Genomic epidemiology of Rift Valley fever in East Africa

by

John A. Juma

A thesis submitted in conformity with the requirements for the
degree of Doctor of Philosophy

in the
Faculty of Science
South African National Bioinformatics Institute

Supervisors: Prof. Alan Christoffels, University of the Western Cape, Cape
Town, South Africa

AND

Dr. Samuel O. Oyola, ILRI, Nairobi, Kenya

Tuesday 24th October, 2023

Declaration

I, John A. Juma, declare that this thesis "Genomic Epidemiology of Rift Valley fever in East Africa" is my own work, that it has not been submitted before for any degree or assessment at any other university, and that all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

Signature:.........................   Date: Tuesday 24th October, 2023

"Don't be afraid to give up the good to go for the great."

John D. Rockefeller

Abstract

Rift Valley fever (RVF) is a climate-driven zoonotic disease of significant importance to public and livestock health given its epidemic potential. The disease was first identified in 1930 in the great rift valley region of Kenya following massive abortion among pregnant ewes in a sheep farm. RVF is caused by the Rift Valley fever virus (RVFV) belonging to the genus *Phlebovirus*, family *Phenuiviridae* and order *Bunyavirales*. RVF primarily affects domestic ruminants, wild animals and humans with varying degrees of fatalities.

In this PhD research, tools and methods that will be of immense importance in conducting genomic surveillance of RVFV were developed. Furthermore, we have applied phylodynamic and phylogeographic approaches to understand the transmission dynamics of the virus in East Africa. In the second chapter, we utilized publicly available genetic sequence data of the virus to build a tool that rapidly detects circulating lineages. A general observation made in this work is the paucity of RVFV genetic data globally. Only a handful of sequence data is available from countries that have previously experienced RVF outbreaks such as South Africa, Kenya and Madagascar. The genetic information in the sequence data is crucial in identifying lineage defining single nucleotide polymorphisms (SNPs) or mutations. Using these mutations, we developed a command line tool that rapidly characterizes RVFV isolates using the 15 alphabet (A to O) nomenclature. A web application version of the tool is also available for users with minimal command line experience. We demonstrated the robustness of the tool by evaluating it against publicly available and clinical outbreak sequences generated during this project. From both datasets, we observed high sensitivity, specificity and accuracy of the tool in classifying RVFV isolates. In the third chapter, we devised a cost effective and time efficient procedure for generating RVFV genomes. The method relies on designing primers across the virus genome that can be used in a single PCR reaction to amplify and recover whole virus genetic material. The proposed amplicon multiplex PCR enrichment (amPCRe) method is inspired by the short fragment sequencing approach applied to Zika, Ebola and Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2). Although tested on a short read sequencing platform, the utility of the method can also be applied in long-read sequencers such as Oxford Nanopore Technologies (ONT). We demonstrate the applicability of the method in generating whole genome sequences of the virus by comparing it to both direct sequencing in an unbiased manner (metagenomics) and cell culture enrichment methods.

In the fourth chapter, we applied phylodynamic and phylogeographic models with incorporation of different data types including climatic, epidemiological and spatiotemporal data to understand the transmission dynamics of RVFV in East Africa. Using molecular clock analysis and discrete phylogeography, we show that lineage C, the predominant lineage in sub-Saharan Africa, originated from Zimbabwe. We demonstrated the value of phylodynamics in measuring the impact of preventive intervention (vaccination) in the epidemic history of RVFV by estimating effective population size of the virus. Through continuous phylogeographic analysis, we described the spread of the virus in East Africa. We applied phylogeographic landscape analysis to characterize how fast or slow the virus spreads. The data shows that the virus spreads at a moderate speed of between $50 - 90$ km/year compared to other zoonotic viruses like Ebola which spread rapidly. From the analysis, we also inferred that the virus has covered a substantial amount of distance, approximately 3000 km (about 1864.11 miles), in East Africa since it was first reported. Finally, We showed that the virus tends to disperse and remain in areas with high human population such as urban areas and farmlands where farming activities are common.

Genomic surveillance has become relevant in this era where the human population is facing unknown dangers of zoonotic diseases. This PhD work has successfully developed tools and methods for generation of RVFV genetic sequence data. With these methods in place, genomic surveillance of RVFV can be effectively implemented to characterize and track circulating variants. Additionally, the sequence data generated is relevant in development of diagnostics and vaccines that can be used as interventions to avert further outbreaks of the disease. Rift Valley fever is a classic example of a disease where One Health approach can be utilized. This work has demonstrated this by integrating various data sources to describe the spread and associated epidemiological characteristics including potential drivers of transmission. Such a model can be further improved by incorporating risk-maps to identify areas that require intervention measures.

Keywords

Amplicon, Bayesian, Enrichment, Epidemiology, Genomic, Lineage, Maximum Likelihood, Metagenomics, Multiplex, Phylodynamics, Phylogenetics, Phylogeography, Rift Valley fever, Segment, Sequencing, Surveillance, Transmission dynamics.

Acknowledgements

My deep gratitude goes to my supervisors Prof. Alan Christoffels and Dr. Samuel Oyola for the invaluable guidance and endless support during my PhD journey. I particularly thank Dr. Oyola for the devotion and dedication in ensuring that I stayed focused on achieving my goals. The informal discussions and shared moments were helpful in shaping me as a young scientist. I take pride as his first PhD student and having interacted with him has allowed me to experience real mentorship in my professional career. Samuel gave me the opportunity to shape my PhD project while allowing me to learn, share and acquire new skills through training, workshops and giving conference talks and presentations. I deeply thank Prof. Christoffels for allowing me to join his lab as a graduate student. Alan has been a reliable supervisor throughout the PhD period, often checking on the status of things and showing genuine concern.

I would also like to thank Dr. Bernard Bett (ILRI), Prof. Rosemary Sang (KEMRI) and Dr. Kristina Roesel (ILRI). They have been of immense significance in guiding and offering invaluable insights into the different studies I undertook. Bernard, Rosemary and Kristina have been so reliable in reviewing my manuscripts and suggesting innovative approaches from diverse perspectives. I also express my gratitude to my PhD buddy, Dr. Limbaso Konongoi (KEMRI) who was extremely helpful in teaching me the techniques of virus cell culture.

I also would like to express my gratitude to Prof. Tulio de Oliveira (Stellenbosch University) and Dr. Vagner Fonseca (WHO-PAHO) for the collaboration and support in developing the web application for RVFV lineage assignment.

My sincere thanks to the SANBI (South Africa National Bioinformatics Institute) family, specifically Peter van Heusden for his guidance and contribution on RVFV tool development, Fungiwe Mpithi in ensuring administrative compliance. I also thank colleagues at the Genomics lab at ILRI, specifically Collins, Edward, Paul, Shebbar, Daniel (currently in Ireland) and Gilbert.

I also acknowledge the sincere mentorship offered to me by Prof. Daniel Masiga (ICIPE). My sincere gratitude also goes to Dr. Jean-Baka Domelevo Entfellner and Alan Orth for their support in availing and updating computational resources on ILRI high performance cluster.

I would also like to appreciate the Capacity Development Unit at ILRI led by Dr. Wellington Ekaya for the conducive environment that nurtured my soft skills and honoring me as the best presenter, PhD category on the 3-minute Research pitch in 2022.

Contents

List of Figures

List of Tables

Abbreviations

aLRT	approximate Likelihood Ratio Test
amPCRe	Amplicon Multiplex Polymerase Chain Reaction Enrichment
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BF	Bayes Factor
BHK	Baby Hamster Kidney
BLAST	Basic Local Alignment Search Tool
BSL	Biosafety Level
BWA	Burrows-Wheeler Aligner
CCHFV	Crimean Congo Hemorrhagic fever virus
CCE	Cell Culture Enrichment
CDC	Centers for Disease Control and prevention
CPE	Cytopathic Effect
cRNA	complementary Ribonucleic Acid
CTMC	Continuous-Time Markov Chain
DIVA	Differentiating between Infected and Vaccinated Animals
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme Linked Immuno Sorbent Assay
ENSO	El Niño southern Oscillation
ESS	Effective Sample Size
FAO	Food and Agriculture Organization
FN	False Negative
FP	False Positive
GBS	Guillain-Barré syndrome
GFP	Green Fluorescent Protein
GPS	Global Positioning System
GTR	General(ized) Time Reversible

HKY	Hasegawa-Kishino-Yano
HPD	Highest Posterior Density
IEP	Inter Epidemic/Epizootic Period
IFN	Interferon
IgG	Immunoglobulin G
IgM	Immunoglobulin M
ILRI	International Livestock Research Institute
kD	kilo Dalton
KEMRI	Kenya Medical Research Institute
LARISSA	Live-Attenuated Rift Valley fever vaccine for Single-Shot Application
LSDV	Lumpy Skin Disease virus
MEP	Measurably Evolving Populations
MAFFT	Multiple Alignment using Fast Fourier Transform
MCMC	Markov Chain Monte-Carlo
MHC	Major Histocompatibility Class
MLE	Maximum Likelihood Estimator
MRCA	Most Recent Common Ancestor
mRNA	messenger Ribonucleic Acid
NCBI	National Center for Biotechnology Information
NJ	Neighbor Joining
nm	nanometer
NSs	Non-structural
NP	Nucleocapsid
OIE	Office International des Epizooties
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PKR	Protein kinase RNA-activated
PFU	Plaque Forming Unit
PP	Posterior Probability
PPE	Personal Protective Equipment
PPR	Peste des petits ruminants
PS	Path Sampling

RdRp	RNA-dependent RNA polymerase
RefSeq	Reference Sequence
RNP	Ribonucleocapsid
RPA	Recombinase Polymerase Amplification
RRW	Relaxed Random Walk
RT-LAMP	Real Time Loop-mediated Isothermal Amplification
RT-PCR	Real Time-Polymerase Chain Reaction
RT-qPCR	Real Time-quantitative Polymerase Chain Reaction
RVF	Rift Valley fever
RVFV	Rift Valley fever virus
SAP30	Sin3A-associated Protein 30
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SNP	Single Nucleotide Polymorphism
SpreaD3	Spatial Phylogenetic Reconstruction and Evolutionary Analysis using Data-Driven Documents
SPV	Sheep poxvirus
SS	Stepping-Stone
TAE	Tris-acetate EDTA
TFIIH	Transcription Factor II H
TN	True Negative
TP	True Positive
TPR	True Positive Rate
UCLN	Uncorrelated Log Normal
UN	United Nations
USD	United States Dollar
UTR	Untranslated region
VHF	Viral Hemorrhagic fever
VLPs	Virus-Like Particles
VNA	Virus Neutralization Assay
vRNA	viral Ribonucleic Acid
WHO	World Health Organization
WOAH	World Organization for Animal Health
XPB	Xeroderma Pigmentosum type B

YY1   Yin Yang 1

List of Publications

1. Juma, J., Fonseca, V., Konongoi, S.L. et al. Genomic surveillance of Rift Valley fever virus: from sequencing to lineage assignment. *BMC Genomics* **23**, 520 **(2022)**. https://doi.org/10.1186/s12864-022-08764-6

2. Juma, J., Konongoi, S.L., Mwangi, Reuben. et al. Using multiplex mplicon PCR technology to efficiently and timely generate Rift Valley Fever virus sequence data for genomic surveillance. *Viruses* **15(2)**, 477 **(2023)**. https://doi.org/10.3390/v15020477

3. Juma, J, Christoffels, A & Oyola, S.O. Genomic epidemiology and evolution of Rift Valley fever virus in East Africa. *In preparation.*

Chapter 1

Introduction and literature review

1.1 Introduction

1.1.1 Etiology

Rift Valley fever (RVF) is an acute febrile mosquito-borne zoonotic disease caused by the Rift Valley fever virus (RVFV) in the genus of *Phlebovirus*, family of *Phenuiviridae* and the order of *Bunyavirales* (Adams et al., 2017), that primarily affects animals and humans. It is a well-known livestock disease in Africa and the Arabian Peninsula and is linked with epizootic events often marked by high mortality and abortion rates among domestic ruminants (Bouloy & Weber, 2010). RVFV is serologically related to other phleboviruses but can be distinguished by virus neutralization tests. There is only one RVFV serotype. Inactivation of the virus can be achieved by dissolution in oxidizing agents such as calcium or sodium hypochlorite or in lipid solvents (Mariner, 2018).

1.1.2 History and global distribution

Rift Valley fever virus was first isolated and characterized in 1930 in the great rift valley region of Kenya following an epidemic among sheep (Daubney et al., 1931). In most countries in Africa, numerous RVF outbreaks have been thoroughly documented, highlighting the health and socioeconomic implications of the disease. Epidemiological analyses of RVF circulation and outbreak investigations have been reported through serological surveys and epidemiological investigations (Clements et al., 2007; Nanyingi et al., 2015; Redding et al., 2017). Recurrent epizootics and epidemics have been reported in Kenya (1951; 1961-1963; 1968; 1977) (Murithi et al., 2011), South Africa (1950-1951; 1974-1975; 2008-2011) (Archer et al., 2013; Coetzer, 1982; Pienaar & Thompson, 2013),

1

Egypt (1977-1978) (Meegan, 1979), Senegal (1989) (Wilson et al., 1994), Kenya and Somalia (1997-1998), Saudi Arabia and Yemen (2000) (Ahmad, 2000; Arishi et al., 2006; Madani et al., 2003), Kenya, Somalia, Sudan and Tanzania (2006-2007) (Anyangu et al., 2010; Bird, Githinji, et al., 2008; Hassan et al., 2011; Mohamed et al., 2010; Munyua et al., 2010; Murithi et al., 2011), Mayotte (2007-2008, 2019) (Balenghien et al., 2013; Cêtre-Sossah et al., 2012; Sissoko et al., 2009; Youssouf et al., 2020), Madagascar (2008-2009) (Andriamandimby et al., 2010; Carroll et al., 2011), Mauritania (1987, 1993, 1998, 2010, 2012, 2015) (Boushab et al., 2016; Digoutte & Peters, 1989; El Mamy et al., 2011; Faye et al., 2007; Nabeth, 2001; Sow et al., 2014; Zeller et al., 1995), Senegal (2013-2014) (Sow et al., 2016), Niger (2016) (FAO, 2017; Lagare et al., 2019), Uganda (2016) (Nyakarahuka et al., 2019; Shoemaker et al., 2019), Kenya (2018) (CDC, 2019), South Sudan (2018) (WHO, 2018) and Sudan (2019) (Ahmed et al., 2020).

1.2 Literature review

1.2.1 Transmission

RVF is a vector-borne disease often transmitted to vertebrates by mosquitoes belonging to the *Culicidae* family and of the order *Diptera* (Arum et al., 2015; Fontenille, 1998; Linthicum et al., 1985; Sang et al., 2017). The transmission of the disease is dependent on numerous factors such as the availability of susceptible hosts, competent vectors, suitable environmental and ecological conditions that enhance survival and breeding of mosquitoes (Arum et al., 2015). RVF is transmitted by a range of mosquito species including *Aedes*, *Culex*, *Anopheles*, *Eretmapoites*, *Mansonia* and *Coquillettidia* (Bird et al., 2009; Diallo et al., 2000; Fontenille, 1998). Floodwater *Aedes* (subgenera *Aedimorphus* and *Neomelaniconion*) are the primary vectors of RVF. These primary vectors (*Ae.cumminsii, Ae. circumluteolus, and Ae. mcintoshi*) can maintain the virus by passing it on to offspring through what is termed as transovarial or vertical transmission (Linthicum et al., 1985). The infected mosquito eggs can tolerate drought and survive in dry soils on low lying depressions ("dambos") on land during the interepidemic period (IEP) (Linthicum et al., 1985). In normal rainfall seasons, the enzootic cycle of RVF is kept in existence by low-level virus activity. During such seasons, the vector population is maintained through vertical transmission with intermittent infection and magnification of the virus in susceptible animal hosts including wildlife (Bird et al., 2009; Evans et al., 2008). This period is marked by cryptic transmission of the virus at undetectable levels in animal and human populations. In sub-Saharan Africa, prolonged and abnormal rainfall in grasslands often precede RVF outbreaks. In the eastern Africa region, stretching from the horn of Africa to Tanzania, epizootics often occur towards the end of the year when

climatic events such as El-Niño southern Oscillation (ENSO) are triggered (Linthicum et al., 1999). The phenomenal climatic event of ENSO is often marked by an increase in the sea surface temperatures. This occurs when there is a simultaneous elevation in sea surface temperatures in the eastern equatorial Pacific Ocean and western equatorial Indian Ocean. The increase in sea surface temperatures caused by warm winds leads to excessive rainfall that is 60-100 times above the seasonal average. The above-normal rainfall often pours into the successive year (Linthicum et al., 1999; Linthicum et al., 2016).

Heavy persistent rainfall accompanied with flooding allows the infected mosquito eggs to hatch. Consequently, the emergence of primary vectors initiates transmission to nearby animals and humans where the infected mosquitoes seek blood meals (Arum et al., 2015). This results in multiplication of the mosquito populations and epizootic or epidemic cycles begin to occur (**Figure 1.1**) (Arum et al., 2015; Fontenille, 1998; Linthicum et al., 2016).
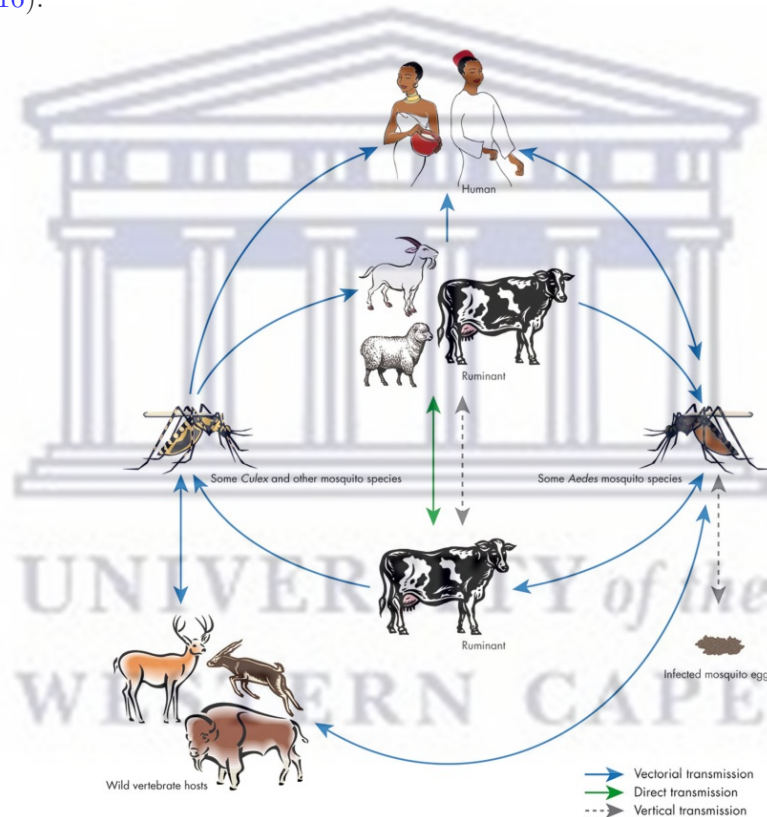
FIGURE 1.1: **Rift Valley Fever transmission cycle**. The virus can be maintained in an enzootic cycle involving *Aedes* mosquitoes, which are able to transmit the virus transovarially. Epizootic outbreaks are often linked with unusual rains, favoring the hatching of infected *Aedes* eggs that are then able to initiate the virus circulation. Subsequently, large numbers of secondary vectors belonging to the *Culex* genus could be infected and induce the emergence of epidemic/epizootic outbreaks. Transmission to humans occurs through direct contact with infected. (Balenghien et al., 2013)

Following the transmission by primary vectors, other mosquito species play the role of secondary vectors. Secondary vectors (genera *Anopheles, Culex, and Mansonia*) act as amplifiers of the virus due to their biting patterns thus leading to outbreaks (Arum et al., 2015; Sang et al., 2017). Domestic ruminants act as the main amplifying hosts of RVFV given the elevated levels of viremia they display once infected (Pepin et al., 2010). Mosquitoes feeding on the ruminants can then initiate animals to human transmission. Humans can acquire an RVF infection through a bite from a virus infected mosquito or from immediate contact with biofluids (saliva, blood or nasal discharges) of infected animals or aborted materials (Fontenille, 1998; Ikegami, 2012; Pepin et al., 2010; Tantely et al., 2015). Placental membranes and aborted fetuses contain large quantities of virus particles which can contaminate the surroundings or neighboring animals (Pepin et al., 2010).

Infection with the virus in humans can also occur through inoculation (e.g., contact with broken skin or needle-stick injuries) and inhalation of aerosolized infectious body fluids (ECDC, 2019; Fontenille, 1998; J. C. Morrill & McClain, 1996). Additionally, human transmission has been linked to various occupational and exposure hazards including skinning or slaughtering of infected animals among abattoir workers, farmers and herders (Anyangu et al., 2010; Cook et al., 2017; ECDC, 2019; Hoogstraal et al., 1979). Veterinary personnel can also acquire infections when helping during calving or abortions because of contact with infected body fluids or material (Clark et al., 2018; LaBeaud, Muiruri, et al., 2011; Nicholas et al., 2014). Other predisposing factors of RVF infection include dietary behaviors such as intake of raw or unpasteurized milk from animals infected with the virus (LaBeaud, Muiruri, et al., 2011; J. C. Morrill & McClain, 1996; Nicholas et al., 2014). There have not been any reports of direct horizontal transmission between humans (Hartman, 2017) although vertical transmission of the infection from mothers to newborn has been reported (I. Adam & Karsany, 2008; Arishi et al., 2006).

Since the identification of the disease in animals, the role of wild vertebrates as amplifier hosts and maintainers of the virus during interepizootic periods has been examined (Evans et al., 2008; Linthicum et al., 2016). During outbreaks in animal populations, mammalian wildlife species become infected although their roles in interepizootic and epizootic periods remain unknown. Numerous serological surveys suggest that many African herbivorous wildlife including the common warthog (*Phacochoerus africanus*), African buffalo (*Syncerus caffer*), giraffe (*Giraffe Camelopardalis*) and other species may be involved in the epidemiology of RVF (Britch et al., 2013; Evans et al., 2008).

1.2.2 Clinical manifestations

Numerous mammalian species show susceptibility towards RVF virus infection with outcomes ranging from subclinical illness to severe hemorrhagic fever including death (Ikegami & Makino, 2011; Pepin et al., 2010). RVF transmission in experimental infection studies have shown that such infections can lead to rapid emergence of acute febrile response and death (J. C. Morrill & McClain, 1996; Pepin et al., 2010). In animals, clinical signs of RVF vary from fever to abortions and are often linked to the age and species of the infected animal (Hartman, 2017; J. C. Morrill & McClain, 1996). There is an inverse relationship between disease progression and severity with age (J. C. Morrill & McClain, 1996; Pepin et al., 2010). Adult sheep and cattle have a case fatality rate of 10-30% or higher, dictated by the nutritional condition of the animal. However, in young animals (less than seven days old), the mortality rates may be up to 100% (J. C. Morrill & McClain, 1996).

The overt hallmark of RVF epizootics has been manifested as 'abortion storm' among domesticated pregnant ruminants (Daubney et al., 1931). The RVF disease has a short-lived period of incubation (one to four days). Often, the disease progresses to a febrile response with rectal temperature ranging between 41 ℃ and 42 ℃. In acutely infected sheep, histopathological findings indicate that there is occurrence of hemorrhage, hepatocellular and lymphoid necrosis (Hartman, 2017; Ikegami, 2012; J. C. Morrill & McClain, 1996; Pepin et al., 2010). Rapid and serious neonatal disease may be the first indicator of RVF in locations with high rates of abortion because of other abortogenic agents such as viruses and bacteria (Cook et al., 2017). Adult animals may die before the occurrence of any clinical signs, a condition termed as peracute death (Hartman, 2017; Ikegami & Makino, 2011; Pepin et al., 2010). In other cases, the disease can progress to acute and chronic phases marked by the onset of inappetence, weakness, bloody nasal discharge, anorexia, rapid decline in milk production, vomiting and diarrhea (Hartman, 2017; J. C. Morrill & McClain, 1996; Pepin et al., 2010). Infected animals show high viremic values ranging from 10^6 PFUs/mL to 10^8 PFUs/mL (J. C. Morrill & McClain, 1996; Pepin et al., 2010). In most infected animal species, the most consistent changes have been observed in the liver (Coetzer, 1977, 1982; J. C. Morrill & McClain, 1996; Peters et al., 1989). Infection with RVFV leads to distinct lesions including hepatic necrosis and gastrointestinal hemorrhage. These characteristic marks give the liver an irregular, fragile and patchy appearance (Coetzer, 1977, 1982; Hartman, 2017; Ikegami, 2012; Pepin et al., 2010). The disease is also marked by elevated levels of viral antigens in the reticulo-endothelial system and in other organs such as the kidneys, gastrointestinal tract, endometrium, adrenal glands and brain parenchyma (Coetzer, 1982; J. C. Morrill & McClain, 1996; Pepin et al., 2010; Van der Lugt et al., 1996).

In humans, RVF infection occurs as a short-term incapacitating illness with an incubation period of between two to six days. The disease is often associated with self-limiting febrile illness manifested with fever, malaise, muscle pains (myalgia), dizziness, weakness, nausea and severe headache (Hartman, 2017; Ikegami, 2012; Meegan, 1979; J. C. Morrill & McClain, 1996; Pepin et al., 2010). Although these signs are reported, they are often resolved within two weeks to a month following medication and treatment. In severe cases of RVF infection, patients develop ocular complications. Such patients exhibit blind or black spots, reduced vision, photophobia, and retro-orbital pain. Assessment of severe cases often reveal retina and eye vessels inflammation as well as retinal hemorrhage (Hoogstraal et al., 1979; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). Fatal cases of infection with RVF involving thrombosis have also been reported (Hartman, 2017; Ikegami, 2012; Ikegami & Makino, 2011). Typical signs of thrombosis-related cases include fatal embolism in the pulmonary vessels and large blood clots in inferior vena cava (Ikegami & Makino, 2011). Fatal cases of RVF infection also involve hemorrhagic symptoms although the time to death varies among the infected patients. Following the short-lived febrile illness phase, such patients acquire macular rash, ecchymosis on arms, eyelids and limbs and bleeding from the gums or gastrointestinal mucosal membrane (Ikegami, 2012; Ikegami & Makino, 2011). Patients showing signs of hemorrhagic disease develop jaundice, low blood pressure, diarrhea, delayed onset encephalitis and liver necrosis (Hartman, 2017; Ikegami, 2012; Ikegami & Makino, 2011; Pepin et al., 2010). Lethal presentation of RVF infection also results in neurological disease, often with a delayed onset. Patients exhibit severe headache, disorientation, hallucination, vertigo, paralysis, excessive salivation and hemiparesis (Hartman, 2017; Ikegami, 2012; Ikegami & Makino, 2011; Pepin et al., 2010).

Currently, the mechanism of species-specific susceptibility to RVF infection is not known. In each animal model, RVF infection shows unique pathogenesis. A thorough understanding of RVF pathogenesis requires identifying and characterizing viral virulence and antiviral factors present in the host. This is particularly important because the replication of the virus and host antiviral responses play an essential role in pathogenicity of the virus (Ikegami & Makino, 2011).

1.2.3 Genetic organization

1.2.3.1 RVFV genome

Like other bunyaviruses, RVFV is an encapsulated RNA virus distinguished by a tripartite genome consisting of three single-stranded RNA segments. The large (L) segment (6404 bp) encodes for the RNA-dependent RNA polymerase (RdRp). The medium

(M) segment (3885 bp) encodes for envelope glycoprotein Gn (encoding achieved by
the amino-terminal sequences of the precursor) and glycoprotein Gc (encoded by the
carboxy-terminal sequences of the precursor) while the small (S) segment (1690 bp)
encodes for non-structural (NSs) and the nucleocapsid (NP) that is encoded in an am-
bisense manner (Hartman, 2017; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010).
In the messenger RNA (mRNA) of the medium segment, four different proteins namely,
78 kD, NSm, Gn and Gc are encoded (Bouloy & Weber, 2010; Ikegami, 2012; Paweska
& Jansen van Vuren, 2014; Pepin et al., 2010). All these proteins are synthesized from
one open reading frame (ORF) of the mRNA at different start codons present in the 5'
untranslated region (UTR) (Ikegami, 2012). The alternate usage of the first or second
start codon present at the 5' UTR encode for the 78 kD and the NSm non-structural
proteins (**Figure 1.2**).
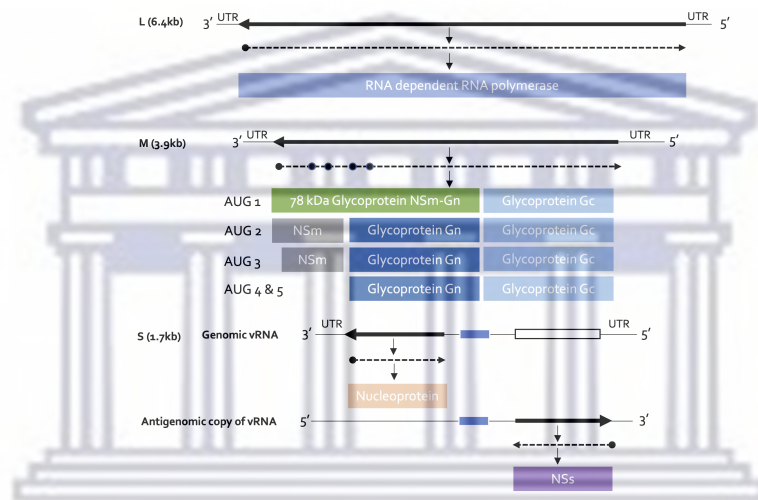
FIGURE 1.2: **Genomic organization of Rift Valley fever virus**. The open
reading frames that are encoded by the different proteins are shown as solid black
arrows on the large (L), medium (M), and small (S) segments of the viral genome.
Untranslated regions (UTR) flank each segment. The size of each segment is indicated
in kilo bases (kb). mRNAs that have been transcribed are depicted as dashed arrows
with the start codons (AUG) indicated as black circles, and the resulting proteins
shown as shaded/colored boxes. AUG1,2, 3, 4 & 5 represent start codons 1, 2, 3, 4 &
5 on the M segment respectively. Post-translational splitting of translated
polyproteins encodes for the different proteins and utilizes the different start codons.
The nucleoprotein is generated by an mRNA transcribed from the negative-sense
genomic viral RNA (vRNA). The non-structural S (NSs) protein is represented on the
genomic-sense S vRNA as a white box and is produced from an mRNA transcribed
from the antigenomic copy vRNA generated during viral replication. The blue colored
box between the nucleoprotein and the non-structural protein represents an intergenic
transcription site (adapted and modified from (Gaudreault et al., 2019)

The initial start codon synthesizes the 78 kD protein composed of the Gn sequence.

The second start codon synthesizes the NSm protein which terminates before the Gn glycoprotein (Bouloy & Weber, 2010; Ikegami, 2012). The second, third and fourth start codons have been postulated to play a role in the synthesis of the glycoprotein, Gn. All the five in-frame start codons upstream the glycoprotein are conserved, indicating the functional requirement that those codons play (Bouloy & Weber, 2010; Ikegami, 2012).

1.2.3.2 Virion structure

Rift Valley fever virus measures between 90 and 110 nanometers (nm) as shown by ultra-structural studies on electron microscope and negative staining (Ellis et al., 1979). A schematic representation of the virus structure is shown in **Figure 1.3**.
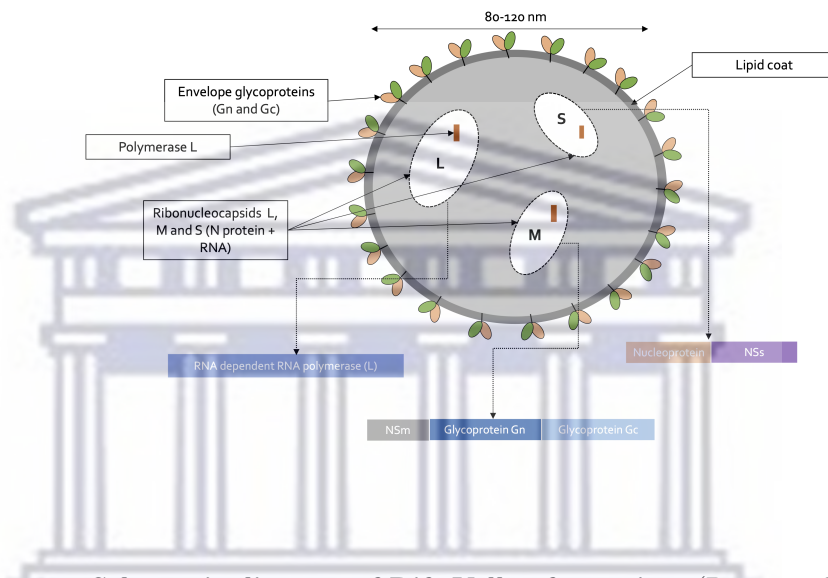
FIGURE 1.3: **Schematic diagram of Rift Valley fever virus** (Pepin et al., 2010)

The virion (virus particle) consists of an envelope and a ribonucleocapsid (RNP) (Ikegami, 2012; Pepin et al., 2010). The envelope is made up of a lipid bilayer having Gn and Gc glycoproteins that make up the surface sub-units, 5-8 nm in length. Compared to other RNA viruses, bunyaviruses do not have the matrix protein connecting the glycoproteins to the viral core. To compensate for the lack of the protein matrix, the cytoplasmic tail of the glycoprotein Gn has a direct interaction with RNPs (Ikegami, 2012; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). For each segment, viral glycoproteins are packaged into the virion (Ikegami, 2012; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010; Sherman et al., 2009).

1.2.3.3 Replication cycle

The cytoplasm of infected cells forms the main compartment where every replication step occurs. Budding of mature virions occurs in the Golgi compartment (Hartman, 2017; Ikegami & Makino, 2011; Pepin et al., 2010). To increase the repertoire of proteins synthesized from their genomes, RNA viruses use several unique RNA translational and editing processes (Gerrard & Nichol, 2007). This is because RNA viruses are constrained by their small genomes (Ikegami, 2012). They lack cap structure at the 5' ends. Like other negative-sense RNA viruses, the replication cycle of RVFV can be divided into three main stages including (i) viral attachment, uptake and fusion, (ii) mRNA transcription, protein translation and replication and (iii) virus assembly and release.

Commencement of the virus-cell interaction occurs through binding of the virus particles to distinct cell surface receptors followed by receptor-mediated endocytosis. The virions' entry mechanism is postulated to employ a class II fusion process turned on by low pH after endocytosis. Following uptake, the virion is channeled along the endocytic pathway towards the perinuclear-localized lysosomes (**Figure 1.4**) (Gaudreault et al., 2019; Paweska & Jansen van Vuren, 2014).
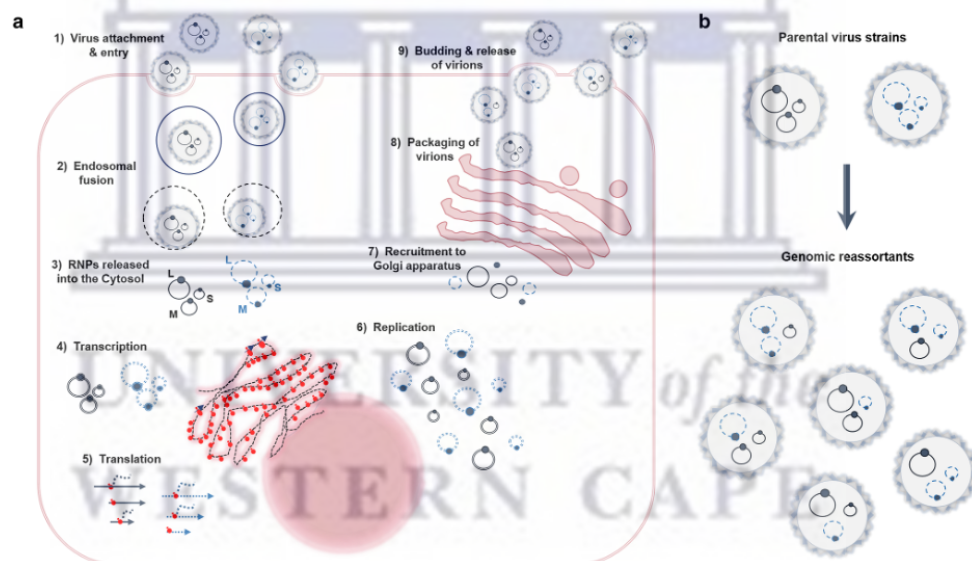
FIGURE 1.4: **Replication cycle of RVFV and emergence of reassortants following co-infection**. **(a)** Depiction of co-infection with different virus strains on the same cell. Once attachment and entry into the host cell has been achieved, viral RNPs comprising of genomic RNAs forming a complex with nucleoprotein and RNA-dependent RNA polymerase (RdRp), the RNPs are released into the cytosol via fusion with endosomes. Linear single-stranded segments are circularized through hybridization of flanking UTRs (Untranslated regions). Assembly and packaging of virus particles occur in the Golgi apparatus. **(b)** Reassortants following co-infection with different strains of the virus are depicted by solid or dotted line segments of the genome (Gaudreault et al., 2019)

Once the viral particles are released into the cytoplasm, primary transcription begins where replication and RNPs assembly occur in viral factories made up of Golgi membranes, actin, and viral proteins. Mature virions are packaged into the Golgi apparatus followed by transportation within vacuoles to cell. Within the cell surface, the binding of the vacuole and plasma membranes facilitates the release of the virus into the host cell. During the replication process, transcription of each viral RNA (vRNA) into messenger RNA (mRNA) and replication occurs through a synthesis step. This results in production of a similar copy of the genome, called complementary RNA (cRNA). The synthesis of non-structural messenger RNA (mRNA) is dependent on a copy of the S ambisense segment (Bouloy & Weber, 2010; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). In the initial virus particles infecting a host cell, there exists a complementary copy of the S segment. This allows for expression of the protein early in the replication cycle, pointing to the key role the segment plays in infection. To initiate messenger RNA synthesis, the virus utilizes an endonuclease activity linked to the L protein. This process also leads to acquisition of a 5' capping mechanism. The cRNA makes use of 5' nucleotide triphosphates (Paweska & Jansen van Vuren, 2014).

1.2.3.4 Cell entry and formation

Glycoproteins play a critical role in the entry of the virus into the host cells. Consequently, efficient and proper processing of glycoproteins is crucial for the formation and budding of virus particles. During infection, glycoproteins are the outermost exposed parts of the virus. As such, the host immune system recognizes these proteins as foreign agents. Upon recognition, the host defense machinery induces the production of neutralizing antibodies which serve to offer immune protection. Additionally, glycoproteins moderate the entry of viral particles into numerous cell types through specific cell surface receptors (Bouloy & Weber, 2010; Pepin et al., 2010). The NSm1 and NSm2 proteins are thought to trigger apoptosis through the caspase 3, 8 and 9 pathways. However, this hypothesis remains inconclusive (Paweska & Jansen van Vuren, 2014).

1.2.3.5 Virus-induced apoptosis and viral virulence

The small segment encodes non-structural protein which has been determined as a major virulence factor blocking host intrinsic viral defense mechanisms (Ikegami, 2012; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010; Terasaki et al., 2016). The NSs gene counteracts the interferon (IFN) system, either at the transcriptional or translational step by downregulating protein kinase RNA-activated (PKR) (Habjan et al., 2009) enzyme.

Primarily, the inhibition counters the antiviral IFN machinery by four main processes including:

i. blocking the transcription of host mRNA by secluding both p44 and Xeroderma Pigmentosum type B (XPB) thereby restraining formation of transcription factor II H (TFIIH) (Kalveram et al., 2011; May et al., 2004; Tirode et al., 1999).

ii. enhancing the breakdown of the TFIIH subunit p62 in the nucleus of the infected host cells (Kalveram et al., 2011; May et al., 2004)

iii. attachment of the Yin Yang 1 (YY1) transcription factor on the interferon-β promoter to obstruct its switching on through the Sin3A-associated protein (SAP30) complex (Ikegami & Makino, 2011; Pepin et al., 2010; Terasaki et al., 2016)

iv. enhancing the suppression of protein kinase RNA-activated (PKR) enzyme early during infection, which averts phosphorylation of initiation factor 2α (Paweska & Jansen van Vuren, 2014; Terasaki et al., 2016)

When the NSs gene is removed, the virus is said to be attenuated. Attenuation of RVFV has been shown in several studies and occurs in naturally attenuated strains and recombinant viruses that lack a functional NSs gene (Bird, Albarino, et al., 2008; Bird et al., 2011; D. R. Smith et al., 2018).

1.2.4 Molecular epidemiology and genetic diversity

RVFV evolutionary history in Africa has been compounded by several factors including environmental and ecological determinants (Pepin et al., 2010). Over the past decades, RVFV gene flow has been influenced by geographical spread and genetic diversity drivers such as reassortment (Pepin et al., 2010). Overall, there are seven major genetic lineages (A to G), although this number has increased to 15 (designated A-O) in the report by Grobbelaar et al. (2011). Different lineages of the virus have been discovered far away from their perceived endemic areas. This implies that the virus can cause large outbreaks outside enzootic areas. High magnitude RVF outbreaks in nations where the disease had not been previously reported, as exemplified by the epidemics/epizootics in Egypt in 1977 (Meegan, 1979), western Africa in 1988 (Wilson et al., 1994), and the Arabian Peninsula in 2000 (Ahmad, 2000; Balkhy & Memish, 2003; Madani et al., 2003) have been documented. The translocation of the virus across large geographic barriers is thought to occur by movement of mosquitoes and livestock mobility through cross-border trades. When the virus adapts to a new environment, it can potentially become established through infection of domestic or wild ruminants. Moreover, in the presence of competent

vectors, virus infected mosquitoes can propagate infection through vertical transmission (Linthicum et al., 1985). Gradual environmental changes, a growing interface between animals and humans, livestock mobility and increased international trade in livestock in and outside Africa are some of the factors that are likely to contribute towards spread of RVFV (Gerdes, 2004; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010).

To promote genetic diversity, segmented viruses like RVFV often undergo genomic rearrangement through reassortment of the segments or recombination. These processes not only contribute towards genetic diversity but also allow for the emergence of potentially new RVFV strains (Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). Within bunyaviruses, reassortment events have been well documented, for example in Batai and Ngari viruses (Briese et al., 2006). Reassortment events in RVFV strains have also been reported (Bird, Githinji, et al., 2008; Freire et al., 2015; Grobbelaar et al., 2011; Sall et al., 1999). However, the effects of reassortment on replication, fitness and virulence in hosts have not been examined in RVFV (Paweska & Jansen van Vuren, 2014).

From phylogenetic studies conducted on RVFV genomic sequence data for all the three segments, there is a tendency of strains from one geographical region to form distinct lineage clades. Interestingly, variants of the virus from distant geographic origins can also cluster within different lineages. This observation offers compelling evidence of widespread circulation and translocation of the virus across sub-Saharan Africa (Grobbelaar et al., 2011). Genetic diversity studies have also indicated that the virus harbors limited or low diversity, approximately 4% and 1% at the nucleotide and amino acid levels respectively (Bird, Khristova, et al., 2007; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). Arboviruses face evolutionary constraints because of the need to replicate in different hosts (mammalian and arthropod). This may partly explain the limited diversity observed in RVFV (Coffey et al., 2008). When compared to other bunyaviruses like the Crimean-Congo hemorrhagic fever virus (CCHFV), RVFV displays a narrow genetic diversity. The close clustering of RVFV isolates observed through phylogenetic analysis can be attributed to the recent emergence of the ancestral virus in Africa and its expansion among introduced colonial sheep and cattle (Pepin et al., 2010). Bayesian phylogenetics has shown that the time of branching of RVFV isolates from the most recent common ancestor (TMRCA) is between 1880-1890 (Bird, Khristova, et al., 2007; Grobbelaar et al., 2011). This concurs with the colonial times when the establishment of susceptible livestock from the European continent would have enabled the emergence of an unknown ancestral virus (Grobbelaar et al., 2011; Samy et al., 2017; Soumaré et al., 2012).

1.2.5 Emergence of endemic lineages, single and multiple introductions

In geographical settings where RVFV spread has not been previously discovered, outbreaks of the disease in livestock and human populations often emanate from a single lineage transmission event. Furthermore, such a virus is distinctively marked by minimum genetic diversity (Bird, Khristova, et al., 2007; Grobbelaar et al., 2011). The early detection of one of the largest RVF outbreaks in 2006/2007 in eastern Africa enabled researchers to perform phylogenetic studies with greater resolution (Bird, Githinji, et al., 2008). In the phylogenetic analysis, an observation highlighted two separate sub-lineages emanating from the 1997/1998 outbreaks and categorized as Kenya-1 and Kenya-2 (Bird, Githinji, et al., 2008; Nderitu et al., 2011). The viruses in the two sub-lineages had increased genomic diversity compared to isolates from the outbreaks in Egypt (1977-1979) and Mauritania (1987) (Bird, Githinji, et al., 2008). Population genetics also revealed that the two sub-lineages of the virus were more linked to the 1997/1998 strains than to each other. Such an observation implies that there could be a continuing and distinct evolutionary pattern following previous outbreaks. Bird and colleagues further highlighted that the Kenya-1 lineage strains underwent recent spatial growth, whereas strains in the Kenya-2 lineage had yet to undergo such an increased evolution (Bird, Githinji, et al., 2008). Molecular clock analysis to infer the time of the most recent common ancestor for Kenya-1 sub-lineage showed that the virus occurred 2-4 years before being detected in 2006/2007 (Bird, Githinji, et al., 2008; Bird, Khristova, et al., 2007). These results indicate that during interepidemic periods, there seems to be ongoing silent transmission and evolution of RVFV. Moreover, the findings emphasize the significance of a cryptic transmission cycle that propagates RVFV endemicity which usually leads to silent outbreaks beyond detection levels. As such, these small-scale outbreaks remain unreported (LaBeaud et al., 2008; Paweska & Jansen van Vuren, 2014). In Kenya, the average interepidemic period (IEP) is approximately four years (Murithi et al., 2011). In this period, it is postulated that immunity in livestock populations declines to levels that allow the virus to circulate without detection (LaBeaud et al., 2008). Cryptic transmission of RVFV during interepidemic period without noticeable outbreak or clinical cases h as been reported in humans in Kenya (LaBeaud et al., 2008; Lichoti et al., 2014; Mbotha et al., 2018), Tanzania (Heinrich et al., 2012; Sumaye et al., 2013), and in wildlife (Evans et al., 2008; LaBeaud, Cross, et al., 2011). It is also during the interepidemic periods that the virus continues to exist in the eggs of Aedes mosquito species or through low-level circulation between mosquitoes and vertebrates (Arum et al., 2015). Currently, the extent to which RVFV has become established in non-endemic countries is still unknown.

1.2.6 Diagnosis

Given the rapid onset and potential of RVF to cause deaths in livestock and humans, there is a need for establishment of national and regional reference diagnostic laboratories with essential skills and expertise. Currently, clinical testing is only available through national or international reference laboratories such the Kenya Medical Research Institute (KEMRI) in Kenya, Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, (United States of America) USA, Onderstepoort Veterinary Institute in South Africa and Unité des Stratégies Antivirales, Institut Pasteur, France (Mariner, 2018). RVFV is one of the viral hemorrhagic fevers (VHF) agents including Ebola, Marburg, Lassa, Crimean-Congo hemorrhagic fever, Dengue, Yellow fever and Hantavirus viruses (Pepin et al., 2010). Consequently, any delays or misdiagnosis of RVF in the initial stages of an outbreak have dire implications. Therefore, diagnosis of RVF is critical since the clinical symptoms of the disease in livestock are not distinctive even though abortion storms are a hallmark of infection (Daubney et al., 1931). As such, differential diagnosis from other infections such as brucellosis and Q fever ought to be conducted as potential co-infecting pathogens (Hartman, 2017; Mansfield et al., 2015). Differential diagnoses can also be undertaken on diseases linked to exotic viral pathogens such as Peste des petits ruminants (PPR), rinderpest, Nairobi sheep disease, Wesselsbron's disease and non-endemic bluetongue (Mansfield et al., 2015). An important characteristic symptom that can distinguish RVF from diseases such as bluetongue, PPR and rinderpest is the occurrence of oral lesions in these infections (Mansfield et al., 2015). According to the World Organization for Animal Health (WOAH) (founded as Office International des Epizooties (OIE)), it is a requirement that confirmatory laboratory tests from at least two positive outcomes are obtained by different diagnostic procedures for RVF incidence reporting in animals. Preferably, these tests should be performed for the same specimen (WOAH, 2022). In animals, infection with RVFV is often exhibited by high viremia levels, and no tissue preference (can affect multiple tissues). Given the high-titer viremia and no tissue specificity during infection, diagnosis of the disease can be performed using several specimen types (Pepin et al., 2010). In settings without cold chain capability, sampling of specimens (tissue such as liver or aborted fetus material) for diagnostic purposes can be preserved in formaldehyde (Mansfield et al., 2015). However, high risk factor for RVF transmission is handling infected animals without personal protective equipment (PPE). Procedures such as necroscopy and histopathology therefore pose serious risk to veterinarians and other personnel. These procedures should be carried out by trained personnel who use appropriate personal protective equipment (PPE). Histopathological assessment of the infected liver often indicates hepatic lesions characteristic of RVFV (Coetzer, 1982). To identify specific RVF viral antigen in a tissue, immune-staining can be performed (Mansfield et al., 2015).

1.2.6.1   Serological assays

Serological assays are usually performed through enzyme-linked immunosorbent assay (ELISA) or virus neutralization assay (VNA) (Meegan, 1979; J. Morrill et al., 1989; WOAH, 2022). Serology methods are routinely applied to confirm infection with RVFV in affected individuals mostly during outbreak interventions and to establish sero-prevalence levels for surveillance purposes (Mansfield et al., 2015; WOAH, 2022). Immunofluorescence assays are still being used although there is a potential limitation of cross-reactivity with other phleboviruses (Mansfield et al., 2015). VNA remains the gold standard serological technique, and often used to determine potency of vaccines. It is also the WOAH prescribed test for international trade in livestock. It is highly specific and can be used for RVFV detection using serum from various hosts (Mansfield et al., 2015; WOAH, 2022). Virus neutralization assays can only be carried out in appropriate biosafety facilities since they involve manipulation of live viruses which pose a risk of RVF infection. Immunoglobulin M (IgM) antibodies are often measurable from days four to six and last for about three months while Immunoglobulin G (IgG) can be typically detected eight days post infection and remain for numerous years (Hartman, 2017; Mansfield et al., 2015; Paweska et al., 2005; Pepin et al., 2010; Williams et al., 2011). IgM detection suggests an ongoing or recent infection whereas IgG-based ELISAs detect previous infections. IgG-based ELISAs can determine the stage of infection when performed in combination with viral RNA molecular detection approaches (Pepin et al., 2010). In addition, determining the concentration of IgM antibodies is critical for detecting acute viral infections. IgM titration in combination with molecular and IgG detection assays can be used to ascertain the time of infection (Paweska et al., 2005). In cattle, the examination of anti-RVFV IgM and IgG antibody titers can be used to distinguish recent from past RVFV infection since the duration of detectable anti-RVFV IgM antibody is transient (50 to 90 days) (Morvan et al., 1992). To detect specific antibodies in human and animal sera, indirect ELISAs developed from recombinant nucleocapsid protein have also gained widespread application (Yu et al., 2018). The dual assay is intended to differentiate between infected and vaccinated animals (DIVA). Consequently, such an assay is extremely useful when considering the regulations required for vaccine licensing and the financial implication of movement constraints on suspect animals (McElroy et al., 2009).

1.2.6.2   Molecular assays

Rapid and economical virus diagnosis during an acute febrile phase when viremia is high can be achieved by detection of viral RNA (Hartman, 2017; Mansfield et al., 2015; Pepin et al., 2010). Numerous highly specific and sensitive nucleic acid-based assays have

been developed for the detection of RVFV, especially during outbreaks. These assays can be exemplified by reverse transcriptase polymerase chain reaction (RT-PCR) (Sall et al., 2002), multiplex PCR-based microarray assay (Drosten et al., 2002), quantitative (real-time) reverse transcriptase polymerase chain reaction (RT-qPCR) (Bird, Bawiec, et al., 2007; Mwaengo et al., 2012), reverse-transcriptase loop-mediated isothermal amplification (RT-LAMP) (Le Roux et al., 2009) and recombinase polymerase amplification (RPA) (Euler et al., 2012). WOAH recommends that viral RNA detection is performed using locally validated methods such as RT-PCR or real-time RT-PCR (WOAH, 2022). Despite their usefulness in detecting RVFV during outbreaks, molecular techniques work efficiently within the febrile phase when viremia is high. Additionally, these assays require specialized skills and equipment which may be lacking in resource-limited settings (Mansfield et al., 2015). Nonetheless, these methods can rapidly identify genetic lineages of the virus. Therefore, these assays can ascertain the pathogenicity of circulating viral strains during outbreaks (Bird, Khristova, et al., 2007). During RVF infection, the viral load is a predictive measure of the disease outcome. Therefore, the application of quantitative real time molecular methods may allow clinicians to practice targeted treatment, especially on individuals with high viral loads. Molecular assays can also be used for detecting RVFV in mosquitoes thereby enhancing RVF surveillance and vector species identification (Lutomiah et al., 2014; Mwaengo et al., 2012).

1.2.7 Prevention and vaccines

Vaccination to offer protection against RVF remains a challenge. While there are no licensed vaccines for human protection against RVF, three licensed vaccines for veterinary use are being administered to offer protection in livestock populations (WOAH, 2022). Despite the availability of both live attenuated (Smithburn) and inactivated vaccines, these preventive medicines have shortcomings in their usage especially in RVF endemic areas. There exists a conditional licensure for the live attenuated MP-12 vaccine in the US while Clone 13 can only be used in South Africa and Zimbabwe **Table 1.1** (Mansfield et al., 2015).

TABLE 1.1: Summary of the current RVF vaccine strains (Mansfield et al., 2015)

	Smithburn	Inactivated	Clone-13	MP-12
Virus origin	Mosquito, Uganda, 1948	Field strains (South Africa and Egypt)	Human isolate 1974	Human strain ZH548 (Egypt, 1977)
Attenuation	> 200 passages in mouse brain	Not applicable	Natural deletion in NSs gene	Multiple mutations in S, M and L segments
Production medium	BHK cells	BHK cells	Vero cells	MRC-5 cells
Regimen	Single dose	Requires booster and annual re-vaccination	Single dose	Single dose
Licensed	Yes	Yes	Yes (South Africa and Zimbabwe)	Conditionally licensed in US

1.2.7.1 Inactivated vaccines

The development of vaccines against RVF began with the utility of inactivated virus material. The initial preparation, the Entebbe vaccine, was developed based on a formalin inactivated mosquito acquired virus that had been subjected to several passages in mice (Mansfield et al., 2015). Named NDBR103, this inactivated vaccine underwent 184 passages of the Entebbe strain. Following the passages, it was amplified in the African green monkey cells to generate massive quantities of sera from mice (Pittman et al., 1999; Randall et al., 1962). The vaccine was used to inoculate a group of United Nations (UN) soldiers on a three-dose regimen. Out of the vaccinated individuals, a majority sero-converted with the maximum antibody titers being recorded six weeks following vaccination (Niklasson et al., 1985). Development of this formalin-inactivated vaccine was followed by production of TSI-GSD 200, a variation of the initial preparation that also underwent trial with inoculation on UN troops. Pittman et al. (1999) indicated that three injections of TSI-GSD 200 administered under the skin on days zero, seven and 28 to human volunteers elicited a neutralizing antibody with a geometric mean titer (GMT) of 237. This response of the three primary doses of TSI-GSD 42 days (about 1 and a half months) post vaccination was particularly good (Pittman et al., 1999). Currently, a regular booster administration is deemed necessary for maintaining maximum protection after a series of primary vaccination (Ikegami & Makino, 2009; Mansfield et al., 2015; Niklasson et al., 1985; WHO, 1983). Vaccination side effects on humans of formalin-inactivated vaccines include local reactions such as erythemas, swelling and pain or tenderness at the injection site. No specific adverse effect or febrile reaction related to vaccination with formalin-inactivated vaccine has been reported (Ikegami & Makino, 2009; Mansfield et al., 2015). However, Niklasson and colleagues reported that a few cases experienced severe effects including cardiac arrhythmia and attack of nerves by the immune system termed as Guillain-Barrè syndrome (GBS) (Niklasson et al., 1985). Formalin-inactivated vaccines obtained from a pantropic RVFV strain that underwent serial passages in baby hamster kidney (BHK) cells have also been used for veterinary purposes. When livestock were vaccinated with the vaccine, they showed a reduction in the number of abortions and mortality in newborn lambs (WHO, 1983). The utilization of inactivated RVF vaccines has the ability of increasing the genetic diversity of circulating RVFV. Often, this occurs when there is a failure to entirely inactivate some batches of vaccine thereby leading to re-introduction of the live virus to livestock (Botros et al., 2006).

1.2.7.2   Live attenuated vaccines

To induce and be able to maintain protective immunity against RVF, vaccinations with inactivated vaccines must be administered frequently. On the contrary, live attenuated vaccines can provide protective immunity without repeated booster immunizations. As such, the development of live attenuated vaccines is critically important (Ikegami & Makino, 2009). First developed in Uganda in the early 1950s, the Smithburn, a live attenuated vaccine has been primarily used in livestock. The vaccine was derived from a virus infected mosquito and subjected to several rounds of passages in the brain of a mouse. The Smithburn vaccine is significantly weakened and has been proactively administered in livestock in several countries including Kenya, Zimbabwe, and South Africa. Despite its success, the Smithburn vaccine has been shown to have residual virulence and pathogenicity. Botros et al. (2006) indicated that, there was retention of virulence of the vaccine. The vaccine was shown to lead to abortions in pregnant ewes and illness in young calves and lambs (Botros et al., 2006). The reversion to virulence and deaths in young lambs has been used as a contraindication in areas where RVFV has never been detected (naïve populations). Consequently, studies that seek to ascertain virulence determinants of the vaccine are necessary (Mansfield et al., 2015). The extensive administration of attenuated livestock vaccines obtained from wild-type RVFV may potentially increase the diversity of circulating RVF viruses (Ikegami & Makino, 2009). When a live attenuated vaccine is administered in an outbreak setting, there is a chance of reassortment between wild/virulent and attenuated viruses (Sall et al., 1999). A similar reassortment event was observed in a reassortant RVF virus (SA184/10) in South Africa. This reassortment was identified in an individual who accidentally injected himself while vaccinating a sheep with Smithburn vaccine (Grobbelaar et al., 2011). Phylogenetic analysis revealed that the medium segment of the virus clustered closely with lineage K strains, of which the vaccine strain Smithburn belongs to while the small and large segments were closely related to lineage H strains. This observation was a clear indication of reassortment and pointed towards an infection with wild type RVFV at the time of accidental injection by the live attenuated vaccine strain (Grobbelaar et al., 2011). Reassortment events between a single attenuated and two virulent segments have been reported to possess weakened pathogenicity (Ikegami et al., 2015). As such, a reassortant between a wild and vaccine strain could lead to a virus with protective ability against RVF. However, this can only be achieved if attenuation markers are present in all the segments of the live vaccine strain (Ikegami et al., 2015; Mansfield et al., 2015). Nonetheless, in-depth genomic characterization is needed to unambiguously identify the virulence or attenuation factors at play.

1.2.7.3 Alternative vaccines

1.2.7.3.1 MP-12

Following the 1977-1979 outbreak in Egypt, Caplen and Bishop tried to develop a live attenuated vaccine for both veterinary and human use. They attempted to achieve this by subjecting the ZH548 and ZH501 human derived strains to serial passages with incorporation of a chemical mutagen, 5-fluorouracil (Caplen et al., 1985). The random mutagenesis yielded a live attenuated virus which was sensitive to temperature and possessed mutations in every segment of the virus genome. The presence of mutations in the live attenuated virus genome was considered a beneficial feature in evasion of reversion to virulence. This live attenuated RVF vaccine was referred to as the MP-12. This vaccine strain has been extensively used to characterize important attenuation regions in the RVFV genome. Conclusively, mutations within the medium and large segments have been identified as the primary contributors of attenuation as shown in a murine model (Takehara et al., 1989; Vialat et al., 1997) whereas those in the small segment to a less extent led to the attenuation (Ikegami et al., 2015). Clinical trials following administration of MP-12 have showed minimal teratogenic effects among ruminants (J. Morrill et al., 1997; Nyundo et al., 2019) and satisfactory immunogenicity in non-human primates (J. C. Morrill & Peters, 2011). However, animal experiments with pregnant sheep in South Africa indicated several abortions and teratogenesis in the early pregnancy stages (Hunter et al., 2002).

