MOLECULAR CHARACTERIZATION OF CHIKUNGUNYA VIRUS.

ANTHONY NDIRANGU KAMAU

MASTER OF SCIENCE
(Biochemistry)

JOMO KENYATTA UNIVERSITY OF

AGRICULTURE AND TECHNOLOGY.

2008

Molecular Characterization Of Chikungunya Virus.

Anthony Ndirangu Kamau

A Thesis Submitted in Partial Fulfillment for the Degree of Master of Science in Biochemistry in the Jomo Kenyatta University of Agriculture and Technology.

2008

DECLARATION

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

Signature:…… Date:…….

Anthony Ndirangu Kamau

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

1. Signature:…… Date:…….

Dr. Moses Kariuki Njenga,

CDC/ IEIP-Kenya

2. Signature:…… Date:……...

Prof. Gabriel Magoma

JKUAT, Kenya.

DEDICATION

This work is dedicated to all those who suffered from Chikungunya fever outbreaks along the Indian Ocean coastlines, (2004-2006).

To my beloved parents George and Lucy Kamau for their distinguished moral and financial support which was a great motivation to complete this work.

ACKNOWLGEMENT

Special thanks to the Almighty God, who gives knowledge and skill in all learning and wisdom and who has brought me this far. To Him be all praise and glory.

My appreciation goes to all teaching and non-teaching staff of the department of biochemistry, JKUAT. They worked hard to ensure that I succeeded.

I wish to express my heartfelt gratitude to my supervisors, Dr. Kariuki M. Njenga and Prof. Gabriel Magoma for the invaluable guidance and support they offered me through the course of this project.

This work described was undertaken at the Centers for Disease Control and Prevention at the Kenya Medical Research Institute, Nairobi. The project was fully funded by International Emerging Infections Program-Kenya. My regards go to them for making the work possible.

My appreciation is to Dr. Rosemary Sang, Dr. Robert Breiman, Dr. Njenga and others who were involved in the collection of the samples from the outbreak areas.

I do acknowledge the dedication and support of Leonard Nderitu, who was instrumental in PCR trouble-shooting and sequencing, Victor Ofula who assisted in the ELISA tests, George K'opiyo and Marion Warigia for their technical assistance. My thanks also go to the IEIP team; Victor Ombala, Rose Wanjala, Solomon Gikundi, Gilbert Kikwai, Dr. Lavoisier Akoolo, Lilian Sheunda, Sylvia Omulo, Newton Wamola, Jackson Musembi, Cyrus Wachira for their kind support in many ways. I wish to thank Dorine Bonyo for administrative support.

Of great spiritual and moral support were my pastors Charles and Nancy Meliyio, Charles and Grace Wanyeki, my regards go out to them as well. I also acknowledge Richard and Glory Gathigi for their assistance in typing and printing facilties. Thanks too to the entire Gospel Revival Center-Kiserian fellowship for they were a great encouragement to me.

I may not have made it if without the tremendous support of my family; Beloved Dad George Kamau and Mom, Lucy Kamau, my brothers Stephen Ngugi and his wife Mary, Samuel Mungai and Susan, my only sister Hildah Wanjiku and little Wanjiru, Kamau and Abby. To all of them I say thank you and God bless you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

A	Adenine
ABTS	2,2′-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)
AW1	Wash buffer 1
AW2	Wash buffer 2
AVL	Lysis buffer
AVE	Elution buffer
C	Cytosine
CDC	Centers for Disease Control and Prevention
cDNA	Complementary DNA
CO_2	Carbon dioxide gas
ddT	dideoxythymidine
ddTTP	Dideoxythymidine triphosphate
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleoside triphosphates
DTT	Dithiothreiotol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylene diamine tetra-acetic acid.
FBS	Fetal Bovine Serum
G	Guanine
HIV	Human immunodeficiency virus
IEIP	International Emerging Infections Program
IgG	Immunoglobulin G (gamma)

IgM	Immunoglobulin M (mu)
PE	Wash buffer
pH	Measure of acidity or basicity of a solution
QG	Solubilization and binding buffer
RNA	Ribonucleic acid
mRNA	messenger RNA
SDS	Sodium dodesyl sulphate
TAE	Tris, acetate and EDTA buffer
Taq	*Thermus aquaticus*
µl.	Micro liter.

ABSTRACT

Outbreaks of Chikungunya fever, a viral disease transmitted by Aedes spp. Mosquito emerged in a number of islands along the Indian Ocean and India in late 2004. Initially, the disease emerged in Kenya's Lamu islands before occurring in Mombasa, both cases being in late 2004. Early 2005, epidemics of the same virus were experienced in the Comoros islands. The Chikungunya fever has re-emerged as an important health problem, affecting many people and causing increasingly severe symptoms. Mutation of the virus, lack of vector control, and globalization of trade and travel might have contributed to the resurgence of the infection. The study was aimed at characterizing the specific strains of Chikungunya present in the Lamu, Mombasa (KPA) and Comoros outbreaks. ELISA tests were performed to detect presence of the virus in patients' sera. From the positive samples, the virus was cultured by inoculation into vero cells. RNA was then isolated from the cultured virus which was reverse transcribed into cDNA and amplified exponentially by PCR. The cDNA was cycle sequenced by termination synthesis and the nucleotide sequences analysed by an automated genetic analyzer. Nucleotide sequence data thus generated was used to detect any differences with reference to the already sequenced S27 African prototype whole genome. The genotypes responsible for the infections were known and their relatedness to each other and to other strains in the world was established by performing phylogenetic analysis. There was over 99% nucleotide sequence homology amongst Comoros 25, KPA 15 and Lamu 33 isolates and they all had about 95% similarity with the S27 African prototype.

CHAPTER ONE: INTRODUCTION & LITERATURE REVIEW

1.1 Introduction

Chikungunya virus (CHIKV) is a re-emerging virus of public health importance in Africa and Asia. CHIKV is a member of the genus Alphavirus in the family Togaviridae, which also includes O'nyong'nyong virus (ONN). This group of viruses collectively forms the Semliki Forest antigenic complex (Blackburn *et al*., 1995; Powers *et al.,* 2000).The complex comprises of positive-strand, enveloped RNA viruses that replicate in the cytoplasms of a variety of vertebrate and insect cells. Studies suggest that nascent RNA molecules of this complex are probably utilized for translation and nucleocapsid assembly (Pekka *et al*., 2001).

CHIKV was first isolated from the blood of a febrile (fever) patient in Tanzania in 1952 (Lumsden, 1955; Marion, 1955). Since then, CHIKV has caused major epidemics in both Africa and Southeast Asia (Powers *et al.,* 2000), and most recently in Kenya in 2004, (Breiman- personal communication).

In Africa, CHIKV is maintained in a sylvatic transmission cycle involving primates and forest-dwelling Aedes mosquitoes (Diallo *et al*., 1999). The most prominent clinical feature of CHIKV is arthralgia (general body ache), which can be debilitating and prolonged (Sam and AbuBakar, 2006).

Chikungunya virus has been responsible for explosive outbreaks in the Indian Ocean in 2006 (Ligon 2006; Pialoux *et al*., 2006; Schuffenecker *et al*., 2006). The first reports of the epidemic in the Indian Ocean occurred during August 2004 when the residents of Lamu, Kenya acquired CHIKV illness. Later, the virus spread to Mombasa, Kenya before been detected in Union of the Comoros in early 2005 (Breiman, 2005). From Comoros, the virus caused phenomenal outbreaks in the island of Reunion (Ligon 2006; Pialoux *et al*., 2006; Schuffenecker *et al*., 2006).

An important question is the origin of the virus that initiated the outbreaks in the Indian Ocean. To investigate whether whether mutations in the virus or novel sequences had resulted in increased virulence of the virus, the complete genome sequencing of isolates obtained from the Lamu, Mombasa and Comoros outbreaks was performed and the genetic sequences compared with the S27 strain.

1.2 LITERATURE REVIEW

1.2 .0 alphaviruses

Alphaviruses are a genus belonging to the group IV Togaviridae family of viruses characterized by a positive sense single stranded RNA having icosahedral symmetry. Virions range from 60nm to 65nm in diameter. Positive-sense viral RNA signifies that a particular viral RNA sequence may be directly translated into the desired viral proteins. Therefore, in positive-sense RNA viruses, the viral RNA genome is identical to the viral mRNA, and can be immediately translated by the host cell. Because of this, these viruses do not need to have an RNA transcriptase packaged into the virion.

Negative-sense viral RNA is complementary to the viral mRNA and thus must be converted to positive-sense RNA by RNA polymerase prior to translation. Negative-sense RNA (like DNA) has a nucleotide sequence complementary to the mRNA that it encodes and thus cannot be translated into protein directly. Instead, it must first be transcribed into a positive-sense RNA which acts as mRNA. Some viruses (Influenza, for example) have negative-sense genomes and so must carry an RNA transcriptase inside the virion (Prescott, 1993).

Generally the alphaviruses have a wide geographic distribution, having been isolated in all continents and from many islands except Antarctica. However, individual alphaviruses have more limited distribution.

The alphaviruses have 26 currently recognized members that are classified into 6 antigenic groups; Venezuelan equine encephalitis group, (VEE), Eastern equine

encephalitis group,(EEE), Semliki forest group, Sinbis group, Recombinant or uncertain group and ungrouped, (Strauss and Strauss, 1994).

The American or eastern equine encephalitides contains 4 viruses which produce encephalitis. These were first isolated from the brains of dead horses in the 1930s. Focal epidemics of eastern equine encephalitis virus have occurred from time to time in the eastern USA. Western equine encephalitis virus, (WEE), of the Sinbis antigenic group is endemic in the Western USA. VEE is endemic in S. America and occasionally causes disease in N. America while the Everglades virus is related to EEE but is restricted to the state of Florida in the USA. Only 5 cases of encephalitis associated with Everglades virus have been reported, despite a high seroprevalence rate in Florida. EEE and WEE are maintained in nature between mosquitoes and birds while VEE between mosquitoes and rodents. The strain of VEE is mainly amplified in horses, producing equine disease prior to the beginning of human disease. This is in contrast to the EEE and WEE viruses, where horses appear to be a dead end host (Strauss and Strauss, 1994; Weston *et al*., 2002).

Each virus in turn consists of numerous geographical variants or strains. Inclusion in the genus Alphavirus is made based on serological cross-reaction with one or more of the existing members of the group. Complete or partial sequences of more than 10 alphaviruses are available, and all share a minimum amino acid sequence identity of approximately 45% in the most divergent structural proteins and approximately 60% in the non-structural proteins.

Eleven alphaviruses are known to be pathogenic for humans, (Strauss and Strauss, 1994). In addition, alphaviruses can infect other vertebrates such as rodents, birds and larger mammals such as horses as well as invertebrates (Weston *et al.*, 2002).

