COMPARING MITOCHONDRIAL GENE SEQUENCES OF *AEDES AEGYPTI* IN EAST AFRICA TO THE GLOBAL ANTHROPOPHILIC STRAIN

Brian Bwanya Edward

A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Bioinformatics of Pwani University

September 2020

DECLARATION

This thesis is my original work and has not been presented in any other university or any other award.

Signature …… Date……14 April 2021……

Brian Bwanya Edward

SG30/PU/36152/18

Pwani University

Declaration by supervisors

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Signature …… Date …16 April 2021 ……

Dr. Jandouwe Villinger (PhD)

International Centre of Insect Physiology and Ecology (*icipe*)

Signature …… Date 16 April 2021……

Dr. David P. Tchouassi (PhD)

International Centre of Insect Physiology and Ecology (*icipe*)

Signature …… Date…19 April 2021……

Dr. David Mburu (PhD)

Pwani University

DEDICATION

To my parents and teachers who taught me how to think and reason, and to be optimistic in every situation in life.

ACKNOWLEDGEMENT

My deepest gratitude goes to my supervisors, Dr. Villinger, Dr. Tchouassi, and Dr. Mburu, for their guidance from conceptualization of the research, the actual research activities to compilation of this thesis.

I appreciate the financial support received from the Fogarty International Center at the U.S. National institutes of health (NIH) through the Eastern Africa Network for Bioinformatics training (EANBiT) project led by Dr. Daniel Masiga at the International Centre of Insect Physiology and Ecology (*icipe*).

I want to thank Tatenda Chiuya, DVM for provision of DNA samples and guidance during laboratory assays.

I would like to acknowledge my ever-loving parents who have always stood by me, giving me the courage to pursue my goals. Special appreciation goes to all my teachers and mentors.

Finally, yet importantly, I thank the Almighty God for good health and wisdom to carry out the scholarly work herein.

ABSTRACT

Advances in genotyping methods have shed more light into mosquito genetics and improved our understanding of vector-borne disease transmission cycles. Outside of Africa, *Aedes aegypti,* the main vector of arboviruses such as dengue, chikungunya, yellow fever, and Zika viruses, prefers humans for a blood meal, resting and breeding in close association with human settlements. However, in its native African ecology, both domestic (*Aedes aegypti aegypti*) and sylvatic (forest) (*Aedes aegypti formosus*) lineages occur. Further, the African *Ae. aegypti* populations have been found to exhibit divergence in typically conserved mitochondrial cytochrome *c* oxidase subunit I (*COI*) genes and traits of epidemiological importance, including developmental time, foraging, oviposition, and resting behaviour. This study tested the hypothesis that mitochondrial variation in *Ae. aegypti* populations correlates with dengue outbreak occurrence patterns. Specifically, molecular investigations were conducted on mitochondrial cytochrome *c* oxidase subunits I and II (*COI, COII*), and cytochrome *b* (*cyt b*) gene sequences of *Ae. aegypti* generated by Sanger sequencing. Additional sequences were retrieved from GenBank and the Barcode of Life Data (BOLD) portal. The vectors were sampled from selected sites in Kenya *viz:* Kilifi and Kwale counties in coastal Kenya (endemic for dengue), Maasai Mara National Reserve in Narok County (sylvatic), and Kakamega, Bungoma, and Busia counties in western Kenya (non-dengue endemic region). The fragment sizes of the generated *Ae. aegypti* sequences were ~1100 bp, 700 bp, and ~350 bp for *COI, COII*, and *cyt b,* respectively. These sequences were analysed by maximum likelihood phylogenetic reconstruction. The *COI* gene sequences from sylvatic and domestic settings resolved *Ae. aegypti* into three lineages. The majority of the *COI* sequences of samples from Bungoma, Busia, and Kakamega counties clustered with referenced sylvatic rather than domestic strains, suggesting occurrence of the forest form

in the domestic sites in these counties in Kenya. On the other hand, most *COI* sequences of samples from Kilifi, Narok, and Kwale counties clustered with referenced domestic than sylvatic strains. Since the domestic form is a better vector of dengue virus than the forest form, these findings, coupled with factors such as temperature and humidity, may explain recurrent dengue outbreaks at the coast of Kenya.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

Aaa:	*Aedes aegypti aegypti*
Aaf:	*Aedes aegypti formosus*
icipe:	International Centre of Insect Physiology and Ecology
COI:	Cytochrome c oxidase subunit 1
DNA:	Deoxyribonucleic acid
*mt*DNA:	Mitochondrial DNA
WHO:	World Health Organization
BOLD:	Barcode of life data system
PCR:	Polymerase Chain Reaction
NADH	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
ND4:	NADH-ubiquinone oxidoreductase chain 4
URL:	Uniform resource locator
PPS:	Protein precipitation solution
bp:	Base pair
COII:	Cytochrome c oxidase subunit II gene
cyt b:	Cytochrome b gene
PCA:	Principal Component Analysis
MSA:	Multiple Sequence Alignment

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background

Mitochondrial genome sequences are useful in inferring phylogenetic relationships, population genetics and molecular evolution (Behura *et al*., 2011). These are molecular pillars that have been widely exploited in insect-vector biology and ecology as molecular markers (Behura *et al.*, 2011). The sequences are highly conserved among phyla and are present in high copy numbers with *mt*DNA being advantageous in genetic diversity studies due to its faster rate of evolution compared to nuclear DNA (Moriyama & Powell, 1997).

Mitochondrial NADH dehydrogenase subunit 4 (*ND4*) gene has been reported to be an excellent genetic marker for phylogenetic analysis and population genetic studies (Ali *et al*., 2016; Bosio *et al*., 2005; Seixas *et al*., 2013). Mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene sequence is also commonly used to discriminate closely related species and characterize diversity within the animal kingdom (Hebert *et al.*, 2003). The *COI* gene sequence is used in DNA barcoding to complement morphological identification of mosquito species (Ajamma *et al*., 2016; Chan *et al*., 2014). Diversity in the gene's sequence is attributed to its high rate of sequence change, especially in the amino acid sequences coded by the 5′ end of the gene in most animal groups (Hebert *et al*., 2003). Both *COI* and *COII* gene sequences have been used for insect species molecular identification and phylogenetic reconstruction (Cook *et al*., 2005; Wang *et al*., 2016), and are well represented in public repositories. The usefulness of mitochondrial cytochrome *b* (*cyt b*) gene in molecular systematics has also been reported (Gimeno *et al*., 1997). Despite the importance and usefulness of the *COI, COII*, and *cyt b*

mitochondrial genes in population genomics, their utility in evolutionary history and discrimination of mosquito subspecies has not been widely exploited.

Aedes aegypti is the principal vector of medically important arboviruses to humans, with a deserved global reputation for transmission of yellow fever, dengue, chikungunya, and Zika viruses (Agha *et al*., 2019; Futami *et al*., 2019; Karungu *et al*., 2019; Rose *et al*., 2020). The distribution of this mosquito species has increased worldwide, making it an important model for studying habitat invasion biology (Gloria-Soria *et al*., 2016). *Aedes aegypti* is native to Africa, and in sub-Saharan Africa, the vector has both domestic and ancestral populations (Powell & Tabachnick, 2013). The ancestral and domestic forms are referred to as *Ae. aegypti formosus* and *Ae. aegypti aegypti* at subspecies level, respectively. The *Ae. aegypti formosus* strain has been observed to lay eggs in tree holes and prefers non-human blood, behaviours that distinguish it from the domestic form, *Ae. aegypti aegypti*, which breeds in close proximity with humans, lays eggs in stored water jars, and prefers human blood meals (Powell & Tabachnick, 2013).

Humans occupy the widest range of habitats on earth and once a species such as *Ae. aegypti* evolves the repertoire to coexist with humans, they will likely be spread with human movement (Powell & Tabachnick, 2013). One consequence of such coexistence is the adaptation of *Ae. aegypti* to breeding in ecological conditions around human habitats, leading to ‘domestication’ of the vector (Crawford *et al*., 2017). Domestication of *Ae. aegypti* occurs when the vector adapts to breeding in human habitats (Powell & Tabachnick, 2013). Adult female *Ae. aegypti* mosquitoes bite host for a bloodmeal, which is needed for provision of proteins for egg development and subsequent viability. *Aedes aegypti aegypti* (Aaa), the domestic form, is thought to have arose when the African subspecies, *Ae. aegypti formosus* (Aaf), evolved from being zoophilic and living in forest

habitats into an anthropophilic form which resides in habitats near human populations (Crawford *et al*., 2017).

The importance of the strains of *Ae. aegypti* in human history in relation to public health emphasizes the need for studies of the mosquito as a long-term national, regional, and global problem rather than as a temporary local threat (Soper, 1970). Therefore, this study leverages on research debt on utility and relevance of mitochondrial genes to resolve the *Ae. aegypti* subspecies according to their ecological niches and global distribution. This is important in unravelling *Ae. aegypti*'s past in order to understand changes in behaviour of the vector that led to its present global distribution which determines human populations at risk of exposure to the arboviral disease pathogens that this vector transmits (Powell *et al*., 2018).

1.2. Statement of the problem

Aedes aegypti aegypti (Aaa) and *Ae. aegypti formosus* (Aaf) are known to co-exist in parts of the endemic coastal niches in East Africa (Gloria-Soria *et al*., 2016). Interestingly, in urban areas, they freely introgress, while in sylvatic settings both forms do not interbreed despite cohabitation in forest habitats (Gloria-Soria *et al.*, 2016). Further, populations outside Africa are more diverse compared to ancestral populations in sub-Saharan Africa (Crawford *et al.*, 2017; Gloria-Soria *et al*., 2016). These earlier findings prompted efforts to investigate whether there could have been co-evolutionary changes in the mitochondrial genome of these mosquitoes due to 'domestication' and their subsequent phylogeography that may be associated with transmission of arboviruses.

1.3 Justification of the study

Studies of the population structure of strains of *Ae. aegypti* are not just of academic interest, but are of key importance to global health (Redmond *et al*., 2020). *Aedes aegypti* is responsible for major arbovirus outbreaks and pandemics including yellow fever, dengue, chikungunya, and Zika viruses (Kotsakiozi *et al.*, 2018). However, this study mainly focuses on dengue virus due to its major health impact on a global scale compared to the other arboviruses which may be region specific. For example, Zika and chikungunya viruses have caused substantially more burden in the Americas than in any other World Health Organization (WHO) region (Puntasecca *et al*., 2021). *Aedes aegypti* is of major concern as a vector of dengue fever (Powell & Tabachnick, 2013). Kenya has recently experienced recurring dengue outbreaks along its coast (Agha *et al*., 2019; Karungu *et al.*, 2019). Brady and colleagues (2012) estimated that 3.97 billion people are at risk of dengue virus infection. Since the two subspecies of *Ae. aegypti* differ in their ability to transmit arbovirus pathogens (Crawford *et al*., 2017), variations in their mitochondrial genes may have important implications in dengue virus transmission. Kenya has experienced repetitive dengue outbreaks (Baba *et al*., 2016) and their management has been at the expense of the country's economy (Karungu *et al*., 2019). Efforts aimed at tracking transmission of these outbreaks have been hampered by limited genomic data (Giovanetti *et al*., 2019). An approach such as phylogenetic analysis of mitochondrial genes of arbovirus vectors may be helpful to understand how the distributional profile of *Ae. aegypti* correlates with dengue disease outbreaks.

Different markers have distinct rates of evolution and levels of variability (Rašić *et al*., 2015). The mitochondrial genes *COI*, *COII*, and *cyt b* are key molecular tools used in evolutionary detections and were selected in this study because they are present in multiple copies in a cell and they evolve faster than nuclear genes in arthropods including

insects, and thus are well suited to the study of closely related taxa (Cook *et al*., 2005). Both *COI* and *COII* have been reported to have the characteristics of a molecular clock, thus variations in their sequences may reveal important information with regards to speciation (Wang *et al*., 2016). The findings of this study were employed in testing the hypothesis that there is correlation between mitochondrial gene sequence variants and dengue outbreak occurrence patterns in East Africa and globally. This report contributes to understanding of the population genomics of this vector and informs on management options that can be employed to keep the vector at bay.

1.4 Research questions

1. Is there variation in mitochondrial *COI* gene sequences of *Ae. aegypti* populations from domestic settings compared to populations from sylvatic habitats in East Africa?
2. Is there variation in mitochondrial *COI, COII*, and *cyt b* gene sequences derived from East African strains of *Ae. aegypti* populations compared to similar sequences of populations from outside Africa?
3. Does the distribution profile of the above-mentioned *Ae. aegypti* gene sequences correlate with global dengue outbreak occurrence patterns?

1.5 Main objective of the study

To compare the mitochondrial gene sequences of sylvatic and domestic strains of *Ae. aegypti* in East Africa to strictly anthropophilic strains in continents outside Africa.

1.5.1 Specific Objectives

1. To establish similarities and differences between *COI* gene sequences of *Ae. aegypti* sampled near human habitation to those sampled in sylvatic settings in East Africa.

2. To assess *COI, COII*, and *cyt b* gene sequence divergence between *Ae. aegypti* populations within and outside Africa.
3. To establish the correlation of the mitochondrial gene sequence variants and global dengue outbreak occurrence patterns.

1.6 Scope of the study

The major assumption in this study is that all *COI*, *COII* and *cyt b* sequences from outside Africa represent anthropophilic strains (Gloria-Soria *et al*., 2016; Kotsakiozi *et al*., 2018). The study utilised archived *Ae. aegypti* samples that had been collected from Rabai in Kilifi County and stored at -81°C for three months prior to start of the research. Specific interest in Rabai is driven by the observation that the two forms of *Ae. aegypti* populations in the area have remained genetically distinct from one another (Mattingly, 1957; Tabachnick *et al*., 1979), and there are higher anthropophilic populations of *Ae. aegypti* at the Kenyan coast compared to other areas within Kenya (Agha *et al*., 2019; Ndenga *et al*., 2017).

The study explores evolution of *COI* gene in *Ae. aegypti* strains and how domestication might have led to *Ae. aegypti* forage specialisation and subsequent disease pathogen transmission. The three mitochondrial genes (*COI*, *COII* and *cyt b*) were examined since analysis of multiple unlinked genes has been shown to be superior in phylogenetic reconstruction compared to single gene analysis (Bevan *et al*., 2005).

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Evolutionary history of *Aedes aegypti*

Documented evidence suggests that *Ae. aegypti* originated in Africa from an ancestral sylvatic and more zoophilic form *Ae. aegypti formosus*, which expanded from tropical forests to urban areas giving rise to a domestic and anthropophilic form known as *Ae. aegypti aegypti* (Bennett *et al*., 2016; Powell & Tabachnick, 2013; Salgueiro *et al*., 2019). It is this domesticated form that successfully invaded countries outside Africa and subsequently formed a monophyletic group (Bennett *et al*., 2016; Brown *et al*., 2014; Gloria-Soria *et al*., 2016). Slave trade in the 16th century might have played a major role in the introduction of *Ae. aegypti* into the Americas from Africa (Kotsakiozi *et al.*, 2018), and the vector subsequently spread globally to tropical and sub-tropical regions of the world (Brown *et al.*, 2014). Although the vector collections from Africa generally correspond to Aaf, it is speculated that East Africa might have experienced back migration of Aaa from outside Africa, leading to differentiation of *Ae. aegypti* (Kotsakiozi *et al*., 2018).

The domestication of *Ae. aegypti* has been hypothetically attributed to the formation of the Sahara desert, presuming that as the North of Africa became dry 4000-6000 years ago, populations along the Northern coast of the Mediterranean were separated from the sub-Saharan forest form, and this compelled the Northern populations to rely on water sources from human settlements, hence adaptation to civic environments (Kropelin *et al*., 2008; Tabachnick, 1991). This scenario is supported by the fact that in dry season, a female mosquito searching for an oviposition site to oviposit, would settle for stored water intended for human use (Powell *et al.,* 2018). This implies that when a female mosquito ecloses from such waters, it would also oviposit back in human stored

water in containers and the cycle would continue, with generations likely to evolve a taste for new food source, the closest being humans (McBride *et al*., 2014; Powell *et al*., 2018). This adaptation allowed for conditions on slave trade ships to select for the domestic type leading to gradual global spread and adaptation of *Ae. aegypti* to human environment (Powell & Tabachnick, 2013).

Populations of Aaa in the Americas were the first to branch of from the African ancestral populations during the transatlantic trade and consequently, reproductive isolation observed between the two forms of *Ae. aegypti* likely arose when they were separated from one another by the Atlantic ocean (Powell *et al*., 2018). With expanding urban centres and subsequent encroachment on natural forests, Aaf populations in sub-Saharan Africa are now breeding in human habitats, with larval preference shifting from tree holes to discarded items such as tyres, implying a change in egg-laying behaviour (Powell, 2016). Domesticity of Aaa is believed to have arose once (Brown *et al*., 2014) and the strong anthropophily exhibited by the domestic form that predominates outside of Africa is partly attributed to changes in an odorant receptor and the ability to lay eggs and survive in nutrient poor water (Bennett *et al*., 2016).

