likelihood phylogenetic tree (Nguyen et al., 2015) using ultrafast bootstrapping with 1000 replicates (Felsenstein, 1985). The appropriate model for the dataset was determined using Model-finder (Kalyaanamoorthy et al., 2017). The resulting tree was annotated and visualized in FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). All trees are mid-point rooted.

3.9 To Determine the Genomics Characteristics of the Isolated Arboviruses

3.9.1 Virus Isolation in Cell Culture

Virus isolation was performed in Vero (African green monkey kidney) cells (CCL-81™), passage 13, obtained from the Viral Haemorrhagic Fever lab at KEMRI, as described before (Lwande et al., 2013). Vero cells were selected because they are highly susceptible to a wide range of viruses, including many arboviruses and haemorrhagic fever viruses, due to their deficient interferon response, which allows for more efficient viral replication and higher viral titres compared to other cell lines (Osada et al., 2014). This broad susceptibility makes Vero cells the preferred choice for isolating both known and novel RNA viruses from field samples, as demonstrated in multiple virology studies and viral surveillance programs (Osada et al., 2014). Briefly, Vero cells grown overnight at 37 °C and 5% CO_2 in minimum essential medium supplemented with 2% glutamine, 2% penicillin/streptomycin/amphotericin, 10% fetal bovine serum, and 7.5% $NaHCO_3$ in 24-well plates (Corning, Incorporated). At 80% confluence, a 50 µL aliquot of the clarified supernatant from individual pools was inoculated into the wells. The plates were incubated for 1 hour in a humidified incubator at 37 °C and 5% CO_2 with gentle rocking of the plates side to side every 15 minutes for virus adsorption. Following incubation, 1 mL of maintenance medium, comprising minimum essential medium supplemented with 2% glutamine, 2% penicillin/streptomycin/amphotericin, 2% fetal bovine serum, and 7.5% $NaHCO_3$, was added. The plates were then cultured in a humidified incubator at 37 °C with 5% CO_2. They were monitored daily under a light microscope for any cytopathic effect (CPE) formation for up to 14 days. Uninfected Vero cells were used as negative controls. Cultures exhibiting CPE were harvested and further passaged by inoculating them onto fresh monolayers of Vero cells (CCL 81™) in 25-cm^2 cell culture flasks. After two successive passages, the supernatants of virus-infected Vero cell cultures exhibiting a cytopathic effect of approximately 70% were harvested from the flasks for virus identification through next-generation sequencing.

3.9.2 Illumina Library Preparation and Sequencing

Viral particles from CPE-positive cultures were recovered through 0.22µm filters (Millipore, Merck). From an aliquot of 140 µl of the supernatant, viral RNA was extracted using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) and eluted in one step with 60 µl of elution buffer. The extracted nucleic acid was quantified using the Qubit RNA fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Sequencing libraries were prepared using Illumina RNA Prep with Enrichment, Tagmentation Kit (Illumina, USA), following the manufacturer's instructions. The obtained libraries were assessed for quantity using the Qubit 2.0 Fluorometer (ThermoFisher, Waltham, MA, USA). Sequencing was performed on an Illumina Iseq platform (Illumina, San Diego, California, USA) using iSeq 100 i1 Reagent v2 (Illumina, San Diego, California, USA) in a 2 × 150 bp paired-ended configuration.

3.9.3 Sequencing Data Analysis

The sequence data was initially processed for quality using Trimmomatic v0.39 (ILLUMINACLIP option enabled, seed mismatch threshold = 2, palindrome clip threshold = 40, simple clip threshold of 15; SLIDING WINDOW option enabled, window size = 4, quality threshold = 30; MINLEN =35; LEADING = 3; TRAILING = 3) (Bolger *et al.,* 2014). This step involved the removal of adapters, duplicates, and low-quality sequences. Sequences with phred scores above 30 (indicating a high probability of accuracy in base calling), were used for further analysis. The cleaned reads were then assembled using Megahit v1.2.9 (Li et al., 2015)with various k-mer values to reconstruct the viral genome. To identify viral contigs within the assembly, a BLASTn (Altschul et al., 1990) search was conducted against the NCBI databases (https://BLAST.ncbi.nlm.nih.gov/BLAST.cgi), using an E-value cutoff of 1e-5 to ensure significant matches to known viral genomes. To determine the sequencing depth of the obtained genomes, sequencing reads were mapped against the genomes using Bowtie2 v2.5.2 (Langmead & Salzberg, 2012) in “end-to-end” mode with “--very-sensitive-local” alignment settings, and the depth of coverage was determined using Samtools depth v1.13 with a minimum mapping quality threshold of 10.

3.9.4 Genome Organization and ORF Prediction

Following genome assembly, viral genomes were analyzed to determine their genetic organization and coding potential. ORFs were predicted using the NCBI ORFfinder tool (https://www.ncbi.nlm.nih.gov/orffinder/). Conserved protein domains and functional motifs were identified through homology searches against the CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Multiple sequence alignments with reference viral proteins were performed using MAFFT (Katoh & Standley, 2013) to assist in identifying protein cleavage sites and structural features. Manual curation of annotations was conducted to ensure consistency with known viral genome architectures.

3.9.5 Amino acid Variations and Computational Protein Modeling

Comparative amino acid analyses were performed on the viral isolates complete coding regions. Amino acid alignment was performed using MEGA11 (Tamura et al., 2021), showing only variable sites in each gene. The unique amino acid variation at the site of interest was mapped onto the Crystal structure of protein (Protein Data Bank) using the PyMOL Molecular Graphics System, Version 2.6 Schrödinger, LLC. Electrostatic surface potential was calculated using the pdb2pqr server and APBS (v1.5) (Jurrus et al., 2018). Graphical representations of electrostatics surface potential were generated using NGL viewer (Rose et al., 2018).

3.9.6 Phylogenetic Analysis

Complete genomes representing all WNV lineages were retrieved from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) https://www.bv-brc.org/ and combined with the genome sequences of the isolated viruses. Alignment was performed using MAFFT v7.490 (Katoh & Standley, 2013) and manually curated using BIOEDIT (T. A. Hall, 1999). Phylogenetic analysis was conducted using the Maximum Likelihood method in IQTREE v2.0.7 (Nguyen et al., 2015) using GTR+F+I+G4 (General Time Reversible + Frequencies + Invariant Sites + Gamma distribution with 4 rate categories) substitution model as determined using Model-

finder (Kalyaanamoorthy et al., 2017). The robustness of the tree topology was assessed using 1000 non-parametric bootstrap analyses (Felsenstein, 1985). The resulting tree was annotated and visualized in FigTree v.1.4.4 (available from http://tree.bio.ed.ac.uk/software/figtree/).

3.9.7 Molecular Clock Analysis

To estimate coalescent times of the most recent common ancestor (tMRCA) for the isolated viruses, molecular clock analysis was performed using the Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v2.1.3 software package (Bouckaert et al., 2019). Complete genomes with country and year of isolation data, spanning continents worldwide, were obtained from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) https://www.bv-brc.org/. These genomes were added to the dataset for maximum likelihood phylogenetic analysis. The analysis employed the uncorrelated lognormal relaxed molecular clock model to account for rate variation across branches in the phylogenetic tree. Geographical traits were assigned to the sequences, and a flexible non-parametric Bayesian Skyline model and an asymmetric trait model were used to estimate past population dynamics and ancestral states, respectively. The MCMC analyses were run for 800 million states and sampled every 100,000th generation. Tracer v1.7.1 (Rambaut et al., 2018) was employed to assess the logs and ensure adequate effective sample sizes (ESS) values > 200 by visual inspection of chains. Maximum clade credibility (MCC) trees were generated using TreeAnnotator v2.1.2, discarding the initial 10% as burnin. The resultant phylogeny was visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), showing posterior probability values in each node and the time to the most recent common ancestors (tMRCA) as median year with 95% Highest Posterior Density (HPD).

3.9.8 Evolutionary Selection Pressure Analysis

Evolutionary pressures acting across the entire coding sequences of isolates' viral genomes were assessed using various methods, including fixed-effects likelihood (FEL), single-likelihood ancestor counting (SLAC), mixed effects model of evolution

(MEME), and fast unconstrained Bayesian approximation (FUBAR) on the Datamonkey selective and evolutionary bioinformatics online server (Weaver et al., 2018). To determine the selective pressure, the ratio (ω) of the rate of non-synonymous substitutions (dN) to the rate of synonymous substitutions (dS) per codon site was calculated using the SLAC and FEL codon-based maximum likelihood approaches. A value of $\omega \geq 1$ indicates positive selection, $\omega \leq 1$ indicates negative selection, and $\omega = 0$ indicates neutral selection. For statistical significance, the following thresholds were applied: $p < 0.05$ for SLAC, FEL, and MEME, and a posterior probability of ≥ 0.9 for FUBAR. Positive selection at a specific site was considered present when at least three methods detected it.

3.9.9 Confirmation of the Presence of Isolated Viruses in Original Pools.

Viral genomic sequences from arboviruses isolated in cell culture were used to develop specific primers and probes for the detection of West Nile Koutango virus (WN-KOUTV) and Bogoria virus (BOGV) as shown in Table 3.3 and Table 3.4. Primer and probe design was performed using the PrimerQuest Tool (Integrated DNA Technologies, IDT). The design parameters included: Primer Length: 18–25 nucleotides, Melting Temperature (Tm): 50–60°C, GC Content: 40–60%, Product Size: 100–300 base pairs, Probe Length: 20–30 nucleotides, Probe Tm: 68–72°C, Avoid Hairpins: ≤ 5% secondary structure, Avoid Dimers: ≤ 2% primer-primer dimer formation.

Table 3.3: Primer and Probe Sequences for Bogoria Virus L Segment

Bogoria virus					
Segment	**Primer Name**	**Sequence 5'-3'**	**Start**	**Length**	**Product**
L	BOGV_L_F	GGTGTGTTCCTTAGCCATCTTA	1655	22	118bp
L	BOGV_L_R	CAGAGAAGCTAGTGAGCGATAAA	1772	23	
L	BOGV_L_Probe	FAM-TCCTGCTCTTCTCAGTTCCAAACCA-BHQ	1725	25	

Table 3.4: Primer and Probe Sequences for WN-KOUTV

WN-KOUTV				
Primer Name	**Sequence 5'-3'**	**Start**	**Length**	**Product**
WN-KOUTV-1-F	ATGGAGTCAGGCCAGGTAA	10407	19	101bp
WN-KOUTV-1-R	GATCTCCTCGCAGCTTTGTT	10507	20	
WN-KOUTV-1-Probe	HEX-ACTGCCACCGGAAGTTGGGT-BHQ	10422	20	

In silico validation was performed using BLAST to ensure the specificity of the designed primers and probes against known viral and non-viral sequences. Viral RNA was isolated from 140 µL of original sandfly supernatant using the QIAamp Viral RNA Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Reverse transcription polymerase chain reaction (RT-PCR) was performed to test the amplification efficiency and optimization of the designed primers and probes. The RT-PCR conditions, including temperature cycles, enzyme concentrations, and reagent volumes, were systematically adjusted to ensure optimal amplification of the target viral sequences

A multiplex real-time RT-PCR assay was optimized for the simultaneous detection of WN-KOUTV and BOGV. The assay was performed using the AgPath-ID One-Step RT-PCR Kit (Applied Biosystems™). The final reaction mixture was 25 µL and comprised 0.8 µM for primers and 0.2 µM for the probe as in the following master mix setup and cycling conditions Table 3.5.

Table 3.5: Multiplex RT-PCR Assay Set up for WN-KOUTV and BOGV

Reaction Components	1x Reaction	X Reactions	PCR Reaction Conditions
2X AgPath RT-PCR Buffer	12.5 µL		**Preholding** 50°C -10 minutes
WN-KOUTV-probe/primers	1 µL		**Holding** 95°C -10 minutes
BOGV-probe/primers	1 µL		**Cycling** (95°C -15 sec 58°C -45 sec) 45 cycles- Data collection
25x AgPath RT-PCR Enzyme	1 µL		
Nuclease free water	4.5 µL		
Template RNA	5 µL		
Total	25 µL		

3.10 To Determine the Taxonomic Identity of Phlebotomine Sandflies Species in Arbovirus Positive Pools.

3.10.1 Phlebotomine Sandflies DNA Extraction and Amplification of Mitochondrial Cytochrome Oxidase I (CO1) Gene

Whole genomic DNA was extracted from homogenate pellets of virus-positive pools using a QIAamp DNA Mini Kit (Qiagen, Germany), according to the manufacturer's instructions. Briefly, pellets were resuspended in 500 µl of homogenization media, which consisted of minimum essential media supplemented with 15% fetal bovine serum (FBS) (Gibco by Life Technologies, Grand Island, NY, USA), 2% L-glutamine (Sigma-Aldrich), and 2% antibiotic/antimycotic (Gibco by Life Technologies, Grand Island, NY, USA), then vortexed. 20 µl Proteinase K and 200 µl Buffer AL were added to 200 µl aliquot of the crude homogenate, vortexed, and incubated at 56°C for 10 minutes for lysis. Next, 200 µl of molecular grade ethanol (96%) was added to precipitate the nucleic acids, vortexed, and the sample was transferred to a QIAamp Mini spin column. After centrifugation at 6000 rpm for 1 minute, the flow-through was discarded. The precipitated DNA in the mini column was washed with 500 µl Buffer AW1, centrifuged, then with 500 µl Buffer AW2, and centrifuged at 14,000 rpm for 3 minutes to ensure thorough washing. DNA was eluted with 200 µl Buffer AE, incubated at room temperature for 1 minute, and centrifuged for 1 minute at 6000 rpm. One extraction blank was included. The extracted DNA was quantified using the Qubit ds DNA HS fluorometer (Thermo

Fisher Scientific, Waltham, Massachusetts, USA). PCR amplification of the CO1 gene was performed using universal metazoan primers, which amplify approximately 710-bp region of the COI gene of arthropod vectors. A 25-µl reaction mix containing 12.5 µl of NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs), 0.5 µl of each respective primer at 25 µM, 2 µl of extracted DNA template, and nuclease-free water was prepared. The PCR conditions were as follows: initial denaturation at 94°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, annealing at 49°C for 30 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 7 minutes.

Table 3.6: Sequences of Primers for PCR Amplification of Phlebotomine Sandfly's CO1 Gene

Gene	Primers ID	Sequence 5-3′	Reference
CO1	LCO1490	GGTCAACAAATCATAAAGATA-TTGG	(Hebert et al., 2003)
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	

The success of the amplification was visualized on a 1.5% agarose gel along with a reference 1Kb DNA ladder and visualized with a UV transilluminator.

3.10.2 MinION Library Preparation and Sequencing

PCR amplicons were purified using Agencourt AMPure XP (Beckman Coulter) according to the manufacturer's instructions. Briefly, AMPure XP beads were brought to room temperature and thoroughly vortexed before adding 1.8× volume of beads to the DNA sample and mixing well. The mixture was incubated for 5 minutes, placed on a magnetic rack until the solution cleared, and the supernatant was discarded. The beads were washed twice with 80% ethanol, air-dried for up to 5 minutes, resuspended in elution buffer, and the cleared DNA was transferred to a new tube, carefully avoiding bead carryover. The DNA concentration was measured using a Qubit Fluorometric Quantitation fluorometer (Thermo Fisher Scientific) using the Qubit dsDNA High Sensitivity kit. The Qubit dsDNA HS Reagent was diluted at 1:200 with Qubit dsDNA HS Buffer, and the standards and samples were prepared in assay tubes and incubated. After a 2-minute incubation at room

temperature, fluorescence was measured using a Qubit Fluorometer, which calculates DNA concentration based on the binding of a fluorescent dye to double-stranded DNA, comparing the sample fluorescence to the standards.

The CO1 amplicon sequencing libraries were prepared using the Native Barcoding Kit 96 (SQK-NBD112.96) from Oxford Nanopore Technologies, following the manufacturer's instructions. The required starting material was 200 fmol, and the amount of DNA was calculated accordingly. The DNA ends were repaired using the NEBNext Ultra II End Repair reagent (New England Biolabs) to prepare for adapter attachment. The end-repaired amplicons were then individually barcoded with the Native Barcoding Expansion 1-12 Kit (Oxford Nanopore Technologies). Native Barcodes were ligated to the end-repaired amplicons using the Blunt/TA Ligase Master Mix (New England Biolabs). The barcoded libraries were then pooled, and adapters were ligated using the NEBNext Quick Ligation Module (New England Biolabs). The prepared libraries were purified using AMPure XP beads, loaded onto the flow cell, and sequenced using the workflows provided on a MinION Mk1B platform.

3.10.3 CO1 Gene Sequences Analysis and Taxonomic Assignment

Base-calling of raw reads was performed using Guppy Version 6.3.8 (Oxford Nanopore Technologies, Oxford, United Kingdom) into FastQ files (Wick et al., 2019). Reads of low quality and sequences identical to those detected in the extraction blanks were removed using PycoQC Version 2.5.2 (Leger & Leonardi, 2019). De-multiplexing, removal of barcodes, and adapters were done in Porechop Version 0.2.4 (Wick, 2017), after which Trimmomatic Version 0.36 (Bolger et al., 2014) was used for quality filtering and trimming. These pre-processing steps were essential for preparing the data for accurate Operational Taxonomic Units OTUs clustering and taxonomic identification. For OTUs generation, dereplicated sequences were clustered at a 98% similarity threshold using VSEARCH, resulting in a set of distinct OTUs. Taxonomic assignment of the generated consensus sequences was performed using the MIDORI server using RDP Classifier with COI reference sequence database (Leray et al., 2013). MIDORI server uses extensive reference

databases to accurately classify sequences by comparing them against GenBank and the Barcode of Life Data Systems Identification engine (BOLD-ID). Taxonomic assignments were further validated by searching the consensus sequences against the NCBI nr database.

3.11 Statistical Analysis

All statistical analyses and data visualizations were performed using R statistical software (R Development Core Team, 2015). Descriptive statistics, including counts and proportions, were calculated to summarize the distribution of Phlebotomine sandfly species and sequencing data. Visualization techniques such as pie charts and stacked bar plots were employed to effectively illustrate the relative abundance of taxonomic groups and viral families across samples.

CHAPTER FOUR

RESULTS

4.1 Phlebotomine Sandflies Distribution

Six thousand three hundred and twelve female Phlebotomine sandflies collected between September 2020 and June 2022 from six different geographical locations in Kenya (Baringo, Nakuru, Kwale, Kilifi, West Pokot, and Kisumu), as shown in Fig. 3.1, were examined for the presence of RNA viruses. They were organized into 700 pools according to specific collection sites, trap numbers, trap types, and collection dates, genus such as either *Sergentomyia spp.* or *Phlebotomus spp*., and the locality of collection. Table 4.1 summarizes the distribution of sandflies according to their genus and geographical region.

The highest numbers were recorded in Baringo and West Pokot compared to the coastal counties of Kwale and Kilifi. This marked difference in sandfly abundance can be attributed to several ecological factors prevalent in the Rift Valley regions. Baringo and West Pokot offer abundant suitable microhabitats such as animal sheds, termite mounds, and rodent burrows, which serve as optimal breeding and resting sites for sandflies, especially during the dry season when their populations peak (Anjili et al., 2011; Hassaballa et al., 2021). The semi-arid climate and bimodal rainfall patterns in these areas further support sandfly survival and reproduction, as the annual pattern of rainfall distribution is more influential than total precipitation in determining local sandfly abundance (Anjili et al., 2011). In contrast, the coastal regions experience higher humidity and different habitat structures, which are less favorable for sandfly proliferation.

Table 4.1: Distribution of Phlebotomine Sandflies from Different Geographical Regions in Kenya

Geographical Region	Genus	
	Sergentomyia spp.	***Phlebotomus spp.***
Baringo	3,527	152
Nakuru (Gilgil)	13	345
Kilifi (Malindi)	89	3
Kwale	42	4
Kisumu	350	0
West Pokot	1683	104
Total	**5704**	**608**
	6312	

4.2 Diversity of Insect Specific Viruses and Arboviruses using Metagenomics

4.2.1 Metagenomics Reads Analysis

A total of approximately 8.7 million raw reads were generated across all samples sequenced in four separate iSeq runs, with per-sample read counts ranging from 467,784 to 1,124,538. After quality filtering, the number of retained reads per sample ranged from 447,334 to 1,110,080 (Table 4.2).

Table 4.2: Number of Reads Before and After Quality Filtering per Sample

County	Super Pool ID	No of Reads	No of reads after Filtering
Baringo	BARM1	833,660	822,462
	BARM3	558,942	547,862
	BARM4	467,784	447,334
	BARM5	616,972	594,894
	BARM6	1,031,552	1,006,653
West Pokot	PKTM1	1,124,538	1,110,080
	PKTM4	854,674	839,984
Kisumu	KSMM1	838,522	826,770
Kwale	KWAM1	783,628	769,176
Kilifi (Malindi)	MALM1	517,494	505,498
Nakuru (Gilgil)	GILM1	780,754	761,250

Taxonomic classification revealed that, on average, over 80% of the classified reads were assigned to Eukaryota, while viral reads accounted for less than 1% (Figure 4.1).

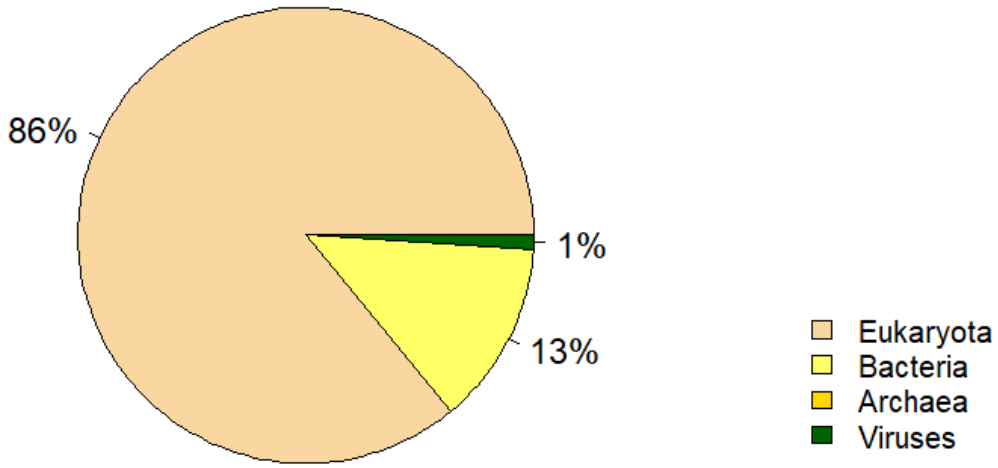


Figure 4.1: Taxonomic Classification of Metagenomic Reads

Legend

Pie chart showing the proportion of classified metagenomic reads assigned to major

taxonomic groups. Eukaryota (light peach), followed by Bacteria (yellow), Viruses (green), and Archaea (light gold). Classification was performed using the DRAGEN Metagenomics pipeline with Kraken 2.