Inter and intra-species transmission occurs via mosquitoes or other hematophagous arthropods making alphviruses one of the arboviruses or arthropod borne viruses, (Strauss and Strauss, 1994).

1.2.1 Chikungunya virus

The virus is the etiologic agent of Chikungunya fever.The name Chikungunya is derived from the Makonde word meaning "that which bends up" in reference to the stooped posture developed as a result of the arthritic symptoms of the disease. The disease was first described by Marion and Lumsden in 1955, following an outbreak on the Makonde Plateau, along the border between Tanganyika and Mozambique, in 1952. The virus was in 1997 given a systematic name as the Chikungunya virus and is alternatively known as the Buggy Creek virus. It has an Acronym, CHIKV, according to the International Committee on Taxonomy of Viruses (Büchen-Osmond, 2004).

Serologically, it is most closely related to ONN. CHIKV is believed to be enzootic throughout much of Africa. Historical evidence demonstrates that CHIKV originated in Africa where it is maintained in a sylvatic cycle involving wild primates and forest-dwelling mosquitoes such as *Aedes furcifer*.

It was subsequently introduced into Asia including India, Sri Lanka, Myanmar, Thailand, Indonesia, the Philippines and Malaysia where, as in the African urban cycle, it is transmitted by *Ae. Aegypti* mosquitoes. Phylogenetic studies support this theory, with CHIKV strains falling into three distinct genotypes based on origin from West Africa, Central/East Africa or Asia, the latter group including Malaysian isolates from the Klang outbreak.

A distinctive feature of CHIKV is that it causes explosive outbreaks, before apparently disappearing for a period of several years to decades. This is in contrast to the endemic nature of dengus, which shares with CHIKV the same mosquito vectors of *Aedes aegypti* and *Ae. Albopictus* (1263 bp) (Sam and AbuBakar, 2006). Strains from Africa and Asia are reported to differ biologically, indicating that distinct lineages may exist (Powers *et al.*, 2000; Dana *et al.*, 2005).

ONN also causes a disease similar to Chikungunya characterized by fever (38°C), periocular erythema, symmetrical polyarthralgia, lymphadenopathy, generalized paular or maculopapular enxanthema, joint pain and pharyngitis (Dana *et al.*, 2005; Bessaud *et al*., 2006). However, ONN is transmitted by anopheles not culex mosquitoes, typically *Anopheles funestus* and *Anopheles gambiae* (Powers *et al*., 2000). In addition to the difference in the mode of transmission, CHIKV and ONN appear to have serological differences, e.g. during an infection of ONN in Chad in 2004, serum samples collected during and after the acute phase were CHIKV negative by enzyme linked immunosorbent assay (ELISA) for immunoglobulin M (IgM) and IgG antibodies to CHIKV by IgM-antibody capture (MAC-ELISA) and antigen-capture ELISA respectively.

1.2.2 Genome organization

Alphaviruses possess conserved sequences at the 5' and 3' ends as well as the intergenic region. Conserved repeated sequence elements (RSEs) are also present in the 3' non-translated region (NTR) among alphaviruses. These conserved domains play an important role in the regulation of viral RNA synthesis (Khan *et al.*, 2002). Complete genome sequences of the Ross (Logue and Atkins, 2002), S27 (Khan *et al.*, 2002), strain LR2006_OPY1 (Parola, *et al.*, 2006) and strain D570/06 (Chamberlain *et al.*, 2006) have been completed.

CHIKV and ONN have similar genomic organization and length (Levinson *et al.*, 1990; Khan *et al.*, 2002). The genome size is 11805 nucleotides, excluding the 5' cap nucleotide, an internal poly (A) tract and the 3' poly (A) tail. The genome comprises of two long open reading frames that encode the non-structural and structural polyproteins.

Nonstructural proteins (nsP) are translated as a polyprotein (nsP1 to -4) from the genomic 42S RNA, and they form essential components of viral RNA replication and transcription complexes. The nsP1 protein is a methyl- and guanylyltransferase, whereas the nsP2 is a proteinase and nucleoside triphosphatase. The nsp3 gene product is a phosphoprotein, which functions together with nsP1 in anchoring the replication complex proteins to cytoplasmic membrane structures. Mutations in the nsP3 protein result in blockage of RNA synthesis, indicating the importance of this protein or the polyprotein component in replication, although the exact mechanism of action remains unknown.

The nsp4 gene displays high similarity to the RNA-dependent polymerase sequences of other RNA viruses (Minna *et al*., 2000).

A subgenomic positive-strand RNA referred to as 26S RNA, identical to the 3' one-third of the genomic RNA, is transcribed from a negative-stranded RNA intermediate. This RNA serves as the mRNA for the synthesis of the viral structural proteins; envelope (E) glycoproteins and capsid (C).According to the genomic organization of other alphaviruses, the genome of CHIKV is considered to be: 5' cap-nsP1-nsP2-nsP3-nsP4-(junction region)-C-E3-E2-6K-E1-poly(A) 3' (Khan *et al*., 2002).

There are secondary structures in the 5´- NTR (non-translated region) and repeated sequence elements (RSEs) within the 3´ NTR. The existence of I-poly (A) fragments with different lengths at identical initiation positions for each clone strongly suggests that the polymerase of the alpha sp. virus has the capacity to create poly (A) (Khan *et al.*, 2002).

1.2.3 Transmission of CHIKV

The isolation of CHIKV from non-human primates and other vertebrates supports the existence of sylvatic transmission cycles in Africa, which may maintain the virus in the wild during inter-epidemic years (Lenglet *et al*., 2006).

There is evidence of mosquitoes as vectors for CHIKV. Chikungunya virus strains have been isolated from forest mosquitoes in Senegal, with most of them isolated from *Ae. furcifer-taylori, Ae. luteocephalus,* and *Ae. dalzieli*. The characteristics of the sylvatic transmission cycle are a circulation periodicity with silent intervals that last approximately three years (Diallo *et al.,* 1999). The Asian tiger mosquito Aedes (stegomyia) albopictus (Skuse) has played a major role in transmission of Chikungunya, a recent example being the outbreak of Chikungunya on the Indian Ocean island of Reunion (Pages *et al.,* 2006).

In Asia, transmission appears to be mainly from *Aedes aegypti* and *Ae. albopictus* to human in urban settings. No animal reservoirs have been definitively identified, although the presence of neutralizing antibodies in Malaysian monkeys suggests that these primates may be hosts.

There may also be mother-to-child transmission (Lenglet *et al*., 2006). In the Reunion Island, from June 2005 through the end of January 2006, 84 pregnant women had acute Chikungunya infections during pregnancy. In 88% of these cases all involving infections relatively distant from delivery--the newborns appeared asymptomatic. Conversely, 10 newborns had severe attacks (4 with meningoencephalitis and 3 with intravascular coagulations) after birth and required

prolonged neonatal hospitalization (6 in the neonatal intensive care unit with intubation and assisted ventilation). No infants died, but there was one case of severe intracerebral hemorrhage after severe thrombocytopenia. These cases were confirmed by specific serology testing or PCR or both for mothers and newborns. This was the first time reported maternal-fetal transmission of this virus (Robillard *et al.*, 2006).

Transovarial transmission of CHIKV has not been demonstrated. Different geographical strains of Aedes mosquitoes vary in their susceptibility to infection and ability to transmit the virus, which may be critical in determining CHIKV endemicity in a given area. This episodic nature of CHIKV outbreaks still cannot be explained, but likely depends on interplay of factors, including human and vector susceptibility to infection, high density of mosquito vectors and the introduction of virus from other endemic areas. The latter has become increasingly likely in this age of increased travel by, for example, immigrants and tourists (Sam and AbuBakar, 2006).

1.2.4 Epidemiology of CHIKV

CHIKV disease occurs in Africa and Asia. Since the 1952 Tanzania outbreak, CHIKV has caused outbreaks in East Africa (Tanzania and Uganda), in Austral Africa (Zimbabwe and South Africa), in West Africa (Senegal and Nigeria), and in Central Africa (Central African Republic and Democratic Republic of the Congo).

Since the first documented Asian outbreak in 1958 in Bangkok, Thailand, outbreaks have been documented in Thailand, Cambodia, Vietnam, Laos, Myanmar, Malaysia, Philippines and Indonesia (Kit, 2002; Laras *et al.*, 2005; Isabelle *et al*., 2006). The first reported case of virus in Thailand was in 1960 (Thaikruea *et al.*, 1997). The virus has also caused numerous large outbreaks in India. No active or passive surveillance has been carried out since the last epidemic which occurred in 1971 (Hunderkar *et al*., 2002).

CHIKV has been found to be re-emerging. In both Africa and Asia, the re-emergence was unpredictable, with intervals of 7-8 years to 20 years between consecutive epidemics (Kit, 2002; Laras *et al.*, 2005; Isabelle *et al*., 2006). Since 1960, CHIKV in Thailand was again reported in 1991. However, the disease surveillance system does not specifically include cases and the signs and symptoms are similar to those of dengue fever/dengue hemorrhagic fever, rubella, and fever of unknown origin; thus cases might often be reported under those diagnoses. During the rainy season of 1995 (Jun-Aug), there were at least 2 reported outbreaks which might indicate that it is a re-emerging disease in Thailand (Thaikruea *et al.*,1997).

Few epidemics of this disease have been reported in Senegal. The most recent one occurred in 1996 in Kaffrine where two Chikungunya virus strains were isolated from *Ae. Aegypti* (Diallo *et al*., 1999). Two outbreaks have been reported in Malaysia, in Klang, Selangor (1998) and Bagan Panchor, Perak (2006). The Klang outbreak was the first time that CHIKV was isolated and reported to cause clinical disease in Malaysia.

Earlier studies in Malaysia showed only the presence of CHIKV antibodies in the human population in northern and eastern States bordering Thailand, where CHIKV is known to be present. Seropositivity has also been found in people in East Malaysia, especially among immigrants from neighboring countries. This suggests that CHIKV has been in existence in certain parts of Malaysia, and that transmission was probably low-level, sporadic and undiagnosed (Kumarasamy *et al*., 2006; Sam and AbuBakar, 2006).

Twenty-four distinct outbreaks of etiology were identified throughout Indonesia from September 2001 to March 2003, after a near 20-year hiatus of epidemic CHIKV activity in the country. Thirteen outbreak reports were based on clinical observations alone, and 11 confirmed by serological/virological methods. Both outbreaks started in the rainy season following increased *Aedes aegypti* and *A. albopictus* densities (Kit, 2002; Laras *et al.*, 2005; Isabelle *et al*., 2006).