1.2.7.3.2 Clone 13

The naturally occurring live attenuated strain, 74HB58, isolated from an RVFV-infected patient in the Central Africa Republic forms the basis of Clone 13. The Clone 13 vaccine lacks a huge portion in its non-structural (NSs) gene, a major virulence factor for RVFV (Muller et al., 1995). It is significantly attenuated, up to 69%. It has been shown that a reassortant virus containing a wild type ZH548 strain and a Clone 13-derived small segment has reduced virulence in mice. On the contrary, reassortant viruses having either Clone 13-derived medium or large segments had high virulence in mice (Bouloy et al., 2001; Vialat et al., 2000). The avirulence activity observed in Clone 13-derived small segment reassortant is a pointer to the role of NSs as the virulence factor in RVF infection. It has been demonstrated that immunization of pregnant sheep with Clone 13 induces protective immunity with no teratogenic defects such as abortions (Dungu et al., 2010).

1.2.7.4 LARISSA candidate vaccine

A novel live-attenuated vaccine for veterinary (vRVFV-4S) recently developed by Schreur and colleagues was demonstrated to be safe for susceptible livestock hosts (Wichgers Schreur et al., 2020). This vaccine was developed by splitting the medium segment into two, one encoding NSm, Gn and large glycoprotein (LGp) and one encoding for Gc. To reduce virulence, the vaccine has its NSs gene deleted. The study demonstrated that a single jab of the vaccine among pregnant sheep offered complete protection against RVF infection. Neither did the pregnant ewes pass on the vaccine virus to the young lambs nor were the fetuses malformed during pregnancy. The vaccine does not disseminate in vaccinated animals and therefore can be used as a DIVA assay. In addition, the vaccine is in full compliance with the European vaccine requirements in that it does not shed into the environment (Wichgers Schreur et al., 2020; Wichgers Schreur et al., 2021). A similar vaccine candidate for human use (hRVFV-4S) is in the 1st phase of clinical trials in Belgium (LARISSA, 2022).

1.2.7.5 Recombinant vaccines

The utilization of viral products such as heterologous and antigen vectors has been examined for RVFV (Bieńkowska-Szewczyk & Szewczyk, 1999; Mansfield et al., 2015). Poxyviruses have been successfully utilized in the expression of glycoproteins (Bieńkowska-Szewczyk & Szewczyk, 1999), even for RVFV proteins (Boshra et al., 2013). Induction of neutralizing antibodies and robust CD8+ T-cell immune response in mice through a chimpanzee adenovirus vector deficient of replication has been reported (Warimwe et al., 2013). This heterologous immunity was capable of encoding for the RVFV envelope glycoproteins Gn and Gc (Warimwe et al., 2013). Similarly, the application of another ruminant pathogen has been used to express RVFV glycoproteins. This system was shown to generate a bivalent DIVA vaccine for RVFV, and which also offered protective immunity against Lumpy Skin Disease virus (LSDV) (Soi et al., 2010; Wallace et al., 2006). After vaccination, the vaccine elicited a protective response against RVFV and sheep poxvirus (SPV) in sheep (Soi et al., 2010).

1.2.7.6 Reverse genetics and Virus-like particles (VLPs)

RVFV vaccines based on reverse genetics have also been explored. The ability to generate cDNA copies and grow live viruses in cell culture conditions has played significant roles in the development of novel vaccines (Kortekaas et al., 2014; Stobart & Moore, 2014). Reverse genetics has been implemented to generate a mutant virus derived from the

human ZH501 strain (Bird et al., 2011). This virus lacked the NSm gene located in the medium segment and possessed a green fluorescent protein (GFP) instead of the NSs gene found in the small segment (Bird et al., 2011). In rats, the mutant virus was shown to be significantly attenuated due to the deletion of NSs gene. An invaluable feature of a live attenuated vaccine having viral gene deletions and non-viral insertions is the ability to differentiate vaccinated animals from naturally infected ones. This is because vaccinated animals cannot mount an immune response against deleted proteins but rather can elicit antibodies against inserted genes. A limitation of a vaccine candidate without NSs gene is that a deletion in the mutant strain may not efficiently replicate in animals with normal inborn immunity. Additionally, such a mutant virus may not be able to elicit long-term protective immunity (Bird et al., 2011).

An in-depth understanding of the molecular biology of RVFV has enabled recovering recombinant virus-like particles (VLPs) and intact viruses (Ludwig & Wagner, 2007). VLPs may be generated as substitutes to RVFV vaccines that were classically inactivated using formalin (Ikegami et al., 2006). A virus-like particle for RVFV is a non-replicable and comprises viral envelope G1, G2 proteins and N protein (Ludwig & Wagner, 2007). Following its fusion with a cell surface receptor, the viral glycoproteins are broken down in endolysosomes. The major histocompatibility complex (MHC) class II molecules present antigens for antigen processing. Virus-like particles have superior antigen stability and immunogenicity characteristics compared to soluble proteins (Ludwig & Wagner, 2007). VLPs generated from the nucleocapsid (Pichlmair et al., 2010) and glycoprotein (Liu et al., 2008) have been developed as potential RVF vaccines. This technology has also been utilized to examine the roles of RVFV proteins through recovery of both virulent and avirulent strains (Billecocq et al., 2008; Habjan et al., 2008; Ikegami et al., 2006). Despite its advantages, the main disadvantage of this technology is the lack of acceptance of using genetically modified products, specifically in humans, and also livestock entering the food-chain (Mansfield et al., 2015).

1.3 Pathogen sequencing and epidemiology

To develop effective public health interventions and disease eradication plans, gaining insights into transmission dynamics of pathogens is crucial (Gardy & Loman, 2018; Ladner et al., 2019). The advancement in genomic technologies has revolutionized numerous scientific fields. For example, whole genome sequencing of the human genome has led to development of catalogs of human variation. These technologies are playing a vital role in response to infectious diseases through genomic surveillance (Ladner et al., 2019). In a global world marked by international trade and travel, this connectedness

has impacted pathogen transmission. With increased globalization and urbanization, a localized outbreak can easily metamorphose into a global pandemic (Pybus et al., 2015). The increased mobility allowing for inter-connections between nations may be in part why in the recent past there have been several large-scale infectious disease epidemics, often from unexpected sources, including a novel coronavirus, Severe Acute Respiratory Syndrome coronavirus-2 (SARS-CoV-2) (WHO, 2020), Middle East Respiratory Syndrome (MERS) coronaviruses (Ksiazek et al., 2003; Zaki et al., 2012), H1N1/A influenza virus (Team, 2009), Ebola (Baize et al., 2014; Gire et al., 2014; Holmes et al., 2016; Park et al., 2015), and Zika viruses (Faria et al., 2017; Grubaugh et al., 2017; Metsky et al., 2017).

1.3.1   Genomic epidemiology

Traditionally, epidemiological applications to infectious disease interventions have relied on incidence or case data and contact-tracing. These approaches have been applied to estimate key epidemiological parameters such as the basic reproduction number (R_0), effective reproduction number, serial interval, and incubation period to build up transmission patterns (Ladner et al., 2019). However, in real world settings, such data can be incomplete owing to the labor demanding nature of tracing infected individuals. Moreover, such approaches may suffer from misreporting because of using clinical symptoms to recognize cases (Ladner et al., 2019). Additionally, in large outbreaks settings or epidemics in remote areas and diseases with long incubation periods, these methods' effectiveness is limited (Wohl et al., 2016). Despite the shortcomings exhibited by classical epidemiological approaches, sequencing of pathogens in a massive-throughput approach during outbreaks is transforming how we respond to infectious diseases in a near real-time manner (Gardy & Loman, 2018; Ladner et al., 2019).

The fundamental principle in the utility of pathogen genomic data to inform infectious disease interventions lies in the genetic evolution of a given pathogen. When pathogens evolve, they accumulate mutations which can be used as markers of transmission events in an outbreak setting. Monitoring of these genetic changes is critical in the design of effective therapeutic agents such as vaccines and development of diagnostics (Grad & Lipsitch, 2014; Ladner et al., 2019). Pathogen genomic data sampled at finer resolution with additional metadata such as time and location can be modelled to reveal epidemic transmission patterns. Genomic sequence data sampled during outbreaks can be used to answer several key questions including origin or source of infection, timing, transmission, and chains of spread in a population (Holmes et al., 2016; Ladner et al., 2019; Volz et al., 2013). The outcomes can be used to channel intervention methods to targeted areas

such as hotspots and to determine the impact of specific control strategies such as border closures (Grubaugh et al., 2018; Holmes et al., 2016; Ladner et al., 2019). For instance, this approach was used to inform policies and infection intervention during the Ebola virus epidemic in West Africa between 2013 and 2016. The outbreak was the first in which genomic data was used to inform policies and control strategies (Gardy & Loman, 2018; Holmes et al., 2016). When coupled with epidemiological data in a phylodynamic model, genomic data can be used to estimate effective population size of pathogens and assess intervention measures (du Plessis & Stadler, 2015; Ladner et al., 2019).

1.3.2 Spatial epidemiology and phylogeography

Increased international trade and travel as evidenced by greater mobility can result in pathogen transmissions to new locations. In their destinations or recipient locations, ecological and social factors may contribute to the entrenchment of the introduced pathogens (Pybus et al., 2015). Geography has a significant role in the epidemiology of infectious diseases. However, the impact of global evolution of pathogens in geographical contexts is still under-appreciated. Host movement patterns are significant, especially for RNA viruses, most of which cannot survive outside their host environments for extended periods of time. Consequently, proximity to vectors or hosts is essential for transmission. RNA viruses exhibit high mutation and evolution rates, often accumulating genetic changes in their genomes while undergoing spatial dissemination during outbreaks (Grenfell, 2004; Pybus et al., 2015). As such, the evolutionary and spatial dynamics of these pathogens are related and exert reciprocal influence on one another. This fundamental principle has numerous important consequences. First, genetic differences observed in viruses sampled from distinct locations will carry information about the spatial activities that gave rise to the virus geographical distribution. Second, rapidly evolving viruses can acclimatize to new environments. As they disperse geographically, they gain abilities to alter vector sensitivity and specificity, host preference, drugs or vaccines or immune responses (Ally et al., 2014; Dellicour et al., 2017). Third, sampling in space offers a concept whereby evolution and movement of a virus can be integrated with other data types such as epidemiological or environmental data. The incorporation of geographical and genetic data in robust statistical analysis holds promise in studies that seek to understand the origins, spread and transmission dynamics of pathogens (Pybus & Rambaut, 2009).

The present-day distribution of fast-evolving pathogens in space is the consequence of ecological and evolutionary interactions (Grenfell, 2004). When either spatial or genetic sequence data is considered in isolation, then only a small fraction of insight in transmission dynamics is gained (Real et al., 2005). Technological advancement in sequencing

and robust statistical approaches has enabled integration of spatial and genetic data (Biek et al., 2015; Pybus & Rambaut, 2009). The arrival of affordable mobile global positioning systems (GPS) and their extensive use in disease surveys has transformed geospatial storing and analysis of infectious disease occurrence and prevalence (Hay et al., 2013). Phylogeography is an interdisciplinary study that combines genetic and spatial data to describe geographic dissemination of pathogens (Dellicour et al., 2019; Real et al., 2005). On the other hand, spatial genetics aims at assessing the spread of genetic variation but does not explicitly use phylogenetics (Real et al., 2005). Other genealogical methods such as population genetics use both phylogeographic and spatial genetics. In studying emerging pathogens, phylogenetics methods are commonly used to understand evolutionary processes. This is mainly because rapidly evolving pathogens accumulate mutations which provide sufficient genetic variation at individual levels even in early outbreak stages (Biek et al., 2015; Lemey, Salemi, et al., 2009; Pybus et al., 2015).

Reconstruction of epidemic dispersal from genetic data in space can either treat sampling location as discrete or continuous trait to represent migration as change in trait along the sampled lineages. This is achieved using stochastic models that are not in isolation from molecular evolution (Lemey, Rambaut, et al., 2009). Therefore, sampling locations and collection dates (ages) are of critical importance. Inference of locations of common ancestry in an epidemic can be applied to measurably evolving populations (MEPs) with complex spatial dynamics (Lemey et al., 2010). Furthermore, deduced changes in location on a phylogeny constitute statistically independent examinations, whereas the sampling locations are related because of common ancestry (Pybus et al., 2015). However, the inferred locations of ancestry can be uncertain, especially those that are only distantly related to the sampled ones. Therefore, when applied to datasets containing sequences sampled continuously over time, viral phylogeography can be more informative (Lemey et al., 2010).

One significant objective in phylogeographic analysis is to help identify potential environmental or social factors that are associated with viral movement and velocity (Dellicour, Rose, & Pybus, 2016; Dellicour et al., 2019; Lemey et al., 2014; Pybus & Rambaut, 2009; Pybus et al., 2015). This can be achieved by qualitative comparison between spatially explicit derived phylogeographic reconstructions with environmental or any geographical data (Dellicour, Rose, & Pybus, 2016). Recently, landscape phylogeographic studies formulated and applied this method by parametrizing location exchange rates as a function of differential potential factors. Consequently, the impact of these factors on spatial dispersal can be quantified and examined using genetic sequence data. This has allowed

the integration of virus genetic data and other data sources (epidemiological, environmental and mobility data) to be modelled in robust phylodynamic and phylogeographic frameworks (Dellicour, Rose, & Pybus, 2016; Dellicour et al., 2019; Lemey et al., 2014; Pybus & Rambaut, 2009; Pybus et al., 2015).

Phylogeographic analysis can suffer from biased sampling. Consequently, achieving an optimal sample composition is important (Frost et al., 2015; Viboud et al., 2013). Although a detailed spatiotemporal record may not be a requirement in addressing questions about spread of a pathogen, the accuracy with which genetic variation can demonstrate key patterns depends on the representativeness of sampling (Frost et al., 2015). If there is under-sampling from key locations, the dispersal inference will be under-estimated. Therefore, the inferred ancestral locations will be biased towards over-represented sampling locations (Pybus et al., 2015). Caution must be taken when interpreting phylogeographic inference results. Validation using robust statistical procedures and combining such analyses with epidemiological data is therefore highly recommended (Frost et al., 2015; Pybus et al., 2015).

1.3.3 Rationale of the thesis

The significance of Rift Valley fever virus as a pathogen of human and livestock underscores the need for integrated national, regional, international preventive and surveillance policies. There is a complex dynamic interaction between the virus, vector, host, environment, ecology, and climate factors that govern spread of RVFV. As such, RVF is one of the most important viral threats to both livestock and human health. Controlling RVF can at best be attained through close collaboration between agricultural, veterinary, entomologic, and medical personnel. Consequently, successful control requires a multifaceted intervention strategy that coordinates the efforts of human and veterinary control, preventive measures and surveillance programs utilizing a One Health approach.

This project proposes to characterize the genetic differences of circulating RVFV strains by performing high throughput genomic sequencing. The genomic information generated will help in developing molecular tools useful in various intervention measures. For example, the data will be useful in identifying genetic markers in specific lineages and therefore understand viral adaptation to specific hosts. The RVFV genome has been shown to be stable with high conservation levels. The observed limited variation in its genetic makeup is attributable to several factors. First, this genetic conservation is imposed by the need for the virus to replicate in mammals and vectors. Second, the brief history of dispersion in Africa could be a pointer to its limited variation. Consequently, the presence of highly conserved regions in the RVFV genome should make it possible to

develop diagnostic probes that can detect all RVFV strains. By integrating genomic, spatiotemporal and environmental data, the study will highlight the transmission dynamics and identify potential drivers of the virus spread in eastern Africa. This research will enhance our understanding of how RVFV circulates within its endemic settings, how it spreads outside such settings and how it establishes enzootic presence in new areas. The study hypothesizes that RVF virus has undergone evolution at a measurably detectable rate. The study also assumes that the different circulating strains in Eastern Africa have accumulated mutations that can be used to further improve lineage assignment of the virus.

1.4 Objectives

1.4.1 Main objective

This project aims to establish the genomic epidemiology of Rift Valley fever virus and control in endemic settings by leveraging pathogen genomics and phylodynamic modelling.

1.4.2 Specific objectives

  i. To genetically characterize the current RVFV strain(s) circulating in East Africa through whole genome sequencing on Illumina and Oxford Nanopore sequencing platforms.

 ii. To develop genomic tools for RVFV surveillance and virus lineage assignment.

iii. To infer transmission dynamics of Rift Valley fever virus through phylodynamic analysis of genomic and spatial-temporal data.

 iv. To develop open-source bioinformatic workflows/pipelines for whole genome reconstruction of the virus and lineage analysis.

1.5 Ethics statement and biosafety measures

The study protocol was approved by the International Livestock Research Institute (ILRI) Institutional Ethical Review Committee (IREC), reference ILRI-IREC2020-07 and Institutional Animal Care and Use Committee (IACUC), reference ILRI-IACUC2021-18. All manipulations of the RVF virus from human samples were done in the highest

containment Biosafety Level 3 (BSL3) facilities at ILRI and Kenya Medical Research Institute (KEMRI). Approval (ILRI-IBC2021-09) was obtained from the Institutional Biosafety Committee (IBC) which is registered and accredited by the National Biosafety Authority (NBA) in Kenya.

1.6 Scope

This thesis is organized into five chapters. Chapter one provides background to Rift Valley fever including its etiology, distribution of outbreaks and literature review which offers rich theory on RVF ranging from transmission to prevention measures. The chapter also provides insights into pathogen sequencing which has informed genomic surveillance strategies in the last few decades. Additionally, the chapter offers some theoretical underpinnings of spatial epidemiology and phylogeography which form a tenet for chapter four. Chapter two offers a method for lineage assignment of RVFV and its classification using either a web application or command line tool. The development of this method is crucial in tracking circulating lineages and thereby enhancing genomic surveillance. In Chapter three, an amplicon multiplex PCR enrichment-based sequencing approach is outlined to enhance timely and cost-effective generation of sequencing data for RVFV. In chapter four, Bayesian phylodynamic and phylogeographic methods was applied to offer insights into expansion of lineage C, transmission history, population dynamics and potential environmental factors that influence the spread of the virus. The study concludes by offering a general discussion, summary, and future considerations in prevention of the virus in chapter five.

Chapter 2

Genomic Surveillance of Rift Valley fever virus

Abstract

The evolutionary history of Rift Valley fever virus (RVFV) is complex and has been influenced by dramatic environmental changes throughout Africa in the past few decades. Over this period, RVFV gene flow has been impacted on various levels such as geographic dispersal and reassortment events. Overall, there are 15 lineages, designated from A to O. In many cases, viruses from these lineages have been detected outside enzootic regions through probable movement of infected animals and/or mosquitoes. This has led to large outbreaks in countries where the disease had not been previously reported. Genomic surveillance of virus diversity is crucial in developing control measures. Therefore, we have developed a user-friendly computational tool for rapidly classifying and assigning lineages based on a partial glycoprotein gene and whole genomic segment sequences of the virus. The lineage assignment procedure can be achieved by using partial glycoprotein (Gn) gene sequence within the medium (M) segment or complete genomic sequences of the small (S), medium (M) and large (L) segments. The computational method is presented both as a command line tool and a web application. Validation of the tool has been performed on a large dataset consisting of partial glycoprotein (Gn) gene and whole genome sequences of the large (L), medium (M) and small (S) segments of

RVFV retrieved from the GenBank database. Using the glycoprotein sequences, the
Rift Valley fever virus typing tool was able to correctly classify all 234 RVFV sequences
at species level with 100% specificity, sensitivity and accuracy. All the sequences in
lineages A($n = 10$), B($n = 1$), C($n = 88$), D($n = 1$), E($n = 3$), F($n = 2$), G$n = 2$),
H($n = 105$), I($n = 2$), J($n = 1$), K($n = 4$), L($n = 8$), M($n = 1$), N($n = 5$) and
O($n = 1$) were correctly classified at phylogenetic level. Assignment of sequences to
lineages using whole genome sequences (L, M and S segments) did not achieve 100%
specificity, sensitivity and accuracy. We further tested our tool using genomic data that
we generated by sequencing 5 samples collected following recent RVF outbreaks in Kenya.
All the 5 samples were assigned to lineage C using both the partial gene (Gn) and whole
genome se uence classifiers. The tool is useful in tracing the origin of outbreaks and
supporting surveillance efforts.

The command line tool can be accessed at https://github.com/ajodeh-juma/rvfvtyping
while the web application is available via https://www.genomedetective.com/app/typingtool/
rvfv/.

2.1 Introduction

Rift Valley fever (RVF) is a zoonotic disease caused by the Rift Valley fever virus (RVFV)
and mostly transmitted by mosquitoes (Bouloy & Weber, 2010). Rift Valley fever (RVF)
was first characterized in 1930 in Naivasha, a town located in the great rift valley region
of Kenya following an epidemic in a sheep farm (Daubney et al., 1931). The disease
is climate-driven and has an epidemic potential, especially in areas where it has not
been previously reported (Pepin et al., 2010). RVF affects both animals and humans
and is responsible for increased mortality in human and livestock populations. In sub-
Saharan Africa where small-scale farmers depend on livestock for their livelihoods, RVF
has negatively impacted their economic status (Mariner, 2018). The huge outbreaks of
the disease in the horn of Africa in 1997 and 2007 led to economic losses of approximately
600 million USD (Muga et al., 2015; Rich & Wanyoike, 2010). The disease has affected
several countries including Egypt, Saudi Arabia, Yemen and the Indian Ocean Island of
Madagascar where it is now considered endemic (Andriamandimby et al., 2010; Balkhy
& Memish, 2003; Meegan, 1979; Morvan et al., 1992). Among animals, the disease leads
to high mortality and abortion rates in a phenomenon termed 'abortion storm' (Bouloy
& Weber, 2010). In humans, the presentation of RVF varies with clinical symptoms
ranging from mild to severe (Ikegami & Makino, 2011; J. C. Morrill & McClain, 1996;
Pepin et al., 2010). Typical signs of severity of RVF can be indicated by hepatitis or liver
disease, hemorrhagic fever, retinitis and encephalitis. In humans, the overall case fatality
ratio is estimated to be between 0.5% and 2.0% (Pepin et al., 2010). As a result of the

increasing spread of the virus outside its endemic settings, high number of competent vectors, increased international trade in livestock and climate change, there is need for coordinated efforts to better prepare for an emergence of RVF in disease-free countries (Mariner, 2018; Pepin et al., 2010). Consequently, RVF has been identified and listed by the World Health Organization (WHO) as likely to cause future epidemics in a new emergency plan developed after the Ebola epidemics of 2018 in West Africa (Mehand et al., 2018; WHO, 2018).

RVFV is a tri-partite genome with three segments, large (L), medium (M) and small (S), encoding for different genes essential for its replication, transcription, attachment/entry and virulence. The RVFV genome has been shown to be highly conserved as elucidated by sequencing and phylogenetic studies (Bird, Githinji, et al., 2008; Bird, Albariño, et al., 2007; Grobbelaar et al., 2011). Irrespective of the genomic segment, the diversity at nucleotide and amino acid levels has been reported to be approximately 4% and 1% respectively (Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). Variations in the genome occur as random single nucleotide polymorphisms (SNPs) with no well-defined variable sites. This makes it difficult to distinguish between strains without genome sequencing since there are no well-defined serotypes (Gaudreault et al., 2019).

Genomic surveillance has become a critical tool for elucidating genetic diversity of viruses and is crucial in understanding the emergence and spread of viruses (Gardy et al., 2015). This is particularly important for the development of effective intervention and prevention measures including diagnosis and vaccine initiatives. Moreover, when such a surveillance is undertaken at fine resolution with consistent classification of reported cases, strains likely to cause epidemic potential or disease severity can be detected and characterized (Gardy & Loman, 2018; Gardy et al., 2015). There is a need for methods that can reliably classify arboviruses based on their genomic sequences. In addition, whole genome sequences are often lacking in routine clinical settings. In turn, short gene sequences are often used to attain classification and lineage assignment at viral species level (Fonseca et al., 2019). Here, we present a computational method for lineage assignment of RVFV sequences. The lineage assignment method is implemented both as a web application and command line tool. The web-based application is built on top of the genome detective software tool (Vilsker et al., 2019) while the command line version of the method is implemented in nextflow language (Di Tommaso et al., 2017). The method was validated with a dataset comprising 234 samples using both partial gene and whole genomic segment sequences. The tool was further evaluated using genomic sequences generated from a recent RVF outbreak in Kenya.

2.2 Materials and methods

2.2.1 Lineage assignment and classification

A method that allows for sequence classification and lineage assignment of consensus partial (glycoprotein gene, Gn) and whole genome sequences (complete L, M and S segments) was developed as shown in (**Figure   2.1**). In order to build a database for virus species classification, we downloaded 10,368 (as of 29-05-2021) virus genome sequences from NCBI RefSeq database (O'Leary et al., 2016). This translated to 501,622 protein sequences. A local database was constructed using DIAMOND (Buchfink et al., 2021) with the provision of taxon names, nodes and protein accession to taxonomic identifier files obtained from NCBI.
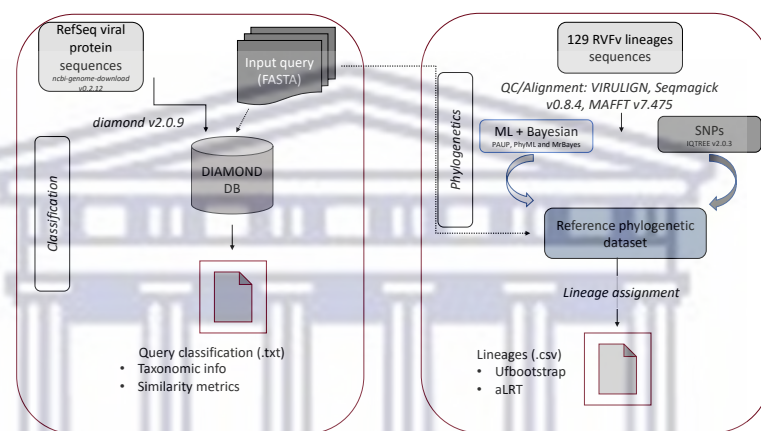
FIGURE 2.1: **Schematic representation of the command line workflow** The workflow begins with virus classification using DIAMOND and reports the output as a text file with taxonomic information and similarity metrics. Phylogenetic analysis is performed using a default phylogenetic reference dataset generated by Neighbor-Joining (NJ), Maximum Likelihood (ML) and Bayesian tree. Users can specify which phylogenetic reference dataset to use. Query sequences are aligned to the reference dataset multiple sequence alignment with MAFFT, and a Maximum Likelihood tree is constructed followed by lineage assignment. An output file with the lineage assignment, bootstrap values and likelihood test ratio is generated in comma-separated values (CSV) format.

2.2.1.1 Glycoprotein (Gn) classifier dataset

Applying the nomenclature implemented in the largest RVF phylogenetic study conducted by Grobbelaar and colleagues (Grobbelaar et al., 2011), we identified representative taxa for each lineage that were used to build the RVFV typing tool. Sequences were compiled from the NCBI nucleotide database (Benson et al., 2012). This dataset, used as reference, comprised 129 partial and whole genome sequences for the M segment. The

annotation of original location, collection date and the originating and submitting laboratory or data generators are shown in the appendix table (Appendix A, **Table A.5**). Recombinant strains as well as duplicated sequences were filtered on sequence composition and an alignment constructed with MAFFT (Katoh & Standley, 2013). Each alignment was manually edited until a codon-correct consensus sequence between positions 815 and 1306 was achieved using seqmagick with the option `--mask 1:815,1306:3885`. These coordinates represent the start and end positions of the 490 base pairs M segment partial glycoprotein (Gn) gene. The suitability of the M segment was chosen due to its high variability 2% and 5% at the amino acid and nucleotide levels respectively compared to the other segments (Gaudreault et al., 2019; Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). The M segment encodes for surface glycoproteins which are targets for neutralizing antibodies and play an essential role in virus attachment. Given this role, positive selection pressure is expected to be responsible for the evolutionary patterns observed in the glycoprotein (Allen et al., 2018).

2.2.1.2 Lineage delineation using single nucleotide polymorphisms (SNPs)

To generate a representative dataset to be used for lineage assignment by phylogenetic inference, we delineated lineages using SNPs. For each lineage sequence, we identified defining SNPs (i.e., those that are shared or conserved within a lineage) using NC_014396 as the reference sequence. Defining SNPs per lineage were considered if they were present in 90% of all the available sequences in each lineage. A target of five sequences or more per lineage was aimed at although some lineages contained only a single strain sequence. Lineages with less than five sequences were all included into the representative dataset. This resulted in 51 unique representative sequences referred to as the Gn classifier. The next step in this exploration involved phylogenetic analysis. A suitable substitution model was identified using ModelTest-Ng (Darriba et al., 2012). Using a general time reversible (Tavaré, 1986) with discrete gamma distributed rate variation among sites (GTR + I + G4) as the optimal model which produced consistent tree topologies with IQ-TREE (Nguyen et al., 2015), a maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE while performing 1000 ultrafast bootstrap approximation and likelihood ratio test to generate a consensus tree.

2.2.1.3 Complete segments classifier datasets

To generate lineage assignment representative SNPs for whole genome segments, 408 sequences were retrieved from NCBI. Out of the 408, only 234 had complete sequences available for each of the three genomic segments. For L, M and S segments, the reference sequences of RVFV strain ZH-548 used were NC_014397, NC_014396 and NC_014395 respectively. The 234 sequences were used to build representative lineage assignment

SNPs from which 47 unique taxa were identified. The 47 unique representative sequences were aligned using MAFFT (Katoh & Standley, 2013) and manually edited followed by construction of a maximum likelihood phylogenetic using a generalized time reversible substitution model with discrete gamma (GTR + I + G4) distribution (Tavaré, 1986).

2.2.2 The Web application

Representative sequences used in the web interface were identified using bootstrap support and posterior probability values. These values were obtained from a maximum likelihood phylogenetic tree using PhyML (Guindon & Gascuel, 2003) and a Bayesian tree constructed using MrBayes (Huelsenbeck & Ronquist, 2001). The trees were visualized in Figtree (Rambaut, 2018). Taxa with bootstrap support of 100 and posterior probability of one were used as the criteria in the selection of reference sequences. The phylogenetic reference dataset using the representative sequences was used to create an automated Rift Valley fever virus typing tool. We selected five to ten sequences that represented the diversity of each virus lineage. We included all the sequences in a lineage where the total number of available sequences was less than five. Sequences that met these selection criteria were checked on quality for the presence of insertions, deletions, frame shifts and non-IUPAC characters using VIRULIGN (Libin et al., 2019). Sequences that passed the quality control were aligned using MAFFT (Katoh, 2002), and subjected to phylogenetic analysis using PAUP* (i.e., Neighbour Joining) (Lemey, Salemi, et al., 2009), MrBayes (i.e., Bayesian) (Huelsenbeck & Ronquist, 2001; Nylander et al., 2008) and PhyML (i.e., maximum likelihood) (Guindon & Gascuel, 2003) using GTR + I + G substitution model. Sequences that gave consistent topologies using all three tree inference methods were retained as potential reference sequences and used in the next step of the evaluation process. A total of 53 unique reference sequences that represent the diversity of each of the lineages were extracted from the dataset.

2.2.3 Similarity search and lineage assignment

In both implementations, sequence classification and lineage assignment involve a similarity search against a database of viral protein sequences followed by phylogenetic analysis. Classification of query sequences was performed using DIAMOND (Buchfink et al., 2021) DIAMOND is a high-throughput program for aligning sequences with high sensitivity against a protein reference database and is up to 20,000 times the speed of BLAST. Phylogenetic analysis for lineage assignment was achieved by construction of a maximum likelihood phylogenetic tree using IQ-TREE. This process was achieved by obtaining an alignment of the query against the reference dataset using the option `--add` in MAFFT. A maximum likelihood phylogenetic tree was constructed using the GTR + I + G4 substitution model with 1000 bootstrap replicates. Single branch test using the

Shimodaira Hasegawa-like approximate likelihood ratio test (SH-aLRT) was performed to assess the bootstrap support values. Polytomies were collapsed if the branch lengths were below a given threshold (default 0.000005). The query sequence was assigned to a particular lineage if it clusters monophyletically with that genotype clade with ultrafast bootstrap support $\geq$ 70. Query sequences whose values were below this cut-off were reported as unassigned. The command line tool was built using nextflow (Di Tommaso et al., 2017), a scalable and reproducible workflow language.

The web implementation of lineage assignment involved construction of a neighbour joining (NJ) phylogenetic tree that was used to make assignments at the lineages level. For this component, the query sequence was aligned against a set of reference sequences using the profile alignment option in the ClustalW software (Larkin et al., 2007), such that the query sequence was added to the existing alignment of reference sequences. Following the alignment, a NJ phylogenetic tree, with 100 bootstrap replicates was inferred. The tree was constructed using the Hasegawa Kishino Yano (HKY) (Hasegawa et al., 1985) distance metric with gamma among-site rate variation, as implemented in PAUP* software (Lemey, Salemi, et al., 2009). The query sequence was assigned to a particular lineage if it clusters monophyletically with a genotype clade with bootstrap support > 70. If the bootstrap support was < 70, the genotype was reported to be unassigned. This web interface is built using the Genome detective framework (Vilsker et al., 2019). Both the command line and the web application produce classification and phylogenetic lineage assignment results as report text files. The report includes static (for command line) and interactive (for web application) phylogenetic trees as data visualization output (**Figure 2.2**).
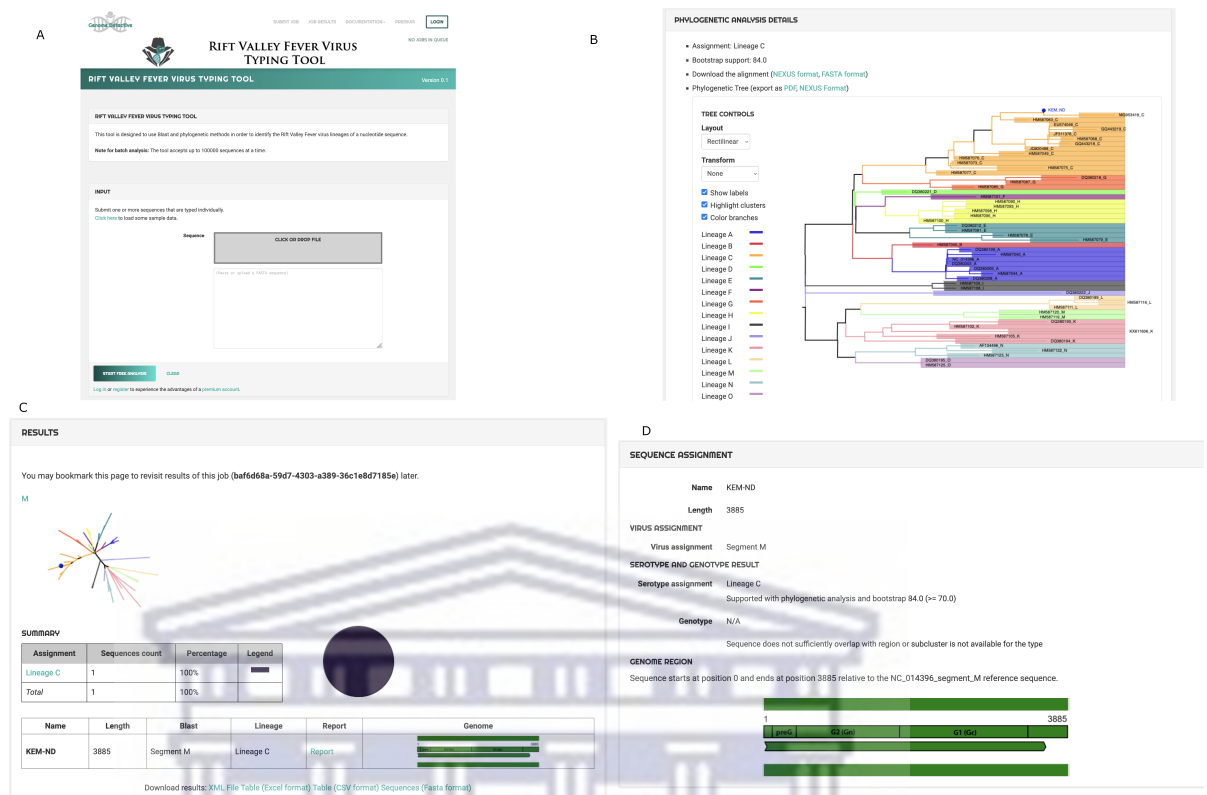
FIGURE 2.2: **Screenshot of the web interface for RVFV typing tool (A)** The web interface offers a portal for users to perform classification and visualize the results. The typing report provides information on the sequence name of the query sequence, the nucleotide length of the sequence, an illustration of the position of the position in the virus' genomic segment, the species assignment and the genotype assignment. A detailed report **(B)** is provided for the phylogenetic analysis that resulted into this classification. All results can be exported to a variety of file formats (XML, CSV, Excel or FASTA format). The detailed HTML report **(C)** contains information on the sequence name, length, assigned virus and genotype, an illustration **(D)** of the position of the sequence in the virus' genomic segment and the phylogenetic analysis section. The alignment section shows the alignment and constructed phylogeny.

2.2.4 Performance evaluation

The True Positive Rate (TPR) or sensitivity, False Positive Rate (FPR) or specificity and accuracy of the proposed method were computed for phylogenetic lineage assignment. These metrics were computed as indicated:

$$TPR = \frac{TP}{TP + FN} \tag{2.1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2.2}$$

$$Accuracy = \frac{TP + TN}{FP + TN + FP + FN} \tag{2.3}$$

where:

$TP =$ True Positives(outcome of lineage classification correctly identified as positive)

$FP =$ False Positives(outcome of lineage classification incorrectly identified as positive)

$TN =$ True Negatives(outcome of lineage classification correctly identified as negative)

$FN =$ False Negatives(outcome of lineage classification incorrectly identified as negative)

2.2.5 Outbreak samples

To assess the functionality of the proposed classification and lineage assignment of the tool, the study utilized samples that were collected following a suspected RVF outbreak in Kiambu county in Kenya in 2021 (**Figure A.1**). From the clinical data collected by veterinary officers, the affected cattle revealed fever (body temperature > 40℃), lack of appetite, abortions and diarrhea that are typical of RVF infection. In one farm, there were reports of abortions in seven cattle and two mortalities.

2.2.6 IgM antibody capture ELISA

All RVF suspected samples were first screened using enzyme linked immunosorbent assay (ELISA) for the detection of anti-nucleoprotein IgM antibodies in sera (Appendix A, **Table A.4**). This was done using the IDVet-ELISA kits as per the manufacturer's instructions (IDVet Innovative Diagnostic, Grabels, France). Briefly, serum samples in a 1:10 dilution were added in plates coated with anti-bovine-ovine-caprine IgM polyclonal antibody. Plates were incubated for 1 hour at 37 ℃, washed and RVF nucleoprotein added and incubated. This was followed by 3 washes, addition of anti-NP-HRP conjugate, and an incubation. A substrate buffer containing 0.1mg/mL 3,3',5,5-tetramethylbenzidine (TMB) (Thermo Scientific, Rockford IL) was then added after the final wash and optical

density (OD) values recorded using an ELISA reader. The presence of IgM antibodies to RVFV was detected by a blue coloration which changed to yellow following the addition of the stop solution. The contents of the wells of the microplate were analyzed at 450nm by the ELISA reader. For each IgM antibody capture ELISA experiment internal controls were included. The optical densities at 450nm obtained from the samples were validated in accordance with the manufacturer's instructions as indicated below:

$$net\ OD = OD_{even\ well}\ -\ OD_{odd\ well} \tag{2.4}$$

The plate was valid if the mean value of the net positive control $OD > 0.35$ and the ratio of the mean values of the net positive and negative control (absolute value of ODs) > 3 ($OD_{positive\ control}/net\ OD_{negative\ control>3}$

For each sample, the percentage of the ratio of sample and positive control (s/p%) was calculated as:

$$S/P\% = net\ OD_{sample}\ /\ net\ OD_{positive\ control} \tag{2.5}$$

Sample presenting a S/P percentage (S/P%):

$$\leq 40\%\ is\ negative$$
$$40\% - 50\%\ is\ Doubtful$$
$$\geq 50\%\ is\ Positive$$

2.2.7 Virus propagation in culture

Virus isolation was performed in containment Biosafety Level 3 (BSL3) facility at ILRI. Vero cells (Gibco-Biocult) were cultured in T25 flasks in growth Media (Minimum essential media, Eagles, Sigma with Earle's salts and reduced NaHCO3) supplemented with 10% heat inactivated Fetal Bovine Serum (FBS), 2% L-glutamine and 2% penicillin/amphotericin and incubated at 37℃ in 5% CO_2. The T25 flasks with confluent monolayers were then inoculated with 200µL serum. Prior to inoculation, the growth media was poured out and after inoculation, the flasks were incubated for 1 hour at 37℃ to allow for virus adsorption. Up to 5mL of maintenance media (Minimum essential media, Eagles, Sigma with Earle's salts and reduced $NaHCO_3$) supplemented with 2% heat inactivated Fetal Bovine Serum (FBS), 2% L-glutamine and 2% penicillin/amphotericin) was then added to each flask and incubated at 37℃ and 5% CO2. Flasks were observed daily for a period of 2 weeks using an inverted microscope for evidence of cytopathic effects (CPE). Flasks showing evidence of CPE were freeze/thawed and cell culture supernatant filtered, harvested, and stored at -80℃ for viral RNA extraction.

2.2.8 RNA extraction

Viral RNA was extracted from 140µL of the filtered cell culture supernatant using QI-Amp Viral RNA kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Cycle threshold (Ct) values were determined for the RNA samples using probe-based reverse transcription quantitative real-time PCR against a highly conserved domain located on the L segment of the virus for RVFV detection (using 5′ Fam reporter dye and 3′ BHQ1 quencher dye). The following primers and probe set were used:

- RVFL-2912fwdGG (`5'-TGAAAATTCCTGAGACACATGG-3'`)

- RVFL-2981revAC (`5'-ACTTCCTTGCATCATCTGATG-3'`)

- RVFL-probe-2950 (`5'-CAATGTAAGGGGCCTGTGTGGACTTGTG-3'`) (Bird, Bawiec, et al., 2007).

A one-step assay composed of 2µL(5-10ng) of the RNA template in a reaction of 15µL using a final concentration of 0.3µM for primers and 0.1µM for probe in a PCR System (Applied Biosystems) was set up. The reaction was carried out in a series of incubation steps as shown in **Table 2.1** below.

TABLE 2.1: **Real-Time Transcriptase quantitative polymerase chain reaction conditions for RVFV detection**

Temperature	Time	
50℃	10 minutes	
95℃	2 minutes	
95℃	3 seconds	40×
60℃	30 seconds	

2.2.9 Library prep

Sequencing libraries were prepared using the total RNA NEBNext Ultra II library prep kit for Illumina (New England Biolabs, Ipswich, USA) following the kit manufacturer's instructions. In brief, removal of ribosomal RNA (rRNA) was performed by hybridization of NEBNext depletion probe to 12µL of sample RNA in a 15µL reaction. The mixture was mixed thoroughly and incubated on a thermocycler with a denaturation step of 95℃ for two minutes followed by a ramp down step to 22℃ (**Table 2.2**).

Ribosomal RNA depleted samples were subjected to digestion using DNase I enzyme. The RNA samples were purified with NEBNext RNA sample purification beads. Purified RNA samples prepared for first strand cDNA synthesis using NEBNext random primers

TABLE 2.2: **Probe hybridization to RNA**

Step	Temperature	Time
Initial denaturation	95℃	2 minutes
Ramp down	22℃	0.1℃/second
Hold	22℃	5 minutes

in a 10µL reaction at 94℃ for 8 minutes. First strand cDNA synthesis was performed using NEBNext first strand synthesis enzyme mix. The reaction was carried in a series of incubation steps as follows: 25℃ for 10 minutes, 42℃ for 15 minutes, 70℃ for 15 minutes and a final hold at 4℃. Second strand cDNA synthesis was immediately performed using NEBNext second strand synthesis buffer and enzyme mix. The samples were incubated for one hour at 16℃ with the thermocycler lid temperature turned off. Double-stranded cDNA products were purified using NEBNext sample purification beads. Clean cDNA samples were end repaired and adaptor ligated using NEBNext adaptors with dilution buffer (25-fold dilution). The NEBNext USER enzyme was added to the ligation reaction mixture. Adaptor-ligated samples were cleaned using sample purification beads. Library PCR enrichment was performed using NEBNext Ultra II Q5 master mix and premixed NEBNext oligos in a 50µL reaction for 15 cycles (**Table ??**).

TABLE 2.3: **Adapter-ligated PCR enrichment**

Step	Temperature	Time	
Initial denaturation	98℃	30 seconds	
Denaturation	98℃	10 seconds	
Annealing/Extension	65℃	75 seconds	15×
Final extension	65℃	5 minutes	
Hold	4℃	∞	

PCR-enriched libraries were cleaned using sample purification beads in a 1:0.9 ratio. Quality assessment of the libraries was performed through gel electrophoresis (1% agarose gel in 1× TAE buffer). The libraries were quantified using KAPA library quantification kit (Roche, Boston, USA) and normalized to obtain a final concentration of 4nM. Pooled libraries were denatured and diluted for loading at a concentration of 1.5pM for sequencing on an Illumina NextSeq550 instrument. The run was set for a paired-end 2×150 bp sequencing using the mid output cartridge. The run took < 26 hours to obtain 47.7Gb of compressed sequencing data.

2.2.10   Virus genome assembly

The consensus genome for each sample was generated using an in-house nextflow pipeline, viclara (Juma et al., 2022). In brief, the demultiplexed data files were merged for each lane for both forward and reverse reads. The raw data was checked for quality metrics using FastQC v0.11.9 (Andrews, 2010). Low quality reads and adaptors were trimmed using fastp (Chen et al., 2018). Alignment to the GenBank reference genome was performed using Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009). Alignment metrics were generated with SAMtools (Li et al., 2009) and only samples surpassing a minimum mapping threshold of 500 were kept for subsequent downstream analyses. Using each segment's reference, ZH-548 strain and genome annotation file, variants were called on the pileup generated from the marked duplicates alignments. Only positions with ≥ 10 $\times$ genome coverage and with ≥ 20 base qualities were used to produce consensus alleles. Regions with lower coverage and base quality were masked with N characters. Genome wide coverage was computed using BEDTools (Quinlan & Hall, 2010). The accession numbers of the consensus genome sequences are available at OM744365-OM744379 on NCBI. Raw sequence reads were deposited to the sequence read archive (SRA) database with the accession PRJNA811331.

2.3 Results

2.3.1 Identification of lineage assignment SNPs

In searching for RVFV sequences on the NCBI database, there is a general paucity of the virus sequence data. The number of viral sequences per country also varied with South Africa, Kenya, Zimbabwe and Egypt having the majority (> 10). This variation in the number of sequences available in public databases can be attributed to the frequency of outbreaks and the effort by individual countries and their partners to sequence and report the RVFV strains (**Figure 2.3**). For each of the four lineage classifying sequence datasets (Gn, S, M and L), defining SNPs were identified for all the available sequences. For the Gn sequences, a total of 121 lineage assignment SNPs distributed across all the 15 RVFV lineages were identified (**Table 2.4**).
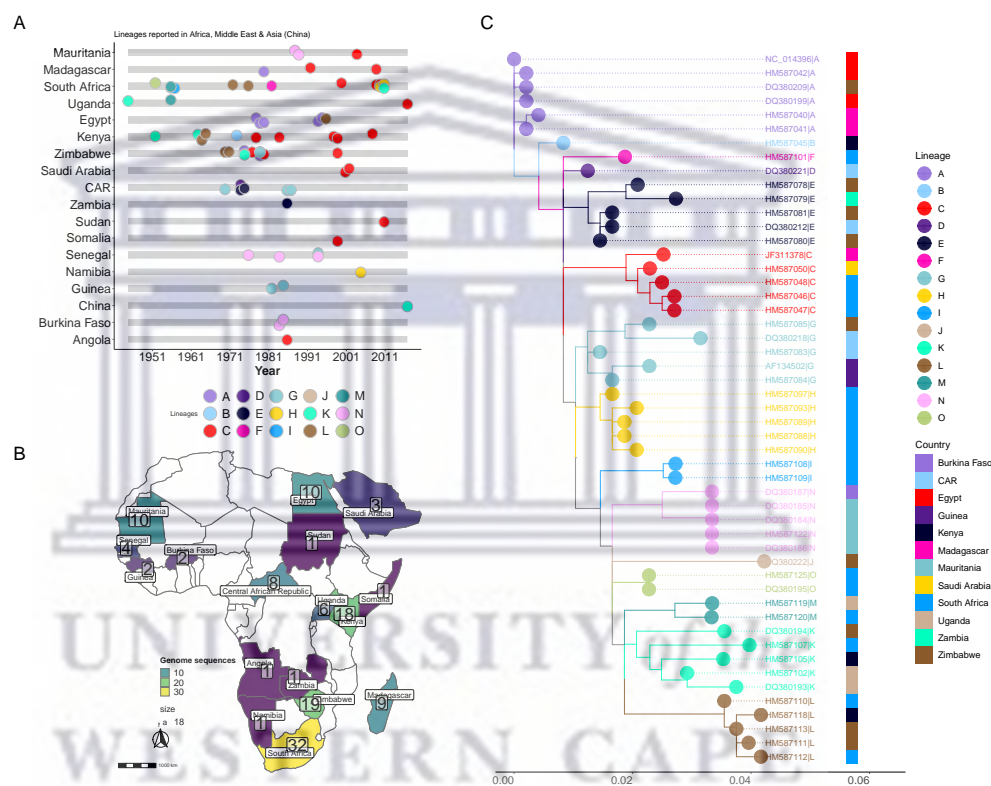
FIGURE 2.3: **Distribution of RVFV lineages in Africa and Middle East. (A)** Lineages reported in Africa and the Middle East (Saudi Arabia) sampled between 1944 to 2016. **(B)** Map of Africa and Saudi Arabia indicating the number RVFV sequences for the M-segment (partial and complete) as of 28th May 2021. **(C)** Maximum likelihood tree representative sequences used in typing and lineage assignment of query sequences indicating the different lineages. The tips of the tree depict RVFV lineages while the vertical bar shows the geographical origins of the representative sequences. **CAR, Central African Republic**.

TABLE 2.4: **RVFV Lineage defining single nucleotide polymorphisms (SNPs)**. For each lineage sequences, SNPs were identified in comparison to the reference (strain ZH-548).

Lineage	SNPs	Count
A	-	1
B	830GA;1103TC;1142TC;1304GA	4
C	836TA;926GA;1103TC;1163CT;1190TC;1241AG	6
D	839TC;926GA;1103TC;1142TC;1163CT;1195GA	6
E	854TA;926GA;1103TC;1142TC;1163CT;1166AG	6
F	816AG;902GA;926GA;1079GA;1103TC;1106GA;1142TC;1163CT;1253GA	9
G	926GA;1103TC;1142TC;1163CT	4
H	920AG;926GA;1103TC;1142TC;1157AG;1163CT	7
I	833CT;920AG;986CT;998TC;1049GA;1103TC;1115GA;1142TC;1163CT;1304GA	10
J	836TC;860CT;920AG;926GA;953AG;995GA;1007CA;1055TC;1115GA;1142TC;1154GA;1160GA;1161TC;1163CT;1190TC;1250TC	16
K	894CT;1091TC;1115GA;1142TC;1250TC	5
L	842GA;866CT;917CT;920AG;926GA;1103TC;1115GA;1122CT;1124AG;1142TC;1163CT;1190TC;1250TC;1274AT;1304GA	15
M	857GA;894CT;920AG;924TC;926GA;992GT;1103TC;1115GA;1142TC;1151TC;1163CT;1250TC;1304GA	13
N	920AG;926GA;1103TC;1112GA;1115GA;1142TC;1163CT;1187GA;1304GA	9
O	920AG;926GA;1103TC;1106GA;1115GA;1142TC;1163CT;1205AG;1243AG;1250TC;1304GA	11

2.3.2 Maximum likelihood phylogenetic trees for lineage clustering

Using the unique representative sequences for both Gn ($n = 51$) and whole genome sequences ($n = 47$) for the three segments (L, M and S), we constructed maximum likelihood phylogenetic trees as shown. We observed striking similarity in the tree topology generated with both the glycoprotein (Gn) gene and RVFV whole genome sequences. As is expected, each lineage formed a distinct cluster shown as monophyletic clades (**Figure 2.4**). This indicates successful classification by the assignment tool.
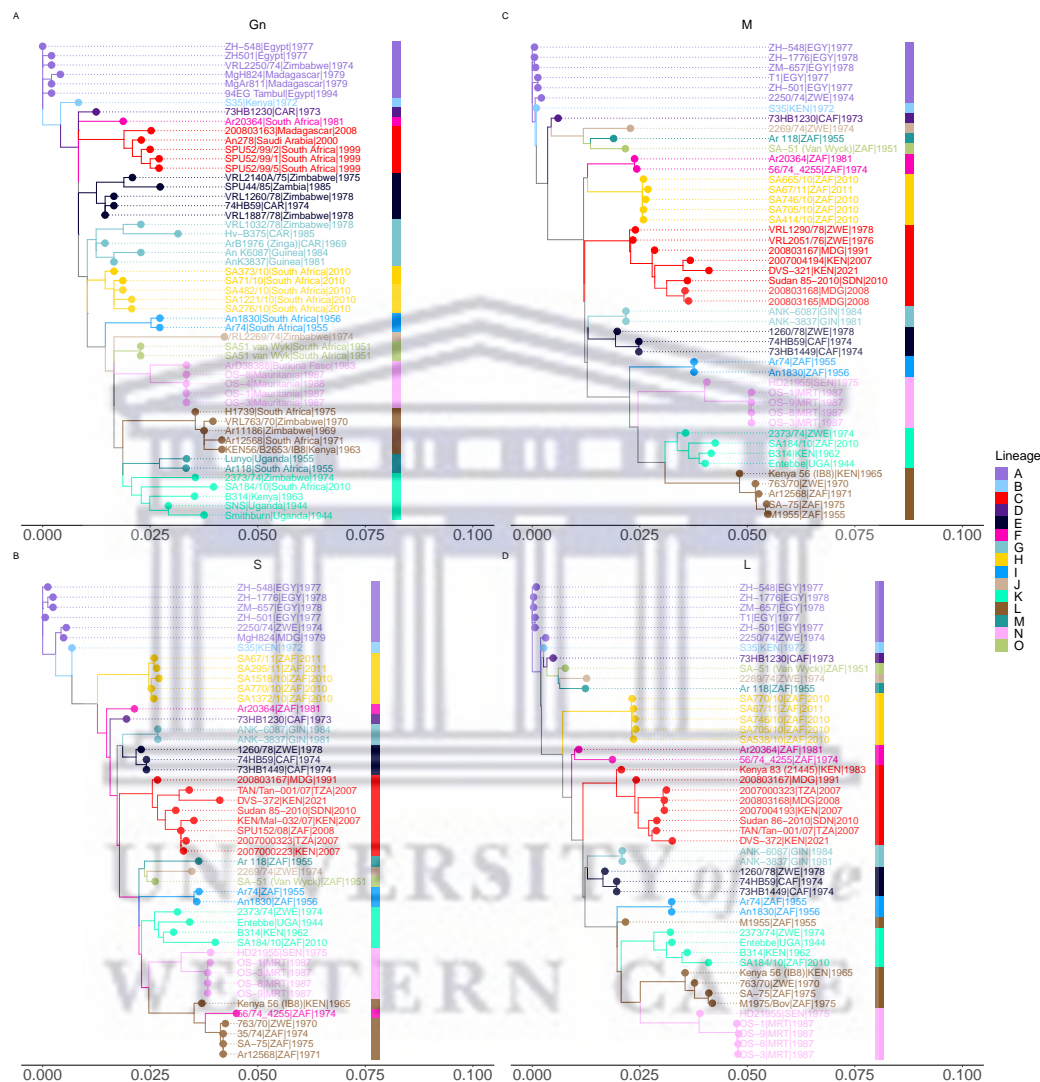
FIGURE 2.4: **Phylogenetic analysis using Gn and whole genome (L, M & S) segment classifiers. (A)** Maximum likelihood (ML) phylogenetic trees inferred from the representative sequences for all lineages within the **(A)** 51 sequences of the glycoprotein (490 bp) gene aligned with MAFFT and ML tree inferred under the GTR+Γ+G substitution model, **(B)** 47 sequences of the Small (S) segment (1690 bp), **(C)** 47 sequences of the Medium (M) segment (3885 bp) and **(D)** 47 sequences of the Large (L) segment (6404 bp). All the trees show similar topology for all the lineages.

2.3.3 Evaluating of lineage assignment using the Glycoprotein gene (Gn) as classifier

To determine the accuracy of the tool, we applied the nomenclature implemented by Grobbelaar and colleagues (Grobbelaar et al., 2011). A total of 129 partial sequences spanning the glycoprotein gene were used, of which 51 were determined to be the representative dataset used as lineage defining sequences (**Table 2.4**). All the 129 sequences distributed in lineages A($n = 13$), B($n = 1$), C($n = 44$), D($n - 1$), E($n = 7$), F($n = 1$), I($n = 2$), J($n = 1$), M($n = 2$), N($n = 13$) and O ($n = 2$) were correctly classified at phylogenetic level with accuracy, sensitivity and specificity of 100%. Near perfect classification of 99% at the phylogenetic level for lineages G($n = 8$) and H($n = 13$) were obtained. Only a single sequence HM587100 could not be assigned as per it's original lineage assignment. Upon the exclusion of this particular sequence, the typing tool assigned all the remaining 128 sequences with 100% sensitivity and accuracy. Representative sequences belonging to lineage G (HM587087, HM587083, AF134499, DQ380218) and J (DQ380222) were correctly assigned but with low bootstrap support values below the set threshold of 70 (**Table 2.5**). The impact of common mutations shared across multiple lineages) was also notable in the lineage assignment process as illustrated in lineages G and H (**Table 2.4**). The presence of shared SNPs in these lineages at positions 926 ($G \rightarrow A$), 1103($T \rightarrow C$) and 1163($C \rightarrow T$), reduced the sensitivity of the assignment process due to low support values.

TABLE 2.5: **Validation/testing of the RVFV Typing tool** to classify partial and whole genome sequences ($n = 128$). The classification results were compared to manual phylogenetic analysis. Abbreviations as used in this table: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives, TPR = True Positive Rate, FPR = False Positive Rate, ACC=Accuracy.

Lineage	Known (no.of strains)	TP	TN	FP	FN	TPR	FPR	ACC
A	13	13	115	0	0	100	0.0	100
B	1	1	127	0	0	100	0.0	100
C	44	44	84	0	0	100	0.0	100
D	1	1	127	0	0	100	0.0	100
E	7	7	121	0	0	100	0.0	100
F	1	1	127	0	0	100	0.0	100
G	8	8	120	1	0	100	0.0	100
H	12	12	116	0	1	100	0.0	100
I	2	2	126	0	0	100	0.0	100
J	1	1	127	0	0	100	0.0	100
K	11	11	118	0	0	100	0.0	100
L	10	10	118	0	0	100	0.0	100
M	2	2	126	0	0	100	0.0	100
N	13	13	115	0	0	100	0.0	100
O	2	2	126	0	0	100	0.0	100

2.3.4 Evaluating lineage assignment using whole RVFV genome sequences as classifiers

Assessment of the performance of the typing tool was further undertaken on the 234 sequences from which representative sequences were extracted for whole genome sequences (L, M and S-segments). Most of these sequences were correctly assigned with 100% accuracy. However, a few sequences, most of which had only a single sequence per lineage were assigned at 99% accuracy (**Table 2.6**, **Table 2.7** and **Table 2.8**).

TABLE 2.6: **Validation/testing of the RVFV Typing tool** to classify whole genome sequences ($n = 234$) using complete L segment sequences. The classification results were compared to manual phylogenetic analysis. Abbreviations as used in this table: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives, TPR = True Positive Rate, FPR = False Positive Rate, ACC=Accuracy.