Among the classified viral reads, members of the family *Dicistroviridae* represented the predominant proportion across most sample collection sites, with *Iflaviridae* and *Chuviridae* also present at notable levels. Additionally, smaller proportions of arbovirus-related families, including Flaviviridae, *Phenuiviridae*, and *Rhabdoviridae*, were detected at certain sites. This distribution highlights the dominance of RNA viruses within the order Picornavirales in the metagenomic profiles, while also confirming the presence of medically important arboviruses as shown in Figure 4.2.

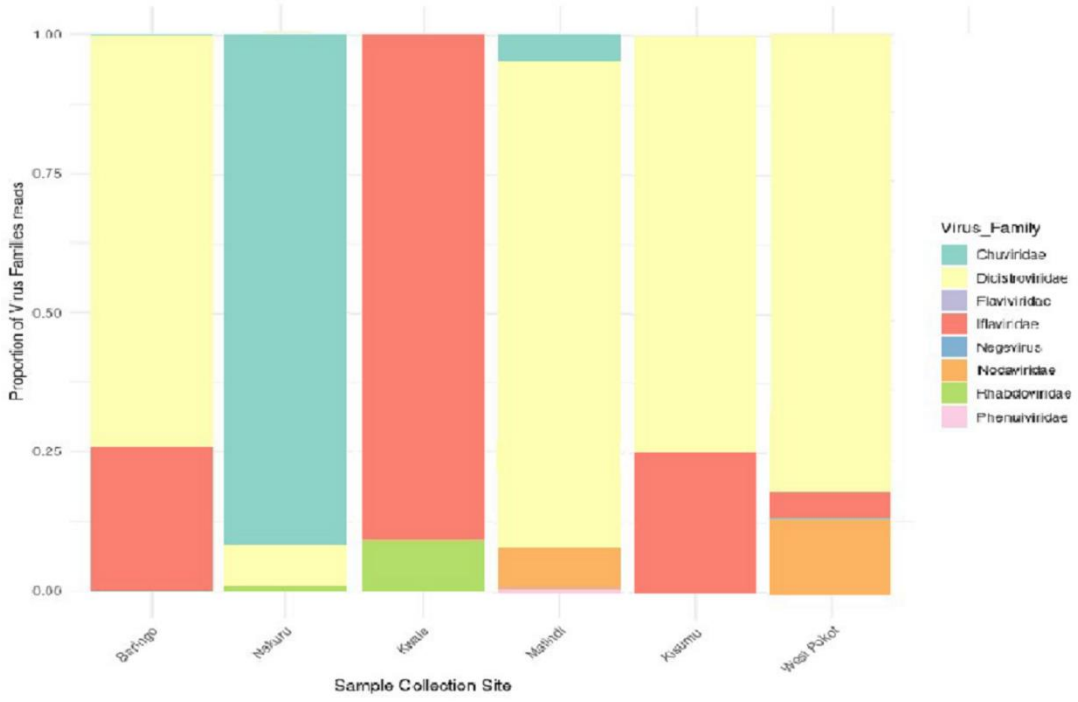


Figure 4.2: Proportion of Virus Family Reads Detected at each Sample Collection Site

Legend

Stacked bar plots show the relative abundance (proportion) of viral sequencing reads assigned to each virus family across different sample collection sites. The y-axis

represents the proportion of total virus family reads per site (ranging from 0 to 1), while the x-axis indicates the geographic location of sample collection. Colors correspond to different virus families as indicated in the legend.

4.2.2 Metagenomics Contigs Analysis

Sequence assembly and analysis revealed the presence of a variety of RNA viruses, identified based on the RNA-dependent RNA polymerase (RdRp) gene. These included positive-sense single-stranded RNA viruses from the families *Dicistroviridae*, *Iflaviridae*, *Nodaviridae*, and *Flaviviridae*, as well as negative-sense single-stranded RNA viruses from the families *Phenuiviridae*, *Chuviridae*, and *Rhabdoviridae*. Most of the identified viruses were insect-viruses; however, several reads associated with mammalian viruses belonging to the families Flaviviridae, *Phenuiviridae*, and *Rhabdoviridae* were recovered. Most of the viruses detected are novel, exhibiting relatively low similarity to known viruses, with some sharing as little as 30% amino acid identity as shown in Table 4.3.

Table 4.3: Putative Viral Sequences and their Closest Hits

Virus Name	Virus Family	Study Region	Contig Length Range (nt)	Closest Hit	Identity (aa %)
Phlebotomine picorna-like virus	Unclassified Picornavirales	Kilifi (Malindi)	10,150	Soybean thrips dicistrovirus 2	30.28
		Baringo	10,134		31.36
		West Pokot	9,934		30.21
Aphis gossypii virus	*Dicistroviridae*	Baringo	9,477	Aphis gossypii virus	98.11
Phlebotomine dicistrovirus		Kisumu	612–2,781	Homalodisca coagulata virus 1	48.03
Phlebotomine nodavirus	*Nodaviridae*	Kilifi (Malindi)	1613-2790	Lutzomyia Nodavirus	60.10
		West Pokot	799-1210		57.22
Phlebotomine chuvirus	*Chuviridae*	Nakuru (Gilgil)	592-4630	Culex mosquito virus 4	41.68
		Kilifi (Malindi)	823-4277		45.40
Phlebotomine iflavirus 1	*Iflaviridae*	Baringo	9,949	Iflavirus sp.	49.57
Phlebotomine iflavirus 2		Kwale	467–1,757 8874*	Sacbrood virus	43.52
Phlebotomine negevirus 1	*Negevirus*	Baringo	504	Utsjoki negevirus 1	65.45
Bogoria Virus	*Phenuiviridae*	Kilifi (Malindi)	20 reads (90–150 nt)	Bogoria virus	98–100
WN-KOUTV	*Flaviviridae*	Baringo	1 read (40 nt)	WN-KOUTV	98
Phlebotomine rhabdovirus	*Rhabdoviridae*	Kwale	7 reads (84–150 nt)	Vesicular stomatitis jersey, Chandipura virus	73–91
			8 contigs (303-1385) *	Chandipura virus	49.46

Legend

* Contig length range obtained from Illumina Miseq resequencing.

4.2.3 Unclassified Virus Within the Order Picornavirales

Pircorna-like genomes of 10,150 nt, 10,134 nt, and 9,934 nt were recovered from the MALM1, BARM4, and PKTM4 superpools, respectively. These genomes are distantly related to unclassified dicistroviruses, sharing only 30.28% to 31.36% amino acid identity with the non-structural protein of Soybean thrips dicistrovirus 2 (Accession no. MT224137.1) (Table 4.3). These genomes feature two overlapping open reading frames (ORFs), deviating from the canonical dicistronic, non-overlapping genome organization characteristic of *Dicistroviridae* (Valles et al., 2017). ORF1 is 6,760 nucleotides long, encoding a 2,162 amino acid polyprotein containing conserved peptidase 3C (residues 1455–1594), helicase (residues 695–815), and RdRp (residues 1751–2119) domains. ORF2 spans 3,248 nucleotides and encodes a 1,099 amino acid polyprotein including a VP4 domain (residues 316–361) (Figure 4.3). Due to these features, this virus is tentatively designated Phlebotomine picorna-like virus. The Phlebotomine picorna-like virus genomes form a sister clade to unclassified dicistrovirus as shown in Figure 4.4.

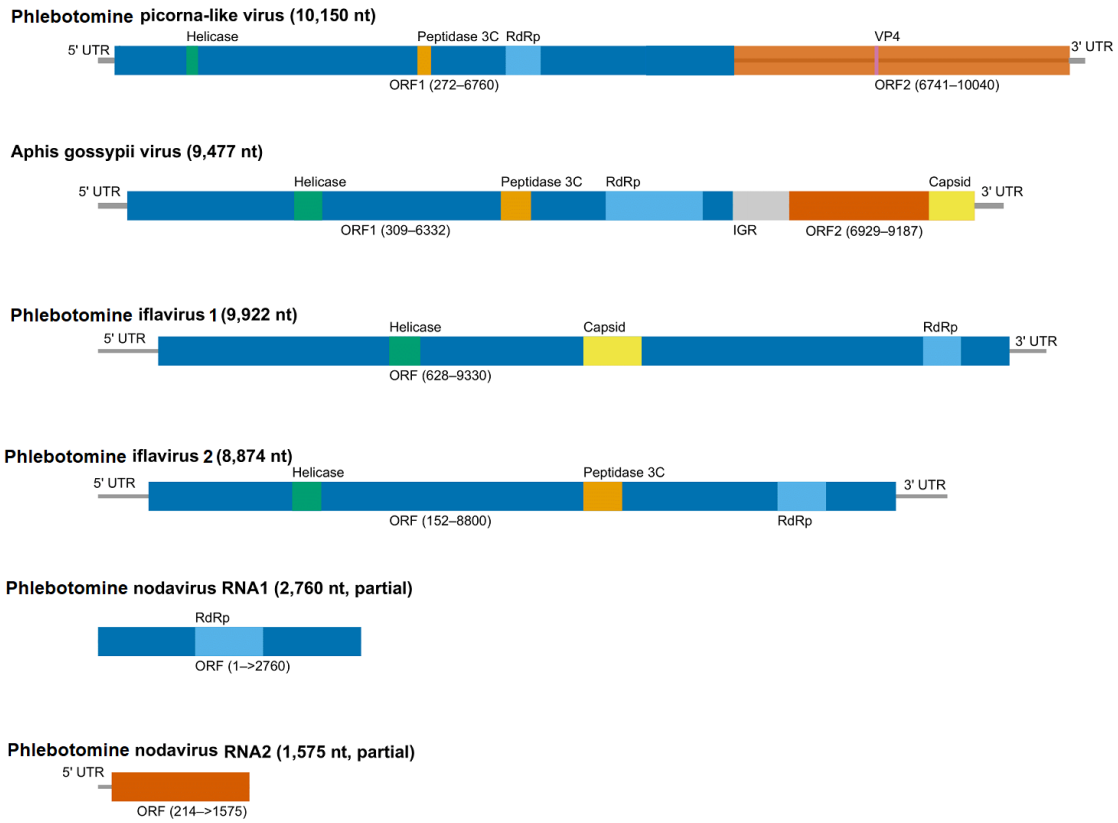


Figure 4.3: Schematic Representation of the Genome Organization of the Identified RNA Viruses

Legend

Linear diagrams represent the genomic architecture of each virus, drawn to scale. Blue and brown bars indicate open reading frames (ORFs), with colored boxes denoting conserved protein domains: green for Helicase, orange for Peptidase 3C, light blue for RNA-dependent RNA polymerase (RdRp), yellow for Capsid, and brown for VP4. Untranslated regions (UTRs) are shown as short gray lines at the 5′ and 3′ ends. The intergenic region (IGR) in Aphis gossypii virus is depicted in light gray. ORF coordinates and partial genome status are indicated where applicable.

4.2.4 *Dicistroviridae* Family

Members of the family *Dicistroviridae* were identified in superpools BARM5 and KSMM1. Dicistroviruses are known to infect invertebrates. The BARM5 superpool contained a nearly complete genome (9,477 nucleotides) highly similar to Aphis gossypii virus, a member of the genus *Cripavirus* (GenBank accession no MH476200.1), sharing 98.11% amino acid identity (Table 4.3). This genome encodes two ORFs separated by an IGR, ORF1 (nt 309–6332) and ORF2 (nt 6929–9187), which translate into proteins of 2,007 and 752 amino acids, respectively. ORF1 harbors conserved motifs characteristic of the peptidase 3C domain (nt 4218–4532), helicase domain (nt 2046–2336), and RdRp domain (nt 5184–6200). ORF2 contains a conserved capsid domain spanning nucleotides 8585–9184 (Figure 4.3).

A second dicistrovirus was detected in the KSMM1 superpool, consisting of three contigs totaling 4,761 nt. This virus shares 45.38% to 49.38% amino acid identity with Homalodisca coagulata virus 1 (GenBank accession no. NC_008029.1), a member of the genus *Triatovirus* (Table 4.3). This virus has been designated Phlebotomine dicistrovirus. Consistent with nucleotide similarity results, phylogenetic analysis of the RdRp gene amino acid sequences placed the BARM5_1 Aphis gossypii virus within the *Cripavirus* genus, and the Phlebotomine dicistrovirus within the *Triatovirus* genus, as shown in Figure 4.4

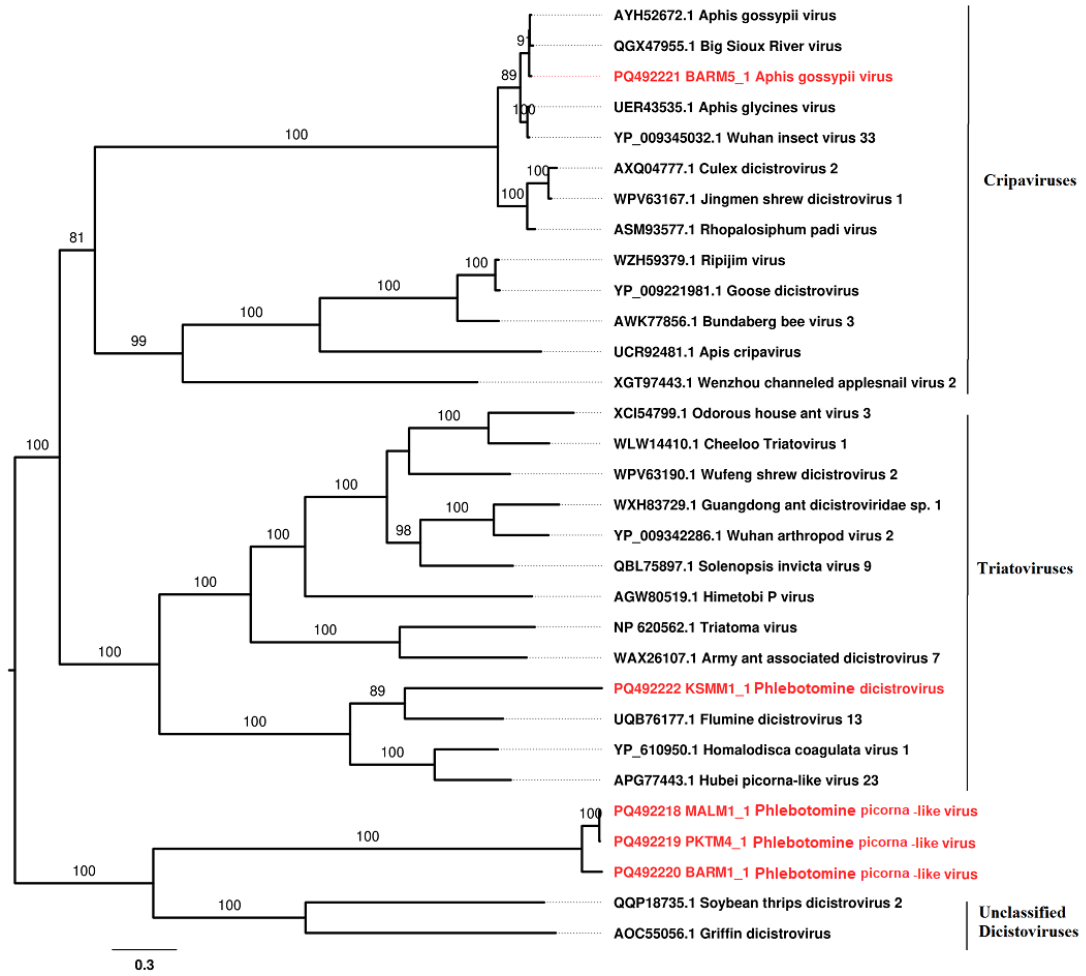


Figure 4.4: Phylogenetic Relationship of Identified Genomes Related to Dicistroviruses and Reference Viruses

Legend

Midpoint-rooted phylogenetic tree based on amino acid sequence alignments of the RdRp gene (~1,990 residues), showing relationships between dicistroviruses identified in this study (highlighted in red) and reference viruses. The tree was generated with 1,000 bootstrap replicates; support values ≥70 are indicated at relevant nodes. The identified viruses cluster within established genera (*Cripavirus*, *Triatovirus*) and unclassified Dicistroviruses.

4.2.5 *Nodaviridae* Family

Genomes were recovered for both RNA1 (2,760 nt) and RNA2 (1,575 nt) segments from the MALM1 superpool, and RNA1 (2,324 nt) and RNA2 (870 nt) from the PKTM4 superpool (Table 4.3). These lengths represent approximately 87% and 77% of the expected full segment sizes for RNA1 and RNA2, respectively, in MALM1, and approximately 73% and 42% for RNA1 and RNA2, respectively, in PKTM4, based on the expected nodavirus segment sizes. The RNA1 of the MALM1 genome contains a single partial ORF spanning approximately nucleotides <1 to >2,760, which includes a conserved RdRp domain located between nucleotides 1,498 and 2,208, while RNA2 also contains one ORF spanning nucleotides 214 to >1,575 as shown in Figure 4.3. The genomes are related to Lutzomyia nodavirus, detected from Phlebotomine sandflies in Brazil in 2012 (Aguiar et al., 2015), sharing 57.22% and 60.10% amino acid identity (Table 4.3).

Based on ICTV species demarcation criteria-less than 80% nucleotide identity for RNA2 and less than 87% amino acid identity, this nodavirus was designated phlebotomine nodavirus. Phylogenetic analysis based on RdRp amino acid sequences revealed that the recovered Phlebotomine nodavirus genomes (accession numbers PQ492225 for MALM1 and PQ492226 for PKTM4) form a distinct cluster within the unclassified nodaviruses clade as shown in Figure 4.5. Notably, these genomes group closely with Lutzomyia nodavirus but remain phylogenetically distinct from the established alphanodaviruses and other classified nodavirus groups. This clustering supports the designation of Phlebotomine nodavirus as a novel, unclassified member of the *Nodaviridae* family.

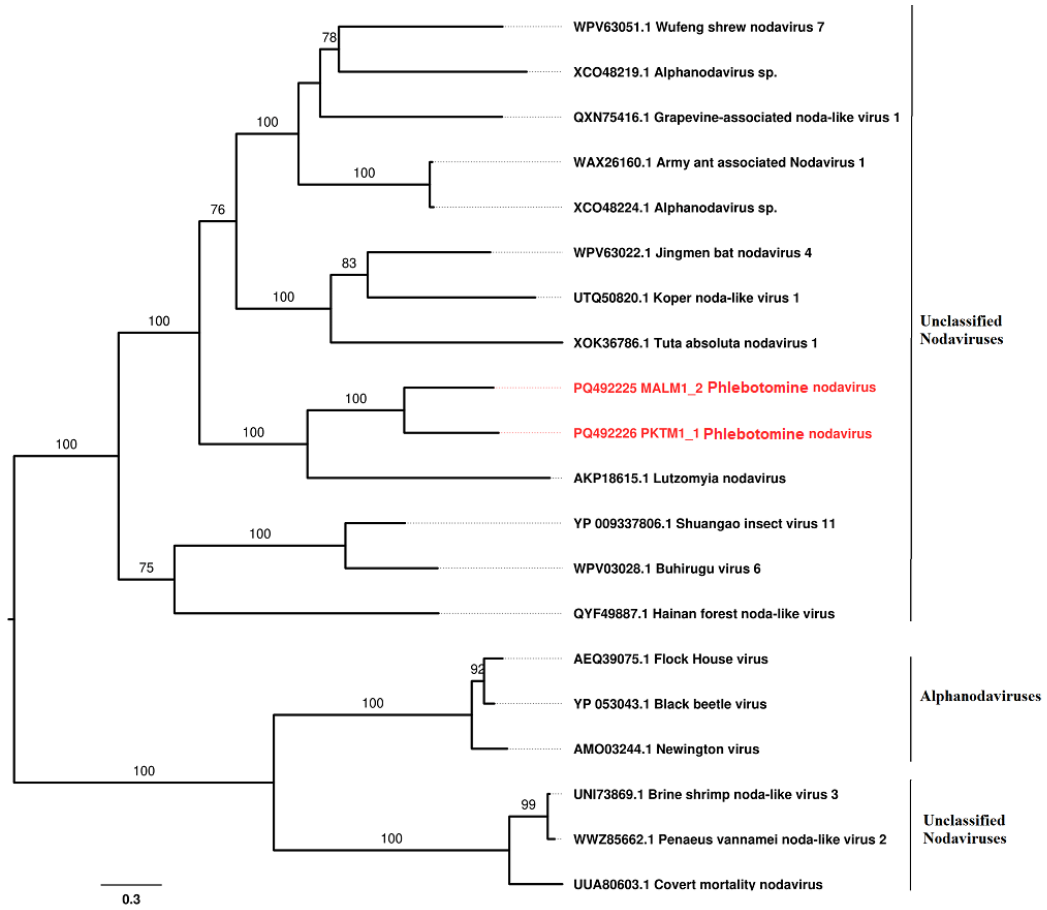


Figure 4.5: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Nodaviridae* RdRp Amino Acid Sequences

Legend

The tree includes sequences from classified alphanodaviruses, unclassified nodaviruses, and related noda-like viruses. Newly identified Phlebotomine nodavirus sequences from this study (MALM1, accession number PQ492225; PKTM4, accession number PQ492226) are highlighted in red and cluster within the unclassified Nodaviruses clade, closely related to Lutzomyia nodavirus (AKP18615.1). Bootstrap support values greater than 70% are shown at the nodes. The scale bar represents the number of amino acid substitutions per site. Genus-level groupings are indicated on the right.

4.2.6 Negevirus

A 554 bp contig corresponding to the RdRP of the *negevirus* taxon was detected in the BARM6 superpool. This contig showed moderate similarity to Utsjoki negevirus 1 (GenBank Accession no. ON949948.1), with an amino acid identity of 65.45% at the RdRp gene (Table 4.3). The virus was tentatively named Baringo negevirus. Maximum likelihood phylogenetic analysis based on RdRP sequences revealed that the Phlebotomine negevirus 1 (PQ492230 BARM6_1) clusters within the Nelorpiviruses clade, grouping with Big Cypress virus and other related nelorpiviruses (Figure 4.6). This placement, supported by high bootstrap values, confirms Phlebotomine negevirus 1 as a novel member of the Nelorpiviruses and expands the known diversity of this group.