The resurgence of Chikungunya virus is described during an urban epidemic in Kinshasa Democratic Republic of the Congo, after 39 years without any isolation of the virus (Pastorino *et al*., 2004). Recent reports of large scale outbreaks of fever caused by virus infection in several parts of Southern India have confirmed the re-emrgence of this virus (Ravi, 2006).

A virus outbreak of unprecedented magnitude occurred in Indian Ocean territories in 2006. In Réunion Island, this alphavirus infected about one-third of the human population (Ligon, 2006; Pialoux *et al*., 2006; Schuffenecker *et al*., 2006).

The epidemic of Chikungunya fever that affected the population of La Reunion since Christmas 2005 will be remembered as the most serious public health crises in the island's history (Boutin, 2006).

Kenya outbreak- Lamu

The Lamu Islands are located in the Indian Ocean close to the northern coast of Kenya. They lie between the towns of Lame and Kiunga, close to the border with Somalia, and are a part of Lamu District, (Wikipedia Encyclopedia). High temperatures and humidty prevail year round; peak rains occur from April through July, (http://www.geographyiq.com/countries/ke/Kenya_climate_f.htm).

Lamu has become a major tourist destination both because of its rich cultural heritage and its beaches. Access to the island is by ferry.

In August 2004, Lamu District Hospital noted a dramatic surge in diagnoses of malaria, and suspected that another etiology might have been responsible for the febrile illnesses. Serum specimens were sent to the Arbovirology/Viral Hemorrhagic Fever Laboratory at the Kenya Medical Research Institute. IgM antibodies to alphaviruses were identified in three of the samples and an investigation was carried out on the outbreak. The subsequent description is based on 58 confirmed cases of Chikungunya illness.

The most common symptoms were joint pain, decreased mobility, and headache, which was retro-orbital in most. Arthralgia was the first presenting symptom for the majority of patients; some identified fever as the first symptom and stil others experienced a combination of symptoms simultaneously. At least 71% had missed school or work because of their illness (Breiman, 2005).

Comoros outbreak

Since the end of 2004, CHIKV emerged in the islands of the south-western Indian Ocean. More than 5,000 cases were reported in Comoros, between January and March. Later in 2005, the virus circulated in the other islands, namely Mayotte, Seychelles, Reunion and Mauritius (Isabelle *et al.*, 2006). Active case surveillance was initiated in March, 2005. Retrospective surveillances showed a steady increase in suspected cases in February, 2005. The number of reported cases peaked towards end of March, 2005 and gradually declined thereafter. There were 5,202 suspected cases reported with no case-related mortality. During the surveillance period, there were also reports of 206 suspected cases on the island Anjouan and one on Mohéli. Treatment of suspected CHIKV infections included paracetamol, herbal treatments, antimalarials, NSAIDs, antibiotics, and aspirin. By definition, all of suspected case-patients were febrile. Joint pain, often severe, was the most commonly reported symptom in suspected case-patients. Many patients had the stooped ambulation pattern which is characteristic of CHIKV infection (Lumsden, 1955; Marion, 1955).

More than half of the patients experienced limited range of joint movement. Affected joints included knee, wrist, ankle, shoulder, elbow, fingers, and hip. 72% case-patients missed work because of their illnesses. Many reported that at least one other member of the household was ill with similar symptoms (Breiman, 2005).

1.2.5 Symptoms of infection

The symptoms of Chikungunya include fever which can reach 39°C, (102.2 °F) a petechial or maculopapular rash usually involving the limbs and trunk, and arthralgia or arthritis affecting multiple joints which can be debilitating. Other symptoms included headache, conjunctival infection and slight photophobia. Fever typically lasts for two days and abruptly comes down; however joint pain, intense headache, insomnia and an extreme degree of prostration lasts for a variable period, usually for about 5 to 7 days (Thaikruea *et al.*, 1997; Fourcade *et al.*, 2006). severe forms, central neurological involvement, hepatic cytolyse, severe lymphopenia, fulminant hepatitis, severe dermatological involvement, deaths, and neonatal infections which are clinical cases rarerly or never described before presented in Reunion Island (Pialoux *et al.*, 2006; Schufffenecker *et al.*, 2006). A fatal hemorrhagic phenomenon was reported in India, (Laine *et al.*, 2004; Afjal *et al.*, 2002).

Due to similarities in clinical presentation with dengue, limited awareness, and a lack of laboratory diagnostic capability, CHIKV is probably often under diagnosed or misdiagnosed as dengue (Sam and AbuBakar, 2006; Chastel 2005; Pialoux *et al.*, 2006).

1.2.6 Diagnosis of the disease

Definitive diagnosis can only be made in the laboratory. CHIKV however, should be suspected when epidemic disease occurs with the characteristic triad of fever, rash and rheumatic manifestations, (Lam and Chua, 2005).Virus isolation is readily accomplished by inoculation of mosquito cell culture, mammalian cell culture or suckling mice where the virus can produce cytopathic effects, (Cunningham *et al.*, 1975).

Antigen/Antibody tests for detection of the virus

Haemagglutination inhibition tests can be usedto check for samples which have the virus. Haemagglutination inhibition antibodies appear with the cessation of viremia. All patients will be positive by day 5 to 7 of illness, (Lam and Chua, 2005).

Alternatively, virus specific IgM antibodies are readily detected by capture-enzyme linked immunosorbent assay (capture-ELISA) in patients recovering from CHIKV infection and they persist in excess of 6 months. ELISA is a biochemical technique used mainly in immunology to detect the presence of an antibody or an antigen in a sample. It uses two antibodies. One antibody is specific to the antigen. The other reacts to antigen-antibody complexes, and is coupled to an enzyme. This second antibody, which accounts for "enzyme-linked" in the test's name, can also cause a chromogenic or fluorogenic substrate to produce a signal. Because the ELISA can be performed to evaluate either the presence of antigen or the presence of antibody in a sample, it is a useful tool both for determining serum antibody concentrations (such as with the Human Immunodeficiency Virus, HIV test or West Nile Virus) and also for detecting the presence of antigen. ELISA may be run in a qualitative or quantitative format. Qualitative results provide a simple positive or negative result for a sample. The cutoff between positive and negative is determined by the analyst and may be statistical. Two or three times the standard deviation is often used to distinguish positive and negative samples. In quantitative ELISA, the optical density or fluorescent units of the sample is interpolated into a standard curve, which is typically serial dilution of the target, (Hongbao Ma, 2006).

Molecular test

The virus can also be diagnosed molecularly by Polymerase Chain Reaction. The Polymerase Chain Reaction (PCR) technique, invented in 1985 by Mullis, makes possible the *in vitro* synthesis of millions of copies of a scarce sample of DNA. It is a molecular biology technique, for enzymatically replicating DNA without using a living organism such as *E.coli* or yeast. Like amplification using living organisms, the technique allows a small amount of the DNA molecule to be amplified exponentially. However, because it is an *in vitro* technique, it can be performed without restrictions on the form of DNA and it can be extensively modified to perform a wide array of genetic manipulations. A pair of oligonucleotides is used as primers for a series of synthetic reactions that are catalyzed by a DNA polymerase. These oligonucleotides typically have different sequences and are complementary to sequences that lie on opposite strands of template DNA and flank the segment of DNA to be amplified. Following exponential amplification, a segment of double stranded DNA whose termini are defined by the oligonucleotide primers is generated as the major product. (Smithsonian Institution Archives; Sambrook and Russel, 2001).

A single tube reaction of both cDNA synthesis and PCR can now be performed on RNA; this reduces the risk of contamination. Titan one tube RT-PCR system is designed for sensitive, quick and reproducible analysis of RNA with high fidelity.

The one step reaction system uses avian myeloblastosis virus reverse transcriptase for first strand synthesis at 42°C and expand high fidelity enzyme blend, which consist of Taq DNA Polymerase and a proof reading Polymerase for PCR (Mallet *et a*l., 1995).

1.2.7 Treatment

Treatment solely addresses the symptoms of the disease such as fever, pain etc. It is usually a combination of analgesic and/or antiinflammatory agents (Paganin *et al*., 2006). The prognosis is generally good, although some patients experience chronic arthritis. With no vaccine or antiviral available, prevention and control depends on surveillance, early identification of outbreaks, and vector control (Sam and AbuBakar, 2006).

1.3 Statement of the problem

CHIKV probably originated in Tropical Africa and subsequently was imported into southern Asia. In Africa, evidence that the virus circulates continually in sylvatic cycles has been documented for decades, (Powers *et al*., 2000). This means that the virus has ability to spread and assume a wide geographical distribution across the globe. The urban cycle of transmission reported in Asia, (Sam and AbuBakar, 2006) portrays a potential in increased chances of infection, thus dramatically increasing the magnitude of epidemics where environments favor mosquito breeding. CHIKV in 2006 caused one of the largest Chikungunya fever outbreaks reported in the last 40years. The magnitude of the epidemics surprised the populations, policy makers and public health specialists (Isabelle *et al*., 2006).

1.3.1 Justification of the study

CHIKV has in the recent past caused massive outbreaks and with increasing severity in presentation including neurological involvement, haemorrhage and causing death. Unfortunately, no clear understanding of what initiates an epidemic of this scope has been recorded. There is, therefore, need to carry out extensive field studies to monitor human illness, non-human seroprevalence rates, and mosquito infection rates would be ideal. However, as field investigations of this nature are unlikely to be supported, particularly during inter-epidemic periods. Molecular genetic approaches using existing viral isolates may provide some clues as to the origin of outbreak strains and help to more effectively assess viral movement and infection patterns.

The genetic make-up of the viruses involved in the Kenya and Comoros outbreaks would provide information to ascertain the relationships among them and with the closely related S27 African Prototype strain. The study would therefore, produce useful data for following the movement of particular strains and help in prediction of future outbreaks. Since recent CHIKV outbreaks have increased in magnitude compared to the previously documented ones, previously obtained CHIKV sequences may not be relied upon entirely for vaccine development. Obtaining complete genetic data of viruses involved in recent outbreaks would provide useful information for development of effective vaccine. By inferring the amnio acid sequences from the nucleotide sequences, polypeptide vaccines could be developed. Alternatively the genes responsible for virulence, either coding for destructive enzymes or toxins could be established. The complimentary RNA sequence could be used to produce antisense drugs which have been found useful against viral diseases because they are designed to enter the cell and eliminate viruses by preventing their replication. The drugs which act by blocking critical viral genetic sequences may be more potent than anti-virals such as protease inhibitors, which seek to inhibit a protein needed for replication (Warfield *et al.*, 2006). No commercial, routinely available prophylaxis exists for chikungunya.