These events highlight the subspeciation of *Ae aegypti,* and the opportunistic and invasive nature of the vector. Proper research and documentation on variation of the domestic populations outside Africa (Bosio *et al*., 2005; Bracco *et al*., 2007; Brown *et al*., 2011; da Silva *et al*., 2012) exists, and there is still room for more research on the understudied range of the ancestral strain.

2.2 Vector Biology and Ecology

Considerable variations exist in terms of ecology and behaviour among *Ae. aegypti* populations found in sub-Saharan Africa (Brown *et al*., 2014). For instance, Aaf is a dark mosquito that breeds in tree holes and prefers non-human bloodmeals (Lounibos,

1981; Powell & Tabachnick, 2013; Tabachnick, 1991). On the other hand, the Aaa subspecies is a paler domestic form that exhibits strong genetic preference for entering houses to obtain bloodmeals through bites on humans, and an ability to survive in relatively clean water in man-made containers within households (Crawford *et al*., 2017; McBride *et al*., 2014). In modern day Africa, the ancestral form can be found in almost all ecologies including sylvatic habitats, cities, and peri-domestic habitats (Kotsakiozi *et al*., 2018; Powell & Tabachnick, 2013). With differences in oviposition choices, larvae of the two ecotypes is consequentially exposed to different bacterial communities, and this likely has influence on vectorial capacity of the two subspecies (Dickson *et al*., 2017). Larvae of the domestic form has been found in septic tanks (Barrera *et al*., 2008), a scenario that demonstrates the ecological flexibility of Aaa.

In spite of the absence of reproductive barriers between Aaa and Aaf strains in East Africa, with hybrids and backcrosses perfectly fertile (Powell & Tabachnick, 2013), the difference in choice of hosts makes Aaf populations exhibit lower vectorial capacity to transmit pathogens compared to Aaa populations (Powell & Tabachnick, 2013). Further, vector competence of Aaf has been observed to vary and some African populations of Aaf show competence for arbovirus transmission as those outside Africa (Kotsakiozi *et al*., 2018).

Previous research postulated that gene flow restriction between Aaf and Aaa is due to differences in their spatial distributions (Mousson *et al*., 2005). Hence, Aaa has been implicated in global transmission of dengue virus, while Aaf is credited with dengue virus transmission in sylvatic settings (Gubler, 1997). The two subspecies display hybridization in civic settings where they co-exist, and practice reproductive isolation in sylvatic setting (Redmond *et al.*, 2020). Vector control efforts aimed at *Ae. aegypti* have been largely inadequate and the mosquito's population has grown in urban areas resulting

in greater risk of humans coming into contact with competent vectors (Wilson & Chen, 2002). As a result, more than half of the world's population lives in areas where *Ae. aegypti* is present as shown in Figure 2.1. In addition, there have been reports of chromosomal inversions among subspecies of *Ae. aegypti,* and how these play a major role in adaptation and speciation of dipteran insects (Redmond *et al*., 2020). These chromosomal inversions lead to transference of distinct phenotypes across the subspecies and provide adaptation to *Ae. aegypti*, leading to successful invasion and global distribution of the vector (Moore *et al.,* 2013; Redmond *et al.*, 2020).

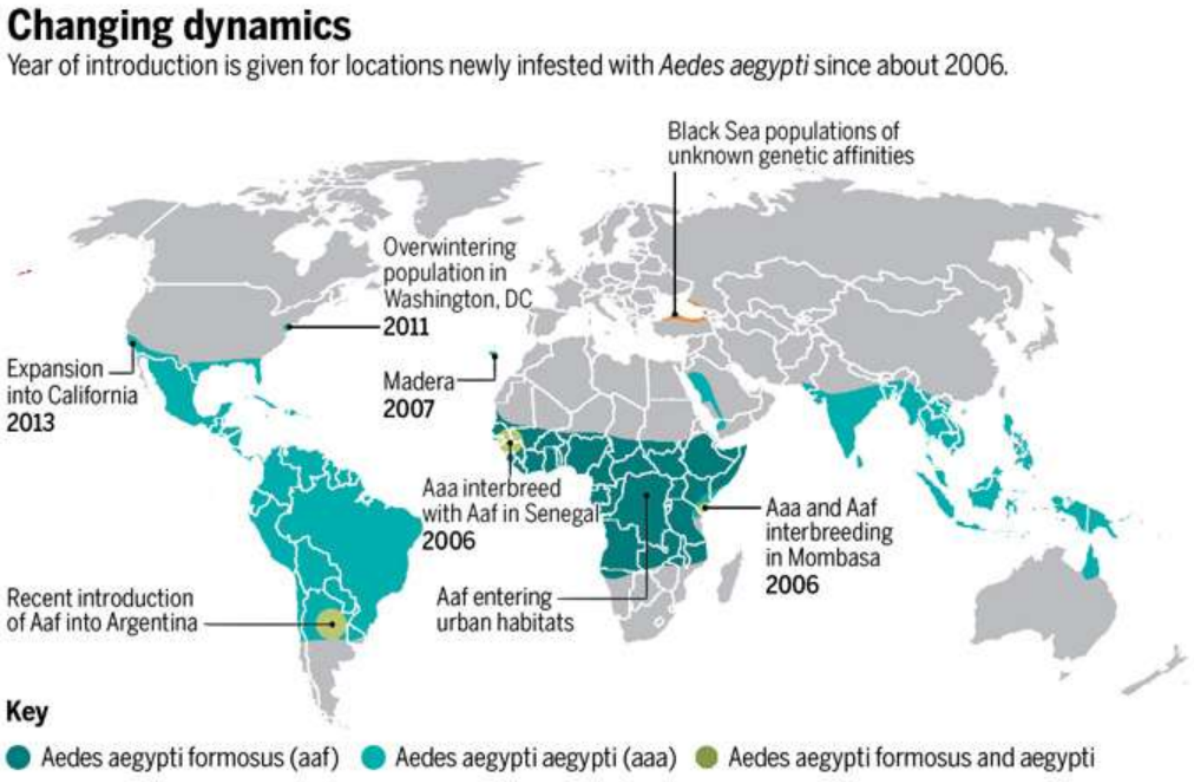


Figure 2.1. Global distribution *of Ae. aegypti* subspecies. Adapted from "Mosquitoes on the move," by Jeffrey R. Powell, 2016, *Science*, 354(6315), 971–972.

From Figure 2.1, it is clear that geographic distribution of the two subspecies has begun to overlap, hence the interbreeding populations in East and West Africa. Interbreeding could consequently lead to the two forms being ecologically

indistinguishable, especially with Aaf increasingly breeding in urban areas in Africa (Powell, 2016). *Aedes aegypti* populations are increasing outside Africa and now the vector can be found year round in non-native regions like Madeira (Almeida *et al*., 2007) and the Black Sea (Akiner *et al.*, 2016).

Aedes aegypti is granted a major boost in arbovirus pathogen transmission due to poorly planned urban structures especially in Africa, and the presence of water in containers that are not sealed among households which act as breeding sites (Baba *et al.*, 2016; Futami *et al.*, 2019). These anthropogenic breeding sites make control of the vector more challenging to stamp out dengue outbreak occurrence patterns and many lives remain at risk of contracting the disease since the vector is on the loose (Futami *et al*., 2019). Thus, efforts aimed at predicting spread patterns such as the Early Warning and Response System (EWRS) by WHO and Training in Tropical Diseases (TTD) proposed by Olliaro *et al*. (2018) are dealt a blow.

Viruses are known to evolve rapidly and have the potential to influence the genetic structures of host population by integration of viral nucleic acid sequences into germline cells, leading to inheritance of the viral integrations by subsequent generations (Blair *et al*., 2019). Viral DNA formation in *Ae. aegypti* occurs after infection of the mosquito by a virus (Nag & Kramer, 2017). Studies conducted on virus susceptibility of Aaa (Black *et al*., 2002; Diallo *et al*., 2005; Mitchell & Miller, 1991; Sylla *et al.*, 2009) have concluded that the strain likely became susceptible to dengue virus outside Africa and was reintroduced into Africa via human mediated transport (Powell & Tabachnick, 2013; Tabachnick, 1991). This observation is supported by often repeated outbreaks of dengue fever within and around African seaports (Futami *et al*., 2019). With the documented difference in susceptibility, it is crucial to understand the geographical distribution of *Ae. aegypti* strains to inform public health interventions. Dispersal of *Ae. aegypti,* has led to

change in arbovirus transmission since viruses are obligate intracellular pathogens and their transmission patterns at the ecological level can be influenced by specific species of vector or host involved in the transmission cycle (Huang *et al*., 2019). The focus of this study is on mitochondrial genes in order to understand *Ae. aegytpi* population genetics is fuelled further by (Paupy *et al*., 2008) who provided evidence of urban and forest African populations being genetically differentiated, but still remain within the larger Aaf genetic group.

2.3 Mitochondrial genomics

The insect mitochondrial genome is a circular molecule consisting of 37 genes that are highly conserved across metazoa, a characteristic that has led to the widespread use of insect mitochondrial genomes as a source of sequence data for phylogenetic analysis (Cameron, 2014). Mitochondrial genomes have played a major role in detecting evolutionary divergence in mosquitoes and understanding their dispersal history (Behura *et al*., 2011; Jones *et al.*, 2018; Kambhampati & Rai, 1991). These genes are advantageous in mosquito phylogenomics due to haploid maternal inheritance, rare occurrence of recombination, and higher mutation rate compared to changes in nuclear genomes (Jones *et al.*, 2018; Moriyama & Powell, 1997). Availability of multiple copies of mitochondria in the cell makes mitochondrial genomes amenable targets for sequencing at high coverage (Jones *et al*., 2018). Therefore, understanding historical gene flow and genetic structure via analysis of mitochondrial gene sequences is a key step towards revealing vector population dynamics and accurate discrimination between lineages.

Although research has shown considerable genetic variations among populations of *Ae. aegypti* (Gloria-Soria *et al*., 2016; Kotsakiozi *et al*., 2018), particularly their ability to transmit dengue viruses, the efforts have almost exclusively utilized genomic DNA (Parimittr *et al*., 2018). This warrants utilization of the mitochondrial genes *COI, COII,*

and *cyt b* for phylogenetic reconstruction of *Ae. aegypti* subspecies. The *COI* gene is the largest of the mitochondrial-encoded cytochrome *c* oxidase subunits and its protein sequence contains highly conserved functional domains and variable regions (Morlais & Severson, 2002). These features make *COI* a suitable gene for phylogenetic reconstruction of species and for documenting global biodiversity, through development of a DNA barcoding system for the animal kingdom based on *COI* gene (Buhay, 2009; Hebert *et al.*, 2003). There are plenty of published *COI* gene sequences in public repositories compared to other mitochondrial genes, and this makes it easier to compare sequences generated from different studies.

The evolutionary dynamics of *cyt b* gene and the biochemistry of its products have been well characterized by (Esposti *et al*., 1993), and *cyt b* is unlikely to be severely compromised by saturation effects involving perimposed nucleotide substitutions (Moritz *et al*., 1987), hence making it a phylogenetically informative gene. Mitochondrial *COII* gene sequence has also proven to be useful in accurately determining speciation and phylogenetic relationships (Wang *et al.*, 2016).

2.4 Incrimination of *Ae. aegypti* in arbovirus transmission

Aedes aegypti is regarded as the most dangerous animal in the arboviral world as it has been the single most important vector of the arboviruses yellow fever, dengue, chikungunya, and Zika viruses (Agha *et al.*, 2019; Brown *et al*., 2014; Kotsakiozi *et al*., 2018; Powell, 2016; Powell & Tabachnick, 2013; Rose *et al*., 2020). Today, *Ae. aegypti* is more of a public health concern as the primary vector of viruses that cause dengue fever, chikungunya and Zika (Musso *et al.*, 2015; Weaver & Lecuit, 2015). As mentioned earlier, this study seeks to understand how mitochondrial gene variation of *Ae. aegypti* strains correlates with global dengue outbreaks. Specific focus on dengue is driven by its health impact on a global scale (Conway *et al*., 2014), and from history, dengue outbreaks

have coincided with invasion and occurrence of *Ae. aegypti* in non-native regions (Smith, 1956; Urdaneta-Marquez & Failloux, 2011).

The dengue disease is caused by dengue viruses (DENV), which belong to the genus *Flavivirus* (family: Flaviviridae) and comprise four related serotypes, DENV1-4 (Baba *et al*., 2016) that cause millions of dengue hemorrhagic fever cases each year globally (Gubler, 2002). The capabilities of *Aedes aegypti aegypti* in transmission of pathogen have largely been influenced by vector abundance, anthropophilic blood feeding behaviour, and its indoor resting habits (Agha *et al*., 2017). Dengue has shown a 30-fold increase in global incidence in recent decades, affecting more than 100 countries throughout the tropical and subtropical regions of the world (Akiner *et al*., 2016). It is estimated that approximately 390 million infections occur yearly in the tropical and subtropical regions (Bennett, 2003; Bhatt *et al*., 2013).

More than 800 million people in Africa alone are at risk of dengue infection (Weetman *et al.*, 2018), and outbreaks of dengue in Africa indicate the likelihood that Aaf in urban areas preferentially choose humans for bloodmeals (Powell, 2016). From earlier studies (Bosio *et al*., 1998; Tabachnick *et al*., 1985), it is well known that the ancestral strain Aaf is less competent for arbovirus transmission compared to Aaa. But with more studies on mosquito arbovirus competence (Diallo *et al*., 2008; Dickson *et al*., 2014; Vazeille *et al*., 2013), there is increasing evidence that vector competence of Aaf varies considerably and is population -specific with some African Aaf populations being as competent as those outside Africa (Kotsakiozi *et al*., 2018).

Europe has experienced sporadic cases of locally acquired dengue cases, with Madeira Island having outbreaks in 2012-2013 (Akiner *et al*., 2016). France also reported dengue cases (Schaffner *et al*., 2014), which demonstrate Europe's vulnerability to

dengue. These cases in Europe have largely been driven by viraemic travellers and invasion of the continent by *Ae. aegypti* (Akiner *et al*., 2016; Langat *et al*., 2020). With advances in technology, mathematical models have estimated that approximately 50 million infections occur yearly in more than 100 countries in the Asia-Pacific region (Solórzano *et al*., 2010). In noting these statistics down, it is important to remember that these represent human lives globally, and therefore control of dengue transmission is a major 21st century challenge for global health. This challenge is exacerbated by widespread insecticide resistance by *Ae. aegypti* (Moyes *et al*., 2017) and relentless growth in urban environments (Alirol *et al*., 2011). Efforts aimed at tracking transmission of these outbreaks have been hampered by limited genomic data (Giovanetti *et al.*, 2019), and therefore, an approach such as phylogenetic analysis of mitochondrial genes of arbovirus vectors may be helpful to understand the distributional profile of *Ae. aegypti* and dengue outbreak occurrence patterns.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Review of bioinformatics tools used

3.1.1 Barcode of Life Data portal (BOLD)

DNA barcoding is a genotyping technique that aids in identification of organisms by assessing their degree of DNA sequence similarity in comparison to their specific references (Virgilio *et al.*, 2010). The Consortium of Barcoding of Life (CBOL) initiative is tasked with coordinating the collection of DNA barcodes and building a worldwide reference database (http://www.barcodinglife.org). The Barcode of Life Data portal (BOLD) is a freely available bioinformatics database that enables querying, storage, analysis, and publication of DNA barcode sequences (Ratnasingham & Hebert, 2007). BOLD has a Management and Analysis System (MAS) that provides an interface for quality assurance by reporting sequence anomalies that may signal presence of a pseudogene, contamination or kingdom level mismatches (Ratnasingham & Hebert, 2007).

The sequence export function in BOLD generates a FASTA file for exporting sequence records for further analysis by a user. In addition, BOLD includes a search functionality that allows users to ascertain sequences for a particular taxonomic group in a specified geographic region (Ratnasingham & Hebert, 2007). These features make BOLD an excellent and reliable source for retrieval of biological sequence data.

3.1.2 High Performance Computer (HPC)

High Performance Computer (HPC) is a parallel and distributed system, which consists of a collection of interconnected stand-alone computers working together as a single integrated computing resource to provide the functionality of supercomputing (Al-

Khazraji *et al*., 2014). The computers are interconnected via a local area network and a user interacts with a master node and submits tasks to it. The master node is the controlling node, with a job scheduler that assigns priority to tasks submitted. For instance, after several tasks have been submitted by different users, the job scheduler assigns order and priority according to how the tasks were submitted and divides the tasks amongst the computer clusters for computation. After the computations of the tasks from each cluster are done, the results are unified & returned back to the master node and the user can see their results (Al-Khazraji *et al*., 2014; Kahanwal & Singh, 2013).