Lineage	Known (no.of strains)	TP	TN	FP	FN	TPR	FPR	ACC
A	10	11	223	1	0	100	0.45	99.57
B	1	1	223	0	0	100	0.0	100
C	88	93	141	5	0	100	3.42	97.91
D	1	0	234	0	1	0.0	0.0	99.57
E	3	3	231	0	0	100	0.0	100
F	2	2	233	0	1	50	0.0	99.57
G	2	2	232	0	0	100	0.83	99.23
H	105	99	135	0	6	94.29	0.0	97.5
I	2	3	231	1	0	100	0.43	99.57
J	1	0	234	0	1	0.0	0.0	99.57
K	4	6	228	2	0	100	0.87	99.15
L	8	10	224	2	0	100	0.88	99.15
M	1	0	234	0	1	0.0	0.0	99.57
N	5	5	229	0	0	100	0.0	100
O	1	0	234	0	1	0.0	0.0	99.57

Lineage defining representative sequences in the S segment identified 31 SNPs that were common to lineages G and H and 33 SNPs common to lineage O and L (Appendix A, **Table A.1**). Representative sequences in the complete M segment showed that 33 SNPs are common to lineages C and D, 13 SNPs common to lineage H and F, 77 SNPs common to lineage I and J, 87 SNPs common to lineage K and M, and 83 SNPs common to lineage O and I (Appendix A, **Table A.2**). Complete L segment showed that there were 39 SNPs common to lineages H and C, 126 SNPs common to lineages J and I and 115 SNPs common to lineages O and L (Appendix A, **Table A.3**).

TABLE 2.7: **Validation/testing of the RVFV Typing tool** to classify whole genome sequences ($n = 234$) using complete M segment sequences. The classification results were compared to manual phylogenetic analysis. Abbreviations as used in this table: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives, TPR = True Positive Rate, FPR = False Positive Rate, ACC=Accuracy.

Lineage	Known (no.of strains)	TP	TN	FP	FN	TPR	FPR	ACC
A	10	12	222	2	0	100	0.89	99.15
B	1	0	234	0	1	0.0	0.0	99.57
C	88	89	145	1	0	100	0.68	99.57
D	1	0	234	0	1	0.0	0.0	99.57
E	3	3	231	0	0	100	0.0	100
F	2	4	230	2	0	100	0.86	99.15
G	2	2	232	0	0	100	0.0	100
H	105	102	132	0	3	97.14	0.0	98.73
I	2	4	234	2	0	100	0.86	99.15
J	1	0	234	0	1	0.0	0.0	99.57
K	4	5	229	1	0	100	0.43	99.57
L	8	8	226	2	0	100	0.0	100
M	1	0	234	0	1	0.0	0.0	99.57
N	5	5	229	0	0	100	0.0	100
O	1	0	234	0	1	0.0	0.0	99.57

TABLE 2.8: **Validation/testing of the RVFV Typing tool** to classify whole genome sequences ($n = 234$) using complete S segment sequences. The classification results were compared to manual phylogenetic analysis. Abbreviations as used in this table: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives, TPR = True Positive Rate, FPR = False Positive Rate, ACC=Accuracy.

Lineage	Known (no.of strains)	TP	TN	FP	FN	TPR	FPR	ACC
A	10	11	223	1	0	100	0.45	99.57
B	1	1	233	0	0	100	0.0	100
C	88	88	146	0	0	100	0.68	100
D	1	1	233	0	0	100	0.0	100
E	3	5	229	2	0	100	0.87	99.15
F	2	2	232	0	0	100	0.0	100
G	2	0	234	0	2	0.0	0.0	99.15
H	105	103	131	0	2	98.1	0.0	99.15
I	2	2	232	0	0	100	0.0	100
J	1	1	233	0	0	100	0.0	100
K	4	5	229	1	0	100	0.43	99.57
L	8	7	227	0	1	87.5	0.0	99.57
M	1	1	233	0	0	100	0.0	100
N	5	5	229	0	0	100	0.0	100
O	1	2	232	1	0	100	0.43	99.57

2.3.5 Lineage assignment of RVF outbreak strains in Kenya

The RVFV typing tool was further tested on RVFV sequences generated from clinical livestock samples from a RVF outbreak in Kenya. Using IgM capture ELISA method (as described in the methods), five Samples were positive indicating a recent infection with RVFV. These samples also displayed low cycle threshold (Ct) values, ranging from 14 to 19 on RT-qPCR, indicating sufficient viral load for whole genome sequencing (**Figure 2.5**). These five samples produced whole genome sequences with coverage of $> 99\%$. The genomic segment segments were subjected to lineage assignment using the glycoprotein gene and whole genome sequence (S, M and L complete segments) classifiers. Both the glycoprotein and whole genome classifiers assigned all the 5 sequences to lineage C.
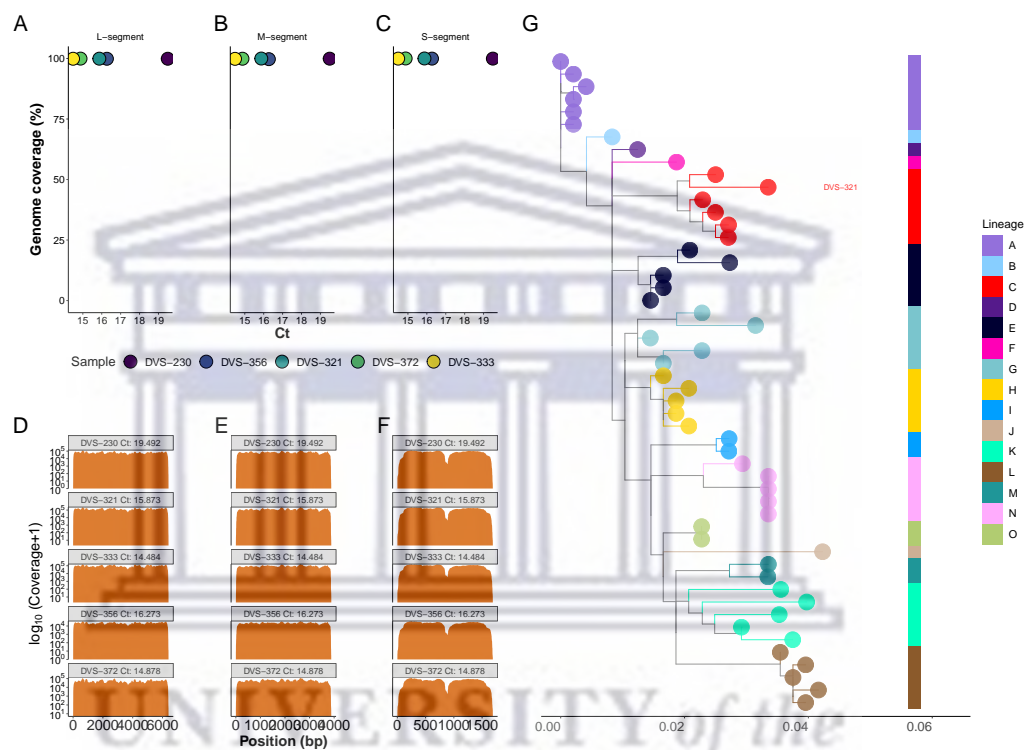
FIGURE 2.5: **Genome sequencing and phylogenetic analysis (A-C)** RT-qPCR cycle threshold (Ct) values ($x - axis$) plotted against percent genome covered ($y - axis$) for L, M and S-segments of the RVFV genome. **(D-F)**, Genome coverage ($log10$ transformed) along genomic positions in base pairs. **(G)** Maximum likelihood phylogenetic tree indicating the different clades corresponding to the fifteen major lineages and showing where a query sequence **(DVS-321)** is clustered in the tree.

TABLE 2.9: **RVFV Typing tool lineage assignment analysis on RVF outbreak samples**. The following terminologies are used: Query, sequence identifier/header in the FASTA file; Lineage, assigned/identified lineage of the query sequence; Bootstrap, ultrafast bootstrap approximation support value; Length, length of the nucleotide sequence; Year_first; Year when the lineage was first reported; Year_last: Year when the lineage was last reported, Countries: Countries where the identified lineage have also been reported.

Lineage	Bootstrap	Length	Year_first	Year_last		Countries
DVS-372	C	98	3885	1976	2016	South Africa; Mauritania; Zimbabwe; Uganda; Somalia; Angola; Madagascar; Sudan; Saudi Arabia; Kenya
DVS-333	C	97	3885	1976	2016	South Africa; Mauritania; Zimbabwe; Uganda; Somalia; Angola; Madagascar; Sudan; Saudi Arabia; Kenya
DVS-356	C	91	3885	1976	2016	South Africa; Mauritania; Zimbabwe; Uganda; Somalia; Angola; Madagascar; Sudan; Saudi Arabia; Kenya
DVS-321	C	93	3885	1976	2016	South Africa; Mauritania; Zimbabwe; Uganda; Somalia; Angola; Madagascar; Sudan; Saudi Arabia; Kenya
DVS-230	C	96	3885	1976	2016	South Africa; Mauritania; Zimbabwe; Uganda; Somalia; Angola; Madagascar; Sudan; Saudi Arabia; Kenya

2.4 Discussion

RVFV has been shown to have low genomic diversity at both the nucleotide and amino acid levels (Bird, Khristova, et al., 2007; Grobbelaar et al., 2011). However, the M segment of the virus has been reported to be slightly more diverse at 5% and 2% at the nucleotide and amino acid levels respectively, compared to the L and S segments which have compositional differences of 4% and 1% respectively (Bird, Khristova, et al., 2007; Grobbelaar et al., 2011). The observed limited diversity in the virus suggests that it may have a low tolerance for mutation (Weaver, 2006). This limited diversity has been captured by the proposed RVFV lineage assignment tool which delineated the clades based on SNPs. The observation of identical or common mutations across multiple lineages also highlights the low mutation rate within the RVFV genome (Bird, Khristova, et al., 2007). Delineating lineages using SNPs highlighted the impact of shared/common mutations in the lineage assignment process. The presence of shared mutations reduces the sensitivity of lineage assignment. Shared mutations between two or more lineages makes it difficult to identify definitive mutations that can be confidently used in lineage assignment.

Based on lineage classification proposed by Grobbelaar (Grobbelaar et al., 2011), lineage assignment using RVFV whole genome sequences for L, M and S segment was relatively less optimal with specificity values ranging between 0.4 and 3.4%, sensitivity ranging from 50-98% and accuracy from 99-100% (**Tables 2.6**, **2.7**, **2.8**) respectively. The less optimal assignment of lineages observed in using whole genome sequences can be, to an extent attributed to the presence of common SNPs among different lineages (**Table A.1**, **Table A.2**, **Table A.3**). In their analysis, Bird and colleagues (Bird, Githinji, et al., 2008) produced only 7 lineages using complete segments sequences, however, Grobbelaar et. al., (Grobbelaar et al., 2011) later generated 15 lineages using the partial glycoprotein sequences, and included a South African strain of 2010 classified as lineage H. In evaluating the proposed tool, we have compared its performance with the latest classification brought forth by Grobbelaar which produced the highest number (n = 15) of lineages. Although whole genome sequences are expected to produce a finer resolution in lineage assignment, we observed low sensitivity for lineages B, D, J, M and O across the three genomic segments of RVFV. This could be due to the few numbers of whole genome sequences that are available for these lineages. There were only single whole genome sequences for these lineages. The limited number of sequences belonging to these lineages made it difficult to identify unique lineage defining SNPs with strong statistical power. However, despite the limited number of sequences for these lineages, lineage assignment using glycoprotein representative sequences produced accurate and optimal assignment for all the 234 sequences with respect to 15 lineages nomenclature.

Furthermore, despite low accuracy in the assignment of sequences belonging to lineage G and J using whole genome sequences, the glycoprotein lineage assignment classifier correctly assigned G((HM587087, HM587083, AF134499, DQ380218) and J(DQ380222)) but wil low support values, ranging from 61 and 64 . Although the glycoprotein classifier does not achieve 100% accuracy in these two lineages, a robust bootstrapping (with more replications) can be undertaken to ensure that the desired bootstrap threshold value is obtained.

Testing the complete M segment sequences using the glycoprotein classifier, lineage assignment achieved 100% sensitivity, specificity and accuracy. However, sequence accessions DQ380215, DQ380216 (lineage G) and DQ380222 (lineage J) were correctly assigned but with low support values of 67, 68 and 35 respectively. Overall, using the glycoprotein classifier with the complete M segment sequences as input, resulted into correct assignment for all the sequences. The glycoprotein gene classifier was able to achieve 100% lineage assignment with respect yo the 15 lineage classification proposed by Grobbelaar (Grobbelaar et al., 2011), with no false positives identified.

The maximum likelihood phylogenetic reference trees generated in the analyses were able to resolve the 15 lineages with support values of > 70. The trees generated using the partial glycoprotein gene and whole genome sequences had similar topologies. This congruence is indicative of low occurrence of reassortment of the Rift Valley fever virus (Bird, Khristova, et al., 2007). From the phylogenetic inference, there is no clear pattern of lineages occurence in specific countries in Africa where the RVF is endemic. This may indicate widespread transmission and dispersal of the virus across the African continent. Most countries that have experienced RVF outbreak(s) have reported more than a single circulating lineage (Pepin et al., 2010) However, since it was first reported in 1976, lineage C continues to be the most predominant lineage in Africa (Grobbelaar et al., 2011).

The RVFV typing tool was also tested using whole RVFV genome sequences that were generated from clinical livestock samples of an outbreak in Kenya (**Figure A.1**). The sequence data was generated using Illumna technologies. These technologies have been widely used to carry out genomic epidemiology of pathogens at varying scales of outbreaks (Ladner et al., 2019). Lineage analysis of the outbreak samples using both glycoprotein (Gn) gene and complete M segment classifiers showed that the strains belong to lineage C. The assignment of lineages was supported by maximum likelihood phylogenetic analysis that produced a monophyletic clustering for all the five samples with high support (> 90) support values, and these sequences clustered with other strains belonging to lineage C.

2.5 Conclusion

We have developed an RVFV typing tool with both command-line and user-friendly web-based interface usability. The tool presented here allows for fast and accurate classification of RVFV species and lineage assignment within a few minutes. Lineages can be confidently assigned using whole genome (L, M and S segment) and partial glycoprotein gene sequences. Based on the 15 lineages proposed by Grobbelaar (Grobbelaar et al., 2011), the glycoprotein gene classifier showed consistency in lineage assignment of the partial and whole genome sequence of the M segment. In resource limited settings where whole genome sequences may not be readily generated, partial sequences of the M segment, particularly the 490 bp glycoprotein gene, can be used for typing. In addition, the glycoprotein classifier can still accurately assign lineages to samples where full length genome segments are provided as input. Although the evaluation of the tool applied the 15 lineages currently in use, further analysis using whole genome sequences as classifiers should provide a finer and higher resolution on lineage assignment, thereby providing comprehensive information that may include possible genetic reassortments. In conclusion, the RVFV Typing tool presented here allows for fast and accurate classification of RVFV species and lineages. Species can be classified using several sequencing products including whole genome sequences/consensus sequences and partial sequences. Lineages can be confidently assigned using complete segment sequences (L, M and S) or the partial glycoprotein Gn sequences. The method accommodates the need for consistency and accuracy in classifying RVFV sequences. This is useful in tracing the origin of outbreaks and supporting surveillance efforts. The study generated 5 livestock genomes, and these have been added to the existing pool of publicly available Rift Valley fever virus whole genome sequences. These sequences will enrich genetic characterization of the virus by allowing phylogenetic inference between known and newly sequenced strains.

Chapter 3

Multiplex PCR method for sequencing Rift Valley fever virus

Abstract

Rift Valley fever (RVF) is a febrile vector-borne disease endemic in Africa and continues to spread in new territories. It is a climate sensitive disease mostly triggered by abnormal rainfall patterns. The disease is associated with high mortality and morbidity in both humans and livestock. RVF is caused by the Rift Valley fever virus (RVFV) belonging to the *Phlebovirus* genus in *Phenuiviridae* family. It is a tripartite RNA virus with three genomic segments: small (S), medium (M) and large (L). Pathogen genomic sequencing is becoming a routine procedure and a powerful tool for understanding the evolutionary dynamics of infectious organisms including viruses. Inspired by the utility of amplicon-based sequencing demonstrated in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), Ebola, Zika and West Nile viruses, we report an RVFV sample preparation approach based on amplicon multiplex polymerase chain reaction (amPCR) for template enrichment and reduction of background host contamination. The technology can be implemented rapidly to characterize and genotype RVFV during outbreaks in a near real-time manner. To achieve this, we designed 74 multiplex primer sets covering the entire RVFV genome to specifically amplify the nucleic acid of RVFV in clinical samples from animal tissue. Using this approach, we demonstrated achieving complete RVFV genome coverage even from samples containing low viral load. We report the first

primer scheme approach of generating multiplex primer sets for a tripartite virus which can be replicated for other segmented viruses.

3.1 Introduction

A powerful tool which has become useful in studying emerging and re-emerging infectious diseases is genomic sequencing (Armstrong et al., 2019; Gwinn et al., 2019). By sequencing pathogens, we can answer different diagnostic questions such as the genetic relationship of viruses and the detection of mutations in viral genomes, which potentially lead to increased virulence, resistance towards antivirals, vaccine failure or immune escape (Deurenberg et al., 2017; Houldcroft et al., 2017). Genome sequencing of pathogens can be performed directly on clinical samples in an unbiased way (Gu et al., 2019) through a metagenomics approach or following enrichment methods that include cell culture (Deurenberg et al., 2017; Quick et al., 2016). In circumstances where the etiological agent of an outbreak is unknown, metagenomics is the option of choice. Metagenomics is therefore well suited when dealing with potentially lethal infections that fail the conventional diagnostic procedures such as immunoenzyme and immunofluorescence methods. However, sequencing an entire sample through metagenomics is less sensitive, generating insufficient pathogen reads due to the abundance of host genome contamination and thereby yielding fragmented and incomplete genome sequences (Deurenberg et al., 2017; Gu et al., 2019; Schlaberg et al., 2017). Genome sequencing directly from clinical samples without isolation is difficult for viruses such as RVFV whose viremia levels drop eight days post-infection (Laughlin et al., 1979; Pepin et al., 2010).

In trying to sequence Zika virus using a metagenomics approach, Quick and colleagues, were unable to recover sufficient sequence reads of the virus even after depleting human ribosomal RNA (rRNA) due to the low levels of viremia (< 1000 copies/µL of RNA) (Quick et al., 2017). In many cases, measuring virus diversity with high accuracy using deep sequencing remains a significant challenge. Factors such as virus titer, sample preparation, sequencing errors and computational inferences can bias measures of genetic diversity (Illingworth et al., 2017; McCrone & Lauring, 2016; Schirmer et al., 2015). Detection of pathogens through metagenomics is made more complex by specificity issues that arise from misclassification or contamination, nucleic acid stability, evolving bioinformatics workflows and excessive costs involved in data generation and analysis (Gardy & Loman, 2018; Schlaberg et al., 2017). Therefore, a timely response in terms of sample collection and processing is extremely critical to enhance recovery of virus genomes, especially for RNA viruses.

To undertake real-time genomic surveillance, there is a need to rapidly sequence the viral material, directly from clinical samples without cell culture enrichment which is laborious and time consuming. To generate complete viral genomes from clinical samples economically, targeted enrichment is needed (Houldcroft et al., 2017; Quick et al., 2017). Targeted enrichment can be achieved indirectly via host nucleic acid depletion or directly using oligonucleotide bait probes targeting the virus of interest (Houldcroft et al., 2017). Sequencing of complete genomic segments of RVFV has mostly relied on isolation of the virus by culturing on VeroE6 (African green monkey) cell lines (Ikegami et al., 2007; M. R. Smith et al., 2019). The manipulation of RVFV cultures is conducted in high containment biosafety laboratories and requires highly skilled personnel. Moreover, the cell culture process is often time consuming (takes up to 14 days) and laborious (Houldcroft et al., 2017). Viral passaging can also introduce mutations that were not present in the original clinical sample therefore leading to incorrect variant determination (Carrillo et al., 2007).

An alternative approach to cell culture is enrichment using the amplicon multiplex polymerase chain reaction (amPCR) method in a single assay. Viral genetic material amplification is achieved using primers complementing a known nucleotide sequence. This targeted sequencing approach has been successfully used in enriching viral genomes such as Ebola (Quick et al., 2016), Zika (Faria et al., 2017; Grubaugh et al., 2017; Quick et al., 2017) and currently SARS-CoV-2 (Quick, 2020) viruses. Heterogeneous and segmented RNA viruses may necessitate the use of multiple overlapping sets of primers to ensure the amplification of all genotypes and segments. Although amplicon sequencing can also generate incomplete coding sequences of viral genomes when the viral load in a sample is low, such sequences can still be used for genotyping and other evolutionary analyses. Amplicon multiplex PCR-based sequencing offers many advantages including the following: (i) high specificity, as most sequence reads will be of pathogen origin and not the host; this significantly reduces the sequencing cost; (ii) high sensitivity, with good coverage even at low pathogen load; and (iii) simplicity in terms of the design and application of new sets of primers for novel strains (Houldcroft et al., 2017). Therefore, amplicon enrichment methods are cheap, readily available and a fast option in comparison to isolation by cell culture, which is time-consuming, expensive, and laborious. To generate RVFV genomes in a cost-effective and timely manner, we have devised an amplicon primer scheme, the very first for a segmented genome, and showed that high quality genomes with high accuracy and good coverage are achievable. Given the advantages of amplicon multiplex PCR-based sequencing, we aimed to reconstruct RVFV genomes with higher coverage and depth, which are crucial for genotyping and evolutionary inference.

3.2 Materials and methods

3.2.1 Samples dataset

For benchmarking the amplicon multiplex PCR-based sequencing approach, we subjected five livestock (bovine) samples to three different treatments namely amplicon multiplex PCR enrichment (amPCRe), cell culture enrichment (CCE) and non-enrichment (Direct) representing the different enrichment and non-enrichment processes undertaken as shown in **Table 3.1**.

In summary, samples were subjected to three categories of treatment including amplicon multiplex PCR (n = 5), cell culture (n = 5) and non-enrichment (n = 5). In the sample set, a single sample, (RU1), was collected during an outbreak event in Rwanda while the rest were all from outbreaks in Kenya. The sample collection years ranged between 2018 and 2021. Clinical outbreak samples from Kenya were linked to a record that collated both the epidemiological and clinical data including the date of sample collection, date of onset of symptoms, geographical locations and demographic characteristics. The affected animals displayed symptoms such as lack of appetite, dyspnea, extreme fever, ecchymosis, conjunctiva, reddish vulva, diarrhea, jaundice, congestion of udder, nasal discharge, and abortion, all typical of RVF infection. Additional information regarding the samples can be found in **Table B.4**.

TABLE 3.1: **Samples used in the amplicon development protocol were subjected to three treatments**. Amplicon multiplex PCR enrichment (amPCRe), cell culture enrichment (CCE) and non-enrichment (Direct). The samples were archived and clinical outbreak samples from bovine species as host for Rift Valley fever virus.

SampleID	Treatment(s)	Host species	Sample type	Country	Location	Collection date
DVS-230	amPCRe,CCE,Direct	Bovine	Serum	Kenya	Kiambu	2021
DVS-356	amPCRe,CCE,Direct	Bovine	Serum	Kenya	Kiambu	2021
DK-B2	amPCRe,CCE,Direct	Bovine	Serum	Kenya	Murang'a	2021
RU1	amPCRe,CCE,Direct	Bovine	Serum	Rwanda	Rulindo	2021
08HAB	amPCRe,CCE,Direct	Bovine	Serum	Kenya	Wajir	2018

3.2.2 IgM antibody capture ELISA

All the RVF suspected livestock samples collected in Kenya in 2021 were first screened using enzyme linked immuno-sorbent assay (ELISA) for the detection of anti-nucleoprotein IgM antibodies in sera. This was conducted using the ID Screen RVF IgM Capture MAC ELISA as per the manufacturer's instructions (IDVet Innovative Diagnostic, Grabels, France) as previously described in chapter 2.

3.2.3 Virus enrichment in culture

Virus enrichment was performed in a containment BSL3 facility as described in chapter 2.

3.2.4 Nucleic acid extraction and RT-qPCR in non-cell culture samples

For the Direct and amPCRe samples, viral RNA was extracted from 140µL IgM positive samples using the Nucleospin RNA Virus Minikit for viral RNA (MACHEREY-NAGEL, Ref 740956, Dueren, Germany), according to the manufacturer's guidelines. Viral RNA was isolated from 140µL of the filtered cell culture (CCE samples) supernatant using a QIAmp Viral RNA kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions.

3.2.5 Diagnosis using RT-qPCR

Reverse transcription quantitative real-time PCR (RT-qPCR) was performed on the RNA samples. The one-step assay consisted of 2µL (5-10ng) of the RNA template in a reaction of 15µL using a final concentration of 0.3µmol for primers and 0.1µmol for the probe in a PCR system (Applied Biosystems, **Table 3.2**). The reaction was carried was carried out in a series of incubation steps as follows: 50℃ for 10 minutes, 95℃ for 2 minutes, 95℃ for 3 seconds, 60℃ for 30 seconds for 40 cycles. This assay uses a highly conserved domain located on the L segment of the virus for RVFV detection (using the 5' Fam reporter dye and 3' BHQ1 quencher dye) (Bird, Bawiec, et al., 2007), as shown in **Table 3.2**.

TABLE 3.2: **Primers and probe set used for RT-qPCR assay**

RVFV segment	Primer name	Sequence
L	RVFL-2912fwdGG	(5'-TGAAAATTCCTGAGACACATGG-3')
L	RVFL-2981revAC	(5'-ACTTCCTTGCATCATCTGATG-3')
L	RVFL-probe-2950	(5'-CAATGTAAGGGGCCTGTGTGGACTTGTG-3')

3.2.6 Designing tiling primers

RVFV primer panels for multiplex PCR were designed using primal scheme (Quick et al., 2017) for each segment of the virus. The primal scheme tool uses Primer3 software (Untergasser et al., 2012) and applies a greedy algorithm to find primers for tiling amplicon generation using multiple reference genomes. For each segment, a set of 104 complete genomic RVFV sequences was obtained from the NCBI RefSeq database (O'Leary et

al., 2016). The complete genome sequences for each segment were separately concatenated into a single FASTA file. The ZH548 strain was used as the coordinate system for the primers. Multiple sequence alignment on the DNA sequences was performed using Clustal Omega (Sievers et al., 2011). Sequences with more than 5% divergence and 99-100% identity to other genomes were removed resulting in a final set of 16, 15 and 19 for the L, M and S segments respectively. The primal scheme tool was executed to generate amplicon primers with an amplicon size of 400bp (**Figure 3.1**). 74 primers in two pools were generated. In-silico prediction of the coverage for whole genome sequences was reported to be 97.31%, 98.84% and 96.80% for the L, M and S segments respectively. The primers were synthesized by Macrogen Europe (Amsterdam, Netherlands) and shipped to ILRI Nairobi, Kenya, followed by reconstitution in tris-ethylenediaminetetraacetic Acid (EDTA) buffer. Detailed information providing the primer sequences and locations can be found in the appendix (Appendix B, **Table B.1**, **Table B.2**, **Table B.3** and **Table B.5**).
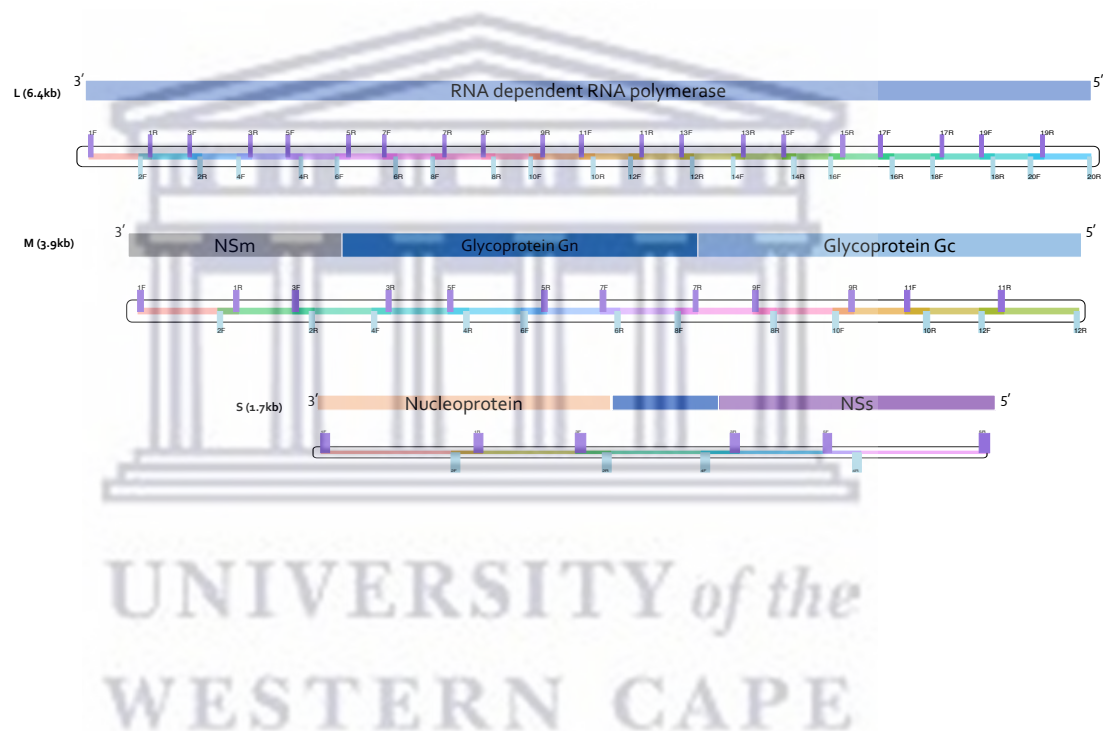
FIGURE 3.1: **Schematic representation of RVFV primer schemes.** The three genomic segments for RVFV and the primers mapping positions are shown. Forward primers have the suffix 'F' and are colored purple, while reverse primers have the suffix 'R' and are colored in light blue. For each segment, primers were designed with Primal Scheme (Quick et al., 2017) for the generation of amplicons with a target size of 400 bp in the genome. In total, there were 74 primers (38 for pool 1 and 36 for pool 2 reactions). There were 20, 12 and 6 primer pairs for the L, M and S segments respectively. The minimum amplicon size for the primers was 355 and the maximum size was 374.

3.2.7 cDNA synthesis

Extracted RNA was converted to cDNA using the LunaScript RT supermix kit (New England Biolabs, Hitchin, UK) in a reaction of 10µL according to the manufacturer's instructions.

3.2.8 Amplicon Multiplex PCR

The multiplex PCR was set up in two separate reactions, each having one of the two primer pools as indicated in **Table 3.3**. The reaction conditions were marked by low primer concentrations, long annealing times and high primer annealing temperatures. This enabled amplification of targets that cover the whole genome in the two reactions (**Figure 3.1**). Briefly, lyophilized primers were reconstituted in tris-ethylenediaminetetraacetic acid (EDTA) (TE) buffer to attain concentrations of 100µmol. Two 1.5mL (about 0.05 oz) Eppendorf tubes were labeled as pool 1 and pool 2. An equal volume (100µmol stocks) of each odd numbered primer was added to pool 1, while all the even ones were added to pool 2. In total, there were 38 and 36 primers for pool 1 and pool 2 respectively. The primer pools were then diluted at a ratio of 1:10 with TE buffer to a working concentration of 10µmol. The volume of primer varied depending on the number of primers per pool.

TABLE 3.3: **Amplicon multiplex PCR components.**(*Initial concentration is 5×)

Component	Amount (µL)	Final concentration
Q5 HotStart Master Mix Buffer*	12.5	1×
Primer pool 1 or 2 (10µmol)	1.4 for pool 1 and pool 2	0.015µmol per primer
Nuclease Free Water	Up to 7	-
cDNA	4.5	-

3.2.9 Library preparation using NEBNext Ultra II DNA Library prep kit

Library preparation of the samples was performed using a NEBNext Ultra II DNA library prep kit (New England Biolabs, Ipswich, USA). PCR products from the amplicon reaction in separate primer pools were cleaned up using AMPure XP purification beads (Beckman Coulter, High Wycombe, UK). The clean products were quantified using fluorimetry with the Qubit dsDNA Broad Sensitivity assay on the Qubit 2.0 instrument (Thermo Fisher Scientific) . The corresponding PCR products from each primer pool were pooled together and quantified using Qubit for library preparation. End repair and adaptor ligation was performed using the NEBNext Ultra II DNA library prep kit for

Illumina (New England Biolabs, Hitchin, UK). Adaptor-ligated cleanup was performed using AMPure XP purification beads followed by library PCR using the NEBNext Ultra II Q5 master mix, TruSeq Index Primers (i7) and Universal PCR Primers (i5) in a 50µL reaction for 15 cycles. Equal volumes were pooled after which a clean-up with AMPure XP purification beads was carried out followed by quantification. Pooled libraries were denatured and diluted for loading on an Illumina sequencing instrument according to the manufacturer's instructions.

For CCE and Direct samples, 12µL of RNA (concentration of 5ng - 1µg) was added to a probe hybridization buffer (2µL) and NEBNext ribosomal RNA (rRNA) depletion buffer (3µL). The mixture was thoroughly mixed and incubated on a thermocycler with a denaturation step at 95℃ for 2 minutes followed by a ramp down step to 22℃ at a rate of 0.1℃ per second and a final hold at 22℃. The samples were then subjected to digestion with DNase I. Purification of the samples was performed using NEBNext RNA sample purification beads (2.2× the sample volume). The purified RNA samples were prepared for first strand synthesis by hybridizing to random primers (NEBNext) and incubated for 8 minutes at 94℃. First strand cDNA was synthesized using NEBNext first strand synthesis enzyme mix and placed on a thermocycler with the following conditions: 10 minutes at 25℃, 15 minutes at 42℃, 15 minutes at 70℃ and a final hold at 4℃. Second strand synthesis was immediately performed using the NEBNext second strand synthesis reaction buffer and enzyme mix. The samples were incubated for 1 hour at 16℃ with the lid temperature off. Double-stranded cDNA was purified using NEBNext sample purification beads. The clean cDNA samples were end repaired and adapter ligated using NEBNext adaptors diluted 200-fold with the dilution buffer. This was conducted since the concentrations of the RNA samples were low. The NEBNext USER enzyme was added to the ligation reaction mixture. The adaptor-ligated sample was cleaned using Ampure purification beads (0.9×). Enrichment of the adaptor ligated cDNA through PCR was performed using NEBNext multiplex primers (forward and reverse primers combined). The thermocycler conditions were as follows: initial denaturation at 98℃ for 30 seconds for 1 cycle, denaturation at 98℃ for 10 seconds and annealing at 65℃ for 15 cycles, final extension at 65℃ for 5 minutes for 1 cycle and a hold at 4℃. Cleanup of the PCR enriched libraries were performed using sample purification beads (0 .9×). The libraries were then quantified using Qubit dsDNA HS assay kit and normalized to obtain a concentration of 4nM. Pooled libraries were denatured and diluted for loading on an Illumina sequencing instrument following the manufacturer's instructions.

3.2.10 Generation of consensus genomes

To statistically compare the mapping metrics of the samples, we normalized each library as a relative proportion of the total library size. We obtained the reads for every sample

in 20 million and randomly subsampled the resulting number from the raw reads. The consensus genome for each barcoded RVFV sample was generated using an in-house rvfv-amplicon-seq (Juma et al., 2022) nextflow pipeline. Raw demultiplexed reads in FASTQ format were assessed on quality using FastQC v0.11.9 (Andrews, 2010). Reads that had low quality scores and adaptor sequences were trimmed using fastp (Chen et al., 2018). Alignment of the trimmed reads to the RVFV reference genome (ZH-548) was conducted using bwa-mem (Li & Durbin, 2009). Alignment statistics were computed with SAMtools (Li et al., 2009), and only samples having a minimum mapping threshold of 200 reads were utilized in subsequent downstream analysis. Alignments containing amplicon primers were trimmed using iVar (Grubaugh et al., 2019). Using each segment reference (ZH-548 strain) genome and the corresponding gene features file, variants were called on the pile-up generated from the filtered alignments. To generate consensus alleles, positions with $\geq 10 \times$ genome coverage and with ≥ 20 base quality were considered. Regions that failed the above criteria and those in primer binding regions were masked with N characters. Genome-wide, amplicon mean and amplicon per base coverages were computed using BEDTools (Quinlan & Hall, 2010). Variant effect prediction was performed using SnpEff (Cingolani, Platts, et al., 2012) and filtered one line per variant using vcfEffOnePerLine.pl that comes bundled with SnpSift (Cingolani, Patel, et al., 2012). Consensus genomes for CCE and direct samples were generated using viclara (Juma, 2022). In the viclara analysis pipeline, we called the variants using bcftools, while maintaining the minimum base quality at ≥ 25 and at $\geq 10\times$ coverage. SNP (single nucleotide polymorphisms) concordance analysis was performed on samples that had high genome coverage ($> 90\%$) using the SnpSift concordance method (Cingolani, Patel, et al., 2012).

3.2.11   Maximum likelihood estimation and molecular clock phylogeny reconstruction

We retrieved 216 complete genome sequences of the medium (M) segment and concatenated them to the 11 consensus genomes generated in this study. We selected the 11 genomes based on a threshold value of 80% coverage on consensus genomes. The sequences were deduplicated to remove those with similar base composition followed by multiple sequence alignment using MAFFT (Katoh & Standley, 2013) after the removal of primer binding sites. After this filtering process, there were 196 sequences used in downstream analyses (Appendix B, **(Table B.6)**). We identified generalized time reversible (GTR) (Tavaré, 1986) with gamma (Γ) substitution rate as the optimal evolutionary model using ModelTest-NG (Darriba et al., 2012). We applied this model to infer a maximum likelihood (ML) phylogenetic tree using IQ-TREE (Nguyen et al., 2015) with single branch support testing (Guindon et al., 2010) and 1000 replications. We also assigned lineages to the complete M segment genome sequences using the RVFV typing

tool (Juma et al., 2022). We assessed the temporal signal of the sequences by regressing the root-to-tip distance against sampling time in decimal years using TempEst (Rambaut et al., 2016). Molecular clock analysis was performed using TreeTime (Sagulenko et al., 2018) on the alignment using the collection years of the sequences as dates to generate a time-scaled phylogenetic tree (Appendix B, **Figure B.1**).

3.3 Results

3.3.1 Sequencing and consensus genomes

The inverse relationship between genome coverage and RT-qPCR cycle threshold (Ct) values was quite distinct, as expected in all the samples. Samples subjected to cell culture enrichment (CCE) yielded genomes with high coverage compared to amplicon multiplex PCR-enriched (amPCRe) and non-enriched (Direct) samples (**Figure 3.2**).
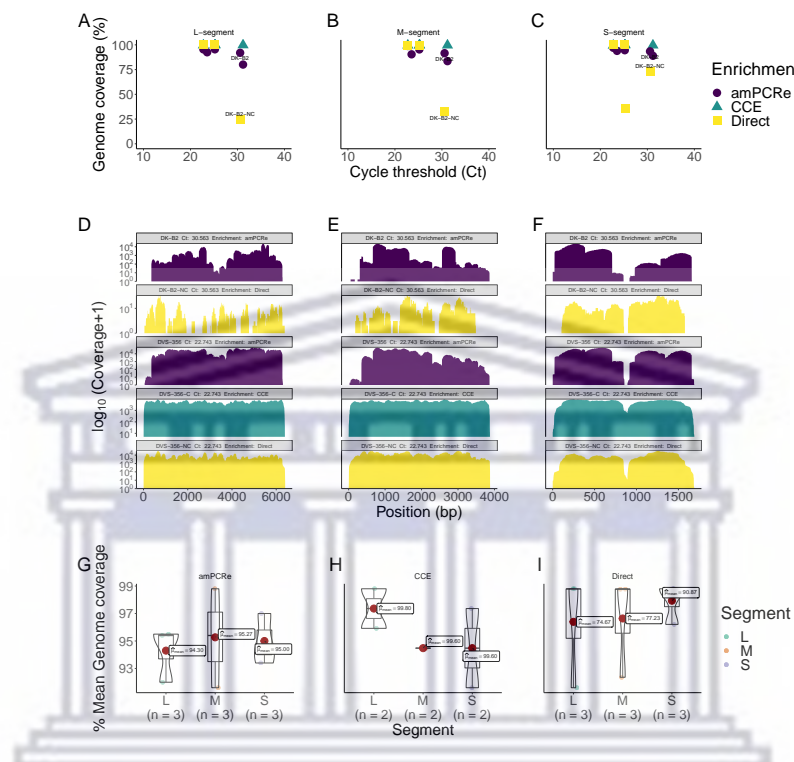
FIGURE 3.2: **Genome coverage and RT-qPCR of samples subjected to am-PCRe, CCE and direct sequencing** RVFV positive samples' cycle threshold (Ct) values ($x - axis$) versus genome coverage ($y - axis$) for the **(A)** large, **(B)** medium and **(C)** small segments. Samples plotted on the graph (DVS-230, DVS-356, DK-B2, RU1 and 08HAB). amPCRe samples are colored purple, CCE samples are colored green while Direct samples are shown as yellow. Lower Ct values indicate higher viral titers while higher Ct values depict low viral titers. CCE, direct and amPCRe samples are represented using triangle, square and circle shapes, respectively. Genome-wide density plots show coverage ($log10$ transformed) across the RVFV genome segments: **(D)**, a 6404 bp L segment **(E)**, a 3885 bp M segment and **(F)** a 1690bp S segment. The coverage density plots are colored as purple, yellow and green to represent the different treatments. There is a significant drop in coverage in the middle of the S segment, an area characterized by homopolymers of C bases. Boxplots and violin plots show the mean genome coverage of the consensus sequences for all the virus segments (L, M and S) in **(G)** amPCRe, **(H)** CCE and **(I)** Direct samples.

We sequenced and generated consensus RVFV genomes in 11 samples, of which four had been subjected to CCE, five to amPCRe and two were from Direct clinical outbreak and archived samples. Samples that were enriched by CCE produced near complete genomes and covered over 99% of the genome while Direct samples displayed genome coverage values ranging from 75% to 90% (**Figure 3.2**). Amplicon enriched (amPCRe) samples yielded genome coverage between 80% and 97%. Samples with low viral load as indicated by high Ct values do not produce enough genome coverage following sequencing due to less viral material in the starting sample.

For samples with Ct > 30, genome recovery from direct sequencing was low, ranging from 25% to 70%, and in one sample (08HAB), no genome was recovered in all three segments. In another sample (RU1), with a Ct value of 25.25, direct sequencing yielded only the S segment genome. Enrichment by cell culture significantly allowed for genome recovery in most samples that had Ct values of > 25. However, in one sample (DK-B2), genome recovery failed due to fewer reads mapping to the virus genome. Amplicon multiplex PCR enriched samples yielded genomes with > 80% coverage. Interestingly, samples with Ct > 30, produced near-complete genomes compared to non-enriched samples. In addition, samples with Ct < 25 showed genome recovery of over 90%. In amPCRe samples, there was a general trend of fewer reads mapping at the extreme ends of the genome segments (both 3' and 5' ends). We also observed a distinctive drop in coverage between positions 840 and 900 in the S segment in all treatments, with amPCRe samples showing a large drop in depth at these positions. The patterns of genome wide depth distribution in amPCRe samples were almost identical across the segments with distinctive drops in specific regions (**Figure 3.2**). The seropositivity rates in the samples ranged from 26 to 138, with 2 samples (DVS-230 and DVS-356) displaying strong positivity through IgM assay, indicating a recent infection (Appendix B, **Table B.4**).

Out of the five samples subjected to CCE, we recovered four (DVS-356, DVS-230, RU1 and 08HAB) RVFV genomes with a coverage of > 99%. Of the five amPCRe samples, we obtained five RVFV genomes with coverage ranging between 80% and 97% in all segments. From non-enriched (Direct) samples, we recovered only two (DVS-356, DVS-230) RVFV genomes with a coverage of > 99%. Two samples (DK-B2 and RU1) yielded fragmented partial genomes for one or all segments. In total, we recovered 11 complete and near complete (> 80% genome coverage) RVFV genomes. All the genomes recovered were from samples with Ct values less than 35. To compare genome coverage metrics, we selected three samples (DVS-356, DVS-230 and DK-B2) that yielded RVFV genomes in amPCRe and Direct treatment. For CCE, we excluded DK-B2. Overall, the genome coverage in CCE samples was higher (99%) compared to that in Direct and amPCRe samples. At the genome segment levels, there was no significant variation in genome

coverage among CCE samples. The CCE samples showed genome coverages in L (n=2, μ=99.8%, σ=0.14), M (n=2, μ=99.6%, σ=0.00), and S (n=2,μ=99.6%, σ=0.28). The observed mean genome coverage in direct samples were 74.7% (n=3, μ=43.7), 77.2% (n=3, μ=39.0), and 90.9% (n=3, σ=15.7) for L, M and S segments respectively. Among amPCRe samples, the genome coverage reported were 94.3%(n=3, μ=1.99), 95.3% (n=3, μ=3.60), and 95.0% (n=3, σ=1.83) for L, M and S segments respectively. Nonetheless, in all the three categories, there were insignificant differences in the mean genome coverage across the three genomic segments as reported by the p-values of 0.89, 1.0 and 0.79 for CCE, Direct and amPCRe respectively (**Figure 3.2**).

3.3.2 Performance of amplicon primers

We examined the performance of each amplicon primer by computing the mean amplicon coverage per segment using BEDTools (Quinlan & Hall, 2010). We observed that the first and last amplicon primers for every genomic segment of the virus reported coverage dropout. The 5' and 3' ends of the genome segments were insufficiently covered due to the minimal number of reads mapping in these regions. Uneven amplicon primer coverage of genomes was also observed with distinct patterns especially in samples with extremely high Ct values ($\geq$ 35). In both L and M segments, amplicon primers that showed significant coverage dropouts were mostly observed in samples with high ($\geq$ 35) Ct values. However, two primers in the S segment, S_2_RIGHT (spanning positions 709-739) and S_4_LEFT (spanning positions 957-979) reported a significant drop in coverage irrespective of a sample's Ct values. Zooming into this region, we identified a homopolymer track which is marked by repetitive stretches of nucleotide C's. Overall, samples with low Ct values were well covered by the amplicon primers compared to samples with high Ct values (**Figure 3.3**).
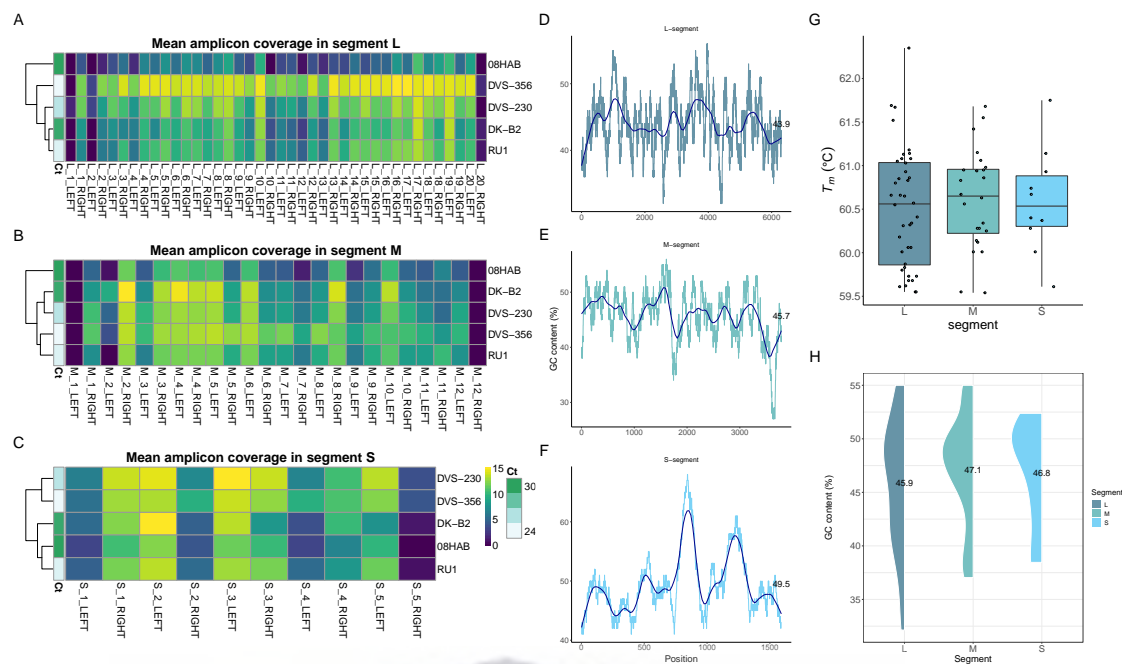
FIGURE 3.3: **Amplicon primer coverage for Rift Valley fever virus** Primal Scheme generated 74 sets of primers: small (n=10), medium (n=24) and large (n=40). Amplicon primers with odd numbers on their prefix belong to pool 1 while those with even numbers are pool 2 primers. Panel (A) shows heatmaps of the mean amplicon coverage across the L segment, (B) M segment and (C) S segment per sample. The first and last amplicon primers did not sufficiently amplify the 5' and 3' ends of the sequences thereby resulting in a drop in coverage. The dark purple color on the heatmaps show primers which resulted in minimal amplification while bright yellow color indicates sufficient coverage following amplification with both primer pools. Panels (D, E and F) show the percentage GC content computed in a sliding window of 100 bp along the reference genome (ZH-548 strain) for the L, M and S segments respectively. The solid blue line in the middle depicts the average GC%. Panel (G) shows the boxplots of the melting temperature and panel (H) indicates the distribution of percent GC content for all the amplicon primers per genomic segment.

3.3.3 Amplification accuracy assessed by SNP concordance analysis

To assess the accuracy of amPCRe against CCE and Direct treatments, we compared the SNPs identified in each treated sample. To identify the mutations within the RVF virus genome, we enumerated changes in the generated consensus genomes. SNP concordance analysis was limited to two samples (DVS-230 and DVS-356) that yielded high (> 94%) coverage genomes in all three treatments. We observed a higher proportion of synonymous to non-synonymous mutations in the RVF virus genome for all three treatments. From the consensus alleles, we identified shared synonymous mutations that were occurring within the RVFV genes in CCE, amPCRe and Direct samples. We found 146, 89 and 39 intragenic SNPs in the L, M and S segments respectively, that were in perfect concordance in all three categories of treatment.

We also observed 27 non-synonymous mutations classified as missense variants in all the genomic segments of RVFV. There was insignificant variation in the number of mutations observed in CCE, amPCRe and Direct samples. In the amPCRe samples, the number of synonymous and non-synonymous mutations reported was 337 and 37, respectively. In the CCE samples, 286 and 31 respective mutations were reported, while in Direct samples, we enumerated 287 synonymous and 29 non-synonymous mutations, respectively. Eleven non-synonymous mutations located in the RNA-dependent polymerase in the L segment were common in all three treatments. All four non-synonymous mutations (T715A, A1717G, A1911G and A1913G) observed in the CCE, amPCRe and Direct samples occurred within the glycoprotein (Gn) gene. We observed 11 non-synonymous mutations occurring in the S segment, of which 10 were found in the non-structural (NSs) gene and only one was observed in the nucleocapsid (NP) gene **(Figure B.2)**.

On evaluating the suitability of amplicon-enriched samples for genetic studies, SNP concordance showed high concordance between direct and CCE samples. Of all the SNPs called, we observed a concordance score of 100% in all three genomic segments of the virus genome. However, there was a slight reduction in the concordance scores between Direct and amplicon-enriched samples. We observed concordance scores of 98.42% (out of 191 SNPs), 99.31% (out of 147 SNPs) and 91.04% (out of 67 SNPs) in the L, M and S segments, respectively.
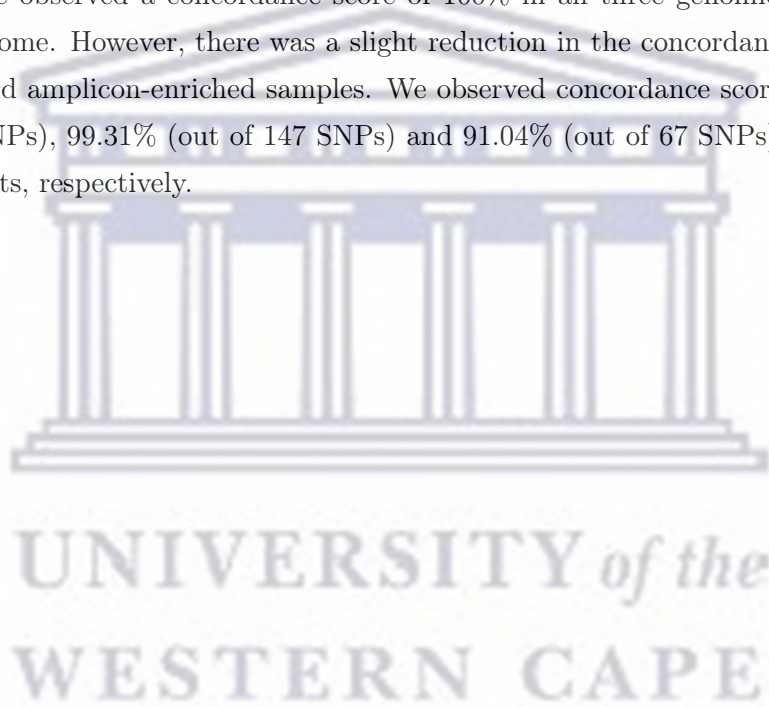
3.3.4 Similar lineage placement in CCE, amPCRe and direct genomes of RVFV

To infer the evolutionary dynamics of RVFV, we performed phylogenetic analysis to determine whether the consensus genomes generated by the three treatment approaches can be used for such inferences. We began by assigning lineages to the M segment genome sequences generated from the study with those obtained from the public database, NCBI GenBank . All the genome sequences (n = 11) generated from this study were confidently assigned to lineage C. In addition to the three (T715A, A1717G and A1913G) non-synonymous mutations found in the M segment, multiple sequence alignment indicated 43 synonymous mutations found in lineage C, that were also present in the CCE, amPCRe and Direct samples (**Figure 3.4**). This observation indicates the suitability of the data generated by amPCRe for evolutionary dynamic studies and inference (**Figure 3.4**).
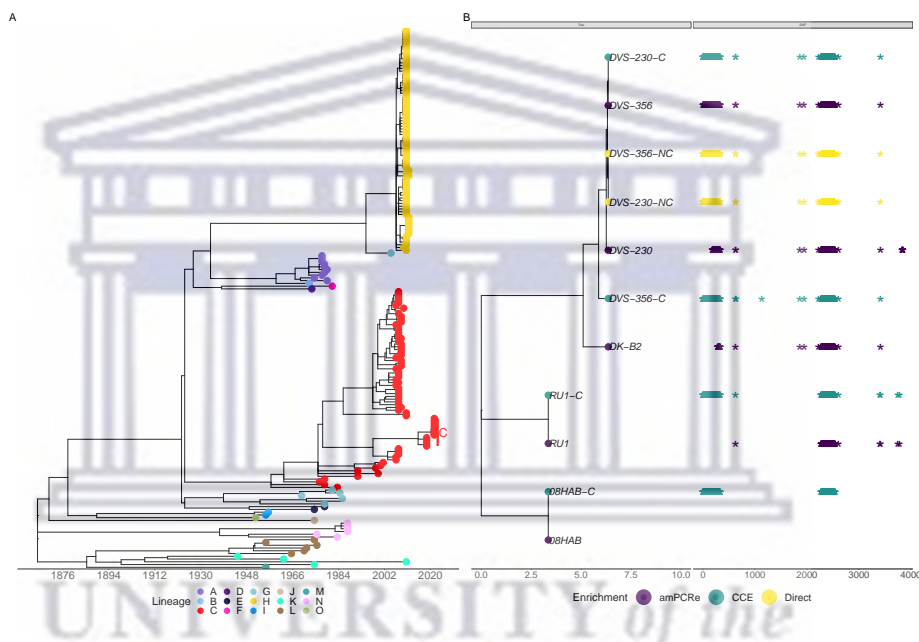
FIGURE 3.4: **Molecular clock phylogenetic reconstruction (A)** Time scaled phylogenetic tree of heterochronous sequences (n=199) belonging to the M segment showing the occurrence of lineages from 1951 to 2021. The tips of the phylogenetic tree are colored according to the lineages. The clade labelled C shows the placement of the 11 genomes from this study within the global context **(B)** sub-tree of the 11 genomes generated in the study indicating that all the genomes irrespective of the treatment option confidently places the sequences in the global tree, and assigned as lineage C.

3.4   Discussion

Genomic surveillance has become a useful tool for studying emerging and re-emerging infectious diseases (Armstrong et al., 2019; Gwinn et al., 2019). Using genomic data, control measures can be developed including diagnostic and vaccination reagents. Virus genome sequence data can be used to explore different areas such as the genetic relationship of viruses and mutations patterns that can potentially lead to increased virulence, resistance towards antivirals, vaccine failure or immune escape (Deurenberg et al., 2017; Houldcroft et al., 2017). The challenge, however, is obtaining pure virus genetic material from a clinical sample that is low in host contamination to generate whole viral genome sequences. Existing methods for enriching viral genetic material and reducing host contamination include culturing techniques that are expensive, laborious and time consuming. Here, we have developed an RVFV sample preparation method based on amplicon multiplex polymerase chain reaction enrichment (amPCRe) that allows for template enrichment and reduction of background host contamination. The technology can be implemented rapidly in a near real-time manner to characterize and genotype RVFV during outbreaks.

Assessment of the levels of antibodies against the nucleoprotein for all the samples used in the study were positive. The results were further corroborated by RT-qPCR which revealed the amount of target nucleic acid present in the samples, reported as Ct values. Ct values of less than 29 often indicate a strong positive reaction with high viral titers in the samples. A Ct value in the range of 30-37 shows positivity with low viral titers in the sample (Bird, Bawiec, et al., 2007). Finally, Ct values > 37 indicate a weak positive reaction with exceptionally low viral titers in the sample. In all the samples used in this study, we encountered all three categories of Ct values. Applying the three treatment options (CCE, amPCRe and Direct), we took note of the recovery success rate of RVFV genomes with respect to the original sample Ct values. We note that when the Ct values of samples are less than 25, enrichment of the samples with CCE or amPCRe does not result in any significant difference in genome recovery of the virus. However, when the Ct values of samples are over 25, our proposed amPCRe method produces an improvement in genome recovery as shown in **Figure 3.2**.

The amPCRe process which amplifies the target genome using overlapping primers can help recover genomes of these samples with a low viral load (Quick et al., 2017). Through amPCRe, the target genome is amplified to generate sufficient DNA for sequencing, resulting in improved whole genome sequence recovery. As shown through comparison by percent genome coverage, CCE remains superior in terms of genome coverage. CCE multiplies the virus in a growth medium, and exponentially increases its viral load to

sufficient levels for optimal genome sequence recovery. However, CCE is an expensive and slow process that cannot be relied upon for near real-time genomic surveillance in a pandemic situation. Furthermore, CCE is unpredictable and tends to occasionally fail to propagate the virus even when samples of high viral titers have been used (Carrillo et al., 2007; Dolskiy et al., 2020).

To capture sufficient RVFV material for genome assembly cost-effectively and in a timely manner, we employed an amPCRe approach before sequencing. This technology relies on nucleic acid amplification of the viral genome fragments in single multiplex PCR reactions. The amPCRe method generated enough reads that aligned to the RVFV reference genome . The use of an optimal annealing temperature (Ta) in a PCR reaction is critical to ensure specificity and efficiency (Frey et al., 2008). Determination of the optimal annealing temperature is often dictated by the PCR primers and the melting temperatures (Tm) of the products. Regions of DNA which have a high melting temperature and are rich in GC content are susceptible to the formation of secondary structures (Frey et al., 2008). Consequently, finding the appropriate melting temperature for primer-template annealing is critical. The average GC content of the S segment as indicated by the amplicon primers is slightly lower than that of the reference genome. Consequently, regions of lower GC content result in optimal amplification of the nucleic acid, thereby yielding sufficient material for sequencing (**Figure 3.3**). Amplicon sequence data are further normalized by trimming off the tiling primers (Grubaugh et al., 2019) which may affect mapping. The implication of the PCR parameters and the normalization of the sequence data can further explain the observed differences in read depth and coverage between the genome segments.