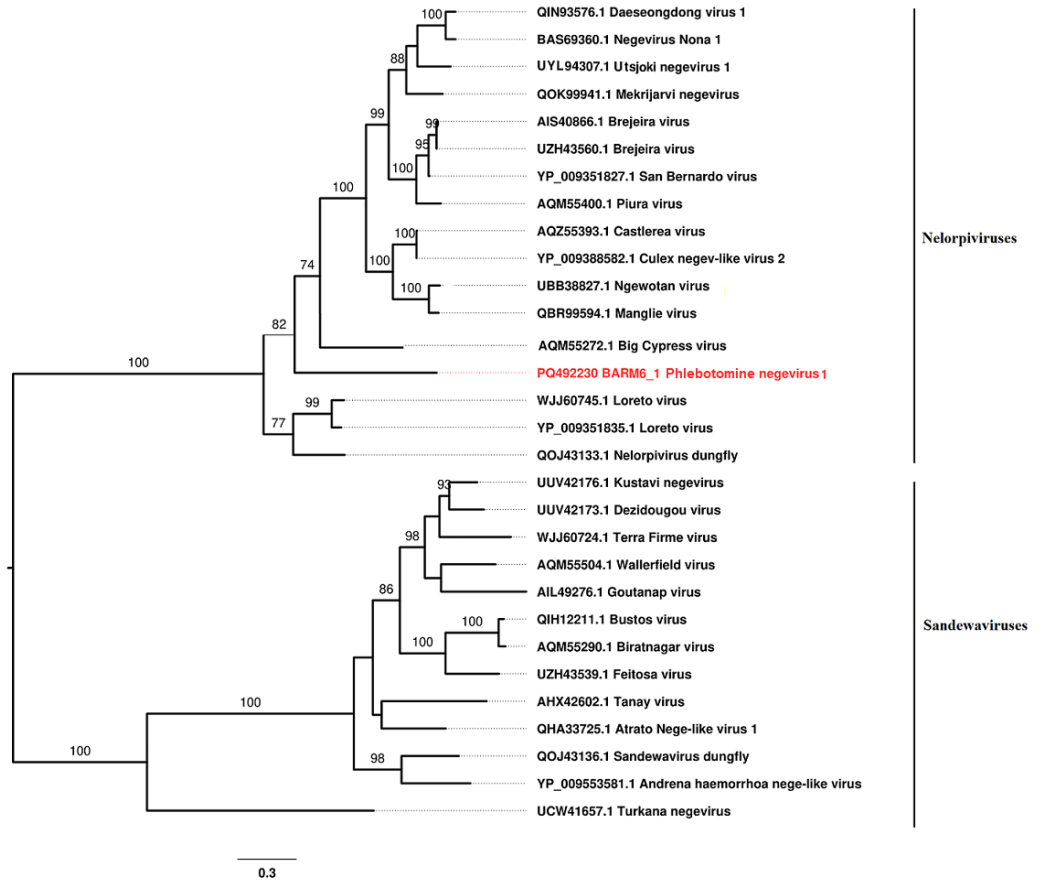


Figure 4.6: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative Negeviruses

The tree depicts the evolutionary relationships among selected *Negevirus* members. The newly identified Phlebotomine negevirus 1 (PQ492230 BARM6_1) from this study is highlighted in red and clusters within the Nelorpiviruses clade. Bootstrap support values greater than 70% are indicated at the nodes. The scale bar represents the number of amino acid substitutions per site. Genus-level groupings (Nelorpiviruses and Sandewaviruses) are indicated on the right.

4.2.7 *Iflaviridae* Family

Two linear and non-segmented genomes, each containing the entire coding region were obtained in BARM5 and KWAM1 samples. The BARM5 iflavirus genome is 9,922 nucleotides in length and encodes a single polyprotein of 2,900 amino acids. The ORF spans nucleotides 628 to 9330. Conserved domain analysis identified key functional motifs within the polyprotein, including a helicase domain located between nucleotides 5302 and 5619, an RdRp domain between nucleotides 8260 and 9225, and a capsid domain spanning amino acid residues 3658 to 4263 (Figure 4.3). This sequence showed the closest similarity to *Iflavirus sp.* with an amino acid identity of 49.57% (Table 4.3) and was tentatively named Phlebotomine iflavirus 1.

In the KWAM1 superpool from the Iseq sequencing, 6 contigs were obtained, totaling 5831nt distantly related to Sacbrood virus (43.52%). In the subsequent Miseq resequencing of the KWAM1 superpool genome of 8,874 nucleotides was recovered, here named Phlebotomine iflavirus 2 (Table 4.3). The Phlebotomine iflavirus 2 genome contains a single ORF spanning nucleotides 152–8800, encoding a polyprotein of 2,882 amino acid residues. Conserved domain analysis revealed the presence of key functional motifs, including a helicase domain located between nucleotides 4070–4390, an RdRp domain between nucleotides 7760–8716, and a peptidase 3C domain spanning residues 6518–7081 (Figure 4.2).

These genomes organization is consistent with typical iflaviruses, which encode a single polyprotein subsequently processed into structural and non-structural proteins. The Phlebotomine iflavirus 1 and Phlebotomine iflavirus 2 cluster within the broader *Iflaviridae* family, as shown in Figure 4.7. The Phlebotomine iflavirus 1 groups

closely with *Iflavirus sp*., while the Phlebotomine iflavirus 2 forms a distinct branch near Sacbrood virus. These placements confirm both viruses as novel members within the *Iflaviridae*, expanding the known diversity of this family.

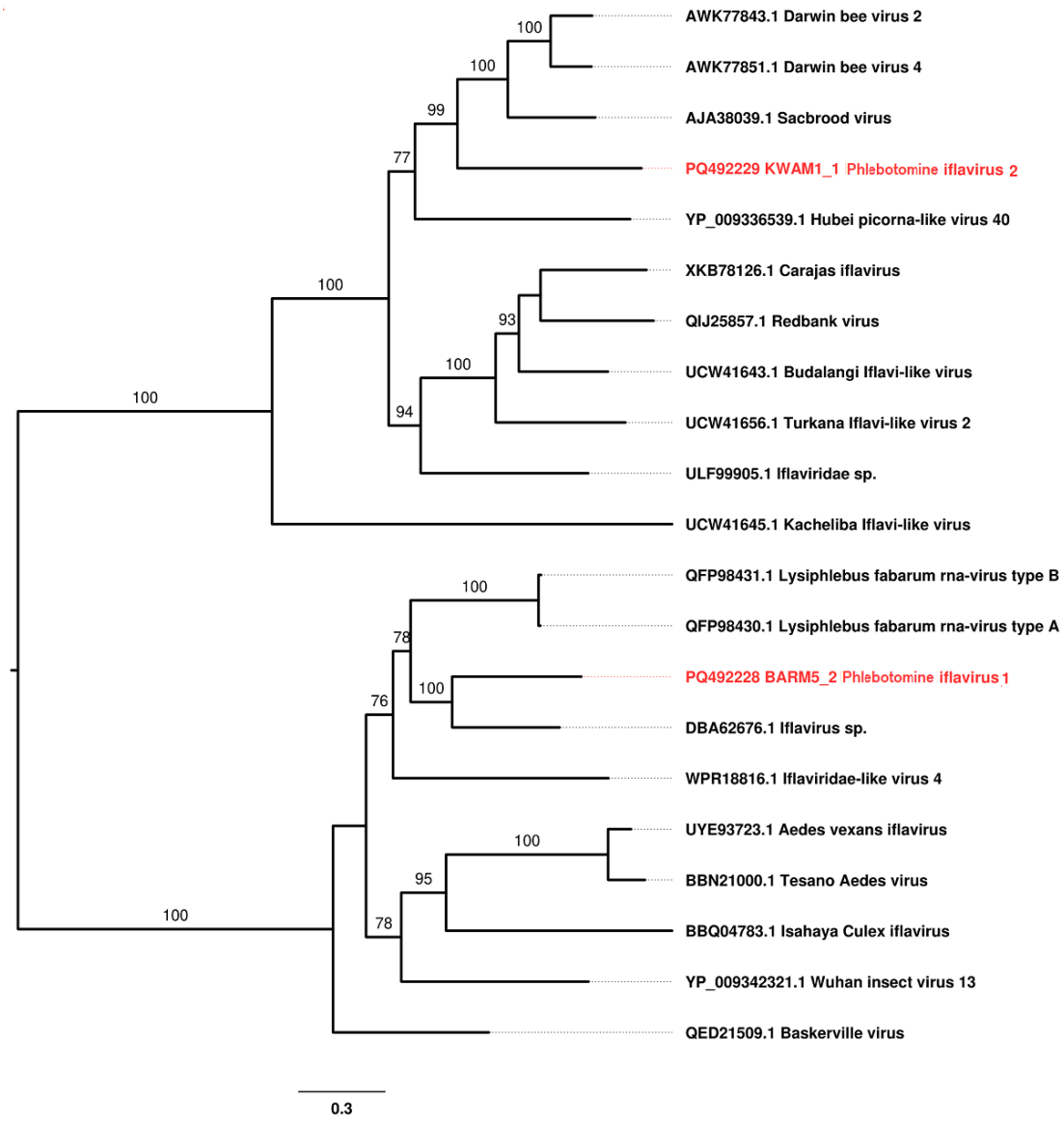


Figure 4.7: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Iflaviridae* Viruses

Legend

The tree illustrates the evolutionary relationships among selected members of the

Iflaviridae family, inferred from RdRp gene sequences (~2000 aa). Sequences obtained in this study Phlebotomine iflavirus 2 (PQ492229 KWAM1_1) and Phlebotomine iflavirus 1 (PQ492228 BARM5_2)-are highlighted in red. Bootstrap support values greater than 70% are shown at the nodes. The scale bar represents the number of amino acid substitutions per site.

4.2.8 *Chuviridae* Family

Contigs related to the family *Chuviridae* were recovered from GILM1 and MALM1 superpools. These contigs had 41.68% and 45.40%, amino acid identity respectively to RdRp of the Culex mosquito virus 4 (Table 4.3). Culex mosquito virus 4 (NC_076240.1) is a nonsegmented virus with a linear genome belonging to the *Culidacivirus* genus and was sequenced from *Culex tritaeniorhynchus* mosquitoes from the USA (Sadeghi et al., 2018). However, only partial sequences corresponding to the large segment of Culex mosquito virus 4 were obtained. This virus was tentatively named Phlebotomine chuvirus. As shown in Figure 4.8, the Phlebotomine chuvirus genomes (PQ492233 GILM1_1 and PQ492235 MALM1_3) cluster within the *Culidacivirus* genus, forming a sister clade to Culex mosquito virus 4, Wuhan mosquito virus 8 and Imjin River virus 1. The placement of these sequences within the Culidacivirus clade, supported by high bootstrap values, confirms their classification as novel members.

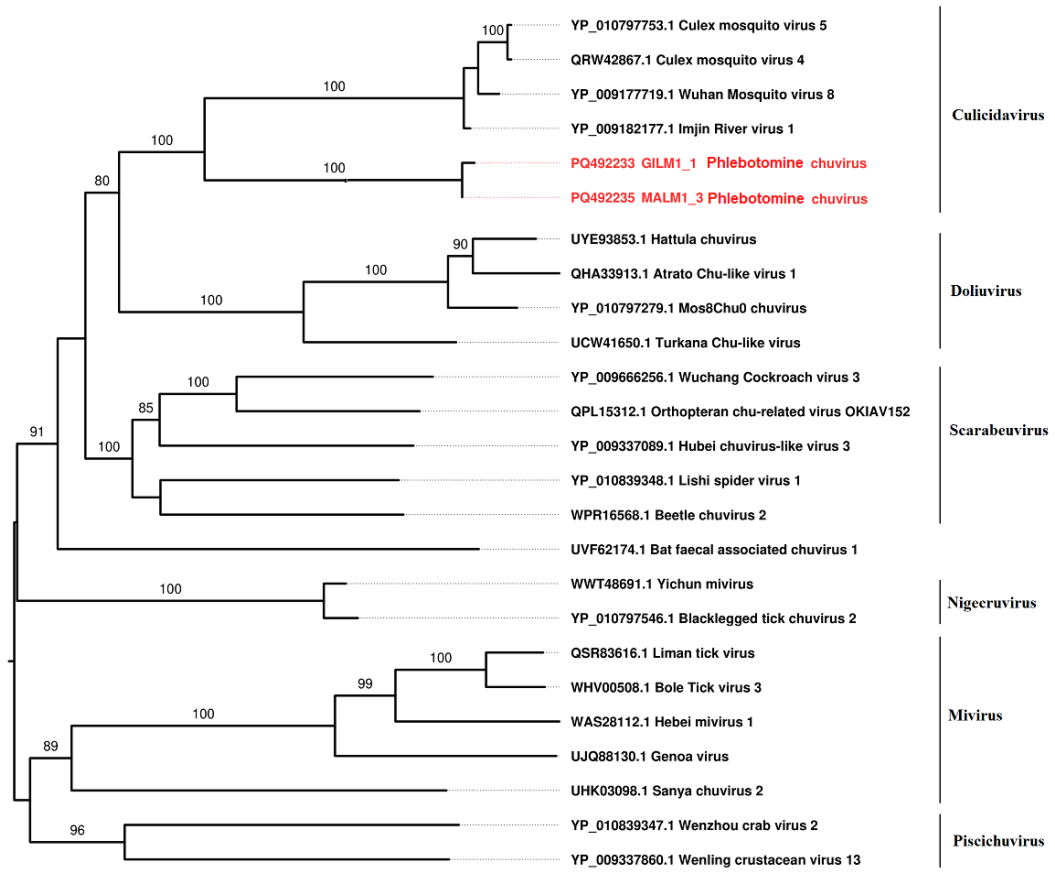


Figure 4.8: Mid-Point Phylogenetic Relationship of Phlebotomine Chuvirus with other Viruses

Legend

The tree depicts the evolutionary relationships among selected *Chuviridae* members based on RdRp gene sequence alignments (~800 aa). Sequences obtained in this study, Phlebotomine chuvirus (PQ492233 GILM1_1 and PQ492235 MALM1_3) are highlighted in red. Bootstrap support values greater than 70% are shown at the nodes. The scale bar represents the number of amino acid substitutions per site. Genus-level groupings are indicated on the right.

4.2.9 *Flaviviridae* Family

One read 40nt aligned to WN-KOUTV, belonging to the family *Flaviviridae*, genus *flavivirus* at 90% (Table 4.3) nucleotide similarity to WN-KOUTV strain AnD5443

with no other significant matches highlighted. This sequence was identified in the BARM1 superpool. Despite the short nucleotide sequence, the results were supported by the isolation of BARM1 WN-KOUTV from the individual pool BAR_S3_S008 genebank PP489328 and qRT-PCR assay with cycle threshold (Ct) values 36, as shown in Table 4.7.

4.2.10 *Phenuiviridae* Family

20 viral sequences read of the family *Phenuiviridae*, ranging between 90-150 nt (Table 4.3) representing the large, medium, and small segments, were detected in the MALM1 superpool. The identified sequences had the highest similarity to Bogoria virus (genus *Phlebovirus*) (98–100%), isolated from Phlebotomine sandflies in Kenya in 2015 (ON158112.1). Efforts to isolate the Malindi Bogoria were not feasible however, qRT-PCR assay further confirmed the existence of Ct values 38 as shown in Table 4.7.

4.2.11 *Rhabdoviridae* Family

From Iseq sequencing of the Kwale sample, 9 reads, ranging from 84 to 150 nt in length, were identified as vesiculoviruses (Table 4.3). Amino acid identity analysis linked these reads to various vesiculoviruses, including Chandipura virus and Vesicular Stomatitis New Jersey (Table 4.3). Unfortunately, it was not possible confirm the presence of this virus in the sample using qRT-PCR due to the lack of a universal vesiculoviruses primer. To increase sequencing depth, the RNA virus-enriched library was re-sequenced on the Illumina Miseq platform. This led to the discovery of 8 contigs totaling 4,759 bp, all annotated to the *Rhabdoviridae* family. These contigs included genes for RdRP, matrix, glycoprotein, and nucleoprotein. The largest contig, 1,385 bp in length, corresponded to the RdRP, with amino acid identity analysis showing its closest match to Chandipura virus (WCC64990.1) at 49.46% (Table 4.3). This virus was tentatively named Phlebotomine rhabdovirus. Phlebotomine rhabdovirus (PQ492232 KWAM1_2) clusters within the unclassified rhabdoviruses, forming a distinct branch separate from the established *Vesiculovirus* genus (Figure 4.9). Notably, Phlebotomine rhabdovirus is most closely related to

Wuhan Louse Fly Virus 11 (AJG39213.1), but remains on a separate lineage, indicating substantial evolutionary divergence. This phylogenetic position, together with the low amino acid identity to known vesiculoviruses (49.46% to Chandipura virus), suggests that Phlebotomine rhabdovirus represents a novel and highly divergent member of the *Rhabdoviridae* family.

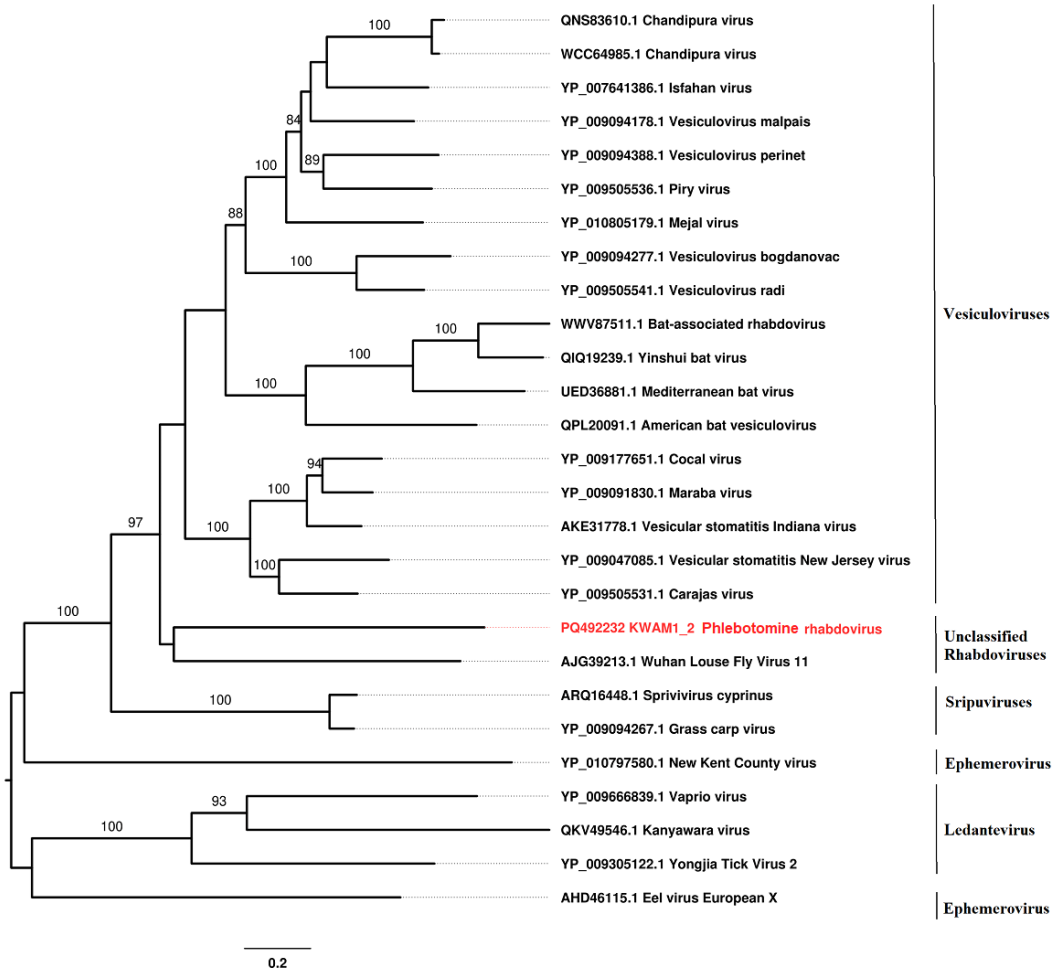


Figure 4.9: Midpoint-Rooted Maximum Likelihood Phylogenetic Tree of Representative *Rhabdoviridae* Viruses

The tree depicts the evolutionary relationships among selected members of the *Rhabdoviridae* family, inferred from RdRp protein sequences (~400 aa). The newly identified phlebotomine rhabdovirus (PQ492232 KWAM1_2) from this study is

highlighted in red. The virus clusters within the unclassified rhabdoviruses and is most closely related to Wuhan Louse Fly Virus 11 (AJG39213.1) but forms a distinct branch. Bootstrap support values greater than 70% are indicated at the nodes. Genus-level groupings are shown on the right. The scale bar represents the number of amino acid substitutions per site.

4.2.12 Metagenomics Diversity Statistical Analysis

Due to the limited number of viral reads obtained from sequencing (less than 1% of total reads, with filtered reads per sample ranging approximately from 447,000 to 1,110,000) (Table 4.2) (Figure 4.1), the effective sequencing depth for viral populations was insufficient to perform robust and reliable formal statistical diversity analyses such as alpha and beta diversity metrics. Robust statistical analyses of viral diversity require enough sequencing depths to accurately measure community richness and evenness (Roux et al., 2017).

Nonetheless, taxonomic profiling of the available metagenomic data revealed a broad spectrum of viral families, including several insect-specific viruses and medically important arboviruses. The detection of multiple distinct viral taxa, some of which are novel and exhibit low similarity to known viruses, provides valuable qualitative insights into the diversity of viruses circulating in Phlebotomine sandflies from the sampled regions.

4.3 Genetic Characterization of the Isolated Arboviruses

4.3.1 Virus Isolation

Pool BAR_S3_S008, which consisted of *Sergentomyia spp.* sandflies trapped in Salabani village (Lat/Lon N0.56187 E36.03038, altitude 3289 ft), and pool BAR_S5_230, which consisted of *Phlebotomus spp.* sandflies trapped in KARLO (Lat/Lon N0.47017 E35.99720, altitude 3361 ft), Baringo County, exhibited a clear cytopathic effect (CPE). The CPE was characterized by rounding of cells, alterations in cell morphology, and disruption of the monolayer, beginning on days 2 and 4 post-inoculation, respectively as shown in Figure 4.10. BAR_S3_S008 CPE was stably

serially passaged in Vero cells (CCL-81™ cells). BAR_S5_230 isolate did not produce CPE in passage 1, however, reproducible CPE was obtained in passage 2. The minimum infection rate of the collected Phlebotomine sandflies was determined to be 0.031% (2 out of 7,646 Phlebotomine sandflies tested).