3.1.3 Sanger Sequencing

Sanger sequencing is a method described by (Sanger *et al*., 1977) for determining the sequence of nucleotide bases in a piece of DNA. The method is used to determine sequences less than 1000 base pairs (bp). Sanger sequencing makes several copies of the target DNA region using the following ingredients:

i. DNA polymerase
ii. Template DNA to be sequenced
iii. Single stranded DNA primer that binds to the template DNA
iv. Four DNA nucleotides (dATP, dTTP, dCTP, dGTP)

During the sequencing process, DNA-dependent DNA polymerase is used to generate a complimentary copy of a single DNA template sequence (Sanger, 1988). Starting at the 3′ end, a new chain of primer DNA complementary to a single-stranded template DNA strand is synthesized and the deoxynucleotides are added to the growing chain, and these are complementary to the nucleotides in the template DNA (Shendure *et al.*, 2011). Sanger sequencing leverages the technology that when DNA polymerase incorporates a chain terminating 2′,3′-dideoxynucleotide monophosphate (ddNMP) at the appropriate

complementary position, sequence synthesis will be stopped because the incorporated nucleotide lacks the required 3′ hydroxyl group for formation of a phosphodiester bond (Sanger *et al*., 1977).

Four reactions, containing template, polymerase, all four dNTPs (one radioactively labelled), and primer, are set up to generate a continuing series of synthesis products that reflects each potential chain termination position. In addition, each reaction also contains one of the four ddNTPs at a specific ratio reflecting the relative probability of incorporation. Many terminated strands of different lengths exist within each of the four reactions. As each reaction contains only one ddNTP species, a set of different-length fragments is generated in each reaction, terminated at all of the positions corresponding to one of the four nucleotides in the template sequence. The four reactions are then individually separated on a large denaturing polyacrylamide gel to yield single-nucleotide resolution. The pattern of bands across the four lanes allows direct readout of the primary sequence of the template under analysis (Sanger *et al.*, 1977; Shendure *et al.*, 2011).

3.1.4 Basic Local Alignment Search Tool (BLASTn)

BLASTn is an algorithm for comparing nucleotide sequence information. It allows a user to compare a query nucleotide sequence with a database of sequences, and the program identifies sequences in the database that are similar to the query sequence above a certain threshold. BLASTn achieves local sequence similarity check by seeking only relatively conserved subsequences, and a single query may yield distinct subsequence alignments (Al-Khazraji *et al*., 2014). For similarity measures, BLASTn begins with a matrix of similarity scores for all conservative residues, while the unlikely replacements in a query have negative scores (Altschul *et al*., 1990). The overall BLAST program implements 3 major algorithmic steps involving compilation of a list of high scoring residues, scanning the database for hits and extending hits (Altschul *et al*., 1990). These

critical steps permit the fast execution of BLASTn making it easier to conduct local sequence similarity searches.

3.1.5 Multiple Alignment using Fast Fourier Transform (MAFFT)

Multiple sequence alignment (MSA) is a key step in molecular biological analyses involving detection of key functional residues and inferring the evolutionary history of a set of organism sequences. The MAFFT method is highly efficient and allows for rapid detection of homologous segments in nucleotide sequences, and is advantageous over other multiple sequence aligners in that it incorporates a simplified scoring system that performs well by drastically reducing CPU time and increases the accuracy of alignments even for sequences having large insertions or extensions (Katoh, 2002). With support for parallel processing and multithreading, MAFFT can easily be used to execute MSA of a large number of sequences on high performance computer clusters (Katoh & Standley, 2013). These merits highlight the choice of MAFFT as the most suitable alignment program for use in this study.

3.1.6 Geneious Prime software

Geneious is an easy-to-use and flexible desktop application software that provides a bioinformatics platform for organization and analysis of biological data using computational resources (Kearse *et al.*, 2012). Geneious has an interactive graphical user interface that makes it easier to visualize sequences and even generate publication quality images, plus it is a flexible software compatible with multiple bioinformatics file formats ranging from the simple fasta file format to more complex binary file formats generated by different sequencing platforms (Kearse *et al.*, 2012).

3.1.7 Randomized Axelerated Maximum Likelihood (RAxML)

Randomized Axelerated Maximum Likelihood (RAxML) is a program for phylogenetic analysis of large sequence datasets using the maximum likelihood principle

(Stamatakis, 2014). The program implements the standard non-parametric bootstrap and also offers an option to calculate the SH-like support values (Guindon *et al*., 2010). RAxML provides superior functionality over other phylogenetic analysis software by implementing parsimony and maximum likelihood flavors (Stamatakis, 2014) adapted from the evolutionary placement algorithm that places short reads into a given reference phylogeny obtained from full-length sequences to determine the evolutionary origin of the reads (Berger *et al*., 2011). Additionally, RAxML provides a functionality that enables testing for bootstrap convergence of the phylogenetic tree obtained (Kozlov *et al.*, 2019).

Maximum likelihood is a computationally feasible method that involves finding an evolutionary tree which yields the highest probability evolving the observed data (Felsenstein, 1981). The maximum likelihood method is superior to the parsimony method in that it incorporates a probabilistic model by assuming the independence of evolution at different sites, and therefore the probability of a given data set arising on a given tree can be computed site by site, and the product of the probabilities is taken across sites at the end of the computation (Felsenstein, 1981). This is a far much better approach than the parsimony method which assumes that change is improbable a priori (Felsenstein, 1973). The latter could only be justified if the amount of change was small over the evolutionary period under consideration (Felsenstein, 1981).

3.2 Samples used in the study

This study had mosquito samples obtained from different sources. First, it involved archived samples that had been collected from Kilifi County (Rabai), two months to the start of this study. These samples had been stored at -81°C. The study also utilized archived DNA of mosquito samples that were collected in Kakamega, Bungoma and Busia Counties, and stored in the same conditions for a period of two years. Sequences of samples from Narok County (The Maasai Mara National reserve) and

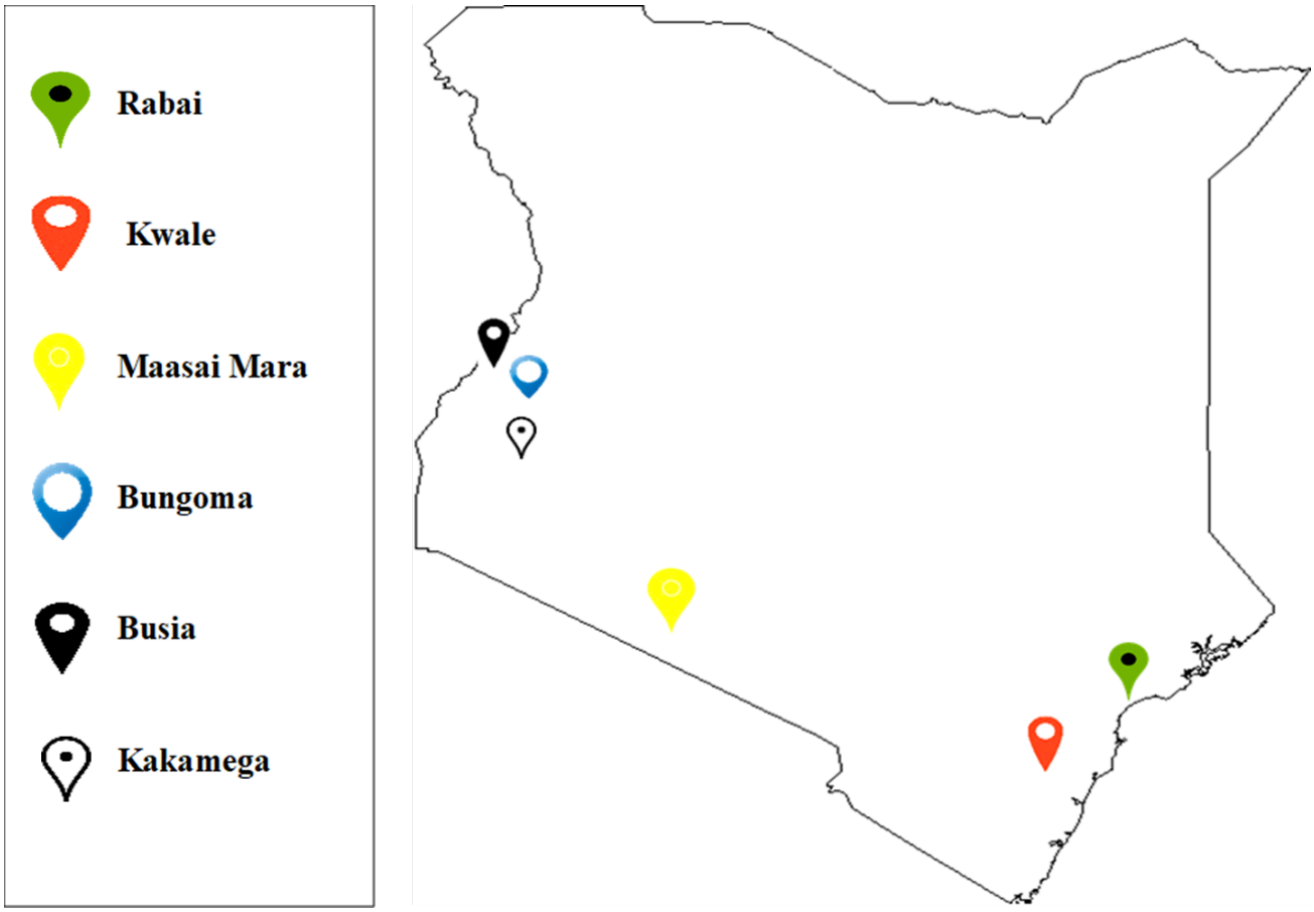
Kwale County (Shimba Hills National reserve) were obtained from GenBank. Moreover, additional sequences and sequences representing strains from outside Africa were retrieved from published sequences hosted in both BOLD and GenBank databases. In total, 1230 *Ae. aegypti mt*DNA sequences (*COI, COII* and *cyt b*) were retrieved for this study.A summary of the mosquito sample breakdown is as provided in Table 3.1. Further breakdown of the sequences by country is provided in Table 3.3 (appendix).

Additional *COII* and *cyt b* sequences from outside Africa were retrieved from GenBank by conducting a BLASTn query search with sequences generated from the laboratory at *icipe*. Default BLASTn program parameters were used, and the highly similar option megablast was selected. *COI*-5P barcode sequences representing East Africa and continents outside Africa were retrieved from the Barcode of Life Data portal (BOLD) (Ratnasingham & Hebert, 2007). BOLD has a resource tab which houses the BOLD application programming interface (BOLD API) that enabled easier data retrieval using specific search criteria. The web service endpoint option selected was sequence data retrieval. The search criteria used included taxon, geo and marker, with respective search terms being *Aedes aegypti*, countries in East Africa, Asia, Australia, North and South America, and Europe, and *COI*-5P. The query returned a URL (https://www.boldsystems.org/index.php/api_urlgenerator/) that was used to download *Ae. aegypti COI* sequence data.

Table 3.1. Number of sequences for each of the mitochondrial genes investigated.

Source	**Marker gene**			
	COI	*COII*	*cyt b*	Total sequences
Archived adult female mosquitoes from Rabai in Kilifi County	41	20	15	76
DNA from mosquitoes collected in Kakamega, Bungoma, and Busia	18	9	9	36
Narok (Maasai Mara) and Kwale counties (Available in GenBank)	10			10
BOLD Database	1068			1068
Additional *COII* and *cyt b* sequences from GenBank		22	18	40
Total Number of sequences				1230

The map of Kenya (Figure 3.1) shows areas from which the archived samples had been collected. Rabai and Maasai Mara represent Kilifi and Narok counties, respectively. Sequences from Kwale and Maasai Mara are samples that had been collected from sylvatic settings while those from Bungoma, Busia, Kakamega and Rabai, are samples that were collected from domestic settings.



https://www.google.com/maps/

Figure 3.1. Map of Kenya showing sites where archived samples were collected.

3.3 Laboratory sample processing

Molecular biology assays involving PCR amplification were conducted on archived samples from Rabai in Kilifi County and on DNA for mosquito samples collected from Kakamega, Bungoma, and Busia counties. These archived samples had been stored at -81°C as mentioned earlier. Figure 3.2 summarises steps that were followed.

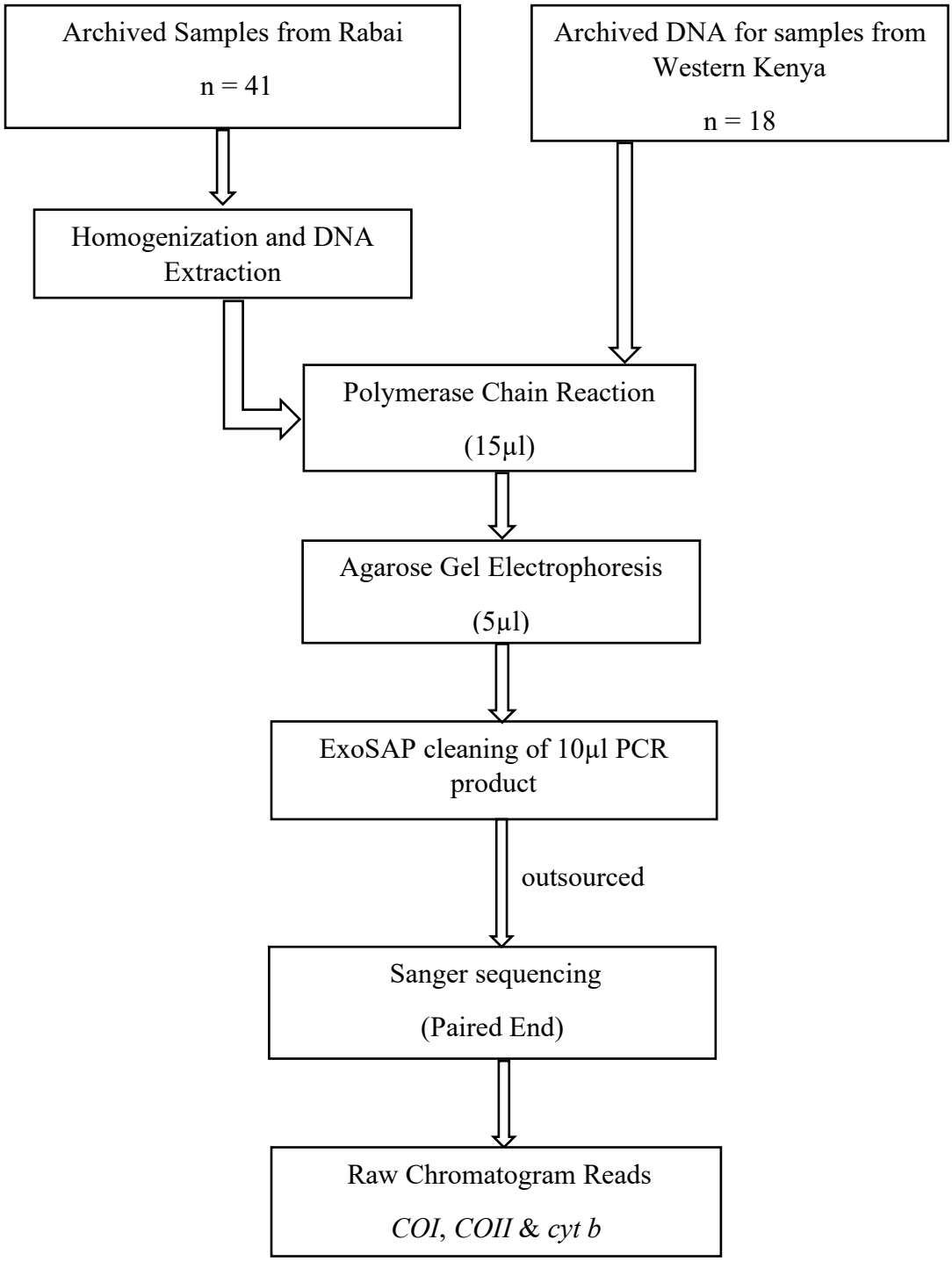


Figure 2.2. Flowchart of molecular assays carried out in the laboratory.

DNA was extracted from the body homogenate of 41 individual adult female *Ae. aegypti* mosquitoes from Rabai in Kilifi County, using an in-house protein precipitation (PPS) protocol for DNA extraction from whole flies. This protocol is provided in the appendix.

COI gene sequences were amplified in two parts with a final length of 1100 bp in 15-µl two-cycle PCR reactions containing 0.67 µM LCO1490 and HCO2198 primers (Table 3.2), 2x MyTaq HS Mix (Bioline), and nuclease free PCR water. The PCR conditions included an initial denaturation step at 95°C for 5 minutes followed by 5 cycles of 40 seconds at 95°C, 1 min at 45°C and 1 min at 72°C. This was then followed by 35 cycles of 40 seconds at 95°C, 1 min at 51°C and 1 min at 72°C, and 7 min elongation at 72°C. The final hold was at 4°C. Similar PCR conditions were used to amplify *COI* UEA3 and Fly 10 primers (Table 3.2).