A hallmark of amplicon sequencing is coverage dropout observed at the ends of the genome segments (5' and 3' untranslated regions (UTRs)) compared to Direct or CCE treated samples. The coverage dropout corresponds to the overlapping nature of amplicons in the multiplex PCR method and is therefore expected (Quick et al., 2017). Such a dropout can also be explained by the fact that retroviruses and RNA viruses lack the 3'-5' exonuclease activity (Kwok et al., 1994). The coverage dropout in the 3' and 5' untranslated regions fall outside the outer primer binding sites and is therefore less likely to be amplified at the same rate as other parts of the genome (Kwok et al., 1994). The 3' and 5' ends of sequences can disproportionately affect the successful recombining and extension of DNA strands. As such, while designing primers for such regions, mismatches are severely penalized. Additionally, we also observed a consistent coverage dropout in the intergenic region (approximately 90 bp) of the S segment in both enriched and non-enriched samples. The primers S_2_RIGHT (spanning positions 709-739) and S_4_LEFT (spanning positions 957-979) which amplify the intergenic region (positions

845-915) in the S segment did not generate sufficient amplicons for sequencing hence the drop in coverage (**Figure 3.3**). These primers amplify the intergenic region separating the non-structural (NSs) and nucleocapsid (NP) genes. This region is regarded as an inaccessible part of the genome characterized by homopolymer stretches of C's that makes it difficult to amplify by PCR due to the repetitions of the same nucleotide (Schirmer et al., 2015).

Reconstructing viral genomes with high coverage and depth is crucial in phylogenetic inference and other genome-based genetic studies. Molecular clock evolutionary analysis revealed that the virus genomes generated by the three treatment options can be used to infer the genetic evolution of the virus (**Figure   3.4**). The amplicon sequencing approach developed in this study generated viral genomes with a mean genome coverage of > 80%, sufficient for genotyping purposes. The high concordance in SNPs called in amplicon enriched and direct or cultured samples shows the suitability of our method in evolutionary analysis including phylogenetics and phylodynamics. Among the SNPs identified with high concordance in the three treatment methods were lineage defining mutations as shown in a previous study by (Juma et al., 2022) (Juma et al., 2022). The lineage-defining SNPs identified in all the genomes generated by the three treatment options were G851A, C881T, G922A, G2005A and G4015A for the L segment, T715A, A1911G and A1913G in the M segment and T101A, C534T, T684C and A758G in the S segment (Grobbelaar et al., 2011; Juma et al., 2022).

3.5   Conclusion

Although primal scheme has been optimized for non-segmented virus genomes, here we showed that a multiplex PCR primer scheme can also be generated for segmented viruses. We have shown that the amplicon enrichment method for sequencing RVFV clinical samples with low viral titers as indicated by Ct values of less than 30 can be reliably used in generating near complete genomes of the virus. Where CCE fails in samples with moderate Ct values, amPCRe has the advantage of recovering the genomes at levels useful for genotyping. However, obtaining full length whole genomes in samples with high Ct values (> 30) remains a challenge. Importantly, the timing of sampling during an outbreak is key in ensuring adequate viral load for successful sequencing. Therefore, there is a need for continued development and implementation of the best procedures for handling samples without disrupting standard clinical workflows for wider adoption in genomic surveillance during outbreaks. Although amPCRe has proved to be successful in the recovery of genomes from isolates with Ct values < 30, further optimization is needed to reduce the coverage dropouts observed in specific regions of the genome. Genome recovery following amPCRe does not ordinarily generate 100%

coverage. However, sufficient sequence coverage is obtained suitable for genotyping and other genomic epidemiological studies such as transmission dynamics.

Chapter 4

Unravelling the transmission and phylogeographic distribution of RVFV using phylodynamic approaches

This chapter is in *Prep Genomic epidemiology and evolution of Rift Valley fever virus in East Africa.*

Abstract

Rift Valley fever (RVF) is an arthropod-borne zoonotic viral disease that is classified as a hemorrhagic fever affecting domestic and wild ruminants, camels and humans. Rift Valley fever virus (RVFV), the causative agent of RVF, can cause large outbreaks in animal and human populations, often crossing significant geographic distances. Here, a detailed genomic, epidemiological and environmental data was used to estimate the epidemic origins of lineage C and reconstruct the demographic and dispersal history of RVFV in East Africa. By estimating the evolutionary rates and dates of lineage C, discrete trait phylogeographic analysis pointed to Zimbabwe as the country of origin. However, this analysis was limited due to the insufficient sampling strategies in the former years. Urbanization, high population density and croplands were associated with lineage direction and velocity. With increased land use, global climate change, increased vector competency and gradual land use, strengthening RVF surveillance activities is crucial especially in urban areas and wetlands which form suitable habitats for urban transmission and vector breeding respectively.

4.1   Introduction

Rift Valley fever (RVF) has been listed and prioritized as one of the diseases that pose great public health risk because of its epidemic potential. As such, the Africa Union (AU) and the World Health Organization (WHO) have earmarked the disease for research and development (A. CDC, 2020; WHO, 2018). The causative agent of RVF, Rift Valley fever virus (RVFV), is a zoonotic agent and a potential biosecurity threat. Consequently, it is of great interest for biosecurity authorities globally (Mandell & Flick, 2010; Rabinowitz et al., 2006). Currently, there are no licensed vaccines available for RVF prevention and treatment in humans although there is a promising live-attenuated Rift Valley fever vaccine hRVFV-4s for single-shot application (LARISSA), which is undergoing human clinical trials (LARISSA, 2022). Veterinary vaccines licensed for use in Africa are based on inactivated and live-attenuated virus (Mansfield et al., 2015). However, these vaccines possess residual virulence making them unsafe for pregnant animals (Botros et al., 2006). In addition, these vaccines have also been linked with teratogenic effects such as abnormal fetus formation and stillbirths in pregnant animals (Makoschey et al., 2016). One of the major outbreaks that had significant public health impact and economic losses was the 1997-1998 epizootic in the horn of Africa. This outbreak was because of abnormal rainfall patterns linked to El-Niño southern Oscillation (ENSO) events (Linthicum et al., 1999; Linthicum et al., 2016). To date, several countries in the eastern Africa region including Uganda (Nyakarahuka et al., 2019; Shoemaker et al., 2019), Tanzania (for Disease Control, (CDC, et al., 1998), Sudan (A. A. Adam et al., 2010; Ahmed et al., 2020; Aradaib et al., 2013; Hassan et al., 2011; WHO, 2018), Zimbabwe (Anderson & Rowe, 1998; Caron et al., 2013; Swanepoel, 2004), Somalia ((for Disease Control, (CDC, et al., 1998; Nderitu et al., 2011) and Madagascar (Andriamandimby et al., 2010; Morvan et al., 1992) have experienced episodes of RVF outbreak at different epidemiological scales. Recently, reports of RVF outbreaks in Burundi (UNOCHA, 2022) and Rwanda (OIE, 2022) have been documented.

Outbreaks of RVF have mostly occurred in populations with low levels of immunity and in close association with prolonged rains or above average rainfall patterns (Anyamba et al., 2009; Linthicum et al., 1985). Following outbreaks, herd immunity is always high and in most endemic settings, there occurs cryptic transmission during interepidemic periods (IEPs). In eastern and southern Africa, outbreaks are associated with prolonged heavy rainfall, a characteristic of ENSO events. ENSO events are characterized by periodic changes in sea surface temperatures (SST) and associated winds in the eastern Pacific Ocean (Linthicum et al., 1999). Early detection and implementation of intervention measures are essential for effective control of RVF outbreaks. These preventive interventions require a thorough understanding of transmission (including cryptic transmission)

and spread of the virus. This is because the epidemiology of the disease is complex. In addition, factors that trigger outbreaks and mechanisms in which the virus persists during interepidemic periods are not fully known. One of the epidemiological systems which has been postulated to maintain transmission in East Africa is the presence of low-lying depressions called 'dambos' where mosquitoes deposit their eggs (Linthicum et al., 2016). Vertical transmission among the mosquitoes allows the infected eggs to persist in these depressions and when favorable climatic conditions occur, mostly heavy rainfall, they hatch to maintain a high density of the vectors (Linthicum et al., 1985; Linthicum et al., 1999).

The ability of the virus to spread to disease-free territories and re-emerge in endemic settings after extended periods of inactivity is a litmus test for public and veterinary health. Global climate change, presence of competent vectors in disease-free nations, increased international trade in livestock, global travel and other anthropogenic factors may potentially contribute to spread of the virus (Pepin et al., 2010). Although several studies have highlighted the spread of RVFV at continental and regional levels using discrete trait phylogeography analysis (Samy et al., 2017; Soumaré et al., 2012), the overall transmission history and potential factors for RVFV spread are yet to be ascertained. Therefore, this study examined the possibility of integrating genetic sequence data and environmental data in a phylogeographic framework to understand the transmission dynamics of RVFV in East Africa. The eastern region of sub-Sahara Africa has been the hardest hit as indicated by the number of outbreaks (Nanyingi et al., 2015). It is also the region where RVF was first reported. As such, studying the transmission history of the virus can shed light into its dispersal patterns.

To deduce the dispersal history of viruses from sequence or any other data, it is essential to model how the pathogens spread through space or host species (Faria et al., 2013; Lemey, Rambaut, et al., 2009). A discrete trait phylogeographic analysis can be used to infer transmission patterns of viruses between discrete locations (Lemey, Rambaut, et al., 2009). However, in most cases, sampling occurs in a continuous manner. When this is the case, geographical coordinates (latitude and longitude) of the ancestral nodes of a phylogenetic tree can be inferred from the sampled locations over time to reconstruct dispersal histories. Landscape phylogeographic frameworks can exploit the spatially explicit phylogeographic reconstructions to formally test the impact of environmental factors on the spread of viruses and estimate dispersal statistics (Dellicour et al., 2019). In this work, we explored the dispersal history of RVFV in East Africa and the potential drivers of outbreaks in the region. Phylodynamic models which consisted of genetic sequence, epidemiological (sampling times and geographical locations) and environmental data was applied. This study described the epidemic origin of lineage C, the transmission pattern

of RVFV between and within East Africa countries as well as potential environmental drivers of RVFV spread in East Africa.

4.2 Methods

4.2.1 In-house generated data

In-house generated genomic data in our lab and publicly available datasets from the National Center for Biotechnology Information (NCBI) GenBank public database were utilized in this study. The in-house datasets were generated by amplicon multiplex PCR and cell culture enrichment sequencing approaches as previously described in chapters two and three. The data included both human and livestock samples that were archived or clinical samples collected during cryptic transmission as part of syndromic surveillance efforts in Kenya, Rwanda and Burundi. The sample collection years ranged between 2007 and 2022. The locations where samples were collected were also appended to the sequences metadata for phylodynamic modelling. The associated metadata of the in-house sample data is shown in **Table 4.1**.

TABLE 4.1: **Sample metadata for in-house generated genomic data** to Samples generated from our lab indicating the sample names, years of collection, hosts and countries of origin. These strains comprised archived and clinical outbreak samples collected as part of syndromic surveillance efforts by relevant country authorities

Sample	Year	Host	Country
BDI0008	2022	cow	Burundi
BDI0007	2022	cow	Burundi
BDI0002	2022	cow	Burundi
BDI0003	2022	cow	Burundi
BDI0006	2022	cow	Burundi
HA-HAR	2007	human	Kenya
08HAB	2018	cow	Kenya
J9	2007	human	Kenya
J10	2007	human	Kenya
J8	2007	human	Kenya
JU-KB	2007	human	Kenya
KEM-BR	2007	human	Kenya
KEM-JC	2007	human	Kenya
KEM-ND	2007	human	Kenya
MSA	2007	human	Kenya
N12	2019	human	Kenya
NA-KOL	2007	human	Kenya
SWA-ISI	2020	human	Kenya
CML	2018	cow	Kenya
AM-M1	2021	cow	Kenya
A-M0	2020	human	Kenya
500618	2007	human	Kenya
RU1	2018	cow	Rwanda
NG10	2018	cow	Rwanda
K2	2007	human	Kenya
DVS-230	2021	cow	Kenya
DVS-321	2021	cow	Kenya
DVS-356	2021	cow	Kenya
DVS-372	2021	cow	Kenya
DVS-333	2021	cow	Kenya

4.2.2 Retrieval and curation of Rift Valley fever viral sequences

In addition to the in-house generated data, Rift Valley fever virus nucleotide sequences were retrieved from NCBI using the keywords: "Rift Valley fever virus segment L AND 6404[SLEN]", "Rift Valley fever virus segment M AND 3885[SLEN]" and "Rift Valley fever virus segment S AND 1520:1695[SLEN]" for complete L (n = 231), M (n = 233) and S (n = 310), respectively, as of November 19, 2022. The search term also included SLEN (Sequence Length), which indicates the total number of nucleotides in the sequence. The retrieved sequences were combined with new sequences generated from our lab as described earlier. New sequences were from RVFV strains collected in Rwanda (unpublished data), Burundi (unpublished data) and Kenya (Juma et al., 2022). The resulting dataset was filtered by: (i) excluding laboratory strains (multiply passaged, reassortants, antivirals or vaccine experiments), (ii) keeping only sequences with specified lengths in the keywords and (iii) excluding sequences with no known geographical locations (no data in terms of administrative unit where the sample was collected from). The filtered complete L (n = 204), M (n = 211) and S (n = 298) sequences formed the global datasets.

To conduct phylodynamic and phylogeographic analysis at regional level, in this case, East Africa, the global sequences were further filtered to retain only sequences from the eastern and southern African countries (Burundi, Kenya, Madagascar, Mauritius, Mayotte, Rwanda, Sudan, Tanzania, Uganda, and Zimbabwe) that previously recorded RVF outbreaks. Zimbabwe was included in the analysis to ascertain the origin of lineage C and to infer the transmission dynamics of the virus. . The resulting sequences were further filtered by deduplication on sequence composition using seqkit (Shen et al., 2016). Using the sample location information, the sequences were geocoded using googleapis https://maps.googleapis.com/maps/api/geocode/ to obtain latitude and longitude coordinates at administrative level one. For each genomic segment dataset, multiple sequence alignment was performed using Multiple Alignment using Fast Fourier Transform (MAFFT) v7.475 (Katoh & Standley, 2013) applying FFT-NS-2 strategy with progressive guide trees. The alignments were manually inspected in AliView v1.28 (Larsson, 2014) and trimmed to correct codon alignment. For the complete S segment, the alignments were trimmed to obtain non-structural (NSs) and nucleocapsid (NP) gene datasets. Given the antisense nature of the S segment, the reverse complement of nucleocapsid gene was obtained. Nucleotide alignments for M and L segments were obtained following trimming of untranslated regions (UTRs) and primer binding sites. For each of the four alignments (L (n = 87), M (n = 92), NSs (n = 105) and NP (n = 105)), the optimal evolutionary model was identified using ModelTest-Ng v0.1.7 (Darriba et al., 2012). For L and M sequence alignments, a generalized time reversible (GTR) (Tavaré, 1986)

with Gamma (Γ) distribution at four sites was the substitution model of choice as determined by Bayesian Information Criterion (BIC), while NSs and NP alignments were better modelled using Hasegawa-Kishino-Yano (HKY) (Hasegawa et al., 1985) with Gamma (Γ) distribution at four sites. Preliminary maximum likelihood (ML) phylogenetic trees were inferred using IQ-TREE v2.0.3 (Nguyen et al., 2015) using the substitution models estimated with bootstrapping of 1000 replicates using the -bb option. TempEst v1.5.3 (Rambaut et al., 2016) was used to estimate the root-to-tip distances by applying the best-fitting root with a residual mean squared function from the resulting ML trees. Regression was also performed in R statistical computing environment (R Core Team, 2017) using a linear model *(lm)* function of distances on the sequence dates to augment root-to-tip regression by TempEst. Therefore, molecular clock analysis was performed based on the relationship between root-to-tip divergence and sample dates. To visualize the genetic diversity, SNPs were extracted from the multiple sequence alignments using SNP-Sites (Page et al., 2016) and summarized using snipit (O'Toole, 2022).

4.2.3 Preliminary time-scaled phylogenetic analysis

Preliminary time-scaled phylogenetic inference was performed on complete L, M and S segments. In these analyses, complete S segment sequences were queried on NCBI by applying the search term, "Rift Valley fever virus segment S AND 1690[SLEN]". For this analysis, there were 204, 211 and 95 sequences for the L, M and S segments respectively. Complete L and M segment sequences were retrieved using the search terms as described in the previous paragraph. The same filtering procedure as described above was applied for these datasets. Multiple sequence alignment was obtained by using MAFFT followed by manual inspection and editing in AliView (Larsson, 2014). Maximum likelihood tree inference was performed using IQ-TREE followed by time-scaled phylogenetic reconstruction using TreeTime (Sagulenko et al., 2018) on the alignment using collection year of the sequences as dates.

4.2.4 Model selection and molecular clock analyses

Bayesian Estimation Analysis by Sampling Trees (BEAST) v1.10.4 (Suchard et al., 2018) was used to perform molecular clock analyses. Preliminary model selection was exclusively conducted using the M segment dataset. The model testing procedure was performed using four coalescent tree priors: two parametric, a constant-size population and exponential growth population and two non-parametric, a Bayesian skyline (ten groups, piecewise-constant model) (Drummond, 2005) and a Bayesian skygrid tree prior (number of parameters = 50, time at last transition point = 200) (Gill et al., 2013). For each tree prior, two clock models were tested, an uncorrelated relaxed clock with log normal distribution (UCLN) (Drummond et al., 2006) and a strict clock. In each model tested,

a continuous time Markov chain (CTMC) rate reference prior (Ferreira & Suchard, 2008) and a GTR with a discrete gamma-distributed rate heterogeneity (Yang, 1994) substitution model were applied. For the eight combinations of models, path-sampling (Lartillot & Philippe, 2006) and stepping-stone (SS) (Xie et al., 2011) were applied to generate marginal likelihood estimators (MLE). The procedures involved sampling for 100 path steps with a chain length of one million, with power posteriors determined for evenly spaced quantiles of a Beta (alpha = 0.3; 1.0) distribution and sampling at every 1000th iteration. BEAGLE 3 library was incorporated in the analyses to speed-up the likelihood approximation process (Ayres et al., 2019). From the model selection testing, Bayesian skyline and skygrid priors offered better fit of the data compared to the other two models as indicated by the log Bayes factors (**Table 4.2**). Therefore, Bayesian skygrid tree prior was used in all the subsequent analyses. A relaxed clock model provided a better model fit using constant or exponential tree prior as indicated by the log Bayes factors. The difference in log Bayes factors between a relaxed and strict clock when using either skyline or skygrid tree prior is smaller than it is when using the other tree priors. The variability in the log maximum likelihood estimates from PS and SS using both skyline and skygrid tree priors is also like the log Bayes Factor difference. A skygrid tree prior with relaxed clock was selected for further analyses since its estimate of the time of the most recent ancestor (TMRCA) closely matched with previous approximation (Bird, Khristova, et al., 2007). Time-scaled phylogenetic trees were reconstructed using skygrid tree prior while accommodating for branch-specific evolutionary rate using a relaxed molecular clock with an underlying log-normal distribution. BEAST was run with 600 million Markov chain Monte-Carlo (MCMC) steps. The convergence and mixing properties of the output were inspected to ensure effective sample size (ESS) values linked with estimated parameters were all > 200 using Tracer v1.7.2 (Rambaut et al., 2018).

TABLE 4.2: **Model selection of priors in BEAST** Marginal likelihood estimators computed using path-sampling (PS) and stepping-stone (SS) for combinations of 4 coalescent tree priors (constant-size population, exponential growth population, Bayesian Skyline and Bayesian Skygrid) and two molecular clock models (strict clock and uncorrelated relaxed clock with log-normal distribution). The Bayes Factor is calculated against the baseline model, a parametric constant-size tree prior and a strict clock. Abbreviations as used in this table: CS = Constant Strict, CR = Constant Relaxed, ES = Exponential Strict, ER = Exponential Relaxed, SLS = Skyline Strict, SLR = Skyline Relaxed, SGS = Skygrid Strict, SGR = Skygrid Relaxed.

Sampling	Variable	CS	CR	ES	ER	SLS	SLR	SGS	SGR
Path Sampling	log(maximum likelihood)	-19032	-18919	-19016	-18907	-18973	-18876	-18893	-18888
	log(Bayes Factor)	-	113	16	125	59	156	39	144
Stepping-stone	log(maximum likelihood)	-19034	-18922	-19019	-18910	-18977	-18879	-18996	-18893
	log(Bayes Factor)	-	112	15	124	57	155	38	141
	tMRCA	1852[1830-1873]	1890[1833-1922]	1859[1840-1877]	1878[1830-1915]	1863[1847-1878]	1873[1812-1913]	1852[1831-1870]	1881[1844-1911]
	Clock rate	1.9E-4[1.6E-4-2.2E-3]	-	2.0E-4[1.7E-4-2.3E-3]	-	2.1E-4[1.8E-4-2.4E-4]	-	1.9E-4[1.6E-4-2.1E-4]	-

4.2.5 Estimating the epidemic origin and discrete phylogeographic analysis of RVFV lineage C

To estimate the dates and origin of lineage C, 405 partial ($\geq$ 490bp) and complete (3885 bp) M segment sequences were retrieved from the NCBI GenBank public database (search term: "Rift Valley fever virus segment M AND 490:3885[SLEN]"). We chose the partial glycoprotein gene (Gn) and complete M segment sequences for this analysis for the following reasons, (i) these sequences captured good geographical representation in terms of sample collection sites and (ii), they could be confidently assigned to lineages using the glycoprotein classifier as reported in chapter 2 of this thesis (Juma et al., 2022). These sequences were combined with the new sequences generated from our lab. Lineage assignment was performed on the 433 sequences using the RVFV typing tool described in chapter 2 of this thesis (Juma et al., 2022). All sequences assigned to lineage C were extracted and filtered to deduplicate on composition resulting in a final dataset of 127 unique sequences. Multiple sequence alignment was performed using MAFFT. Partial glycoprotein gene sequences spanning positions 815 and 1305 were subset and used for subsequent phylogenetic inference. These coordinates represent the start and end positions of the 490 bp M segment partial glycoprotein gene sequence. The HKY+Γ+4 substitution model was identified as the best evolutionary representation using ModelTest-Ng. An initial maximum likelihood tree inference was performed with IQ-TREE using ultra-fast bootstrapping of 1000 replications. Assessment of the temporal signal of the sequences was performed using TempEst. Molecular clock analysis was performed using BEAST with incorporation of the HKY+Γ+4 substitution model and skygid prior while accommodating for branch-specific evolutionary rate using the uncorrelated lognormal relaxed model. To estimate the time of the most common recent ancestor and evolutionary rates of Kenya and Zimbabwe sequences, the alignment data was partitioned into two taxa consisting of sequences collected between 1976 and 1983. BEAST was run with 200 million MCMC steps. Inspection of the output to ensure effective sample size (ESS) values linked with estimated parameters were all > 200 was performed using Tracer v1.7.2. To obtain maximum clade credibility (MCC) trees, TreeAnnotator was used with a burn-in of between 10-30%. Bayesian discrete phylo-

geographic analysis was performed using partial glycoprotein gene sequence data and sampling locations (country and exact locations) to estimate the ancestral locations of lineage C. Ten countries ("BDI", "KEN", "MDG", "MOZ", "MYT", "RWA", "SDN", "TZA", "UGA" and "ZWE") and 55 locations represented the dataset. The location exchange process in the phylogeny was modelled using an asymmetric (non-reversible) continuous time Markov chain (CTMC) (Ferreira & Suchard, 2008; Lemey, Rambaut, et al., 2009). To infer a minimum set of location exchange rates that provide sufficient

description of the viral dissemination process, a Bayesian stochastic search variable selection (BSSVS) procedure was applied (Lemey, Rambaut, et al., 2009). A 'robust counting' procedure to approximate the expected number of location changes along the branches of a posterior tree distribution (Minin & Suchard, 2008a, 2008b; O'Brien et al., 2009) was applied in the analysis. BEAST was run with 100 million MCMC steps, sampling every 10,000 steps, resulting in 10,000 samples. Convergence and mixing properties of the output were examined to ensure effective sample size (ESS) values linked with estimated parameters were all greater than 200 using Tracer v1.7.2. TreeAnnotator was used to obtain MCC trees with a burn-in of between 10-30%. Visualization of the MCC trees and computed Bayes factor support for rates was performed using Spatial Phylogenetic Reconstruction and Evolutionary Dynamics using Data-Driven Documents (SpreaD3) (Bielejec et al., 2016).

4.2.6   Phylogeographic inference in continuous space

To characterize the dispersal history and dynamics of RVFV in a continuous diffusion approach, a relaxed random walk (RRW) model was applied (Lemey et al., 2010; Pybus et al., 2012). This model allows for inference of spatially and temporally referenced phylogenies while accommodating variation in dispersal velocity among branches (Baele et al., 2017). To model among-branch heterogeneity in diffusion velocity, a Cauchy model was used (**Table 4.3**). The bivariate traits were represented as latitude and longitude with addition of a random jitter to tips at a window size of 0.05. BEAST was run with $1.5{\times}10^4$ MCMC steps, sampling every $150,000^{th}$ step for the L segment, $8.0{\times}10^8$ MCMC steps, sampling $80,000^{th}$ step for M segment and NP gene sequences and $6.0{\times}\ 10^8$ MCMC steps, sampling at $60,000^{th}$ step for the NSs gene sequences. The analyses were run until all the ESS values associated with the estimated parameters were greater than or equal to 200. To obtain spatially explicit phylogeographic reconstruction, TreeAnnotator was used to obtain the MCC trees representation with a burn-in of 30%. The phylogeographic reconstructions were performed on the four datasets with sequences originating from East Africa comprising of L (n=87), M (n=92) and both NSs and NP (n=105).

TABLE 4.3: **Model selection for phylogeographic inference in BEAST** Marginal likelihood estimators computed using path-sampling (PS) and stepping-stone (SS) for combinations of 2 coalescent tree priors (Bayesian Skyline and Bayesian Skygrid) and 3 diffusion models (Cauchy, Gamma and Brownian distribution). The Bayes Factor is calculated against the Brownian diffusion model as the baseline.

Sampling	Factor	Cauchy Skyline	Cauchy Skygrid	Gamma Skyline	Gamma Skygrid	Brownian Skyline	Brownian Skygrid
Path Sampling	log(maximum likelihood)	-19331	-19439	-19087	-19042	-19914	-20062
	log(Bayes Factor)	731	623	975	1020	148	-
Stepping-stone	log(maximum likelihood)	-19316	-19442	-19086	-19044	-19910	-20058
	log(Bayes Factor)	742	616	972	1014	148	-

UNIVERSITY *of the*

WESTERN CAPE

4.2.7 Landscape phylogeography

Phylogenetic branches from spatiotemporal referenced trees can be treated as conditionally independent movement vectors (Pybus et al., 2012). To summarize virus diffusion over time and space, 1000 post-burn-in trees sampled at regular intervals from the posterior distribution were extracted from the MCC trees. The R package seraphim (Dellicour, Rose, Faria, et al., 2016) was used to extract the spatiotemporal information (start and end spatial coordinates and start and end dates) in each of the trees. Spatial dissemination statistics such as dispersal velocity, diffusion coefficients and evolution of the maximal wavefront distance from the epidemic origin were estimated from the extracted trees.

The environmental factors tested for the association on dispersal direction of RVFV were elevation, temperature, precipitation, normalized difference vegetative index (NDVI), human population, croplands, pastures, urban areas, forests (primary and secondary) and non-forests (primary and secondary) (**Figure** 4.1). The estimation procedure begins by computation of the E statistic, which calculates the mean environmental values at tree node positions. The R ratio is also computed, which is the proportion of branches for which the environmental value recorded at the oldest node position is higher than the environmental value recorded at the youngest node position. Therefore, whereas E measures the tendency of viral lineages to remain in higher or lower environmental values, R determines the tendency of the lineages to disperse towards these areas. To check if posterior distributions of both E and R were significantly different by chance under a null dispersal model, these statistical values were also computed based on simulations following randomization of the phylogenetic tree nodes. When the inferred distribution is significantly lower than the simulated distribution, there is evidence for the environmental factor to repulse RVFV while if the inferred distribution is higher than the simulated distribution, the environmental factor is deemed to attract RVFV viral lineages.
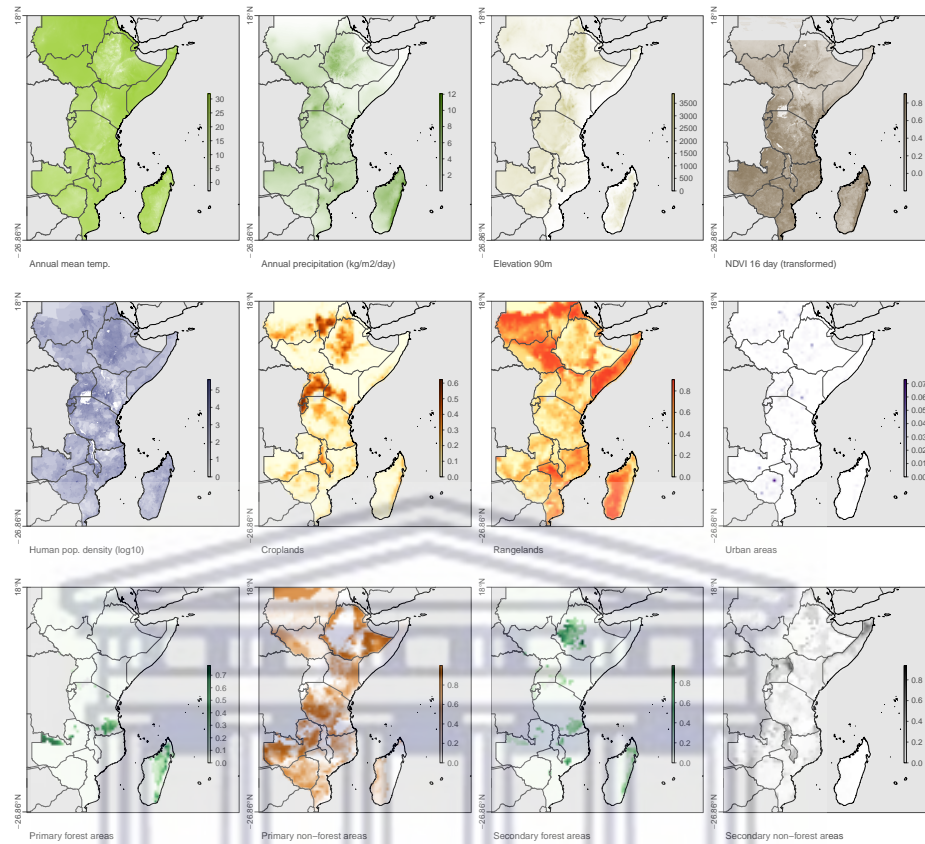
FIGURE 4.1: **Environmental variables tested for their impact on RVFV dispersal in eastern Africa.** Assessment of the impact of environmental factors on the dispersal dynamics of RVFV in eastern Africa using seraphim.

4.3 Results

4.3.1 Molecular clock analyses of lineage C splits the lineage into two sublineages

The predominant lineage of RVFV circulating in Africa is lineage C as reported in our previous analyses (Juma et al., 2022). To ascertain the evolutionary characteristics of this lineage, molecular clock phylogenetics analysis was carried out using BEAST (Suchard et al., 2018). Specifically, the main objective of the analysis was to determine the geographic origin, time of most recent ancestor (TMRCA) as well as evolution rate of lineage C using partial glycoprotein gene sequences. Preliminary maximum likelihood reconstruction on glycoprotein (n = 127) sequences revealed 1963 as the time of the most recent ancestor following regression of the root-to-tip distances against time (years). The evolution rate was determined to be 3.2×10^{-4} substitutions per site per year (subs/per/site/year). Bayesian phylogenetics estimated a mean rate of 5.3×10^{-4}, 95% Highest Posterior Density (HPD) [3.1×10^{-4} - 7.7×10^{-4}] substitutions per site per year. Molecular clock phylogenetic analysis using a skygrid prior revealed that the most recent common ancestor of lineage C appeared at the same time in both Kenya and Zimbabwe (1970.706, 95% HPD [1959-1975] in Kenya and 1970.406, 95%HPD [1961-1975] in Zimbabwe). This was further supported by high posterior probability values (1 at 95% HPD) in the maximum clade credibility trees, placing the Kenya and Zimbabwe samples at the root of the tree. Furthermore, an expansion of lineage C with two dominant sublineages represented by C.1 and C.2 was observed (**Figure** 4.2). This is a significant finding, that introduces a sub-lineage classification into lineage C that will enable finer genotyping studies of the virus epidemiology.
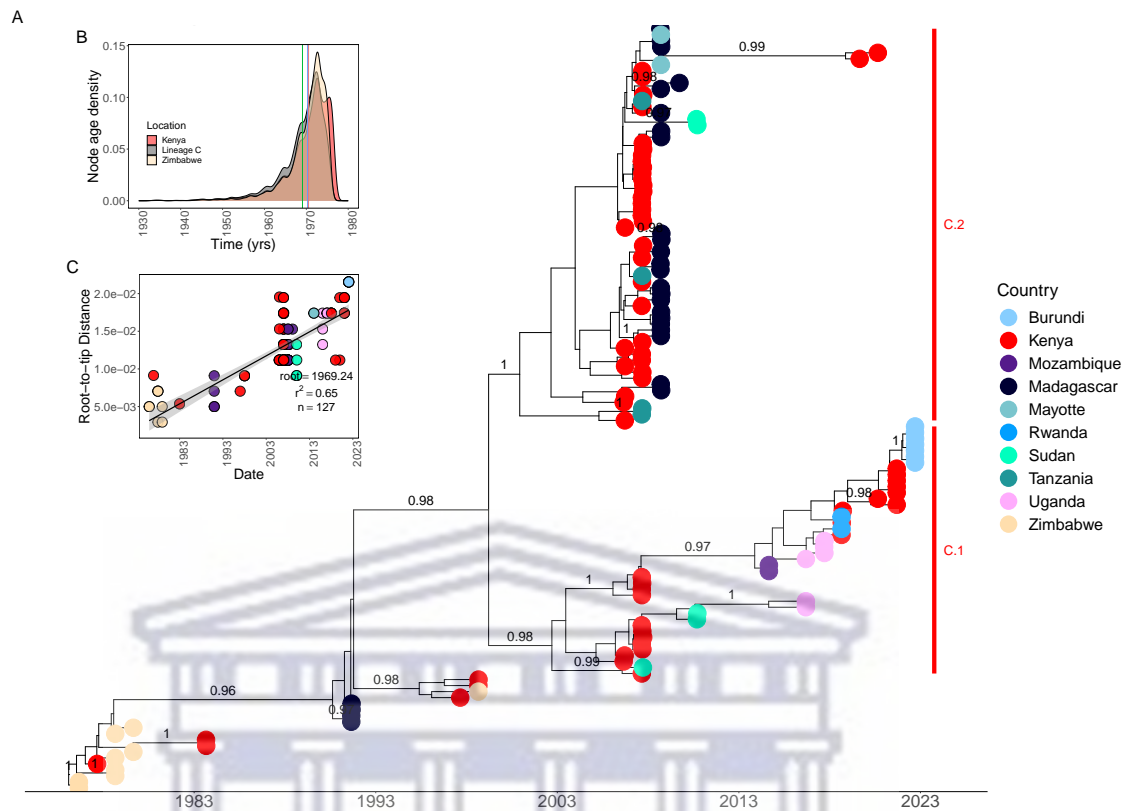
FIGURE 4.2: **Maximum clade credibility (MCC) trees based on the glyco-protein gene sequences (n = 127).** Time-scaled phylogeographic history of lineage C in East Africa. Panel (**A**) shows the time-scaled phylogenetic tree. The tips of the tree are colored according to country of origin of the isolates. The posterior probabilities for their ancestral nodes are also indicated. The top right inset plot (**B**) shows the node density of time of the most recent common ancestor for sequences obtained in Kenya, Zimbabwe and all the other isolates. The lower inset plot (**C**) is the regression of distances and sequence dates indicating a low temporal signal.

4.3.2 Discrete phylogeographic analysis indicates Zimbabwe as the origin of lineage C

Discrete phylogeographic analysis placed the origin of lineage C in Zimbabwe. Estimation of transmission rates of RVFV lineage migration in the eastern African region using a robust counting procedure identified statistically significant epidemiological links between countries and locations. Earlier spread of lineage C occurred in several locations in Zimbabwe including Gatooma (presently called Kadoma), Salisbury (presently known as Harare) and Sinoia (presently known as Chinhoyi) in 1975. The virus then spread to the coastal town of Mombasa in Kenya in 1978 followed by another separate introduction to the central Kenyan town of Thika in 1985. A separate introduction to Madagascar from Zimbabwe occurred in 1988. In the early 1990s, in-country transmission of the virus was observed in the Indian Ocean Island of Madagascar. From the late 2000s, the virus has spread from Kenya to other East African countries including Tanzania (Dodoma and Tanga) in 2006/2007, Madagascar (2008) (Andriamandimby et al., 2010) and Sudan (White Nile and Gezira) in 2010. More recently viral exchanges of lineage C have occurred in Uganda (Sembabule) from Sudan (2016) and from Uganda to Rwanda (2018) (**Figure 4.3** and **Figure 4.4**).
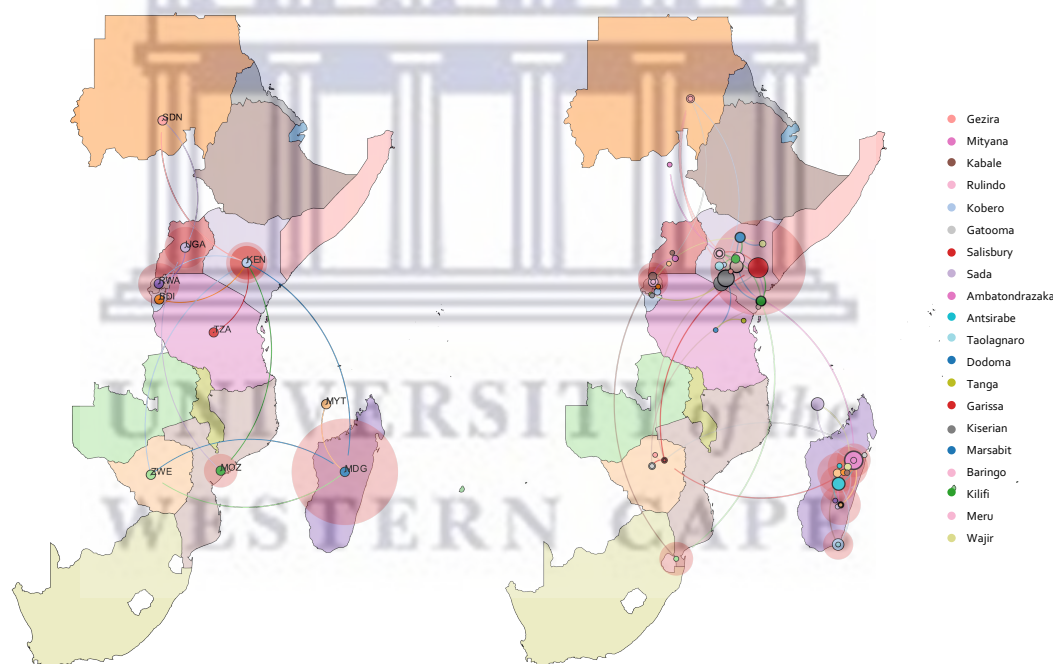
FIGURE 4.3: **Spatiotemporal diffusion of RVFV lineage C in Eastern Africa Left panel** shows the diffusion of RVFV lineage C in eastern Africa countries under a discrete phylogeographic analysis. Tree branches are colored according to the parental countries (n = 10). The circular polygon areas are proportional to the number of tree lineages maintaining that location (country). **Right panel** is a similar data driven document (D3) visualization generated by SpreaD3 (Bielejec et al., 2016). In this image, the tree branches depict actual locations (n = 56). (KEN=Kenya, MDG=Madagascar, SDN=Sudan, UGA=Uganda, ZWE=Zimbabwe, BDI=Burundi, RWA=Rwanda, TZA=Tanzania, MOZ=Mozambique, MYT=Mayotte).
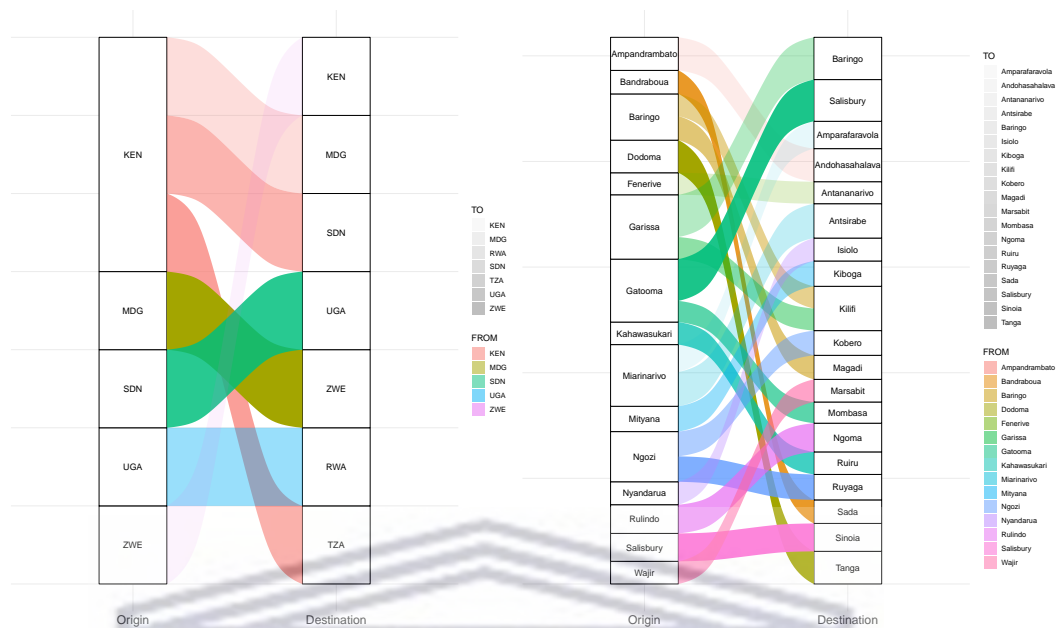
FIGURE 4.4: **Exchanges of lineage C in Eastern and Southern Africa** Number of lineage C viral exchanges between eastern Africa countries **(Left panel)** and locations **(Right panel)**. The intensity of the are represented by Bayes factor values summarized by SpreaD3 from state counts using BSSVS robust counting procedure. (KEN=Kenya, MDG=Madagascar, SDN=Sudan, UGA=Uganda, ZWE=Zimbabwe, RWA=Rwanda, TZA=Tanzania). For viral exchanges between countries, Bayes Factor threshold following robust counting with BSSVS was set to a threshold of > 10, while for viral exchange events in locations, the threshold was set to > 50

4.3.3   The genetic diversity of lineage C

Genetic diversity of lineage C was first inspected following multiple sequence alignment of codon corrected M and L segments as well as NSs and NP gene sequences on the S segment. SNPs (single nucleotide polymorphisms) were extracted and categorized as singletons, conserved or multiple SNPs for every position in the reference sequence. Many singleton SNPs in M and L segments and in gene sequences in the S segment were identified. Within the M segment, there was evidence of increased diversity with time indicating an ongoing evolution process (**Figure 4.5**). There was an accumulation of SNPs in lineage C in compared to other lineages (**Figure 4.5**).
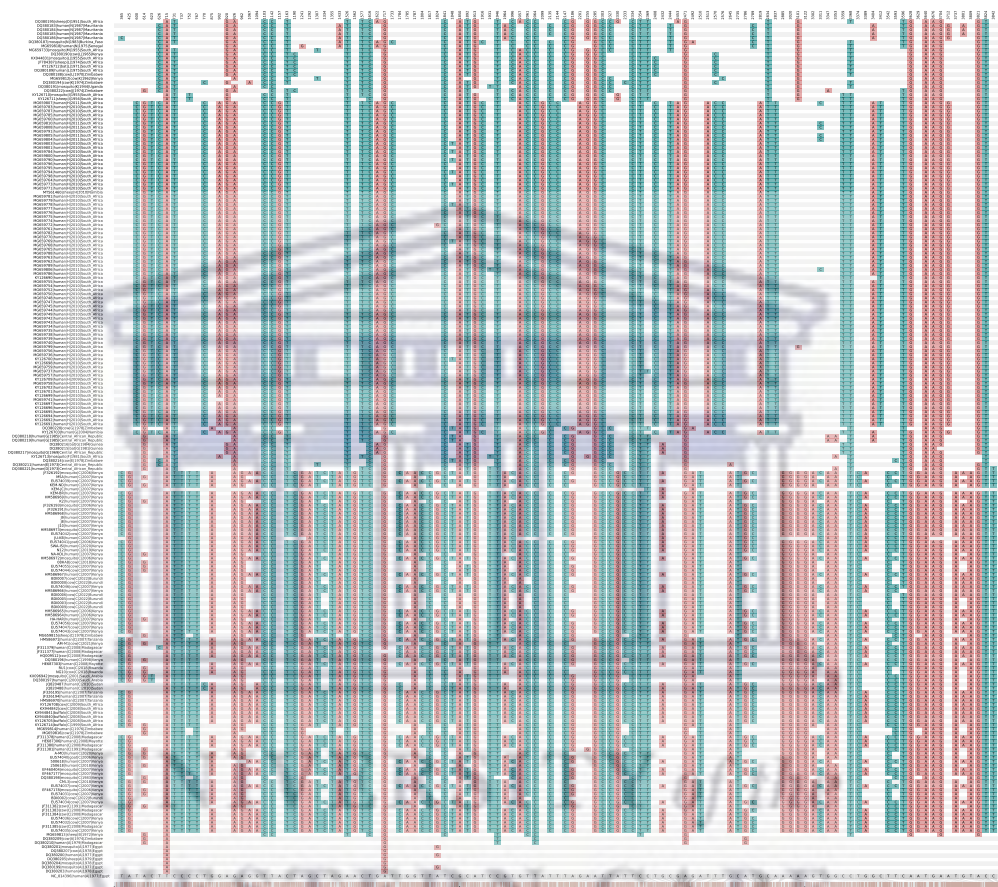
FIGURE 4.5: **SNPs summary relative to the reference sequence of RVFV M segment.** Multiple sequence alignment of singleton SNPs. Rows represent sequence identifiers and relevant metadata (host species, lineage, year of collection and country), and are ordered by lineage. The columns represent positions within the genome where SNPs occur colored as green - pyrimidines (C and T) and light maroon - purines (G and A).

Preliminary molecular clock analyses of complete L, M and S segment genome sequences showed expansion of lineage C with the most predominant sub-lineages represented by C.1 and C.2 as indicated by time-scaled phylogenetic trees (**Figure 4.6**, **Figure 4.7**, **Figure 4.8**).
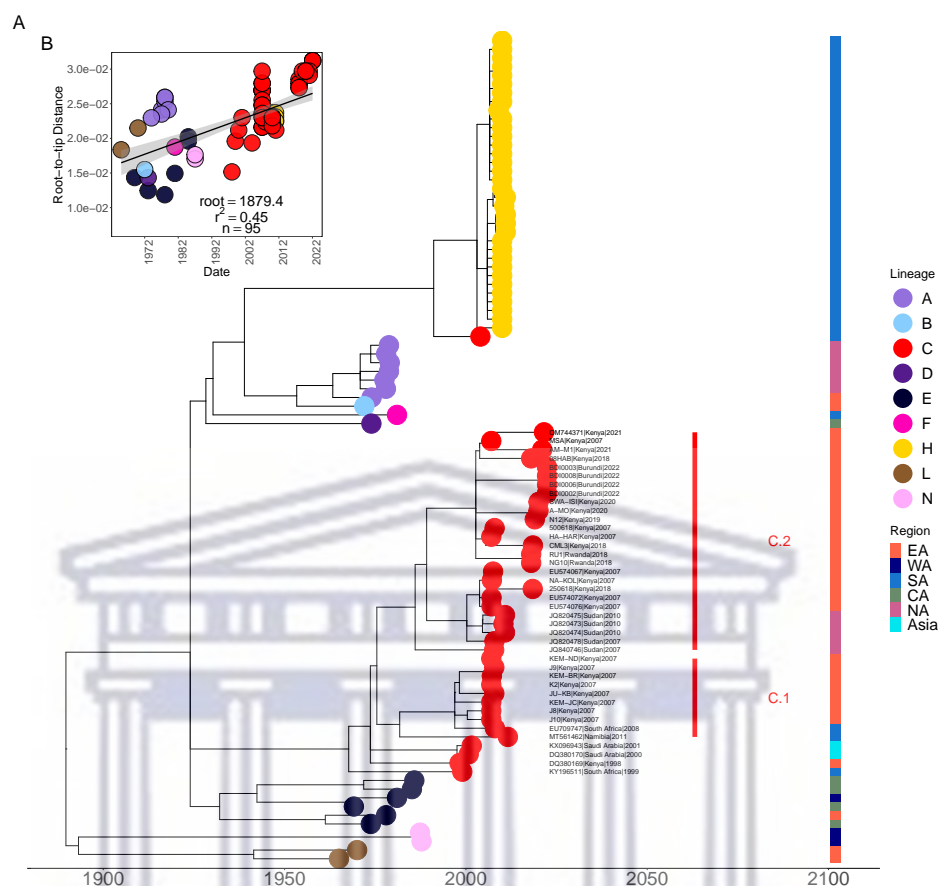
FIGURE 4.6: **Time-scaled phylogenetic tree showing lineage expansion using the complete S segment genome sequences.** Molecular clock phylogeny (**A**) using complete S segment genome sequences (n = 95) indicating expansion of lineage C through time and geographical occurrence mostly in Africa and Western Asia. Inset the plot (**B**) is the root-to-tip regression of distances against sampling dates for sequences containing information on the correlation of the time and sequence occurrence (r^2) and the inferred evolution rate as the number of substitutions per site per year. Tip labels are colored by lineage. The Inset horizontal bar shows the categories of the sub-regions in Africa and Asia. ***EA=East Africa, WA=West Africa, SA=South Africa, NA=North Africa, CA=Central Africa**.
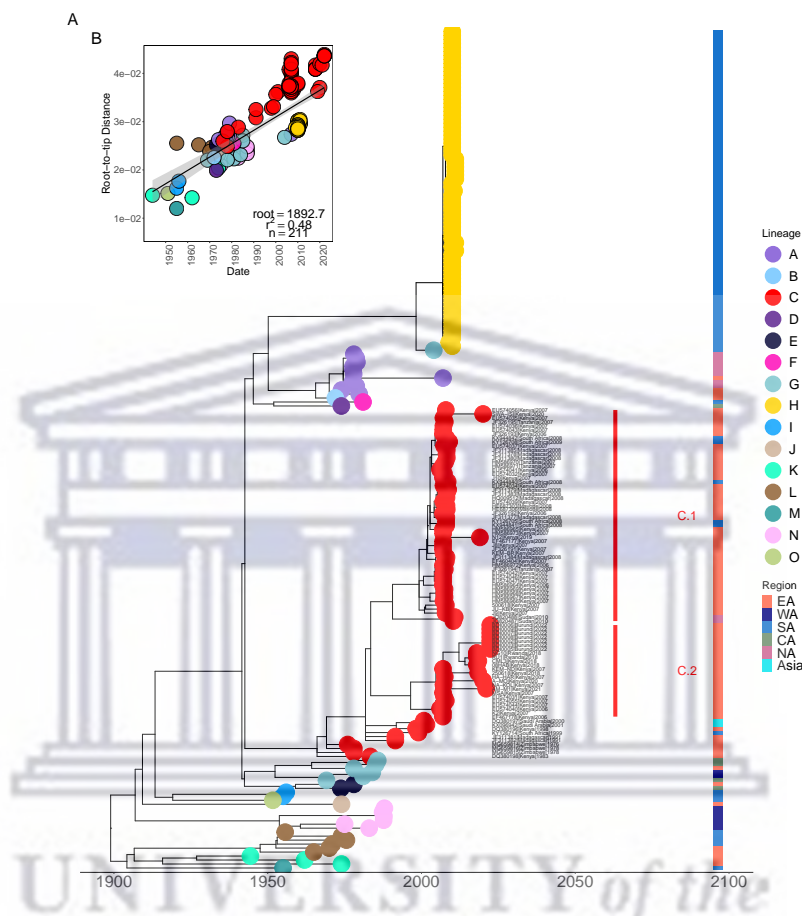
FIGURE 4.7: **Time-scaled phylogenetic tree showing lineage expansion using the complete M segment genome sequences.** Molecular clock phylogeny (**A**) using complete M segment genome sequences (n = 211) indicating expansion of lineage C through time and geographical occurrence mostly in Africa and Western Asia. Inset the plot (**B**) is the root-to-tip regression of distances against sampling dates for sequences containing information on the correlation of the time and sequence occurrence (r^2) and the inferred evolution rate as the number of substitutions per site per year. Tip labels are colored by lineage. The Inset horizontal bar shows the categories of the sub-regions in Africa and Asia. ***EA=East Africa, WA=West Africa, SA=South Africa, NA=North Africa, CA=Central Africa**.
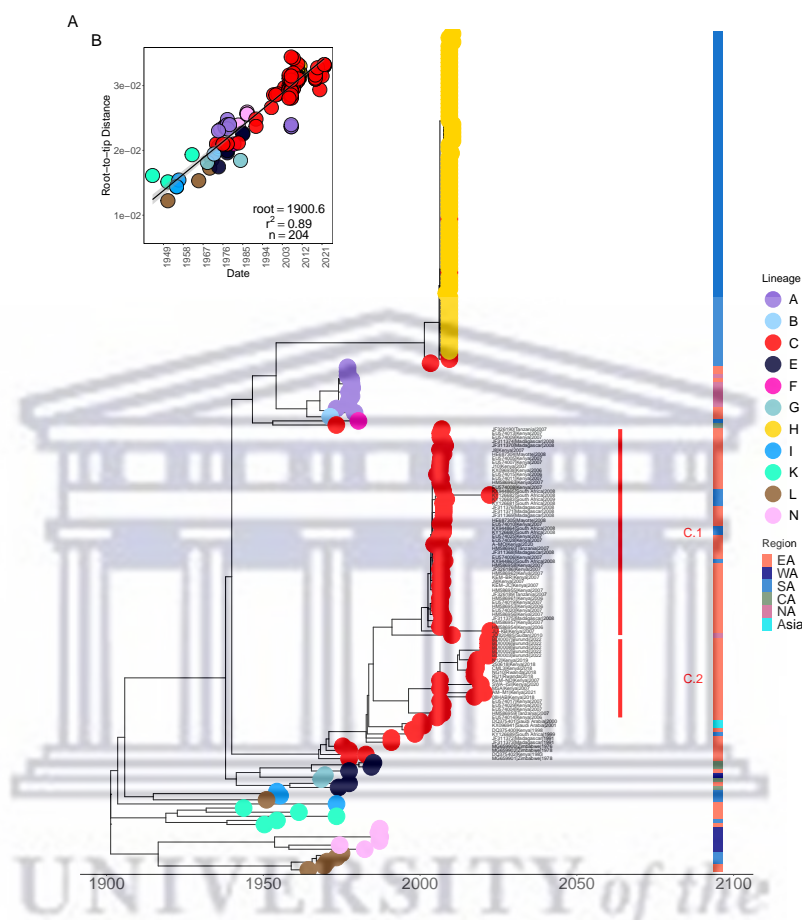
FIGURE 4.8: **Time-scaled phylogenetic tree showing lineage expansion using the complete L segment genome sequences.** Molecular clock phylogeny (**A**) using complete L segment genome sequences (n = 204) indicating expansion of lineage C through time and geographical occurrence mostly in Africa and Western Asia. Inset the plot (**B**) is the root-to-tip regression of distances against sampling dates for sequences containing information on the correlation of the time and sequence occurrence (r^2) and the inferred evolution rate as the number of substitutions per site per year. Tip labels are colored by lineage. The Inset horizontal bar shows the categories of the sub-regions in Africa and Asia. ***EA=East Africa, WA=West Africa, SA=South Africa, NA=North Africa, CA=Central Africa**.

Within the C.1 sub-lineage, there was emergence of a distinct subclade from the Sudan (2010) outbreak isolates. In both sub-lineage C.1 and C.2, the 2010 outbreak isolates from Sudan may have been responsible for the Uganda outbreak in 2016 and the cryptic transmission of the virus observed in Kenya in 2019/2020 (Appendix C, **Figure C.2**, **Figure C.3**, **Figure C.4** and **Figure C.5**). The root-to-tip distance regression against time can provide an evolutionary rate of the viral genomes. In this preliminary analysis, the substitution rates inferred were: 1.91×10^{-4}, 2.70×10^{-4} and 2.89×10^{-4} for S, M and L segment genome sequences, respectively. As described in our previous study in chapter 2 (Juma et al., 2022), these analyses revealed that lineage C was mostly predominant in the Eastern and Southern African region. Sub-lineages of lineage C were observed in the MCC trees obtained from continuous phylogeographic analysis (Appendix C, **Figure C.2**, **Figure C.3**, **Figure C.4** and **Figure C.5**). The analyses indicated a slow molecular clock for RVFV using all the complete segment genomic data as well as the NSs and NP genes sequence data from eastern Africa. The substitution rates observed were: 7.39×10^{-4}[95% HPD: 4.28×10^{-4}-1.0×10^{-3}], 8.12×10^{-4}[95% HPD: 4.52×10^{-4}-1.13×10^{-4}], 5.29×10^{-4}[95% HPD: 2.27×10^{-4}-8.19×10^{-4}], 3.09×10^{-4}[95% HPD: 1.58×10^{-4}-4.83×10^{-4}] substitutions per site per year for NSs, NP, M and L sequences respectively. Extracted single nucleotide polymorphisms for the L segment, NSs and NP gene sequences can be found in Appendix C (**Figure C.6**, **Figure C.7** and **Figure C.8**).

4.3.4   Continuous phylogeographic inference

To reconstruct the dispersal history of RVFV in East Africa, a spatially explicit Bayesian phylogeographic analysis using genomic sequences obtained from public databases (n = 61, 66 and 79 for L, M and S segments) and those generated from our lab (n = 26). The RVFV genome is tripartite having large (L), medium (M) and small (S) segments that may reassort as has been previously shown (Sall et al., 1999). Since reassortment may lead to distinct evolutionary profiles for the L, M and S segments, these genomic sequences (n = 87, 92 and 105 respectively) were analyzed separately for phylogeographic inference. For the S segment, phylogeographic analysis was performed using the non-structural and nucleocapsid gene sequences independently. The spatiotemporal spread of RVFV was captured in the spatially explicit trees with each branch indicating dispersal between an estimated start and end location and linked with a duration estimate (**Figure 4.9**, **Figure 4.10**, **Figure 4.11**, **Figure 4.12**).

To estimate how fast RVFV circulates in eastern Africa, spatiotemporal information contained in annotated phylogenetic trees sampled from the posterior distribution of continuous phylogeographic analysis was extracted. From the extracted spatially explicit trees, estimation of various dispersal statistics was computed using seraphim (Dellicour et al., 2017; Dellicour, Rose, & Pybus, 2016) Specifically, the mean and weighted dispersal velocities and maximal wavefront distance were quantified from the movement data. Mean dispersal velocity is the total geographic distance (in kilometers) travelled by the virus divided by the time elapsed in years. The weighted dispersal velocity is the total distance covered by the dispersal events in the trees divided by the summation of the corresponding durations. From the analysis, RVFV circulates with a mean dispersal velocity of 319.5 (**Figure 4.9**), 356.8 (**Figure 4.10**), 406.3 (**Figure 4.11**) and 308.7 (**Figure 4.12**) km/year based on the NSs, NP, M and L genome sequences respectively. The weighted dispersal velocities of the virus were: 77.6, 92.5, 66.6 and 50.3 km/year for the NSs, NP, M and L segments respectively. There were several peaks of dispersal velocity with distinct spikes occurring between 1930s to 1970s, 1997 and 2007. The early peak corresponds to the expansion phase of RVFV spread in its endemic settings while the 1997 and 2007 peaks correspond to the massive outbreaks that ravaged East Africa following abnormal rainfall patterns termed El-Niño southern oscillation (ENSO). This is further supported by the estimation of the maximal wavefront distance from the epidemic origin through time. The virus traversed great distances in its early expansion phase between the 1930s and 1970s.