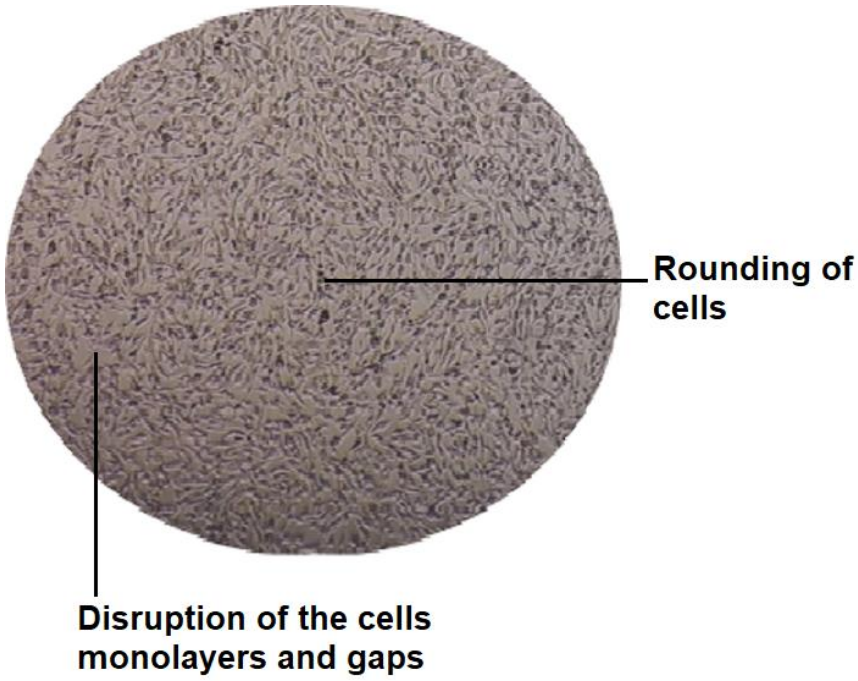


Figure 4.10: Cytopathic Effect Observed in a Cell Culture Monolayer

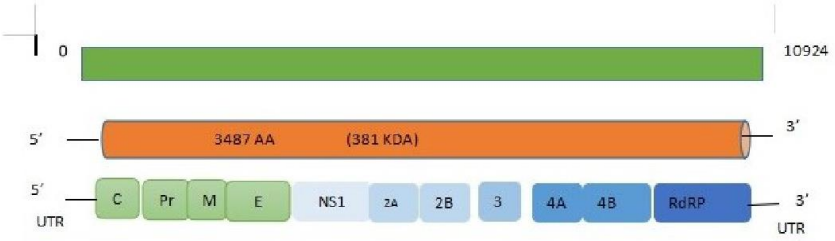
Legend

The image shows characteristic rounding of cells and disruption of the cell monolayer, resulting in gaps between cells. These morphological changes are indicative of viral infection and cell damage.

4.3.2 Genetic Organization of the Isolated Viruses

The complete genomes of BAR_S3_S008 and BAR_S5_230 isolates were determined. The BAR_S3_S008 isolate genome comprises 10,924 nucleotides (nt), with a G+C content of 50.6%. The genome contains a single open reading frame located between 30 and 10,343 nucleotides, encoding a 3437 amino acid (aa) long

polyprotein (Figure 4.11). BLAST analysis revealed that the genome belongs to a flavivirus of West Nile virus Koutango lineage (WN-KOUTV). Nucleotide similarity to available ORF sequences of other WN-KOUTV strains ranged between 90.11% - 99.25%. At the amino acid levels the percentage identity ranges from 97.63-99.71 to other WN-KOUTV strains as shown in Table 4.4. The entire genome had an average coverage depth of 3864x, indicating thorough sequencing coverage. (Appendix 111).



Protein	Amino Residues
Viral capsid	1-123
Pre-Membrane	124-290
Envelope	291-791
NOn-Structural 1	792-1143
Non-Structural 2A	1144-1374
Non -Structural 2B	1375-1505
Non-Structural 3	1506-2124
Non-Structural 4A	2125-2273
Non-Structural 4B	2274-2528
Non- Structural 5-RdRP	2529-3437

Figure 4.11: Genome Organization of the WN-KOUTV Strain BAR_S3_S008

Legend

The genome of the WN-KOUTV strain BAR_S3_S008 contains a single ORF that encodes a 381kilo-dalton polyprotein, 3437 amino acids in length, which is post translationally cleaved into three structural proteins and seven non-structural proteins.

Table 4.4: Percentage Identities of Nucleotide and Amino Acid Sequences Between BAR_S2_S008 WN-KOUTV and other WN-KOUTV Strains

Strain	Genbank Number	Origin	Genome (nt)	Year of Isolation	Host	Reference	BAR_S3_S008	
							nt %	aa %
PM148	MN057643.1	Niger	10,948	2016	Phlebotomine Sandflies	(Fall et al., 2021)	90.11	98.52
ArD96655	KY703855.1	Senegal	10,302	1993	*Rhipicephalus guilhoni*	(Fall et al., 2017)	90.15	98.23
AnD5443	EU082200.2	Senegal	10988	1968	Tatera Kempi	(Coz et al., 1976)	91.20	98.49
AnD95153	OP846972.1	Senegal	10,104	1993	Mastomys erythroleucus	(Fall et al., 2021)	89.98	97.63
BAR_SS4_020	ON158112.1	Kenya	9,826	2015	Phlebotomine Sandflies	(Koskei et al., 2024)	99.25	99.71

BAR_S5_230 viral isolate exhibited a tripartite genome organization comprising a large (6331 nucleotides), medium (4217 nucleotides), and small (1870 nucleotides) segment. The genome was identified as Bogoria Virus (BOGV), a phlebovirus (Figure 4.12).

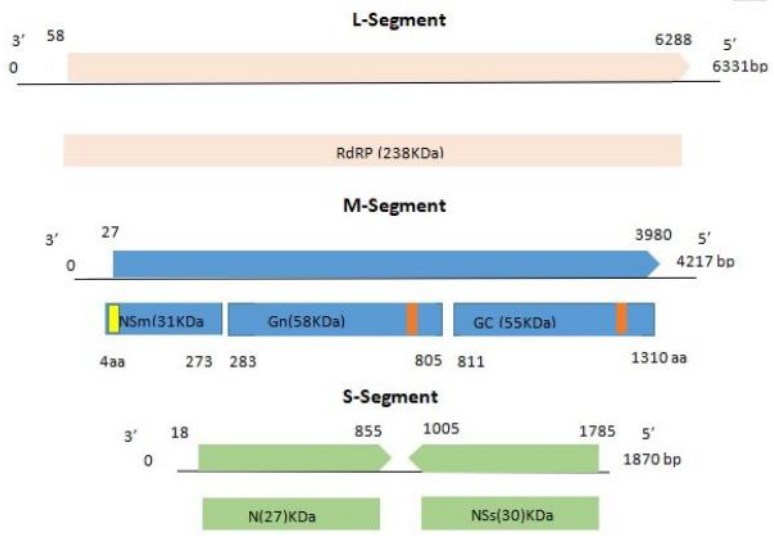


Figure 4.12: Genome Organization of the BAR_S5_230 Bogoria Virus

Legend

Bogoria virus strain BAR_S5_230 has a tripartite genome organization comprising a large (L), a medium (M), and a small (S) segment. The L segment has a single open

reading frame (ORF) of 6,231 nucleotides that encodes the RNA-dependent RNA polymerase (RdRp). The M segment also contains a single ORF, which is post translationally cleaved into the glycoproteins Gn and Gc, and the non-structural protein NSm. The S segment contains two ORFs in an antisense orientation that encode the nucleocapsid protein (N) and the nonstructural protein (NSs).

The average depth of coverage for the large, medium, and small segments was 114x, 128x, and 182x respectively as shown in Appendix IV. All three segments of the BOGV showed a high similarity to previously isolated BOGV, KCH_SS1_089 (GenBank Accession numbers: ON158113.1-ON158115.1) in West Pokot County and SP105-KE-2016 reported in Kapkuikui, Baringo County, with amino acid similarity of over 98% as shown in Table 4.5.

Table 4.5: Percentage Identities of Nucleotide and Amino Acid Sequences Between BAR_S5_230 Bogoria virus and other Bogoria Viruses

BAR_S5_230 BOGV	ON158114.1 BOGV		NC_078349.1 BOGV	
Protein	nt %	aa %	nt %	aa %
RdRp	97.95	99.86	88.58	99.42
Glycoprotein	97.75	99.01	96.59	98.94
Nonstructural protein (NSs)	98.33	98.84	98.46	99.61
Nucleocapsid	98.37	99.59	99.32	100

4.3.3 Amino Acid Variation among WN-KOUTV Strains

The amino acid analysis of WN-KOUTV lineage strains' open reading frame showed 76 out of 3437 (2.21%) variable sites among the strains. The NS2b gene is highly conserved without any variation in the amino acids (Figure 4.13). The WN-KOUTV BAR_S3_SOO8 strain showed 41 out of 76 (53.95%) amino acid variations in genes to the other WN-KOUTVs. Amino acid variations were found in the capsid gene 4 (9.75%), envelope gene 6 (14.63), NS1 gene 4 (9.75%), NS2a gene 4 (9.75%), NS3 gene 3 (7.31%), NS4a gene 3(7.31%), NS4b gene 10 (6 amino acid variation and 4 amino acid insertion accounting for 24.39%), and NS5 gene 7 (17.03%) (Figure 4.13). The insertion of the four amino acids (LNWE) at NS4b resulted in a 3437-

amino-acid-long polyprotein, compared to the 3433-amino-acid polyprotein observed in other WN-KOUTV strains. This characteristic is shared with a previous isolate from Kenya, ON158112.1.

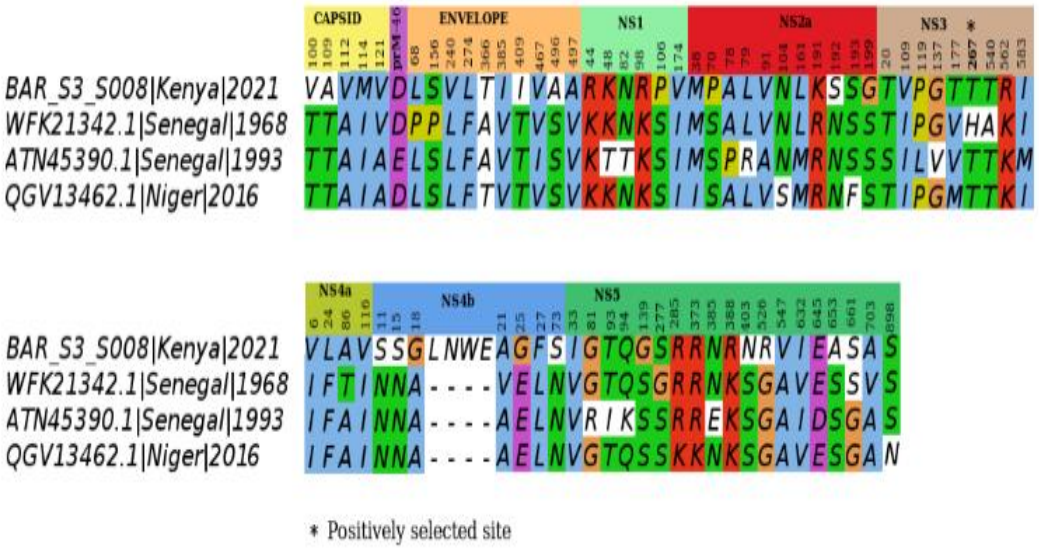


Figure 4.13: Amino Acids Variation among the WN-KOUTV Strains

Legend

Mutations characteristic of the WN-KOUTV lineage, in reference to the WNV NY99 strain, were identified in the WN-KOUTV BAR_S3_S008 sequence. These mutations, which are known molecular determinants of WNV virulence and replication, include: S72M in the pre-membrane protein, Y155F within the glycosylation site of the envelope protein, and a threonine (T) residue at position 249 of the NS3 protein (Beasley *et al.,* 2005; Langevin *et al.,* 2014; Maharaj *et al.,* 2019; Setoh *et al.,* 2012). However, in WN-KOUTV BAR_S3_S008, an alanine residue was observed at position 653 of the NS5 protein, whereas West African WN-KOUTV strains typically have a serine at this position, as shown in Figure 4.14.

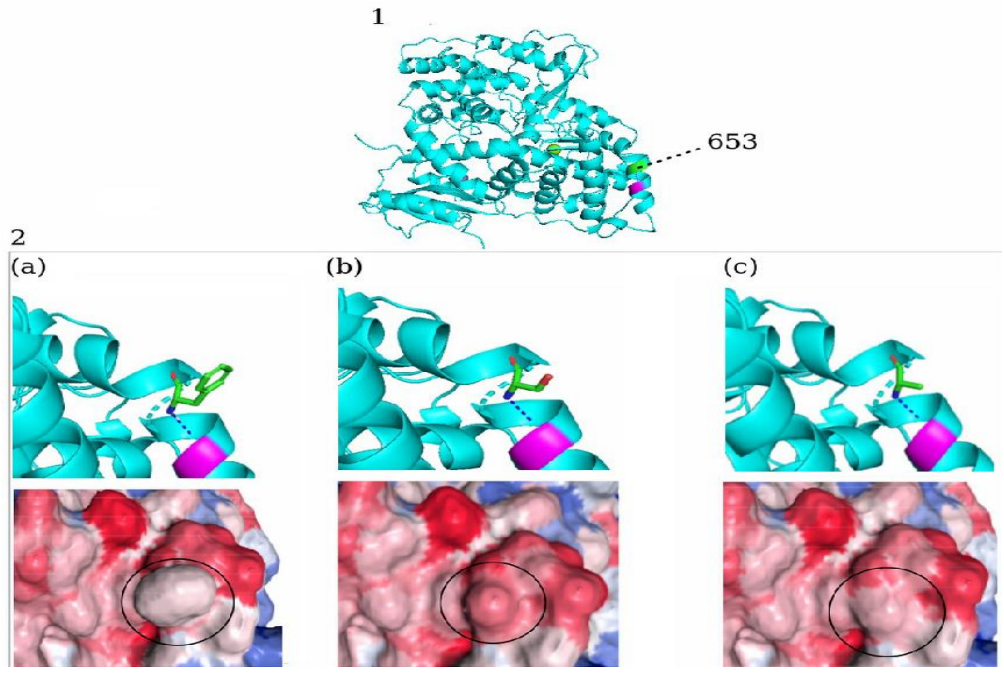


Figure 4.14: Amino Acid Variation at Position 653 within NS5 Protein

Legend

West Nile virus NS5 protein showing residue 653 colored in green and closely associated amino acid (Magenta) mapped on the WNV KUN RdRp structure (Protein Data Bank ID: 2HFZ) (Malet et al., 2007). 2. Close-up views of NS5 653: (a) Phenylalanine amino acid found in most WNV lineages strain, (b) Serine found in WN-KOUTVs from West Africa, and (c) Alanine found in WN-KOUTVs from

Kenya, with their corresponding surface structure showing the surface charge distributions. Positive and negative electrostatic potential are indicated in blue and red, respectively, while white color represents neutral. Protein structure was modeled in PyMOL Version 2.6.0 (Schrödinger, LLC). Electrostatic potential was calculated using the pdb2pqr server and Adaptive Poisson-Boltzmann Solver (APBS) (v1.5). Images were generated using the NGL viewer.

4.3.4 Positive Selection Pressure at Codon Site 1772 of the WN-KOUTV Lineage

A dataset consisting of four WN-KOUTV sequences was assessed for selection pressure. The overall ratio of non-synonymous to synonymous substitutions (dN/dS or ω) was determined to be 0.0216. Significant evidence of positive selection was observed at one site (1772-NS3:267), supported by three of the employed methods, namely, FUBAR, FEL, and MEME, as shown in Table 4.6.

Table 4.6: Selection Pressure Analysis Table Indicating Positively Selected Site 1772 (NS3:267) of the WN-KOUTV Lineage

West Nile Virus Dataset	Overall value ω	FUBAR	FEL	MEME
Lineage 7 (Koutango) n=4	0.0216	1772(NS3:267)	1772(NS3:267)	1772(NS3:267)

4.3.5 Phylogeny of the WN-KOUTV Isolate

Maximum likelihood phylogenetic analysis of the WN-KOUTV isolate, alongside representative strains from various West Nile Virus (WNV) lineages, revealed that the isolate forms a well-supported, single clade with ON158112.1, the WN-KOUTV isolate from Kenya in 2015. This cluster forms a sister clade to the West African WN-KOUTVs, as illustrated in Figure 4.15.

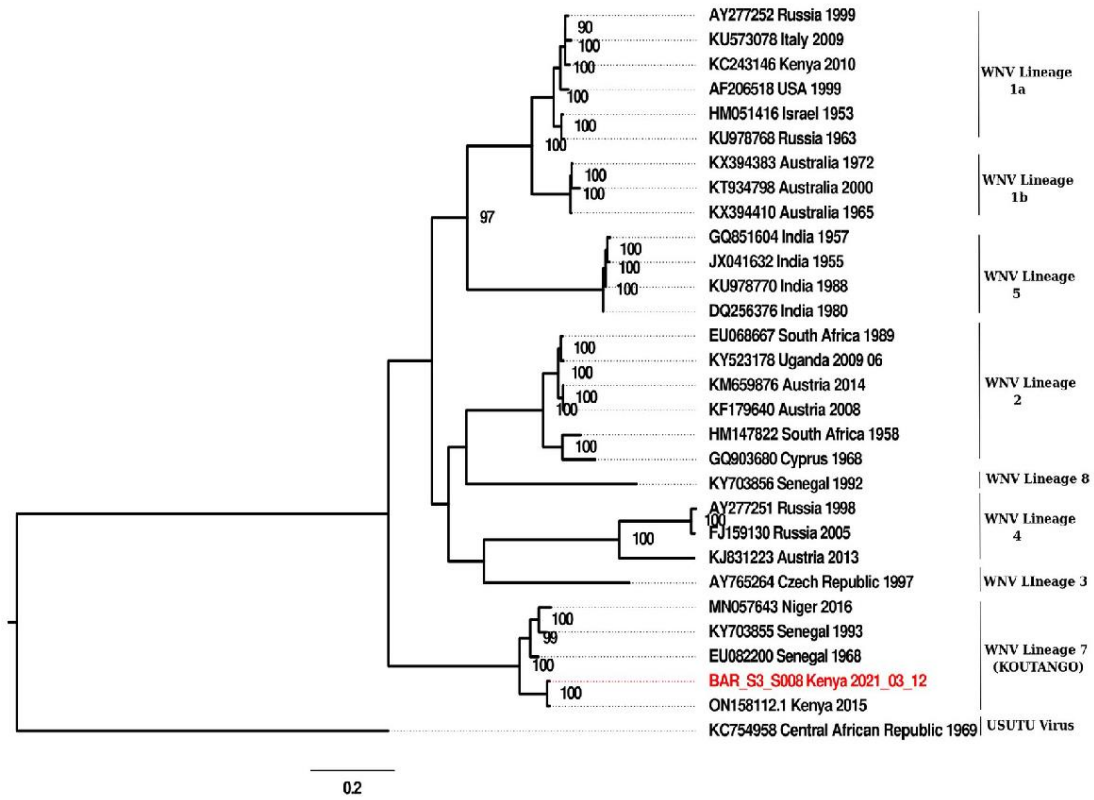


Figure 4.15: Maximum-Likelihood Phylogenetic Trees Showing the Relationship of the WN-KOUTV Strains BAR_S3_S008, with Select WNV Strains

Legend

The robustness of the tree was assessed using bootstrap analysis. Only bootstrap values >70% are displayed at important nodes. Usutu virus was considered as the out-group. The scale bar represents nucleotide substitutions per site.

4.3.6 The Most Common Recent Ancestor of the Isolated West Nile Koutango Virus Potentially Existed in Kenya in 1930

The molecular clock analysis estimated the time to the most recent common ancestor (tMRCA) of the WN-KOUTV isolate to be in the early 20th century (1930.91, 95% HPD: 1904.56, 1957.36) as shown in Figure 4.16 making this the most likely latest time point of introduction of this particular strain in Kenya. The estimated tMRCA of this lineage was dated back to 1885.39.