COII gene sequences were amplified in a multiplex 15-µl PCR reaction containing 0.67 µM of the primers SCTL2-J-3037, SC1-TD-N-3862 and TK-N-3785 (Table 3.2). The PCR conditions included an initial denaturation step at 95°C for 5 minutes followed by 35 cycles of 95°C for 1 min, 1 min at 52°C, 2 min at 72°C, and 7 min elongation at 72°C. The final hold was at 4°C.

Table 3.2. Published primers used for PCR amplification and sequencing.

Gene	Primer Name	Primer Sequence (5′ to 3′)	Amplicon size (bp)	Reference
COI	LCO 1490	GGTCAACAAATCATAAAGATATTGG	~ 710	(Folmer *et al.*, 1994)
	HCO 2198	TAAACTTCAGGGTGACCAAAAAATCA		
	UEA3	TATRGCWTTYCCWCGAATAAATAA	~ 1250	(Cook *et al.*, 2006)
	Fly10	ASTGCACTAATCTGCCATATTAG		
COII	SCTL2-J-3037	ATGGCAGATTAGTGCAATGA	~ 800	*(*Cook *et al.*, 2006)
	SC1-TD-N-3862	TTTAGTTTGACATACTAAKGTTAT		
	TK-N-3785	GTTTAAAGACCAGTACTTG		
cyt-b	Cyt-b J1	TCTTTTTGAGGAGCWACWGTWATTAC	~ 438	(Belshaw & Quicke, 1997)
	Cyt-b N	AATTGAACGTAAAATWGTRTAAGCAA		

Cytochrome b (*cyt b*) gene sequences were amplified in 15 µl PCR reaction containing 0.67 µM *cyt b* J1 and *cyt b* N primers (Table 3.2), 2X MyTaq HS Mix (Bioline), and nuclease free PCR water. The PCR conditions included an initial denaturation step at 95°C for 5 min followed by 40 cycles of 1 min at 95°C, 1 min at 45°C and 2 min at 72°C, and a final 7 min elongation at 72°C. The final hold was at 4°C.

Five microliters of the PCR products for *COI*, *COII*, and *cyt b* genes were size fractionated by electrophoresis for 30 minutes at 300 A/130V in 1.5% ethidium bromide-stained agarose gel in Tris-borate ethylenediaminetetraacetic acid (EDTA) buffer. The DNA bands were visualized under ultraviolet light in a transilluminator. Ten microliters of the remaining *COI, COII,* and *cyt b* amplicons were purified using ExoSAP (New England BioLabs inc. Ipswich, Massachusetts, USA). Purified *COI* amplicons were outsourced for sequencing in both directions, while half of the amplicons for *COII* and *cyt b* were sequenced in the forward direction only and the other half in the reverse direction at Macrogen (Netherlands). The same primer sequences were used for amplification and sequencing.

3.4 Quality control and Chromatogram assembly

Raw mitochondrial *COI*, *COII*, and *cyt b* sequences were analysed in Geneious Prime software version 2020.1.2 (created by Biomatters) (Kearse *et al*., 2012). Low quality reads with more than 2% chance of an error per base were trimmed from the sequences. Primer sequences were removed from the 5′and 3′ ends. Forward and reverse sequences were pairwise aligned and assembled to produce consensus sequences. The cleaned sequences were queried against the nucleotides in GenBank database using the Basic Local Alignment Search Tool (BLASTn). Geneious mapper (created by Biomatters) (Kearse *et al*., 2012), was used to map the sequences to a reference *Ae.*

aegypti mitochondrial genome accession number NC_035159, which is 16,790bp. Figure 3.3 summarises the processes in downstream analysis.

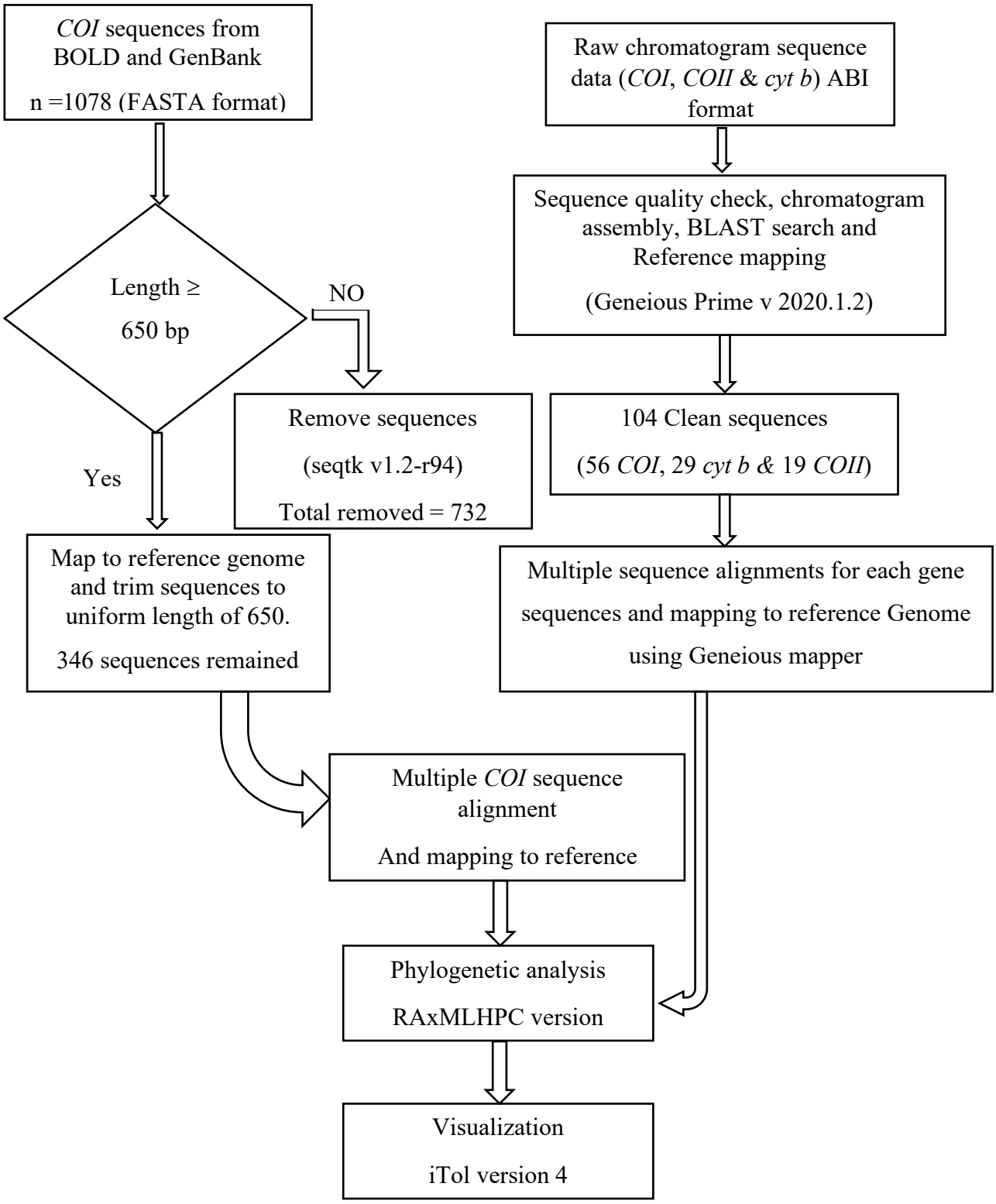


Figure 3.3. Flowchart showing analyses conducted on all sequence data.

COI sequences from BOLD and GenBank were subjected to quality check to ensure they conform to the desired length of 650 bp. For each gene, multiple sequence alignments were conducted before being subjected to phylogenetic analysis in

RAxMLHPC on a high-performance computer. The maximum likelihood trees obtained were then visualized in iTol version 4. The total number of sequenced samples utilized for downstream analysis after performing quality check and chromatogram assembly was 56 *COI,* 29 *cyt b* and 19 *COII* sequences. In total, 104 sequences were generated from the laboratory. 5′ end *COI* sequences, 1068 in number, were downloaded from BOLD (Ratnasingham & Hebert, 2007). These represented three East African countries (Kenya, Uganda, and Tanzania), North and South America, Asia, Australia, and Europe. Sequences less than 650 bp were removed from the downloaded sequences since the *COI* barcoding region covers ~650 base pairs (bp) from the 5' end. Consequently, available sequences of other East African countries which were less than 650 bp were also removed. As a result, 337 sequences out of the downloaded 1068 were retained. These were further mapped to a reference genome and trimmed to a uniform length of 650 bp and used in subsequent multiple sequence alignments and phylogenetic analyses. *COI* sequences of samples from Narok (Maasai Mara), and Kwale are from a published study (Musa *et al*., 2020). For Narok (Maasai Mara) and Kwale counties *COI* sequences from GenBank, only one sequence less than 650 bp was removed, and the remaining nine were also mapped to a reference genome and trimmed to a uniform length of 650 bp. For the overall downstream phylogenetic analyses, this study utilized 490 sequences comprising 402 *COI,* and 47 *cyt b* and 41 *COII* sequences, respectively.

3.5 Phylogenetic analyses

In line with the study objectives, phylogenetic analysis was conducted on:

i. *COI* sequences of samples from domestic (Busia, Bungoma, Kilifi, and Kakamega Counties) and sylvatic (Kwale and Narok Counties) settings.
ii. *COI, COII*, and *cyt b* sequences from East Africa compared against global sequences of the anthropophilic strain.

3.5.1 Comparison of *COI*, *COII* and *cyt b* sequences

Multiple sequence alignments (MSA) for each of the gene sequences were generated using MAFFT plugin v7.450 (Katoh & Standley, 2013) in Geneious Prime v 2020.1.2 (Kearse *et al*., 2012). To assess evolutionary relationship among strains of *Ae. aegypti*, phylogenetic trees were reconstructed on the basis of maximum likelihood (ML) in Randomized Axelerated Maximum Likelihood (RAxMLHPC) version 8.2.12 (Stamatakis, 2014). The general time reversible (GTR) model of evolution was used along with the gamma model rate of heterogeneity (Yang, 1996), which was determined as the best-fit model using ModelFinder (Kalyaanamoorthy *et al*., 2017) implemented in IQ-TREE multicore version 1.6.1 (Minh *et al.*, 2020). This substitution model was determined based on Bayesian Information Criterion (BIC) (Cavanaugh, 1997). Bootstrap branch support of 5000 was used to obtain the trees and bootstrap convergence implemented in RAXML-NG (Kozlov *et al*., 2019) was used to check convergence and stability of the best maximum likelihood tree obtained. *Aedes albopictus* was employed as the outgroup species. Maximum likelihood trees for each of the gene sequences were visualized and edited in Interactive tree of life (iTol) v4 (Letunic & Bork, 2019).

CHAPTER FOUR

4.0 RESULTS

4.1 Assembly of *COI* sequences from different primer sets

COI sequences from the different primer sets were pairwise aligned and assembled to form uniform contiguous sequences of 1100 bp. These sequences covered different portions of the reference genome and overlapped. For instance, *COI* sequences from LCO 1490 and HCO 2198 primers (Table 3.2) cover the *COI*-5′ end of the reference genome starting from ~1340 bp to ~1996 bp. *COI* sequences from the primers UEA-3 and Fly-10 (Table 3.2) cover the reference genome starting from ~1632 bp to ~ 2737 bp. Figure 4.1 shows overlapping of the different primer set *COI* sequences when mapped to a reference *Ae. aegypti* mitochondrial genome accession number NC_035159.

Figure 4.1. Overlapping *COI* sequences from different primer sets mapped to an *Ae. aegypti* reference mitochondrial genome.

4.2 Phylogenetic analyses

4.2.1 Comparison of *COI* sequences of *Ae. aegypti* mosquitoes sampled in domestic vs sylvatic settings

Maximum likelihood tree (ML) (Fig 4.2) shows sequences of samples from domestic settings in Kakamega, Bungoma, Busia, and Kilifi (Rabai), and from sylvatic settings in Narok (Maasai Mara), and Kwale counties. The ML tree resolves *Ae. aegypti* into 3 lineages, with the first lineage having sequences of samples from Kilifi (Rabai), Bungoma, Kakamega, Busia and Narok (Maasai Mara) counties clustering with the reference sequence representing Aaf. The second lineage consists of sequences from all the sampling locations within Kenya clustering together with the reference GenBank accessions of *Ae. aegypti aegypti* strain. Sequences of samples from sylvatic settings (Kwale and Maasai Mara) cluster with the referenced Aaa sequences in lineage 2, while one sequence from Maasai Mara clusters with the referenced sylvatic strain sequence in lineage 1. Lineage 3 consisted of sequences of mosquito samples from Rabai only.

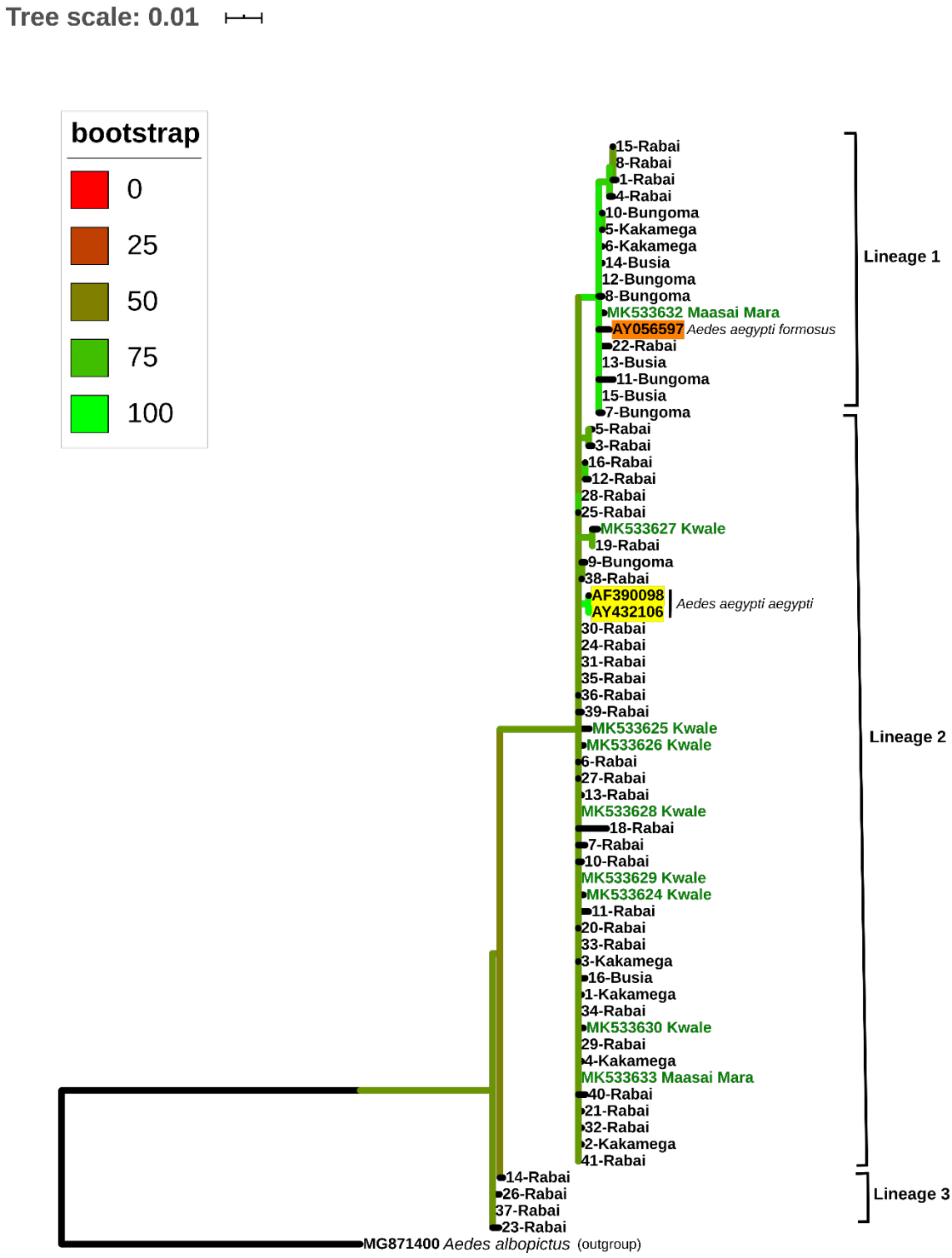


Figure 4.2. Maximum likelihood phylogenetic tree of *COI* sequences of *Ae. aegypti* mosquitoes from sylvatic and domestic settings in Kenya. Sequences from sylvatic settings (Maasai Mara and Kwale) are represented in green. GenBank reference of Aaf is highlighted in orange while Aaa references are highlighted in yellow. Aedes albopictus was used as the outgroup. The branches are colored according to bootstrap measure

4.2.2 *COII* sequences from East Africa in comparison to global anthropophilic strains

Phylogenetic tree of *COII* sequences showed sequence accessions from Senegal clustering with sequence accessions from Puerto Rico and India. Sequences of samples from Rabai in Kilifi County formed a cluster of their own, and also with the other global sequences. One sequence accession from Senegal appeared to form its own lineage.