105 NSs sequences with known geographical locations.

Mean branch dispersal velocity = 319.48 km/year, 95% HPD = [176.2 − 1349.5].

Weighted branch dispersal velocity = 77.6 km/year, 95% HPD = [42.0 − 95.2].
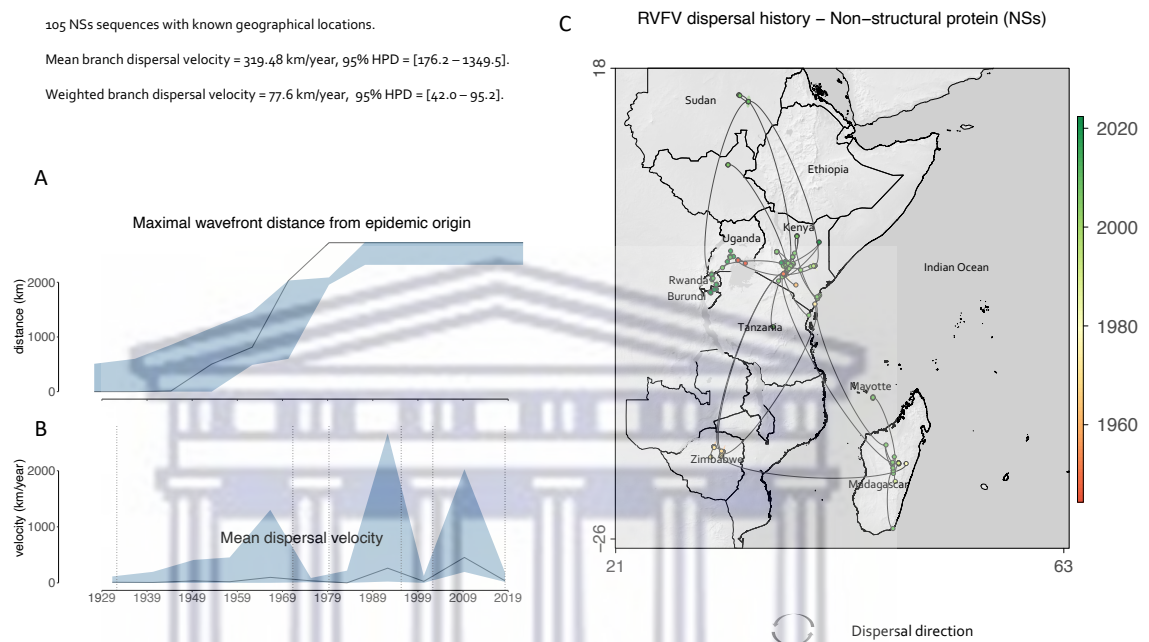
FIGURE 4.9: **Spatiotemporal diffusion and dispersal dynamics of RVFV in Eastern African region using the non-structural (NSs) sequences.** RVFV dispersal statistics showing the evolution of the maximal wavefront distance from the epidemic origin (**A**) and mean dispersal velocity over the entire virus spread through time (**B**). The shaded blue area corresponds to the 95% credible regions of the estimated wavefront position. Dispersal history of RVFV obtained by continuous phylogeographic inference based a Maximum clade credibility (MCC) tree sampled on 1000 posterior trees (**C**) . Nodes of the trees are colored according to the time of occurrence, and oldest nodes are plotted on top of youngest nodes. The trees are superimposed on maps displaying features such as the ocean and lakes.
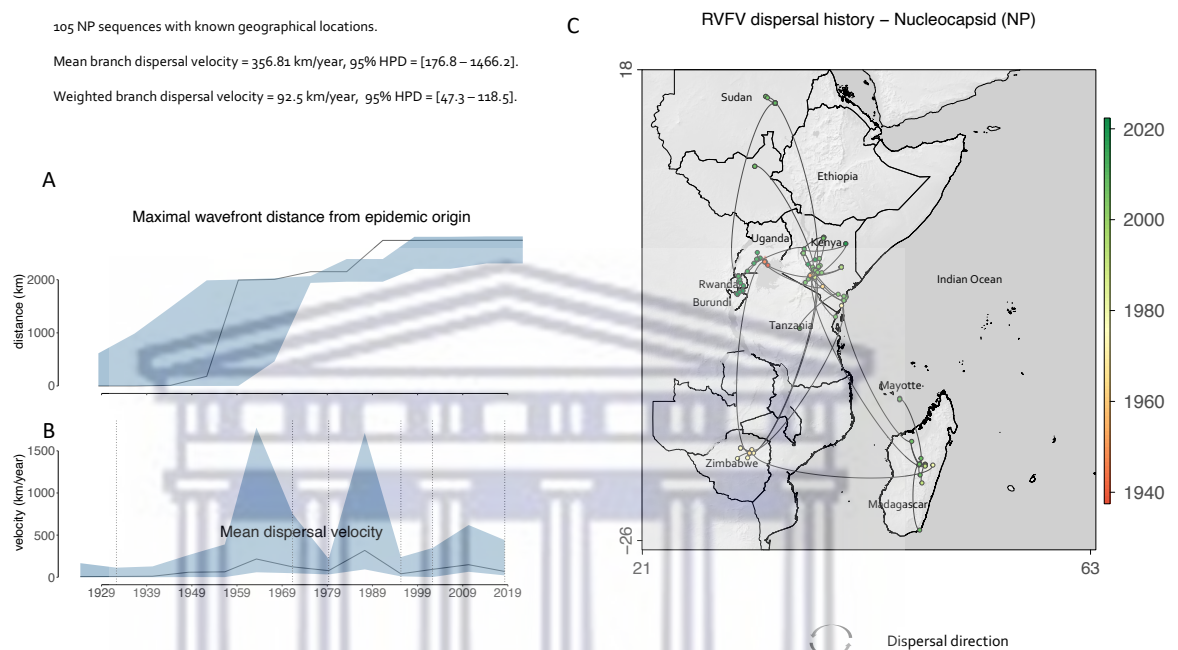
FIGURE 4.10: **Spatiotemporal diffusion and dispersal dynamics of RVFV in Eastern Africa using the nucleocapsid (NP) sequences.** RVFV dispersal statistics showing the evolution of the maximal wavefront distance from the epidemic origin (**A**) and mean dispersal velocity over the entire virus spread through time (**B**). The shaded blue area corresponds to the 95% credible regions of the estimated wavefront position. Dispersal history of RVFV obtained by continuous phylogeographic inference based a Maximum clade credibility (MCC) tree sampled on 1000 posterior trees (**C**) . Nodes of the trees are colored according to the time of occurrence, and oldest nodes are plotted on top of youngest nodes. The trees are superimposed on maps displaying features such as the ocean and lakes.
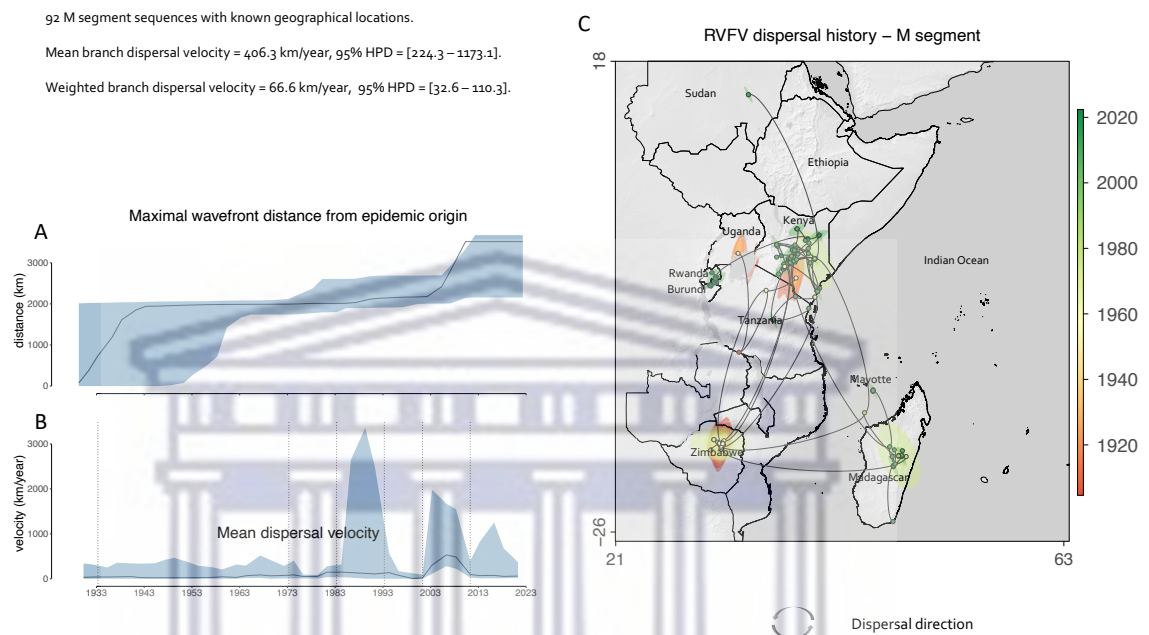
FIGURE 4.11: **Spatiotemporal diffusion and dispersal dynamics of RVFV in Eastern Africa using the complete M segment sequences.** RVFV dispersal statistics showing the evolution of the maximal wavefront distance from the epidemic origin (**A**) and mean dispersal velocity over the entire virus spread through time (**B**). The shaded blue area corresponds to the 95% credible regions of the estimated wavefront position. Dispersal history of RVFV obtained by continuous phylogeographic inference based a Maximum clade credibility (MCC) tree sampled on 1000 posterior trees (**C**) . Nodes of the trees are colored according to the time of occurrence, and oldest nodes are plotted on top of youngest nodes. The trees are superimposed on maps displaying features such as the ocean and lakes.

87 L-segment sequences with known geographical locations.

Mean branch dispersal velocity = 308.7 km/year, 95% HPD = [157.3 – 1144.7].

Weighted branch dispersal velocity = 50.3 km/year, 95% HPD = [28.9 – 82.5].

FIGURE 4.12: **Spatiotemporal diffusion and dispersal dynamics of RVFV in Eastern Africa using the complete L segment sequences.** RVFV dispersal statistics showing the evolution of the maximal wavefront distance from the epidemic origin (**A**) and mean dispersal velocity over the entire virus spread through time (**B**). The shaded blue area corresponds to the 95% credible regions of the estimated wavefront position. Right Panel Dispersal history of RVFV obtained by continuous phylogeographic inference based a Maximum clade credibility (MCC) tree sampled on 1000 posterior trees (**C**). Nodes of the trees are colored according to the time of occurrence, and oldest nodes are plotted on top of youngest nodes. The trees are superimposed on maps displaying features such as the ocean and lakes.

4.3.5 Population dynamics of RVFV

To estimate changes in the effective population size of RVFV over time, non-parametric skygrid reconstructions on NSs, NP, M and L segment sequence data were conducted in BEAST (Suchard et al., 2018). This flexible coalescent model allows accurate estimation of population trajectories. With convergence of the reconstruction, similar demographic trajectories in all four sequence datasets were observed when estimating the effective population size over time. Distinctively, there was an increase in the effective population size of the virus between 1930s to 1970s with a gradual drop in the 1980s followed by fluctuations in the late 1990s and 2000s (**Figure** 4.13, **Figure** 4.14). The period between 1930 and 1970 corresponds to the expansion phase of the epidemic. The peaks around 1997/1998, 2006/2007 are consistent with the major outbreaks of RVF experienced in the horn of Africa.
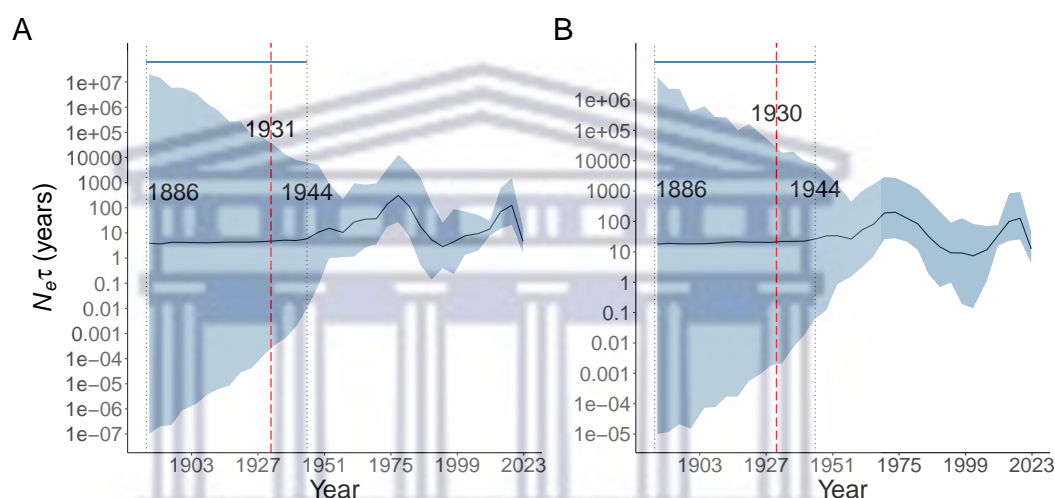
FIGURE 4.13: **Demographic history of RVFV in East Africa using nonstructural and nucleocapsid gene sequences.** Effective population size of RVFV (Nẽ) through time as inferred by skygrid coalescent prior model (Gill et al., 2013). Posterior mean estimates of the viral effective population size based on sequence data are represented by dark colored blue curves and the corresponding shaded light blue colored polygons reflect the 95% HPD region. The dashed grey vertical lines represent the 95% HPD time of the root of the tree and the upper highest posterior density. The dashed red line indicates an estimate of the time of the most recent ancestor (time of the root of the tree). The skygrid reconstructions are represented by **(A)** non-structural gene sequences, **(B)** nucleocapsid sequences.
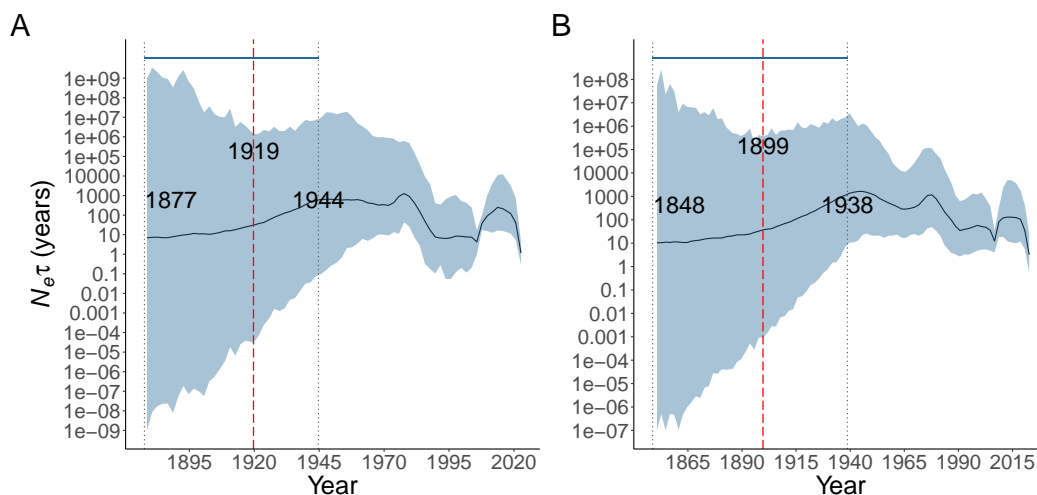
FIGURE 4.14: **Demographic history of RVFV in East Africa using using complete M and L segments sequences.** Effective population size of RVFV (Nẽ) through time as inferred by skygrid coalescent prior model (Gill et al., 2013). Posterior mean estimates of the viral effective population size based on sequence data are represented by dark colored blue curves and the corresponding shaded light blue colored polygons reflect the 95% HPD region. The dashed grey vertical lines represent the 95% HPD time of the root of the tree and the upper highest posterior density. The dashed red line indicates an estimate of the time of the most recent ancestor (time of the root of the tree). The skygrid reconstructions are represented by **(A)** complete M segment and **(B)** complete L segment sequences.

4.3.6 Environmental factors associated with the spread of RVFV

Assessment of the impact of environmental factors on the dispersal dynamics of RVFV in eastern Africa was also performed using seraphim (Dellicour, Rose, Faria, et al., 2016). Using a landscape phylogeographic framework, a formal test to ascertain if RVFV dispersal locations were correlated to specific environmental conditions was performed. These environmental values were extracted from geo-referenced grids or rasters (Hijmans & van Etten, 2012) that summarize distinct climatic factors considered for testing. Different trends in the association of RVFV dispersal locations and environmental factors were observed. While positive or moderate evidence (BF value 3-20) was shown for RVFV to disperse in areas with higher or lower elevation for the L and M segments, RVFV also displayed a tendency to remain in such areas. Overall, the statistical association tests showed that RVFV tended to disperse and remain in certain areas including those with higher or lower elevation, human population, croplands, urban areas but with moderate evidence (**Figure** 4.4).

Evaluation of the impact of environmental factors on RVFV velocity in eastern Africa using a different landscape phylogeographic framework for all the virus dispersal events was also conducted. To examine the correlations, scaled environmental distances for each environmental factor tested were computed using the least-cost path model as implemented in seraphim (Dellicour et al., 2017). The static of interest, Q, was determined to measure the correlation between RVFV durations and environmentally scaled distances. A simulation procedure to generate a null model of dispersal was also performed to allow for comparison between inferred and simulated Q distributions. This was performed to allow for approximation of Bayes factor support. The analysis revealed weak support for elevation as a resistance factor in the dispersal velocity of RVFV using all dispersal events in spatially explicit phylogeographic reconstruction using nucleocapsid sequences. Additionally, there was weak support for annual mean temperature as a conductance factor in RVFV dispersal events in eastern Africa. However, there was moderate support (BF value 3-20) for annual precipitation as a resistance factor in RVFV dispersal (**Table** 4.5).

TABLE 4.4: **Impact of various environmental factors on the dispersal locations of RVFV in Eastern Africa.** The table indicates the Bayes Factor (BF) support for the association between environmental values on geo-referenced grids and tree node locations. These results are computed on 1000 posterior trees obtained by continuous phylogeographic framework in BEAST. "C" indicates if the tested environmental variable was considered as a factor aiding/attracting while "R" corresponds to a factor that was considered as a resisting/repulsing RVFV dispersal. Applying (Kass & Raftery, 1995) guidelines, a BF value > 20 is considered strong evidence for an association between an environmental variable and dispersal location while a BF value of between 3 and 20 is a positive support for association between the environmental values and the dispersal locations.

Environmental Factor	L segment		M segment		NSs		NP	
	Disperse	Remain	Disperse	Remain	Disperse	Remain	Disperse	Remain
Elevation (C)	**4.2**	2.8	**3.2**	0.8	0.5	**7.5**	0.5	**10.1**
Elevation (R)	0.2	0.4	0.3	1.2	2.1	0.1	2.1	0.1
Temperature (C)	1.5	0.1	0.7	2.4	0.6	0.3	2.1	0.2
Temperature (R)	0.6	**7.0**	1.4	0.4	1.6	**3.8**	0.5	**5.5**
Precipitation (C)	2.2	3.9	3.7	0.7	1.8	0.1	1.8	0.3
Precipitation (R)	0.5	0.3	0.3	1.4	0.5	6.9	0.6	0.1
NDVI (C)	1.6	4.5	1.6	0.6	0.8	1.4	1.3	0.8
NDVI (R)	0.6	0.2	0.6	1.6	1.3	0.7	0.8	1.2
Human population (C)	**89.9**	54.6	**28.4**	0.8	2.5	7.6	2.0	**6.2**
Human population (R)	0.1	0.2	0	1.2	0.4	0.1	0.5	0.2
Croplands (C)	**5.5**	24.6	**7.2**	**13.3**	0.9	1.5	0.7	2.3
Croplands (R)	0.2	0.1	0.1	0.1	1.1	0.6	1.5	0.4
Pastures (C)	4.5	0.1	2.5	0.3	0.6	0.8	1.0	0.4
Pastures (R)	0.2	9.3	0.4	3.0	1.6	1.2	1.0	2.8
Urban areas (C)	0.8	**996**	0.9	0.2	0.3	**>99**	0.1	**>99**
Urban areas (R)	1.3	0	1.0	4.6	3.3	0	10.6	0
Primary forests (C)	0.5	**3.2**	0.6	2.9	0.1	**5.7**	0.1	2.6
Primary forests (R)	2.0	0.3	1.6	0.3	11	0.2	7.5	0.4
Secondary forests (C)	0.4	**5.5**	0.3	2.4	0.1	**25.3**	0.1	**12.5**
Secondary forests (R)	2.7	0.2	3.0	0.4	17	0	11.8	0.1
Primary non-forests (C)	2.2	0.6	0.7	0.6	8.4	0.8	**5.2**	0.4
Primary non-forests (R)	0.5	1.8	1.4	1.7	0.1	1.2	0.2	2.3
Secondary non-forests (C)	2.6	0.3	0.8	0.5	8.5	0.2	3.6	0.3
Secondary non-forests (R)	0.4	3.9	1.2	2.0	0.1	5.2	0.3	2.9

TABLE 4.5: **Impact of various environmental factors on dispersal velocity of RVFV in Eastern Africa.** The table shows Bayes Factor support for the association between environmental distances computed using the least-cost path model and dispersal durations. The environmental factors were independently considered as resistance or conductance and k is the rescaling factor used for transforming the initial raster values in the geo-referenced grids. Here, only positive evidence (BF value of 3-20) according to (Kass & Raftery, 1995), 1995 is highlighted.

Path model	Factor	k	L segment		M segment		NSs		NP	
			Resistance	Conductance	Resistance	Conductance	Resistance	Conductance	Resistance	Conductance
Least-cost	Elevation	10	-	-	-	-	-	-	2.2	-
Least-cost	Elevation	100	-	-	-	-	-	-	1.9	-
Least-cost	Elevation	1000	-	-	-	-	-	-	-	-
Least-cost	Temperature	10	-	2.4	-	2.4	-	-	-	2.5
Least-cost	Temperature	100	-	2.4	-	2.4	-	-	-	2.5
Least-cost	Temperature	1000	-	2.4	-	2.3	-	-	-	2.4
Least-cost	Precipitation	10	**8.9**	-	3	-	-	-	-	-
Least-cost	Precipitation	100	**9.5**	-	3.2	-	-	-	-	-
Least-cost	Precipitation	1000	**9.6**	-	3.3	-	-	-	-	-
Least-cost	NDVI	10	-	-	-	-	-	-	-	-
Least-cost	NDVI	100	-	-	-	-	-	-	-	-
Least-cost	NDVI	1000	-	-	-	-	-	-	-	-

4.4 Discussion

In this study, spatially phylogeographic and phylodynamic inference methods were applied to reconstruct the spatiotemporal diffusion and dispersal dynamics of RVFV in the eastern region of sub-Saharan Africa. RVFV is considered endemic in sub-Saharan Africa with Kenya, Zimbabwe and South Africa reported as countries with highest endemicity (Nielsen et al., 2020; Paweska & Jansen van Vuren, 2014). The virus has caused huge economic losses and increased mortality and morbidity both in livestock and humans in the Eastern Africa region (Bouloy & Weber, 2010; Pepin et al., 2010; Rich & Wanyoike, 2010). The virus circulates in other neighboring countries through movement of vectors, increased international trade in livestock and climate conditions (Paweska & Jansen van Vuren, 2014; Pepin et al., 2010). Although numerous evolutionary studies to ascertain transmission history of RVFV have been undertaken both at country and continent levels in sub-Saharan Africa (Samy et al., 2017; Soumaré et al., 2012), this study has applied phylodynamic modeling that incorporates genomic sequence and environmental data. Through this approach, quantification of the evolution and dispersal dynamics of the virus in eastern Africa, a region that has been experiencing numerous outbreaks of RVF since the early 1900s has been attained (Nanyingi et al., 2015; Paweska & Jansen van Vuren, 2014).

A substantial period of RVFV history is still very unclear, particularly in the 1970s when countries like South Africa and Zimbabwe were facing major outbreaks of RVF (Swanepoel, 2004). The evolutionary origins of lineage C, the most predominant one in sub-Saharan Africa has not yet been investigated. This lineage continues to ravage new territories in sub-Saharan Africa where RVF is endemic. Investigation on the emergence and geographic origins of lineage C was achieved by analyzing partial glycoprotein (Gn) gene sequence data. Molecular clock analysis estimated that the time of the most recent ancestor appeared at the same time in Kenya and Zimbabwe (1970.706, 95% HPD [1959-1975] in Kenya and 1970.406, 95% HPD [1961-1975] in Zimbabwe). However, there was an intermingling of sequences from both countries at the root of the MCC (Maximum Clade Credibility) trees, an indication that the epidemic origin can be from either of the two countries. The co-occurrence of sequences for two non-neighboring countries suggests that the virus must at least have circulated between intermediate countries, although no sampling is available to support this hypothesis. This highlights the general issue of sampling bias that in earlier years must have been a problem. During the 1970s, newer genotypes and lineages of the virus began to emerge, as shown by our Bayesian analyses. In this period, several lineages including A (DQ380209), G (DQ380220), J (DQ380222), K (DQ380194), L (DQ380188) were observed in Zimbabwe (Grobbelaar et al., 2011), all of which were introduced to new locations within the continent (Samy

et al., 2017; Soumaré et al., 2012). Discrete phylogeographic analysis incorporating statistically supported viral migration revealed exchanges between eastern Africa countries. With significant support values (Bayes factors (BF) and posterior probabilities (PP)) (Appendix C, **Table C.1** and **Table C.2**), the analysis showed that lineage C originated from Zimbabwe. Furthermore, phylogeographic reconstructions revealed that most exchanges to neighboring countries originated from Kenya and Zimbabwe (**Figure 4.3** and **Figure 4.4**).

Initially considered endemic to a few countries in Africa, RVFV has spread to new territories and caused deaths and economic losses (Pepin et al., 2010). The emergence of RVFV in Indian Ocean islands of Mayotte and Madagascar from mainland Africa is thought to have occurred by transport of live animals rather than enzootic maintenance (Carroll et al., 2011). Cross-border livestock routes and markets have been identified in Kenya, Somalia, Ethiopia, Djibouti, Sudan and Uganda (IGAD, 2012). Our analysis identified major towns which play important roles in the transit of livestock between and within countries (**Figure 4.4**). Garissa, a town located in northeastern Kenya acts as a hub for animals entering and leaving Kenya to Somalia (lower Juba). Strong Bayes factor values for movement of the virus from one location to the other was reported. For example, the movement of the virus from Garissa to other towns such as Kilifi was supported by strong support values (BF=59, PP=0.52), Baringo ($BF > 1000$, PP=1), Wajir to Marsabit (BF=62, PP=0.53) **(Appendix C, Table C.2)**. In search of pasture and water, pastoralist communities traverse great distances to new territories (Turner & Schlecht, 2019). The occurrence of RVFV in perceived pastoralist areas in Kenya highlight the significance of livestock mobility in spread of RVFV. The discrete trait phylogeography inference also highlighted the spread of the virus. This was observed in Kenya, Madagascar and Zimbabwe. It has been postulated that once established in a new environment, recurrent epizootics/epidemics occur as silent transmission leading to its endemicity (Linthicum et al., 1985; Linthicum et al., 2016). Outbreaks are often observed once favorable environmental conditions (increased or abnormal rainfall)are met (LaBeaud et al., 2008; LaBeaud, Muiruri, et al., 2011; Linthicum et al., 1985).

Time-scaled phylogenetic analysis showed that the 2010 outbreak in South Africa formed a larger clade classified as lineage H. Unlike other RVFV lineages, lineage C is undergoing evolution at a slightly different rate. Using all the RVFV segments, time-scaled phylogenetic analyses revealed that lineage C is undergoing expansion with two distinct sub-lineages that are named here as C.1 and C.2 (**Figure 4.6**, **Figure 4.7**, **Figure 4.8**). Similarly, this observation was highlighted in the discrete phylogeographic analysis using partial glycoprotein gene sequences. Overall, these analyses indicated that both sub-lineages arose from 1997/1998 outbreak isolates in eastern Africa as earlier reported by (Bird, Githinji, et al., 2008) Bird and colleagues (Bird, Githinji, et al., 2008). Within the C.2 sub-lineage, there was an occurrence of a distinct ladder-like monophyletic clade with

isolates from Mozambique (2014) forming the root. The diversity of these virus sequences was further assessed by examining single nucleotide polymorphisms in every position of the reference sequence. There was an accumulation of SNPs in lineage C (**Figure 4.5**) in comparison to other lineages. Since most sequences in eastern Africa belong to lineage C, a comparison of the substitution rates with the mean rates previously reported was performed. In a study utilizing 33 ecologically distinct RVFV genomes, Bird and colleagues (Bird, Khristova, et al., 2007) reported substitution rates for S, M and L segments as 2.35×10^{-4} [95% HPD: 1.28×10^{-4}-3.4×10^{-4}], 3.4×10^{-4} [95% HPD: 1.8×10^{-4}-3.0×10^{-4}] and 2.78×10^{-4} [95% HPD: 2.0×10^{-4}-3.5×10^{-4}] respectively. Although negligible, there was an increase in the substitution rates for L and M segments in comparison to previously estimated values (Bird, Khristova, et al., 2007). The M segment showed a higher mutation rate with 1.5×10^{-7} mutations compared to segment L which accumulated fewer mutations (5.0×10^{-8}). Overall, the virus displays a slow molecular clock as indicated by the few numbers of mutations and low substitution rates per site per year. In general, the virus accumulates less than one mutation per site per year. Arboviruses face evolutionary constraints because of the need to replicate in different hosts (mammalian and arthropod). This may partly explain the limited diversity observed in RVFV (Coffey et al., 2008). When compared to other bunyaviruses like the Crimean-Congo hemorrhagic fever virus (CCHFV), RVFV displays a narrow genetic diversity. The close clustering of RVFV isolates observed through phylogenetic analysis can be attributed to the recent emergence of the ancestral virus in Africa and its expansion among introduced colonial sheep and cattle (Pepin et al., 2010).

The transmission of RVFV is complex since its evolution is impacted at various levels, from molecular (reassortment) to geographical (climatic and environmental factors) and ecological (Pepin et al., 2010). It has been postulated that RVFV can maintain cryptic transmission at undetectable levels during interepidemic/epizootic periods (IEPs) (Linthicum et al., 1985). During low rainfall seasons, mosquito vectors can deposit their eggs in low-lying depressions. Once favorable environmental conditions are met, infected eggs hatch and initiate transmission to domestic ruminants. From time-scaled phylogenetics and discrete phylogeographic analyses, the 1997/1998 outbreak in the horn of Africa led to two separate sub-lineages that occurred at the same time (2006/2007). This clearly indicates that the virus was undergoing an evolutionary process in the period between 1997/1998 and 2006/2007, thus augmenting the hypothesis of cryptic transmission during IEPs. Since the 2006/2007 outbreak, the eastern Africa region has not had any major RVF outbreaks. However, local transmission beyond detection levels have been reported in various countries (Juma et al., 2022; OIE, 2022; UNOCHA, 2022). With continuous sampling following the 2006/2007 outbreak, a major finding from this study is the evolution of lineage C. Lineage C seems to undergo expansion resulting in sublineages in comparison to other lineages (Appendix C, **(Figure C.1)**). This observation's

implication is that sub-lineages of the virus can be seen after IEPs indicating persistence in endemic areas.

To describe the population characteristics and assess the impact of environmental factors on dispersal velocity and locations of RVFV, this study relied on phylodynamic and phylogeographic analytical frameworks. The epidemic/epizootic events of RVF in eastern Africa is a perfect example of viral invasion and emergence in new territories. Through phylodynamic reconstruction, this study demonstrated that RVFV spread in eastern Africa has undergone three main phases denoted as expansion, lag and maintenance (**Figure 4.13**) stages. The expansion phase was marked by an increase in the virus effective population size followed by a lag phase that was depicted by a decline in the virus population and finally the maintenance phase marked by IEPs. Phylodynamic analysis illustrated how intervention can be measured during or following outbreaks. The lag phase in virus spread between 1970s and 1980s is a period when countries were undertaking massive vaccinations in livestock to avert RVF spread (Swanepoel, 2004). Formal evaluation on intervention (i.e., vaccination) was performed by linking the observed demographic history of the virus with epidemiological data. Both the Smithburn (Live Smithburn neurotropic strain (SNS) obtained after intracranial passage in mouse, Source: virus isolated from mosquito in Uganda, Manufacturer: Onderstepoort Biological Products, Country: South Africa) and the inactivated RVFV vaccines (95EG vaccine, Source: RVFV SNS strain, Country: Egypt) were sold during the period of 1960s to 1980s to reduce RVF occurrence in livestock (Grobbelaar et al., 2011). During this period, the annual sales of the vaccines were higher in eastern and southern Africa countries (Swanepoel, 2004). The administration of the vaccines during this period led to a gradual decline in the effective population size of the virus as observed in the coalescent demographic estimates. Descriptive dispersal statistics provided summarized trends in RVFV spread in eastern Africa. From the analysis, it was observed that the virus has covered close to 3000 kilometers (about 1864.11 miles) from the epidemic origin with much distance covered in the expansion and epidemic/interepidemic periods. RVFV spreads on a moderate spatial scale with numerous long-distance dispersal events (**Figure 4.9**, **Figure 4.10**, **Figure 4.11**, **Figure 4.12**). Although mosquitoes have a short distance range of three kilometers (Verdonschot & Besse-Lototskaya, 2014), the large distances travelled by the virus from eastern Africa mainland to the Indian Ocean Islands of Madagascar and Mayotte may have occurred because of international trade in livestock as previously reported (Carroll et al., 2011). In eastern and southern Africa, large RVF outbreaks often occur at irregular intervals and are linked with heavy rainfall and floods (Linthicum et al., 2016; Murithi et al., 2011; Paweska & Jansen van Vuren, 2014). When these climate-driven outbreaks occur, there is an increase in the number of vectors. Not so much is known about the virus during interepizootic periods. However, it is thought that there is some level of cryptic maintenance through transovarial

transmission mostly in *Aedes* mosquitoes (Linthicum et al., 1985).

To test for association between environmental factors and dispersal locations of RVFV, there was strong evidence for the virus to disperse and remain in areas with high urbanization, human population, and crops. The rapid urbanization rates in low- and middle-income countries coupled with an increase in human population have created avenues for the emergence of infectious diseases (Baker et al., 2022; Rose et al., 2020). Urbanization has led to the surfacing and spread of arboviral diseases transmitted by *Aedes* mosquito species adapted to urban areas. Furthermore, there appears to be a correlation between preference for *Aedes aegypti* and human odour (Baker et al., 2022). Globally, there are over 65 mosquito species that are potential vectors of RVFV (Fontenille, 1998; Linthicum et al., 2016). These vectors have been identified in several countries including Kenya (Arum et al., 2015; Hanafi et al., 2010; Sang et al., 2017) and Madagascar (Balenghien et al., 2013; Tantely et al., 2015). The need to achieve food security at country and global level has led to agricultural intensification (Rohr et al., 2019). Agricultural intensification leads to a reduction in biodiversity and an increase in zoonotic diseases (Jones et al., 2013). Agricultural practices on natural ecosystems result in the expansion of ecological boundaries leading to species assemblage (Jones et al., 2013). When this phenomenon occurs, infectious pathogens can spillover to new areas and gain genetic diversity for their survival and adaptation (Grace, 2019). Farming practices in substantial portions of land may incorporate irrigation which offer prime environmental conditions for vector breeding as exemplified by the 2010 outbreak in the White Nile region in Sudan (Aradaib et al., 2013). On the association between environmental factors and dispersal velocity, there was weak evidence for annual mean precipitation as a factor impeding the spread of the virus. Overall, there were no strong environmental factors that impacted the spread of RVFV. Our ability to achieve consistent results on testing the association of environmental values and dispersal velocity and locations of RVFV were limited by sampling patterns in the four datasets. Additionally, the current framework does not allow testing the impact of time-series environmental variables on the dispersal velocity of viruses. Nonetheless, our findings suggest enhanced mosquito control programs should be undertaken to reduce RVF transmission particularly in urban and densely populated areas and agricultural lands.

4.5 Conclusion

Urbanization, climate change, increased trade in livestock and the ever-changing land-use patterns continue to increase the risk of RVF and many other zoonotic diseases. Therefore, considerable research on the ecology and spread of vector-borne diseases is urgently required in anticipation of potential future outbreaks. By placing RVFV genomic data in a spatiotemporal context, this study has investigated the emergence and epidemic origins

of the virus and its spread in the East African region through phylogeographic inference. With robust statistical procedures, the study has quantified viral exchanges between East African countries and locations and showed that lineage C originated from Zimbabwe. More significantly, this analysis has defined two sub-lineage clades within the lineage C, providing a finer genotyping tool for detailed epidemiological studies of the virus transmission dynamics and control interventions. Further, this study has linked phylogeographic analyses with epidemiological data to understand viral population changes and the impact of interventions such as vaccinations on RVFV control. By testing associations between environmental variables and RVFV dispersal locations, the study has shown that areas with high urban and population density as well as wetlands or farming can be targets for enhanced vector surveillance. The phylogeographic and phylodynamic approaches implemented in this study can be implemented in an outbreak setting when coupled with real-time virus genome sequencing to inform intervention strategies and surveillance decisions.

Chapter 5

General summary, conclusions and perspectives

5.1 General summary

RVF is a climate-driven disease which has shown its epidemic potential outside endemic settings. RVFV is an important emerging zoonotic pathogen of great interest to both public and livestock health in sub-Saharan Africa and the rest of the world. The World Health organization (WHO) has listed and prioritized RVF as a disease that may cause future pandemic (WHO, 2018). RVF poses a substantial threat to communities in Africa which depend on livestock for their livelihoods. When outbreaks of RVF occur, prompt banning on international livestock trade comes with heavy economic implications. In addition, massive abortion rates among pregnant animals lead to mortality in livestock. This negatively affects farmers who depend on meat and milk as both dietary and income sources. In public health, the disease may lead to hemorrhagic fever, blindness, and renal failure, all which can lead to death. Furthermore, there are no licensed vaccines for human treatment while the available livestock vaccines possess residual virulence and pathogenicity which can be passed to unborn animals leading to abnormal fetus formation.

The complex nature of RVF transmission in terms of epidemiology and ecology warrants an integrated approach that goes beyond conventional epidemiological methods such as contact-tracing or enumeration of cases. These conventional approaches have particularly been limited since they are laborious and time intensive in large outbreaks settings. Additionally, these methods may fail to capture the actual information in silent/cryptic RVF transmissions that occur during interepidemic periods. One of the significant achievements in augmenting classical epidemiological approaches is pathogen sequencing. Through sequencing of RVFV detected during outbreaks, genomic based surveillance can allow us to extract information regarding its transmission patterns and

112

evolution. With sampling at finer resolution, genomic sequencing can estimate epidemic characteristics such as basic reproductive number and transmission chains in a population. These procedures constitute genomic epidemiology, whose foundational tenet is the rapid accumulation of mutations in virus genomes. Genomic sequencing of pathogens holds much promise in the development of new control and therapeutics tools against RVF. By monitoring changes at the genomic level, differences between vaccine strains and clinical isolates may trigger the development of new vaccines and diagnostics based on observed mutations. Spatial epidemiology plays a critical role in understanding the dispersal history and transmission dynamics of pathogens. During outbreaks, pathogens get disseminated to new environments thereby allowing transmission to proceed. During such movements, pathogens accumulate mutations through adaptation mechanisms in new environments and hosts. The evolutionary and spatial dynamics of the pathogens therefore constitute an interdependent linkage. As such, sampling of pathogens at different time points and locations offers a common frame of reference where integration of virus genetic data, spatiotemporal information, epidemiological and environmental data can be linked to understand the origin, spread and transmission dynamics. Consequently, effective control of RVF through an integrated surveillance mechanism is paramount, especially in Africa where the disease has ravaged several countries including those that were previously considered RVF-free.

In this study, a lineage classification method that allow us to rapidly characterize and track circulating lineages was developed. We built a lineage classification tool that incorporates phylogenetic methods to determine lineage defining single nucleotide polymorphisms among RVFV strains by assessing diversity in partial and whole genetic sequences. The lineage assignment tool incorporated several classifiers using whole genome sequences (complete large, medium and small genomic segments) as well as one that uses partial glycoprotein gene sequences. Higher sensitivity, specificity and accuracy in the tool validation using a large dataset obtained from a public repository was achieved. Further evaluation and testing of the tool was performed using clinical samples obtained from an outbreak in Kenya. Similarly, high accuracy, sensitivity and specificity was observed among the tested outbreak samples. The tool is available both as a command line (https://github.com/ajodeh-juma/rvfvtyping) and a web application (http://krisp.ukzn.ac.za/app/typingtool/rvfv/) for use by those with programming or little (or no) programming skills.

Secondly, we developed a sequencing methodology that enriches RVFV genetic material from clinical outbreak samples. This method takes advantage of overlapping primers that can amplify the entire RVFV genome with increased genome coverage. We compared this method, the amplicon multiplex PCR enrichment (amPCRe), to other approaches

commonly used in sequencing. Specifically, we evaluated this method on the same clinical samples that were enriched through cell culture and non-enriched (directly sequenced). We showed that the amplicon multiplex PCR enrichment method is an alternative approach to generating RVFV viral genomes from samples with low viral starting material (low viral load). We further demonstrated that this method is affordable compared to metagenomics which is expensive since synthesis of the primers is cost effective. As opposed to metagenomics or unbiased sequencing of a sample, the amPCRe method is specific and sensitive given that it targets genetic material of a specific pathogen and eliminates host contamination that forms up to 99.5% of the clinical sample genetic material. Finally, we demonstrated that the amPCRe method is time sensitive, and less labor intensive as opposed to enrichment of the virus by cell culture (takes up to 14 days). We showed that the genome sequences generated by the amplicon PCR multiplex enrichment can be used for evolutionary studies to understand the evolution of RVFV. A major limitation with our amplicon multiplex PCR enrichment method is the need to regularly monitor diversity to identify mutations for the design of updated primer sets. However, given the low genetic diversity in RVFV, our method can be utilized to generate genomic data for all the RVFV lineages. The other drawback encountered in the method was the dropout in coverage in the untranslated regions and the intergenic region in the small segment of the RVF virus genome. We aim to optimize the primers, particularly in regions where there are dropouts to cover a larger breadth of the virus genome.

Finally, in chapter 4, phylodynamic and phylogeographic models were used to understand the transmission dynamics, spread and potential drivers of RVFV in East Africa. The study utilized four distinct datasets to arrive at conclusions that were robustly tested. By employing molecular clock analysis and discrete phylogeographic inference, our analyses robustly placed Zimbabwe as the geographic origin of lineage C, the most predominant lineage in sub-Sahara Africa. However, the placement of both Kenya and Zimbabwe virus isolates at the root of the time-scaled phylogenetic tree suggests that the virus may have been in circulation between intermediate countries given that Kenya and Zimbabwe are non-neighboring. This analysis also highlighted the biased or lack of sampling in the earlier years. Using phylodynamic analysis, the study determined a classification that splits lineage C into two sub-lineages . This result is critical for future genomic epidemiological studies of the virus by providing a finer genotyping tool.

The study also utilized continuous phylogeographic analysis to describe the transmission of RVFV in Eastern Africa region using spatiotemporal information. The study further characterized the dispersal dynamics to estimate the maximal wavefront distance from the epidemic origin and dispersal velocity. It was noted that RVFV has a

moderate dispersal spread compared to other zoonotic pathogens like Ebola. By examining the population dynamics and incorporating phylodynamic models, we observed the distinct phases of effective virus population over time that were in congruence with major epidemiological observations and interventions such as vaccinations. The final analyses incorporated environmental data (climate, land use and population) to test for association between RVFV dispersal direction and velocity. A significant association on urbanization, population density and croplands in dispersal direction was observed. This observation calls for heightened surveillance efforts in areas with high human population, urban areas and farmlands which are prime fields for vector breeding.

5.2 Implications and recommendations

RVF is a regional problem and collaboration between the affected nations is critical for surveillance, prevention, and intervention measures. Systematic surveillance in sentinel herds to assess RVF infections among susceptible animals is one element of prevention and control. Additionally, prompt reporting of confirmed cases and implementation of hygienic measures to curb spread constitute elements of prevention and control (Mariner, 2018). Both the public and livestock health service providers have the mandate to co-operate and share information for effective intervention measures. RVF is primarily transmitted by vectors, which can be controlled through spraying of insecticides especially following heavy rainfall. These control measures can be limited by factors such as costs and geographical extent or coverage of infection or risk. This thesis highlighted the need for genomic surveillance in contributing towards RVF control and intervention. Genomic epidemiology can augment the above intervention measures by monitoring genetic changes of the virus over space and time. When coupled with epidemiological data and other data sources, genomic surveillance can inform policies. For example, by studying the genetic variants in circulation, implementing institutions can be advised on which geographical areas they need to focus more on to cost effectively implement control strategies. Consequently, there is a need for scaling up of RVF sequencing efforts in sub-Saharan Africa. The data stored in genetic sequence information can allow us to compare and match vaccine strains with variants circulating within a population. With this insight, better and improved diagnostics and vaccines can be developed to fight the disease. Additionally, genomic data offers a layer of surveillance towards controlling RVF. When incorporated with other datasets, deeper insights can be unearthed regarding its transmission dynamics as shown in this work.

Although there are two universally applied lineage nomenclature, this study highlighted that some lineages are undergoing expansions and therefore a robust nomenclature that gives a finer resolution at sub-lineage level will be highly beneficial to the scientific

community. To achieve this, regional sequencing efforts can be attained by prompt reporting of outbreaks and continuous sampling to have a pool of genetic data that can be used to inform intervention measures.

In trying to achieve near real-time genomic sequencing which translates to real-time genomic surveillance, availability of methods that rapidly generate the required genetic data is critical. Our proposed method using amplicon sequencing is a momentous change that was inspired by its application in monitoring the global variants of SARS-CoV-2. This method offers a simple approach to timely and cost effectively generate viral genomic data in an outbreak situation.

The changing environmental conditions may interfere with species range and density, increasing risk of zoonotic emergence. Increased international trade in livestock, human travel and vector movement are avenues for transmission of RVFV. These spatial fingerprints coupled with mutations in the virus can offer spatially explicit information that can be integrated with genetic data to understand the spread of RVFV. In this project, phylogeographic models that incorporate various data types was used to describe transmission dynamics and drivers of RVFV. These models can be applied to the global dataset consisting of data from sub-Saharan Africa to have a broader overview of RVFV spread both locally and regionally. These models can also be tested to assess intervention measures such as vaccination.

Taken all together, the study has laid ground for genomic epidemiology of Rift Valley fever virus by providing tools and methods for sustained genomic surveillance. By examining the genetic differences between circulating RVFV strains, the study sets stage for additional research in vaccine development and diagnostics. Enhanced molecular surveillance of RVFV will therefore require scaling up of sequencing efforts during outbreaks and interepidemic periods to have a clear picture on the evolution of the virus. Furthermore, a clear sampling approach that does not introduce bias should be prioritized to ensure that genetic and spatiotemporal dynamics are well captured.

Appendix A

Genomic Surveillance of Rift Valley fever virus
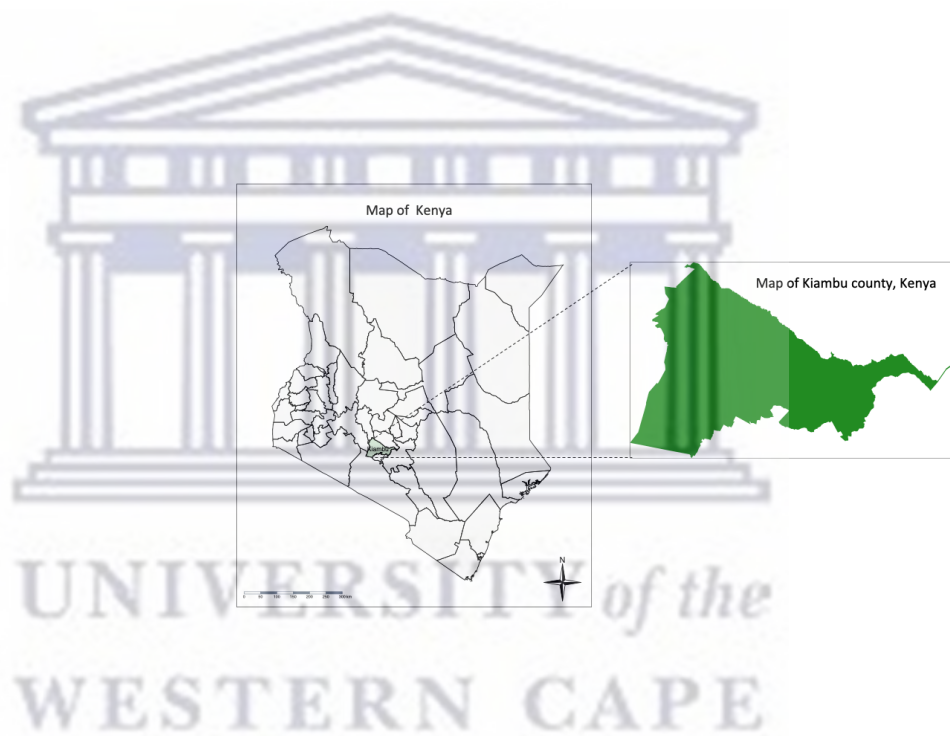
FIGURE A.1: **Sampling areas in Kenya where suspected cases of Rift Valley fever were reported.** A map of the sampling area in Kimabu county, central Kenya where livestock sera samples were collected.

117

TABLE A.1: **RVFV Lineage defining single nucleotide polymorphisms (SNPs) for complete S segment** For each lineage sequences, SNPs were identified in comparison to the reference (NC_014395).

Lineage	SNPs	Total
A	1233CT;1325AG	2
B	40TC;427TC;538TC;684TC;724CT;758AG;796GA;893CT;945GA;978GA;990TC;1089TC;1194AG;1233CT;1325AG;1451TG	16
C	101TA;109AT;427TC;478TC;487CT;534CT;538TC;544CT;586TC;616CT;625TC;631AG;684TC;758AG;945GA;990TC;996GA; 1089TC;1180CT;1194AG;1325AG;1451TG;1545TC	23
D	100CT;101TA;109AT;169CT;233TC;283GA;431TA;427TC;439CA;451TC;478TC;534CT;538TC;574GA;586TC;616CT;625TC; 631AG;684TC;724CT;758AG;784GA;796GA;884TC;893CT;945GA;978GA;990TC;996GA;1089TC;1128GA;1152AG;1180CT; 1194AG;1233CT;1299CT;1325AG;1449TC;1451TG;1545TC	40
E	101TA;109AT;115TC;283GA;427TC;478TC;487CT;534CT;538TC;544CT;586TC;616CT;619TA;625TC;631AG;684TC;724CT; 758AG;796GA;823GA;884TC;892TA;893CT;945GA;978GA;990TC;996GA;1089TC;1098AG;1180CT;1194AG;1233CT;1254GA; 1301AG;1325AG;1451TG	36
F	101TA;109AT;283GA;534CT;538TC;616CT;625TC;631AG;684TC;724CT;758AG;945GA;990TC;996GA;1089TC;1180CT;1194AG; 1233CT;1325AG;1451TG;1545TC	21
G	101TA;109AT;283GA;392CT;406CT;409AT;427TC;439CT;478TC;487CT;534CT;538TC;544CT;586TC;616CT;625TC;631AG; 673CG;675CG;684TC;724CT;758AG;796GA;823GA;884TC;890CT;893CT;945GA;952AT;966CT;978GA;990TC;996GA;1002CT; 1089AG;1098AG;1180CT;1194AG;1233CT;1325AG;1362GA;1428CT;1451TG;1545TC;1548AG	45
H	101TA;214CT;283GA;310AG;427TC;442TC;445AG;478TC;534CT;538TC;607CT;616CT;620CA;625TC;684TC;724CT;758AG; 760TC;787AT;796GA;814AG;863AG;893CT;942AG;945GA;978GA;990TC;996GA;1029TA;1089TC;1180CT;1185TC;1194AG; 1215GA;1233CT;1299CT;1325AG;1425TC;1451TG;1530TC;1539GA	41
I	4CT;101TA;109AT;115TC;208TC;214CT;229CT;283GA;310AG;394GA;427TC;478TC;526GA;534CT;538TC;544CT;586TC; 607CT;610AT;616CT;625TC;631AG;649TC;664TC;679CT;684TC;700CT;724CT;758AG;796GA;839AG;884TC;885NA;893CT; 902CA;936AC;945GA;960AC;966CT;990TC;996GA;1077AG;1089TC;1119GA;1158TC;1180CT;1194AG;1212GA;1233CT; 1254GA;1325AG;1329GA;1374TC;1451TG;1512AT;1545TC;1560CT	57
J	101TA;109AT;199AG;283GA;361AG;406CT;427TC;478TC;487CT;490CT;493TC;534CT;538TC;544CT;586TC;589CA;610AT; 616CT;625TC;630AG;631AG;684TC;700CT;724CT;758AG;796GA;864AG;884TC;885NA;887CT;893CT;902CA;936AC;945GA; 960AC;975AG;987AG;990TC;996GA;1014CT;1065CT;1089TC;1125GA;1158TC;1180CT;1194AG;1203GA;1233CT;1254GA; 1290CT;1325AG;1344TC;1350CT;1374TC;1380TC;1401AG;1451TG;1482AG;1542AG;1545TC;1560CT;1563TC	62
K	101TA;109AT;283GA;427TC;502GA;534CT;538TC;547CT;586TC;610AC;616CT;631AG;684TC;724CT;758AG;884TC;936AC; 945GA;978GA;990TC;996GA;1089TC;1158TC;1180CT;1194AG;1233CT;1254GA;1325AG;1374TC;1428CT;1451TG;1545TC	32
L	101TA;109AT;283GA;478TC;487CT;534CT;544CT;610AT;616CT;625TC;631AG;684TC;724CT;758AG;796GA;885NA;893CT; 902CA;936AC;945GA;990TC;996GA;1089TC;1158TC;1180CT;1194AG;1233CT;1254GA;1325AG;1374TC;1451TG;1545TC; 1560CT	33
M	101TA;109AT;283GA;310AG;319TC;385AG;391AC;427TC;478TC;487CT;502GA;518TC;520AG;534CT;538TC;544CT;547CT; 586TC;593TC;610AC;616CT;625TC;631AG;684TC;724CT;758AG;796GA;838AG;869GA;872CT;875AT;884TC;885NA;886NC; 893CT;902CA;936AC;945GA;960AC;972CT;978GA;990TC;996GA;1032AG;1083CT;1089TC;1158TC;1167CT;1180CT; 1233CT;1254GA;1301AG;1325AG;1329GA;1374TC;1386AG;1428CT;1451TG;1536CT;1545TC;1560CT	62
N	101TA;109AT;232AC;283GA;409AG;415TC;427TC;451TC;534CT;538TC;544CT;586TC;607CT;610AT;616CT;625TC;631AG; 684TC;721TC;724CT;758AG;787AT;796GA;884TC;885NA;893CT;902CA;916AG;936AC;945GA;960AC;990TC;996GA;1089TC; 1158TC;1180CT;1194AG;1233CT;1254GA;1325AG;1340GT;1374TC;1416TC;1451TG;1458GA;1545TC;1560CT;1596CT	48
O	42AT;85TC;101TA;109AT;226AG;283GA;382AG;427TC;478TC;487CT;534CT;538TC;544CT;586TC;610AT;616CT;625TC; 631AG;649TC;684TC;724CT;742CT;758AG;796GA;884TC;885NA;893CT;902CA;936AC;945GA;960AC;978GA;990TC; 996GA;999TC;1089TC;1152AG;1158TC;1180CT;1194AG;1233CT;1254GA;1325AG;1332AG;1374TC;1451TG;1542AG; 1545TC;1560CT	49

TABLE A.2: **RVFV Lineage defining single nucleotide polymorphisms (SNPs) for complete M segment** For each lineage sequences, SNPs were identified in comparison to the reference (NC_014396).