1.4 Hypothesis

The hypothesis of the study is that the Chikungunya viruses involved in the Kenya and Comoros islands outbreaks are genetically conserved; suggesting virus spread could be from a single source.

1.5 Main goal

The main goal of the study is to compare the nucleotide sequences of the Chikungunya viruses from the Kenya and Comoros outbreaks with each other and with the already sequenced S27 strain.

1.5.1 Specific objectives

1. To sequence the entire Chikungunya virus genome of one of each virus involved in the Kenya and Comoros outbreaks.

2. To perform phylogenetic comparison among the isolates of CHIKV from Lamu, Mombasa, Comoros and with the S27 strain sequence in the Genbank.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Collection of samples

Serum samples were collected following an outbreak of Chikungunya fever among the people of Lamu Island, Mombasa (KPA) and Comoros islands and were kept at -80° C at Centers for Disease Control based at the Kenya Medical Research Institute in Nairobi.

There were two samples from Lamu coded as Lamu 33 and Lamu 75, four from KPA Mombasa coded as KPA 11, KPA 15, KPA 16 and KPA 23, and seven from Comoros coded as Com1, Com3, Com11, Com18, Com19, Com20 and Com 25. One representative sample from each of these outbreak areas was picked at random and used for sequencing. These included Lamu 33, KPA 15 and Com 25. The experiments were done in replicates of three for each of the isolates.

2.1.1 Detection of the virus

Diagnosis was done by Immunoglobulin-M (IgM) capture-ELISA (Hongbao Ma, 2006). The steps of the general, "indirect," ELISA for determining serum antibody concentrations for a battery of alpha viruses were: Goat anti-human IgM which was diluted 1:1000 with serum dilution buffer (PBS, pH 7.4, 0.01% Merthiolate, 0.1% Tween-20, 5% skim milk) and used to coat the whole plate, 75µl per well. The plate was incubated for 18 hours at 4^0C. This was then washed five times with 300µl per well volumes of wash buffer. To avoid unspecific bindings, the antibodies were blocked using skimmed milk 0.5% PBS-Tween. 30 minute incubation followed at room temperature before washing five times with 300µl per well volumes of wash buffer (PBS, pH 7.4, 0.01% Merthiolate, 0.1% Tween-20).

Test sera (positive and negative control) diluted 1:400 with the dilution buffer was added and incubated at 37^0C for 60 minutes. The plates were then washed 5 times with wash buffer. The positive and mock antigens were diluted with the dilution buffer, 1:40, and 50µl added to each well. The plate was then incubated at 4^0C for 18 hours. The plate was then washed 5 times with 300µl per well volumes of wash buffer. 50µl of anti-alpha virus conjugate, diluted 1:6000 with dilution buffer, were added per well and the plate incubated for 1 hour at 37^0C. The plate was then washed 10 times with 300µl per well of wash buffer. 75µl of ABTS substrate was added to each well. The plate was incubated again at room temperature for 15 minutes. The plate was then read using an ELISA readerat a wave length of 405nm.

In this study, serum samples that were CHIKV positive following diagnosis during the outbreaks were characterized.

2.2 Virus culture

Virus was cultured in Vero cells. (a cell line developed from African green monkey nephrocytes byYasamura and Kawakita in 1963). 200µl of supernatant of each of the isolate samples were inoculated into Vero cells. The aliquots made for seed virus were maintained at – 80^o C in a biosafety level 3+ laboratory at CDC.

2.2.1 Culture media for Vero cells

Medium used was the modified minimum essential medium from Sigma Aldrich company (Eagle, 1957). The cells were grown in growth medium (GM), a high nutrient medium, which constituted 500ml of minimum essential medium supplemented with 10% fetal bovine serum– 50ml, 2% L-glutamine- 10ml, 1% antibiotic/antimycotic- 5ml. A low nutrient medium was also prepared to maintain the cells without overgrowing. This maintenance medium constituted 2% fetal bovine serum- 10ml, 2% L-Glutamine- 10ml, 2% antibiotic/antimycotic- 10ml, all in minimum essential medium - 500ml.

2.2.2 Vero cells sub-culturing

A fully confluent T25 flask of Vero cells was obtained from Kenya Medical Research Institute. For aseptic technique, all procedures were undertaken in a biosafety cabinet. The T25 flask was passaged by first decanting the medium and washing twice to thrice with about 3ml of 1% sterile Phosphate-buffered Saline (PBS) pre-warmed to 37°C in a water bath.

The cells were detached from the substratum/monolayer by the addition of warm (37°C) trypsin (0.5ml), rocking the flask gently to ensure entire coverage of the cell monolayer by the enzyme and briefly incubating in a 5% CO_2 at 37°C. (Care was taken since over exposure to trypsin would kill the cells). The flask was tapped and shaken to aid detaching of cells. Detached cells were seen to be floating when viewed using the inverted light microscope. Once detached, 5ml of GM (pre-warmed to 37°C) were added immediately to stop the trypsin action. This was purged by pipetting up and down so as to break any clumps of clustered cells.

The cells were viewed under the microscope to ensure that the cells appeared singly. To split in a ratio of 1:4, GM was topped up to 20ml. Each of the 4 new sterile T25 flasks was labeled with the cell type, date, subsequent passage number and user name. 5ml of cell suspension was pipetted into each of the 4 flasks. The flasks were tightly closed and incubated at 37°C in a 5% CO_2 incubator. The cells were observed daily when at around 70% confluent, they were infected with the virus (serum). To have more of Vero cells, some flasks were allowed to grow to 100% confluence and passaged into more flasks.Confluent flasks that could not be split were maintained at halted growth by decanting GM and washing twice to thrice with 1% sterile PBS pre-warmed to 37°C and adding maintenance medium also pre-warmed to 37°C. These were kept in a 5% CO_2 incubator.

2.2.3 Inoculation of Vero cells with virus

Inoculation of Vero cells was done in T25 flasks that were about 70-80% confluent. The medium was first decanted and the cells washed with about 3ml of 1% sterile PBS pre-warmed to 37°C twice or thrice. 200µl of serum were added to each flask while 200µl of maintenance medium were added to one control. The flasks were gently rocked to spread the inoculum over the cell-sheet and incubated in a 5% CO_2 incubator for 1 hour while rocking at 15 minutes intervals to allow attachment of virus. 5ml of maintenance medium were added and the flasks kept in the same incubator. The cells were observed daily for cytopathogenic effect (C.P.E.). After two to three days, Vero cells inoculated with serum rounded and some died off, thus, lifting from the monolayer. The control cell-sheet remained intact showing that that cell rounding was real virus effect which was actually C.P.E.

After day three or four, the C.P.E. was evident in about 90% of the monolayer and the cells were frozen in a -80°C freezer. For the study, the virus titer was increased by re-infecting the other cells using T75 flasks. The flasks were freeze thawed to facilitate cell lysis and virus release, followed by centrifugation at 3000 rpm for 10 minutes to remove cellular debris. The virus-containing supernatant was aliquoted into cryovials and stored at -80 ^{o}C.

2.2.4 Dilution of virus titer for re-inoculation

High concentrations of the virus isolate could easily kill the Vero cells during the 1 hour incubation in the infection process. This would render virus isolation in the flask unsuccessful since viruses can only grow in living cells. To avoid this, the isolates were diluted in a way that they would not immediately kill cells neither would they be too dilute to be effective in infection. To infect the T75 flasks, isolates were thawed on ice. Six culture flasks (T75) that were 80% confluent were washed twice with 5ml each of 1% sterile PBS equilibrated at 37^{o}C in a water bath. PBS was finally aspirated. 4ml of maintenance medium were put in a 15ml tube and a 1.5ml vial of virus added to it. These were mixed by pipetting up. Afterwards five flasks were inoculated with 1.5ml.of the diluted virus (isolate). A sixth flask was inoculated with 1.5ml.maintenance medium as a control. The inoculum and the cells were incubated for 1 hour at 37^{o}C in a CO_2 incubator, while rocked after every 15 minutes. After incubation, 20 ml of maintenance medium were added to the flasks. The cultures were observed daily using an inverted light microscope for signs of cytopathic effects (CPE).

CPE included cell rounding and disruption of the monolayer (lytic CPE). When CPE involving 75-100% of monolayer was noticed, the flasks were frozen at -80^{o}C to preserve the virus and also for complete lysis of the Vero cells. The flasks were thawed and the contents transferred into a 50 ml falcon tube (centrifuge tube). Centrifugation was performed at 1,500g for 10 min and the supernatant decanted into clean 50ml. falcon tube. The pellet was discarded. Supernatants were aliquoted for virus seed.

2.3 Isolation of RNA

Infected Vero cells were used for RNA extraction by a modified isolation method of acid guanidinium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi, 1987; Sambrook *et al*., 1989). QIAamp Viral RNA mini kit from Applied Biosystems which combines both the selective binding properties of a silica-gel based membrane with the high-speed microspin or vacuum technology was used following manufacturers instructions **(**QIAgen Inc, Valencia, C.A).

An aliquot of supernatant from infected Vero cells was equilibrated to room temperature (15-25^{o}C). 140µl of it was mixed with lysis buffer AVL (560µl) containing carrier RNA in 1.5 ml micro-centrifuge tube. To allow for complete lysis of the viral particals, a 10 minute incubation period ensued at room temperature (15-25^{o}C). The 1.5 ml micro-centrifuge tubes were then briefly centrifuged.

To precipitate RNA, absolute ethanol (560 µl) was added to the samples and mixed by pulse-vortexing for 15 seconds. Solution (630 µl) from above step was carefully applied to the QIAamp spin column (in a 2 ml collection tube) without wetting the rim. This was centrifuged at 8000 r.p.m. for 1 min. This bound the RNA to the silica-gel based membrane. The QIAamp spin column was placed into a clean 2 ml collection tube, and the tube containing the flow-through discarded.

In order to obtain pure RNA, it was necessary to wash away other unbound cell components such as proteins. 500ul of buffer AWI were added to the QIAamp spin column. This was centrifuged at 8000 r.p.m. for 1 min. The QIAamp spin column was placed in a clean 2 ml collection tube, and the tube containing the filtrate discarded.

500 µl of buffer AW2 were added and centrifuged at full speed (14,000 r.p.m.) for 3 min.

The flow-through was discarded and the QIAamp spin column was centrifuged for an additional minute at full speed.

The QIAamp spin column was placed in a clean 1.5 micro-centrifuge tube, the collection tube containing the filtrate discarded. 60µl of buffer AVE equilibrated to room temperature added. The cap was closed, incubated at room temperature for 1 min. This was centrifuged at 8000 r.p.m. for 1 min. The eluent contained the RNA.