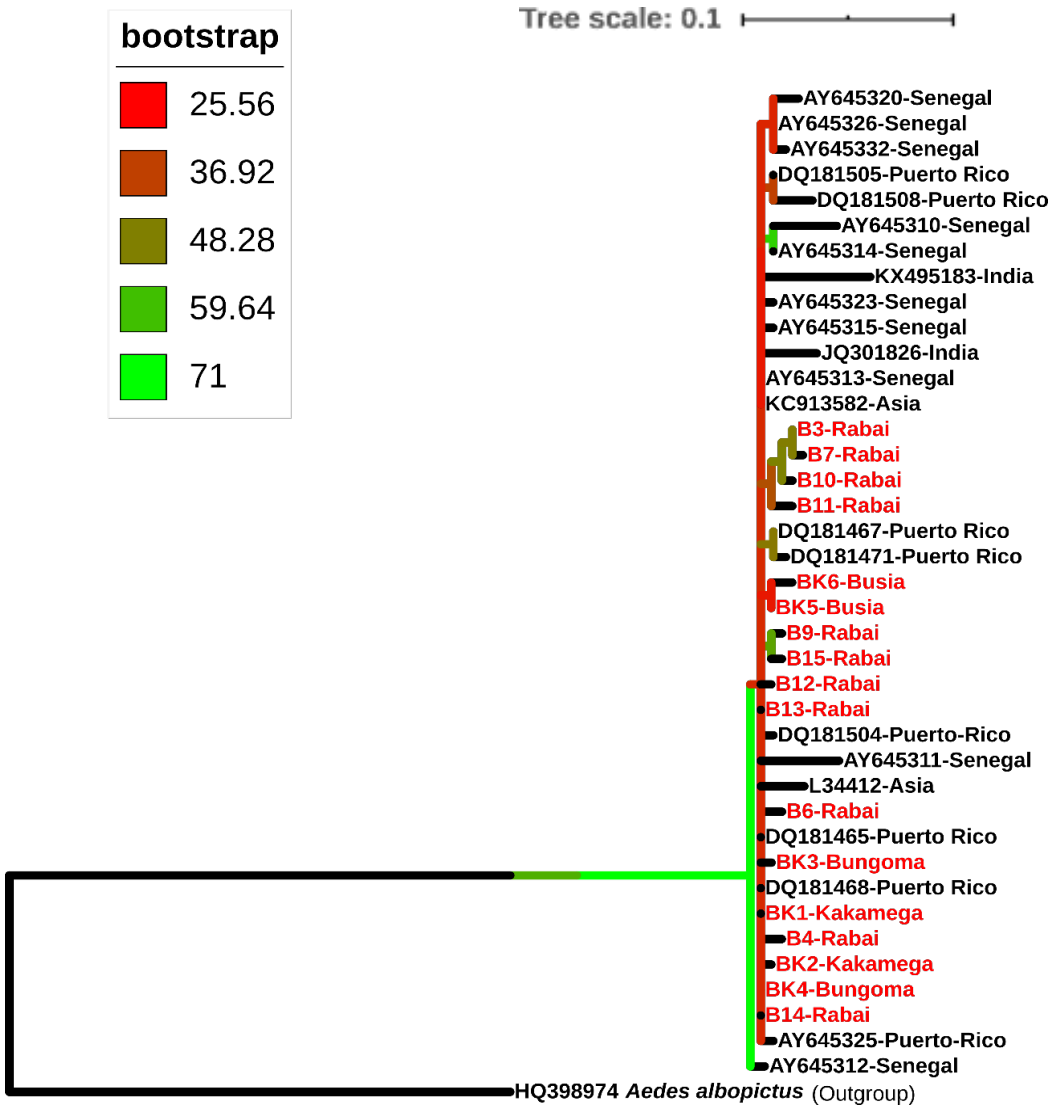


Figure 4.3. Maximum likelihood phylogenetic tree of *COII* sequences from Kenya (red) and globally available sequences in GenBank. The tree branches are coloured by bootstrap as shown. *Ae. albopictus* was used as outgroup.

4.2.3 *cyt b* sequences from East Africa in comparison to global anthropophilic strains

Phylogenetic analysis of *cyt b* sequences resolves *Ae. aegypti* into two distinct lineages. One lineage is composed of *cyt b* sequences generated from this study clustering together while the other lineage mainly consists of sequences from outside Africa. One sequence from Brazil clusters together with sequences generated from this study.

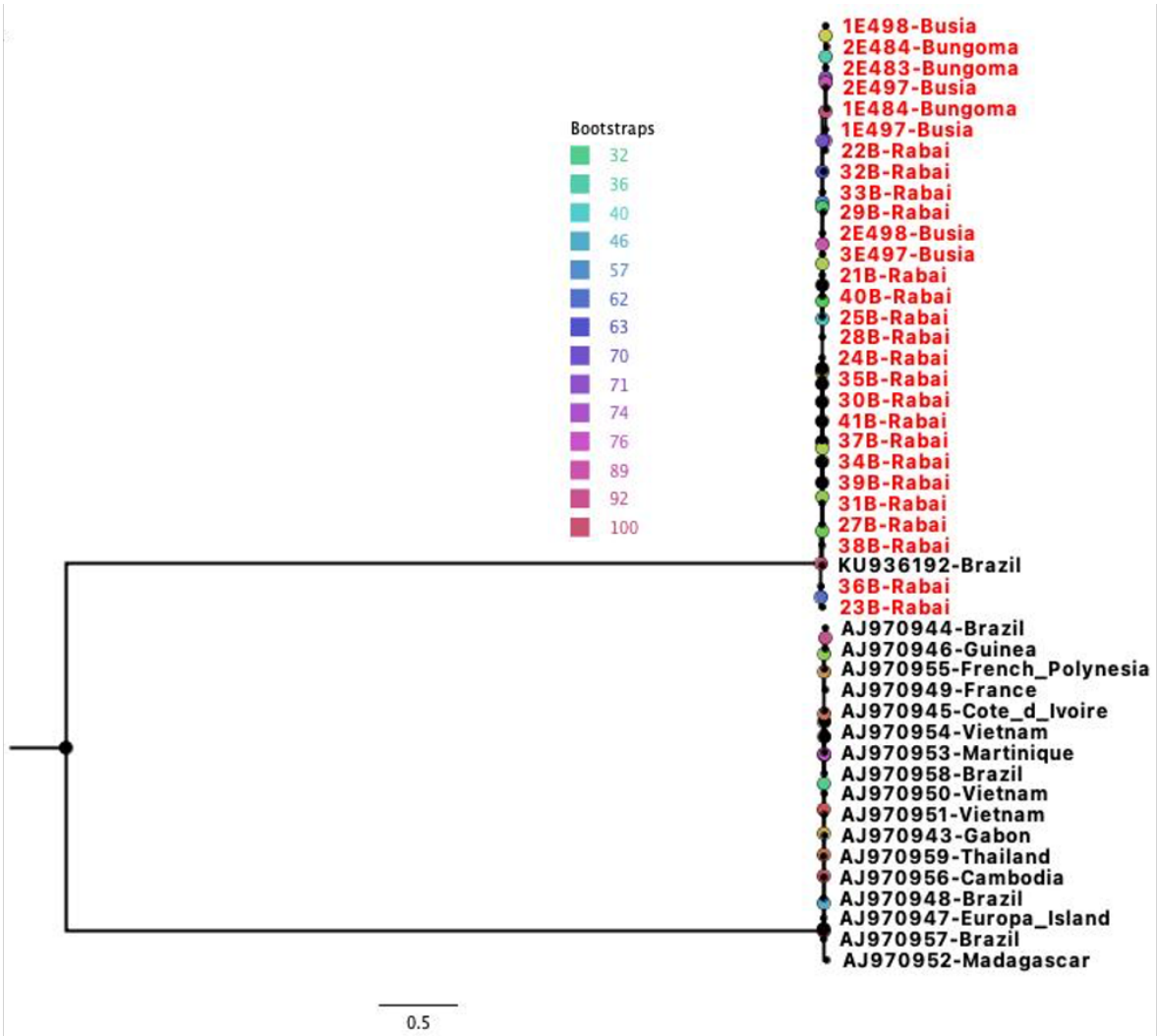


Figure 4.4. Maximum likelihood phylogenetic tree of *cyt b* sequences rooted at midpoint. Sequences generated from this study are labelled in red, while the rest are from other countries as shown.

4.2.4 East African *Ae. aegypti COI* sequences compared against global anthropophilic strain

COI sequences of *Ae. aegypti* from Asia, North and South America cluster with the GenBank reference of *Ae. aegypti aegypti* (Aaa) (Fig 4.5). North and South America sequences show varied patterns where they cluster together and also cluster with sequences from Kenya and Asia. South America sequences again cluster with the *Ae. aegypti formosus* Genbank reference. The Tanzanian sequence clusters with sequences from Kenya (Kilifi County) while the Uganda sequence clusters with sequences from Asia and South America. A majority of sequences from Asia show mixed clustering with the rest of the sequences, with some forming clusters with Kilifi County (Rabai) sequences. Sequences from Europe and Australia all cluster together with the Asian sequences.

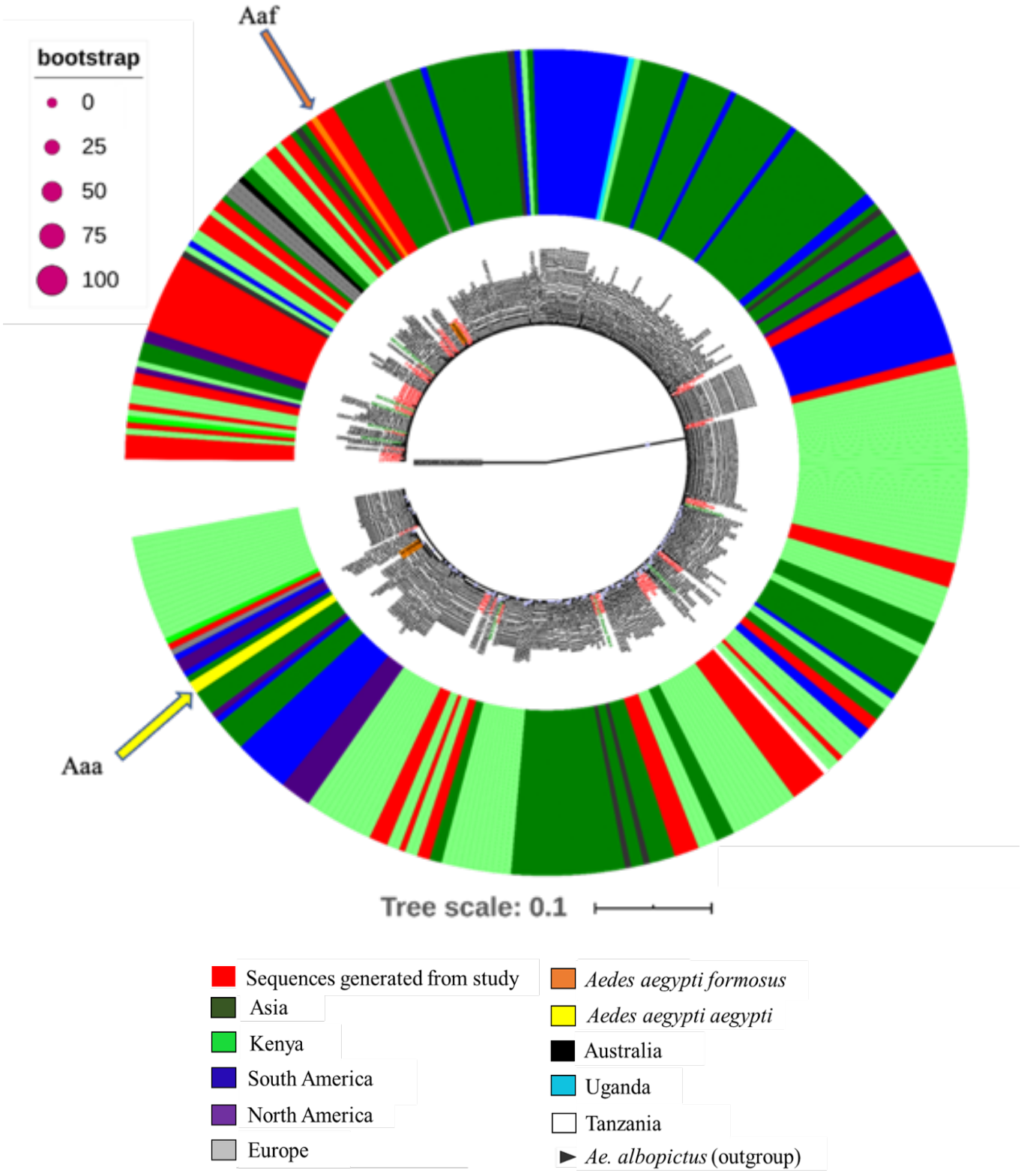


Figure 4.5. Maximum likelihood phylogenetic tree of *Ae. aegypti COI* sequences from East Africa compared against global anthropophilic strains. GenBank references of Aaf and Aaa are shown on the tree.

CHAPTER FIVE

5.0 DISCUSSION

5.1 Mitochondrial genomics

In this study, mosquito samples that were processed using molecular laboratory assays produced *COI* sequences covering different important segments of the *Ae. aegypti* mitochondrial genome. The universal primers by (Folmer *et al*., 1994) generated *COI* sequences covering the standard barcode region of ~ 650 bp. These sequences were elongated and complemented by sequencing using the *COI* primers described further by (Cook *et al*., 2005). Sequences from these two primer sets overlap when mapped to a reference genome (Fig 4.1) and together they were combined to form much longer contiguous sequences of ~ 1100 bp. One keen observation from these two sequences is that based on BLASTn searches using the highly similar option (megablast), sequences comprising lineage 1 as shown in Figure 4.2, give 99% BLAST identity to both Aaf and Aaa at 1100 bp, while the same sequences at 650 bp still showed 99% identity, but only to Aaa. The observed percentage identities not only provide caution for reliance on barcode region sequences as discussed in-depth by (Beebe, 2018), but also likely supports the observation that there is better differentiation of the two *Ae. aegypti* subspecies with longer *COI* sequences compared to just the standard barcode region. However, this assertion requires further elucidation as there exists only one defined Aaf *COI* sequence available in GenBank, and most of the published *Ae. aegypti COI* sequences in public repositories such as BOLD (Ratnasingham & Hebert, 2007) are not classified to the subspecies level.

Contrary to the observation that *cyt b* is a poor resolution marker for answering deep evolutionary questions (Meyer, 1994), data here (Fig 4) showed the potential for *cyt*

b utilisation in population genomics. From a global perspective, *cyt b* sequences seemingly placed *Ae. aegypti* into two major lineages, an observation that supports earlier studies which showed that *Ae. aegypti* resolves into two major clades (Bracco *et al.*, 2007; Brown *et al*., 2011). Additionally, *COII* gene sequences appeared to invoke the classic discussion of West African Senegalese populations of *Ae. aegypti* being different from the rest of African populations. Exception to this difference being populations of *Ae. aegypti* from domestic settings in Rabai in East Africa (Gloria-Soria *et al.*, 2016; Powell, 2016).

5.2 Comparison of sylvatic and domestic *Ae. aegypti* populations

Clustering of samples of Aaa collected from domestic settings with Aaf (Lineage 1, Fig 4.2) suggests that there could be cohabitation of the two *Ae. aegypti* forms in domestic settings in the sampled areas, an observation that is consistent with the findings of Gloria-Soria *et al*. (2016). This cohabitation phenomenon may also imply that there is independent colonization of human habitats by Aaf as reported by Brown *et al*. (2011) and Gloria-Soria *et al*. (2016). A huge mix of sequences from the different sample site sources within Kenya is observed in lineage 2 (Fig 4.2) that clustered with the Aaa GenBank reference sequences. Given that *Ae. aegypti* samples from Kwale and Narok counties came from sylvatic settings and the rest from domestic settings within Kenya, the observed clustering likely suggests ecological expansion of Aaa. This ecological expansion could have been caused by anthropogenic impacts on natural habitats and animal food resource (Brown *et al.*, 2014). A pointer to this suggestion is the peri-sylvatic settings from which Kwale County samples had been collected by Musa *et al.* (2020). A linkage to this ecological expansion is the fact that international tourists especially in ships from all over the world frequent Narok (Maasai Mara National Reserve) and Kwale (Shimba Hills National Reserve). Further, harbour activities attract ships globally for

trade, and together these activities may be pathways that facilitate *Ae. aegypti* introductions/ reintroductions and spread in the coastal Counties.