Lineage	SNPs	Total
A	715TA;1717AG	2
B	269TC;614AG;715TA;830GA;1103TC;1142TC;1526AT;1616TC;1717AG;1841TC;1946TC;2061TC;2064TC;2354CT;2471AG; 2682GA;3195TA;3368CT;3628AG;3704AG;3821AG	21
C	110CT;179GA;194TC;212AG;303GA;323CT;635GA;715TA;731CT;737CT;767CT;836TA;926GA;1103TC;1163CT;1190TC; 1241AG;1355AT;1526AT;1577CT;1580CT;1616TC;1787GA;1817TC;1841TC;1850AG;1911AG;1913AG;1931CT;2057GA; 2061TC;2064TC;2147TC;2303TC;2327TC;2330AG;2334TC;2354CT;2369CT;2441GA;2471AG;2735CT;2885AG;3020AG; 3053AG;3101AG;3122AG;3195TA;3236CT;3239TC;3254AG;3263CT;3347GA;3353GA;3368CT;3374TC;3628AG;3640TA	58
D	56GA;110CT;125TC;158AG;179GA;212AG;239AG;323CT;396AG;614AG;715TA;731CT;839TC;926GA;1103TC;1142TC;1163CT; 1195GA;1325AG;1337GT;1526AT;1577CT;1616TC;1717AG;1841TC;1850AG;1859CT;1892CT;1911AG;1913AG;1931CT; 1946CT;1997TC;2057GA;2061TC;2064TC;2135TC;2186AG;2285AG;2303TC;2327TC;2334CT;2354CT;2360CT;2363GA; 2366AG;2369CT;2443CT;2471AG;2918AG;3098CT;3195TA;3236CT;3263CT;3290GA;3353GA;3368CT;3512CT;3596CT; 3621AG;3628AG;3640TA;3676GA;3681GT;3691GA;3704AG;3720AT;3740AT;3813GA;3821AG;3824CT;3842CT	72
E	125TC;179GA;323CT;614AG;626TC;715TA;731CT;854TA;926GA;1103TC;1142TC;1151TC;1163CT;1166AG;1316CT;1526AT; 1616TC;1717AG;1793AG;1841TC;1850AG;1892CT;1911AG;1913AG;1946TC;2024TC;2036CT;2057GA;2061TC;2064TC; 2135TC;2147TC;2186AG;2214TC;2237AC;2285AG;2303TC;2334TC;2354CT;2369CT;2471AG;2486GA;2600TA;2643TC; 2805TC;2903TC;2921AG;3017TC;3195TA;3236CT;3263CT;3266GA;3269TC;3290GA;3296AG;3326TC;3368CT;3551TC; 3596CT;3628AG;3640TA;3641GA;3676GA;3704AG;3745TC;3824CT	66
F	110CT;125TC;179GA;212AG;323CT;497AG;614AG;659AG;662GA;677CT;715TA;731CT;816AG;902GA;926GA;1079GA;1103TC; 1106GA;1142TC;1163CT;1253GA;1319TC;1337GT;1454TC;1526AT;1577CT;1616TC;1649CT;1691TC;1717AG;1813TC; 1841TC;1850GA;1868AG;1874GA;1892CT;1911AG;1913AG;1931CT;1946TC;1985AG;2057GA;2061TC;2064TC;2135TC; 2186AG;2216GA;2249GA;2255AG;2285AG;2303TC;2327TC;2334TC;2363GA;2369CT;2471AG;2495TC; 2534TC;2599CG;2708CT;2822CT;2936CT;2981CA;3011AG;3152AG;3155GA;3195TA;3230TC;3236CT;3263CT;3290GA; 3353GA;3368CT;3596CT;3628AG;3640TA;3645TC;3650AG;3676GA;3704AG;3720AT;3721AG;3733AT;3740AT;3824CT	87
G	110CT;125TC;179GA;212AG;323CT;344CT;608AG;614AG;620TC;715TA;731CT;920AG;926GA;1085TC;1103TC;1142TC; 1163CT;1171CA;1173AT;1406GA;1490AG;1526AT;1616TC;1622GA;1715TC;1717AG;1841TC;1850AG;1892CT;1911AG; 1913AG;1918AG;1946CT;1967AG;2029CT;2030CT;2036CT;2057GA;2061TC;2064TC;2067CT;2126CT;2135TC;2147TC; 2186AG;2285AG;2303TC;2327TC;2334TC;2354CT;2369CT;2432GA;2471AG;2603AG;2855AG;2933TC;2945CA; 2993AT;3041AG;3195TA;3236CT;3263CT;3290GA;3344GA;3352AG;3359CT;3368CT;3513CT;3551TC;3596CT;3628AG; 3640TA;3676GA;3706TC;3725AC;3732AG;3821AG;3824CT	79
H	92CT;125TC;179GA;212AG;323CT;515AG;600TC;614AG;623CT;626TC;635GA;715TA;731CT;770TC;920AG;926GA;1103TC; 1142TC;1157AG;1163CT;1169AT;2303TC;3628AG	23
I	33TC;110CT;125TC;179GA;212AG;236GA;239AG;305TC;323CT;386TC;515GC;557CT;605TC;635GT;715TA;731CT;752CT; 794TC;806TC;833CT;920AG;980GA;986CT;998TC;1049GA;1103TC;1115GA;1142TC;1163CT;1304GA;1343GA;1349GA; 1427TA;1466TC;1481GA;1493TC;1520AC;1526AT;1553CT;1616TC;1691TC;1717AG;1790TC;1799TC;1803AG;1823CT; 1834GA;1841TC;1850AG;1853GA;1892CT;1895AT;1911AG;1913AG;1946TC;2001GA;2030CT;2036CT;2057GA; 2061TC;2064TC;2135TC;2141GA;2147TC;2162AG;2180AG;2186AG;2240TC;2255AG;2285AG;2303TC;2330AG;2334TC; 2354CT;2369CT;2468GA;2471AG;2489TC;2531GA;2555GA;2582CT;2603AG;2633AT;2678TG;2831GA;2834CT;2906AG; 3005AG;3008TC;3011AG;3089GA;3101AG;3195TA;3197TC;3213AG;3236CT;3263CT;3290GA;3296AG;3305GA;3359CT; 3368CT;3437TC;3548CT;3572CT;3596CT;3621AG;3628AG;3638GA;3640TA;3659CT;3664CG;3676GA;3679TC;3681GA; 3695AG;3697TC;3702GA;3704AG;3714CT;3725AT;3761TC;3797GA;3798CT;3815TC;3821AG;3824CT	128
J	33TC;110CT;125TC;131AC;158AG;179GA;212AG;236GA;239AG;305TC;323CT;350GA;386TC;557CT;605TC;614AG;620TC; 626TC;677CT;695TC;715TA;731CT;749AG;785GA;800CT;836TC;860CT;920AG;926GA;953AG;995GA;1007CA;1055TC; 1115GA;1142TC;1154GA;1160GA;1161TC;1163CT;1190TC;1250TC;1343GA;1361AG;1412GA;1478TC;1491TC;1526AT; 1577CT;1604GA;1610CT;1616TC;1652TC;1691TC;1717AG;1790TC;1793AG;1799TC;1803AG;1820AG;1823TC;1834GA; 1841TC;1850GA;1864GA;1880AG;1892CT;1895AT;1911AG;1943TC;1946CT;1949TA;1996TC;2027TC;2030CT;2042TC; 2057GA;2061TC;2064TC;2067TC;2084AG;2117CT;2120AG;2123CT;2135TC;2141GA;2147TC;2162AG;2180AG;2183CT; 2186AG;2222GA;2240TC;2255AG;2285AG;2303TC;2306TC;2327TC;2334TC;2369CT;2384TC;2385CT;2471AG; 2486GA;2534TC;2582CT;2732TA;2805TC;2825CT;2831GA;2834CT;2894CT;2906AG;2975GA;2987CT;3011AG;3041AG; 3101AG;3195TA;3212TC;3236CT;3239TC;3248GA;3263CT;3296AG;3342CT;3347GA;3401GA;3458TC;3494AG;3497TC; 3509TA;3548CT;3572CT;3596CT;3608CT;3621AG;3628AG;3638GA;3640TA;3658GA;3665TC;3669TC;3676GA;3691GA; 3697TC;3703GA;3704AG;3714CT;3792CA;3797GA;3819AG;3821AG;3824CT;3829TC;3840AG	156
K	33TC;104GA;110CT;125TC;179GA;197TC;236GA;488TC;605TC;626TC;715TA;894CT;1091TC;1103TC;1115GA;1142TC; 1250TC;1349GA;1364AC;1526AT;1571TC;1577CT;1616TC;1649CT;1691TC;1707TC;1717AG;1740CT;1790TC;1799TC; 1820AG;1823TC;1834GA;1841TC;1850GA;1895AT;1911AG;1913AG;1946TC;1949TC;1988TC;2036CT;2057GA;2061TC; 2064TC;2135TC;2141GA;2147TC;2162GA;2180GA;2186AG;2222GA;2240TC;2255AG;2285AG;2303TC;2327TC;2334TC; 2354CT;2369CT;2417GA;2471AG;2486GA;2579CT;2582CT;2810GA;2831GA;2834CT;2906AG;3011AG;3101AG;3143TC; 3161GA;3195TA;3236CT;3263CT;3290GA;3296AG;3299AG;3392GA;3467CT;3494AG;3506CA;3572CT;3605GA;3638GA; 3640TA;3651GA;3663CT;3676GA;3697TC;3704AG;3714CT;3727GA;3733AG;3797GA;3804AG;3824CT	100
L	48TA;89CT;125TC;179GA;197TC;212AG;221GA;236GA;242CT;314CT;323CT;386TC;401GA;488CT;581AT;605TC;626TC; 641TC;647TC;653TC;662GA;715TA;731CT;743AG;842GA;866CT;872TC;917CT;920AG;926GA;1061GA;1103TC;1115GA; 1122CT;1124AG;1142TC;1163CT;1190TC;1250TC;1274AT;1304GA;1349GA;1364AC;1385AG;1403AG; 1430CT;1511GT;1526AT;1556CT;1571TC;1577CT;1616TC;1649CT;1700CT;1703AG;1707TC;1717AG;1790TC;1796CT; 1799TC;1803AG;1805TC;1820AG;1823TC;1834GA;1835GA;1838GT;1841TC;1850GA;1892CT;1895AT;1898CT;1911AG; 1913AG;1946CT;1949TC;1955AG;1988TC;1992TC;1996TC;2030CT;2036CT;2057GA;2061TC;2064TC;2135TC;2141GA; 2147TC;2162AG;2180GA;2186AG;2222GA;2240TC;2252CT;2255AG;2285AG;2303TC;2334TC;2354CT;2369CT;2396GA; 2417GA;2450AG;2471AG;2486GA;2546TC;2570TC;2582CT;2609CT;2612AT;2643CT;2723TC;2831GA;2834CT;2882CT; 2906AG;2957TC;3011AG;3044AG;3056TA;3101AG;3143TC;3161GA;3195TA;3215GA;3236CT;3263CT;3290GA;3296AG; 3356GA;3368CT;3392GA;3494AG;3521TA;3599TC;3605GA;3628AG;3640TA;3651GA;3663CT;3676GA;3689AG;3697TC; 3704AG;3714CT;3721AT;3727GA;3733AG;3786CT;3797GA;3799TA;3804AG;3821AG;3824CT	156
M	33TC;54CT;110CT;125TC;149TG;161CT;179GA;182GA;236GA;323CT;386TC;488CT;500CT;557CT;626TC;659AG;671CT; 715TA;731CT;857GA;894CT;920AG;924TC;926GA;992GT;1034TC;1103TC;1115GA;1142TC;1151TC;1163CT;1195GA; 1214TC;1250TC;1304GA;1307CT;1322TC;1337GA;1343GA;1349GA;1364AC;1526AT;1571TC;1577CT;1616TC;1649CT; 1667CT;1707TC;1717AG;1790TC;1799TC;1803AG;1820AG;1823CT;1834GA;1841TC;1850GA;1871TA;1892CT;1895AT; 1911AG;1913AG;1946CT;1949TC;1996TC;2030CT;2036CT;2057GA;2061TC;2064TC;2082AG;2135TC;2141GA;2147TC; 2153CT;2162AG;2180GA;2186AG;2222GA;2240TC;2255AG;2285AG;2303TC;2315GA;2327TC;2334TC;2354CT;2369CT; 2417GA;2444CT;2459GA;2471AG;2486GA;2582CT;2609CT;2831GA;2834CT;2906AG;2963GA;3011AG;3050TC;3092AG; 3101AG;3143TC;3161GA;3191TC;3195TA;3236CT;3263CT;3290GA;3296AG;3299AG;3368CT;3437TC;3488TC;3494AG; 3506CT;3521TA;3548CT;3572CT;3605GA;3625GC;3628AG;3638GA;3640TA;3651GA;3654AG;3663CT;3676GA;3689AG; 3697TC;3704AG;3714CT;3720AC;3721AT;3727GA;3734TC;3754CT;3797GA;3804AG;3821AG;3824CT;3841GA	143
N	33TC;110CT;125TC;179GA;194TC;212AG;236GA;245GA;323CT;326TC;386TC;500CT;557CT;605TC;620TC;626TC;671CT; 704TC;715TA;734CT;794TC;821TC;920AG;926GA;1025AG;1103TC;1112GA;1115GA;1142TC;1163CT;1187GA;1250TC; 1304GA;1343GA;1364AC;1433GA;1463TC;1526AT;1535TA;1553CT;1577CT;1586TC;1616TC;1707TC;1717AG;1742GA; 1790TC;1799TC;1803AG;1823TC;1834GA;1841TC;1850GA;1874GA;1892CT;1895AT;1911AG;1913AG;1946CT;1982AG; 1996TC;2036CT;2061TC;2064TC;2123CA;2141GA;2147TC;2162AG;2180GA;2186AG;2213TC;2222GA;2240TC;2246CA; 2285AG;2294GA;2303TC;2315GA;2324CT;2327TC;2334TC;2354CT;2369CT;2444CT;2459GA;2471AG;2486GA;2495TC; 2519AG;2582CT;2624CT;2789TC;2831GA;2834CT;2865AC;2879TC;2903TC;2906AG;2966AT;2981CT;3011AG;3051CT; 3101AG;3126TC;3137CT;3161GA;3195TA;3236CT;3263CT;3290GA;3296AG;3494AG;3548CT;3572CT;3596CT;3614GA; 3620TC;3628AG;3640TA;3648AG;3676GA;3690GA;3692TC;3697TC;3704AG;3714CT;3753GA;3755TC;3756AG;3794AG; 3797GT;3821AG;3824CT	133
O	33TC;110CT;125TC;179GA;212AG;236GA;320TC;323CT;350GA;386TC;515GC;557CT;605TC;629CT;715TA;731CT;806TC; 920AG;926GA;1103TC;1106GA;1115GA;1142TC;1163CT;1205AG;1243AG;1250TC;1304GA;1343GA;1394GA;1415GA;1526AT; 1571CT;1580CA;1598TC;1616TC;1622GA;1691TC;1697GA;1717TC;1790TC;1799TC;1803AG;1814TC;1823TC; 1834GA;1841TC;1850AG;1892CT;1895AT;1911AG;1913AG;1946TC;1967AG;1996TC;2006TC;2030CT;2036CT;2057GA; 2061TC;2064TC;2093TC;2141GA;2147TC;2162AG;2180AG;2186AG;2240TC;2255AG;2285AG;2303TC;2327TC;2330AG; 2334TC;2369CT;2399CT;2453GA;2471AG;2486GA;2582CT;2603AG;2607AG;2615TC;2699CT;2831GA;2834CT; 2840CT;2849TC;2852TC;2906AG;2960TC;3008TC;3011AG;3101AG;3113CT;3134TC;3195TA;3197TC;3236CT;3263CT; 3290GT;3305GA;3368CT;3548CT;3572CT;3575CT;3596CT;3616TC;3628AG;3640TA;3666AC;3676GA;3692TC;3693CT; 3697TC;3703GA;3704AG;3714CT;3725AT;3797GA;3821AG;3824CT;3845GA;3846GA;3866AG	127

TABLE A.3: **RVFV Lineage defining single nucleotide polymorphisms (SNPs) for complete L segment** For each lineage sequences, SNPs were identified in comparison to the reference (NC_014397).

Lineage	SNPs	Total
A	822AC;1057CT;1427GA;3768TC;6303GA	5
B	86TA;208CT;822AC;851GA;1020TC;1057CT;1227CT;1353CT;1427GA;1605AG;1710TC;2005GA;2112TC;2178TC;2421TG;2487GA;2943GA;3072GA;3090CT;3210CT; 3348GA;3489TC;3642TA;3723AG;3768TC;3906CT;4108TC;5070CT;5079AG;5157CT;5175TC;5239TC;5406GA;5490TC;5968GA;6255AG;6303GA	37
C	198CT;207GA;366TC;519TC;544CT;744CT;753AG;822AC;851GA;855CT;876TC;879GA;881CT;922GA;1020TC;1026TC;1057CT;1065GA;1090TC;1119CT;1227CT;1290TC; 1427GA;1500TC;1509GA;1524CT;1533CT;1695TC;1830CT;1932TC;2005GA;2106TC;2208GA;2331GA;2406CT;2421TC;2424CT;2448CT;2520GA;2529CT;2571GA;2598TC; 2628GA;2637AG;2943CT;3042GA;3072GA;3108AG;3210CT;3283CT;3291TC;3318CT;3348GA;3426TC;3465TC;3642TA;3675AG;3768TC;3819GA;3835CT;3894AG;4015GA; 4108TC;4110GA;4272TC;4557TC;4584GA;4689AG;4812TC;4815AT;4884TC;4887TC;4953AG;4968AC;5061AT;5067CA;5070CT;5079AG;5157CT;5187GA;5217TC;5349AG; 5406GA;5433CT;5596CT;5664GA;5730CT;5784GA;5916AG;5934GA;5955CT;5970CT;5994CT;6051TC;6303GA	95
D	86TA;450CT;516TC;519TC;544CT;666GA;699CT;744CT;822AC;851GA;922GA;930AC;1017TC;1020TC;1057CT;1090TC;1119CT;1161AG;1227CT;1350TC;1353CT;1401GA; 1427GA;1497GA;1500TC;1533CT;1588CT;1605AG;1626CT;1695TC;1710TC;1794TC;2005GA;2025TC;2136CT;2211TA;2274CT;2280CT;2358CT;2385GA;2421TG;2424CT; 2463CT;2475TC;2520GA;2628AG;2637AG;2751CT;2805CT;2838AG;2853AT;2886TC;2943CT;3042GA;3072GA;3090CT;3210CT;3243AT;3250TC;3318CT;3348GA;3426TC; 3489TC;3627CT;3642TA;3648CT;3684GA;3723AG;3768TC;3831AG;3894AG;3903TG;3906CT;4015GA;4053CT;4077AG;4086AG;4108TC;4167GA;4272TC;4584GA;4689AG; 4815AT;4818AG;4872CT;5002AG;5070CT;5079AG;5157CT;5175TC;5181TC;5187GA;5215AG;5217TC;5253AT;5268CT;5334TC;5349AG;5406GA;5490TC;5604CT;5664GA; 5730CT;5766GA;5784GA;5796AG;5850AG;5853CT;5934GA;5968GA;5994CT;6117TC;6147GA;6231AG;6255AG;6303GA	116
E	54GA;86TA;132CT;243CT;447AG;453TC;516TC;519TC;544CT;744CT;822AC;851GA;879GA;881CT;922GA;1017TC;1020TC;1057CT;1090TC;1119CT;1188GA;1227CT;1350CT; 1353CT;1427GA;1500TC;1533CT;1572CT;1605AG;1626CT;1680TC;1710TC;1830CT;1911TC;2005GA;2106TC;2232CT;2358CT;2364AG;2406CT;2421TG;2424CT;2487GA; 2520GA;2569CT;2571GA;2628AG;2637AG;2658AG;2751CT;2895TC;2943CT;3042GA;3072GA;3090CT;3135TC;3210CT;3283CT;3318CT;3348GA;3426TC;3489TC;3642TA; 3660AG;3723AG;3768TC;3894AG;3906CT;3996GA;4015GA;4108TC;4207TC;4272TC;4551GA;4554CT;4584GA;4620TC;4689AG;4696TC;4725TC;4731CT;4815AT;4926GA; 4953AG;4968AG;5070CT;5079AG;5157CT;5187GA;5217TC;5253AT;5349AG;5406GA;5490TC;5596CT;5730CT;5784GA;5925AG;5934GA;5968GA;5970CT;5994CT;6240GA; 6255AG;6276GA;6303GA	106
F	86TA;340TC;516TC;519TC;544CT;744TC;822AC;851GA;922GA;1020TC;1057CT;1090TC;1119CT;1227CT;1350CT;1353CT;1427GA;1500TC;1533CT;1605AG;1614CT;1626CT; 2005GA;2358CT;2424CT;2487GA;2628AG;2637AG;2751CT;2943CT;3042GA;3072GA;3090CT;3210CT;3348GA;3426TC;3489TC;3642TA;3723AG;3768TC;3894AG;3906CT; 4108TC;4272TC;4584GA;4689AG;4815AT;5070CT;5079AG;5157CT;5187GA;5217TC;5253AT;5349AG;5406GA;5490TC;5730CT;5784GA;5934GA;5940CT;5994CT;6117TC;6303GA	62
G	54GA;86TA;285GA;357CA;447AG;468TC;516TC;519TC;544CT;585GA;658CT;744CT;786TC;822AC;851GA;879GA;881CT;922GA;1017TC;1020TC;1057CT;1084TA;1090TC;1119CT; 1128GA;1227CT;1249AG;1350CT;1353CT;1427GA;1458GA;1488GA;1500TC;1503CT;1533CT;1551TA;1572CT;1605AG;1626CT;1701TC;1704GA;1710TC;1740TC;1767CT;1770CT; 1830CT;1929GA;2005GA;2028TA;2106TC;2232CT;2268AG;2271AT;2358CT;2406CT;2421TG;2424CT;2487GA;2520GA;2569TC;2571GA;2628AG;2637AG;2643CT;2703TC;2751CT; 2943CT;3042GA;3072GA;3090CT;3210CT;3291TC;3318CT;3348GA;3426TC;3489TC;3516GA;3642TA;3681CT;3723AG;3744TC;3768TC;3894AG;3895TC;3906CT;4015GA;4053CT; 4108TC;4125CT;4272TC;4359TC;4444TC;4551GA;4584GA;4689AG;4731CT;4812TC;4815AT;4833GA;4839TC;4953AG;4968AG;5025GA;5070CT;5079AG;5157CT;5181TC;5187GA; 5217TC;5253AT;5307CT;5335AG;5349AG;5406GA;5490TC;5517GA;5596TC;5628TC;5661GA;5664GA;5730CT;5766GA;5784GA;5904TC;5913GA;5934GA;5940CT;5968GA;5994CT; 6255AG;6303GA;6304TC;6332TC	133
H	86TA;102AG;147AG;150TC;156AT;377TC;465TC;516TC;519TC;540TC;544CT;549GT;637CT;744CT;780TC;798CT;822AC;827AG;846CT;851GA;897AG;900CT;912TC;924AG;981CT; 1020TC;1047TC;1057CT;1119CT;1185AG;1188GA;1227CT;1290TC;1353CT;1374AG;1427GA;1456TC;1496GA;1605AG;1620TC;1638AG;1647AG;1671AG;1671CA;1710TC;1788GA;1863CT; 1971AG;1986CT;2005GA;2019AG;2193TC;2217CT;2247GA;2340TC;2358CT;2364AG;2421TG;2424CT;2487GA;2520GA;2526CT;2550TC;2568GA;2637AG;2700AG;2820AG;2907TC; 2943CT;2961GA;2979AT;3018TC;3042GA;3072GA;3090CT;3210CT;3231CT;3250TC;3297CT;3318CT;3348GA;3426TC;3468CT;3489CT;3513GA;3642TA;3675AG;3723AG;3732AG; 3768TC;3816AG;3891CT;3894AG;3906CT;4108TC;4305AG;4335AT;4374GA;4470TC;4530CT;4575AG;4662TC;4689AG;4722AG;4755TC;4773AG;4812TA;4833GA; 4863CT;4911CT;5016AG;5070CT;5079AG;5082CT;5139CT;5154AG;5157CT;5181TC;5187GA;5253AT;5296GA;5304TC;5406GA;5442AG;5490TC;5538AG;5541GA;5572CT;5730CT; 5795GA;5856AG;5859TA;5914TC;5934GA;5967TC;5968GA;5994CT;6018TC;6048TC;6075TC;6210AG;6255GA;6303GA;6331AG	145
I	86TA;114TC;129TC;174TC;231GA;237TC;360TC;396CT;483TC;519TC;534AT;544CT;550CT;693TC;744CT;774CT;777CT; 786TC;801GA;822AC;851GA;879GA;881CT;922GA;957CT;966GA;997CT;1020TC;1044CT;1057CT;1067GA;1090TC;1098GA;1107CT;1119CT;1143CT;1179GA;1182AG;1227CT; 1230GA;1238GA;1278TC;1299CT;1344CT;1350CT;1353CT;1407AT;1427GA;1470GA;1488GA;1500TC;1533CT;1563GA;1572CT;1602TA;1605AG;1626CT;1668AG;1710TC; 1737TC;1764GA;1791AG;1812TC;1815GA;1821CT;1827TA;1830CT;1851AG;1929GA;2005GA;2019AG;2070AG;2142GA;2145CT;2176GA;2187CT;2199GA;2232CT;2322TA; 2358CT;2406CT;2421TG;2424CT;2487GA;2520GA;2562AG;2569TC;2571GA;2628AG;2667AG;2724TC;2751CT;2826AG;2832TC;2835AG;2862CT;2910AG;2937CT;2943CT; 2955TC;2958GA;2979AG;3036CT;3042GA;3072GA;3087AG;3090CT;3210CT;3237AG;3279AG;3283CT;3285GA;3304TC;3306AG;3318CT;3348GA;3426TC;3453TC;3465TC; 3489TC;3591AG;3615TC;3642TA;3652CT;3699CT;3723AG;3768TC;3888GA;3894AG;3906CT;4014CT;4015GA;4108TC;4204TC;4206GA;4272TC;4323GA;4341GA;4392AC; 4444TC;4527CT;4563TC;4584GA;4644CT;4659GA;4689AG;4704CT;4731CT;4794AG;4815AT;4827GA;4839TC;4905TC;4938TC;4953AG;4968AG;4977CT;5070CT;5079AG; 5097AC;5157CT;5172AG;5187GA;5217TC;5253AT;5349AG;5406GA;5409GA;5433CT;5472TC;5490TC;5517GA;5526GA;5578CT;5596CT;5607TA;5664GA;5715AG;5730CT; 5784GA;5800AC;5841AG;5916AG;5934GA;5967TC;5968GA;5970CT;5994CT;6024TA;6087TC;6093AG;6115GA;6126GA;6189TC;6231AG;6255AG;6258AG	197
J	42GA;81GA;86TA;114TC;129TC;150TC;174TC;210AG;237TC;340TC;360TC;411AG;501TC;516TC;519TC;525GA;534AT;544CT;600TG;619AC;630CT;684AG;720CT;744CT; 774CT;777CT;780TC;783CT;786TC;822AC;834AG;851GA;879GA;922GA;927AG;948AG;990TC;1053GA;1056AG;1057CT;1067GA;1090TC;1098GA;1119CT;1179GA;1182AG; 1227CT;1230AG;1238GA;1299CT;1329TC;1344CT;1350CT;1353CT;1427GA;1455AG;1470GA;1488GA;1500TC;1524CT;1572CT;1584TA;1602TA;1605AG;1623CT;1626CT; 1659TA;1661TA;1665GA;1668AG;1674GA;1704GA;1707TC;1710TC;1764GA;1812TC;1815GA;1821CT;1830CT;1839CT;2005GA;2019AG;2137TC;2160TA;2187CT;2193TC; 2232CT;2262TC;2268AG;2358CT;2370GA;2406CT;2418AG;2421TG;2424CT;2487GA;2493CT;2520GA;2535TC;2544GA;2568GA;2571GA;2598TC;2604GA;2616CT;2619TC; 2628GA;2637AG;2652TC;2724TC;2733GA;2751CT;2808GA;2835AG;2862CT;2886TC;2907TC;2910AG;2915AG;2943CT;3042GA;3072GA;3073CT;3090CT;3124TC;3168TC; 3192CT;3210CT;3234TC;3237AG;3279GA;3283CT;3318CT;3348GA;3489TC;3513GA;3642TA;3699CT;3723AG;3732AT;3768TC; 3798AG;3888GA;3894AG;3895TC;3906CT;3936AG;3993TG;3996GA;4014TC;4015GA;4038CT;4092AG;4095GA;4101TC;4108TC;4272TC;4296AG;4323GA;4380AG;4389TC; 4425GA;4428AG;4444TC;4482TC;4527CT;4557TA;4581CT;4584GA;4602GA;4623GC;4659GA;4689AG;4696TC;4710AG;4731CT;4746TC;4794AG;4815AT;4839TC;4905TC; 4938TC;4953AG;4968AG;4977CT;5043GA;5070CT;5079AG;5133GA;5157CT;5172AG;5187GA;5217TC;5253AT;5285GA;5346GA;5349AG;5361CT;5406GA;5409GA;5418TC; 5433CT;5490TC;5517GA;5596CT;5628TC;5637TC;5643GA;5664GA;5673CT;5715AG;5730CT;5784GA;5877CT;5934GA;5967TC;5968GA;5970CT;5980TA;6033TC;6060CT; 6063CT;6075TC;6159AG;6231AG;6255AG;6279GA;6303GA;6327AG	234
K	86TA;129TC;174TC;243CT;360TC;516TC;519TC;544CT;588CT;606GA;729TC;744TC;774CT;777CT;786TC;822AC;851GA;879GA;906AG;922TC;933CT;948AG;1041GA;1090TC; 1098GA;1119CT;1179GA;1230AG;1234AG;1238GA;1344CT;1350CT;1353CT;1416GA;1427GA;1452TC;1470GA;1488GA;1500TC;1533CT;1569CT;1572CT;1573TC;1605AG;1626CT; 1668AG;1671CT;1710TC;1716CT;1764GA;1812TC;1815GA;1821CT;1830CT;1932TC;2005GA;2019AG;2106TC;2109GA;2187CT;2232CT;2262TC;2358CT;2388AT;2406CT;2424CT; 2508TC;2520GA;2554TC;2569TC;2571GA;2622GA;2628AG;2637AG;2658AG;2664GA;2838GA;2862CT;2880TA;2910AG;2937CT;2943CT;3042GA;3072GA;3090CT;3210CT;3283CT; 3318CT;3369TA;3426TC;3453TC;3465TC;3489TC;3591AG;3621AT;3642TA;3684GA;3699CT;3738CT;3768TC;3894AG;3960TC;4023CT;4272TC;4332AG;4389TC;4419TC;4444TC; 4455TC;4500AT;4584GA;4689AG;4731CT;4794AG;4815AT;4839TC;4938TC;4953AG;4968AG;4977CT;5070CT;5079AG;5091AG;5142AG;5187GA;5217TC;5253AT;5349AG;5406GA; 5478TC;5490TC;5509CT;5517GA;5580TC;5596CT;5664GA;5730CT;5784GA;5967TC;5968GA;5970CT;6033TC;6144TC;6159AG;6231AG;6255AG;6303GA	149
L	86TA;114TC;129TC;174TC;360TC;516TC;519TC;534AT;544CT;744CT;774CT;777CT;822AC;879GA;881CT;922GA;1020TC;1057CT;1090TC;1098GA;1107CT;1119CT;1179GA; 1182AG;1227CT;1238GA;1344CT;1350CT;1353CT;1427GA;1470GA;1488GA;1500TC;1533CT;1572CT;1602TA;1605AG;1626CT;1668AG;1674GA;1812TC;1815GA;1821CT; 2005GA;2106TC;2187CT;2232CT;2358CT;2406CT;2424CT;2487GA;2532GA;2571GA;2628AG;2637AG;2724TC;2751CT;2862CT;2910AG;2937CT;2943CT;3042GA;3072GA; 3090CT;3237AG;3279AG;3283CT;3318CT;3348GA;3426TC;3453TC;3465TC;3489TC;3501CA;3642TA;3699CT;3723AG;3768TC;3894AG;3906CT;4108TC;4272TC;4444TC; 4527CT;4584GA;4689AG;4731CT;4794AG;4815AT;4839TC;4905TC;4938TC;4953AG;4968AG;5070CT;5079AG;5172AG;5187GA;5217TC;5253AT;5349AG;5406GA;5409GA; 5490TC;5517GA;5596CT;5730CT;5784GA;5934GA;5967TC;5970CT;5994CT;6159AG;6231AG;6255AG	115
M	86TA;114TC;129TC;174TC;243CT;270CT;360TC;366TC;516TC;519TC;534AT;544CT;582CT;588CT;606GA;621GA;624CT;658CT;729TC;744CT;774CT;777CT;786TC;822AC; 823GA;851GA;879GA;881CT;906AG;922GA;933CT;948AG;1020TC;1054GA;1057CT;1090TC;1098GA;1107CT;1119CT;1179GA;1182AG;1227CT;1230AG;1234AG;1238GA; 1299CT;1329TC;1344CT;1350CT;1353CT;1427GA;1452TC;1467GA;1470GA;1488GA;1500TC;1533CT;1539AT;1569CT;1572CT;1602TA;1605AG;1626CT;1668AG; 1695TC;1710TC;1716CT;1764GA;1812TC;1815GA;1821CT;1830CT;1929GA;2005GA;2019AG;2079AG;2106TC;2109GA;2142AG;2187CT;2232CT;2262TC;2358CT;2388AT; 2400GA;2406CT;2421TG;2424CT;2487GA;2508TC;2520GA;2554TC;2569TC;2571GA;2622GA;2628AG;2637AG;2658AG;2706CT;2724TC;2739GA;2751CT;2838GA;2858GA; 2862CT;2910AG;2915AG;2937CT;2943CT;2944CA;2979AG;2994AG;3042GA;3072GA;3090CT;3138GA;3168TC;3210CT;3237AG;3279AG;3283CT;3318CT;3348GA;3369TA; 3426TC;3432AG;3453TC;3465TC;3489TC;3537CT;3579GA;3585TC;3591AG;3621AT;3642TA;3699CT;3714TC;3723AG;3738CT;3768TC;3810AG;3894AG;3900AG;3903TG; 3906CT;3924AG;3960TC;4014TC;4015GA;4023CT;4108TC;4272TC;4287CT;4323GA;4332AG;4362TC;4389TC;4419TC;4444TC;4455TC;4500AT;4527CT;4584GA;4689AG; 4698GA;4722AG;4731CT;4737GA;4755TC;4794AG;4797CT;4815AT;4839TC;4905TC;4938TC;4944GA;4953AG;4968AG;4977CT;5025GC;5070CT;5079AG;5142AG;5157CT; 5172AG;5187GA;5217TC;5253AT;5280CT;5349AG;5361CT;5406GA;5409GA;5414GA;5433CT;5478TC;5490TC;5509CT;5517GA;5580TC;5596CT;5628TC;5634CT;5664GA; 5730CT;5784GA;5787AG;5923CT;5934GA;5964TA;5967TC;5968GA;5970CT;5994CT;6033TC;6054CT;6144TC;6159AG;6231AG;6255AG;6303GA;6312GA;6341AG	224
N	60TC;86TA;114TC;129TC;174TC;216TC;327CT;330GT;360TC;516TC;519TC;522TC;534AT;543TC;544CT;549GA;570GA;738AT;744TC;765GA;774CT;777CT;786TC;822AC; 879GA;906AG;912TC;922GA;933CT;948AG;969CT;972CT;1017TC;1020TC;1025TC;1057CT;1090TC;1098GA;1107CT;1119CT;1179GA;1182AG;1185AG;1224CT;1227CT; 1230AG;1238GA;1257CT;1272AG;1290TC;1299CT;1338TC;1344CT;1350CT;1353CT;1427GA;1434CT;1458GA;1470GA;1473TC;1488GA;1500TC;1533CT;1572CT;1602TA; 1605AG;1608AG;1626CT;1677CT;1710TC;1764GA;1812TC;1821CT;1830CT;1845CT;1866CT;1875CT;1932TA;1953TC;1990TC;2005GA;2070AG;2106TC;2115CT;2232CT; 2262TC;2307TC;2334CT;2358CT;2406CT;2421TG;2424CT;2463CT;2475TC;2487GA;2493CT;2520GA;2568GA;2569TC;2571GA;2580TC;2589AG;2598TC;2628AG;2637AG; 2724TC;2751CT;2862CT;2910AG;2937CT;2943CT;3042GA;3063AG;3072GA;3144GA;3210CT;3237AG;3279AG;3283CT;3318CT;3336CT;3354CT;3390GA;3426TC; 3453TC;3465TC;3489TC;3597TC;3621AG;3627CT;3630GA;3642TA;3699CT;3723AG;3753TC;3768TC;3786GA;3813CT;3814TC;3894AG;3906CT;4011AG;4108TC;4182TC; 4207TC;4230TC;4257TC;4272TC;4323GA;4353GA;4362TC;4389TC;4444TC;4446AG;4509CT;4527CT;4584GA;4647AG;4671AG;4689AG;4704CT;4731CT;4794AG;4815AT; 4902CT;4905TC;4938TC;4953AG;4968AG;4977CT;5013CT;5070CT;5172AG;5187GA;5253AT;5335AG;5391CT;5406GA;5409GA;5487TC;5490TC;5511GA;5517GA;5565CT; 5580TC;5583GA;5596CT;5664GA;5679AG;5730CT;5754TC;5784GA;5793CT;5850AG;5883CT;5892CT;5904TC;5934GA;5967TC;5968GA;5970CT;5982GA;5994CT;6033TC; 6075TC;6087TC;6099TC;6159AG;6252GA;6255AG;6315AT	212
O	86TA;114TC;129TC;153CT;162CT;174TC;360TC;363AG;387CT;420AG;516TC;519TC;534AT;544CT;546GA;550CT;558CT;744CT;749GA;765GA;774CT;777CT;786TC;822AC; 851GA;864TC;879GA;881CT;922GA;1020TC;1057CT;1090TC;1098GA;1107CT;1119CT;1143CT;1179GA;1182AG;1227CT;1230AG;1238GA;1295CG;1299CT;1344CT;1350CT; 1353CT;1427GA;1470GA;1488GA;1500TC;1533CT;1542TC;1548AT;1572CT;1602TA;1605AG;1626CT;1647AG;1668AG;1710TC;1764GA;1806TC;1812TC;1815GA;1821CT; 1830CT;1923GA;2005GA;2019AG;2070AG;2106TC;2142AG;2157AG;2187CT;2226CT;2232CT;2262TC;2358CT;2406CT;2421TG;2424CT;2454GA;2487GA;2499TC;2532GA; 2554TC;2556GA;2569TC;2571GA;2628AG;2637AG;2646AG;2724TC;2751CT;2826AG;2835AG;2862CT;2886TC;2910AG;2937CT;2943CT;2979AG;3000GA;3042GA;3060TC; 3072GA;3090CT;3237AG;3279AG;3283CT;3291TC;3304TC;3318CT;3348GA;3426TC;3453TC;3465TC;3489TC;3501CA;3642TA;3699CT;3723AG;3768TC;3867CT;3879TC; 3894AG;3897GA;3906CT;3994AC;4014TC;4015GA;4053CT;4077AG;4108TC;4143TC;4173CT;4185CT;4272TC;4296AG;4447CT;4584GA;4689AG;4827CT;4584GA;4689AG; 4794AG;4815AT;4839TC;4896AG;4905TC;4938TC;4953AG;4968AG;5014TC;5070CT;5079AG;5097AC;5157CT;5172AG;5187GA;5189AG;5217TC;5253AT;5328CT;5349AG; 5406GA;5409GA;5472TC;5490TC;5517GA;5557TC;5562CT;5596CT;5664GA;5730CT;5784GA;5800AC;5934GA;5967TC;5968GA;5970CT;5994CT;6159AG;6231AG;6237TC; 6249TC;6255AG;6258AG;6303GA;6326TC;6347GA	191

TABLE A.4: **IgM antibody capture ELISA results** Optical Density values with dilution buffer (OD1) and with Nucleoprotein added to the dilution buffer (OD2) measured at 450nm.

Tested sample	OD1	OD2	Final OD	Sample positivity	Interpretation
DVS-333	0.057	0.783	0.726	87.05035971	Positive
DVS-372	0.062	0.804	0.742	88.96882494	Positive
DVS-356	0.079	1.23	1.151	138.0095923	Positive
DVS-230	0.121	1.228	1.107	132.7338129	Positive
DVS-321	0.067	1.277	1.21	145.0839329	Positive

TABLE A.5: **RVFV representative sequence data generators** Metadata of the data generators of RVFV representative sequence in building lineage assignment tool for Gn classifier.

Accession	Lineage	source	Strain	Country	Year	Authors
NC_014396	A	Human	ZH-548	Egypt	1977	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol, S.T.
HM587040	A	Human	MgH824	Madagascar	1979	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587045	B	Ovine	S35	Kenya	1972	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
DQ380197	C	Human	Saudi 2000-10911	Saudi Arabia	2000	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T .G. and Nichol,S.T.
DQ380196	C	Human	Kenya 9800523	Kenya	1998	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
JF311378	C	Homo sapiens	200803163	Madagascar	Feb-2008	Carroll,S.A., Reynes,J.M., Khristova,M.L., Andriamandimby,S.F., Rollin,P.E. and Nichol,S.T.
HQ009512	C	Bovine	M48/08	Madagascar	2008	Potgieter,C.
JQ820488	C	Homo sapiens	Sudan 85-2010	Sudan	Oct-2010	Aradaib,I.E., Erickson,B.R., Elageb,R.M., Khristova,M.L., Carroll,S.A., Elkhidir,I.M., Karsany,M.E., Karrar,A.E., Elbashir,M.I. and Nichol,S.T.
DQ380221	D	Human	73HB1230	Central African Republic	1973	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380212	E	Homo sapiens	74HB59	Central African Republic	1974	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380211	E	Human	73HB1449	Central African Republic	1973	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
HM587080	E	Bovine	VRL1887/78	Zimbabwe	1978	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587080	E	Bovine	VRL1887/78	Zimbabwe	1978	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587080	E	Bovine	VRL1887/78	Zimbabwe	1978	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587080	E	Bovine	VRL1887/78	Zimbabwe	1978	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587078	E	Bovine	VRL1290A/78	Zimbabwe	1978	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587079	E	Human	SPU44/85	Zambia	1985	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587079	E	Human	SPU44/85	Zambia	1985	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587101	F	Cx. zombaensis mosquito	Ar20364	South Africa	1981	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587101	F	Cx. zombaensis mosquito	Ar20364	South Africa	1981	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587083	G	Mansonia africana mosquito	ArB1976	Central African Republic	1969	Grobbelaar,A.A., Weyer,J ., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
DQ380218	G	Human	Hv-B375	Central African Republic	1985	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
AF134502	G	Micropteropus pusillus bat	An K6087	Central African Republic	-	Sall,A.A., Zanotto,P.M., Sene,O.K ., Zeller,H.G., Digoutte,J.P., Thiongane,Y. and Bouloy,M.
HM587084	G	Hipposideros caffer bat	AnK3837	Guinea	1981	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A ., Paweska,J.T. and Swanepoel,R.
AF134499	G	Ae. vexans arabiensis mosquito	Ar D104769	Central African Republic	-	Sall,A.A., Zanotto,P.M., Sene,O.K., Zeller,H.G., Digoutte,J.P., Thiongane,Y. and Bouloy,M.
HM587088	H	Human	SA71/10	South Africa	2010	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587093	H	Human	SA276/10	South Africa	2010	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587089	H	Human	SA482/10	South Africa	2010	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587090	H	Human	SA1221/10	South Africa	2010	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska, J.T. and Swanepoel,R.
HM587097	H	Human	SA373/10	South Africa	2010	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska, J.T. and Swanepoel,R.
HM587108	I	Ovine	An1830	South Africa	1956	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
HM587109	I	Ae. circumluteolus mosquito	Ar74	South Africa	1955	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
DQ380222	J	Bovine	2269/74	Zimbabwe	1974	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380194	K	Homo sapiens	2373/74	Zimbabwe	1974	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380193	K	RVFV strain Entebbe	Smithburn	Rift Valley fever virus genomics		Bird,B.H., Khristova,M.L. and Nichol,S.T.
DQ380191	K	Eretmapodites spp. bat	Entebbe	Uganda	1944	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
HM587105	K	Bovine	B314	Kenya	1962	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
KX611606	K	Homo sapiens	Beijing-01	China	22-Jul-2016	Pan,Y., Cui,S., Sun,Y., Li,J., Lv,Y., Dou,X., Li,X., Tian,L., He,Z., Li,X., Chen,L. and Wang,Q.
HM587119	M	Aedes spp. mosquito	Lunyo	Uganda	1955	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587120	M	Ae. circumluteolus mosquito	Ar118	South Africa	1955	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587110	L	Human	H1739	South Africa	1975	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
HM587118	L	Bovine	KEN56/B2653/IB8	Kenya	1963	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
JF784387	L	Bovine	35/74	South Africa	1974	Kortekaas,J., Oreshkova,N., Cobos-Jimenez,V., Vloet,R.P., Potgieter,C.A. and Moormann,R.J.
DQ380189	L	Human	SA-75	South Africa	1975	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
HM587111	L	Bovine	VRL763/70	Zimbabwe	1970	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J .T. and Swanepoel,R.
DQ380186	N	Human	OS-1	Mauritania	1987	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380183	N	Human	OS-9	Mauritania	1987	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380185	N	Human	OS-8	Mauritania	1987	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380184	N	Human	OS-3	Mauritania	1987	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380187	N	Ae. cuminsi mosquito	ARD-38388	Burkina Faso	1983	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
HM587125	O	Ovine	SA51	South Africa	1951	Grobbelaar,A.A., Weyer,J., Leman,P.A., Kemp,A., Paweska,J.T. and Swanepoel,R.
DQ380195	O	Ovine	SA-51 (Van Wyck)	South Africa	1951	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T .G. and Nichol,S.T.
DQ380209	A	Bovine	2250/74	Zimbabwe	1974	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.
DQ380199	A	Human	T-46 (228113)	Egypt	1977	Bird,B.H., Khristova,M.L. and Nichol,S.T.
DQ380208	A	RVF isolate ZH548	MP-12	Rift Valley fever virus genomics		Bird,B.H., Khristova,M.L. and Nichol,S.T.
DQ380205	A	Ovine	ZS-6365	Egypt	1979	Bird,B.H., Khristova,M.L., Rollin,P.E., Ksiazek,T.G. and Nichol,S.T.

Appendix B

Multiplex PCR method for sequencing Rift Valley fever virus
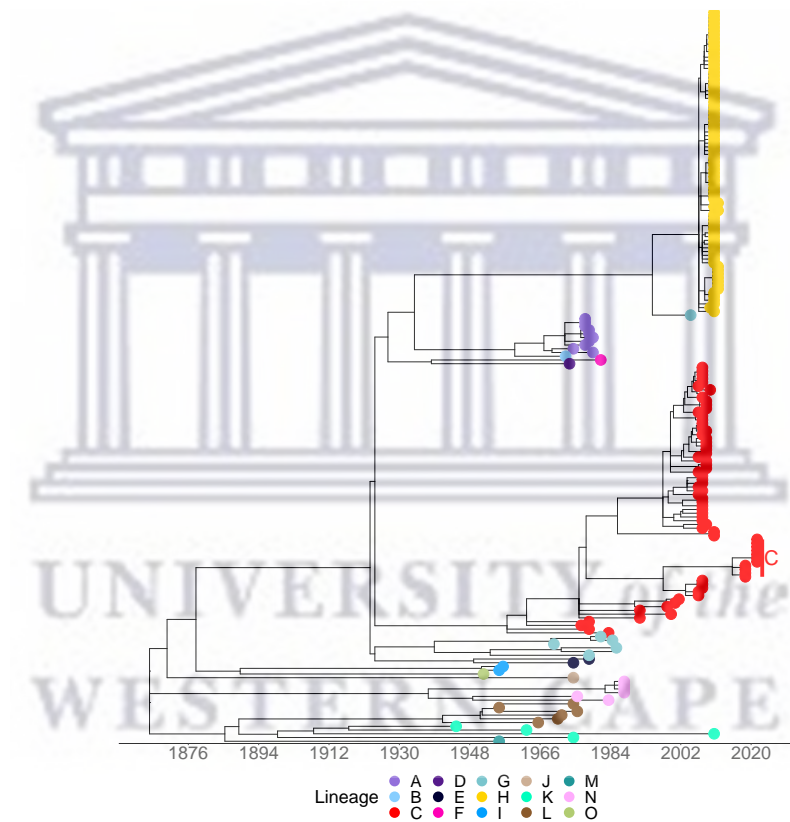
FIGURE B.1: **Molecular clock phylogenetic reconstruction.** Time scaled phylogenetic tree showing the occurrence of lineages from 1951 to 2021. The tips of the phylogenetic tree are colored according to the lineages and are labeled as accession, country code and year.

[1] qRT-PCR cycle threshold values [2] Percentage number of reads aligning to the Rift Valley fever reference genome segment [3] Mean genome depth of coverage per genomic segment
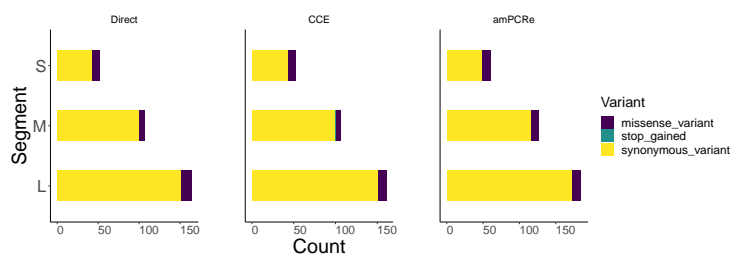
<table>
<tr><th></th><th>Chrom</th><th>Pos</th><th>Ref</th><th>Alt</th><th>Variant</th><th>SNP</th><th>Gene</th><th>AA.change</th><th>Segment</th></tr>
<tr><td>1</td><td>NC_014397.1</td><td>764</td><td>G</td><td>A</td><td>missense_variant</td><td>G764A</td><td>RdRp</td><td>Arg256Lys</td><td>L</td></tr>
<tr><td>2</td><td>NC_014397.1</td><td>851</td><td>G</td><td>A</td><td>missense_variant</td><td>G851A</td><td>RdRp</td><td>Ser285Asn</td><td>L</td></tr>
<tr><td>3</td><td>NC_014397.1</td><td>922</td><td>G</td><td>A</td><td>missense_variant</td><td>G922A</td><td>RdRp</td><td>Val308Ile</td><td>L</td></tr>
<tr><td>4</td><td>NC_014397.1</td><td>1063</td><td>G</td><td>A</td><td>missense_variant</td><td>G1063A</td><td>RdRp</td><td>Val355Met</td><td>L</td></tr>
<tr><td>5</td><td>NC_014397.1</td><td>1250</td><td>G</td><td>A</td><td>missense_variant</td><td>G1250A</td><td>RdRp</td><td>Ser418Asn</td><td>L</td></tr>
<tr><td>6</td><td>NC_014397.1</td><td>1354</td><td>G</td><td>A</td><td>missense_variant</td><td>G1354A</td><td>RdRp</td><td>As452Asn</td><td>L</td></tr>
<tr><td>7</td><td>NC_014397.1</td><td>2005</td><td>G</td><td>A</td><td>missense_variant</td><td>G2005A</td><td>RdRp</td><td>Ala669Thr</td><td>L</td></tr>
<tr><td>8</td><td>NC_014397.1</td><td>5335</td><td>A</td><td>G</td><td>missense_variant</td><td>A5335G</td><td>RdRp</td><td>Ile1779Val</td><td>L</td></tr>
<tr><td>9</td><td>NC_014397.1</td><td>5735</td><td>A</td><td>G</td><td>missense_variant</td><td>A5735G</td><td>RdRp</td><td>Lys1913Arg</td><td>L</td></tr>
<tr><td>10</td><td>NC_014397.1</td><td>5968</td><td>G</td><td>A</td><td>missense_variant</td><td>G5968A</td><td>RdRp</td><td>As1990Asn</td><td>L</td></tr>
<tr><td>11</td><td>NC_014396.1</td><td>715</td><td>T</td><td>A</td><td>missense_variant</td><td>T715A</td><td>Gn</td><td>Leu239Gln</td><td>M</td></tr>
<tr><td>12</td><td>NC_014396.1</td><td>1717</td><td>A</td><td>G</td><td>missense_variant</td><td>A1717G</td><td>Gn</td><td>As573Gly</td><td>M</td></tr>
<tr><td>13</td><td>NC_014396.1</td><td>1911</td><td>A</td><td>G</td><td>missense_variant</td><td>A1911G</td><td>Gn</td><td>Ile638Val</td><td>M</td></tr>
<tr><td>14</td><td>NC_014396.1</td><td>1913</td><td>A</td><td>G</td><td>missense_variant</td><td>A1913G</td><td>Gn</td><td>Ile639Met</td><td>M</td></tr>
<tr><td>15</td><td>NC_014396.1</td><td>3568</td><td>C</td><td>T</td><td>missense_variant</td><td>C3568T</td><td>Gc</td><td>Thr1190Ile</td><td>M</td></tr>
<tr><td>16</td><td>NC_014395.1</td><td>101</td><td>T</td><td>A</td><td>missense_variant</td><td>T101A</td><td>NSs</td><td>Phe35Ile</td><td>S</td></tr>
<tr><td>17</td><td>NC_014395.1</td><td>105</td><td>G</td><td>A</td><td>missense_variant</td><td>G105A</td><td>NSs</td><td>Arg36Lys</td><td>S</td></tr>
<tr><td>18</td><td>NC_014395.1</td><td>432</td><td>A</td><td>G</td><td>missense_variant</td><td>A432G</td><td>NSs</td><td>Asn145Ser</td><td>S</td></tr>
<tr><td>19</td><td>NC_014395.1</td><td>534</td><td>C</td><td>T</td><td>missense_variant</td><td>C534T</td><td>NSs</td><td>Ala179Val</td><td>S</td></tr>
<tr><td>20</td><td>NC_014395.1</td><td>651</td><td>C</td><td>T</td><td>missense_variant</td><td>C651T</td><td>NSs</td><td>Ser218Phe</td><td>S</td></tr>
<tr><td>21</td><td>NC_014395.1</td><td>684</td><td>T</td><td>C</td><td>missense_variant</td><td>T684C</td><td>NSs</td><td>Val229Ala</td><td>S</td></tr>
<tr><td>22</td><td>NC_014395.1</td><td>758</td><td>A</td><td>G</td><td>missense_variant</td><td>A758G</td><td>NSs</td><td>Ile254Val</td><td>S</td></tr>
<tr><td>23</td><td>NC_014395.1</td><td>764</td><td>C</td><td>T</td><td>missense_variant</td><td>C764T</td><td>NSs</td><td>Pro256Ser</td><td>S</td></tr>
<tr><td>24</td><td>NC_014395.1</td><td>792</td><td>A</td><td>G</td><td>missense_variant</td><td>A792G</td><td>NSs</td><td>Glu265Gly</td><td>S</td></tr>
<tr><td>25</td><td>NC_014395.1</td><td>1177</td><td>C</td><td>T</td><td>missense_variant</td><td>C1177T</td><td>NP</td><td>Gly393Glu</td><td>S</td></tr>
</table>

FIGURE B.2: **SNPs annotation and summary.** The upper panel shows bar graphs of the SNPs observed in Direct, CCE and amPCRe treated samples. The graphs show the number of SNPs per segment as well as their annotations as synonymous or non-synonymous (stop-gained and missense). The lower panel is a summarized table showing the shared/common SNPs identified in all the three treatments for all the genomic segments.

TABLE B.1: **Multiplex PCR amplicon primer scheme for the L segment.** The first column is the reference sequence accession, second column is the start site, third column is the end site, fourth column is the primer name (depicted as either forward, LEFT suffix or reverse depicted with the RIGHT suffix), the fifth column is the pool, and the sixth column is the strand.

Reference accession	Start	End	Primer name	Pool	Strand
NC_014397.1	19	47	rvfv-L-400_1_LEFT	1	+
NC_014397.1	394	416	rvfv-L-400_1_RIGHT	1	-
NC_014397.1	332	355	rvfv-L-400_2_LEFT	2	+
NC_014397.1	704	731	rvfv-L-400_2_RIGHT	2	-
NC_014397.1	642	664	rvfv-L-400_3_LEFT	1	+
NC_014397.1	1022	1044	rvfv-L-400_3_RIGHT	1	-
NC_014397.1	948	970	rvfv-L-400_4_LEFT	2	+
NC_014397.1	1339	1361	rvfv-L-400_4_RIGHT	2	-
NC_014397.1	1256	1280	rvfv-L-400_5_LEFT	1	+
NC_014397.1	1636	1660	rvfv-L-400_5_RIGHT	1	-
NC_014397.1	1564	1587	rvfv-L-400_6_LEFT	2	+
NC_014397.1	1932	1954	rvfv-L-400_6_RIGHT	2	-
NC_014397.1	1861	1883	rvfv-L-400_7_LEFT	1	+
NC_014397.1	2239	2259	rvfv-L-400_7_RIGHT	1	-
NC_014397.1	2164	2186	rvfv-L-400_8_LEFT	2	+
NC_014397.1	2545	2568	rvfv-L-400_8_RIGHT	2	-
NC_014397.1	2480	2504	rvfv-L-400_9_LEFT	1	+
NC_014397.1	2853	2876	rvfv-L-400_9_RIGHT	1	-
NC_014397.1	2778	2800	rvfv-L-400_10_LEFT	2	+
NC_014397.1	3168	3193	rvfv-L-400_10_RIGHT	2	-
NC_014397.1	3094	3120	rvfv-L-400_11_LEFT	1	+
NC_014397.1	3474	3495	rvfv-L-400_11_RIGHT	1	-
NC_014397.1	3403	3427	rvfv-L-400_12_LEFT	2	+
NC_014397.1	3790	3812	rvfv-L-400_12_RIGHT	2	-
NC_014397.1	3724	3746	rvfv-L-400_13_LEFT	1	+
NC_014397.1	4110	4133	rvfv-L-400_13_RIGHT	1	-
NC_014397.1	4047	4069	rvfv-L-400_14_LEFT	2	+
NC_014397.1	4423	4452	rvfv-L-400_14_RIGHT	2	-
NC_014397.1	4363	4387	rvfv-L-400_15_LEFT	1	+
NC_014397.1	4731	4758	rvfv-L-400_15_RIGHT	1	-
NC_014397.1	4657	4681	rvfv-L-400_16_LEFT	2	+
NC_014397.1	5041	5067	rvfv-L-400_16_RIGHT	2	-
NC_014397.1	4968	4993	rvfv-L-400_17_LEFT	1	+
NC_014397.1	5355	5377	rvfv-L-400_17_RIGHT	1	-
NC_014397.1	5296	5319	rvfv-L-400_18_LEFT	2	+
NC_014397.1	5674	5697	rvfv-L-400_18_RIGHT	2	-
NC_014397.1	5602	5626	rvfv-L-400_19_LEFT	1	+
NC_014397.1	5984	6008	rvfv-L-400_19_RIGHT	1	-
NC_014397.1	5906	5935	rvfv-L-400_20_LEFT	2	+
NC_014397.1	6279	6302	rvfv-L-400_20_RIGHT	2	-

TABLE B.2: **Multiplex PCR amplicon primer scheme for the M segment.** The first column is the reference sequence accession, second column is the start site, third column is the end site, fourth column is the primer name (depicted as either forward, LEFT suffix or reverse depicted with the RIGHT suffix), the fifth column is the pool, and the sixth column is the strand.

Reference accession	Start	End	Primer name	Pool	Strand
NC_014396.1	0	24	rvfv-M-400_1_LEFT	1	+
NC_014396.1	396	418	rvfv-M-400_1_RIGHT	1	-
NC_014396.1	329	351	rvfv-M-400_2_LEFT	2	+
NC_014396.1	708	730	rvfv-M-400_2_RIGHT	2	-
NC_014396.1	638	664	rvfv-M-400_3_LEFT	1	+
NC_014396.1	1025	1046	rvfv-M-400_3_RIGHT	1	-
NC_014396.1	965	987	rvfv-M-400_4_LEFT	2	+
NC_014396.1	1345	1368	rvfv-M-400_4_RIGHT	2	-
NC_014396.1	1279	1301	rvfv-M-400_5_LEFT	1	+
NC_014396.1	1667	1690	rvfv-M-400_5_RIGHT	1	-
NC_014396.1	1581	1608	rvfv-M-400_6_LEFT	2	+
NC_014396.1	1969	1991	rvfv-M-400_6_RIGHT	2	-
NC_014396.1	1908	1932	rvfv-M-400_7_LEFT	1	+
NC_014396.1	2292	2312	rvfv-M-400_7_RIGHT	1	-
NC_014396.1	2217	2239	rvfv-M-400_8_LEFT	2	+
NC_014396.1	2612	2636	rvfv-M-400_8_RIGHT	2	-
NC_014396.1	2537	2560	rvfv-M-400_9_LEFT	1	+
NC_014396.1	2934	2957	rvfv-M-400_9_RIGHT	1	-
NC_014396.1	2867	2892	rvfv-M-400_10_LEFT	2	+
NC_014396.1	3242	3268	rvfv-M-400_10_RIGHT	2	-
NC_014396.1	3164	3186	rvfv-M-400_11_LEFT	1	+
NC_014396.1	3551	3576	rvfv-M-400_11_RIGHT	1	-
NC_014396.1	3471	3493	rvfv-M-400_12_LEFT	2	+
NC_014396.1	3864	3885	rvfv-M-400_12_RIGHT	2	-

TABLE B.3: **Multiplex PCR amplicon primer scheme for the S segment.** The first column is the reference sequence accession, second column is the start site, third column is the end site, fourth column is the primer name (depicted as either forward, LEFT suffix or reverse depicted with the RIGHT suffix), the fifth column is the pool, and the sixth column is the strand.

Reference accession	Start	End	Primer name	Pool	Strand
NC_014395.1	0	22	rvfv-S-400_1_LEFT	1	+
NC_014395.1	386	408	rvfv-S-400_1_RIGHT	1	-
NC_014395.1	328	350	rvfv-S-400_2_LEFT	2	+
NC_014395.1	709	731	rvfv-S-400_2_RIGHT	2	-
NC_014395.1	641	667	rvfv-S-400_3_LEFT	1	+
NC_014395.1	1031	1054	rvfv-S-400_3_RIGHT	1	-
NC_014395.1	957	979	rvfv-S-400_4_LEFT	2	+
NC_014395.1	1339	1361	rvfv-S-400_4_RIGHT	2	-
NC_014395.1	1265	1286	rvfv-S-400_5_LEFT	1	+
NC_014395.1	1658	1684	rvfv-S-400_5_RIGHT	1	-

TABLE B.4: **Sample metadata sheet containing sequencing metrics, geographic information, and ELISA results for outbreak samples.** Samples' metadata for archived and clinical isolates subjected to the three treatment options.