2.4 Designing of primers

RT-PCR primers were designed using previously published partial sequences for CHIKV Ross strain (Logue and Atkins, 2002) S27 African prototype strain (Khan *et al.*, 2002). Internal primers were designed from the nucleotide sequences of CHIKV analyzed in the study.

Primer sequence was verified using Primer 3 software in the internet, (Rozen and Skaletsky, 2000). The software allows for entry of a sequence by pasting source sequence in 5′->3′ direction. Values for desired primer size, melting temperature (primer Tm) and guanine and cytosine (GC) nucleotides concentration in percentage are entered, each with a minimum, maximum and optimum value. Primers with 55-60% GC content have been found to be optimal. The primer3 software selects out of the pasted sequence the region giving the optimal size (20-24 nucleotides) with the optimal GC%.

Primer sets were designed in such a way that they would yield overlapping sequences so as to ensure that every nucleotide in sequence was captured. The overlap was about 200bp.

*Nucleotide sequences of the primers and their utility are shown in **Table 1**.

Other ptimers designed and used as the work was under-way are shown in appendix 4.

Table 1: Oligonucleotide primers used to amplify CHIKV genome in the Study were derived from S27 and Ross strains.

Primer	Sequence	Size	Gene
Chik921 R	ACGAAACCACTGTATCACAGCGGCATGTGAAG	32	nsP1
Chik413F	CTCTGGAAAGATCGGGGACTTACAAGCAG	29	nsP1
Chik726F	CTGACGGAAGGTAGACGAGGCAAATTGT	28	nsP1
Chik1521F	CCATACAGCGGGGACGCCCAAGAAGCCCGG	30	nsP2
Chik1550R	CCGGGCTTCTTGGGCGTCCCCGCTGTATGG	29	nsP1
Chik1387F	CCCAGTCAATTCAGAAGGTTCAGGCCGAGTT	30	nsP1
Chik2283F	ACCAGGCAAGACCTGGTGACTAGCGGAAAG	30	nsP2
Chik2172R	ATTCGTGGTAGGGCGGATTAGTCAAGTCGCCCACCAG	36	nsP2
Chik3008F	CTTCAAAGCAACTATTAAGGAGTGGGA	27	nsP2
Chik3034R	TCCCACTCCTTAATAGTTGCTTTGAAG	26	nsP2
Chik2902F	CAGAGCACGTCAACGTACTCCTAACGCGTACGGAA	35	nsP2
Chik3990F	GACAATGGCAGAAGGAATTTCACAACTCAT	30	nsP3
Chik3840R	CACCCGGTTTGAGCAGTCTCAATGAGTC	28	nsP2
Chik4927R	ATGCTTGTGACATGGTTCATGCGAAGTCGGGTAACGCGTTCTGG	44	nsP3
Chik4705F	CAGTGGATATGGCAGAGATATATACTATGTG	31	nsP3
Chik5747F	GGAGGAAGTCCACGAGGAGAAGTG	24	nsP4
Chik5579R	TAGGAAGTCTCCGAAAGTTAGTAGCTCAG	29	nsP3
Chik6320F	GGACTCAGCAGTATTCAACGTGGAGTG	27	nsP4
Chik6346R	CACTCCACGTTGAATACTGCTGAGTCC	27	nsP4
Chik7308F	CGAGCGCTGGCTGACGAAGGGTCAGATGGCAACGA	35	C
Chik7267R	TTGCCCAGTTTAAATAGCCTTTTTAGCG	28	nsP4
Chik8128R	GTACTGCTCCGTGGTGCCAGTTGTAGTACCCCTCCGG	37	nsP4
Chik7902F	CGAAGGTAAGGTAACAGGTTACGCGTGCT	29	E3
Chik8991F	GCAGAGCAACGCCGCAACTGCCGAGGAGATAGAG	34	E2
Chik8753R	GCGTCTGCTGGTATGTGATTGTCCATGTAAC	31	E3
Chik9758F	GCCTAATATGCTGCATCAGAACAGCTAAAGCGGCCA	36	E2
Chik9571F	CACCCGCATGAGATAATCTTGTACTAT	27	6K
Chik9597R	ATAGTACAAGATTATCTCATGC GGGTG	27	E2
Chik10497R	GTTGTCRAAAGGTGTCCAGGCTGA	24	E1
Chik10272F	CACGCAATTGAGCGAAGCACATGTGGAG	28	E1
Chik261F	TGGATATTGGTAGTGCGCCAGCAAGGAGGATGATGTCGGACAG	43	nsP1
Chik280R	GGCGCACTACCAATATCCA	19	nsP1
Chik7755F	GAATCGGAAGAATAAGAAGCAAAAGCAAAAGCAGCA	36	C
Chik7787R	TGCTTTTGCTTTTGCTTCTTATTCTTCCGATTCTTG	36	nsP4
Chik10965R	ATGCACCGCACACTTGCCTTTCTTGCT	27	E1
Chik520R	CGCGACGTCTGCTCTCTG	18	nsP1
Chik11278F	TTCAGCAGGCACTAACTTGAC	21	3'End
Chik11254F	CTAATCGTGGTGCTATG	17	3'End
Chik11701R	TGAGTTCGGCTGCTTTTAGG	20	E1
Chik11747R	CGGAGAATCGTGGAAGAGT	19	E1

KEY

F- Forward sequence
R- Reverse sequence
nsP-non-structural protein gene
C- Capsid
E- Envelope protein gene

2.5 Polymerase Chain Reaction (PCR)

The one step reverse-transcription PCR (rt-PCR) was optimized using ABI QIAamp® kit for the study (QIAgen Inc., Valencia, CA).

Forward and reverse primer sets (Table 1) were used to amplify portions of the genome to produce amplicons ranging from about 400bp to a maximum of 1500 bp. The 1500 bp limit was determined by the *Taq* polymerase used in the experiments which was unable to amplify larger amplicons. A forward primer was coupled to any reverse primer at random to make a set as long as it produced an amplicon of not more than 1500 bp. The size of the amplicon expected was the difference of the forward and reverse primer. Each set of primers was used to amplify RNA extracts from 1st passage Vero cell supernatant of each of the three isolates i.e. Lamu, KPA and Comoros.

RT-PCR reaction mix

Into each of 200 ul MicroAmp autoclaved reaction tubes the following reagents were added;5X buffer,10µl, 3.5 µl of MgCl (25mM), 4.0 µl of dNTP's (10mM), 2.5 µl of DTT (5mM), Forward primer(50pmol), 0.15µl, and a same amount of equally concentrated reverse primer. To protect the RNA template from nucleases, 0.5µl of Rnase inhibitor (20U/µl) were added and for reverse transcription and ultimate amplification of the nucleic acid, 0.3µl of Multiscribe Reverse transcriptase (50U/µl) and Amplitaq Gold DNA polymerase (5U/µl), 0.5 µl were added respectively.

Finally 10.0 µl of template RNA were added and the volume topped to a final volume of 50 µl using RNase free water.

Thermocycling conditions

The reaction mixtures were placed in the ABI 9700 thermo cycler and reaction volume set to 50 µl. The thermocyclic process entailed two pre-PCR steps; reverse transcription hold at 45^{o}C for 45 minutes after which was a rapid thermal ramp to 94^{o}C. Amplitaq Gold DNA polymerase activation occurred at this temperature for three minutes.

The following parameters were repeated for 35 cycles: 94°C for 30 seconds (Denaturation of DNA), rapid thermal ramp to 48^{o}C, and 48^{o}C for 1 minute for the annealing of primers to template cDNA, rapid thermal ramp to 72^{o}C, which was the optimal temperature for elongation process for 1 minute.

The final elongation stage i.e. the 35th cycle, was at 72^{o}C for 7 minutes and was a non cyclic stage. Finally there was a rapid thermal ramp to 4^{o}C and held until ready to purify (non cyclic stage).

PCR products were loaded on a gel for an electrophoretic run to determine the specificity of amplification that is by size of the amplicon observed.

2.5.1 Gel electrophoresis

Electrophoretic analysis of the PCR products was performed using agarose gels. The gel solution was prepared by mixing 100 ml of 1x TAE buffer (contains; 0.04m Tris Base, 0.04M acetate and 0.001M EDTA) with 1.5g of electrophoresis grade agarose. These were boiled by heating until the solution was clear (no particles observed). After cooling the agarose gel to about 55^0 C, 10 µl Ethidium Bromide, which complexes with DNA and allows its visualization, was added according to principle of method (Severini and Morgan, 1991). This was gently swirled avoiding formation of air bubbles. The gel was poured (approximately ¼ to ½ inch. in thickness) and a proper comb for particular gel inserted. The gel was allowed to cool until it set (turned whitish and opaque when ready). The comb was removed and the gel placed in the gel rig with the wells closest to the cathode (black) end. The gel was covered with 1X TAE running buffer in an electrophoretic tank. Using a clean sample-loading tray, 2 µl of 6X Loading dye was added into loading tray wells. Keeping samples on ice, 10 µl of PCR product was mixed together with bromophenol blue (loading dye) in the loading tray. 123 b.p. DNA ladder was also prepared in a separate well in the loading tray by adding 2 µl of 123 b.p. DNA ladder (1 ug/ul), 2 µl of loading dye and 8 µl of distilled water and mixing them.

Samples were loaded; placing 10 µl of dye mixed sample in each well with 123 bp ladder at the extreme left hand side after the last sample was loaded.

The electrophoretic tank cover was placed on the gel rig and the samples ran towards the anode (red) end at 108 volts for 1 hour. The run was stopped after the dye front had migrated about two thirds or three quarters of the gel.

The separated bands were visualized under ultra violet light source at a wavelength of 312 nm. Using a hand held 0.7 X electrophoresis hood with a Wratten 22A filter and a lens of *f* = 103 nm, photographs of gel were taken using Polaroid 667 (ASA 3000) film.

2.5.2 Fragment recovery

Fragments with the expected base pairs, according to the primer set used, were excised and purified using QIAgen Min-elute gel extraction kit (QIAgen Inc., Valencia, CA).

DNA fragments were excised from the agarose gel with a clean, scalpel. Removal of extra agarose from the DNA fragments minimized the size of gel slices. The gel slice was weighed in colorless tube and 3 volumes of buffer QG added to 1 volume of gel (100 mg ~ 100 µl). For example, 300µl of buffer QG were added to 100mg of gel slice. The gel slice in buffer QG was incubated at 50^0 C for 10 min (or until the gel is completely dissolved). To help dissolution of the gel, tubes were mixed by vortexing every 2-3 min during the incubation. After the gel slice had dissolved completely and with yellow color confirmation, 1 gel volume of isopropanol was added to the sample and mixed by inversion of the tube several times. To bind DNA, the samples were applied to the MinElute column with a 2 ml collection tube attached, and centrifuged for 1 minute at 13,000 rpm. The flow-through was discarded and the MinElute column placed in a clean collection tube. 500 µl of Buffer QG were added to the spin column and centrifuged for 1 minute at 13,000 rpm. The flow-through was discarded and the MinElute column placed in a clean collection tube. To wash, 750µl of Buffer PE were added to the MinElute column. The column was let to stand for 5 minutes and then centrifuged for 1 minute at 13,000 rpm.