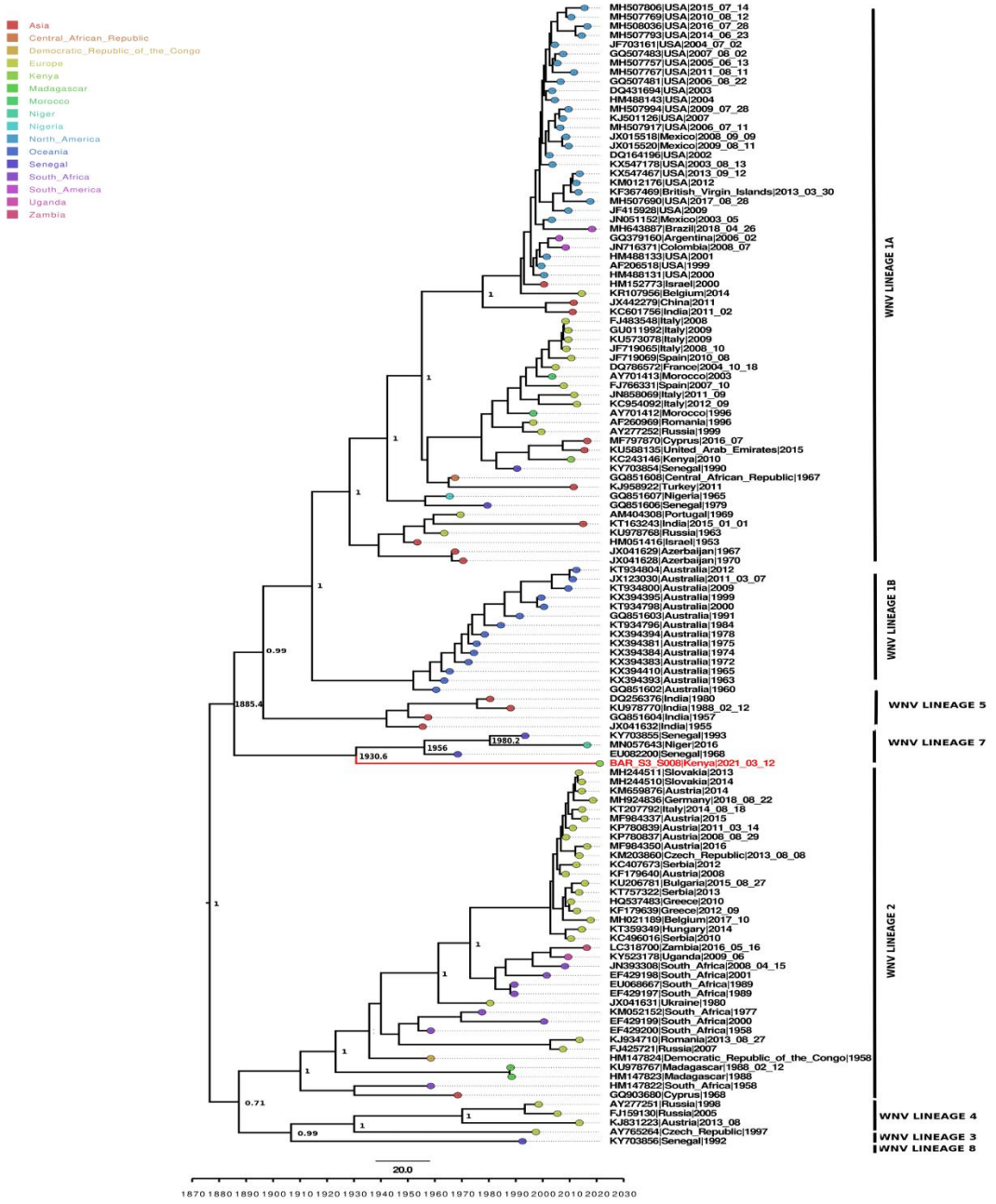


Figure 4.16: Time-calibrated Phylogeny of Different WNV Lineages from Different Countries

Legend

The WN-KOUTV BAR_S3_S008 strain is in red. Posterior probability > 0.7 are displayed and tree tip-nodes are inferred to geographical location of origin. Branches are scaled in years before 2021.28.

4.3.7 The Phylogeny of Isolated Bogoria Virus

The maximum likelihood phylogenetic analyses of the isolated Bogoria virus, based on four different viral proteins (A-D), consistently place our isolate (highlighted in red) within the Bogoria virus lineage with strong bootstrap support (Figure 4.17). However, the precise branching patterns and relationships with closely related viruses, including the Perkerra and Embossos viruses, vary across the different protein-based trees. These differences likely reflect protein-specific evolutionary histories or varying levels of phylogenetic signal. Despite these minor topological differences, all four analyses confirm the taxonomic classification of our isolates as Bogoria virus and illustrate their evolutionary relationships within this virus group.

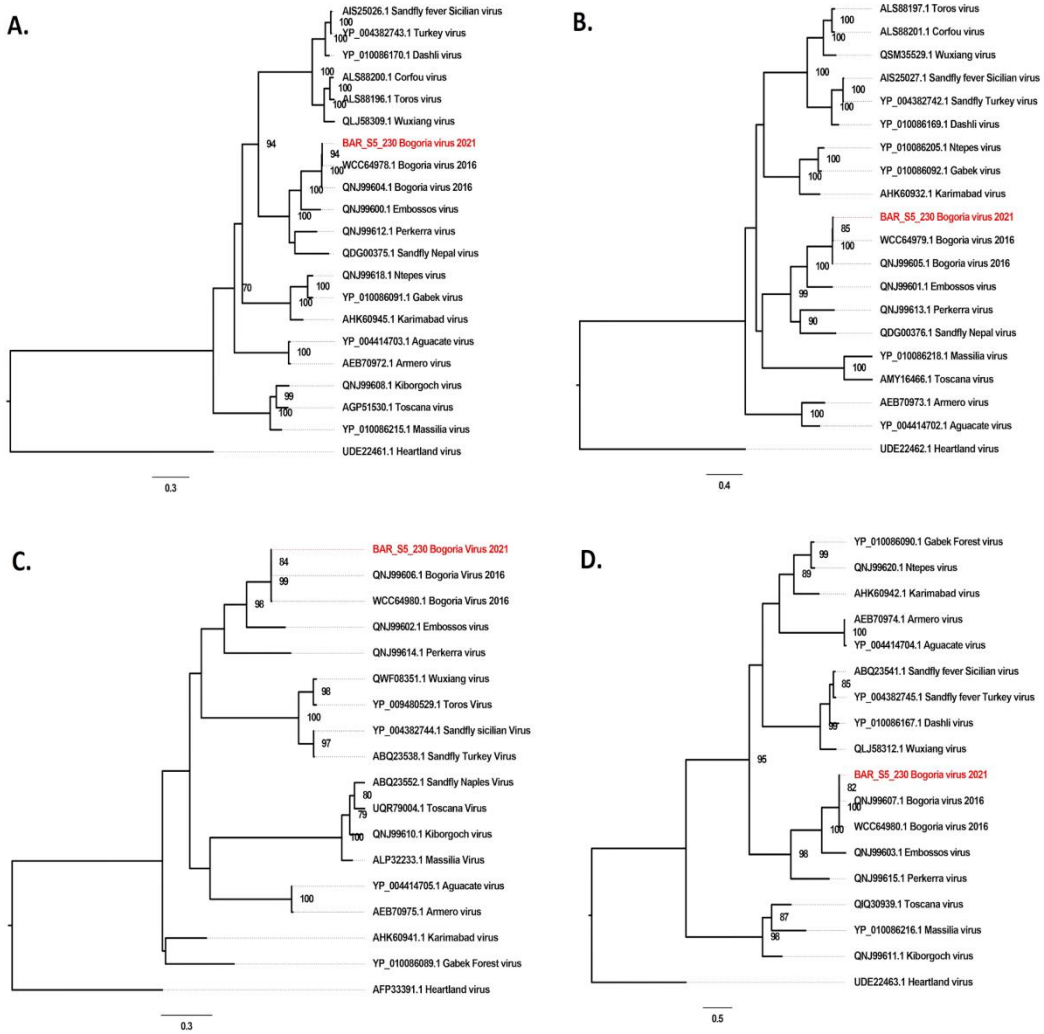


Figure 4.17: Phylogenetic Relationship of BAR_S5_230 with Select Phleboviruses

Legend

Maximum likelihood phylogeny tree based amino acid sequences. A. RNA dependent RNA polymerase; B. Glycoprotein precussor; C. Nucleocapsid protein;D. Small segment Non-structural protein. The trees demonstrate that the Bogoria virus isolates from this study consistently group within the Bogoria virus clade, with varying relationships to Perkerra virus, Embossos virus, and other related phleboviruses depending on the protein analyzed.

4.3.8 Confirmation of Identified Arboviruses using RT-PCR

The optimized multiplex RT-PCR assay effectively amplified and detected the target viral sequences, confirming the presence of WN-KOUTV and BOGV in the analyzed samples. The ct values ranged from 28 to 38, indicating varying viral loads across samples Table 4.7. No amplification was observed in the negative template control.

Table 4.7: RT-PCR Amplification Results Confirming Presence of WN-KOUTV and BOGV in Original Pools

Pool Number	Virus	Estimated Ct Value	Interpretation
BAR_S3_230	BOGV	28	Strong positive; high viral load
BAR_S3_S008	KOUTV	32	Positive; moderate viral load
BARM1	KOUTV	36	Positive; moderate viral load
MALM1	BOGV	38	Positive; moderate viral load
NTC	N/A	0	Negative control: no viral RNA detected

4.4 Taxonomic Classification of Phlebotomine Sandflies Species Associated with the Isolated Arboviruses

PCR amplification of the COI gene was performed on the WN-KOUTV-positive pool BAR_S3_S008 and the BOGV-positive pool BAR_S5_230. Successful amplification of the expected COI gene fragment (~700 bp) was confirmed by agarose gel electrophoresis (Appendix V). Clear bands corresponding to COI amplicons were observed in lanes B and C, representing the WN-KOUTV and BOGV positive pools, respectively. Lane A and lane D, the negative controls, showed no amplification, confirming the specificity of the PCR assay.

4.4.1 Distribution of Phlebotomine Sandflies Species in the WN-KOUTV Positive Pool

Sequence analysis of the COI gene of the sandflies from the WN-KOUTV-positive pool revealed that sandflies predominantly belonged to the genus *Sergentomyia*, accounting for 93.3% of the reads. The most abundant species identified was

Sergentomyia sp., comprising 49% of the reads. *Sergentomyia schwetzi* represented 20%, followed by *Sergentomyia inermis* with 11%, *Sergentomyia sp. SERG.UNK. F.1-8* with 7.30%, and *Sergentomyia squamipleuris* at 5%. *Phlebotominae* accounted for 6.7% of the reads as shown in Figure 4.18.

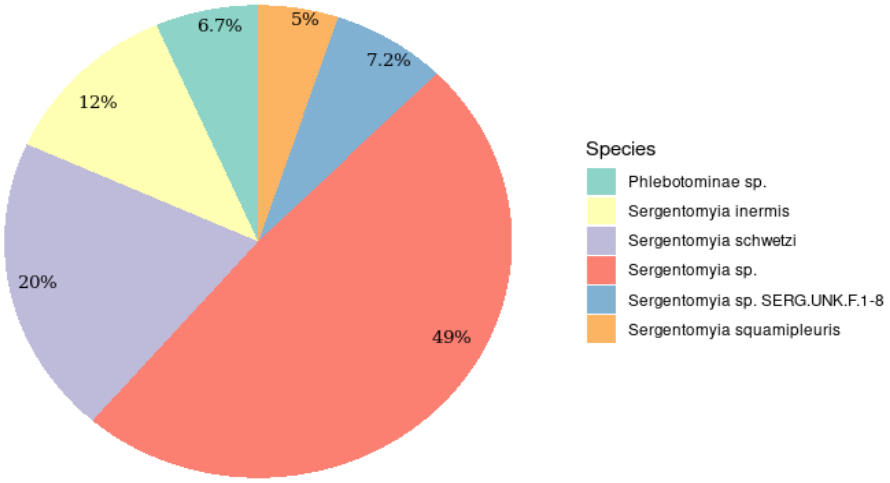


Figure 4.18: Distribution of Phlebotomine Sandflies Species in Pool BAR_S3_S008

4.4.2 Distribution of Phlebotomine Sandflies Species in the BOGV Positive Pool

COI gene sequence analysis for the BAR_S5_230 pool, which was BOGV positive, revealed that unclassified *Sergentomyia sp.* was the most abundant, representing 63.29% of the sequence reads. *Sergentomyia schwetzi* accounted for 18.35% of the reads, followed by unclassified *Phlebotominae sp.* at 12.66%, and *Sergentomyia inermis* at 5.70%, as shown in Figure 4.19.

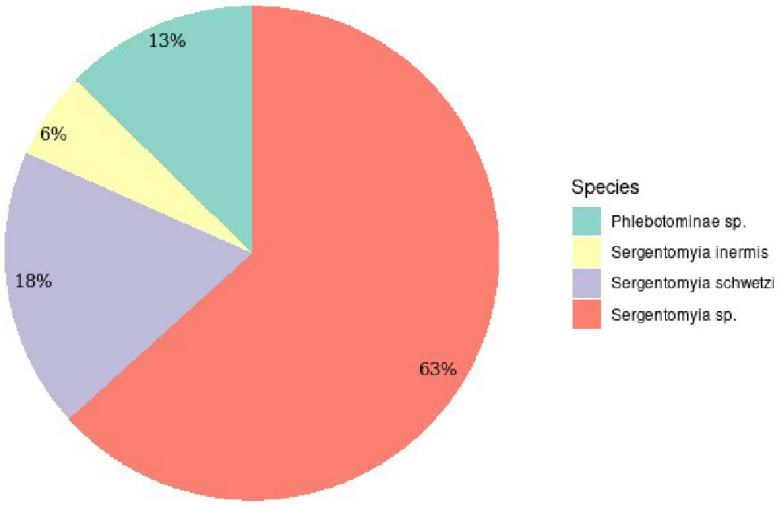


Figure 4.19: Distribution of Phlebotomine Sandflies Species in Pool BAR_S5_230

4.5 Data Availability

The raw sequence reads generated in this study are available at the NCBI Sequence Read Archive (SRA) database under BioProject accession PRJNA1139200. The sequence data has been assigned BioSample accession numbers SAMN42766516, SAMN42766517 and SAMN44294167- SAMN44294177. The sequences of viruses were deposited at Genbank under the accession codes: GenBank: PP105528-PP105530, PP489328, PQ492218-PQ492235.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Discussion

A comprehensive understanding of arboviruses and insect-specific viruses in hematophagous disease-transmitting vectors, as well as their associated vector species is crucial for elucidating the potential risk of arbovirus transmission to humans and animals (Vasilakis & Tesh, 2015; Viglietta et al., 2021). It is also pivotal for developing effective arbovirus and vector control strategies (Wilson et al., 2020). This study employed metagenomics to identify diverse insect-specific viruses and arboviruses, utilized virus isolation in cell culture and whole genome sequencing to characterize genomes from isolates, and determined the species of Phlebotomine sandflies associated with virus isolate pools using metabarcoding.

The metagenomic approach employed in this study allowed the characterization of several insect-specific RNA viruses and arboviruses in Phlebotomine sandflies from select regions in Kenya. In total, eight genomes of phylogenetically distinct insect-specific RNA viruses and sequence reads of related to three arboviruses were identified, significantly expanding the known virome of Phlebotomine sandflies. These viruses belong to diverse families, including positive-sense single-stranded RNA viruses from *Dicistroviridae*, *Iflaviridae*, *Flaviviridae*, Nodaviridae, and *Negevirus*, as well as negative-sense single-stranded RNA viruses from *Phenuiviridae*, *Chuviridae*, and *Rhabdoviridae*.

The metagenomic analysis revealed viral diversity across families, many of which represent novel or poorly characterized species, reflecting both geographical and intra-family variation. Notably, a group of picorna-like genomes provisionally designated asphlebotomine picorna-like virus are distantly related to Soybean thrips dicistrovirus 2 (Accession no. MT224137.1). Soybean thrips dicistrovirus 2 is an insect-specific member of the *Dicistroviridae,* originally identified in thrips, like many dicistroviruses it possesses a similar genome structure but remains poorly characterized regarding its biology and host interactions (Thekke-Veetil et al., 2020).

Importantly, these genomes feature two overlapping ORFs, deviating from the canonical dicistronic, non-overlapping genome organization that defines the family *Dicistroviridae* (Valles et al., 2017).

In contrast, Aphis gossypii virus from Baringo, exhibited high similarity to previously identified Aphis gossypii viruses. Aphis gossypii virus is a positive-sense single-stranded RNA virus classified within *Dicistroviridae* and was first reported in Aphis gossypii sampled in Israel from 2015 to 2017 (GenBank: MH476200 to MH476204), and has since been detected in other insect hosts such as *Culex pipiens* and honey bees (Marzoli et al., 2021). This virus is insect-specific and demonstrates the ability to replicate in both its aphid host and in honeybees, although its pathogenicity in non-aphid hosts remains to be fully elucidated (Marzoli et al., 2021). The detection of Aphis gossypii virus in Phlebotomine sand flies suggests these vectors can acquire and harbor viruses typically associated with other insect hosts, highlighting their potential role in the broader ecology and transmission of arthropod-borne viruses.

The distribution of these viruses varied, with some, like the Phlebotomine picorna-like virus, being widely distributed, while others, such as the Phlebotomine dicistrovirus and Phlebotomine iflavirus 2, were more region-specific. Other viruses are related to viruses previously identified in Phlebotomine sandflies, including Phlebotomine nodavirus and Phlebotomine rhabdovirus, suggesting possible host-specific adaptations. The detection of high-similarity sequence reads Bogoria virus, and West Nile Virus Koutango lineage highlights the presence of well-established viruses, while the identification of Phlebotomine rhabdovirus with lower similarity to Chandipura virus suggests the discovery of potentially new viral strains related to known arbovirus. This diversity illustrates the complex RNA virome in Phlebotomine sandflies.

Majority of the viruses identified were insect-specific viruses, while not directly pathogenic to vertebrates, they play a crucial role in shaping the viral ecology of vectors and may influence the transmission dynamics of arboviruses (Vasilakis & Tesh, 2015; H. Wang et al., 2024). Highly pathogenic insect viruses in the family

Dicistroviridae have been suggested for use as biological control agents to reduce populations of virus vectors (Bonning & Miller, 2010; Marti et al., 2015). Co-infections of mosquito cells with a negevirus and Chikungunya virus (CHIKV) demonstrated that negeviruses could inhibit the replication of CHIKV (Patterson *et al.,* 2020). Other studies have investigated on the effect of insect specific viruses on the propagation and transmission of arboviruses, highlighting the importance of the identification and characterization of the virome of hematophagous vectors in trying to understand insect-associated viral diseases (H. Wang et al., 2024).

In addition to insect-specific viruses mammalian-associated viruses were detected, including WN-KOUTV, BOGV, and the novel Kwale rhabdovirus distantly related to Chandipura virus. The detection of just a few reads of these viruses in comparison to insect viruses highlights the low abundances of RNA arboviruses within these complex RNA/DNA populations (Vasilakis & Gubler, 2016). This finding also highlights the capability of metagenomic techniques to detect and identify low-abundance viral sequences, some of which might otherwise go undetected in traditional screening methods (Mokili et al., 2012).

Negative control data suggest a low level of cross-contamination between samples during the library preparation as none of the identified viral genomes were present in the negative controls (Salter et al., 2014; Strong et al., 2014).

Using cell culture and next-generation sequencing, both WN-KOUTV and BOGV were successfully isolated, and their complete genomes were obtained from Phlebotomine sandflies sampled in Baringo, Kenya. WN-KOUTV has been mainly isolated from wild rodents, ticks and mosquitoes (Fall et al., 2017), and recently from Phlebotomine sandflies in Senegal in 2016 (Fall et al., 2021) and Baringo County, Kenya in 2015 (Koskei et al., 2024). Re-isolation of WN-KOUTV in this study in the same region after five years implies ongoing viral circulation, and highlights the potential involvement of Phlebotomine sandflies, in maintaining the virus within its enzootic transmission cycle.

Sequence comparison revealed that WN-KOUTV BAR_S3_S008 possesses mutations previously observed in West Africa WN-KOUTV strains: S72M in the pre-membrane, Y155F within the glycosylation site in the envelope, and a threonine (T) residue at position 249 of the NS3 protein (Fall et al., 2017, 2021; Prow et al., 2014). However, it exhibits F653A mutation in the NS5 protein, in place of the F653S mutation exhibited by West Africa WN-KOUTV strains. In the naturally attenuated WNV Kunjin subtype, site-specific mutagenesis studies at NS5 653, where serine is replaced by phenylalanine as in WNV NY99, increased sensitivity to type I interferon-mediated JAK-STAT signaling, revealing the role of site NS5 653 in interferon suppression, potentially enhancing the virus's virulence (Laurent-Rolle et al., 2010). Amino acid changes at this site represent a change in hydrophobicity and charge distribution. Therefore, further studies are needed to determine the implications of A653 in Kenya WN-KOUTV strains on interferon resistance.

This study observed a strong purifying pressure acting on the WN-KOUTV lineage, aligning with observations in other vector-borne RNA viruses. This strong purifying pressure is attributed to the interaction of the arbovirus with its dual hosts, the arthropod and the vertebrate, each equipped with distinct defense mechanisms (Woelk & Holmes, 2002). Nevertheless, there was significant diversifying pressure at the codon site 1772 of the polyprotein gene (position 267 of the NS3 protein). NS3-267 is located within the NS3 helicase domain, which is crucial to viral replication (Mastrangelo et al., 2007). However, no documentation has been attributed to this site in WNV. Therefore, further investigation may elucidate the precise functional implication of this variation.

Phylogenetic analysis of the WN-KOUTV BAR_S3_S008 revealed that the isolate forms a well-supported single clade with WN-KOUTV BAR_SS4_020 and a sister clade to the West African WN-KOUTVs. Molecular clock inference suggested that the most recent common ancestor of WN-KOUTV BAR_S3_S008 and the West African WN-KOUTVs dates to 1930 (95% HPD: 1904.56, 1957.36), indicating a possible introduction time of the WN-KOUTV in the region. The extended branch leading to the WN-KOUTV isolate underscores the historical continuity of the virus's

circulation and highlights potential gaps in surveillance data. This may be attributed to the limited attention given to sandfly-borne viruses until recently in the Sub-Saharan region. (Marklewitz et al., 2020; Tchouassi et al., 2019). Historically, surveillance efforts have predominantly focused on vectors such as mosquitoes and ticks (Junglen & Drosten, 2013; Philbert et al., 2022), potentially allowing the circulation of WN-KOUTV to go undetected for an extended period. Notably, WNV lineage 1a primarily transmitted by mosquitoes, has been documented through serological evidence in this region (Tigoi et al., 2015). Given that birds serve as the natural reservoir hosts for WNVs (Komar et al., 2003), and considering the documented adaptation of WN-KOUTV to avian cells (Varelas-Wesley & Calisher, 1982), there is a strong basis to explore their role in WN-KOUTV spread. Additionally, the observed phylogenetic and genetic relationship between West and East African WN-KOUTV further supports this connection. Intra-African migratory birds between these regions may have contributed to the movement and potential spread of the virus.

Phlebotomine sandflies in the WN-KOUTV-positive pool were morphologically identified as belonging to the genus *Sergentomyia*. This identification was further supported by metabarcoding analysis of the COI gene, which confirmed the presence of multiple Sergentomyia species within the pool. The species composition indicated a mixture of *Sergentomyia spp.,* including *Sergentomyia schwetzi*, *Sergentomyia inermis*, and other unclassified Sergentomyia species. Traditionally, *Sergentomyia spp* are known for their preference to feed on cold-blooded animals (Abbate et al., 2020; Latrofa et al., 2018). (Koskei et al., 2024)However, recent studies have revealed DNA from various vertebrates, including humans, in blood-fed females of different Sergentomyia species, suggesting a potential vector role (Azizi et al., 2016; Berdjane-Brouk et al., 2012; Tchouassi et al., 2019). This isolation also emphasizes the possible wider range of vectors of WN-KOUTV, with previous isolation's from mosquitoes and ticks in Senegal (Fall et al., 2017) and from sandflies in Niger in 2015 (Fall et al., 2021).