The third lineage (Fig 4.2) raises several questions regarding population genetics of *Ae. aegypti.* For instance, could these sequences represent the East African subspecies of *Ae. aegypti* that was suggested by Mattingly (1967) to conform to *Ae. queenslandensis* from the Mediterranean region? This question may get support from the proximity of the Suez Canal region connecting the Kenyan coast to the Mediterranean region *(*Gloria-Soria *et al.*, 2016). Another important question is whether the sequences in lineage 3 represent the reinvasion of Africa by global anthropophilic *Ae. aegypti* as described earlier (Brown *et al*., 2011, 2014; Powell, 2016).

5.3 East African strains compared to global anthropophilic strains

COI gene sequences reported in this study (Fig 4.5) show similar trend with regards to the reported dynamic pattern of *Ae. aegypt*i outside Africa (Kotsakiozi *et al*., 2018). The association of Kenyan *COI* gene sequences with those from South and North America suggests rapid expansion of this mosquito leading to adaptation in non-native areas. The close association of these sequences also supports previous conclusions that *Ae. aegypti* was introduced into the Americas from Africa (Powell *et al.,* 2018; Powell & Tabachnick, 2013). Further, clustering of sequence accessions from South America with the GenBank reference of Aaf sequence concurs with the observation that Aaf could have been introduced in South America recently (Powell, 2016).

Asian sequences show varied clusters (Fig 4.5). The sequences cluster with sequences from North and South America, Europe, and Australia. Previous research based on Single Nucleotide Polymorphism (SNP) genotyping also observed association of South American and Asian *Ae. aegypti* populations (Kotsakiozi *et al.*, 2018). These

observations support the explanation that *Ae. aegypti* was very likely introduced into Asia from the Americas (Bennett *et al*., 2016; Powell & Tabachnick, 2013). However, the introduction of *Ae. aegypti* into Asia from the Americas can be challenged by the discussion that *Ae. aegypti* became domesticated as a result of the expansion of the Sahara desert, and only human habitats would have provided reliable water sources and forage to support breeding of the vector (Powell & Tabachnick, 2013; Tabachnick, 1991). This implies that adaptation of Mediterranean *Ae. aegypti* populations would have led to Aaa survival aboard ships, and opening of the Suez Canal would have aided their invasion of European countries from Africa before invasion of the Americas (Tabachnick, 1991). The sequences from Asia also cluster with sequences from East Africa, and this might indicate exchange of vector haplotypes between the two regions especially due to the fact that East African vector populations were drawn from Rabai, near the port of Mombasa.

5.4 Correlation of *Ae. aegypti* mitochondrial variation with dengue outbreaks

In this study, the correlation of *Ae. aegypti* mitochondrial population genomics with dengue fever occurrence patterns stems from the fact that human populations at risk of infections with dengue virus haplotypes coincides with the distribution of *Ae. aegypti* (Gloria-Soria *et al*., 2016; Powell, 2016*)*. From the phylogenetic tree (Fig 4.2), the sequences collected from domestic settings in Bungoma, Busia, and Kakamega Counties seem to be more closely related to Aaf than Aaa. On the contrary, sequences from counties at Kenyan coast (Kilifi and Kwale counties) show close relationship with Aaa than Aaaf. In line with the evidence that Aaa is more susceptible to dengue virus than Aaf (Futami *et al*., 2019), this observation potentially explains why there is low dengue prevalence in Western Kenya, compared to the Kenyan Coast (Grossi-Soyster *et al*., 2017). The data reported in this study (Fig 4.2) shows sequences of samples derived from sylvatic settings clustering with the more dengue susceptible form Aaa. The observation trigers

speculation that the domestic form Aaa could likely be responsible for sylvatic dengue transmissions. This speculation is anchored on corroborated report of potential risk of sylvatic dengue transmissions by the Aaf form in Kenya (Musa *et al.*, 2020).

South America has witnessed repeated dengue outbreaks in recent times (Romero *et al.*, 2019), and the close association of sequences from South America with the GenBank Aaa reference sequences (Figs 4.5) suggests dengue virus transmissions are likely driven by the more dengue virus competent form, Aaa. Moreover, expansion of vector range, such as *Ae. aegypti* introduction into the Americas (Powell & Tabachnick, 2013), is highly correlated with the emergence of outbreaks of arbovirus diseases including dengue (Powell, 2016).

CHAPTER SIX

6.0 Conclusions and recommendations

Aedes aegypti continues to be a threat to global public health, and data analysed in this study suggests continued ecological expansion of this vector. *COI* gene data indicates clustering of *Ae. aegypti* sequences of samples collected from domestic settings with the referenced forest form, Aaf. This clustering may indicate that there are domestication events occurring in Kenya, particularly due to increased urbanization as reported by Baba *et al*. (2016). Dynamic *COI* patterns observed when comparing East African *Ae. aegypti* with global anthropophilic strains reveals how *Ae. aegypti* may be a very genetically diverse insect species. This conclusion stems from the observation of sequences from outside Africa forming clusters with the ancestral form, which is known to predominantly reside in Africa (Powell & Tabachnick, 2013).

A close association of Asian sequences was observed with those from Kilifi (Rabai) County at the coast of Kenya. There has been increased interconnectedness between Mombasa, East Africa's main shipping port and an international tourist destination, and Asian countries (Bita, 2015). Thus, tourism activities and international trade might have led to vector introduction/reintroductions between the two regions.

This study also highlights the usefulness of mitochondrial genes in understanding population genomics at the subspecies level. Evident little focus on these gene sequences is exhibited by the existence of only one defined *Ae. aegypti formosus COI* gene sequence in the GenBank database and few *Ae. aegypti cyt b* and *COII* sequences in GenBank. This sort of neglect may be attributed to reported evolutionary constraints of the *COII* gene (Beckenbach *et al*., 1993; Liu & Beckenbach, 1992).

6.1 Recommendations

From the sequence data examined in this study, it is evident that mitochondrial DNA barcode *COI* sequence has the inherent propensity to answer deep evolutionary questions. It is therefore recommended that future research using mitochondrial DNA for barcoding of a given set of species of interest should not just focus on the standard barcoding region but also try to cover much of the gene as possible. It would also be better to have subspecies level classification for sequences deposited in important public databases especially in the GenBank and the BOLD repositories.

Based on study findings, there is need for step up of pest control measures such as sustained surveillance systems targeting the vectors at ports of entry and frequented sites of tourism attraction because these may play a role in *Ae. aegypti* introductions/reintroductions and subsequent spread.

References

Agha, S. B., Tchouassi, D. P., Bastos, A. D. S., & Sang, R. (2017). Dengue and yellow fever virus vectors: Seasonal abundance, diversity and resting preferences in three Kenyan cities. *Parasites & Vectors*, *10*(1), 628. https://doi.org/10.1186/s13071-017-2598-2

Agha, S. B., Tchouassi, D. P., Turell, M. J., Bastos, A. D. S., & Sang, R. (2019). Entomological assessment of dengue virus transmission risk in three urban areas of Kenya. *PLOS Neglected Tropical Diseases*, *13*(8), e0007686. https://doi.org/10.1371/journal.pntd.0007686

Ajamma, Y. U., Villinger, J., Omondi, D., Salifu, D., Onchuru, T. O., Njoroge, L., Muigai, A. W. T., & Masiga, D. K. (2016). Composition and Genetic Diversity of Mosquitoes (Diptera: Culicidae) on Islands and Mainland Shores of Kenya's Lakes Victoria and Baringo. *Journal of Medical Entomology*, *53*(6), 1348–1363. https://doi.org/10.1093/jme/tjw102

Akiner, M. M., Demirci, B., Babuadze, G., Robert, V., & Schaffner, F. (2016). Spread of the Invasive Mosquitoes *Aedes aegypti* and *Aedes albopictus* in the Black Sea Region Increases Risk of Chikungunya, Dengue, and Zika Outbreaks in Europe. *PLOS Neglected Tropical Diseases*, *10*(4), e0004664. https://doi.org/10.1371/journal.pntd.0004664

Ali, K. H. A., El-Badry, A. A., Ali, M. A., El-Sayed, W. S. M., & El-Beshbishy, H. A. (2016). Phylogenetic Analysis of *Aedes aegypti* Based on Mitochondrial ND4 Gene Sequences in Almadinah, Saudi Arabia. *Iranian Journal of Biotechnology*, *14*(2), 58–62. https://doi.org/10.15171/ijb.1329

Alirol, E., Getaz, L., Stoll, B., Chappuis, F., & Loutan, L. (2011). Urbanisation and infectious diseases in a globalised world. *The Lancet Infectious Diseases*, *11*(2), 131–141. https://doi.org/10.1016/S1473-3099(10)70223-1

Al-Khazraji, S. H. A. A., Al-Sa'ati, M. A. Y., & Abdullah, N. M. (2014). Building High Performance Computing Using Beowulf Linux Cluster. *International Journal of Computer Science and Information Security*, *12*(4), 1.

Almeida, A. P., Gonçalves, Y. M., Novo, M. T., Sousa, C. A., Melim, M., & Gracio, A. J. (2007). Vector monitoring of *Aedes aegypti* in the Autonomous Region of Madeira, Portugal. *Weekly Releases (1997–2007)*, *12*(46). https://doi.org/10.2807/esw.12.46.03311-en

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

Baba, M., Villinger, J., & Masiga, D. K. (2016). Repetitive dengue outbreaks in East Africa: A proposed phased mitigation approach may reduce its impact: Repetitive dengue outbreaks in East Africa. *Reviews in Medical Virology*, *26*(3), 183–196. https://doi.org/10.1002/rmv.1877

Barrera, R., Amador, M., Diaz, A., Smith, J., Munoz-Jordan, J. L., & Rosario, Y. (2008). Unusual productivity of *Aedes aegypti* in septic tanks and its implications for dengue control. *Medical and Veterinary Entomology*, *22*(1), 62–69. https://doi.org/10.1111/j.1365-2915.2008.00720.x

Beckenbach, A. T., Wei, Y. W., & Liu, H. (1993). Relationships in the Drosophila obscura species group, inferred from mitochondrial cytochrome oxidase II sequences. *Molecular Biology and Evolution*, *10*(3), 619–634. https://doi.org/10.1093/oxfordjournals.molbev.a040034

Beebe, N. W. (2018). DNA barcoding mosquitoes: Advice for potential prospectors. *Parasitology*, *145*(5), 622–633.

Behura, S. K., Lobo, N. F., Haas, B., deBruyn, B., Lovin, D. D., Shumway, M. F., Puiu, D., Romero-Severson, J., Nene, V., & Severson, D. W. (2011). Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochemistry and Molecular Biology*, *41*(10), 770–777. https://doi.org/10.1016/j.ibmb.2011.05.006

Belshaw, R., & Quicke, D. L. J. (1997). A Molecular Phylogeny of the Aphidiinae (Hymenoptera: Braconidae). *Molecular Phylogenetics and Evolution*, *7*(3), 281–293. https://doi.org/10.1006/mpev.1996.0400

Bennett, K. L., Shija, F., Linton, Y.-M., Misinzo, G., Kaddumukasa, M., Djouaka, R., Anyaele, O., Harris, A., Irish, S., Hlaing, T., Prakash, A., Lutwama, J., & Walton, C. (2016). Historical environmental change in Africa drives divergence and admixture of *Aedes aegypti* mosquitoes: A precursor to successful worldwide colonization? *Molecular Ecology*, *25*(17), 4337–4354. https://doi.org/10.1111/mec.13762

Bennett, S. N. (2003). Selection-Driven Evolution of Emergent Dengue Virus. *Molecular Biology and Evolution*, *20*(10), 1650–1658. https://doi.org/10.1093/molbev/msg182

Berger, S. A., Krompass, D., & Stamatakis, A. (2011). Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood. *Systematic Biology*, *60*(3), 291–302. https://doi.org/10.1093/sysbio/syr010

Bevan, R. B., Lang, B. F., & Bryant, D. (2005). Calculating the Evolutionary Rates of Different Genes: A Fast, Accurate Estimator with Applications to Maximum Likelihood Phylogenetic Analysis. *Systematic Biology*, *54*(6), 900–915. https://doi.org/10.1080/10635150500354829

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., & Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, *496*(7446), 504–507. https://doi.org/10.1038/nature12060

Bita, C. (2015). Historical Period Stone Anchors from Mombasa, Kenya: Evidence of Overseas Maritime Trade Contacts with Asia and Middle East. *International Journal of Environment and Geoinformatics*, *2*(3), 15–26. https://doi.org/10.30897/ijegeo.303557

Black, W. C., Bennett, K. E., Gorrochótegui-Escalante, N., Barillas-Mury, C. V., Fernández-Salas, I., de Lourdes Muñoz, M., Farfán-Alé, J. A., Olson, K. E., & Beaty, B. J. (2002). Flavivirus Susceptibility in *Aedes aegypti*. *Archives of Medical Research*, *33*(4), 379–388. https://doi.org/10.1016/S0188-4409(02)00373-9

Blair, C. D., Olson, K. E., & Bonizzoni, M. (2019). The Widespread Occurrence and Potential Biological Roles of Endogenous Viral Elements in Insect Genomes. *Current Issues in Molecular Biology*, *34*, 13–30. https://doi.org/10.21775/cimb.034.013

Bosio, C F, Beaty, B. J., & Black, W. C. (1998). Quantitative genetics of vector competence for dengue-2 virus in *Aedes aegypti*. *The American Journal of*

Tropical Medicine and Hygiene, *59*(6), 965–970. https://doi.org/10.4269/ajtmh.1998.59.965

Bosio, Christopher F., Harrington, L. C., Jones, J. W., Sithiprasasna, R., Norris, D. E., & Scott, T. W. (2005). Genetic structure of *Aedes aegypti* populations in Thailand using mitochondrial DNA. *The American Journal of Tropical Medicine and Hygiene*, *72*(4), 434–442.

Bracco, J. E., Capurro, M. L., Lourenço-de-Oliveira, R., & Sallum, M. A. M. (2007). Genetic variability of *Aedes aegypti* in the Americas using a mitochondrial gene: Evidence of multiple introductions. *Memórias Do Instituto Oswaldo Cruz*, *102*(5), 573–580. https://doi.org/10.1590/S0074-02762007005000062

Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., Moyes, C. L., Farlow, A. W., Scott, T. W., & Hay, S. I. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Neglected Tropical Diseases*, *6*(8), e1760. https://doi.org/10.1371/journal.pntd.0001760

Brown, J. E., Evans, B. R., Zheng, W., Obas, V., Barrera-Martinez, L., Egizi, A., Zhao, H., Caccone, A., & Powell, J. R. (2014). Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution; International Journal of Organic Evolution*, *68*(2), 514–525. https://doi.org/10.1111/evo.12281

Brown, J. E., McBride, C. S., Johnson, P., Ritchie, S., Paupy, C., Bossin, H., Lutomiah, J., Fernandez-Salas, I., Ponlawat, A., Cornel, A. J., Black, W. C., Gorrochotegui-Escalante, N., Urdaneta-Marquez, L., Sylla, M., Slotman, M., Murray, K. O., Walker, C., & Powell, J. R. (2011). Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti* , a

major vector of human diseases. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1717), 2446–2454. https://doi.org/10.1098/rspb.2010.2469

Buhay, J. E. (2009). "*COI*-like" Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. *Journal of Crustacean Biology*, *29*(1), 96–110. https://doi.org/10.1651/08-3020.1

Cameron, S. L. (2014). Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. *Annual Review of Entomology*, *59*(1), 95–117. https://doi.org/10.1146/annurev-ento-011613-162007

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, *33*(2), 201–208. https://doi.org/10.1016/S0167-7152(96)00128-9

Chan, A., Chiang, L.-P., Hapuarachchi, H. C., Tan, C.-H., Pang, S.-C., Lee, R., Lee, K.-S., Ng, L.-C., & Lam-Phua, S.-G. (2014). DNA barcoding: Complementing morphological identification of mosquito species in Singapore. *Parasites & Vectors*, *7*(1), 569. https://doi.org/10.1186/s13071-014-0569-4

Conway, M. J., Colpitts, T. M., & Fikrig, E. (2014). Role of the Vector in Arbovirus Transmission. *Annual Review of Virology*, *1*(1), 71–88. https://doi.org/10.1146/annurev-virology-031413-085513

Cook, S., Diallo, M., Sall, A. A., Cooper, A., & Holmes, E. C. (2005). Mitochondrial markers for molecular identification of Aedes mosquitoes (Diptera: Culicidae) involved in transmission of arboviral disease in West Africa. *Journal of Medical Entomology*, *42*(1), 19–28. https://doi.org/10.1093/jmedent/42.1.19

Cook, S., Diallo, M., Sall, A. A., Cooper, A., & Holmes, E. C. (2006). Mitochondrial Markers for Molecular Identification of <I>Aedes</I> Mosquitoes (Diptera: Culicidae) Involved in Transmission of Arboviral Disease in West Africa.