Sample	Reads	Ct[1]	%RVFV[2]	Coverage	Depth[3]	Treatment	Segment	Host	Country	OD1	OD2
DK-B2	30.6	2940000	5.1	92	2105	amPCRe	L	Cow	Kenya	0.067	0.29
DK-B2	30.6	886000	0.1	24.2	6	Direct	L	Cow	Kenya	0.067	0.29
DVS-230	25.2	3960000	13.9	95.5	6032	amPCRe	L	Cow	Kenya	0.121	1.228
DVS-230	25.2	1668000	29.9	99.9	16019	CCE	L	Cow	Kenya	0.121	1.228
DVS-230	25.2	2848000	14.2	99.9	6811	Direct	L	Cow	Kenya	0.121	1.228
DVS-356	22.7	3640000	48.4	95.4	24230	amPCRe	L	Cow	Kenya	0.079	1.151
DVS-356	22.7	832000	21.4	99.7	5618	CCE	L	Cow	Kenya	0.079	1.151
DVS-356	22.7	1118000	41.9	99.9	8061	Direct	L	Cow	Kenya	0.079	1.151
DK-B2	30.6	2940000	87.1	91.6	61286	amPCRe	M	Cow	Kenya	0.067	0.29
DK-B2	30.6	886000	0	32.2	8	Direct	M	Cow	Kenya	0.067	0.29
DVS-230	25.2	3960000	25.7	95.4	18996	amPCRe	M	Cow	Kenya	0.121	1.228
DVS-230	25.2	1668000	10	99.6	8825	CCE	M	Cow	Kenya	0.121	1.228
DVS-230	25.2	2848000	6	99.8	4738	Direct	M	Cow	Kenya	0.121	1.228
DVS-356	22.7	3640000	34.4	98.8	28541	amPCRe	M	Cow	Kenya	0.079	1.228
DVS-356	22.7	832000	6.2	99.6	2651	CCE	M	Cow	Kenya	0.079	1.151
DVS-356	22.7	1118000	18	99.7	5608	Direct	M	Cow	Kenya	0.079	1.151
DK-B2	30.6	2940000	5.8	93.4	9172	amPCRe	S	Cow	Kenya	0.067	0.29
DK-B2	30.6	886000	0	72.7	18	Direct	S	Cow	Kenya	0.067	0.29
DVS-230	25.2	3960000	8.6	94.6	14117	amPCRe	S	Cow	Kenya	0.121	1.228
DVS-230	25.2	1668000	24	99.8	48512	CCE	S	Cow	Kenya	0.121	1.228
DVS-230	25.2	2848000	42.5	100	78662	Direct	S	Cow	Kenya	0.121	1.228
DVS-356	22.7	3640000	4	97	7306	amPCRe	S	Cow	Kenya	0.079	1.151
DVS-356	22.7	832000	15	99.4	14815	CCE	S	Cow	Kenya	0.079	1.151
DVS-356	22.7	1118000	22.4	99.9	16335	Direct	S	Cow	Kenya	0.079	1.151

TABLE B.5: **Multiplex PCR amplicon primers for sequencing RVFV.** S (n=10), M (n=24), L (n=40). The odd numbered primers belong to pool 1 while even numbered primers belong to pool 2. The primers were designed to amplify a target sequence of approximately 400 bases within the RVFV genome. Reference genomes are used to select primer pairs that maximize the likelihood of successful amplification for known virus diversity. A pairwise local alignment score between each candidate and the reference is calculated to ensure the most 'universal' candidate primers are picked for the scheme.

Primer name	Pool	Seq	Size	%GC	Melting temperature (use 65)
rvfv-S-400_1_LEFT	1	ACACAAAGACCCCCTAGTGCTT	22	50	61.75
rvfv-S-400_1_RIGHT	1	AGGTCAAAGAAAGCCAGTGAGG	22	50	60.67
rvfv-S-400_2_LEFT	2	AGTGATTTCACAAGGTTCCCCA	22	45.45	60.01
rvfv-S-400_2_RIGHT	2	TCATGCTAGGAAGTGATGAGCG	22	50	60.4
rvfv-S-400_3_LEFT	1	AGGACATTTCTAATGCTGTAGTTCCA	26	38.46	60.74
rvfv-S-400_3_RIGHT	1	TGAATGCAGCAGTGAATAGCAAC	23	43.48	60.37
rvfv-S-400_4_LEFT	2	GCTGACGGCTTCCCATTAGAAT	22	50	60.93
rvfv-S-400_4_RIGHT	2	TCCCGGGATGAGTTGACTCTAT	22	50	60.28
rvfv-S-400_5_LEFT	1	AGTGACAGGAAGCCACTCACT	21	52.38	61.14
rvfv-S-400_5_RIGHT	1	AGCTCCCTAGAGATACAAACACTATT	26	38.46	59.61
rvfv-M-400_1_LEFT	1	ACACAAAGACGGTGCATTAAATGT	24	37.5	60.28
rvfv-M-400_1_RIGHT	1	CCTGCCATGGTTTCTCTCCCTA	22	54.55	61.68
rvfv-M-400_2_LEFT	2	AAGGGTCCTCCTCACAAAAAGC	22	50	60.94
rvfv-M-400_2_RIGHT	2	TTCCCATGAGCACTCTGAAGGT	22	50	61.55
rvfv-M-400_3_LEFT	1	GCTCATCATAGAACTCTACTAGAGGC	26	46.15	60.56
rvfv-M-400_3_RIGHT	1	AACCTCATGTGCAGTGCTGAG	21	52.38	60.98
rvfv-M-400_4_LEFT	2	ACAAAAATGAAAGGGGTCTGCG	22	45.45	60.14
rvfv-M-400_4_RIGHT	2	CCCCTGATACTTGTATCTGCACA	23	47.83	60.12
rvfv-M-400_5_LEFT	1	CTCTGCTTATGAGTGCACTGCT	22	50	60.86
rvfv-M-400_5_RIGHT	1	GGGCAGTGAGCTACTATTTTGGA	23	47.83	60.63
rvfv-M-400_6_LEFT	2	AGATTACACTCAAGTATCCAGGGATAT	27	37.04	59.55
rvfv-M-400_6_RIGHT	2	CTGACCTCCTTCCATCCATCCT	22	54.55	61.15
rvfv-M-400_7_LEFT	2	GGGTGTATAAGAAGATGGTTGCCA	24	45.83	60.95
rvfv-M-400_7_RIGHT	1	AGTCCAATAGCTCTGGCCCT	20	55	60.34
rvfv-M-400_8_LEFT	2	TGTTGAAGGGGGTCAAGGAAGA	22	50	61.42
rvfv-M-400_8_RIGHT	2	TGTTGAAACAGAGCCATCAAAGTC	24	41.67	60.28
rvfv-M-400_9_LEFT	1	CAGTCAGTTAGAAAAGAGGCCCT	23	47.83	60.25
rvfv-M-400_9_RIGHT	1	ACCCCTCTCAAAGACAACAAAGG	23	47.83	60.95
rvfv-M-400_10_LEFT	2	GCACCAAAACCTTATCTCATACAAGC	25	44	60.83
rvfv-M-400_10_RIGHT	2	CACTGATCTTTTGTTCCATTCTCTGA	26	38.46	60.01
rvfv-M-400_11_LEFT	1	TGCCTGTCTATCACATCCACAG	22	50	60.01
rvfv-M-400_11_RIGHT	1	AGAGGCCTGTTCTTCCAAGATATAT	25	40	59.54
rvfv-M-400_12_LEFT	2	GTTGGTTTGGAGGGCCTCTTAA	22	50	60.67
rvfv-M-400_12_RIGHT	2	ACACAAAGACCGGTGCAACTT	21	47.62	61.06
rvfv-L-400_1_LEFT	1	TGGATTCTATATTATCAAAACAGCTGGT	28	32.14	59.61
rvfv-L-400_1_RIGHT	1	TCATCTCCCCTAAAGGTGGTGA	22	50	60.41
rvfv-L-400_2_LEFT	2	TGATCATTTGAGCCCTGACATGA	23	43.48	60.31
rvfv-L-400_2_RIGHT	2	ACAGATTCTGTGACTGACCAATTTATT	27	33.33	59.83
rvfv-L-400_3_LEFT	1	TCCACAGATGAGGAACTAGGGA	22	50	59.87
rvfv-L-400_3_RIGHT	1	GCACAGATCTTTGCCATCGACA	22	50	61.69
rvfv-L-400_4_LEFT	2	GCAGAAGACAACCTTAGGGACC	22	54.55	60.8
rvfv-L-400_4_RIGHT	2	TCCACATCATGATCTGGGGAGA	22	50	60.55
rvfv-L-400_5_LEFT	1	CATAGCTGCTTTAGGGGTTAATGG	24	45.83	59.8
rvfv-L-400_5_RIGHT	1	AAATCCCAGAACCTAGTAGTCGTT	24	41.67	59.55
rvfv-L-400_6_LEFT	2	TGGTCAGTTTGATAGGAGCTGAG	23	47.83	60.06
rvfv-L-400_6_RIGHT	2	CATCCATTGCTGCAGAGTCTGA	22	50	60.86
rvfv-L-400_7_LEFT	1	CCTCCTTCTCATTCTGGGCAGA	22	54.55	61.67
rvfv-L-400_7_RIGHT	1	AAGCAGTGGACGCCCTGAAA	20	55	62.35
rvfv-L-400_8_LEFT	2	AGAGAATAGCAGGTGAGCCCTT	22	50	61.08
rvfv-L-400_8_RIGHT	2	CCTTTCTAGAGTCAACCCACTGA	23	47.83	59.68
rvfv-L-400_9_LEFT	1	AGTCTTAAGGAGCCTCTATGGACA	24	45.83	60.67
rvfv-L-400_9_RIGHT	1	AAGAACTTCCCTATAGACCTGGC	23	47.83	59.55
rvfv-L-400_10_LEFT	2	CAGCATGGAGGTCTGAGAGAGA	22	54.55	61.13
rvfv-L-400_10_RIGHT	2	AGAGATCCATCACAAAGTCATCTCT	25	40	59.73
rvfv-L-400_11_LEFT	1	TGTTTACCAGGAAAAGGATGATGATG	26	38.46	60.06
rvfv-L-400_11_RIGHT	1	GCATATAGCTGCGGCCACTTT	21	52.38	61.52
rvfv-L-400_12_LEFT	2	TGCAAGGATCAGATGATAGTAGCA	24	41.67	59.73
rvfv-L-400_12_RIGHT	2	AATAGCTCAGACACTCCCATGC	22	50	60.34
rvfv-L-400_13_LEFT	1	GTGGGTCATTCTCCTTAGCTGC	22	54.55	61.18
rvfv-L-400_13_RIGHT	1	TCTGGTATGTTCAAACGGTCTCT	23	43.48	59.68
rvfv-L-400_14_LEFT	2	GGTGGCGCTCTAATTCTTAGCT	22	50	60.66
rvfv-L-400_14_RIGHT	2	CGGGAATAGGAATAATATATCTTCTGGCT	29	37.93	61.03
rvfv-L-400_15_LEFT	1	AGTACAGCTTGCTACAAAAGATGG	24	41.67	59.62
rvfv-L-400_15_RIGHT	1	CCTAACTTGAATGTGGTTATTCAGTGG	27	40.74	60.85
rvfv-L-400_16_LEFT	2	TGGCTGAATGGGAAAAGCTAAAGA	24	41.67	61.08
rvfv-L-400_16_RIGHT	2	GGTTTTTGTCTTACCATCTGATTCCT	26	38.46	60.01
rvfv-L-400_17_LEFT	1	GGCCCATATAGTGAGAGCTATAAGC	25	48	60.83
rvfv-L-400_17_RIGHT	1	TGTTTGTGACCCCCATGTCTTC	22	50	60.93
rvfv-L-400_18_LEFT	2	TCACAACTAACTGTCAGAGCAGC	23	47.83	61.05
rvfv-L-400_18_RIGHT	2	TGGCTCCCTCTCAAATAATTGCA	23	43.48	60.57
rvfv-L-400_19_LEFT	1	ACACTGCATCAGACAATGATCTCA	24	41.67	60.65
rvfv-L-400_19_RIGHT	1	TCCATCTCGTAAGGACCACTAACA	24	45.83	61.14
rvfv-L-400_20_LEFT	2	TGATGCGTTAATGGATCTAATGATAGAAG	29	34.48	60.32
rvfv-L-400_20_RIGHT	2	TGGTCTTAGCCTAGCATGTCATC	23	47.83	60.18

TABLE B.6: **Sequence accessions and metadata (n = 196) used in molecular phylogenetic analyses.** The Accession column is separated by '|' with the first string showing the accession from NCBI or strain of sequenced data in this study, the second part is the 3-letter country code and the third is the year of sample collection. The second column additionally provides the year of sample collection.

Accession	Country	Accession	Country	Accession	Country	Accession	Country
DQ380187|BFA|1983	BFA	MG659779|ZAF|2010	ZAF	MG659812|KEN|1962	KEN	KY126702|ZAF|2011	ZAF
KJ782453|CAF|1985	CAF	MG659777|ZAF|2010	ZAF	DQ380196|KEN|1998	KEN	KY126701|ZAF|2011	ZAF
DQ380212|CAF|1974	CAF	MG659776|ZAF|2010	ZAF	DQ380198|KEN|1983	KEN	KY126700|ZAF|2010	ZAF
KJ782456|CAF|1969	CAF	MG659775|ZAF|2010	ZAF	MG659813|KEN|1972	KEN	KY126711|ZAF|1956	ZAF
DQ380219|CAF|1985	CAF	MG659774|ZAF|2010	ZAF	DVS-230-NC|KEN|2021	KEN	KY126699|ZAF|2010	ZAF
DQ380221|CAF|1973	CAF	MG659785|ZAF|2010	ZAF	JF326193|KEN|2006	KEN	KY126696|ZAF|2010	ZAF
DQ380205|EGY|1979	EGY	MG659772|ZAF|2010	ZAF	DVS-356-C|KEN|2021	KEN	KY126695|ZAF|2010	ZAF
DQ380203|EGY|1978	EGY	MG659771|ZAF|2010	ZAF	DVS-356|KEN|2021	KEN	KY126694|ZAF|2010	ZAF
DQ380201|EGY|1977	EGY	MG659770|ZAF|2010	ZAF	DVS-356-NC|KEN|2021	KEN	KY126692|ZAF|2010	ZAF
DQ380199|EGY|1977	EGY	MG659769|ZAF|2010	ZAF	DVS-230|KEN|2021	KEN	KY126691|ZAF|2010	ZAF
DQ380207|EGY|1978	EGY	MG659767|ZAF|2010	ZAF	DVS-230-C|KEN|2021	KEN	KY126690|ZAF|2010	ZAF
DQ380204|EGY|1978	EGY	MG659765|ZAF|2010	ZAF	JF311381|MDG|1991	MDG	KX944841|ZAF|2008	ZAF
NC_014396|EGY|1977	EGY	MG659764|ZAF|2010	ZAF	JF311380|MDG|2008	MDG	KX944831|ZAF|1955	ZAF
DQ380216|GIN|1984	GIN	MG659762|ZAF|2010	ZAF	JF311379|MDG|2008	MDG	JF784387|ZAF|1974	ZAF
DQ380215|GIN|1981	GIN	MG659773|ZAF|2010	ZAF	JF311384|MDG|2008	MDG	KY126698|ZAF|2010	ZAF
EU574044|KEN|2007	KEN	MG659787|ZAF|2010	ZAF	JF311382|MDG|1991	MDG	KY126712|ZAF|1971	ZAF
EU574043|KEN|2007	KEN	MG659788|ZAF|2010	ZAF	JF311378|MDG|2008	MDG	KY126708|ZAF|2009	ZAF
EU574042|KEN|2007	KEN	DQ380195|ZAF|1951	ZAF	JF311377|MDG|2008	MDG	KY126714|ZAF|1999	ZAF
EU574041|KEN|2006	KEN	MG659810|ZAF|2011	ZAF	JF311383|MDG|2008	MDG	MG659756|ZAF|2010	ZAF
EU574040|KEN|2006	KEN	MG659809|ZAF|2011	ZAF	HQ009512|MDG|2008	MDG	MG659755|ZAF|2010	ZAF
EU574039|KEN|2007	KEN	MG659808|ZAF|2011	ZAF	DQ380210|MDG|1979	MDG	MG659754|ZAF|2010	ZAF
EU574038|KEN|2007	KEN	MG659807|ZAF|2011	ZAF	JF311385|MDG|2008	MDG	MG659753|ZAF|2010	ZAF
EU574045|KEN|2007	KEN	MG659806|ZAF|2011	ZAF	DQ380184|MRT|1987	MRT	MG659752|ZAF|2010	ZAF
EU574047|KEN|2007	KEN	MG659805|ZAF|2011	ZAF	DQ380186|MRT|1987	MRT	MG659749|ZAF|2010	ZAF
08HAB|KEN|2018	KEN	MG659803|ZAF|2010	ZAF	DQ380183|MRT|1987	MRT	KY126713|ZAF|1981	ZAF
EU574034|KEN|2007	KEN	MG659801|ZAF|2010	ZAF	DQ380185|MRT|1987	MRT	MG659748|ZAF|2010	ZAF
HM586974|KEN|2007	KEN	MG659800|ZAF|2010	ZAF	HE687306|MYT|2008	MYT	MG659747|ZAF|2010	ZAF
HM586973|KEN|2007	KEN	MG659799|ZAF|2010	ZAF	HE687303|MYT|2008	MYT	MG659745|ZAF|2010	ZAF
HM586972|KEN|2006	KEN	MG659798|ZAF|2010	ZAF	MT561460|NAM|2010	NAM	KY126693|ZAF|2010	ZAF
JF326191|KEN|2007	KEN	MG659797|ZAF|2010	ZAF	MG659811|NAM|2011	NAM	MG659743|ZAF|2010	ZAF
JF326192|KEN|2006	KEN	MG659796|ZAF|2010	ZAF	KY126703|NAM|2004	NAM	MG659742|ZAF|2010	ZAF
HM586969|KEN|2007	KEN	MG659795|ZAF|2010	ZAF	RU1|RWA|2018	RWA	MG659741|ZAF|2010	ZAF
EU574048|KEN|2007	KEN	DQ380189|ZAF|1975	ZAF	RU1-C|RWA|2018	RWA	MG659740|ZAF|2010	ZAF
HM586968|KEN|2007	KEN	MG659794|ZAF|2010	ZAF	DQ380197|SAU|2000	SAU	MG659739|ZAF|2010	ZAF
HM586966|KEN|2007	KEN	MG659793|ZAF|2010	ZAF	KX096942|SAU|2001	SAU	MG659738|ZAF|2010	ZAF
HM586965|KEN|2006	KEN	MG659791|ZAF|2010	ZAF	JQ820488|SDN|2010	SDN	MG659737|ZAF|2010	ZAF
HM586964|KEN|2006	KEN	MG659761|ZAF|2010	ZAF	JQ820487|SDN|2010	SDN	MG659736|ZAF|2010	ZAF
EU574056|KEN|2007	KEN	MG659789|ZAF|2010	ZAF	MG659818|SEN|1975	SEN	MG659734|ZAF|2010	ZAF
EU574055|KEN|2007	KEN	MG659786|ZAF|2010	ZAF	HM586970|TZA|2007	TZA	MG659744|ZAF|2010	ZAF
EU574051|KEN|2007	KEN	MG659790|ZAF|2010	ZAF	HM586971|TZA|2007	TZA	MG659733|ZAF|1955	ZAF
HM586967|KEN|2007	KEN	MG659759|ZAF|2010	ZAF	JF326195|TZA|2007	TZA	DQ380188|ZWE|1970	ZWE
EU574031|KEN|2007	KEN	MG659758|ZAF|2010	ZAF	JF326194|TZA|2007	TZA	DQ380222|ZWE|1974	ZWE
EF460404|KEN|2007	KEN	KY126710|ZAF|1955	ZAF	DQ380191|UGA|1944	UGA	MG659816|ZWE|1978	ZWE
EF467177|KEN|2007	KEN	KY126709|ZAF|2009	ZAF	MG659763|ZAF|2010	ZAF	DQ380214|ZWE|1978	ZWE
08HAB-C|KEN|2018	KEN	MG659760|ZAF|2010	ZAF	MG659784|ZAF|2010	ZAF	MG659814|ZWE|1976	ZWE
DK-B2|KEN|2021	KEN	KY126707|ZAF|2008	ZAF	MG659782|ZAF|2010	ZAF	DQ380220|ZWE|1978	ZWE
DQ380190|KEN|1965	KEN	KY126705|ZAF|2008	ZAF	MG659781|ZAF|2010	ZAF	DQ380194|ZWE|1974	ZWE
EF467178|KEN|2006	KEN	KY126704|ZAF|2008	ZAF	MG659780|ZAF|2010	ZAF	MG659815|ZWE|1978	ZWE
						DQ380209|ZWE|1974	ZWE

Appendix C

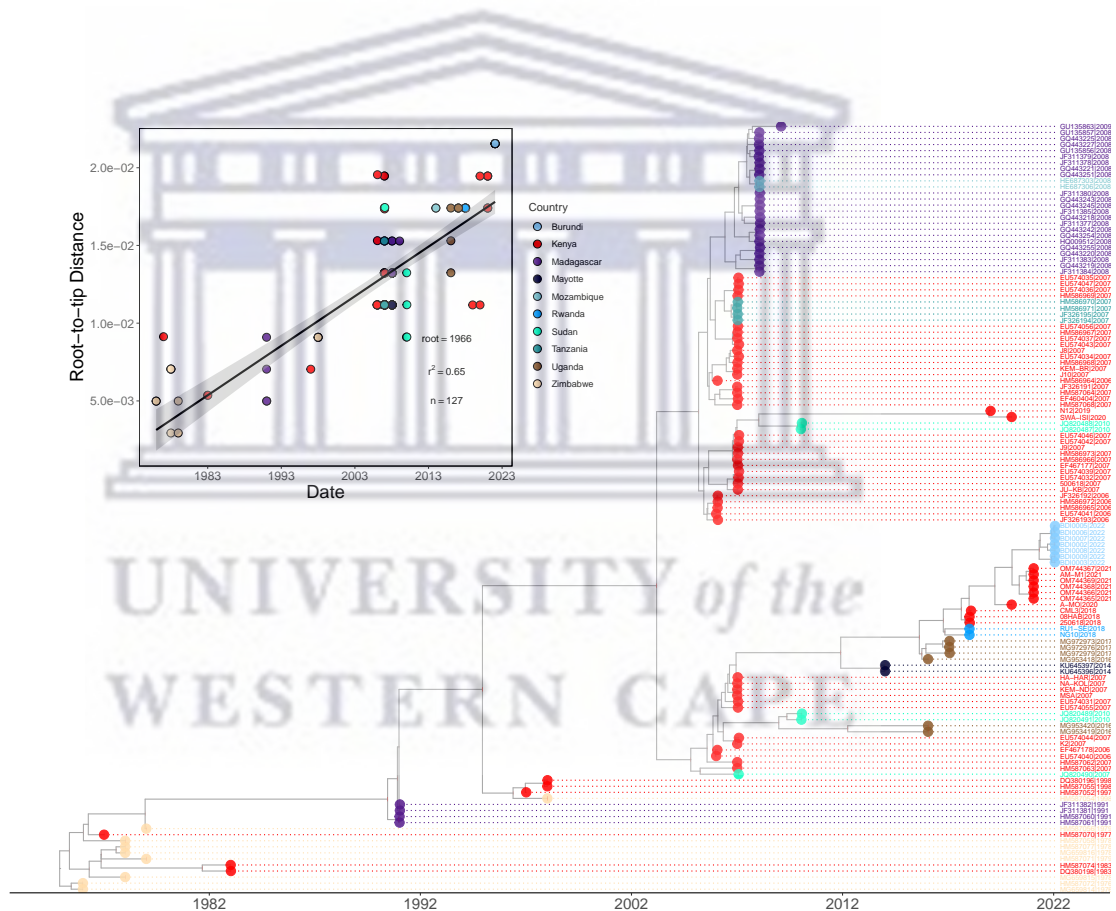# Transmission and phylogeographic distribution of RVFV using phylodynamic approaches

FIGURE C.1: **Time-calibrated maximum clade credibility tree of discrete phylogeographic analysis using partial glycoprotein sequences.** Tips and tips labels in the phylogeny are colored according to countries. The inset plot shows regression of root-to-distance against sampling times from the maximum likelihood phylogeny inferred using IQ-TREE. The plot was generated using the *lm* function in R.
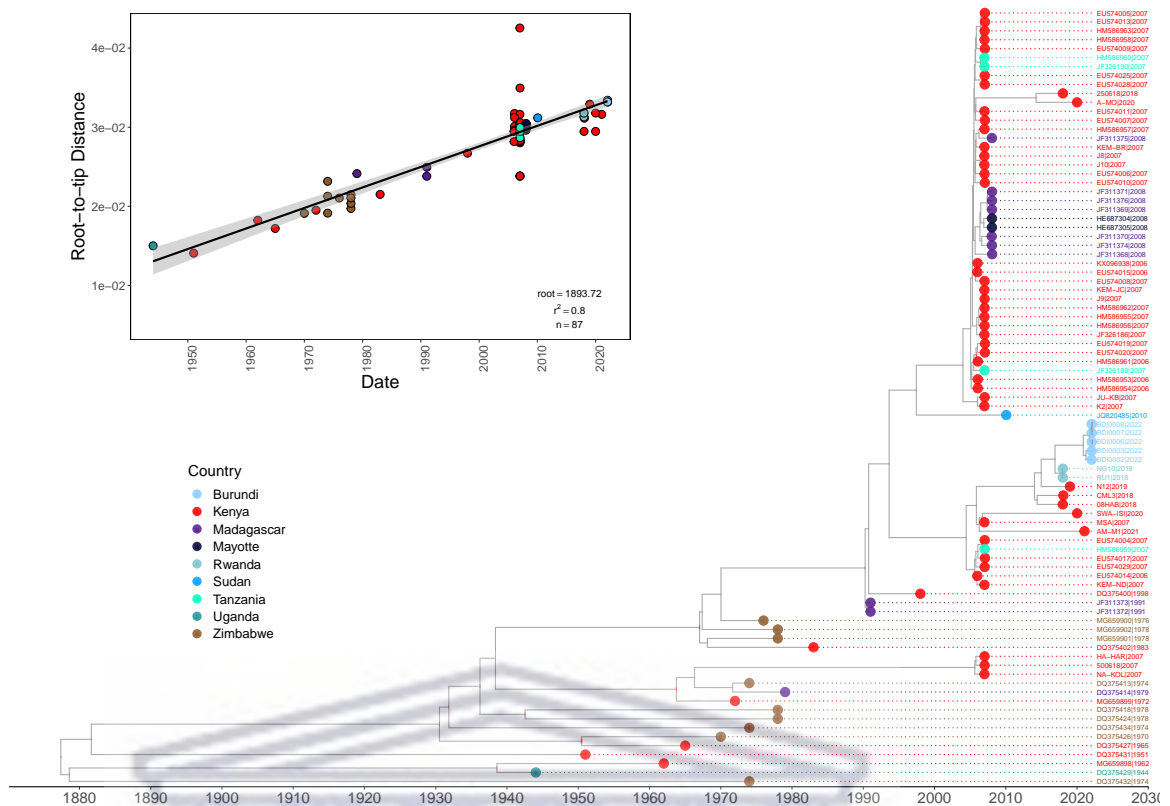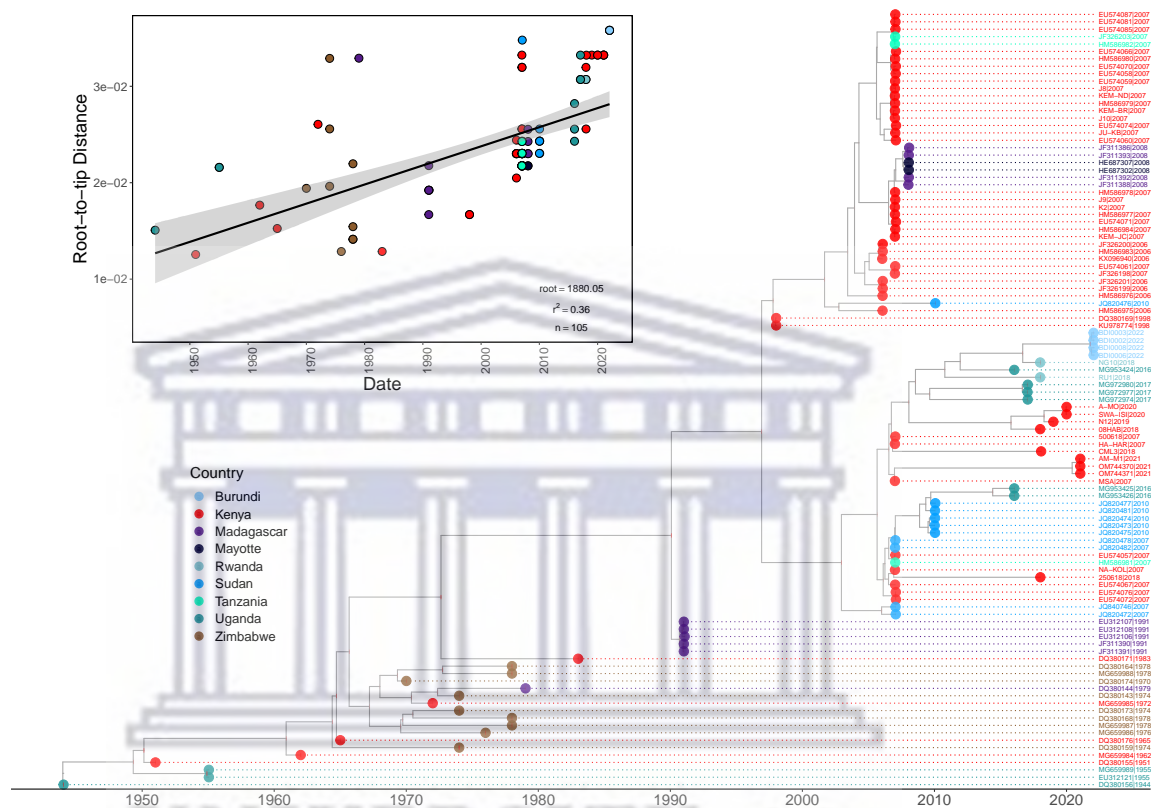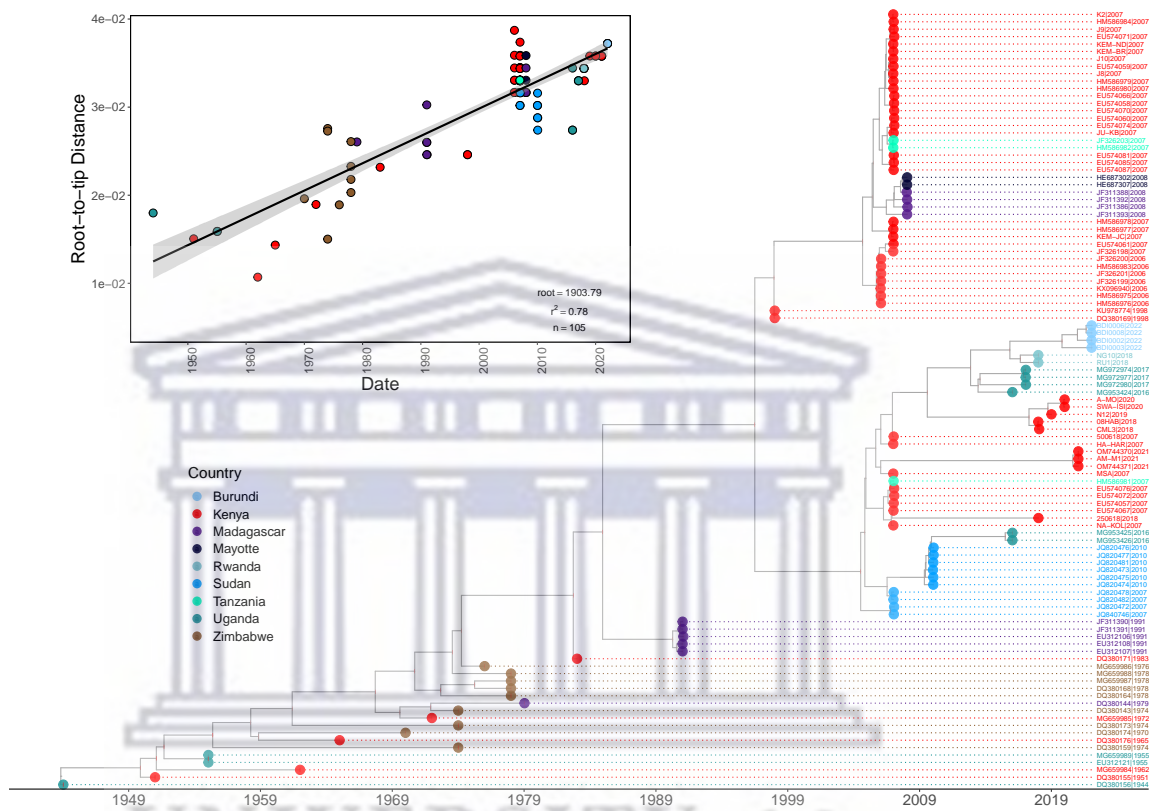
130

FIGURE C.2: **Time-calibrated maximum clade credibility tree of continuous phylogeographic analysis using complete L segment sequences (codon-aligned).** Tips and tips labels in the phylogeny are colored according to countries. The inset plot shows regression of root-to-distance against sampling times from the maximum likelihood phylogeny inferred using IQ-TREE. The plot was generated using the *lm* function in R.

TABLE C.1: **The most significant supported pathways of RVFV lineage C spread between East Africa countries.** Bayes Factor support for non-zero viral migration pathways between pairs of geographic locations was obtained using the BSSVS procedure. Bayes Factors were summarized using the SpreaD3 v.0.9.7.1 application. (KEN=Kenya, MDG=Madagascar, SDN=Sudan, TZA=Tanzania, ZWE=Zimbabwe, UGA=Uganda and RWA=Rwanda).

From	To	Bayes Factor	Posterior probability
KEN	MDG	166.421883579794	0.952578203113841
KEN	SDN	1602.89996481333	0.994857877446079
KEN	TZA	44.7339706258486	0.843736609055849
MDG	ZWE	13.8788864375757	0.626196257677475
SDN	UGA	28.1260475410483	0.772461076989002
ZWE	KEN	2628.19943065693	0.996857591772604
UGA	RWA	57.7773399120356	0.874589344379374
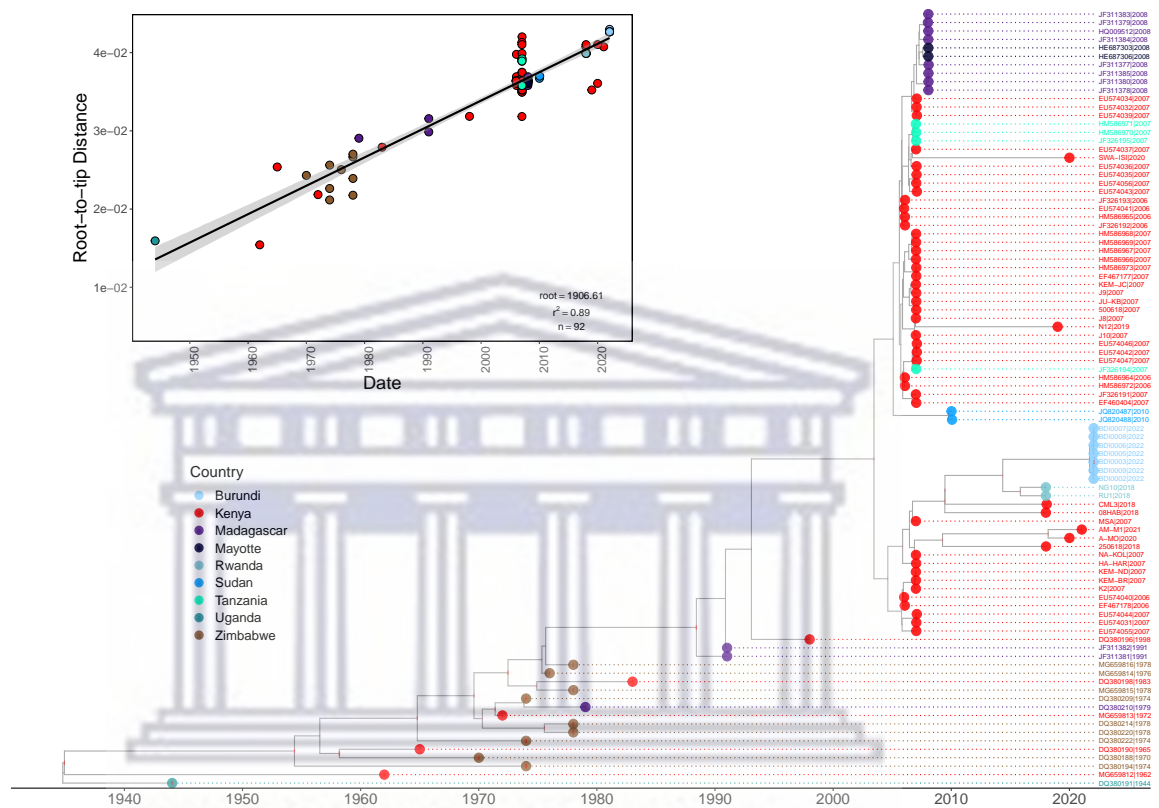
FIGURE C.3: **Time-calibrated maximum clade credibility tree of continuous phylogeographic analysis using complete M segment sequences (codon-aligned).** Tips and tips labels in the phylogeny are colored according to countries. The inset plot shows regression of root-to-distance against sampling times from the maximum likelihood phylogeny inferred using IQ-TREE. The plot was generated using the *lm* function in R.

FIGURE C.4: **Time-calibrated maximum clade credibility tree of continuous phylogeographic analysis using NSs gene sequences.** Tips and tips labels in the phylogeny are colored according to countries. The inset plot shows regression of root-to-distance against sampling times from the maximum likelihood phylogeny inferred using IQ-TREE. The plot was generated using the *lm* function in R.

FIGURE C.5: **Time-calibrated maximum clade credibility tree of continuous phylogeographic analysis using NP gene sequences.** Tips and tips labels in the phylogeny are colored according to countries. The inset plot shows regression of root-to-distance against sampling times from the maximum likelihood phylogeny inferred using IQ-TREE. The plot was generated using the *lm* function in R.
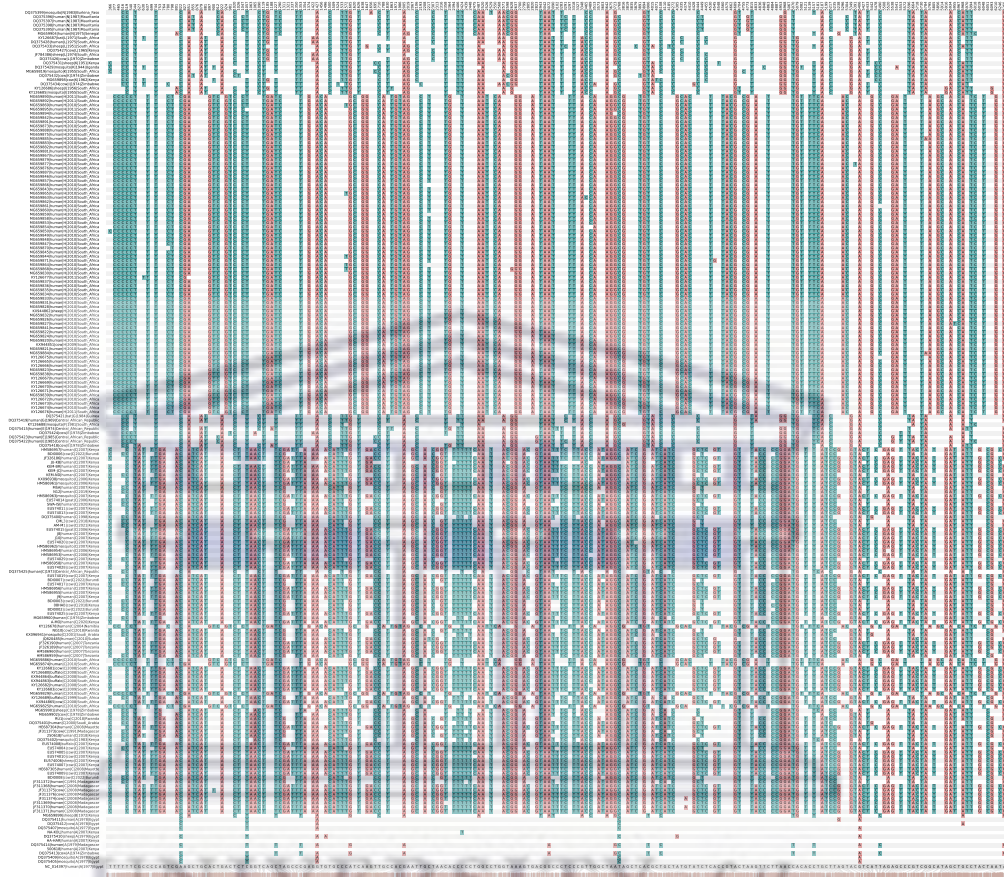
FIGURE C.6: **SNPs summary relative to the reference sequence of RVFV L segment.** Multiple sequence alignment of singleton SNPs. Rows represent sequence identifiers and relevant metadata (host species, lineage, year of collection and country), and are ordered by lineage. The columns represent positions within the genome where SNPs occur colored as green - pyrimidines (C and T) and light maroon - purines (G and A).
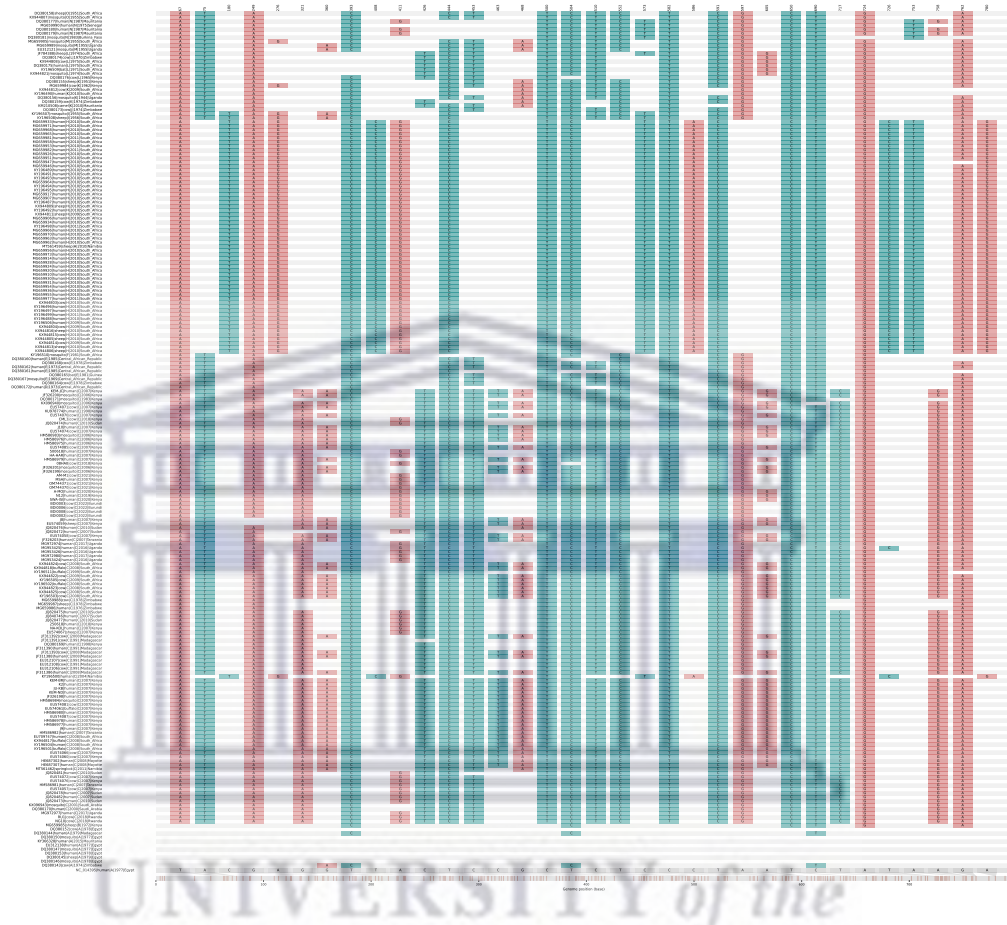
Figure C.7: **SNPs summary relative to the reference sequence of RVFV NSs gene sequences in the S segment.** Multiple sequence alignment of single-ton SNPs. Rows represent sequence identifiers and relevant metadata (host species, lineage, year of collection and country), and are ordered by lineage. The columns represent positions within the genome where SNPs occur colored as green - pyrimidines (C and T) and light maroon - purines (G and A).
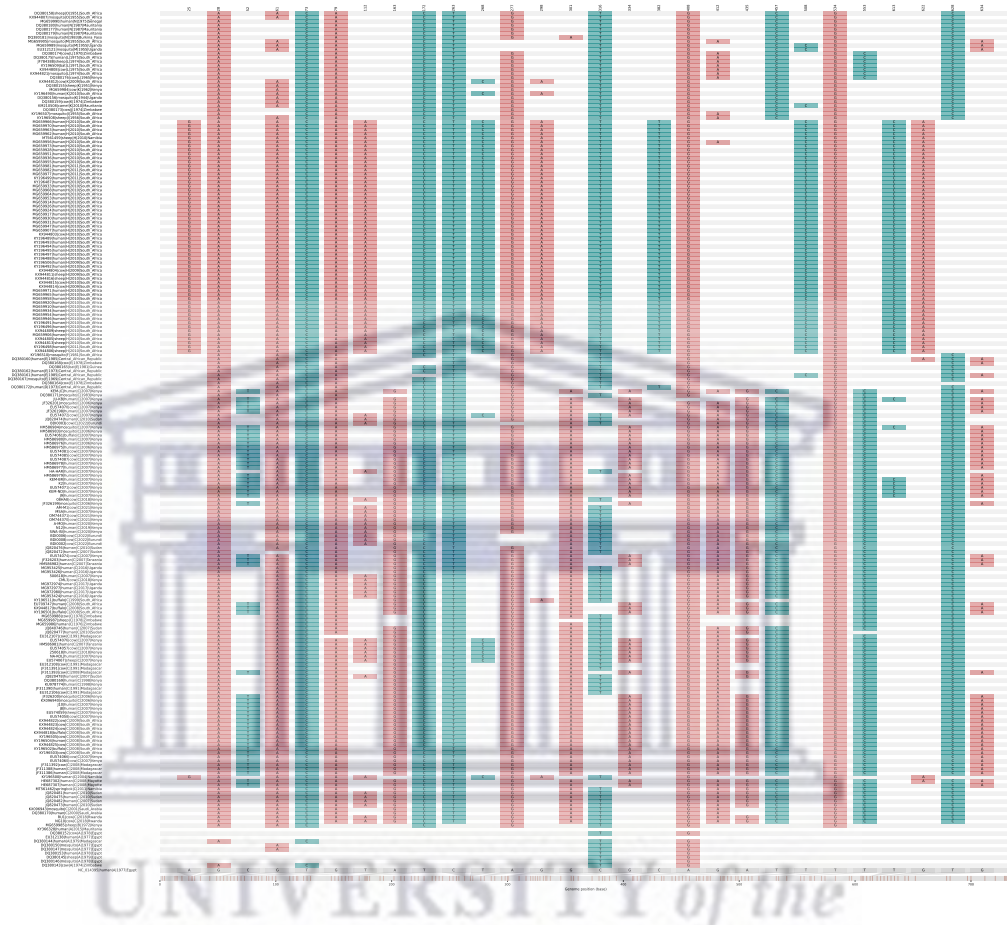
FIGURE C.8: **SNPs summary relative to the reference sequence of RVFV NP gene sequences in the S segment.** Multiple sequence alignment of singleton SNPs. Rows represent sequence identifiers and relevant metadata (host species, lineage, year of collection and country), and are ordered by lineage. The columns represent positions within the genome where SNPs occur colored as green - pyrimidines (C and T) and light maroon - purines (G and A).

TABLE C.2: **The most significant supported pathways of RVFV lineage C spread between East Africa geographic locations.** Bayes Factor support for non-zero viral migration pathways between pairs of geographic locations was obtained using the BSSVS procedure. Bayes Factors were summarized using the SpreaD3 v.0.9.7.1 application.

From	To	Bayes Factor	Posterior probability
Ampandrambato	Andohasahalava	195.648011262356	0.782745322096843
Bandraboua	Sada	68.215670539983	0.556777603199543
Baringo	Kilifi	60.1042907774117	0.525353520925582
Baringo	Magadi	71.7080830596263	0.569061562633909
Dodoma	Tanga	183.902191476397	0.772032566776175
Garissa	Kilifi	59.6243079953189	0.523353806599057
Gatooma	Mombasa	54.6922470410459	0.501785459220111
Gatooma	Salisbury	4366.34277451673	0.987716040565633
Kahawasukari	Ruiru	62.0297506460821	0.533209541494072
Ngozi	Ruyaga	82.4015830865323	0.602771032709613
Salisbury	Sinoia	105.703506416541	0.660619911441223
Miarinarivo	Amparafaravola	99.4274714794847	0.646764747893158
Fenerive	Antananarivo	59.0130733747505	0.520782745322097
Miarinarivo	Antsirabe	236.573431032574	0.813312383945151
Garissa	Baringo	380121.236479875	1
Nyandarua	Isiolo	64.4647073960786	0.542779602913869
Mityana	Kiboga	81.1353486588145	0.599057277531781
Ngozi	Kobero	76.5213087889847	0.584916440508499
Wajir	Marsabit	62.6380734191648	0.535637766033424
Rulindo	Ngoma	115.04019094488	0.67933152406799

Bibliography

Adam, A. A., Karsany, M. S., & Adam, I. (2010). Manifestations of severe rift valley fever in sudan. *International Journal of Infectious Diseases*, *14*(2), e179–e180.

Adam, I., & Karsany, M. S. (2008). Case report: Rift valley fever with vertical transmission in a pregnant sudanese woman. *Journal of medical virology*, *80*(5), 929–929.

Adams, M. J., Lefkowitz, E. J., King, A. M. Q., Harrach, B., Harrison, R. L., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Mushegian, A. R., Nibert, M., Sabanadzovic, S., Sanfaçon, H., Siddell, S. G., Simmonds, P., Varsani, A., Zerbini, F. M., Gorbalenya, A. E., & Davison, A. J. (2017). Changes to taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2017). *Archives of Virology*, *162*(8), 2505–2538. https://doi.org/10.1007/s00705-017-3358-5

Ahmad, K. (2000). More deaths from rift valley fever in saudi arabia and yemen. *The Lancet*, *356*(9239), 1422. https://doi.org/10.1016/S0140-6736(05)74068-X

Ahmed, A., Ali, Y., Elduma, A., Eldigail, M. H., Mhmoud, R. A., Mohamed, N. S., Ksiazek, T. G., Dietrich, I., & Weaver, S. C. (2020). Unique Outbreak of Rift Valley Fever in Sudan, 2019. *Emerg Infect Dis*, *26*(12), 3030–3033. https://doi.org/10.3201/eid2612.201599

Allen, E. R., Krumm, S. A., Raghwani, J., Halldorsson, S., Elliott, A., Graham, V. A., Koudriakova, E., Harlos, K., Wright, D., Warimwe, G. M., Brennan, B., Huiskonen, J. T., Dowall, S. D., Elliott, R. M., Pybus, O. G., Burton, D. R., Hewson, R., Doores, K. J., & Bowden, T. A. (2018). A protective monoclonal antibody targets a site of vulnerability on the surface of rift valley fever virus [Publisher: Elsevier]. *Cell Reports*, *25*(13), 3750–3758.e4. https://doi.org/10.1016/j.celrep.2018.12.001

Ally, D., Wiss, V. R., Deckert, G. E., Green, D., Roychoudhury, P., Wichman, H. A., Brown, C. J., & Krone, S. M. (2014). The impact of spatial structure on viral genomic diversity generated during adaptation to thermal stress. *PLoS One*, *9*(2), e88702.

Anderson, E., & Rowe, L. (1998). The prevalence of antibody to the viruses of bovine virus diarrhoea, bovine herpes virus 1, rift valley fever, ephemeral fever and

bluetongue and to leptospira sp in free-ranging wildlife in zimbabwe. *Epidemiology & Infection*, *121*(2), 441–449.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Andriamandimby, S. F., Randrianarivo-Solofoniaina, A. E., Jeanmaire, E. M., Ravololomanana, L., Razafimanantsoa, L. T., Rakotojoelinandrasana, T., Razainirina, J., Hoffmann, J., Ravalohery, J.-P., Rafisandratantsoa, J.-T., Rollin, P. E., & Reynes, J.-M. (2010). Rift valley fever during rainy seasons, madagascar, 2008 and 2009. *Emerg. Infect. Dis.*, *16*(6), 963–970. https://doi.org/10.3201/eid1606.091266

Anyamba, A., Chretien, J.-P., Small, J., Tucker, C. J., Formenty, P. B., Richardson, J. H., Britch, S. C., Schnabel, D. C., Erickson, R. L., & Linthicum, K. J. (2009). Prediction of a rift valley fever outbreak. *Proceedings of the National Academy of Sciences*, *106*(3), 955–959.

Anyangu, A. S., Gould, L. H., Sharif, S. K., Nguku, P. M., Omolo, J. O., Mutonga, D., Rao, C. Y., Lederman, E. R., Schnabel, D., Paweska, J. T., Katz, M., Hightower, A., Njenga, M. K., Feikin, D. R., & Breiman, R. F. (2010). Risk factors for severe Rift Valley fever infection in Kenya, 2007 [Publisher: The American Society of Tropical Medicine and Hygiene]. *Am J Trop Med Hyg*, *83*(2 Suppl), 14–21. https://doi.org/10.4269/ajtmh.2010.09-0293

Aradaib, I. E., Erickson, B. R., Elageb, R. M., Khristova, M. L., Carroll, S. A., Elkhidir, I. M., Karsany, M. E., Karrar, A. E., Elbashir, M. I., & Nichol, S. T. (2013). Rift valley fever, sudan, 2007 and 2010. *Emerging Infectious Diseases*, *19*(2), 246.

Archer, B. N., Thomas, J., Weyer, J., Cengimbo, A., Landoh, D. E., Jacobs, C., Ntuli, S., Modise, M., Mathonsi, M., Mashishi, M. S., Leman, P. A., le Roux, C., Jansen van Vuren, P., Kemp, A., Paweska, J. T., & Blumberg, L. (2013). Epidemiologic investigations into outbreaks of rift valley fever in humans, south africa, 2008–2011. *Emerg. Infect. Dis.*, *19*(12). https://doi.org/10.3201/eid1912.121527

Arishi, H. M., Aqeel, A. Y., & Al Hazmi, M. M. (2006). Vertical transmission of fatal rift valley fever in a newborn. *Annals of Tropical Paediatrics*, *26*(3), 251–253. https://doi.org/10.1179/146532806X120363

Armstrong, G. L., MacCannell, D. R., Taylor, J., Carleton, H. A., Neuhaus, E. B., Bradbury, R. S., Posey, J. E., & Gwinn, M. (2019). Pathogen genomics in public health. *New England Journal of Medicine*, *381*(26), 2569–2580.

Arum, S. O., Weldon, C. W., Orindi, B., Landmann, T., Tchouassi, D. P., Affognon, H. D., & Sang, R. (2015). Distribution and diversity of the vectors of rift valley fever along the livestock movement routes in the northeastern and coastal regions of kenya. *Parasites Vectors*, *8*(1), 294. https://doi.org/10.1186/s13071-015-0907-1

Ayres, D. L., Cummings, M. P., Baele, G., Darling, A. E., Lewis, P. O., Swofford, D. L., Huelsenbeck, J. P., Lemey, P., Rambaut, A., & Suchard, M. A. (2019). Beagle 3:

Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Systematic biology*, *68*(6), 1052–1061.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M. S., Keïta, S., De Clerck, H., Tiffany, A., Dominguez, G., Loua, M., Traoré, A., Kolié, M., Malano, E. R., Heleze, E., Bocquin, A., Mély, S., . . . Günther, S. (2014). Emergence of zaire ebola virus disease in guinea. *N Engl J Med*, *371*(15), 1418–1425. https://doi.org/10.1056/NEJMoa1404505

Baker, R. E., Mahmud, A. S., Miller, I. F., Rajeev, M., Rasambainarivo, F., Rice, B. L., Takahashi, S., Tatem, A. J., Wagner, C. E., Wang, L.-F., et al. (2022). Infectious disease in an era of global change. *Nature Reviews Microbiology*, *20*(4), 193–205.

Balenghien, T., Cardinale, E., Chevalier, V., Elissa, N., Failloux, A.-B., Jean Jose Nipomichene, T., Nicolas, G., Rakotoharinome, V., Roger, M., & Zumbo, B. (2013). Towards a better understanding of rift valley fever epidemiology in the south-west of the indian ocean. *Vet Res*, *44*(1), 78. https://doi.org/10.1186/1297-9716-44-78

Balkhy, H. H., & Memish, Z. A. (2003). Rift valley fever: An uninvited zoonosis in the arabian peninsula. *International Journal of Antimicrobial Agents*, *21*(2), 153–157. https://doi.org/10.1016/S0924-8579(02)00295-9

Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *40*, D48–D53. https://doi.org/10.1093/nar/gkr1202

Biek, R., Pybus, O. G., Lloyd-Smith, J. O., & Didelot, X. (2015). Measurably evolving pathogens in the genomic era. *Trends in Ecology & Evolution*, *30*(6), 306–313. https://doi.org/10.1016/j.tree.2015.03.009

Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). Spread3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Molecular biology and evolution*, *33*(8), 2167–2169.

Bieńkowska-Szewczyk, K., & Szewczyk, B. (1999). Expression of genes coding for animal virus glycoproteins in heterologous systems. *Acta Biochimica Polonica*, *46*(2), 325–339.

Billecocq, A., Gauliard, N., Le May, N., Elliott, R. M., Flick, R., & Bouloy, M. (2008). Rna polymerase i-mediated expression of viral rna for the rescue of infectious virulent and avirulent rift valley fever viruses. *Virology*, *378*(2), 377–384.

Bird, B. H., Bawiec, D. A., Ksiazek, T. G., Shoemaker, T. R., & Nichol, S. T. (2007). Highly sensitive and broadly reactive quantitative reverse transcription-PCR assay for high-throughput detection of rift valley fever virus. *Journal of Clinical Microbiology*, *45*(11), 3506–3513. https://doi.org/10.1128/JCM.00936-07

Bird, B. H., Githinji, J. W. K., Macharia, J. M., Kasiiti, J. L., Muriithi, R. M., Gacheru, S. G., Musaa, J. O., Towner, J. S., Reeder, S. A., Oliver, J. B., Stevens, T. L., Erickson, B. R., Morgan, L. T., Khristova, M. L., Hartman, A. L., Comer, J. A.,

Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2008). Multiple virus lineages sharing recent common ancestry were associated with a large rift valley fever outbreak among livestock in kenya during 2006-2007. *Journal of Virology*, *82*(22), 11152–11166. https://doi.org/10.1128/JVI.01519-08

Bird, B. H., Albarino, C. G., Hartman, A. L., Erickson, B. R., Ksiazek, T. G., & Nichol, S. T. (2008). Rift valley fever virus lacking the nss and nsm genes is highly attenuated, confers protective immunity from virulent virus challenge, and allows for differential identification of infected and vaccinated animals. *Journal of virology*, *82*(6), 2681–2691.

Bird, B. H., Albariño, C. G., & Nichol, S. T. (2007). Rift valley fever virus lacking NSm proteins retains high virulence in vivo and may provide a model of human delayed onset neurologic disease. *Virology*, *362*(1), 10–15. https://doi.org/10.1016/j.virol.2007.01.046

Bird, B. H., Khristova, M. L., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2007). Complete genome analysis of 33 ecologically and biologically diverse rift valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *Journal of virology*, *81*(6), 2805–2816.

Bird, B. H., Ksiazek, T. G., Nichol, S. T., & MacLachlan, N. J. (2009). Rift Valley fever virus [Publisher: American Veterinary Medical Association Section: Journal of the American Veterinary Medical Association]. *Journal of the American Veterinary Medical Association*, *234*(7), 883–893. https://doi.org/10.2460/javma.234.7.883

Bird, B. H., Maartens, L. H., Campbell, S., Erasmus, B. J., Erickson, B. R., Dodd, K. A., Spiropoulou, C. F., Cannon, D., Drew, C. P., Knust, B., et al. (2011). Rift valley fever virus vaccine lacking the nss and nsm genes is safe, nonteratogenic, and confers protection from viremia, pyrexia, and abortion following challenge in adult and pregnant sheep. *Journal of virology*, *85*(24), 12901–12909.

Boshra, H., Truong, T., Nfon, C., Gerdts, V., Tikoo, S., Babiuk, L. A., Kara, P., Mather, A., Wallace, D., & Babiuk, S. (2013). Capripoxvirus-vectored vaccines against livestock diseases in africa. *Antiviral research*, *98*(2), 217–227.

Botros, B., Omar, A., Elian, K., Mohamed, G., Soliman, A., Salib, A., Salman, D., Saad, M., & Earhart, K. (2006). Adverse response of non-indigenous cattle of european breeds to live attenuated smithburn rift valley fever vaccine. *J. Med. Virol.*, *78*(6), 787–791. https://doi.org/10.1002/jmv.20624

Bouloy, M., & Weber, F. (2010). Molecular biology of rift valley fever virus. *The Open Virology Journal*, *4*, 8–14.