The re-isolation of WN-KOUTV in Phlebotomine sandflies highlights its persistence in these vectors, which is further exemplified by the re-isolation of Bogoria virus within the same ecological setting in this study. Previously, BOGV was isolated from Phlebotomine sandflies in West Pokot County (Koskei et al., 2024) and identified in Baringo County in Kapkuikui (Marklewitz et al., 2020). This re-isolation of BOGV in Baringo County over different years, coupled with the detection of BOGV sequence fragments in ticks from goats in Sandai and Logumgum (Ogola et al., 2023), indicates its widespread and persistent circulation. Sequence comparison revealed no significant genetic differences between the BOGV BAR_S5_230 and other Bogoria viruses, confirming their genetic similarity.

The sandfly pool that tested positive for BOGV was morphologically identified as Phlebotomus spp.; however, COI gene metabarcoding analysis revealed a predominance of Sergentomyia species in the sample. This discrepancy highlights the inherent limitations and sensitivities of both identification methods. Morphological identification relies on visible physical traits, which can be error-prone due to subtle differences and overlap among sandfly genera (Galati & Rodrigues, 2023; Ready, 2013). In contrast, COI metabarcoding provides more precise species-level resolution and can detect mixed-species samples (Andújar et al., 2018). The predominance of Sergentomyia reads suggests contamination by Sergentomyia debris, likely because the small Phlebotomus pool was derived from a larger collection dominated by Sergentomyia without prior washing. The high sensitivity of high-throughput sequencing can detect even minor contamination, explaining the divergence from morphological results. Previous *phlebotomus spp*. identifications in the region included *Phlebotomus martini* and *Phlebotomus dubosqci* (Anjili et al., 2011), suggesting potential vectors for further investigation

Efforts to isolate Bogoria virus from Malindi, which was identified through metagenomics, were unsuccessful. However, its presence in the MALM1 superpool was confirmed using qRT-PCR. This contrasts with the successful re-isolation of BOGV from sandflies in Baringo County within the same ecological setting in this study, as well as previous isolations from West Pokot County (Koskei et al., 2024).

The inability to isolate BOGV from Malindi may be attributed to several factors. First, in cell culture, the occurrence of a visible cytopathic effect is the primary indicator of viral infection, yet the absence of CPE does not necessarily rule out infection (Céspedes-Tenorio & Arias-Arias, 2023; Marklewitz et al., 2020). Notably, the Bogoria virus strain BAR_S5_230 did not induce CPE during the first passage, but reproducible CPE was observed in the second passage. This underscores a significant limitation of cell culture methods, where certain viruses may fail to replicate in the selected cell lines or under specific conditions, thereby escaping detection. Second, the retrospective nature of the study means there was a prolonged interval between sample collection and inoculation, increasing the likelihood of viral inactivation and reducing the chances of recovering infectious virus (Dangsagul et al., 2021).

Similarly, attempts to isolate phlebotomine rhabdovirus, which is distantly related to Chandipura virus (Koskei et al., 2024), were also unsuccessful, despite vesiculoviruses typically propagating readily in cell culture (Jadi et al., 2010). Moreover, due to the high genetic divergence among vesiculoviruses, it was not feasible to develop a consensus genome for primer design to confirm the presence of phlebotomine rhabdovirus, highlighting the challenges inherent in family-specific PCR approaches (Kuzmin et al., 2009; Walker et al., 2015). The successful generation of a larger contig after resequencing with MiSeq highlights the importance of sufficient coverage depth in metagenomic studies, particularly for detecting and assembling viral genomes that exhibit high genetic variability (Aguirre De Cárcer et al., 2014).

The application of these findings extends to several critical areas. The presence of insect-specific viruses, which may influence arbovirus transmission dynamics, offers insights into potential biological control measures, such as exploring certain *Dicistroviridae* family viruses for their potential in reducing vector populations and Negeviruses for interfering with arbovirus replication.

Though the pathogenicity of BOGV remains unknown, its consistent isolation in regions endemic with febrile illness warrants attention. In contrast, WN-KOUTV is

known for high virulence in animal models and has documented serological evidence and a reported human case with febrile symptoms in Africa (Fall et al., 2017, 2021; Shope, 2003). The presence of these viruses in vectors capable of biting humans warrants further investigations into their roles in febrile illnesses within these regions.

The results advocate enhanced and ongoing surveillance of Phlebotomine sandflies and the integration of metagenomics in arbovirus surveillance to facilitate the early detection of emerging viruses, providing essential data for timely intervention. Future research should focus on elucidating interactions between insect-specific viruses and arboviruses, as well as conducting vector competence and serological studies for WN-KOUTV and BOGV, to better understand their potential public health impacts and offer new insights into virus spread and persistence mechanisms.

5.2 Limitations of the Study

Although metagenomics proved effective for detecting potential arboviruses, as confirmed by RT-PCR and virus isolation, the statistical diversity analysis of the metagenomic data was constrained primarily by the low number of sequencing reads obtained from the samples. This limited sequencing depth prevented the reliable application of formal statistical diversity metrics, such as alpha and beta diversity, thereby restricting quantitative assessment of viral community richness and composition. Despite these limitations, taxonomic profiling of the available reads still enabled identification of multiple viral families and taxa. To overcome these challenges in future studies, increasing sequencing read depth-either by reducing the number of samples per sequencing run or by utilizing higher-capacity platforms, will be essential to improve sensitivity and accuracy in detecting arbovirus diversity, as supported by our resequencing results. Additionally, the high ribosomal RNA (rRNA) content in the samples and the use of retrospective sandfly samples collected over two years may have introduced biases in viral diversity and abundance may have reduced the sensitivity of virus detection; incorporating an rRNA depletion step could improve sensitivity in future studies.

Discrepancies between metabarcoding and morphological identification of Phlebotomine sandflies made it difficult to accurately associate viruses with specific sandfly species. This issue could be attributed to contamination with the DNA of more abundant species. Addressing these challenges in future research could improve the correlation between viral detection and host identification, leading to more reliable conclusions regarding virus-host relationships.

5.3 Conclusion

The metagenomic analysis revealed a broad spectrum of insect-specific viruses and arboviruses circulating in Phlebotomine sandflies, expanding the known virome associated with these vectors. Although the limited number of viral reads obtained precluded robust statistical diversity analyses such as alpha and beta diversity metrics, taxonomic profiling provided valuable qualitative insights into viral diversity, including the detection of novel viral taxa. This highlights the complex viral ecosystem within sandflies and underscores the sensitivity of metagenomics in detecting viruses present even at low abundance. These findings demonstrate the potential of metagenomic approaches for surveillance of sandfly-associated arboviruses, while also emphasizing the need for deeper sequencing coverage to enable comprehensive and statistically robust diversity assessments.

Virus isolation in cell culture provided direct evidence of actively circulating viruses and enabled genome characterization. The genomic characterization of WN-KOUTV and BOGV provided key insights into their genetic diversity and evolutionary history. The identified mutations in a key viral protein of WN-KOUTV may influence virulence and interfere with immune responses, such as interferon resistance. The re-isolation of WN-KOUTV and BOGV in the same region after several years suggests ongoing circulation and potential adaptation of this virus within the local Phlebotomine sandfly.

Metabarcoding helped narrow down the possible Phlebotomine sandfly hosts of these viruses. However, there remains a challenge in associating individual viruses with specific Phlebotomine sandfly species, as many species were identified per pool, and

morphological identification did not align with metabarcoding results, especially for the BOGV-positive pool. Therefore, further improvement of this method is necessary to address potential biases from library preparation, sequencing, and bioinformatic analysis. Given that high-throughput sequencing costs are continuing to decline, coupled with optimization of library preparation, metabarcoding might be the most effective solution for large-scale studies in the future. Nevertheless, reliable taxonomic backgrounds are still needed, and methods like co-occurrence networks could help infer the insect hosts associated with the identified viruses.

5.4 Recommendations

1. Implement combined metagenomics and virus isolation techniques for arbovirus surveillance, alongside morphological identification complemented by metabarcoding for accurate vector species identification. This integrated approach will enhance virus discovery and improve the effectiveness of vector control and disease prevention programs.
2. Conduct systematic and periodic monitoring of sandfly populations and their associated viromes to enable early detection of emerging viruses and to track temporal changes in viral diversity. This will support timely public health responses.
3. Use research findings to inform and update vector control strategies and public health policies, especially in regions with known arbovirus circulation, to reduce disease transmission risk.
4. Invest in capacity building by equipping laboratories and training personnel in virus detection, isolation, and genomic characterization techniques. Enhancing these capabilities will improve preparedness and response to potential arbovirus outbreaks

REFERENCES

Abbate, J. M., Maia, C., Pereira, A., Arfuso, F., Gaglio, G., Rizzo, M., Caracappa, G., Marino, G., Pollmeier, M., Giannetto, S., & Brianti, E. (2020). Identification of trypanosomatids and blood feeding preferences of Phlebotomine sand fly species common in Sicily, Southern Italy. *PLoS ONE*, *15*(3), e0229536. https://doi.org/10.1371/journal.pone.0229536

Aguiar, E. R. G. R., Olmo, R. P., Paro, S., Ferreira, F. V., de Faria, I. J. da S., Todjro, Y. M. H., Lobo, F. P., Kroon, E. G., Meignin, C., Gatherer, D., Imler, J.-L., & 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

Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., & Sereno, D. (2016). A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLOS Neglected Tropical Diseases*, *10*(3), e0004349. https://doi.org/10.1371/journal.pntd.0004349

Alexander, B., & Maroli, M. (2003). Control of Phlebotomine sandflies. *Medical and Veterinary Entomology*, *17*(1), 1–18. https://doi.org/10.1046/j.1365-2915.2003.00420.x

Alexander, B., & Maroli, M. (2003). Control of Phlebotomine sandflies. *Medical and Veterinary Entomology*, *17*(1), 1–18. https://doi.org/10.1046/j.1365-2915.2003.00420.x

Alkan, C., Bichaud, L., Lamballerie, X. D., Alten, B., Gould, E. A., & Charrel, R. N. (2013). Sandfly-borne phleboviruses of Eurasia and Africa: Epidemiology, genetic diversity, geographic range, control measures. *Antiviral Research*, *100*(1), 54–74. https://doi.org/10.1016/j.antiviral.2013.07.005

Alkan, C., Erisoz Kasap, O., Alten, B., de Lamballerie, X., & Charrel, R. N. (2016). Sandfly-Borne Phlebovirus Isolations from Turkey: New Insight into the Sandfly fever Sicilian and Sandfly fever Naples Species. *PLoS Neglected Tropical Diseases*, *10*(3), e0004519. https://doi.org/10.1371/journal.pntd.0004519

Alkan, C., Zapata, S., Bichaud, L., Moureau, G., Lemey, P., Firth, A. E., Gritsun, T. S., Gould, E. A., de Lamballerie, X., Depaquit, J., & Charrel, R. N. (2015). Ecuador Paraiso Escondido Virus, a New Flavivirus Isolated from New World Sand Flies in Ecuador, Is the First Representative of a Novel Clade in the Genus Flavivirus. *Journal of Virology*, *89*(23), 11773–11785. https://doi.org/10.1128/JVI.01543-15

Alten, B., Ozbel, Y., Ergunay, K., Kasap, O. E., Cull, B., Antoniou, M., Velo, E., Prudhomme, J., Molina, R., Banuls, A. L., Schaffner, F., Hendrickx, G., Van Bortel, W., & Medlock, J. M. (2015). Sampling strategies for Phlebotomine sand flies (Diptera: Psychodidae) in Europe. *Bulletin of Entomological Research*, *105*(6), 664–678. https://doi.org/10.1017/S0007485315000127

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anjili, C., Ngumbi, P. M., & Cirindi, J. (2011a). The Phlebotomine sandfly fauna (Diptera: Psychodidae) of Kenya. *Article in Journal of Vector Borne Diseases*, *48*, 183–189.

Anjili, C., Ngumbi, P. M., & Cirindi, J. (2011b). The Phlebotomine sandfly fauna (Diptera: Psychodidae) of Kenya Effects of glucocorticoids in leishmania View project Leishmania View project. *Article in Journal of Vector Borne Diseases*. https://www.researchgate.net/publication/221798331

Asuquo, D., Attai, K., Obot, O., Ekpenyong, M., Akwaowo, C., Arnold, K., & Uzoka, F.-M. (2024). Febrile disease modeling and diagnosis system for

optimizing medical decisions in resource-scarce settings. *Clinical eHealth, 7*, 52-76. https://doi.org/10.1016/j.ceh.2024.05.001

Ayhan, N., & Charrel, R. N. (2017). Of Phlebotomines (sandflies) and viruses: A comprehensive perspective on a complex situation. *Current Opinion in Insect Science*, *22*, 117–124. https://doi.org/10.1016/J.COIS.2017.05.019

Ayhan, N., & Charrel, R. N. (2018). Sandfly-Borne Viruses of Demonstrated/Relevant Medical Importance. *Vectors and Vector-Borne Zoonotic Diseases*. https://doi.org/10.5772/INTECHOPEN.81023

Azizi, K., Askari, M. B., Kalantari, M., & Moemenbellah-Fard, M. D. (2016). Molecular detection of Leishmania parasites and host blood meal identification in wild sand flies from a new endemic rural region, south of Iran. *Pathogens and Global Health*, *110*(7–8), 303–309. https://doi.org/10.1080/20477724.2016.1253530

Balaska, S., Fotakis, E. A., Chaskopoulou, A., & Vontas, J. (2021). Chemical control and insecticide resistance status of sand fly vectors worldwide. *PLOS Neglected Tropical Diseases*, *15*(8), e0009586. https://doi.org/10.1371/journal.pntd.0009586

Beasley, D. W. C., Whiteman, M. C., Zhang, S., Huang, C. Y.-H., Schneider, B. S., Smith, D. R., Gromowski, G. D., Higgs, S., Kinney, R. M., & Barrett, A. D. T. (2005). Envelope Protein Glycosylation Status Influences Mouse Neuroinvasion Phenotype of Genetic Lineage 1 West Nile Virus Strains. *Journal of Virology*, *79*(13), 8339–8347. https://doi.org/10.1128/JVI.79.13.8339-8347.2005

Benallal, K. E., Garni, R., Harratfft, Z., Benallal, K. E., Volf, P., & Dvorak, V. (2022). Phlebotomine sand flies (Diptera: Psychodidae) of the Maghreb region: A systematic review of distribution, morphology, and role in the transmission of the pathogens. *PLoS Neglected Tropical Diseases*, *16*(1), e0009952. https://doi.org/10.1371/journal.pntd.0009952

Berdjane-Brouk, Z., Koné, A. K., Djimdé, A. A., Charrel, R. N., Ravel, C., Delaunay, P., Giudice, P. del, Diarra, A. Z., Doumbo, S., Goita, S., Thera, M. A., Depaquit, J., Marty, P., Doumbo, O. K., & Izri, A. (2012). First Detection of Leishmania major DNA in Sergentomyia (Spelaeomyia) darlingi from Cutaneous Leishmaniasis Foci in Mali. *PLOS ONE*, *7*(1), e28266. https://doi.org/10.1371/journal.pone.0028266

Bolger, A. M., Lohse, M., & Usadel, B. (2014). *Trimmomatic: A flexible trimmer for Illumina sequence data | Bioinformatics | Oxford Academic*. *30*(15), 2114–2120.

Bonning, B. C., & Miller, W. A. (2010). Dicistroviruses. *Annual Review of Entomology*, *55*(Volume 55, 2010), 129–150. https://doi.org/10.1146/annurev-ento-112408-085457

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N. D., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Plessis, L. du, Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Butenko, A. M. (1996). [Arbovirus circulation in the Republic of Guinea]. *Meditsinskaia parazitologiia i parazitarnye bolezni*, *2*, 40–45.

Butenko, A. M., Semashko, I. V., Skvortsova, T. M., Gromashevskiĭ, V. L., & Kondrashina, N. G. (1986). [Detection of the Koutango virus (Flavivirus, Togaviridae) in Somalia]. *Meditsinskaia Parazitologiia I Parazitarnye Bolezni*, *3*, 65–68.

Chadha, M. S., Arankalle, V. A., Jadi, R. S., Joshi, M. V., Thakare, J. P., Mahadev, P. V. M., & Mishra, A. C. (2005). An outbreak of Chandipura virus encephalitis in the eastern districts of Gujarat State, India. *American Journal of Tropical Medicine and Hygiene*, *73*(3), 566–570. https://doi.org/10.4269/ajtmh.2005.73.566

Charrel, R. N., Brault, A. C., Gallian, P., Lemasson, J.-J., Murgue, B., Murri, S., Pastorino, B., Zeller, H., de Chesse, R., de Micco, P., & de Lamballerie, X. (2003). Evolutionary relationship between Old World West Nile virus strains: Evidence for viral gene flow between africa, the middle east, and europe. *Virology*, *315*(2), 381–388. https://doi.org/10.1016/S0042-6822(03)00536-1

Charrel, R. N., Gallian, P., Navarro-Marí, J.-M., Nicoletti, L., Papa, A., Sánchez-Seco, M. P., Tenorio, A., & de Lamballerie, X. (2005). Emergence of Toscana Virus in Europe. *Emerging Infectious Diseases*, *11*(11), 1657–1663. https://doi.org/10.3201/eid1111.050869

Cholleti, H., Berg, M., Hayer, J., & Blomström, A. L. (2018). Vector-borne viruses and their detection by viral metagenomics. *Infection Ecology and Epidemiology*, *8*(1). https://doi.org/10.1080/20008686.2018.1553465

Cochran, W.G. (1977) Sampling Techniques. 3rd Edition, John Wiley & Sons, New York

Coz, J., Le Gonidec, G., Cornet, M., Valade, M., Lemoine, M., & Gueve, A. (1975). Transmission experimentale d'un arbovirus du groupe b le virus Koutango par Aedes Aegypti L. *Pascal and Francis*.

Coz, J., Valade, M., Cornet, M., & Robin, Y. (1976). Transovarian transmission of a Flavivirus, the Koutango virus, in Aedes aegypti L. *Comptes rendus hebdomadaires des seances de l'Academie des sciences Serie D: Sciences naturelles*, *283*(1), 109–110.

De Carvalho, M. S., De Lara Pinto, A. Z., Pinheiro, A., Rodrigues, J. S. V., Melo, F. L., Da Silva, L. A., Ribeiro, B. M., & Dezengrini-Slhessarenko, R. (2018). Viola phlebovirus is a novel Phlebotomus fever serogroup member identified in Lutzomyia (Lutzomyia) longipalpis from Brazilian Pantanal. *Parasites and Vectors*, *11*(1), 1–10. https://doi.org/10.1186/s13071-018-2985-3

De Souza Pinto, I., Das Chagas, B. D., Rodrigues, A. A. F., Ferreira, A. L., Rezende, H. R., Bruno, R. V., Falqueto, A., Andrade-Filho, J. D., Galati, E. A. B., Shimabukuro, P. H. F., Brazil, R. P., & Peixoto, A. A. (2015). DNA barcoding of neotropical sand flies (*Diptera, Psychodidae*, Phlebotominae): Species identification and discovery within Brazil. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140636

Dvorak, V., Kasap, O. E., Ivovic, V., Mikov, O., Stefanovska, J., Martinkovic, F., Omeragic, J., Pajovic, I., Baymak, D., Oguz, G., Hlavackova, K., Gresova, M., Gunay, F., Vaselek, S., Ayhan, N., Lestinova, T., Cvetkovikj, A., Soldo, D. K., Katerinova, I., … Alten, B. (2020). Sand flies (Diptera: Psychodidae) in eight Balkan countries: Historical review and region-wide entomological survey. *Parasites and Vectors*, *13*(1), 1–15. https://doi.org/10.1186/s13071-020-04448-w

Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nature Reviews Microbiology*, *3*(6), 504–510. https://doi.org/10.1038/nrmicro1163

Elrefaey, A. M., Abdelnabi, R., Rosas, A. L. R., Wang, L., Basu, S., & Delang, L. (2020). Understanding the Mechanisms Underlying Host Restriction of Insect-Specific Viruses. *Viruses*, *12*(9), 1–19. https://doi.org/10.3390/v12090964

Ergunay, K., Ayhan, N., & Charrel, R. N. (2017). Novel and emergent sandfly-borne phleboviruses in Asia Minor: A systematic review. *Reviews in Medical Virology*, *27*(2), e1898. https://doi.org/10.1002/rmv.1898

Ergunay, K., Kaplan, B., Okar, S., Akkutay-Yoldar, Z., Kurne, A., Arsava, E. M., & Ozkul, A. (2015). Urinary detection of toscana virus nucleic acids in neuroinvasive infections. *Journal of Clinical Virology*, *70*, 89–92. https://doi.org/10.1016/j.jcv.2015.07.297

Fall, G., Diallo, D., Soumaila, H., Ndiaye, E. H., Lagare, A., Sadio, B. D., Ndione, M. H. D., Wiley, M., Dia, M., Diop, M., Ba, A., Sidikou, F., Ngoy, B. B., Faye, O., Testa, J., Loucoubar, C., Sall, A. A., Diallo, M., & Faye, O. (2021). First Detection of the West Nile Virus Koutango Lineage in Sandflies in Niger. *Pathogens*, *10*(3), 257. https://doi.org/10.3390/pathogens10030257

Fall, G., Diallo, M., Loucoubar, C., Faye, O., & Sall, A. A. (2014). Vector Competence of Culex neavei and Culex quinquefasciatus (Diptera: Culicidae) from Senegal for Lineages 1, 2, Koutango and a Putative New Lineage of West Nile virus. *The American Journal of Tropical Medicine and Hygiene*, *90*(4), 747–754. https://doi.org/10.4269/ajtmh.13-0405

Fall, G., Paola, N. D., Faye, M., Dia, M., Freire, C. C. de M., Loucoubar, C., Zanotto, P. M. de A., Faye, O., & Sall, A. A. (2017). Biological and phylogenetic characteristics of West African lineages of West Nile virus. *PLOS Neglected Tropical Diseases*, *11*(11), e0006078. https://doi.org/10.1371/journal.pntd.0006078

Fan, J., Huang, S., & Chorlton, S. D. (2021). BugSeq: A highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinformatics*, *22*(1), 160. https://doi.org/10.1186/s12859-021-04089-5

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using bootstraps. https://academic.oup.com/evolut/article/39/4/783/6872498

Galati, E. A. B., & Rodrigues, B. L. (2023). A Review of Historical Phlebotominae Taxonomy (Diptera: Psychodidae). *Neotropical Entomology*, *52*(4), 539–559. https://doi.org/10.1007/s13744-023-01030-8

Goh, V. S. L., Mok, C. K., & Chu, J. J. H. (2020). Antiviral Natural Products for Arbovirus Infections. *Molecules*, *25*(12). https://doi.org/10.3390/MOLECULES25122796

Hall, T. A. (1999). BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic Acids Symposium Series*.