The flow-through was discarded, the MinElute column put in a clean collection tube and centrifuged for an additional 1 min at 10,000 x *g* (~13,000 rpm). The MinElute column was placed into a clean RNase free 1.5 ml microcentrifuge tube. To elute cDNA, 10 µl of Buffer EB (10mM Tris·Cl, pH 8.5) were added to the center of the membrane, letting the column to stand for 1 min, and then centrifuged for 1 minute at 13,000 rpm.

2.6 Genome Sequencing

Genomic sequencing was done using the dideoxy-synthesis termination method (Sanger *et al.*, 1977).

The method makes use of the 2′,3′-dideoxy and arabinonucleoside analogs of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. Because the ddT contains no 3′-hydroxyl group, the chain cannot be extended further, so that termination occurs specifically at position where dT should be incorporated. If a primer and template are incubated with DNA polymerase in the presence of a mixture of ddTTPs and dTTPs, as well as the other three deoxyribonucleoside triphosphates, a mixture of fragments all having the same 5′ and with ddT residues at the 3′ ends is obtained. When this mixture is fractionated by electrophoresis on denaturing acrylamide gels the pattern of bands shows the distribution of dTs in the newly synthesized DNA. By using analogous terminators for the other nucleotides in separate incubators and running the samples in parallel on the gel, a pattern of bands is obtained from which the sequence can be read off, (Atkinson *et al.*, 1969; Sanger *et al.*, 1977).

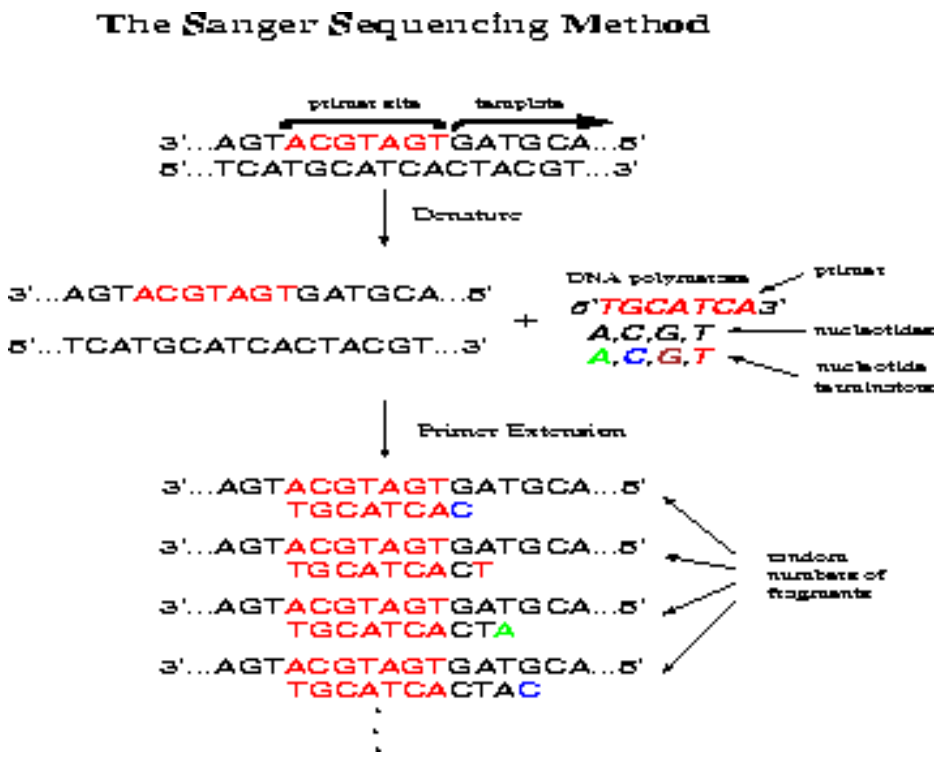

Fig.1. Schematic representation of the DNA sequencing reaction according to Sanger.

A modified method which incorporates all dideoxynucleotdes in one reaction but labeling them with four different dyes (Big Dye Terminator), was used in the experiments. In the ready Reaction format, the dye terminators, deoxynucleoside triphosphates, AmpliTaq DNA Polymerase, magnesium chloride, and buffer were premixed into a single tube of Ready reaction Mix. These reagents were suitable for performing fluorescence-based cycle sequencing reactions on single-stranded DNA templates, on PCR fragments, and on large templates, (ABI Prism 310 Genetic Analyzer, Perkin-Elmer/Applied Biosystems).

The complete sequences of Comoros 25, Lamu 33 and KPA 15 were determined bi-directionally from cDNA.

The reagents were added as follows; Big-dye Terminator ready reaction mix, 4.0 µl, 5x Sequencing Buffer, 2.0 µl, Template dsDNA, 2.0µl, 1.5µl of forward primer (50 pmol) and the reaction brought to a final volume of 20 µl using deionised water. A similar mix was prepared using the reverse primer. The tubes were sealed, mixed well and spun briefly.

The samples were placed in a thermocycler and reaction volume set to 20 µl and subjected to 25 cycles of amplification as follows; rapid thermal ramp to 96^0C, 96^0C for 10 seconds (Denaturation of double strands). Rapid thermal ramp to 50^0C and holding for 5 seconds where primers to anneal. Rapid thermal ramp to 60^0C allowing for elongation/termination synthesis for 4 minutes. Finally a rapid thermal ramp to 4^0C and held until ready to purify (non cyclic).

2.6.1 Ethanol/ Sodium Acetate Precipitation of sequenced DNA fragments

This was to purify sequencing reaction products by removing unincorporated dye terminators which would otherwise obscure data in the early part of sequence and interfere with base calling by the 3100 genetic analyzer (ABI prism 3100 Genetic Analyzer-User Bulletin).

50µl of cold absolute ethanol were mixed with 2µl of 3.0 M sodium acetate, pH 4.6 in 1.5 ml eppendorf tubes and 20µl of sequencing product added and mixed.

The mixture was incubated in the dark for 15 minutes at room temperature after which was centrifuged at 14000 rpm for 30 minutes at 4^o C. The supernatant was pipetted out and the pellet washed with 500µl of 70% ethanol.

Centrifugation followed at 14000 rpm for 20 minutes at 4^{o} C and the pellet was air-dried in a laminar flow after which it was re-suspended in 10µl of Hi-Di formamide and loaded in a 96 well plate. For the genetic analyzer to read the sequence, the double strands of the cDNA had to be unwound. Denaturation of the re-suspended pellet was done by heating at 95^{o} C for 2 minutes. The 96 well- plate was then loaded into the ABI 3100 genetic analyzer where electrophoresis and basecalling was done.

2.6.2 Electrophoretic separation of cycle sequencing fragments

Electrophoretic analysis gave a distribution of the dyes each showing the sequence of the particular nucleotides in respect to the label. cDNA (from rt-PCR) was analyzed following an automated ABI prism® Big Dye Terminator™ v3.1 Cycle Sequencing kit for Capillary Sequencers ABI 3100 Genetic Analyzer (ABI Prism 310 Genetic Analyzer, Perkin-Elmer/Applied Biosystems).

Genome sequencing resulted in many fragments of different lengths depending on the point of termination of synthesis of a new strand. The position of termination thus shows the ddNTP that was incorporated, due to the characteristic fluorescing of dye label, which is used to infer the complemntary nucleotide and as such the nucleotide sequence is established e.g. if a ddT is incorporated at position 314 of the genome, it means that the template has an adenosine nucleotide at that position. The fragments were separated electrophoretically by the automated ABI 3100 genetic analyzer. This used a 50cm. capillary system where fragments were separated according to their sizes and were detected by an internal charge-coupled device camera (CCD).

A CCD is an electrical device that is used to create images of objects, store information (analogous to the way a computer stores information), or transfer electrical charge (as part of larger device). It receives as input light from an object or an electrical charge. The CCD takes this optical or electronic input and converts it into an electronic signal - the output. The electronic signal is then processed by some other equipment and/or software to either produce an image or to give the user valuable information (Courtney Peterson, 2001). Since each ddNTP was labeled with a specific dye, the CCD camera detected each dye and gave an output of it to analyzing software which processed the information giving the template nucleotide sequence and storing that data automatically.

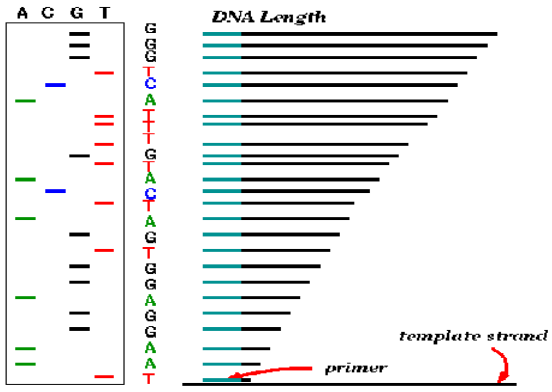

Fig.2. An illustration of electrophoresis of the cycle sequencing products. Gel electrophoresis spells out the sequence (Source: www-math.mit.edu/~lippert/18.417/).

The data was then retrieved and edited with the help of the International Union of Pure and Applied Chemistry diagram which guided in areas where the machine was unable to determine the nucleotide. By viewing an already generated electropherogram, the highest peak between the two nucleotides qualified for the nucleotide in sequence.

2.7 Data Analysis

The CHIKV sequences were analyzed with DNASTAR computer program, (DNASTAR Inc., MadisonWI). DNASTAR provides tools for sequence assembly, editing, alignment and analysis. It consists of Lasergene, comprehensive software for DNA and protein sequence analysis, contig assembly and sequence project management. DNASTAR Lasergene software consists of an integrated suite of seven modules. The modules of Lasergene are: SeqBuilder - visualization and sequence editing Video, SeqMan Pro - sequence assembly and single nucleotide polymorphism discovery, MegAlign - sequence alignment, PrimerSelect - oligo primer design, Protean - protein structure analysis & prediction, GeneQuest - gene finding, EditSeq - utility for importing unusual file types.

The Data Manager enables data integration between the Lasergene modules so that edits, additions and deletions made to a sequence in one module will synchronize and automatically update when opened in most other modules. CHIKV analysis employed use of SeqMan Pro, MegAlign and EditSeq.