Bouloy, M., Janzen, C., Vialat, P., Khun, H., Pavlovic, J., Huerre, M., & Haller, O. (2001). Genetic evidence for an interferon-antagonistic function of rift valley fever virus nonstructural protein nss. *Journal of virology*, *75*(3), 1371–1377.

Boushab, B. M., Fall-Malick, F. Z., Ould Baba, S. E. W., Ould Salem, M. L., Belizaire, M. R. D., Ledib, H., Ould Baba Ahmed, M. M., Basco, L. K., & Ba, H. (2016). Severe human illness caused by rift valley fever virus in mauritania, 2015. *OFIDS*, *3*(4), ofw200. https://doi.org/10.1093/ofid/ofw200

Briese, T., Bird, B., Kapoor, V., Nichol, S. T., & Lipkin, W. I. (2006). Batai and ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in east africa. *JVI*, *80*(11), 5627–5630. https://doi.org/10.1128/JVI.02448-05

Britch, S. C., Binepal, Y. S., Ruder, M. G., Kariithi, H. M., Linthicum, K. J., Anyamba, A., Small, J. L., Tucker, C. J., Ateya, L. O., Oriko, A. A., Gacheru, S., & Wilson, W. C. (2013). Rift Valley Fever Risk Map Model and Seroprevalence in Selected Wild Ungulates and Camels from Kenya [Publisher: Public Library of Science]. *PLOS ONE*, *8*(6), e66626. https://doi.org/10.1371/journal.pone.0066626

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Caplen, H., Peters, C. J., & Bishop, D. H. (1985). Mutagen-directed attenuation of rift valley fever virus as a method for.pdf. *J. gen. Virol*, *66*, 2271–2277.

Caron, A., Miguel, E., Gomo, C., Makaya, P., Pfukenyi, D. M., Foggin, C., Hove, T., & de Garine-Wichatitsky, M. (2013). Relationship between burden of infection in ungulate populations and wildlife/livestock interfaces. *Epidemiology & Infection*, *141*(7), 1522–1535.

Carrillo, C., Lu, Z., Borca, M. V., Vagnozzi, A., Kutish, G. F., & Rock, D. L. (2007). Genetic and phenotypic variation of foot-and-mouth disease virus during serial passages in a natural host [Publisher: American Society for Microbiology]. *Journal of Virology.* https://doi.org/10.1128/JVI.00930-07

Carroll, S. A., Reynes, J.-M., Khristova, M. L., Andriamandimby, S. F., Rollin, P. E., & Nichol, S. T. (2011). Genetic evidence for Rift Valley fever outbreaks in Madagascar resulting from virus introductions from the East African mainland rather than enzootic maintenance [Edition: 2011/04/20 Publisher: American Society for Microbiology]. *J Virol*, *85*(13), 6162–6167. https://doi.org/10.1128/JVI.00335-11

CDC. (2019). Rift valley fever (rvf): Outbreak summaries. Retrieved January 27, 2021, from https://www.cdc.gov/vhf/rvf/outbreaks/summaries.html

CDC, A. (2020). Eastern africa rcc. https://africacdc.org/rcc/eastern-africa-rcc/

Cêtre-Sossah, C., Zeller, H., Grandadam, M., Caro, V., Petinelli, F., Bouloy, M., Cardinale, E., & Albina, E. (2012). Genome analysis of rift valley fever virus, mayotte. *Emerg. Infect. Dis.*, *18*(6). https://doi.org/10.3201/eid1806.110994

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

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, snpsift. *Frontiers in genetics*, *3*, 35.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

Clark, M. H. A., Warimwe, G. M., Di Nardo, A., Lyons, N. A., & Gubbins, S. (2018). Systematic literature review of rift valley fever virus seroprevalence in livestock, wildlife and humans in africa from 1968 to 2016 (C. M. Barker, Ed.). *PLoS Negl Trop Dis*, *12*(7), e0006627. https://doi.org/10.1371/journal.pntd.0006627

Clements, A. C., Pfeiffer, D. U., Martin, V., & Otte, M. J. (2007). A rift valley fever atlas for africa. *Preventive Veterinary Medicine*, *82*(1), 72–82. https://doi.org/10.1016/j.prevetmed.2007.05.006

Coetzer, J. A. (1977). The pathology of rift valley fever. i. lesions occurring in natural cases in new-born lambs. *Onderstepoort J Vet Res.*, 11.

Coetzer, J. A. (1982). The pathology of rift valley fever. ii. lesions occurring in field cases in adult cattle, calves and aborted foetuses. *Onderstepoort J Vet Res.*, 7.

Coffey, L. L., Vasilakis, N., Brault, A. C., Powers, A. M., Tripet, F., & Weaver, S. C. (2008). Arbovirus evolution in vivo is constrained by host alternation. *Proceedings of the National Academy of Sciences*, *105*(19), 6970–6975. https://doi.org/10.1073/pnas.0712130105

Cook, E. A. J., Grossi-Soyster, E. N., de Glanville, W. A., Thomas, L. F., Kariuki, S., Bronsvoort, B. M. d. C., Wamae, C. N., LaBeaud, A. D., & Fèvre, E. M. (2017). The sero-epidemiology of rift valley fever in people in the lake victoria basin of western kenya (A. K. McElroy, Ed.). *PLoS Negl Trop Dis*, *11*(7), e0005731. https://doi.org/10.1371/journal.pntd.0005731

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods*, *9*(8), 772–772. https://doi.org/10.1038/nmeth.2109

Daubney, R., Hudson, J. R., & Garnham, P. C. (1931). Enzootic hepatitis or rift valley fever. an undescribed virus disease of sheep cattle and man from east africa. *J. Pathol.*, *34*(4), 545–579. https://doi.org/10.1002/path.1700340418

Dellicour, S., Rose, R., Faria, N. R., Lemey, P., & Pybus, O. G. (2016). Seraphim: Studying environmental rasters and phylogenetically informed movements. *Bioinformatics*, *32*(20), 3204–3206.

Dellicour, S., Rose, R., Faria, N. R., Vieira, L. F. P., Bourhy, H., Gilbert, M., Lemey, P., & Pybus, O. G. (2017). Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *Molecular Biology and Evolution*, *34*(10), 2563–2571. https://doi.org/10.1093/molbev/msx176

Dellicour, S., Rose, R., & Pybus, O. G. (2016). Explaining the geographic spread of emerging epidemics: A framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics*, *17*(1), 82. https://doi.org/10.1186/s12859-016-0924-x

Dellicour, S., Troupin, C., Jahanbakhsh, F., Salama, A., Massoudi, S., Moghaddam, M. K., Baele, G., Lemey, P., Gholami, A., & Bourhy, H. (2019). Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages — application to rabies virus spread in iran. *Mol Ecol*, *28*(18), 4335–4350. https://doi.org/10.1111/mec.15222

Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., Kooistra-Smid, A. M., Raangs, E. C., Rosema, S., Veloo, A. C., et al. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *Journal of biotechnology*, *243*, 16–24.

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Diallo, M., Lochouarn, L., Ba, K., Sall, A. A., Mondo, M., Girault, L., & Mathiot, C. (2000). First isolation of the rift valley fever virus from culex poicilipes (diptera: Culicidae) in nature. *The American journal of tropical medicine and hygiene*, *62*(6), 702–704.

Digoutte, J., & Peters, C. (1989). General aspects of the 1987 rift valley fever epidemic in mauritania. *Research in virology*, *140*, 27–30.

Dolskiy, A. A., Grishchenko, I. V., & Yudkin, D. V. (2020). Cell cultures for virology: Usability, advantages, and prospects. *International journal of molecular sciences*, *21*(21), 7978.

Drosten, C., Gottig, S., Schilling, S., Asper, M., Panning, M., Schmitz, H., & Günther, S. (2002). Rapid detection and quantification of rna of ebola and marburg viruses, lassa virus, crimean-congo hemorrhagic fever virus, rift valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-pcr. *Journal of clinical microbiology*, *40*(7), 2323–2330.

Drummond, A. J. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/molbev/msi103

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence (D. Penny, Ed.). *PLoS Biol*, *4*(5), e88. https://doi.org/10.1371/journal.pbio.0040088

Dungu, B., Louw, I., Lubisi, A., Hunter, P., von Teichman, B. F., & Bouloy, M. (2010). Evaluation of the efficacy and safety of the rift valley fever clone 13 vaccine in sheep. *Vaccine*, *28*(29), 4581–4587.

du Plessis, L., & Stadler, T. (2015). Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends in Microbiology*, *23*(7), 383–386. https://doi.org/10.1016/j.tim.2015.04.007

ECDC. (2019). Rift valley fever outbreak in mayotte, france [Accessed: 2021-09-30].

El Mamy, A. B. O., Baba, M. O., Barry, Y., Isselmou, K., Dia, M. L., Hampate, B., Diallo, M. Y., El Kory, M. O. B., Diop, M., Lo, M. M., Thiongane, Y., Bengoumi, M., Puech, L., Plee, L., Claes, F., de La Rocque, S., & Doumbia, B. (2011). Unexpected rift valley fever outbreak, northern mauritania. *Emerg. Infect. Dis.*, *17*(10), 1894–1896. https://doi.org/10.3201/eid1710.110397

Ellis, D. S., Simpson, D. I. H., Stamford, S., & Wahab, K. S. E. A. (1979). Rift valley fever virus: Some ultrastructural observations on material from the outbreak in egypt 1977. *Journal of General Virology*, *42*(2), 329–337. https://doi.org/10.1099/0022-1317-42-2-329

Euler, M., Wang, Y., Nentwich, O., Piepenburg, O., Hufert, F. T., & Weidmann, M. (2012). Recombinase polymerase amplification assay for rapid detection of rift valley fever virus. *Journal of clinical virology*, *54*(4), 308–312.

Evans, A., Gakuya, F., Paweska, J. T., Rostal, M., Akoolo, L., Van Vuren, P. J., Manyibe, T., Macharia, J. M., Ksiazek, T. G., Feikin, D. R., Breiman, R. F., & Kariuki Njenga, M. (2008). Prevalence of antibodies against rift valley fever virus in kenyan wildlife. *Epidemiol. Infect.*, *136*(9), 1261–1269. https://doi.org/10.1017/S0950268807009806

FAO. (2017). Food and agriculture organization of theunited nations (fao of the un). rift valley fever in niger: Risk assessment. [Accessed: 2021-09-30].

Faria, N. R., Quick, J., Claro, I. M., Thézé, J., de Jesus, J. G., Giovanetti, M., Kraemer, M. U. G., Hill, S. C., Black, A., da Costa, A. C., Franco, L. C., Silva, S. P., Wu, C.-H., Raghwani, J., Cauchemez, S., du Plessis, L., Verotti, M. P., de Oliveira, W. K., Carmo, E. H., . . . Pybus, O. G. (2017). Establishment and cryptic transmission of zika virus in brazil and the americas [Number: 7658 Publisher: Nature Publishing Group]. *Nature*, *546*(7658), 406–410. https://doi.org/10.1038/nature22401

Faria, N. R., Suchard, M. A., Rambaut, A., Streicker, D. G., & Lemey, P. (2013). Simultaneously reconstructing viral cross-species transmission history and identifying

the underlying constraints. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120196.

Faye, O., Diallo, M., Diop, D., Bezeid, O. E., Bâ, H., Niang, M., Dia, I., Mohamed, S. A. O., Ndiaye, K., Diallo, D., Ly, P. O., Diallo, B., Nabeth, P., Simon, F., Lô, B., & Diop, O. M. (2007). Rift valley fever outbreak with east-central african virus lineage in mauritania, 2003. *Emerg. Infect. Dis.*, *13*(7), 1016–1023. https://doi.org/10.3201/eid1307.061487

Ferreira, M. A., & Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time markov chains. *Canadian Journal of Statistics*, *36*(3), 355–368.

Fonseca, V., Libin, P. J. K., Theys, K., Faria, N. R., Nunes, M. R. T., Restovic, M. I., Freire, M., Giovanetti, M., Cuypers, L., Nowé, A., Abecasis, A., Deforche, K., Santiago, G. A., Siqueira, I. C. d., San, E. J., Machado, K. C. B., Azevedo, V., Filippis, A. M. B.-d., Cunha, R. V. d., . . . de Oliveira, T. (2019). A computational method for the identification of dengue, zika and chikungunya virus species and genotypes (I. Rodriguez-Barraquer, Ed.). *PLoS Negl Trop Dis*, *13*(5), e0007231. https://doi.org/10.1371/journal.pntd.0007231

Fontenille, D. (1998). New vectors of rift valley fever in west africa. *Emerg. Infect. Dis.*, *4*(2), 289–293. https://doi.org/10.3201/eid0402.980218

for Disease Control, C., (CDC, P., et al. (1998). Rift valley fever–east africa, 1997-1998. *MMWR. Morbidity and mortality weekly report*, *47*(13), 261–264.

Freire, C. C. M., Iamarino, A., Soumaré, P. O. L., Faye, O., Sall, A. A., & Zanotto, P. M. A. (2015). Reassortment and distinct evolutionary dynamics of rift valley fever virus genomic segments. *Sci Rep*, *5*(1), 11353. https://doi.org/10.1038/srep11353

Frey, U. H., Bachmann, H. S., Peters, J., & Siffert, W. (2008). Pcr-amplification of gc-rich regions:'slowdown pcr'. *Nature protocols*, *3*(8), 1312–1317.

Frost, S. D., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., & Bedford, T. (2015). Eight challenges in phylodynamic inference. *Epidemics*, *10*, 88–92. https://doi.org/10.1016/j.epidem.2014.09.001

Gardy, J., & Loman, N. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*, *19*(1), 9–20. https://doi.org/10.1038/nrg.2017.88

Gardy, J., Loman, N., & Rambaut, A. (2015). Real-time digital pathogen surveillance — the time is now. *Genome Biology*, *16*(1), 155. https://doi.org/10.1186/s13059-015-0726-x

Gaudreault, N. N., Indran, S. V., Balaraman, V., Wilson, W. C., & Richt, J. A. (2019). Molecular aspects of rift valley fever virus and the emergence of reassortants. *Virus Genes*, *55*(1), 1–11. https://doi.org/10.1007/s11262-018-1611-y

Gerdes, G. H. (2004). Rift valley fever. *Rev. Sci. Tech*, *23*(2), 613–623.

Gerrard, S. R., & Nichol, S. T. (2007). Synthesis, proteolytic processing and complex formation of n-terminally nested precursor proteins of the rift valley fever virus glycoproteins. *Virology*, *357*(2), 124–133. https://doi.org/10.1016/j.virol.2006.08.002

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., & Suchard, M. A. (2013). Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Molecular biology and evolution*, *30*(3), 713–724.

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., Wohl, S., Moses, L. M., Yozwiak, N. L., Winnicki, S., Matranga, C. B., Malboeuf, C. M., Qu, J., Gladden, A. D., Schaffner, S. F., ... Sabeti, P. C. (2014). Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science*, *345*(6202), 1369–1372. https://doi.org/10.1126/science.1259657

Grace, D. (2019). Infectious diseases and agriculture. *Encyclopedia of food security and sustainability*, 439.

Grad, Y. H., & Lipsitch, M. (2014). Epidemiologic data and pathogen genome sequences: A powerful synergy for public health. *Genome biology*, *15*(11), 1–14.

Grenfell, B. T. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, *303*(5656), 327–332. https://doi.org/10.1126/science.1090727

Grobbelaar, A. A., Weyer, J., Leman, P. A., Kemp, A., Paweska, J. T., & Swanepoel, R. (2011). Molecular epidemiology of rift valley fever virus. *Emerg. Infect. Dis.*, *17*(12), 2270–2276. https://doi.org/10.3201/eid1712.111035

Grubaugh, N. D., Faria, N. R., Andersen, K. G., & Pybus, O. G. (2018). Genomic insights into zika virus emergence and spread. *Cell*, *172*(6), 1160–1162. https://doi.org/10.1016/j.cell.2018.02.027

Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*, *20*(1), 8. https://doi.org/10.1186/s13059-018-1618-7

Grubaugh, N. D., Ladner, J. T., Kraemer, M. U. G., Dudas, G., Tan, A. L., Gangavarapu, K., Wiley, M. R., White, S., Thézé, J., Magnani, D. M., Prieto, K., Reyes, D., Bingham, A. M., Paul, L. M., Robles-Sikisaka, R., Oliveira, G., Pronty, D., Barcellona, C. M., Metsky, H. C., ... Andersen, K. G. (2017). Genomic epidemiology reveals multiple introductions of zika virus into the united states [Number: 7658 Publisher: Nature Publishing Group]. *Nature*, *546*(7658), 401–405. https://doi.org/10.1038/nature22400

Gu, W., Miller, S., & Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. *Annual Review of Pathology: Mechanisms of Disease*, *14*, 319–338.

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.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Gwinn, M., MacCannell, D., & Armstrong, G. L. (2019). Next-generation sequencing of infectious pathogens. *Jama*, *321*(9), 893–894.

Habjan, M., Penski, N., Spiegel, M., & Weber, F. (2008). T7 rna polymerase-dependent and-independent systems for cdna-based rescue of rift valley fever virus. *Journal of General Virology*, *89*(9), 2157–2166.

Hanafi, H., Warigia, M., Breiman, R. F., Godsey, M., Hoel, D., Lutomiah, J., Koka, H., O'Guinn, M., Miller, B., Ochieng, C., Lee, J. S., Schnabel, D., Kioko, E., Richardson, J., & Sang, R. (2010). Rift valley fever virus epidemic in kenya, 2006/2007: The entomologic investigations. *The American Journal of Tropical Medicine and Hygiene*, *83*(2), 28–37. https://doi.org/10.4269/ajtmh.2010.09-0319

Hartman, A. (2017). Rift valley fever. *Clinics in Laboratory Medicine*, *37*(2), 285–301. https://doi.org/10.1016/j.cll.2017.01.004

Hasegawa, M., Kishino, H., & Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

Hassan, O. A., Ahlm, C., Sang, R., & Evander, M. (2011). The 2007 rift valley fever outbreak in sudan (S. Brooker, Ed.). *PLoS Negl Trop Dis*, *5*(9), e1229. https://doi.org/10.1371/journal.pntd.0001229

Hay, S. I., Battle, K. E., Pigott, D. M., Smith, D. L., Moyes, C. L., Bhatt, S., Brownstein, J. S., Collier, N., Myers, M. F., George, D. B., et al. (2013). Global mapping of infectious disease. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120250.

Heinrich, N., Saathoff, E., Weller, N., Clowes, P., Kroidl, I., Ntinginya, E., Machibya, H., Maboko, L., Löscher, T., Dobler, G., & Hoelscher, M. (2012). High seroprevalence of rift valley fever and evidence for endemic circulation in mbeya region, tanzania, in a cross-sectional study (R. J. Kent Crockett, Ed.). *PLoS Negl Trop Dis*, *6*(3), e1557. https://doi.org/10.1371/journal.pntd.0001557

Hijmans, R., & van Etten, J. (2012). *Raster: Geographic analysis and modeling with raster data. r package version 2.0-12.* http://CRAN.R-project.org/package=raster

Holmes, E. C., Dudas, G., Rambaut, A., & Andersen, K. G. (2016). The evolution of ebola virus: Insights from the 2013–2016 epidemic. *Nature, 538*(7624), 193–200. https://doi.org/10.1038/nature19790

Hoogstraal, H., Meegan, J. M., Khalil, G. M., & Adham, F. K. (1979). The rift valley fever epizootic in egypt 1977–1978 2. ecological and entomological studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 73*(6), 624–629. https://doi.org/10.1016/0035-9203(79)90005-1

Houldcroft, C. J., Beale, M. A., & Breuer, J. (2017). Clinical and biological insights from viral genome sequencing. *Nature Reviews Microbiology, 15*(3), 183–192.

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics, 17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Hunter, P., Erasmus, B., & Vorster, J. H. (2002). Teratogenicity of a mutagenised rift valley fever virus (mvp 12) in sheep. *The Onderstepoort journal of veterinary research.*

IGAD. (2012). Identification and mapping of key cross-border livestock routes and markets, services and priority transboundary animal diseases including zoonotics for regional and international trade. Retrieved January 27, 2022, from https://icpald.org/wp-content/uploads/2016/01/Cross-border-livestock-routes-and-markets-TADs-and-zoonoses-study-7.pdf

Ikegami, T. (2012). Molecular biology and genetic diversity of rift valley fever virus. *Antiviral Research, 95*(3), 293–310. https://doi.org/10.1016/j.antiviral.2012.06.001

Ikegami, T., Hill, T. E., Smith, J. K., Zhang, L., Juelich, T. L., Gong, B., Slack, O. A., Ly, H. J., Lokugamage, N., & Freiberg, A. N. (2015). Rift valley fever virus mp-12 vaccine is fully attenuated by a combination of partial attenuations in the s, m, and l segments. *Journal of virology, 89*(14), 7262–7276.

Ikegami, T., & Makino, S. (2009). Rift valley fever vaccines. *Vaccine, 27*, D69–D72.

Ikegami, T., & Makino, S. (2011). The pathogenesis of rift valley fever. *Viruses, 3*(5), 493–519. https://doi.org/10.3390/v3050493

Ikegami, T., Won, S., Peters, C. J., & Makino, S. (2007). Characterization of rift valley fever virus transcriptional terminations. *JVI, 81*(16), 8421–8438. https://doi.org/10.1128/JVI.02641-06

Ikegami, T., Won, S., Peters, C., & Makino, S. (2006). Rescue of infectious rift valley fever virus entirely from cdna, analysis of virus lacking the nss gene, and expression of a foreign gene. *Journal of virology, 80*(6), 2933–2940.

Illingworth, C. J., Roy, S., Beale, M. A., Tutill, H., Williams, R., & Breuer, J. (2017). On the effective depth of viral sequence data. *Virus Evolution*, *3*(2), vex030.

Jones, B. A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M. Y., McKeever, D., Mutua, F., Young, J., McDermott, J., et al. (2013). Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the national academy of sciences*, *110*(21), 8399–8404.

Juma, J. (2022). *Viclara is a bioinformatics analysis pipeline for classification and reference guided assembly of segmented viruses from metagenomics reads obtained on illumina platform. (1.0)*. https://github.com/ajodeh-juma/viclara

Juma, J., Fonseca, V., Konongoi, S. L., van Heusden, P., Roesel, K., Sang, R., Bett, B., Christoffels, A., de Oliveira, T., & Oyola, S. O. (2022). Genomic surveillance of rift valley fever virus: From sequencing to lineage assignment. *BMC genomics*, *23*(1), 1–13.

Kalveram, B., Lihoradova, O., & Ikegami, T. (2011). Nss protein of rift valley fever virus promotes posttranslational downregulation of the tfiih subunit p62. *Journal of virology*, *85*(13), 6234–6243.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773–795.

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. *Mol Biol Evol*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kortekaas, J., Oreshkova, N., Van Keulen, L., Kant, J., Bosch, B., Bouloy, M., Moulin, V., Goovaerts, D., & Moormann, R. (2014). Comparative efficacy of two next-generation rift valley fever vaccines. *Vaccine*, *32*(39), 4901–4908.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A.-E., Humphrey, C. D., Shieh, W.-J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., . . . Anderson, L. J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*, *348*(20), 1953–1966. https://doi.org/10.1056/NEJMoa030781

Kwok, S., Chang, S., Sninsky, J., et al. (1994). A guide to the design and use of mismatched and degenerate. *Genome Res*, *3*, S39–47.

LaBeaud, A. D., Muchiri, E. M., Ndzovu, M., Mwanje, M. T., Muiruri, S., Peters, C. J., & King, C. H. (2008). Interepidemic rift valley fever virus seropositivity, northeastern kenya. *Emerg. Infect. Dis.*, *14*(8), 1240–1246. https://doi.org/10.3201/eid1408.080082

LaBeaud, A. D., Cross, P. C., Getz, W. M., Glinka, A., & King, C. H. (2011). Rift valley fever virus infection in african buffalo (syncerus caffer) herds in rural south africa: Evidence of interepidemic transmission. *The American Journal of Tropical Medicine and Hygiene*, *84*(4), 641–646. https://doi.org/10.4269/ajtmh.2011.10-0187

LaBeaud, A. D., Muiruri, S., Sutherland, L. J., Dahir, S., Gildengorin, G., Morrill, J., Muchiri, E. M., Peters, C. J., & King, C. H. (2011). Postepidemic analysis of rift valley fever virus transmission in northeastern kenya: A village cohort study (T. W. Geisbert, Ed.). *PLoS Negl Trop Dis*, *5*(8), e1265. https://doi.org/10.1371/journal.pntd.0001265

Ladner, J. T., Grubaugh, N. D., Pybus, O. G., & Andersen, K. G. (2019). Precision epidemiology for infectious disease control. *Nat Med*, *25*(2), 206–211. https://doi.org/10.1038/s41591-019-0345-2

Lagare, A., Fall, G., Ibrahim, A., Ousmane, S., Sadio, B., Abdoulaye, M., Alhassane, A., Mahaman, A. E., Issaka, B., Sidikou, F., Zaneidou, M., Bienvenue, B., Djingarey Mamoudou, H., Bailo Diallo, A., Kadadé, G., Testa, J., Boubacar Mainassara, H., & Faye, O. (2019). First occurrence of Rift Valley fever outbreak in Niger, 2016 [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/vms3.135]. *Veterinary Medicine and Science*, *5*(1), 70–78. https://doi.org/10.1002/vms3.135

LARISSA, C. (2022). Novel rift valley fever vaccine dosed in phase i human trial. Retrieved 2022, from https://www.larissa.online/post/novel-rift-valley-fever-vaccine-dosed-in-phase-i-human-trial

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., & Higgins, D. (2007). Clustal w and clustal x version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Larsson, A. (2014). Aliview: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278.

Lartillot, N., & Philippe, H. (2006). Computing bayes factors using thermodynamic integration. *Systematic biology*, *55*(2), 195–207.

Laughlin, L. W., Meegan, J. M., Strausbaugh, L. J., Morens, D. M., & Watten, R. H. (1979). Epidemic rift valley fever in egypt: Observations of the spectrum of human illness. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *73*(6), 630–633.

Le Roux, C. A., Kubo, T., Grobbelaar, A. A., van Vuren, P. J., Weyer, J., Nel, L. H., Swanepoel, R., Morita, K., & Paweska, J. T. (2009). Development and evaluation of a real-time reverse transcription-loop-mediated isothermal amplification assay for rapid detection of rift valley fever virus in clinical specimens. *Journal of Clinical Microbiology*, *47*(3), 645–651.

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., & Suchard, M. A. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2 (N. M. Ferguson, Ed.). *PLoS Pathog*, *10*(2), e1003932. https://doi.org/10.1371/journal.ppat.1003932

Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS computational biology*, *5*(9), e1000520.

Lemey, P., Rambaut, A., Welch, J. J., & Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular biology and evolution*, *27*(8), 1877–1885.

Lemey, P., Salemi, M., & Vandamme, A.-M. (2009). *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing. 2 ed.* Cambridge University Press.

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

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

Libin, P. J. K., Deforche, K., Abecasis, A. B., & Theys, K. (2019). VIRULIGN: Fast codon-correct alignment and annotation of viral genomes. *Bioinformatics*, *35*(10), 1763–1765. https://doi.org/10.1093/bioinformatics/bty851

Lichoti, J. K., Kihara, A., Oriko, A. A., Okutoyi, L. A., Wauna, J. O., Tchouassi, D. P., Tigoi, C. C., Kemp, S., Sang, R., & Mbabu, R. M. (2014). Detection of Rift Valley Fever Virus Interepidemic Activity in Some Hotspot Areas of Kenya by Sentinel Animal Surveillance, 2009–2012 [Publisher: Hindawi]. *Veterinary Medicine International*, *2014*, e379010. https://doi.org/10.1155/2014/379010

Linthicum, K. J., Davies, F. G., Kairo, A., & Bailey, C. L. (1985). Rift valley fever virus (family bunyaviridae, genus phlebovirus). isolations from diptera collected during an inter-epizootic period in kenya. *J Hyg (Lond)*, *95*(1), 197–209. Retrieved February 7, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2129511/

Linthicum, K. J., Anyamba, A., Tucker, C. J., Kelley, P. W., Myers, M. F., & Peters, C. J. (1999). Climate and satellite indicators to forecast rift valley fever epidemics in kenya. *Science*, *285*(5426), 397–400.

Linthicum, K. J., Britch, S. C., & Anyamba, A. (2016). Rift valley fever: An emerging mosquito-borne disease. *Annu. Rev. Entomol.*, *61*(1), 395–415. https://doi.org/10.1146/annurev-ento-010715-023819

Liu, L., Celma, C. C., & Roy, P. (2008). Rift valley fever virus structural proteins: Expression, characterization and assembly of recombinant proteins. *Virology journal*, *5*(1), 1–13.

Ludwig, C., & Wagner, R. (2007). Virus-like particles—universal molecular toolboxes. *Current opinion in biotechnology*, *18*(6), 537–545.

Lutomiah, J., Omondi, D., Masiga, D., Mutai, C., Mireji, P. O., Ongus, J., Linthicum, K. J., & Sang, R. (2014). Blood meal analysis and virus detection in blood-fed mosquitoes collected during the 2006–2007 rift valley fever outbreak in kenya. *Vector-Borne and Zoonotic Diseases*, *14*(9), 656–664.

Madani, T. A., Al-Mazrou, Y. Y., Al-Jeffri, M. H., Mishkhas, A. A., Al-Rabeah, A. M., Turkistani, A. M., Al-Sayed, M. O., Abodahish, A. A., Khan, A. S., Ksiazek, T. G., & Shobokshi, O. (2003). Rift valley fever epidemic in saudi arabia: Epidemiological, clinical, and laboratory characteristics. *Clinical Infectious Diseases*, *37*(8), 1084–1092. https://doi.org/10.1086/378747

Makoschey, B., van Kilsdonk, E., Hubers, W. R., Vrijenhoek, M. P., Smit, M., Wichgers Schreur, P. J., Kortekaas, J., & Moulin, V. (2016). Rift valley fever vaccine virus clone 13 is able to cross the ovine placental barrier associated with foetal infections, malformations, and stillbirths. *PLoS neglected tropical diseases*, *10*(3), e0004550.

Mandell, R. B., & Flick, R. (2010). Rift valley fever virus: An unrecognized emerging threat? *Human vaccines*, *6*(7), 597–601.

Mansfield, K. L., Banyard, A. C., McElhinney, L., Johnson, N., Horton, D. L., Hernández-Triana, L. M., & Fooks, A. R. (2015). Rift valley fever virus: A review of diagnosis and vaccination, and implications for emergence in europe. *Vaccine*, *33*(42), 5520–5531. https://doi.org/10.1016/j.vaccine.2015.08.020

Mariner, J. (2018). *Rift valley fever surveillance* [OCLC: 1130680509].

May, N. L., Dubaele, S., Santis, L. P. D., Billecocq, A., Bouloy, M., & Egly, J.-M. (2004). TFIIH Transcription Factor, a Target for the Rift Valley Hemorrhagic Fever Virus. *Cell*, *116*(4), 541–550. https://doi.org/https://doi.org/10.1016/S0092-8674(04)00132-1

Mbotha, D., Bett, B., Kairu-Wanyoike, S., Grace, D., Kihara, A., Wainaina, M., Hoppenheit, A., Clausen, P.-H., & Lindahl, J. (2018). Inter-epidemic rift valley fever virus seroconversions in an irrigation scheme in bura, south-east kenya. *Transbound Emerg Dis*, *65*(1), e55–e62. https://doi.org/10.1111/tbed.12674

McCrone, J. T., & Lauring, A. S. (2016). Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *Journal of virology*, *90*(15), 6884–6895.

McElroy, A. K., Albariño, C. G., & Nichol, S. T. (2009). Development of a RVFV ELISA that can distinguish infected from vaccinated animals. *Virol J*, *6*(1), 125. https://doi.org/10.1186/1743-422X-6-125

Meegan, J. M. (1979). The rift valley fever epizootic in egypt 1977–1978 1. description of the epizootic and virological studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *73*(6), 618–623. https://doi.org/10.1016/0035-9203(79)90004-X

Mehand, M. S., Al-Shorbaji, F., Millett, P., & Murgue, B. (2018). The WHO r&d blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Research*, *159*, 63–67. https://doi.org/10.1016/j.antiviral.2018.09.009

Metsky, H. C., Matranga, C. B., Wohl, S., Schaffner, S. F., Freije, C. A., Winnicki, S. M., West, K., Qu, J., Baniecki, M. L., Gladden-Young, A., Lin, A. E., Tomkins-Tinch, C. H., Ye, S. H., Park, D. J., Luo, C. Y., Barnes, K. G., Shah, R. R., Chak, B., Barbosa-Lima, G., . . . Sabeti, P. C. (2017). Zika virus evolution and spread in the americas [Number: 7658 Publisher: Nature Publishing Group]. *Nature*, *546*(7658), 411–415. https://doi.org/10.1038/nature22402

Minin, V. N., & Suchard, M. A. (2008a). Counting labeled transitions in continuous-time markov models of evolution. *Journal of mathematical biology*, *56*, 391–412.

Minin, V. N., & Suchard, M. A. (2008b). Fast, accurate and simulation-free stochastic mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1512), 3985–3995.

Mohamed, M., Mosha, F., Mghamba, J., Zaki, S. R., Shieh, W.-J., Paweska, J., Omulo, S., Gikundi, S., Mmbuji, P., Bloland, P., Zeidner, N., Kalinga, R., Breiman, R. F., & Njenga, M. K. (2010). Epidemiologic and clinical aspects of a Rift Valley fever outbreak in humans in Tanzania, 2007 [Publisher: The American Society of Tropical Medicine and Hygiene]. *Am J Trop Med Hyg*, *83*(2 Suppl), 22–27. https://doi.org/10.4269/ajtmh.2010.09-0318

Morrill, J. C., & McClain, D. J. (1996). Epidemiology and pathogenesis of rift valley fever and other phleboviruses. In R. M. Elliott (Ed.), *The bunyaviridae* (pp. 281–293). Springer US. https://doi.org/10.1007/978-1-4899-1364-7_12

Morrill, J., Knauert, F., Ksiazek, T., Meegan, J., & Peters, C. (1989). Rift valley fever infection of rhesus monkeys: Implications for rapid diagnosis of human disease. *Research in Virology*, *140*, 139–146. https://doi.org/10.1016/S0923-2516(89)80091-3

Morrill, J., Mebus, C., & Peters, C. (1997). Safety and efficacy of a mutagen-attenuated rift valley fever virus vaccine in cattle. *American journal of veterinary research*, *58*(10), 1104–1109.

Morrill, J. C., & Peters, C. (2011). Protection of mp-12–vaccinated rhesus macaques against parenteral and aerosol challenge with virulent rift valley fever virus. *Journal of Infectious Diseases*, *204*(2), 229–236.

Morvan, J., Rollin, P., Laventure, S., Rakotoarivony, I., & Roux, J. (1992). Rift valley fever epizootic in the central highlands of madagascar. *Research in Virology*, *143*, 407–415. https://doi.org/10.1016/S0923-2516(06)80134-2

Muga, G. O., Onyango-Ouma, W., Sang, R., & Affognon, H. (2015). Sociocultural and economic dimensions of rift valley fever. *The American journal of tropical medicine and hygiene*, *92*(4), 730.

Muller, R., Saluzzo, J.-F., Lopez, N., Dreier, T., Turell, M., Smith, J., & Bouloy, M. (1995). Characterization of clone 13, a naturally attenuated avirulent isolate of rift valley fever virus, which is altered in the small segment. *The American journal of tropical medicine and hygiene*, *53*(4), 405–411.

Munyua, P., Murithi, R. M., Wainwright, S., Githinji, J., Hightower, A., Mutonga, D., Macharia, J., Ithondeka, P. M., Musaa, J., Breiman, R. F., Bloland, P., & Njenga, M. K. (2010). Rift Valley fever outbreak in livestock in Kenya, 2006-2007 [Publisher: The American Society of Tropical Medicine and Hygiene]. *Am J Trop Med Hyg*, *83*(2 Suppl), 58–64. https://doi.org/10.4269/ajtmh.2010.09-0292

Murithi, R. M., Munyua, P., Ithondeka, P. M., Macharia, J. M., Hightower, A., Luman, E. T., Breiman, R. F., & Njenga, M. K. (2011). Rift valley fever in kenya: History of epizootics and identification of vulnerable districts. *Epidemiol. Infect.*, *139*(3), 372–380. https://doi.org/10.1017/S0950268810001020

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

Nabeth, P. (2001). Rift valley fever outbreak, mauritania, 1998: Seroepidemiologic, virologic, entomologic, and zoologic investigations. *Emerg. Infect. Dis.*, *7*(6), 1052–1054. https://doi.org/10.3201/eid0706.010627

Nanyingi, M. O., Munyua, P., Kiama, S. G., Muchemi, G. M., Thumbi, S. M., Bitek, A. O., Bett, B., Muriithi, R. M., & Njenga, M. K. (2015). A systematic review of rift valley fever epidemiology 1931–2014. *Infection Ecology & Epidemiology*, *5*(1), 28024. https://doi.org/10.3402/iee.v5.28024

Nderitu, L., Lee, J. S., Omolo, J., Omulo, S., O'Guinn, M. L., Hightower, A., Mosha, F., Mohamed, M., Munyua, P., Nganga, Z., Hiett, K., Seal, B., Feikin, D. R., Breiman, R. F., & Njenga, M. K. (2011). Sequential rift valley fever outbreaks in eastern africa caused by multiple lineages of the virus. *The Journal of Infectious Diseases*, *203*(5), 655–665. https://doi.org/10.1093/infdis/jiq004

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10. 1093/molbev/msu300

Nicholas, D. E., Jacobsen, K. H., & Waters, N. M. (2014). Risk factors associated with human rift valley fever infection: Systematic review and meta-analysis. *Trop Med Int Health*, *19*(12), 1420–1429. https://doi.org/10.1111/tmi.12385

Nielsen, S. S., Alvarez, J., Bicout, D. J., Calistri, P., Depner, K., Drewe, J. A., Garin-Bastuji, B., Rojas, J. L. G., Schmidt, C. G., Michel, V., Chueca, M. Á. M., Roberts, H. C., Sihvonen, L. H., Stahl, K., Calvo, A. V., Viltrop, A., Winckler, C., Bett, B., Cetre-Sossah, C., . . . Zancanaro, G. (2020). Rift valley fever – epidemiological update and risk of introduction into europe. *EFSA Journal*, *18*(3), e06041. https://doi.org/https://doi.org/10.2903/j.efsa.2020.6041

Niklasson, B., Peters, C., Bengtsson, E., & Norrby, E. (1985). Rift valley fever virus vaccine trial: Study of neutralizing antibody response in humans. *Vaccine*, *3*(2), 123–127.

Nyakarahuka, L., Balinandi, S., Mulei, S., Kyondo, J., Tumusiime, A., Klena, J., Lutwama, J., & Shoemaker, T. (2019). Ten outbreaks of rift valley fever in uganda 2016-2018: Epidemiological and laboratory findings. *International Journal of Infectious Diseases*, *79*, 4. https://doi.org/10.1016/j.ijid.2018.11.029

Nylander, J. A., Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L. (2008). AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in bayesian phylogenetics. *Bioinformatics*, *24*(4), 581–583. https://doi.org/10. 1093/bioinformatics/btm388

Nyundo, S., Adamson, E., Rowland, J., Palermo, P. M., Matiko, M., Bettinger, G. E., Wambura, P., Morrill, J. C., & Watts, D. M. (2019). Safety and immunogenicity of rift valley fever mp-12 and armp-12Δnsm21/384 vaccine candidates in goats (capra aegagrus hircus) from tanzania. *Onderstepoort Journal of Veterinary Research*, *86*(1), 1–8.

O'Brien, J. D., Minin, V. N., & Suchard, M. A. (2009). Learning to count: Robust estimates for labeled distances between molecular sequences. *Molecular biology and evolution*, *26*(4), 801–814.

OIE. (2022). Rwanda—outbreaks of rift valley fever in cattle and possibly human (oie, april 09, 2022). https://flutrackers.com/forum/forum/emerging-diseases-other-health-threats-alphabetical-i-thru-z/rift-valley-fever/945819-rwanda-outbreaks-of-rift-valley-fever-in-cattle-and-possibly-human-oie-april-09-2022

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin,

A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189

O'Toole, Á. (2022). *Snipit: Summarise snps relative to a reference sequence.* https://github.com/aineniamh/snipit

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi- FASTA alignments. *Microbial Genomics*, 5.

Park, D. J., Dudas, G., Wohl, S., Goba, A., Whitmer, S. L., Andersen, K. G., Sealfon, R. S., Ladner, J. T., Kugelman, J. R., Matranga, C. B., et al. (2015). Ebola virus epidemiology, transmission, and evolution during seven months in sierra leone. *Cell*, *161*(7), 1516–1526.

Paweska, J. T., & Jansen van Vuren, P. (2014). Rift valley fever virus. In *The role of animals in emerging viral diseases* (pp. 169–200). Elsevier. https://doi.org/10.1016/B978-0-12-405191-1.00008-9

Paweska, J. T., Mortimer, E., Leman, P. A., & Swanepoel, R. (2005). An inhibition enzyme-linked immunosorbent assay for the detection of antibody to Rift Valley fever virus in humans, domestic and wild ruminants. *Journal of Virological Methods*, *127*(1), 10–18. https://doi.org/https://doi.org/10.1016/j.jviromet.2005.02.008

Pepin, M., Bouloy, M., Bird, B. H., Kemp, A., & Paweska, J. (2010). Rift valley fever virus (*Bunyaviridae: Phlebovirus*): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet. Res.*, *41*(6), 61. https://doi.org/10.1051/vetres/2010033

Peters, C. J., Liu, C.-T., Anderson, G. W., Morrill, J. C., & Jahrling, P. B. (1989). Pathogenesis of viral hemorrhagic fevers: Rift valley fever and lassa fever contrasted. *Reviews of Infectious Diseases*, *11*, S743–S749. http://www.jstor.org/stable/4454961

Pichlmair, A., Habjan, M., Unger, H., & Weber, F. (2010). Virus-like particles expressing the nucleocapsid gene as an efficient vaccine against rift valley fever virus. *Vector-Borne and Zoonotic Diseases*, *10*(7), 701–703.

Pienaar, N. J., & Thompson, P. N. (2013). Temporal and spatial history of rift valley fever in south africa: 1950 to 2011. *Onderstepoort J Vet Res*, *80*(1), 13 pages. https://doi.org/10.4102/ojvr.v80i1.384

Pittman, P. R., Liu, C., Cannon, T. L., Makuch, R. S., Mangiafico, J. A., Gibbs, P. H., & Peters, C. J. (1999). Immunogenicity of an inactivated rift valley fever vaccine in humans: A 12-year experience. *Vaccine*, *18*(1-2), 181–189.

Pybus, O. G., & Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*, *10*(8), 540–550. https://doi.org/10.1038/nrg2583

Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W., Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P., et al. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the national academy of sciences*, *109*(37), 15066–15071.

Pybus, O. G., Tatem, A. J., & Lemey, P. (2015). Virus evolution and transmission in an ever more connected world. *Proc. R. Soc. B.*, *282*(1821), 20142878. https://doi.org/10.1098/rspb.2014.2878

Quick, J. (2020). nCoV-2019 sequencing protocol v3 (LoCost) v.3. protocols.io. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., et al. (2017). Multiplex pcr method for minion and illumina sequencing of zika and other virus genomes directly from clinical samples. *Nature protocols*, *12*(6), 1261–1276.

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J.-P., Cabeza-Cabrerizo, M., Camino-Sanchez, A., Carter, L. L., ... Carroll, M. W. (2016). Real-time, portable genome sequencing for ebola surveillance. *Nature*, *530*(7589), 228–232. https://doi.org/10.1038/nature16996

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2017). R: A language and environment for statistical computing. r foundation for statistical computing. https://www.R-project.org/

Rabinowitz, P., Gordon, Z., Chudnov, D., Wilcox, M., Odofin, L., Liu, A., & Dein, J. (2006). Animals as sentinels of bioterrorism agents. *Emerging Infectious Diseases*, *12*(4), 647.

Rambaut, A. (2018). FigTree v1.4.4. *FigTree v1.4.4 Institute of Evolutionary Biology. University of Edinburgh, Edinburgh.* Retrieved June 3, 2021, from http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Systematic biology*, *67*(5), 901–904.

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-o-gen). *Virus Evol*, *2*(1), vew007. https://doi.org/10.1093/ve/vew007

Randall, R., Gibbs, C., Aulisio, C., Binn, L., & Harrison, V. (1962). The development of a formalin-killed rift valley fever virus vaccine for use in man. *The Journal of Immunology*, *89*(5), 660–671.

Real, L. A., Henderson, J. C., Biek, R., Snaman, J., Jack, T. L., Childs, J. E., Stahl, E., Waller, L., Tinline, R., & Nadin-Davis, S. (2005). Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. *Proceedings of the National Academy of Sciences*, *102*(34), 12107–12111.

Redding, D. W., Tiedt, S., Lo Iacono, G., Bett, B., & Jones, K. E. (2017). Spatial, seasonal and climatic predictive models of rift valley fever disease across africa. *Phil. Trans. R. Soc. B*, *372*(1725), 20160165. https://doi.org/10.1098/rstb.2016.0165

Rich, K. M., & Wanyoike, F. (2010). An assessment of the regional and national socio-economic impacts of the 2007 rift valley fever outbreak in kenya. *The American journal of tropical medicine and hygiene*, *83*(2 Suppl), 52.

Rohr, J. R., Barrett, C. B., Civitello, D. J., Craft, M. E., Delius, B., DeLeo, G. A., Hudson, P. J., Jouanard, N., Nguyen, K. H., Ostfeld, R. S., et al. (2019). Emerging human infectious diseases and the links to global food production. *Nature sustainability*, *2*(6), 445–456.

Rose, N. H., Sylla, M., Badolo, A., Lutomiah, J., Ayala, D., Aribodor, O. B., Ibe, N., Akorli, J., Otoo, S., Mutebi, J.-P., et al. (2020). Climate and urbanization drive mosquito preference for humans. *Current Biology*, *30*(18), 3570–3579.

Sagulenko, P., Puller, V., & Neher, R. A. (2018). Treetime: Maximum-likelihood phylodynamic analysis. *Virus evolution*, *4*(1), vex042.

Sall, A. A., Macondo, E. A., Sène, O. K., Diagne, M., Sylla, R., Mondo, M., Girault, L., Marrama, L., Spiegel, A., Diallo, M., Bouloy, M., & Mathiot, C. (2002). Use of reverse transcriptase pcr in early diagnosis of rift valley fever. *Clinical and Vaccine Immunology*, *9*(3), 713–715. https://doi.org/10.1128/CDLI.9.3.713-715.2002

Sall, A. A., Zanotto, P. M. d. A., Sene, O. K., Zeller, H. G., Digoutte, J. P., Thiongane, Y., & Bouloy, M. (1999). Genetic reassortment of rift valley fever virus in nature. *J. Virol.*, *73*(10), 8196–8200. https://doi.org/10.1128/JVI.73.10.8196-8200.1999

Samy, A. M., Peterson, A. T., & Hall, M. (2017). Phylogeography of rift valley fever virus in africa and the arabian peninsula. *PLoS neglected tropical diseases*, *11*(1), e0005226.

Sang, R., Arum, S., Chepkorir, E., Mosomtai, G., Tigoi, C., Sigei, F., Lwande, O. W., Landmann, T., Affognon, H., Ahlm, C., & Evander, M. (2017). Distribution and abundance of key vectors of rift valley fever and other arboviruses in two ecologically distinct counties in kenya (B. Bird, Ed.). *PLoS Negl Trop Dis*, *11*(2), e0005341. https://doi.org/10.1371/journal.pntd.0005341

Schirmer, M., Ijaz, U. Z., DAmore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the illumina miseq platform. *Nucleic acids research*, *43*(6), e37–e37.

Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., Weinstock, G., Committee, P. P., on Laboratory Practices of the American Society for Microbiology, C., & of the College of American Pathologists, M. R. C. (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Archives of Pathology and Laboratory Medicine*, *141*(6), 776–786.

Shen, W., Le, S., Li, Y., & Hu, F. (2016). Seqkit: A cross-platform and ultrafast toolkit for fasta/q file manipulation. *PloS one*, *11*(10), e0163962.

Sherman, M. B., Freiberg, A. N., Holbrook, M. R., & Watowich, S. J. (2009). Single-particle cryo-electron microscopy of rift valley fever virus. *Virology*, *387*(1), 11–15. https://doi.org/10.1016/j.virol.2009.02.038

Shoemaker, T. R., Nyakarahuka, L., Balinandi, S., Ojwang, J., Tumusiime, A., Mulei, S., Kyondo, J., Lubwama, B., Sekamatte, M., Namutebi, A., Tusiime, P., Monje, F., Mayanja, M., Ssendagire, S., Dahlke, M., Kyazze, S., Wetaka, M., Makumbi, I., Borchert, J., . . . Lutwama, J. (2019). First laboratory-confirmed outbreak of human and animal rift valley fever virus in uganda in 48 years. *The American Journal of Tropical Medicine and Hygiene*, *100*(3), 659–671. https://doi.org/10.4269/ajtmh.18-0732

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, *7*(1), 539.

Sissoko, D., Giry, C., Gabrie, P., Tarantola, A., Pettinelli, F., Collet, L., D'Ortenzio, E., Renault, P., & Pierre, V. (2009). Rift valley fever, mayotte, 2007–2008. *Emerg. Infect. Dis.*, *15*(4), 568–570. https://doi.org/10.3201/eid1504.081045

Smith, D. R., Johnston, S. C., Piper, A., Botto, M., Donnelly, G., Shamblin, J., Albariño, C. G., Hensley, L. E., Schmaljohn, C., Nichol, S. T., & Bird, B. H. (2018). Attenuation and efficacy of live-attenuated rift valley fever virus vaccine candidates in non-human primates (M. V. G. Lacerda, Ed.). *PLoS Negl Trop Dis*, *12*(5), e0006474. https://doi.org/10.1371/journal.pntd.0006474

Smith, M. R., Schirtzinger, E. E., Wilson, W. C., & Davis, A. S. (2019). Rift valley fever virus: Propagation, quantification, and storage. *Current Protocols in Microbiology*, *55*(1). https://doi.org/10.1002/cpmc.92

Soi, R. K., Rurangirwa, F. R., McGuire, T. C., Rwambo, P. M., DeMartini, J. C., & Crawford, T. B. (2010). Protection of sheep against rift valley fever virus and sheep poxvirus with a recombinant capripoxvirus vaccine. *Clinical and Vaccine Immunology*, *17*(12), 1842–1849.

Soumaré, P. O. L., Freire, C. C. M., Faye, O., Diallo, M., Oliveira, J. V. C. d., Zanotto, P. M. A., & Sall, A. A. (2012). Phylogeography of rift valley fever virus in africa reveals multiple introductions in senegal and mauritania (T. Ikegami, Ed.). *PLoS ONE*, *7*(4), e35216. https://doi.org/10.1371/journal.pone.0035216

Sow, A., Faye, O., Ba, Y., Ba, H., Diallo, D., Faye, O., Loucoubar, C., Boushab, M., Barry, Y., Diallo, M., & Sall, A. A. (2014). Rift valley fever outbreak, southern mauritania, 2012. *Emerg. Infect. Dis.*, *20*(2), 296–299. https://doi.org/10.3201/eid2002.131000

Sow, A., Faye, O., Ba, Y., Diallo, D., Fall, G., Faye, O., Bob, N. S., Loucoubar, C., Richard, V., Dia, A. T., Diallo, M., Malvy, D., & Sall, A. A. (2016). Widespread rift valley fever emergence in senegal in 2013–2014. *Open Forum Infect Dis*, *3*(3), ofw149. https://doi.org/10.1093/ofid/ofw149

Stobart, C. C., & Moore, M. L. (2014). Rna virus reverse genetics and vaccine design. *Viruses*, *6*(7), 2531–2550.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *Virus evolution*, *4*(1), vey016.

Sumaye, R. D., Geubbels, E., Mbeyela, E., & Berkvens, D. (2013). Inter-epidemic transmission of rift valley fever in livestock in the kilombero river valley, tanzania: A cross-sectional survey (B. Bird, Ed.). *PLoS Negl Trop Dis*, *7*(8), e2356. https://doi.org/10.1371/journal.pntd.0002356

Swanepoel, R. (2004). Rift valley fever, in. infectious diseases of livestock, edited by jaw coetzer & rc tustin.

Takehara, K., Min, M.-K., Battlesa, J. K., Sugiyama, K., Emery, V. C., Dalrymple, J. M., & Bishop, D. H. (1989). Identification of mutations in the m rna of a candidate vaccine strain of rift valley fever virus. *Virology*, *169*(2), 452–457.

Tantely, L. M., Fontenille, D., & Boyer, S. (2015). A review of mosquitoes associated with rift valley fever virus in madagascar. *The American Journal of Tropical Medicine and Hygiene*, *92*(4), 722–729. https://doi.org/10.4269/ajtmh.14-0421

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of dna sequences. *Lectures on mathematics in the life sciences*, *17*(2), 57–86.

Team, N. S.-O. I. A. (V. I. (2009). Emergence of a novel swine-origin influenza a (h1n1) virus in humans. *N Engl J Med*, *360*(25), 2605–2615. https://doi.org/10.1056/NEJMoa0903810

Terasaki, K., Ramirez, S. I., & Makino, S. (2016). Mechanistic Insight into the Host Transcription Inhibition Function of Rift Valley Fever Virus NSs and Its Importance in Virulence [Publisher: Public Library of Science]. *PLOS Neglected Tropical Diseases*, *10*(10), 1–22. https://doi.org/10.1371/journal.pntd.0005047

Tirode, F., Busso, D., Coin, F., & Egly, J.-M. (1999). Reconstitution of the transcription factor tfiih: Assignment of functions for the three enzymatic subunits, xpb, xpd, and cdk7. *Molecular cell*, *3*(1), 87–95.

Turner, M. D., & Schlecht, E. (2019). Livestock mobility in sub-saharan africa: A critical review. *Pastoralism*, *9*(1), 1–15.

UNOCHA. (2022). The government of burundi, fao, and who continue to combat rift valley fever (rvf. https://reports.unocha.org/en/country/burundi/card/7q7NfsPS0N/

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115–e115.

Zeller, H. G., Akakpo, A. J., & Ba, M. M. (1995). Rift valley fever epizootic in small ruminants in southern mauritania (october 1993): Risk of extensive outbreaks. *Ann. Soc. Belg. Med. Trop."*, *75*(2), 135–140.

Van der Lugt, J. J., Smit, M., & Coetzer, J. A. (1996). Distribution of viral antigen in tissues of new-born lambs infected with rift valley fever virus.

Verdonschot, P. F., & Besse-Lototskaya, A. A. (2014). Flight distance of mosquitoes (culicidae): A metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. *Limnologica*, *45*, 69–79.

Vialat, P., Billecocq, A., Kohl, A., & Bouloy, M. (2000). The s segment of rift valley fever phlebovirus (bunyaviridae) carries determinants for attenuation and virulence in mice. *Journal of virology*, *74*(3), 1538–1543.

Vialat, P., Muller, R., Vu, T. H., Prehaud, C., & Bouloy, M. (1997). Mapping of the mutations present in the genome of the rift valley fever virus attenuated mp12 strain and their putative role in attenuation. *Virus research*, *52*(1), 43–50.

Viboud, C., Nelson, M. I., Tan, Y., & Holmes, E. C. (2013). Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120199.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L. C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., & de Oliveira, T. (2019). Genome detective: An automated system for virus identification from high-throughput sequencing data (I. Birol, Ed.). *Bioinformatics*, *35*(5), 871–873. https://doi.org/10.1093/bioinformatics/bty695

Volz, E. M., Koelle, K., & Bedford, T. (2013). Viral phylodynamics (S. Wodak, Ed.). *PLoS Comput Biol*, *9*(3), e1002947. https://doi.org/10.1371/journal.pcbi.1002947

Wallace, D., Ellis, C., Espach, A., Smith, S., Greyling, R., & Viljoen, G. (2006). Protective immune responses induced by different recombinant vaccine regimes to rift valley fever. *Vaccine*, *24*(49-50), 7181–7189.

Warimwe, G. M., Lorenzo, G., Lopez-Gil, E., Reyes-Sandoval, A., Cottingham, M. G., Spencer, A. J., Collins, K. A., Dicks, M. D., Milicic, A., Lall, A., et al. (2013). Immunogenicity and efficacy of a chimpanzee adenovirus-vectored rift valley fever vaccine in mice. *Virology journal, 10*(1), 1–9.

Weaver, S. (2006). Evolutionary influences in arboviral disease. *Quasispecies: concept and implications for virology*, 285–314.

WHO. (1983). The use of veterinary vaccines for prevention and control of rift valley fever: Memorandum from a WHO/FAO meeting. *Bulletin ofthe World Health Organization, 61*(2), 261–268.

WHO. (2018). A research and development blueprint for action to prevent epidemics. Retrieved January 27, 2020, from https://www.who.int/blueprint/priority-diseases/en/

WHO. (2020). Covid-19 - china [Accessed: 2020-01-12].

Wichgers Schreur, P. J., Oreshkova, N., van Keulen, L., Kant, J., van de Water, S., Soós, P., Dehon, Y., Kollár, A., Pénzes, Z., & Kortekaas, J. (2020). Safety and efficacy of four-segmented rift valley fever virus in young sheep, goats and cattle. *npj Vaccines, 5*(1), 65.

Wichgers Schreur, P. J., Oymans, J., Kant, J., van de Water, S., Kollár, A., Dehon, Y., Soós, P., Pénzes, Z., van Keulen, L., & Kortekaas, J. (2021). A single vaccination with four-segmented rift valley fever virus prevents vertical transmission of the wild-type virus in pregnant ewes. *npj Vaccines, 6*(1), 8.

Williams, R., Ellis, C. E., Smith, S. J., Potgieter, C. A., Wallace, D., Mareledwane, V. E., & Majiwa, P. A. O. (2011). Validation of an IgM antibody capture ELISA based on a recombinant nucleoprotein for identification of domestic ruminants infected with Rift Valley fever virus. *Journal of Virological Methods, 177*(2), 140–146. https://doi.org/https://doi.org/10.1016/j.jviromet.2011.07.011

Wilson, M. L., Chapman, L. E., Hall, D. B., Dykstra, E. A., Ba, K., Zeller, H. G., Traore-Lamizana, M., Hervy, J.-P., Linthicum, K. J., & Peters, C. J. (1994). Rift valley fever in rural northern senegal: Human risk factors and potential vectors. *The American Journal of Tropical Medicine and Hygiene, 50*(6), 663–675. https://doi.org/10.4269/ajtmh.1994.50.663

WOAH. (2022). *Rift valley fever. in: Manual of diagnostic tests and vaccines for terrestrial animals 2022.* Version Version 2.2.x. WOAH.

Wohl, S., Schaffner, S. F., & Sabeti, P. C. (2016). Genomic analysis of viral outbreaks. *Annu. Rev. Virol., 3*(1), 173–195. https://doi.org/10.1146/annurev-virology-110615-035747

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M.-H. (2011). Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic biology, 60*(2), 150–160.

Yang, Z. (1994). Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. *Systematic biology*, *43*(3), 329–342.

Youssouf, H., Subiros, M., Dennetiere, G., Collet, L., Dommergues, L., Pauvert, A., Rabarison, P., Vauloup-Fellous, C., Le Godais, G., Jaffar-Bandjee, M.-C., et al. (2020). Rift valley fever outbreak, mayotte, france, 2018–2019. *Emerging Infectious Diseases*, *26*(4), 769.

Yu, F., Adungo, F., Konongoi, S. L., Inoue, S., Sang, R., Ashur, S., Kwallah, A. o., Uchida, L., Buerano, C. C., Mwau, M., Zha, Y., Nie, Y., & Morita, K. (2018). Comparison of enzyme-linked immunosorbent assay systems using rift valley fever virus nucleocapsid protein and inactivated virus as antigens. *Virology Journal*, *15*(1), 178. https://doi.org/10.1186/s12985-018-1071-y

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in saudi arabia. *N Engl J Med*, *367*(19), 1814–1820. https://doi.org/10.1056/NEJMoa1211721