Hang, J., Forshey, B. M., Kochel, T. J., Li, T., Solórzano, V. F., Halsey, E. S., & Kuschner, R. A. (2012). Random Amplification and Pyrosequencing for Identification of Novel Viral Genome Sequences. *Journal of Biomolecular Techniques : JBT*, *23*(1), 4. https://doi.org/10.7171/JBT.12-2301-001

Hebert, P. D. N., Ratnasingham, S., & DeWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(SUPPL. 1). https://doi.org/10.1098/rsbl.2003.0025

Huemer, H., Prudhomme, J., Amaro, F., Baklouti, A., Walder, G., Alten, B., Moutailler, S., Ergunay, K., Charrel, R. N., & Ayhan, N. (2017). Practical Guidelines for Studies on Sandfly-Borne Phleboviruses: Part II: Important Points to Consider for Fieldwork and Subsequent Virological Screening. *Vector-Borne and Zoonotic Diseases*, *17*(1), 81–90. https://doi.org/10.1089/vbz.2016.1965

Izri, A., Temmam, S., Moureau, G., Hamrioui, B., De Lamballerie, X., & Charrel, R. N. (2008). Sandfly fever sicilian virus, Algeria. *Emerging Infectious Diseases*, *14*(5). https://doi.org/10.3201/eid1405.071487

Jaafar, F. M., Belhouchet, M., Belaganahalli, M., Tesh, R. B., Mertens, P. P. C., & Attoui, H. (2014). Full-genome characterisation of orungo, lebombo and changuinola viruses provides evidence for co-evolution of orbiviruses with their arthropod vectors. *PLoS ONE*, *9*(1), e86392. https://doi.org/10.1371/journal.pone.0086392

Jadi, R. S., Sudeep, A. B., Kumar, S., Arankalle, V. A., & Mishra, A. C. (2010). Chandipura virus growth kinetics in vertebrate cell lines, insect cell lines & embryonated eggs. *INDIAN J MED RES*.

Jancarova, M., Polanska, N., Volf, P., & Dvorak, V. (2023). The role of sand flies as vectors of viruses other than phleboviruses. *Journal of General Virology*, *104*(4), 001837. https://doi.org/10.1099/jgv.0.001837

Jatzold, R., & Kutsch, H. (1982). Agro-ecological zones of the tropics, with a sample from Kenya. *Tropenlandwirt*, *83*, 15–34.

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Junglen, S., & Drosten, C. (2013). Virus discovery and recent insights into virus diversity in arthropods. *Current Opinion in Microbiology*, *16*(4), 507. https://doi.org/10.1016/J.MIB.2013.06.005

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, *27*(1), 112–128. https://doi.org/10.1002/pro.3280

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Karabatsos, N., & Taylor, T. O. B. R. M. (1985). International Catalogue of Arboviruses Including Certain Other Viruses of Vertebrates. *The American Journal of Tropical Medicine and Hygiene*, *25*(1), 204–205. https://doi.org/10.4269/ajtmh.1976.25.204

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

Killick-Kendrick, R. (1999). The biology and control of Phlebotomine sand flies. *Clinics in Dermatology*, *17*(3), 279–289. https://doi.org/10.1016/S0738-081X(99)00046-2

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R., & Bunning, M. (2003). Experimental Infection of North American Birds with the New York 1999 Strain of West Nile Virus. *Emerging Infectious Diseases*, *9*(3), 311–322. https://doi.org/10.3201/eid0903.020628

Koskei, E. C., Langat, S. K., Mutisya, J., Mulwa, F., Lutomiah, J., Koka, H., Oyola, S., Waihenya, R., & Sang, R. (2024). Isolation and phylogenetic characterization of arboviruses circulating among Phlebotomine sandflies in parts of North Rift, Kenya. *Frontiers in Virology*, *4*. https://doi.org/10.3389/fviro.2024.1289258

Kumar, N. P., Srinivasan, R., & Jambulingam, P. (2012). DNA barcoding for identification of sand flies (*Diptera: Psychodidae*) in India. *Molecular*

Ecology Resources, *12*(3), 414–420. https://doi.org/10.1111/j.1755-0998.2012.03117.x

Kwong, J. C., Mccallum, N., Sintchenko, V., & Howden, B. P. (2015). Whole genome sequencing in clinical and public health microbiology. *Pathology*, *47*(3), 199–210. https://doi.org/10.1097/PAT.0000000000000235

Langevin, S. A., Bowen, R. A., Reisen, W. K., Andrade, C. C., Ramey, W. N., Maharaj, P. D., Anishchenko, M., Kenney, J. L., Duggal, N. K., Romo, H., Bera, A. K., Sanders, T. A., Bosco-Lauth, A., Smith, J. L., Kuhn, R., & Brault, A. C. (2014). Host Competence and Helicase Activity Differences Exhibited by West Nile Viral Variants Expressing NS3-249 Amino Acid Polymorphisms. *PLOS ONE*, *9*(6), e100802. https://doi.org/10.1371/journal.pone.0100802

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), Article 4. https://doi.org/10.1038/nmeth.1923

Latrofa, M. S., Iatta, R., Dantas-Torres, F., Annoscia, G., Gabrielli, S., Pombi, M., Gradoni, L., & Otranto, D. (2018). Detection of *Leishmania infantum* DNA in Phlebotomine sand flies from an area where canine leishmaniosis is endemic in southern Italy. *Veterinary Parasitology*, *253*, 39–42. https://doi.org/10.1016/j.vetpar.2018.02.006

Laurent-Rolle, M., Boer, E. F., Lubick, K. J., Wolfinbarger, J. B., Carmody, A. B., Rockx, B., Liu, W., Ashour, J., Shupert, W. L., Holbrook, M. R., Barrett, A. D., Mason, P. W., Bloom, M. E., García-Sastre, A., Khromykh, A. A., & Best, S. M. (2010). The NS5 Protein of the Virulent West Nile Virus NY99 Strain Is a Potent Antagonist of Type I Interferon-Mediated JAK-STAT Signaling. *Journal of Virology*, *84*(7), 3503–3515. https://doi.org/10.1128/JVI.01161-09

Lawyer, P., Killick-Kendrick, M., Rowland, T., Rowton, E., & Volf, P. (2017). Laboratory colonization and mass rearing of Phlebotomine sand flies

(Diptera, Psychodidae). *Parasite*, *24*, 42. https://doi.org/10.1051/parasite/2017041

Leger, A., & Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software*, *4*(34), 1236. https://doi.org/10.21105/joss.01236

Leray, M., Ho, S.-L., Lin, I.-J., & Machida, R. J. (2018). MIDORI server: a webserver for taxonomic assignment of unknown metazoan mitochondrial encoded sequences using a curated database. Bioinformatics, 34(21), 3753 3754. doi: 10.1093/bioinformatics/bty454

Lewis, D. J., Young, D. G., Fairchild, G. B., & Minter, D. M. (1977). Proposals for a stable classification of the Phlebotomine sandflies (Diptera: Psychodidae). *Systematic Entomology*, *2*(4), 319–332. https://doi.org/10.1111/j.1365-3113.1977.tb00381.x

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Lobo, J. M. de A., Christofferson, R. C., & Mores, C. N. (2014). Investigations of Koutango Virus Infectivity and Dissemination Dynamics in Aedes aegypti Mosquitoes. *Environmental Health Insights*, *8s2*, EHI.S16005. https://doi.org/10.4137/EHI.S16005

Lwande, O. W., Lutomiah, J., Obanda, V., Gakuya, F., Mutisya, J., Mulwa, F., Michuki, G., Chepkorir, E., Fischer, A., Venter, M., & Sang, R. (2013). Isolation of Tick and Mosquito-Borne Arboviruses from Ticks Sampled from Livestock and Wild Animal Hosts in Ijara District, Kenya. *Vector-Borne and Zoonotic Diseases*, *13*(9), 637–642. https://doi.org/10.1089/vbz.2012.1190

Maharaj, P. D., Langevin, S. A., Bolling, B. G., Andrade, C. C., Engle, X. A., Ramey, W. N., Bosco-Lauth, A., Bowen, R. A., Sanders, T. A., Huang, C. Y.-H., Reisen, W. K., & Brault, A. C. (2019). N-linked glycosylation of the West Nile virus envelope protein is not a requisite for avian virulence or vector competence. *PLOS Neglected Tropical Diseases*, *13*(7), e0007473. https://doi.org/10.1371/journal.pntd.0007473

Malet, H., Egloff, M.-P., Selisko, B., Butcher, R. E., Wright, P. J., Roberts, M., Gruez, A., Sulzenbacher, G., Vonrhein, C., Bricogne, G., Mackenzie, J. M., Khromykh, A. A., Davidson, A. D., & Canard, B. (2007). Crystal Structure of the RNA Polymerase Domain of the West Nile Virus Non-structural Protein 5 *. *Journal of Biological Chemistry*, *282*(14), 10678–10689. https://doi.org/10.1074/jbc.M607273200

Marklewitz, M., Tchouassi, D. P., Hieke, C., Heyde, V., Torto, B., Sang, R., & Junglen, S. (2020). Insights into the Evolutionary Origin of Mediterranean Sandfly Fever Viruses. *mSphere*, *5*(5), 1–12. https://doi.org/10.1128/msphere.00598-20

Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N., & Gradoni, L. (2013). Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology*, *27*(2), 123–147. https://doi.org/10.1111/j.1365-2915.2012.01034.x

Marti, G. A., Balsalobre, A., Susevich, M. L., Rabinovich, J. E., & Echeverría, M. G. (2015). Detection of triatomine infection by Triatoma virus and horizontal transmission: Protecting insectaries and prospects for biological control. *Journal of Invertebrate Pathology*, *124*, 57–60. https://doi.org/10.1016/j.jip.2014.10.008

Mastrangelo, E., Milani, M., Bollati, M., Selisko, B., Peyrane, F., Pandini, V., Sorrentino, G., Canard, B., Konarev, P. V., Svergun, D. I., de Lamballerie, X., Coutard, B., Khromykh, A. A., & Bolognesi, M. (2007). Crystal Structure

and Activity of Kunjin Virus NS3 Helicase; Protease and Helicase Domain Assembly in the Full Length NS3 Protein. *Journal of Molecular Biology*, *372*(2), 444–455. https://doi.org/10.1016/j.jmb.2007.06.055

Medlin, S., Deardorff, E. R., Hanley, C. S., Vergneau-Grosset, C., Siudak-Campfield, A., Dallwig, R., Travassos da Rosa, A., Tesh, R. B., Pia Martin, M., Weaver, S. C., Vaughan, C., Ramirez, O., Sladky, K. K., & Paul-Murphy, J. (2016). Serosurvey of selected arboviral pathogens in free-ranging, two-toed sloths (Choloepus hoffmanni) and three-toed sloths (Bradypus variegatus) in Costa Rica, 2005–07. *Journal of Wildlife Diseases*, *52*(4), 883–892. https://doi.org/10.7589/2015-02-040

Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, *2*(1), 63–77. https://doi.org/10.1016/j.coviro.2011.12.004

Moureau, G., Ninove, L., Izri, A., Cook, S., Lamballerie, X. D., & Charrel, R. N. (2010). Flavivirus RNA in Phlebotomine sandflies. *Vector-Borne and Zoonotic Diseases*, *10*(2), 195–197. https://doi.org/10.1089/VBZ.2008.0216/ASSET/IMAGES/LARGE/FIG-1.JPEG

Mourya, D. T., Yadav, P. D., Ullas, P. T., Bhardwaj, S. D., Sahay, R. R., Chadha, M. S., Shete, A. M., Jadhav, S., Gupta, N., Gangakhedkar, R. R., Khasnobis, P., & Singh, S. K. (2019). Emerging/re-emerging viral diseases & new viruses on the Indian horizon. *Indian Journal of Medical Research*, *149*(4), 447. https://doi.org/10.4103/ijmr.IJMR_1239_18

Mukhwana, D. W., Makwali, J., Ngeiywa, M. M., Anjili, C. A., & Oyoo-Okoth, E. (2018). Spatial-temporal distribution and risk factors associated with Phlebotomine sand flies in Mt. Elgon Region, Western Kenya. *Journal of Entomology and Zoology Studies*, *January*.

Munstermann, L. E. (2018). Phlebotomine sand flies and moth flies (psychodidae). In *Medical and Veterinary Entomology* (pp. 191–211). Academic Press. https://doi.org/10.1016/B978-0-12-814043-7.00012-1

Munstermann, L. E. (2019). Phlebotomine Sand Flies and Moth Flies (Psychodidae). In G. R. Mullen & L. A. Durden (Eds.), *Medical and Veterinary Entomology (Third Edition)* (pp. 191–211). Academic Press. https://doi.org/10.1016/B978-0-12-814043-7.00012-1

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

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., & Koopmans, M. P. G. (2018). Overview of Virus Metagenomic Classification Methods and Their Biological Applications. *Frontiers in Microbiology*, *0*, 749. https://doi.org/10.3389/FMICB.2018.00749

Nzelu, C. O., Cáceres, A. G., Arrunátegui-Jiménez, M. J., Lañas-Rosas, M. F., Yañez-Trujillano, H. H., Luna-Caipo, D. V., Holguín-Mauricci, C. E., Katakura, K., Hashiguchi, Y., & Kato, H. (2015). DNA barcoding for identification of sand fly species (Diptera: Psychodidae) from leishmaniasis-endemic areas of Peru. *Acta Tropica*, *145*. https://doi.org/10.1016/j.actatropica.2015.02.003

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Ogola, E. O., Kopp, A., Bastos, A. D. S., Slothouwer, I., Omoga, D. C. A., Osalla, J., Sang, R., Torto, B., Junglen, S., & Tchouassi, D. P. (2023). Phlebovirus diversity in ticks from livestock in arid ecologies in Kenya. *Ticks and Tick-Borne Diseases*, *14*(1), 102087. https://doi.org/10.1016/j.ttbdis.2022.102087

Ondrejicka, D. A., Locke, S. A., Morey, K., Borisenko, A. V., & Hanner, R. H. (2014). Status and prospects of DNA barcoding in medically important parasites and vectors. *Trends in Parasitology*, *30*(12), 582–591. https://doi.org/10.1016/j.pt.2014.09.003

Patterson, E. I., Villinger, J., Muthoni, J. N., Dobel-Ober, L., & Hughes, G. L. (2020). Exploiting insect-specific viruses as a novel strategy to control vector-borne disease. *Current Opinion in Insect Science*, *39*, 50–56. https://doi.org/10.1016/j.cois.2020.02.005

Peralta, P. H., Shelokov, A., Vogel, J. E., & Longfellow, D. (1966). Isolation and Characterization of Arboviruses from Almirante, Republic of Panama. *The American Journal of Tropical Medicine and Hygiene*, *15*(3), 369–378. https://doi.org/10.4269/ajtmh.1966.15.369

Pérez-Ramírez, E., Llorente, F., del Amo, J., Fall, G., Sall, A. A., Lubisi, A., Lecollinet, S., Vázquez, A., & Jiménez-Clavero, M. Á. (2017). Pathogenicity evaluation of twelve West Nile virus strains belonging to four lineages from five continents in a mouse model: Discrimination between three pathogenicity categories. *Journal of General Virology*, *98*(4), 662–670. https://doi.org/10.1099/jgv.0.000743

Phan, T., Tesh, R. B., Guzman, H., & Delwart, E. (2020). Genomic characterization of Changuinola viruses from Panama: Evidence for multiple genome segment reassortment. *Virus Genes*, *56*(4), 527–530. https://doi.org/10.1007/s11262-020-01758-0

Philip, M. (1995). *Phlebotomine Sandfly (Diptera: Psychodidae) Behaviour and Utilization of Termite Hells and Animal Burrows in Baringo District, Kenya*.

Phumee, A., Wacharapluesadee, S., Petcharat, S., Tawatsin, A., Thavara, U., & Siriyasatien, P. (2021). Detection of Changuinola virus (Reoviridae: Orbivirus) in field-caught sand flies in southern Thailand. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *115*(9), 1039–1044. https://doi.org/10.1093/trstmh/traa203

Polseela, R., Jaturas, N., Thanwisai, A., Sing, K. W., & Wilson, J. J. (2016). Towards monitoring the sandflies (Diptera: Psychodidae) of Thailand: DNA barcoding the sandflies of Wihan Cave, Uttaradit. *Mitochondrial DNA*, *27*(5), 3795–3801. https://doi.org/10.3109/19401736.2015.1082085

Prow, N. A., Setoh, Y. X., Biron, R. M., Sester, D. P., Kim, K. S., Hobson-Peters, J., Hall, R. A., & Bielefeldt-Ohmann, H. (2014). The West Nile Virus-Like Flavivirus Koutango Is Highly Virulent in Mice due to Delayed Viral Clearance and the Induction of a Poor Neutralizing Antibody Response. *Journal of Virology*, *88*(17), 9947–9962. https://doi.org/10.1128/JVI.01304-14

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rodhain, F., Gonzalez, J. P., Mercier, E., Helynck, B., Larouze, B., & Hannoun, C. (1989). Arbovirus infections and viral haemorrhagic fevers in Uganda: A serological survey in Karamoja district, 1984. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *83*(6), 851–854. https://doi.org/10.1016/0035-9203(89)90352-0

Rodrigues, B. L., Baton, L. A., & Shimabukuro, P. H. F. F. (2020). Single-locus DNA barcoding and species delimitation of the sandfly subgenus Evandromyia (Aldamyia). *Medical and Veterinary Entomology*, *34*(4), 420–431. https://doi.org/10.1111/MVE.12458

Rodríguez, L. L. (2002). Emergence and re-emergence of vesicular stomatitis in the United States. *Virus Research*, *85*(2), 211–219. https://doi.org/10.1016/S0168-1702(02)00026-6

Rosa, A. P. A. T. da, Tesh, R. B., Pinheiro, F. P., Rosa, J. F. S. T. da, Peralta, P. H., & Knudson, D. L. (1984). Characterization of the Changuinola Serogroup Viruses (Reoviridae: Orbivirus). *Intervirology*, *21*(1), 38–49. https://doi.org/10.1159/000149501

Rose, A. S., Bradley, A. R., Valasatava, Y., Duarte, J. M., Prlić, A., & Rose, P. W. (2018). NGL viewer: Web-based molecular graphics for large complexes. *Bioinformatics*, *34*(21), 3755–3758. https://doi.org/10.1093/bioinformatics/bty419

Roux, S., Matthijnssens, J., & Dutilh, B. E. (2021). Metagenomics in Virology. In *Encyclopedia of Virology* (pp. 133–140). Elsevier. https://doi.org/10.1016/b978-0-12-809633-8.20957-6

Rozo-Lopez, P., Drolet, B. S., & Londoño-Renteria, B. (2018). Vesicular Stomatitis Virus Transmission: A Comparison of Incriminated Vectors. *Insects*, *9*(4), Article 4. https://doi.org/10.3390/insects9040190

Sadeghi, M., Altan, E., Deng, X., Barker, C. M., Fang, Y., Coffey, L. L., & Delwart, E. (2018). Virome of > 12 thousand *Culex* mosquitoes from throughout California. *Virology*, *523*, 74–88. https://doi.org/10.1016/j.virol.2018.07.029

Sánchez-Seco, M. P., Vázquez, A., Collao, X., Hernández, L., Aranda, C., Ruiz, S., Escosa, R., Marqués, E., Bustillo, M. A., Molero, F., & Tenorio, A. (2010). Surveillance of arboviruses in spanish wetlands: Detection of new flavi-and phleboviruses. *Vector-Borne and Zoonotic Diseases*, *10*(2), 203–206. https://doi.org/10.1089/VBZ.2008.0188

Sánchez-Seco, M.-P., Echevarría, J.-M., Hernández, L., Estévez, D., Navarro-Marí, J.-M., & Tenorio, A. (2003). Detection and identification of Toscana and

other phleboviruses by RT-nested-PCR assays with degenerated primers. *Journal of Medical Virology*, *71*(1), 140–149. https://doi.org/10.1002/JMV.10465

Scherret, J. H., Poidinger, M., Mackenzie, J. S., Broom, A. K., Deubel, V., Lipkin, W. I., Briese, T., Gould, E. A., & Hall, R. A. (2001). The relationships between West Nile and Kunjin viruses. *Emerging Infectious Diseases*, *7*(4), 697–705. https://doi.org/10.3201/eid0704.010418

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864. https://doi.org/10.1093/bioinformatics/btr026

Setoh, Y. X., Prow, N. A., Hobson-Peters, J., Lobigs, M., Young, P. R., Khromykh, A. A., & Hall, R. A. (2012). Identification of residues in West Nile virus pre-membrane protein that influence viral particle secretion and virulence. *Journal of General Virology*, *93*(9), 1965–1975. https://doi.org/10.1099/vir.0.044453-0

Shope, R. E. (2003). Epidemiology of Other Arthropod-Borne Flaviviruses Infecting Humans. In T. J. Chambers & T. P. Monath (Eds.), *Advances in Virus Research* (Vol. 61, pp. 373–391). Academic Press. https://doi.org/10.1016/S0065-3527(03)61009-2

Sudeep, A. B., Gurav, Y. K., & Bondre, V. P. (2016). Changing clinical scenario in Chandipura virus infection. *Indian Journal of Medical Research*, *143*(JUNE), 712–721. https://doi.org/10.4103/0971-5916.191929

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.xTamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Tandale, B. V., Tikute, S. S., Arankalle, V. A., Sathe, P. S., Joshi, M. V., Ranadive, S. N., Kanojia, P. C., Eshwarachary, D., Kumarswamy, M., & Mishra, A. C. (2008). Chandipura virus: A major cause of acute encephalitis in children in North Telangana, Andhra Pradesh, India. *Journal of Medical Virology*, *80*(1), 118–124. https://doi.org/10.1002/jmv.21041

Tchouassi, D. P., Marklewitz, M., Chepkorir, E., Zirkel, F., Agha, S. B., Tigoi, C. C., Koskei, E., Drosten, C., Borgemeister, C., Torto, B., Junglen, S., & Sang, R. (2019). Sand fly–associated phlebovirus with evidence of neutralizing antibodies in humans, Kenya. *Emerging Infectious Diseases*, *25*(4), 681–690. https://doi.org/10.3201/eid2504.180750

Tesh, R. B., Saidi, S., Gajdamovič, S. J., Rodhain, F., & Vesenjak-Hirjan, J. (1976). Serological studies of the epidemiology of sandfly fever in the Old World. *Bulletin of the World Health Organization*, *54*(6), 663.