Journal of Medical Entomology, *42*(1), 19–28. https://doi.org/10.1603/0022-2585(2005)042[0019:mmfmio]2.0.co;2

Crawford, J. E., Alves, J. M., Palmer, W. J., Day, J. P., Sylla, M., Ramasamy, R., Surendran, S. N., Black, W. C., Pain, A., & Jiggins, F. M. (2017). Population genomics reveals that an anthropophilic population of *Aedes aegypti* mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biology*, *15*(1), 16. https://doi.org/10.1186/s12915-017-0351-0

Diallo, M., Ba, Y., Faye, O., Soumare, M. L., Dia, I., & Sall, A. A. (2008). Vector competence of *Aedes aegypti* populations from Senegal for sylvatic and epidemic dengue 2 virus isolated in West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(5), 493–498. https://doi.org/10.1016/j.trstmh.2008.02.010

Diallo, M., Sall, A. A., Moncayo, A. C., Ba, Y., Fernandez, Z., Ortiz, D., Coffey, L. L., Mathiot, C., Tesh, R. B., & Weaver, S. C. (2005). Potential role of sylvatic and domestic African mosquito species in dengue emergence. *The American Journal of Tropical Medicine and Hygiene*, *73*(2), 445–449.

Dickson, L. B., Jiolle, D., Minard, G., Moltini-Conclois, I., Volant, S., Ghozlane, A., Bouchier, C., Ayala, D., Paupy, C., Moro, C. V., & Lambrechts, L. (2017). Carryover effects of larval exposure to different environmental bacteria drive adult trait variation in a mosquito vector. *Science Advances*, *3*(8), e1700585. https://doi.org/10.1126/sciadv.1700585

Dickson, L. B., Sanchez-Vargas, I., Sylla, M., Fleming, K., & Black, W. C. (2014). Vector Competence in West African *Aedes aegypti* Is Flavivirus Species and

Genotype Dependent. *PLoS Neglected Tropical Diseases*, *8*(10), e3153. https://doi.org/10.1371/journal.pntd.0003153

Esposti, M. D., De Vries, S., Crimi, M., Ghelli, A., Patarnello, T., & Meyer, A. (1993). Mitochondrial cytochrome b: Evolution and structure of the protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1143*(3), 243–271. https://doi.org/10.1016/0005-2728(93)90197-N

Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, *25*(5), 471–492.

Felsenstein, Joseph. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376. https://doi.org/10.1007/BF01734359

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Futami, K., Iwashita, H., Higa, Y., Lutiali, P. A., Sonye, G. O., Mwatele, C., Njenga, S. M., & Minakawa, N. (2019). Geographical Distribution of *Aedes aegypti* aegypti and *Aedes aegypti* formosus (Diptera: Culicidae) in Kenya and Environmental Factors Related to Their Relative Abundance. *Journal of Medical Entomology*, *57*(3), 772–779. https://doi.org/10.1093/jme/tjz233

Gimeno, C., Belshaw, R., & Quicke, D. L. J. (1997). Phylogenetic relationships of the Alysiinae/Opiinae (Hymenoptera: Braconidae) and the utility of cytochrome b, 16S and 28S D2 rRNA. *Insect Molecular Biology*, *6*(3), 273–284. https://doi.org/10.1046/j.1365-2583.1997.00181.x

Giovanetti, M., de Mendonça, M. C. L., Fonseca, V., Mares-Guia, M. A., Fabri, A., Xavier, J., de Jesus, J. G., Gräf, T., Dos Santos Rodrigues, C. D., Dos Santos, C. C., Sampaio, S. A., Chalhoub, F. L. L., de Bruycker Nogueira, F., Theze, J., Romano, A. P. M., Ramos, D. G., de Abreu, A. L., Oliveira, W. K., do Carmo Said, R. F., … de Filippis, A. M. B. (2019). Yellow Fever Virus Reemergence and Spread in Southeast Brazil, 2016-2019. *Journal of Virology*, *94*(1). https://doi.org/10.1128/JVI.01623-19

Gloria-Soria, A., Ayala, D., Bheecarry, A., Calderon-Arguedas, O., Chadee, D. D., Chiappero, M., Coetzee, M., Elahee, K. B., Fernandez-Salas, I., Kamal, H. A., Kamgang, B., Khater, E. I. M., Kramer, L. D., Kramer, V., Lopez-Solis, A., Lutomiah, J., Martins, A., Micieli, M. V., Paupy, C., … Powell, J. R. (2016). Global genetic diversity of *Aedes aegypti*. *Molecular Ecology*, *25*(21), 5377–5395. https://doi.org/10.1111/mec.13866

Gonçalves da Silva, A., Cunha, I. C. L., Santos, W. S., Luz, S. L. B., Ribolla, P. E. M., & Abad-Franch, F. (2012). Gene flow networks among American *Aedes aegypti* populations: Gene flow networks among American *Aedes aegypti* populations. *Evolutionary Applications*, *5*(7), 664–676. https://doi.org/10.1111/j.1752-4571.2012.00244.x

Grossi-Soyster, E. N., Cook, E. A. J., de Glanville, W. A., Thomas, L. F., Krystosik, A. R., Lee, J., Wamae, C. N., Kariuki, S., Fèvre, E. M., & LaBeaud, A. D. (2017). Serological and spatial analysis of alphavirus and flavivirus prevalence and risk factors in a rural community in western Kenya. *PLOS Neglected Tropical Diseases*, *11*(10), e0005998. https://doi.org/10.1371/journal.pntd.0005998

Gubler, D. J. (1997). Dengue and dengue hemorrhagic fever. *Seminars in Pediatric Infectious Diseases*, *8*(1), 3–9. https://doi.org/10.1016/S1045-1870(97)80003-9

Gubler, D. J. (2002). The Global Emergence/Resurgence of Arboviral Diseases As Public Health Problems. *Archives of Medical Research*, *33*(4), 330–342. https://doi.org/10.1016/S0188-4409(02)00378-8

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biological Sciences*, *270 Suppl 1*, S96-99. https://doi.org/10.1098/rsbl.2003.0025

Huang, Y.-J. S., Higgs, S., & Vanlandingham, D. L. (2019). Arbovirus-Mosquito Vector-Host Interactions and the Impact on Transmission and Disease Pathogenesis of Arboviruses. *Frontiers in Microbiology*, *10*, 22. https://doi.org/10.3389/fmicb.2019.00022

Jones, C. M., Lee, Y., Kitchen, A., Collier, T., Pringle, J. C., Muleba, M., Irish, S., Stevenson, J. C., Coetzee, M., Cornel, A. J., Norris, D. E., & Carpi, G. (2018). Complete Anopheles funestus mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. *Scientific Reports*, *8*(1), 9054. https://doi.org/10.1038/s41598-018-27092-y

Kahanwal, D., & Singh, D. T. (2013). The distributed computing paradigms: P2P, grid, cluster, cloud, and jungle. *ArXiv Preprint ArXiv:1311.3070*.

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), 587–589. https://doi.org/10.1038/nmeth.4285

Kambhampati, S., & Rai, K. S. (1991). Mitochondrial DNA variation within and among populations of the mosquito *Aedes albopictus*. *Genome*, *34*(2), 288–292.

Karungu, S., Atoni, E., Ogalo, J., Mwaliko, C., Agwanda, B., Yuan, Z., & Hu, X. (2019). Mosquitoes of Etiological Concern in Kenya and Possible Control Strategies. *Insects*, *10*(6). https://doi.org/10.3390/insects10060173

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kearse, Matthew, Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kotsakiozi, P., Evans, B. R., Gloria-Soria, A., Kamgang, B., Mayanja, M., Lutwama, J., Goff, G. L., Ayala, D., Paupy, C., Badolo, A., Pinto, J., Sousa, C. A., Troco, A.

D., & Powell, J. R. (2018). Population structure of a vector of human diseases: *Aedes aegypti* in its ancestral range, Africa. *Ecology and Evolution*, *8*(16), 7835–7848. https://doi.org/10.1002/ece3.4278

Kotsakiozi, P., Gloria-Soria, A., Schaffner, F., Robert, V., & Powell, J. R. (2018). *Aedes aegypti* in the Black Sea: Recent introduction or ancient remnant? *Parasites & Vectors*, *11*(1), 396. https://doi.org/10.1186/s13071-018-2933-2

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kropelin, S., Verschuren, D., Lezine, A.-M., Eggermont, H., Cocquyt, C., Francus, P., Cazet, J.-P., Fagot, M., Rumes, B., Russell, J. M., Darius, F., Conley, D. J., Schuster, M., von Suchodoletz, H., & Engstrom, D. R. (2008). Climate-Driven Ecosystem Succession in the Sahara: The Past 6000 Years. *Science*, *320*(5877), 765–768. https://doi.org/10.1126/science.1154913

Langat, S. K., Eyase, F. L., Berry, I. M., Nyunja, A., Bulimo, W., Owaka, S., Ofula, V., Limbaso, S., Lutomiah, J., Jarman, R., Distelhorst, J., & Sang, R. C. (2020). Origin and evolution of dengue virus type 2 causing outbreaks in Kenya: Evidence of circulation of two cosmopolitan genotype lineages. *Virus Evolution*, *6*(1), veaa026. https://doi.org/10.1093/ve/veaa026

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Liu, H., & Beckenbach, A. T. (1992). Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Molecular Phylogenetics and Evolution*, *1*(1), 41–52. https://doi.org/10.1016/1055-7903(92)90034-E

Lounibos, L. P. (1981). Habitat segregation among African treehole mosquitoes. *Ecological Entomology*, *6*(2), 129–154. https://doi.org/10.1111/j.1365-2311.1981.tb00601.x

Mattingly, P. F. (1957). Genetical aspects of the *Aedes aegypti* problem. I. Taxonom: And bionomics. *Annals of Tropical Medicine and Parasitology*, *51*(4), 392–408.

Mattingly, P. F. (1967). Taxonomy of *Aedes aegypti* and related species. *Bulletin of the World Health Organization*, *36*(4), 552–554.

McBride, C. S., Baier, F., Omondi, A. B., Spitzer, S. A., Lutomiah, J., Sang, R., Ignell, R., & Vosshall, L. B. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*, *515*(7526), 222–227. https://doi.org/10.1038/nature13964

Meyer, A. (1994). Shortcomings of the cytochrome b gene as a molecular marker. *Trends in Ecology & Evolution*, *9*(8), 278–280. https://doi.org/10.1016/0169-5347(94)90028-0

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mitchell, C. J., & Miller, B. R. (1991). Genetic Selection of a Flavivirus-Refractory Strain of the Yellow Fever Mosquito *Aedes aegypti*. *The American Journal of Tropical Medicine and Hygiene*, *45*(4), 399–407. https://doi.org/10.4269/ajtmh.1991.45.399

Moore, M., Sylla, M., Goss, L., Burugu, M. W., Sang, R., Kamau, L. W., Kenya, E. U., Bosio, C., Munoz, M. de L., Sharakova, M., & Black, W. C. (2013). Dual African Origins of Global *Aedes aegypti* s.l. Populations Revealed by Mitochondrial DNA. *PLoS Neglected Tropical Diseases*, *7*(4), e2175. https://doi.org/10.1371/journal.pntd.0002175

Moritz, C., Dowling, T. E., & Brown, W. M. (1987). EVOLUTION OF ANIMAL MITOCHONDRIAL DNA: RELEVANCE FOR POPULATION BIOLOGY AND SYSTEMATICS. *Annual Review of Ecology and Systematics*, *18*(1), 269–292. https://doi.org/10.1146/annurev.es.18.110187.001413

Moriyama, E. N., & Powell, J. R. (1997). Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. *Journal of Molecular Evolution*, *45*(4), 378–391. https://doi.org/10.1007/PL00006243

Morlais, I., & Severson, D. W. (2002). Complete Mitochondrial DNA Sequence and Amino Acid Analysis of the Cytochrome C Oxidase Subunit I (*COI*) from *Aedes aegypti*. *DNA Sequence*, *13*(2), 123–127. https://doi.org/10.1080/10425170290030051

Mousson, L., Dauga, C., Garrigues, T., Schaffner, F., Vazeille, M., & Failloux, A.-B. (2005). Phylogeography of Aedes (Stegomyia) aegypti (L.) and Aedes (Stegomyia) albopictus (Skuse) (Diptera: Culicidae) based on mitochondrial DNA variations. *Genetical Research*, *86*(1), 1–11. https://doi.org/10.1017/S0016672305007627

Moyes, C. L., Vontas, J., Martins, A. J., Ng, L. C., Koou, S. Y., Dusfour, I., Raghavendra, K., Pinto, J., Corbel, V., David, J.-P., & Weetman, D. (2017). Contemporary status of insecticide resistance in the major Aedes vectors of

arboviruses infecting humans. *PLoS Neglected Tropical Diseases*, *11*(7), e0005625. https://doi.org/10.1371/journal.pntd.0005625

Musa, A. A., Muturi, M. W., Musyoki, A. M., Ouso, D. O., Oundo, J. W., Makhulu, E. E., Wambua, L., Villinger, J., & Jeneby, M. M. (2020). Arboviruses and Blood Meal Sources in Zoophilic Mosquitoes at Human-Wildlife Interfaces in Kenya. *Vector-Borne and Zoonotic Diseases*, *20*(6), 444–453. https://doi.org/10.1089/vbz.2019.2563

Musso, D., Cao-Lormeau, V. M., & Gubler, D. J. (2015). Zika virus: Following the path of dengue and chikungunya? *The Lancet*, *386*(9990), 243–244. https://doi.org/10.1016/S0140-6736(15)61273-9

Nag, D. K., & Kramer, L. D. (2017). Patchy DNA forms of the Zika virus RNA genome are generated following infection in mosquito cell cultures and in mosquitoes. *Journal of General Virology*, *98*(11), 2731–2737. https://doi.org/10.1099/jgv.0.000945

Ndenga, B. A., Mutuku, F. M., Ngugi, H. N., Mbakaya, J. O., Aswani, P., Musunzaji, P. S., Vulule, J., Mukoko, D., Kitron, U., & LaBeaud, A. D. (2017). Characteristics of *Aedes aegypti* adult mosquitoes in rural and urban areas of western and coastal Kenya. *PloS One*, *12*(12), e0189971. https://doi.org/10.1371/journal.pone.0189971

Olliaro, P., Fouque, F., Kroeger, A., Bowman, L., Velayudhan, R., Santelli, A. C., Garcia, D., Skewes Ramm, R., Sulaiman, L. H., Tejeda, G. S., Morales, F. C., Gozzer, E., Garrido, C. B., Quang, L. C., Gutierrez, G., Yadon, Z. E., & Runge-Ranzinger, S. (2018). Improved tools and strategies for the prevention and control of arboviral diseases: A research-to-policy forum. *PLOS Neglected*

Tropical Diseases, *12*(2), e0005967. https://doi.org/10.1371/journal.pntd.0005967

Parimittr, P., Chareonviriyaphap, T., Bangs, M. J., & Arunyawat, U. (2018). Genetic variation of *Aedes aegypti* mosquitoes across Thailand based on nuclear DNA sequences. *Agriculture and Natural Resources*, *52*(6), 596–602. https://doi.org/10.1016/j.anres.2018.11.021

Paupy, C., Brengues, C., Kamgang, B., Hervé, J.-P., Fontenille, D., & Simard, F. (2008). Gene Flow Between Domestic and Sylvan Populations of <I>*Aedes aegypti*</I> (Diptera: Culicidae) in North Cameroon. *Journal of Medical Entomology*, *45*(3), 391–400. https://doi.org/10.1603/0022-2585(2008)45[391:GFBDAS]2.0.CO;2

Powell, J. R. (2016). Mosquitoes on the move. *Science*, *354*(6315), 971–972. https://doi.org/10.1126/science.aal1717

Powell, J. R., Gloria-Soria, A., & Kotsakiozi, P. (2018). Recent History of *Aedes aegypti*: Vector Genomics and Epidemiology Records. *BioScience*, *68*(11), 854–860. https://doi.org/10.1093/biosci/biy119

Powell, J. R., & Tabachnick, W. J. (2013). History of domestication and spread of *Aedes aegypti*—A Review. *Memórias Do Instituto Oswaldo Cruz*, *108*(suppl 1), 11–17. https://doi.org/10.1590/0074-0276130395

Puntasecca, C. J., King, C. H., & LaBeaud, A. D. (2021). Measuring the global burden of chikungunya and Zika viruses: A systematic review. *PLOS Neglected Tropical Diseases*, *15*(3), e0009055. https://doi.org/10.1371/journal.pntd.0009055

Rašić, G., Schama, R., Powell, R., Maciel-de Freitas, R., Endersby-Harshman, N. M., Filipović, I., Sylvestre, G., Máspero, R. C., & Hoffmann, A. A. (2015).