SeqMan Pro enabled the assembly of sequence data from Sanger sequencing. A few fragments to genome assemblies can be done at the push of a button. SeqMan Pro provides two different assembly methods; the Classic Assembler, and the Pro Assembler for assemblies of 100 sequences or more. SeqMan Pro removes unreliable data, including poor quality ends, sub-minimal length reads, and vector and contaminating host sequences in a single pass, then assembles the trimmed data and calls the consensus.

SeqMan Pro uses DNASTAR's unique trace quality evaluation method to call the most accurate consensus sequence possible. This method reduces the depth of coverage needed for accurate sequence determination, saving time. SeqMan Pro also uses dual-end sequence data when available to group and order contigs, allowing you to fill gaps quickly with minimal additional sequencing efforts.

To evaluate the assembly, a single alignment window allowed viewing key aspects including; two consensus sequences for the same contig, called by different methods, multiple sequences and their underlying trace data, candidate SNPs and sequence and consensus conflicts.

EditSeq interfaces with other programs including GenBank, FASTA, MacVector, GCG®, Text, ABI®, and word processing. Dynamic links between sequence and annotations exhibited automatic updating of feature coordinates with sequence editing and inclusion of features with copying/pasting of sequences. Available functions included reverse complement, invert, translate, back-translate, and open reading frame identification, (DNASTAR Inc., MadisonWI).

Pure sequences were analyzed using Basic Local Alignment Sequence Tool (BLAST) to confirm their identity (Altschul *et al.* 1997). Once contigs had been obtained from the SeqMan Pro, next was to simply BLAST* them against National Center for Biotechnology Information (NCBI's) data to gather information already known about related sequences. The NCBI BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

MegAlign offered choice of four pairwise and four multiple sequence alignment methods for aligning nucleic acid or polypeptide sequences. Own sequences or public data directly from NCBI* was entered. To find more related sequences for alignment, a BLAST query was ran, and then the sequences wanted from the list of matches dropped in. Views of alignments were easily customized to highlight the similarities or differences of the sequences. Differences in chemical, structural or functional characteristics between sequences were also displayed as well as groupings or consensus. MegAlign also enabled construction of phylogenetic trees, generation of detailed numerical reports or export data of sequence comparisons, (DNASTAR Inc., MadisonWI).

CHAPTER 3: RESULTS

3.1 RT-PCR

The RT- PCR yielded cDNA of sizes ranging from 615 to 1353 base pairs as shown in figure 3 and 4. The samples include Comoros 25 (Figure 3) and KPA 15 (Figure 4). Eight primer sets were used which were: Set1: 1031F-2049R (1018 bp); Set2: 3161F-4870R (1709 bp); Set3: 5712F-7045R (1333 bp); Set 4: 7308F-8128R (820 bp); Set5: 7902F-8753R (851 bp); Set 6: 8646F-9647R (1001 bp); Set7: 9199F-10357 (1158 bp); Set 8: 10272F-11535R (1263 bp).

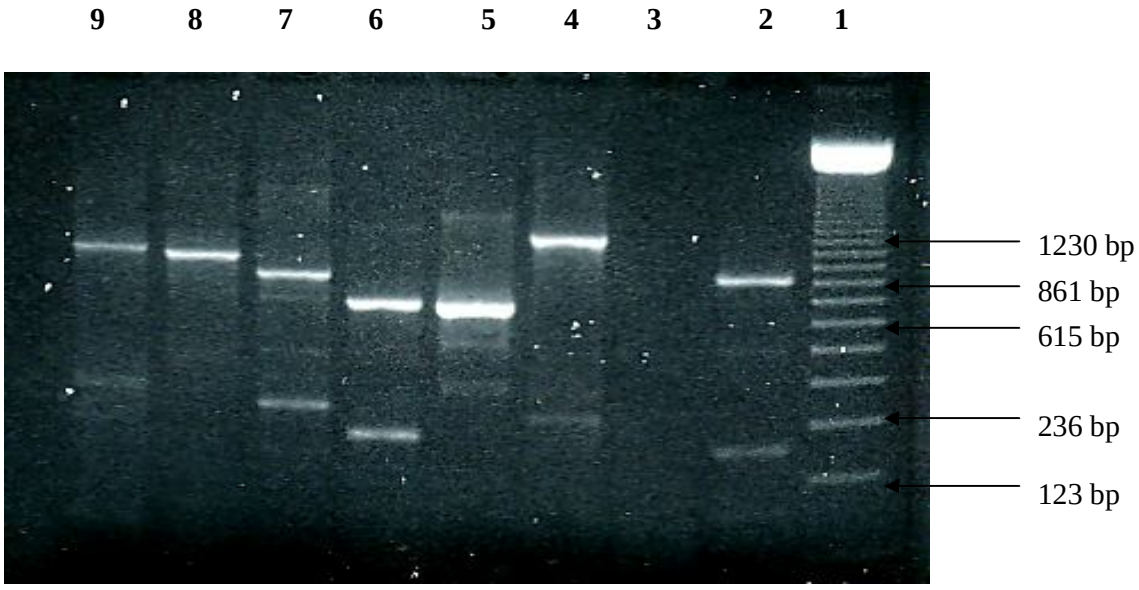

Fig.3. A photograph of a 1.5% agarose gel, containing PCR products of Comoros 25. In lane 1 was loaded 123 bp DNA ladder (2µl); lane 2 was comoros 25 using primer set 1(1018 bp); lane 3, using set 2(no band) ; lane 4, set 3 (1333 bp); lane 5, primer set 4 (820 bp); lane 6, primer set 5 (851 bp); lane 7, set 6 (1001 bp); lane 8, set 7 (1158 bp) and lane 9 was primer set 8 (1263 bp).

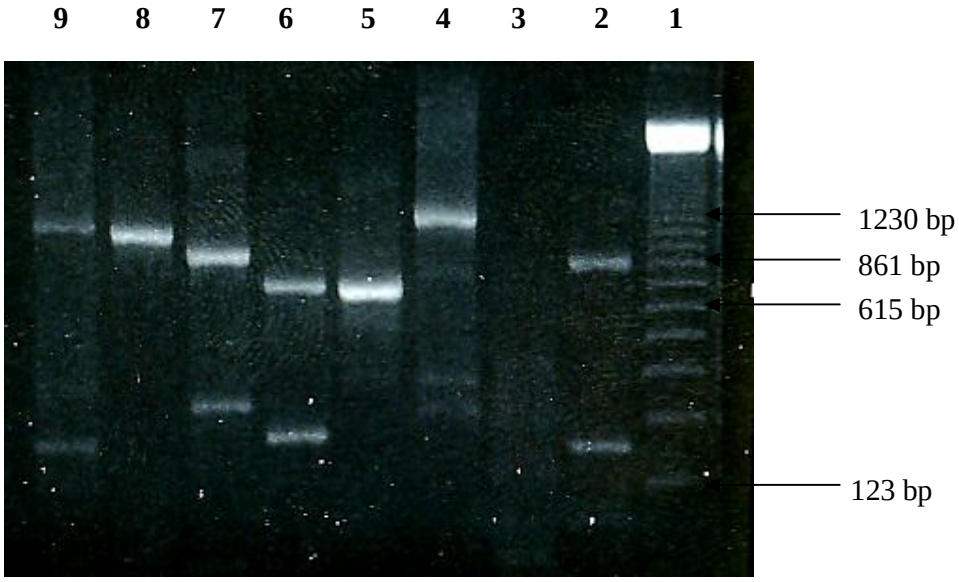

Fig.4. A photograph of a 1.5% agarose gel containing RT-PCR products of KPA 15. In lane 1 was 123 bp DNA ladder; lane 2 was KPA 15 using primer set 1(1018 bp); lane 3 using set 2 (no band); lane 4, using set 3 (1333 bp); lane 5 using primer set 4(820 bp); lane 6 using set 5(851 bp); lane 7 using set 6 (1001 bp); lane 8 using set 7 (1158 bp); lane 9 using set 8 (1263 bp).

The primer set 2; (3161F-4870R) which was expected to give an amplicon of 1709 base pairs did not yield any amplification products in both Comoros and KPA (Fig. 3 & 4).

3.1.1 Preparative RT-PCR electrophoresis.

The separation of the RT-PCR fragments to be used for sequencing reactions was achieved by loading 40µl of the RT- PCR products and in the gel wells. The wells for preparative RT-PCR electrophoresis were made using larger (midi) combs to allow accommodation of larger volumes of RT-PCR poducts. These gels, however, did not give a clear cut DNA ladder. Thus, there was need for such gels as in figures 3 and 4 which allowed determination of bands with the correct size. cDNA bands with the expected number of basepairs were excised for purification. The representative gel for the preparative RT-PCR electrophoresis is shown in figure 5.

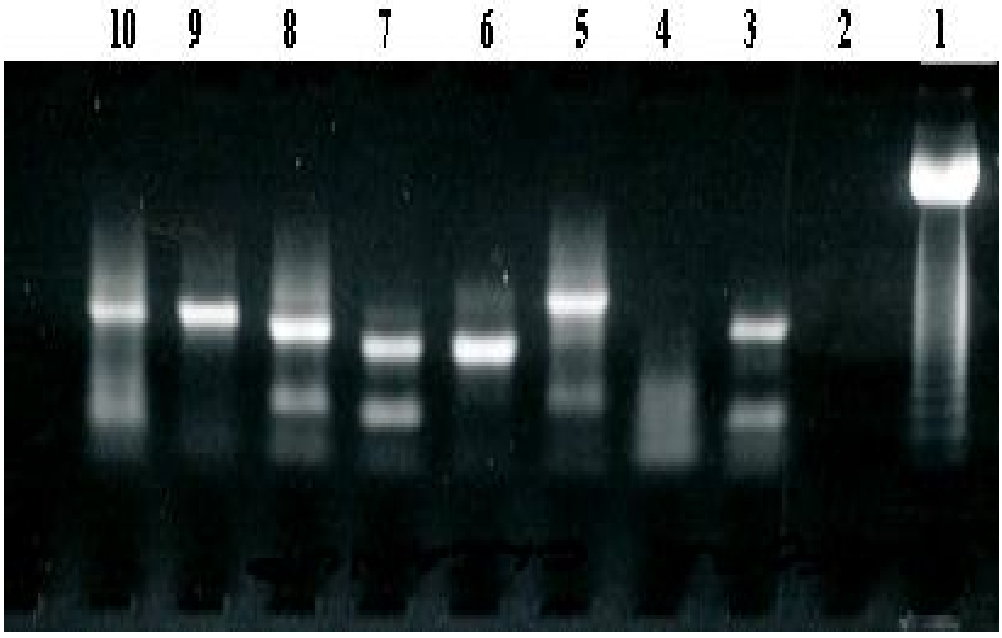

Fig.5. An agar gel electrophoresis photograph showing cDNA fragments of expected sizes. In lane 1 was 123 bp ladder; lane 2, no band; lane3, 1016 bp fragment; lane 4, 706 bp; lane 5, 1711 bp; lane 6, 1084 bp; lane 7, 1389 bp; lane 8, 1208 bp; lane 9, 1502 bp; lane 10, 1429 bp.