Tigoi, C., Lwande, O., Orindi, B., Irura, Z., Ongus, J., & Sang, R. (2015). Seroepidemiology of Selected Arboviruses in Febrile Patients Visiting Selected Health Facilities in the Lake/River Basin Areas of Lake Baringo, Lake Naivasha, and Tana River, Kenya. *Vector-Borne and Zoonotic Diseases*, *15*(2), 124–132. https://doi.org/10.1089/vbz.2014.1686

Traoré-lamizana, M., Fontenille, D., Diallo, M., Bâ, Y., Zeller, H. G., Mondo, M., Adam, F., Thonon, J., & Maïga, A. (2001). Arbovirus Surveillance from 1990 to 1995 in the Barkedji Area (Ferlo) of Senegal, a Possible Natural Focus of Rift Valley Fever Virus. *Journal of Medical Entomology*, *38*(4), 480–492. https://doi.org/10.1603/0022-2585-38.4.480

Travassos da Rosa, A. P., Tesh, R. B., Pinheiro, F. P., Travassos da Rosa, J. F., Peralta, P. H., & Knudson, D. L. (1984). Characterization of the Changuinola serogroup viruses (Reoviridae: Orbivirus). *Intervirology*, *21*(1), 38–49. https://doi.org/10.1159/000149501

Varelas-Wesley, I., & Calisher, C. H. (1982). Antigenic relationships of flaviviruses with undetermined arthropod-borne status. *The American Journal of Tropical Medicine and Hygiene*, *31*(6), 1273–1284. https://doi.org/10.4269/ajtmh.1982.31.1273

Vasilakis, N., & Tesh, R. B. (2015). Insect-specific viruses and their potential impact on arbovirus transmission. *Current Opinion in Virology*, *15*, 69–74. https://doi.org/10.1016/j.coviro.2015.08.007

Vasilakis, N., Forrester, N. L., Palacios, G., Nasar, F., Savji, N., Rossi, S. L., Guzman, H., Wood, T. G., Popov, V., Gorchakov, R., González, A. V., Haddow, A. D., Watts, D. M., da Rosa, A. P. A. T., Weaver, S. C., Lipkin, W. I., & Tesh, R. B. (2013). Negevirus: A Proposed New Taxon of Insect-Specific Viruses with Wide Geographic Distribution. *Journal of Virology*, *87*(5), 2475–2488. https://doi.org/10.1128/jvi.00776-12

Volf, P., & Volfova, V. (2011). Establishment and maintenance of sand fly colonies. *Journal of Vector Ecology*, *36*(s1), S1–S9. https://doi.org/10.1111/j.1948-7134.2011.00106.x

Wang, H., Chen, Q., & Wei, T. (2024). Complex interactions among insect viruses-insect vector-arboviruses. *Insect Science*, *31*(3), 683–693. https://doi.org/10.1111/1744-7917.13285

Watts, D. M., El-Tigani, A., Botros, B. A. M., Salib, A. W., Olson, J. G., McCarthy, M., & Ksiazek, T. G. (1994). Arthropod-borne viral infections associated with a fever outbreak in the Northern Province of Sudan. *Journal of Tropical Medicine and Hygiene*, *97*(4), 228–230.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution*, *35*(3), 773–777. https://doi.org/10.1093/molbev/msx335

Wick, R. R. (2017). *Porechop: Adapter trimmer for Oxford Nanopore reads*. https://github.com/rrwick/Porechop

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 129. https://doi.org/10.1186/s13059-019-1727-y

Woelk, C. H., & Holmes, E. C. (2002). Reduced Positive Selection in Vector-Borne RNA Viruses. *Molecular Biology and Evolution*, *19*(12), 2333–2336. https://doi.org/10.1093/oxfordjournals.molbev.a004059

Woyessa, A. B., Omballa, V., Wang, D., Lambert, A., Waiboci, L., Ayele, W., Ahmed, A., Abera, N. A., Cao, S., Ochieng, M., Montgomery, J. M., Jima, D., & Fields, B. (2014). An outbreak of acute febrile illness caused by sandfly fever sicilian virus in the Afar Region of Ethiopia, 2011. *American Journal of Tropical Medicine and Hygiene*, *91*(6), 1250–1253. https://doi.org/10.4269/ajtmh.14-0299

APPENDICES

Appendix I: SERU Ethical Approval for the Study

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/RD/22 **February 3, 2023**

TO: **THIRU JANE WAMBUI,**
PRINCIPAL INVESTIGATOR.

THROUGH: **THE DEPUTY DIRECTOR, CVR,**
NAIROBI.

Dear Madam,

RE: **PROTOCOL NO. KEMRI/SERU/CVR/007-2022/4570 *(RESUBMISSION II OF INITIAL SUBMISSION)*: MOLECULAR CHARACTERIZATION OF RNA VIRUSES IN SANDFLIES FROM SELECTED COUNTIES IN KENYA *(VERSION 1.6 DATED 18 JANUARY 2023)***

Reference is made to your letter dated January 24, 2023. The KEMRI Scientific and Ethics Review Unit (SERU) acknowledges receipt of the revised study documents on January 24, 2023.

This is to inform you that the Committee notes that the following issues raised during the 328th Committee A meeting of the KEMRI Scientific and Ethics Review Unit (SERU) held on **October 11, 2022** have been adequately addressed.

Consequently, the study is **granted approval** for implementation effective this day, **February 3, 2023** for a period of **one year**. Please note that authorization to conduct this study will automatically expire on **February 2, 2024**. If you plan to continue data collection or analysis beyond this date, please submit an application for continuation approval by **December 22, 2023.**

Please note that only approved documents including (informed consents, study instruments, Material Transfer Agreement) will be used. You are required to submit any proposed changes to this study to SERU for review and the changes should not be initiated until written approval from SERU is received. Any unanticipated problems resulting from the implementation of this study should be brought to the attention of SERU and you should advise SERU when the study is completed or discontinued.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology and Innovation (NACOSTI) https://oris.nacosti.go.ke and also obtain other clearances needed.

Yours faithfully,

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

Appendix II: SERU Ethical Approval for the Study

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **415474**

Date of Issue: **14/April/2023**

RESEARCH LICENSE

This is to Certify that Miss.. Jane Wambui Thiiru of Jomo Kenyatta University of Agriculture and Technology, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Baringo, Isiolo, Kilifi, Kisumu, Kwale, Nakuru, Westpokot on the topic: MOLECULAR CHARACTERIZATION OF RNA VIRUSES IN SANDFLIES FROM SELECT COUNTIES IN KENYA for the period ending : 14/April/2024.

License No: **NACOSTI/P/23/24603**

415474

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

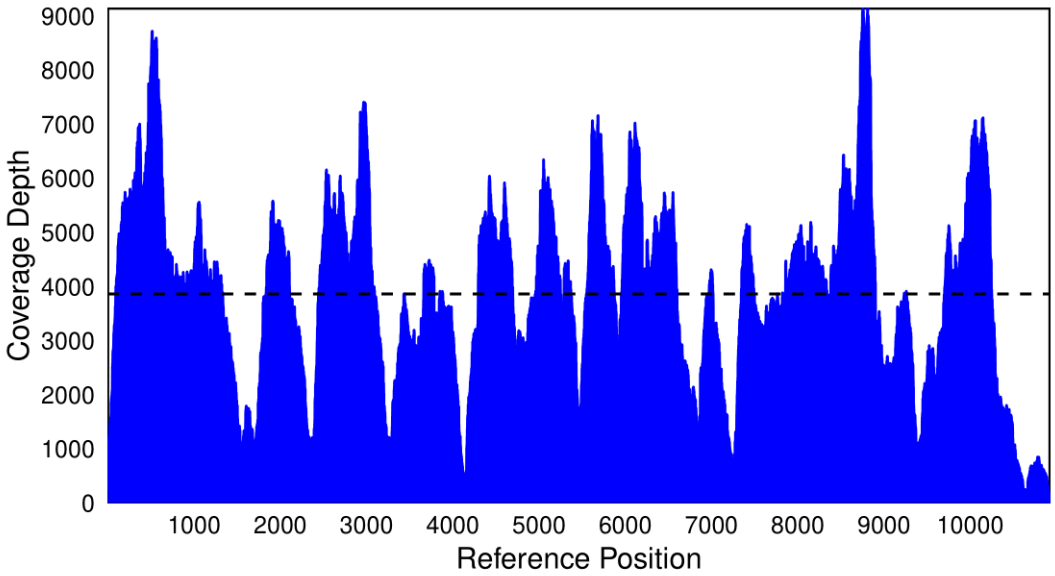
Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

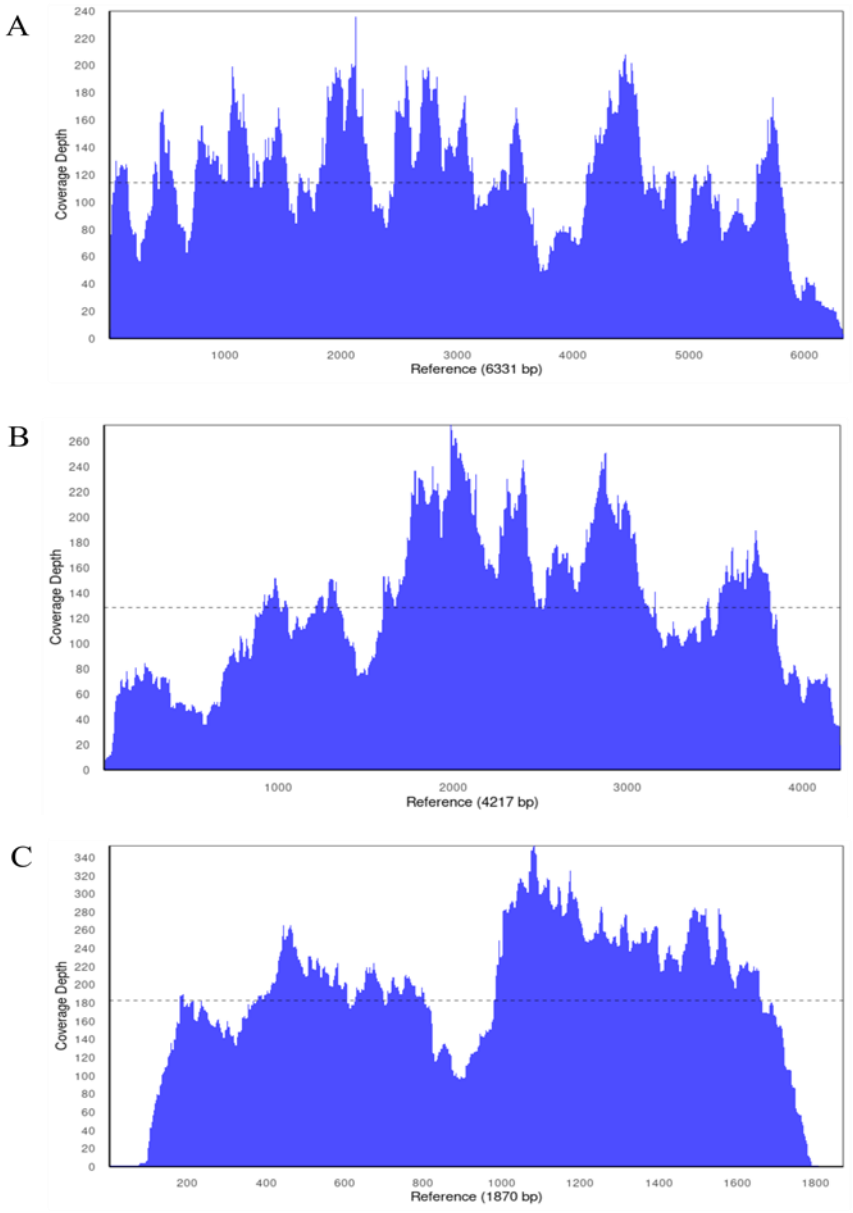
See overleaf for conditions

Appendix III: Plot Showing the Depth of Coverage of WN-KOUTV Strain

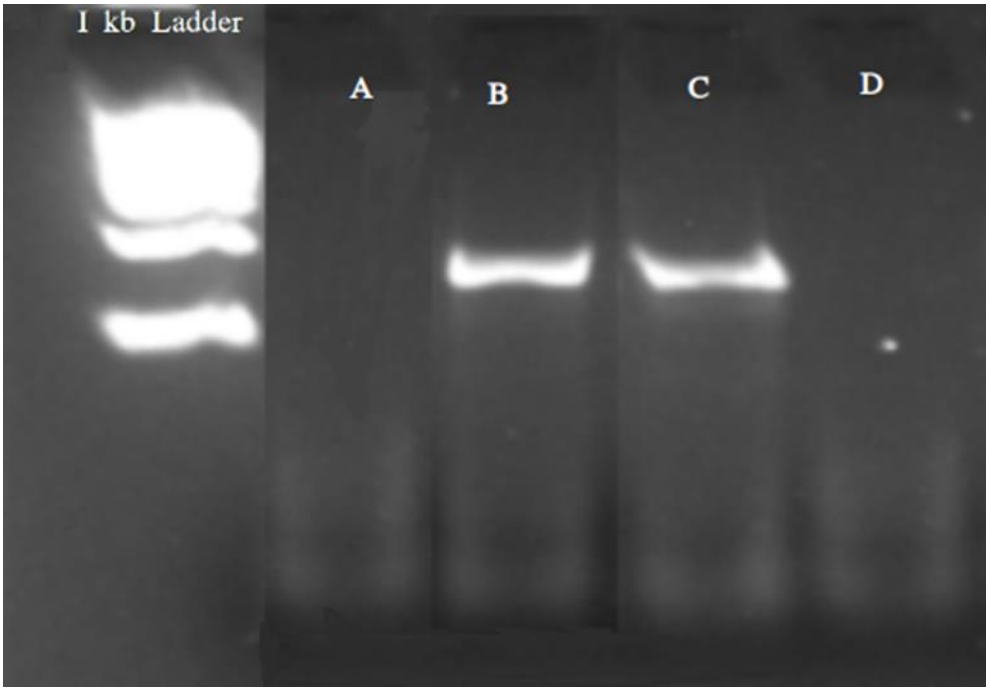
BAR_S3_S008


9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Coverage Depth
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Reference Position

Appendix IV: Plots Showing the Depth of Coverage of Bogoria Virus BAR_S5_230



Appendix V: Gel Photo of the Vector COI Amplicons



LANES:

1 Kb ladder

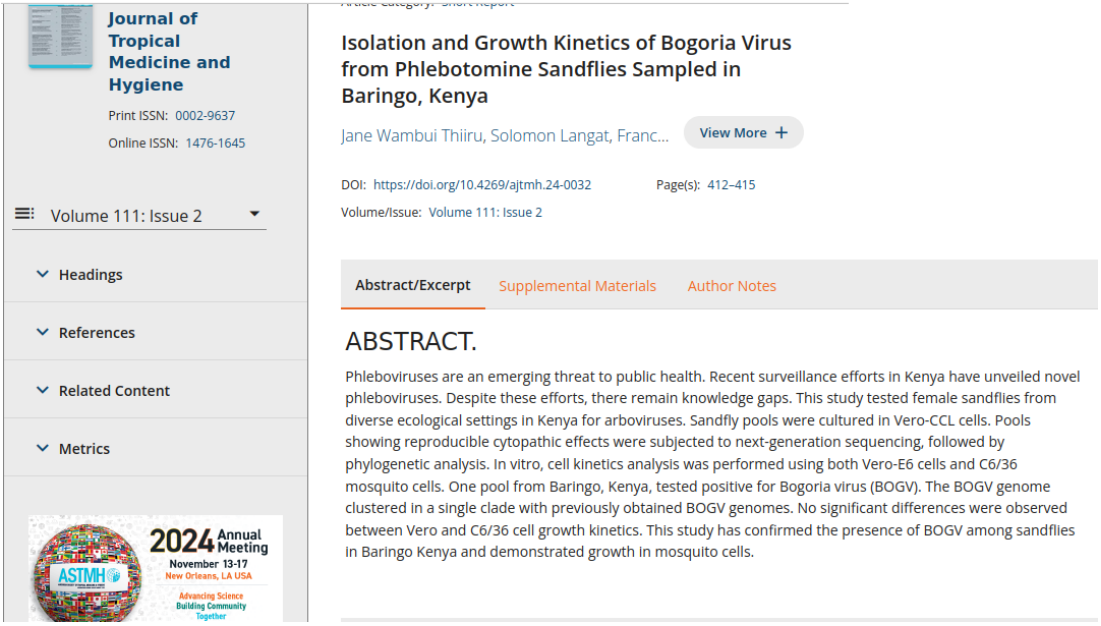
A – No-template control

B – BAR_S3_S008

C – BAR_S5_230

D – No-template control

Appendix VI: Publication 1



Journal of Tropical Medicine and Hygiene

Print ISSN: 0002-9637

Online ISSN: 1476-1645

Volume 111: Issue 2

Headings

References

Related Content

Metrics

2024 Annual Meeting November 13-17 New Orleans, LA USA Advancing Science Building Community Together

Isolation and Growth Kinetics of Bogoria Virus from Phlebotomine Sandflies Sampled in Baringo, Kenya

Jane Wambui Thiiru, Solomon Langat, Franc... View More +

DOI: https://doi.org/10.4269/ajtmh.24-0032 Page(s): 412–415

Volume/Issue: Volume 111: Issue 2

Abstract/Excerpt Supplemental Materials Author Notes

ABSTRACT.

Phleboviruses are an emerging threat to public health. Recent surveillance efforts in Kenya have unveiled novel phleboviruses. Despite these efforts, there remain knowledge gaps. This study tested female sandflies from diverse ecological settings in Kenya for arboviruses. Sandfly pools were cultured in Vero-CCL cells. Pools showing reproducible cytopathic effects were subjected to next-generation sequencing, followed by phylogenetic analysis. In vitro, cell kinetics analysis was performed using both Vero-E6 cells and C6/36 mosquito cells. One pool from Baringo, Kenya, tested positive for Bogoria virus (BOGV). The BOGV genome clustered in a single clade with previously obtained BOGV genomes. No significant differences were observed between Vero and C6/36 cell growth kinetics. This study has confirmed the presence of BOGV among sandflies in Baringo Kenya and demonstrated growth in mosquito cells.

Appendix VII: Publication 2

PLOS ONE

RESEARCH ARTICLE

Characterization of West Nile virus Koutango lineage from phlebotomine sandflies in Kenya

Jane Wambui Thiiru[1,2,3]*, Solomon Langat[2], Francis Mulwa[2], Stephanie Cinkovich[4], Hellen Koka[2], Santos Yalwala[1], Samoel Khamadi[2], Justus Onguso[3], Nicholas Odemba[1], Francis Ngere[1], Jaree Johnson[5], Timothy Egbo[1], Eric Garges[1], Elly Ojwang[1], Fredrick Eyase[1,2]

1 Department of Emerging Infectious Diseases, United States Army Medical Research Directorate-Africa, Nairobi, Kenya, **2** Centre for Virus Research, Kenya Medical Research Institute, Nairobi, Kenya, **3** Institute for Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, **4** Global Emerging Infections Surveillance Branch, United States Armed Forces Health Surveillance Division, Silver Spring, Maryland, United States of America, **5** United States Armed Forces Pest Management Board, Silver Spring, Maryland, United States of America

* janethiiru@gmail.com

OPEN ACCESS

Citation: Thiiru JW, Langat S, Mulwa F, Cinkovich S, Koka H, Yalwala S, et al. (2024) Characterization of West Nile virus Koutango lineage from phlebotomine sandflies in Kenya. PLoS ONE 19(8): e0301956. https://doi.org/10.1371/journal.pone.0301956

Editor: Rhys Harold Parry, The University of Queensland, AUSTRALIA

Received: March 25, 2024

Accepted: July 16, 2024

Published: August 22, 2024

Data Availability Statement: The raw sequence reads generated in this study are available at the NCBI Sequence Read Archive (SRA) database under BioProject accession PRJNA1139200. The sequence data has been assigned BioSample accession numbers SAMN42766516 and SAMN42766517. The genomes of viruses were deposited at GenBank under the accession codes: PP489328 and PP885725.

Funding: This work was funded by the Armed Forces Health Surveillance Branch (AFHSB) and its

Abstract

The West Nile virus (WNV), primarily transmitted by mosquitoes, is one of the most widespread flaviviruses globally, with past outbreaks occurring in the USA and Europe. Recent studies in parts of Africa, including Kenya, have identified the West Nile virus Koutango lineage (WN-KOUTV) among phlebotomine sandfly populations, however, our understanding of this virus remains limited. This study aimed to characterize WN-KOUTV from phlebotomine sandflies. Sandflies were sampled between 12th -16th March 2021 and 16th -20th March 2023 from six villages each in Baringo and Isiolo Counties, using CDC light traps. Female sandflies were taxonomically identified and pooled based on genus and site of collection. Virus isolation was performed in Vero cells. Viral genomes were determined using next-generation sequencing. Phylogenetic and molecular clock analyses were done to decipher the virus's evolutionary relationships. Comparative analyses of amino acid sequences were performed to determine variations. Protein modeling in Pymol was conducted to elucidate variations in key protein regions. Evolutionary pressure analysis investigated the selection pressures on the virus. *In vitro* experiments were done to investigate the virus growth kinetics in mammalian Vero E6 and mosquito C6/36 cells. We report the isolation of WN-KOUTV from Salabani in Baringo and Aremet in Isiolo, Kenya. The isolated WN-KOUTVs clustered with previously identified WN-KOUTV strains. Comparative analysis revealed a unique amino acid at NS5 653. The WN-KOUTV lineage as a whole is under purifying selective pressure, with diversifying pressure acting at site NS3 267. The current WN-KOUTV replicated in Vero E6 and C6/36 cells comparable to West Nile virus Lineage 1a, isolated from mosquitoes. Subsequent isolations of WN-KOUTV in phlebotomine sandflies suggest potential vectors, however, vector competence studies would confirm this. Replication in mammalian and insect cell lines suggests there may exist a vector/host relationship. We speculate the close genetic relationship of WN-KOUTV strains from East and West Africa