Contrasting genetic structure between mitochondrial and nuclear markers in the dengue fever mosquito from Rio de Janeiro: Implications for vector control. *Evolutionary Applications*, *8*(9), 901–915. https://doi.org/10.1111/eva.12301

Ratnasingham, S., & Hebert, P. D. N. (2007). BARCODING: Bold: The Barcode of Life Data System (http://www.barcodinglife.org): BARCODING. *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Redmond, S. N., Sharma, A., Sharakhov, I., Tu, Z., Sharakhova, M., & Neafsey, D. E. (2020). Linked-read sequencing identifies abundant microinversions and introgression in the arboviral vector *Aedes aegypti*. *BMC Biology*, *18*(1), 26. https://doi.org/10.1186/s12915-020-0757-y

Romero, D., Olivero, J., Real, R., & Guerrero, J. C. (2019). Applying fuzzy logic to assess the biogeographical risk of dengue in South America. *Parasites & Vectors*, *12*(1), 428. https://doi.org/10.1186/s13071-019-3691-5

Rose, N. H., Sylla, M., Badolo, A., Lutomiah, J., Ayala, D., Aribodor, O. B., Ibe, N., Akorli, J., Otoo, S., Mutebi, J.-P., Kriete, A. L., Ewing, E. G., Sang, R., Gloria-Soria, A., Powell, J. R., Baker, R. E., White, B. J., Crawford, J. E., & McBride, C. S. (2020). Climate and Urbanization Drive Mosquito Preference for Humans. *Current Biology*, *30*(18), 3570-3579.e6. https://doi.org/10.1016/j.cub.2020.06.092

Salgueiro, P., Serrano, C., Gomes, B., Alves, J., Sousa, C. A., Abecasis, A., & Pinto, J. (2019). Phylogeography and invasion history of *Aedes aegypti* , the Dengue and Zika mosquito vector in Cape Verde islands (West Africa). *Evolutionary Applications*, *12*(9), 1797–1811. https://doi.org/10.1111/eva.12834

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sanger, Frederick. (1988). Sequences, Sequences, and Sequences. *Annual Review of Biochemistry*, *57*(1), 1–29. https://doi.org/10.1146/annurev.bi.57.070188.000245

Schaffner, F., Fontenille, D., & Mathis, A. (2014). Autochthonous dengue emphasises the threat of arbovirosis in Europe. *The Lancet Infectious Diseases*, *14*(11), 1044. https://doi.org/10.1016/S1473-3099(14)70925-9

Seixas, G., Salgueiro, P., Silva, A. C., Campos, M., Spenassatto, C., Reyes-Lugo, M., Novo, M. T., Ribolla, P. E. M., Pinto, J. P. S. da S., & Sousa, C. A. (2013). *Aedes aegypti* on Madeira Island (Portugal): Genetic variation of a recently introduced dengue vector. *Memórias Do Instituto Oswaldo Cruz*, *108*(suppl 1), 3–10. https://doi.org/10.1590/0074-0276130386

Shendure, J. A., Porreca, G. J., Church, G. M., Gardner, A. F., Hendrickson, C. L., Kieleczawa, J., & Slatko, B. E. (2011). Overview of DNA Sequencing Strategies. *Current Protocols in Molecular Biology*, *96*(1). https://doi.org/10.1002/0471142727.mb0701s96

Smith, C. E. G. (1956). The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. *The Journal of Tropical Medicine and Hygiene*, *59*(10), 243–251.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, *147*(1), 195–197. https://doi.org/10.1016/0022-2836(81)90087-5

Solórzano, J. O., Guzmán, M. G., Brathwaite, O., Bouckenooghe, A., Zambrano, B., Dayan, G. H., & San Martín, J. L. (2010). The Epidemiology of Dengue in the Americas Over the Last Three Decades: A Worrisome Reality. *The American Journal of Tropical Medicine and Hygiene*, *82*(1), 128–135. https://doi.org/10.4269/ajtmh.2010.09-0346

Soper, D. F. (1970). Building the Health Bridge, Selections from the Works of Fred L. Soper, M.D.: Edited by J. Austin Kerr. xxxvii + 567 pages. Indiana University Press, Bloomington, Indiana. 1970. $17.50. *The American Journal of Tropical Medicine and Hygiene*, *19*(6), 1056–1057. https://doi.org/10.4269/ajtmh.1970.19.1056

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sylla, M., Bosio, C., Urdaneta-Marquez, L., Ndiaye, M., & Black, W. C. (2009). Gene Flow, Subspecies Composition, and Dengue Virus-2 Susceptibility among *Aedes aegypti* Collections in Senegal. *PLoS Neglected Tropical Diseases*, *3*(4), e408. https://doi.org/10.1371/journal.pntd.0000408

Tabachnick, W. J., Wallis, G. P., Miller, B. R., Powell, J. R., Lorenz, L., Aitken, T. H. G., Beaty, B. J., & Amato, G. D. (1985). Oral Infection of *Aedes aegypti* with Yellow Fever Virus: Geographic Variation and Genetic Considerations. *The American Journal of Tropical Medicine and Hygiene*, *34*(6), 1219–1224. https://doi.org/10.4269/ajtmh.1985.34.1219

Tabachnick, Walter J. (1991). Evolutionary Genetics and Arthropod-borne Disease: The Yellow Fever Mosquito. *American Entomologist*, *37*(1), 14–26. https://doi.org/10.1093/ae/37.1.14

Tabachnick, Walter J., Munstermann, L. E., & Powell, J. R. (1979). GENETIC DISTINCTNESS OF SYMPATRIC FORMS OF *AEDES AEGYPTI* IN EAST AFRICA. *Evolution*, *33*(1Part2), 287–295. https://doi.org/10.1111/j.1558-5646.1979.tb04682.x

Urdaneta-Marquez, L., & Failloux, A.-B. (2011). Population genetic structure of *Aedes aegypti*, the principal vector of dengue viruses. *Infection, Genetics and Evolution*, *11*(2), 253–261. https://doi.org/10.1016/j.meegid.2010.11.020

Vazeille, M., Yébakima, A., Lourenço-de-Oliveira, R., Andriamahefazafy, B., Correira, A., Rodrigues, J. M., Veiga, A., Moreira, A., Leparc-Goffart, I., Grandadam, M., & Failloux, A.-B. (2013). Oral Receptivity of *Aedes aegypti* from Cape Verde for Yellow Fever, Dengue, and Chikungunya Viruses. *Vector-Borne and Zoonotic Diseases*, *13*(1), 37–40. https://doi.org/10.1089/vbz.2012.0982

Virgilio, M., Backeljau, T., Nevado, B., & De Meyer, M. (2010). Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics*, *11*(1), 206. https://doi.org/10.1186/1471-2105-11-206

Wang, G., Li, C., Zheng, W., Song, F., Guo, X., Wu, Z., Luo, P., Yang, Y., He, L., & Zhao, T. (2016). An evaluation of the suitability of *COI* and *COII* gene variation for reconstructing the phylogeny of, and identifying cryptic species in, anopheline mosquitoes (Diptera Culicidae). *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*, *28*(5), 769–777. https://doi.org/10.1080/24701394.2016.1186665

Weaver, S. C., & Lecuit, M. (2015). Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. *New England Journal of Medicine*, *372*(13), 1231–1239. https://doi.org/10.1056/NEJMra1406035

Weetman, D., Kamgang, B., Badolo, A., Moyes, C., Shearer, F., Coulibaly, M., Pinto, J., Lambrechts, L., & McCall, P. (2018). Aedes Mosquitoes and Aedes-Borne Arboviruses in Africa: Current and Future Threats. *International Journal of Environmental Research and Public Health*, *15*(2), 220. https://doi.org/10.3390/ijerph15020220

Wilson, M., & Chen, L. (2002). Dengue in the Americas. *Dengue Bulletin*, *26*, 44–61.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution*, *11*(9), 367–372. https://doi.org/10.1016/0169-5347(96)10041-0

Appendix

Table 3.3. Details of sequences retrieved from BOLD and GenBank

BOLD ID and GenBank Accession Number	Database	Country of Origin	Gene	Sequence Length (Bp)
GQ165783	GenBank	Uganda	*COI*	1152
KU186990	GenBank	Kenya	*COI*	1151
AY432106	GenBank	USA	*COI*	1656
AF380835	GenBank	USA	*COI*	1537
AF390098	GenBank	USA	*COI*	1537
AY056597	GenBank	USA	*COI*	1537
MK533624 to MK533630	GenBank	Kenya	*COI*	658
GBDCU1594-15 to GBDCU1596-15	BOLD	Singapore	*COI*	679
GBDCU337-12	BOLD	India	*COI*	665
GBDCU345-12	BOLD	India	*COI*	668
GBDP50512-19 to GBDP50538-19	BOLD	India	*COI*	710
MADIP260-10	BOLD	Sri Lanka	*COI*	683
MADIP261-10	BOLD	Pakistan	*COI*	658
MADIP378-10	BOLD	Pakistan	*COI*	658
MAMOS017-12	BOLD	Pakistan	*COI*	658

MAMOS019-12	BOLD	Pakistan	*COI*	658
MAMOS113-12	BOLD	Pakistan	*COI*	658
MAMOS115-12	BOLD	Pakistan	*COI*	658
MAMOS1858-13	BOLD	Pakistan	*COI*	658
MAMOS729-12	BOLD	Pakistan	*COI*	658
MAMOS797-12	BOLD	Pakistan	*COI*	650
GBMIN56161-17 to	BOLD	Colombia	*CO*	695
GBMIN56196-17	BOLD	Vietnam	*COI*	764
GBDP12689-12	BOLD	Thailand	*COI*	764
GBDP12695-12	BOLD	Thailand	*COI*	658
AGIRI262-17	BOLD	India	*COI*	660
GBDCU341-12	BOLD	Cambodia	*COI*	764
GBDP12694-12	BOLD	India	*COI*	708
SPLID001-13	BOLD	Pakistan	*COI*	658
MAMOS1860-13	BOLD	India	*COI*	654
GBDP4072-07	BOLD	Cambodia	*COI*	764
GBDP12692-12	BOLD	Tanzania	*COI*	765
GBDP12708-1	BOLD	Kenya	*COI*	~784
GBMNA23247-19 - GBMNA23277-19	BOLD	Kenya	*COI*	~710

GBMNA44480-19 - GBMNA44488-19	BOLD	Kenya	*COI*	~659
MBIEA159-10 - MBIEA197-10	BOLD	Mexico	*COI*	765
GBDP12702-12	BOLD	Canada	*COI*	680
GBDP50287-19 - GBDP50289-19	BOLD	Mexico	*COI*	659
MOSEM010-18 - MOSEM014-18	BOLD	Mexico	*COI*	658
MOSMX044-18	BOLD	Mexico	*COI*	765
GBDP12703-12	BOLD	Canada	*COI*	681
GBDP50288-19	BOLD	Mexico	*COI*	659
MOSEM011-18 - MOSEM013-18	BOLD	Mexico	*COI*	659
MOSMX043-18 - MOSMX047-18	BOLD	Brazil	*COI*	1501
GBDCU118-12 - GBDCU119-12	BOLD	Brazil	*COI*	1501
GBDP0800-06- GBDP0814-06	BOLD	Bolivia	*COI*	765
GBDP12680-12 - GBDP12687-12	BOLD	Brazil	*COI*	765
GBDP12707-12	BOLD	Colombia	*COI*	697
GBDP48742-19	BOLD	Brazil	*COI*	1501
GBDP9667-12 - GBDP9669-12	BOLD	Colombia	*COI*	696
GBMIN56161-17GBMIN56186-17	BOLD	Colombia	*COI*	697
GBMIN56195-17 - GBMIN56200-17	BOLD	Brazil	*COI*	1537
	BOLD	Ecuador	*COI*	659

GBMNA46232-19	BOLD	Chile	*COI*	659
MBIB833-10 - MBIB835-10	BOLD	Bolivia	*COI*	765
MBICL034-10 - MBICL037-10	BOLD	Chile	*COI*	659
GBDP12682-12 - GBDP12684-12	BOLD	Portugal	*COI*	1096
MBICL033-10 - MBICL035-10	BOLD	France	*COI*	1501
GBDCU1217-14	BOLD	Germany	*COI*	710
GBDP0805-06	BOLD	England	*COI*	696
GBDP48812-19	BOLD	Germany	*COI*	710
GBMIN56192-17	BOLD	England	*COI*	1538
GBMIN56225-17	BOLD	Australia	*COI*	780
GTENK039-11	BOLD	Australia	*COI*	658
GBDP10706-12	BOLD	Australia	*COI*	658
DIQTB385-11	BOLD	Australia	*COI*	658
DIQT510-09	BOLD	Australia	*COI*	658
DIQT572-09	BOLD	Australia	*COI*	687

AJ970947	GenBank	Europa Island	*cyt b*	360
AJ970952	GenBank	Madagascar	*cyt b*	360
AJ970959	GenBank	Thailand	*cyt b*	360

AJ970951	GenBank	Vietnam	*cyt b*	360
AJ970945	Genbank	Ivory Coast	*cyt b*	360
AJ970954	GenBank	Vietnam	*cyt b*	360
AJ970953	GenBank	Martinique	*cyt b*	360
AY645312	GenBank	Senegal	*COII*	603
AY645325	GenBank	Senegal	*COII*	603
DQ181465	GenBank	Puerto Rico	*COII*	603
DQ181504	GenBank	Puerto Rico	*COII*	603
AY645311	GenBank	Senegal	*COII*	603
DQ181467-DQ181468	GenBank	Puerto Rico	*COII*	603
DQ181471	GenBank	Puerto Rico	*COII*	603
KC913582	GenBank	China	*COII*	681
JQ301826	GenBank	India	*COII*	564
KX495183	GenBank	India	*COII*	406
AY645323	GenBank	Senegal	*COII*	603
AY645315	GenBank	Senegal	*COII*	603
AY645332	GenBank	Senegal	*COII*	603
AY645320	GenBank	Senegal	*COII*	603
AY645326	GenBank	Senegal	*COII*	603

AY645313	GenBank	Senegal	*COII*	603
DQ181505	GenBank	Puerto Rico	*COII*	603
DQ181508	GenBank	Puerto Rico	*COII*	583
AY645310	GenBank	Senegal	*COII*	603
AY645314	GenBank	Senegal	*COII*	603

Protein Precipitation (PPS) protocol for DNA extraction

Required solutions

Cell Lysis buffer (CLB)

Protein Precipitation Solution (PPS)

70% ethanol

Isopropanol

Water

Protocol

1. Squash one whole fly in liquid nitrogen before adding 300 µl of Cell Lysis Buffer (CLB) on ice.

   Alternatively, the fly can be crushed in 1.5 ml screw-cap Eppendorf tubes (Sarstedt, Newton, NC) filled with 750 mg of 2.0 mm, 150 mg of 0.1 mm Yttria-stabilized zirconium (YSZ) oxide beads (Glen Mills, Clifton, New Jersey, USA) and 50 µl of phosphate-buffered saline using a Mini-BeadBeater 16® (BioSpec, Bartlesville, UK) instead of liquid nitrogen.
2. Add 300ul of Cell Lysis Buffer on ice

3. Incubate the homogenate at 65°C for 30-50 mins
4. At room temperature, add 100 µl of PPS
5. Vortex for 3 x 10 seconds
6. Place on ice for 5 – 10 mins
7. Centrifuge at a maximum speed for 5 mins at 4 °C.
8. Transfer the supernatant to a fresh tube with 300ul isopropanol (discard pellets)
9. Mix by inverting 100 times
10. Centrifuge at maximum speed for 30 mins - 1 hour.
11. Pipette off supernatant (pour off)
12. Add 300ul of ice cold 70% ethanol, mix by inverting several times – 50 times
13. Centrifuge for 15 - 30 minutes
14. Pipette off ethanol (pour off)
15. Invert on the tissue paper overnight
16. Elute with 50 µl of single-distilled sterile water or TE buffer (pH 8.0) or buffer AE; leave in fridge (maintained at 4 ºC) for 3 hours then use of store.
17. Check the DNA quality using Nano-drop and store sample after aliquoting 25ul in different tube

Solution recipe

Cell Lysis Buffer (CLB**)**

10mM Tris (pH 8.0)

0.5% SDS

5mM EDTA

Protein Precipitation Solution (PPS**)**

8M Ammonium Acetate

1mM EDTA

Preliminary

Prepare the ice

Put on the incubator

Put on the centrifuge

Put the CLB and ethanol in fridge (-20 °C)

Prepare another set of tubes for transferring the supernatants