3.2.0 Analysis of CHIKV genome

The analysis done was computational.

3.2.1 Analysis non-structural gene 1 (nsp1) gene

Table 2: Summary of nsp1

Strain	Nucleotide translated (No.)				Mol % of Nucleotides				G+C%	A+T%
	A	G	T	C	A	G	T	C		
S27	465	440	308	386	29.08	27.52	19.26	24.14	51.66	48.34
Com25	461	442	304	392	28.83	27.64	19.01	24.52	52.16	47.84
KPA15	461	442	304	392	28.83	27.64	19.01	24.52	52.16	47.84
Lamu33	461	442	304	392	28.83	27.64	19.01	24.52	52.16	47.84

The nsp1 gene was 1605 base pairs long. Compared to S27 strain, which coded for a 59.76 kDa protein, the Comoro, KPA and Lamu strain nsp1 genes were inferred to translate to a 59.81 kDa protein. The latter 3 had an isoelectric point of 7.647 which differed from S27 with 6.870 and their charge at pH 7.0 was 3.301 against -0.695 of S27 strain. There was no significant difference in the G+C and A+T contents (Table 2).

3.2.2 Analysis of nsp2 gene

Table 3: Summary of nsp2 gene.

Strain	Nucleotide translated (No.)				Mol % of nucleotides				G+C%	A+T%	Isoelectric pt.	Charge at pH 7.0
	A	G	T	C	A	G	T	C				
S27	654	532	446	492	30.79	25.05	21.00	23.16	48.21	51.79	9.004	25.171
Com25	650	534	444	496	30.60	25.14	20.90	23.35	48.49	51.51	9.033	25.034
KPA15	650	534	443	497	30.60	25.14	20.86	23.40	48.54	51.46	9.033	25.034
Lamu33	650	534	444	496	30.60	25.14	20.90	23.35	48.49	51.51	9.033	25.034

The nsp2 gene was 2394 base pairs long. Nsp2 gene coded for a protein of 79.8 kDa in the S27 strain while Comoros 25, KPA 15 and Lamu 33 strains had a 79.9 kDa protein each encoded in this gene. All had 708 amino acids with a total number of 2124 translated bases each. The G+C and A+T contents had insignificant variations, (Table 3) In the nsp2 gene, Comoros 25 and Lamu33 were similar based on the above attributes. Closer to the two was KPA 15 which differed with two bases i.e. T and C but agreed with the two in isoelectric point and charge. A little distant from the three was S27 though exhibited a very close relationship.

3.2.3 Analysis of nsp3 gene.

Table 4: Summary of nsp3 gene.

Strain	Nucleotide translated (No.)				Mol % of nucleotides				G+C%	A+T%	Isoelectric pt.	Charge at pH 7.0
	A	G	T	C	A	G	T	C				
S27	437	418	303	408	27.91	26.69	19.35	26.05	52.75	47.25	4.702	-25.767
Com25	439	418	306	403	28.03	26.69	19.54	25.73	52.43	47.57	4.678	-25.597
KPA15	435	409	305	399	28.10	26.42	25.78	25.78	52.20	47.80	4.642	-28.597
Lamu33	435	410	303	400	28.10	26.49	19.57	25.84	52.33	47.67	4.642	-28.597

The nsp3 gene had 1590 base pairs in length. The molecular weights of the proteins inferred from the nucleotide sequences of this gene were 57.1 kDa, in S27 and Comoros 25 strains while KPA15 and Lamu 33 had a 56.3 kDa protein each. In this gene, there seemed to be two groups, S27 and Comoros having a closer relationship while KPA and Lamu were like each other but slightly different from the other two. There were no exact matches in this gene and every strain had its own composition although, there existed closeness among the four virus strains.

3.2.4 Analysis of Capsid gene.

Table 5: Summary of the capsid gene.

Strain	Nucleotide translated (No.)				Mol % of nucleotides				G+C %	A+T %	Isoelectric pt.	Charge at pH 7.0
	A	G	T	C	A	G	T	C				
S27	139	136	84	112	29.51	28.87	17.83	23.78	52.65	47.35	8.146	2.795
Com25	137	137	83	114	29.09	29.09	17.62	24.20	53.29	46.71	8.146	2.795
KPA15	137	137	83	114	29.09	29.09	17.62	24.20	53.29	46.71	8.146	2.795
Lamu33	137	137	84	113	29.09	29.09	17.83	23.99	53.08	46.92	8.146	2.795

The capsid gene was 782 base pairs long spanning along different portions of the whole genome in each strain. All strains the capsid had a protein of the same molecular weight of 17.1 kDa. The isoelectric point, charge at pH 7.0, C+G and A+T content also were very similar especially among the Comoros25, KPA15 and Lamu 33, (Table 5).

For the A and G content, all the three strains were similar. Comoros25 and KPA15 were exact matches in this gene. Lamu differed slightly from them by having an extra T and one less C. All of the strains including S27 had the same isoelectric point and charge at pH 7.0 besides the earlier noted same weight. This is one gene that revealed a very close relationship among the 4 strains. As would be expected, there was no variability in the amino acid sequences.

3.3 Analysis of structural genes- Envelope (E) glycoprotein genes.

3.3.1 Analysis of E3 gene.

Table 6: Summary of E3 gene.

Strain	Nucleotide translated (No.)				Mol % of nucleotides				G+C %	A+T %	Isoelectric pt.	Charge at pH 7.0
	A	G	T	C	A	G	T	C				
S27	43	37	35	35	24.71	21.26	20.11	33.91	55.17	44.83	6.894	-0.099
Com25	41	38	33	62	23.56	21.84	18.97	35.63	57.47	42.53	6.894	-0.099
KPA15	41	38	33	62	23.56	21.84	18.97	35.63	57.47	42.53	6.894	-0.099
Lamu33	41	38	33	62	23.56	21.84	18.97	35.63	57.47	42.53	6.896	-0.099

E3 was a short gene of 192 bases. The gene exhibited a high level of similarity across the strains, Comoros, KPA and Lamu being exact matches and S27 being at a little distance away. The molecular weight of the protein encoded in this gene was 67.5 kDa for the 3 strains but S27 had a heavier protein weighing 67.6 kDa. In all the 4 strains, the isoelectric point, charge at pH 7.0, are the same. Compared to the S27 strain, A+T and G+C content had no significant difference, (Table 6).

3.3.2 Analysis of E1 gene

Table 7: Summary of E1 gene.

Strain	Nucleotide translated (No.)				Mol % of nucleotides				G+C%	A+T %	Isoelectric pt.	Charge at pH 7.0
	A	G	T	C	A	G	T	C				
S27	323	314	263	339	26.07	25.34	21.23	27.36	52.70	47.30	6.449	-3.591
Com25	320	316	264	339	25.83	25.50	21.31	27.36	52.87	47.13	6450	-3.590
KPA15	320	316	264	339	25.83	25.50	21.31	27.36	52.87	47.13	6450	-3.590
Lamu33	321	315	265	338	25.91	25.42	21.39	27.28	52.70	47.30	6450	-3.590

E1 gene was 1317 base pairs long. The Comoros25, KPA15 and Lamu33 strains had sequences inferring a protein of 44.4 kDa molecular weight each, while S27 had 44.5 kDa. There were no significant differences in the G+C and A+T content, (Table 7).

3.4 Variations in amino acids among the strains analyzed.

TABLE 8: Comparative analysis of No. of amino acids in all genes among isolates.

Size (no. of amino acids)				
Gene	S27 African Prototype	Comoros 25	KPA 15	Lamu 33
Nsp1	533	533	533	533
Nsp2	708	708	708	708
Nsp3	522	522	515	515
Nsp4	554	554	554	554
Capsid	157	157	157	157
E3	58	58	58	58
E2	354	354	354	354
6K	61	61	61	61
E1	413	413	413	413

Apart from the nsp3 gene, all other genes had equal number of amino acids. The nsp3 gene was different across the scope in that S27 and Comoros25 strains translated gave 522 amino acids while KPA15 and Lamu33 had 515 amino acids. This was caused by an early stop codon in the latter two which was due to substitution of cytosine nucleotide with thymine in the latter two strains. Consequently, a total of 21 nucleotides in this gene were locked out of the open reading frame thus lesser amino acids translated.

TABLE 9. List of amino acid differences associated with CHIKV sequences analysed.

Gene	Nucleotide position	KPA 15	Lamu 33	Comoros 25	S27
Nsp1	72	met	met	arg	met
	82	cys	cys	ser	cys
	171	arg	arg	gln	arg
	172	val	val	val	leu
	234	Lys	Lys	Lys	glu
	379	thr	met	thr	thr
	383	leu	leu	Leu	met
	384	leu	leu	Leu	ile
	458	thr	ser	ser	ser
	481	ile	ile	ile	thr
	488	arg	arg	arg	gln
	507	arg	arg	arg	leu
Nsp2	63	ala	val	val	val
	284	tyr	tyr	tyr	his
	552	tyr	tyr	tyr	cys
	553	asn	asn	asn	ser
	703	val	val	val	ala
Nsp3	167	ile	ile	ile	val
	209	his	his	his	tyr
	224	leu	gln	leu	leu
	318	ser	ser	ser	pro
	323	ala	ala	ala	val
	329	ile	ile	ile	thr
	344	glu	glu	glu	lys
	368	thr	thr	thr	ile
	374	thr	thr	thr	ala
	453	pro	pro	pro	leu
	454	asn	asn	asn	ser
	463	ser	ser	ser	pro
Nsp4	18	ala	ala	ala	thr
	197	ala	ala	ala	thr
	410	tyr	tyr	asp	asp
	443	leu	leu	leu	gln
	457	thr	thr	thr	ile
	498	ile	ile	ile	val
	547	ile	ile	ile	val
E3	17	thr	thr	thr	ile
E2	5	met	met	met	ile
	10	glu	glu	glu	gly
	91	thr	thr	thr	asn
	95	thr	thr	thr	ala
	112	met	met	met	leu
	125	gly	gly	gly	ser
	142	thr	thr	ile	ile
	198	arg	arg	arg	met
	230	asn	asn	asn	ser
	243	met	met	met	thr
	275	thr	thr	thr	ala
	306	thr	thr	thr	ser
	317	ala	ala	ala	val
6K	8	ile	ile	ile	val
	54	val	val	val	ile
E1	243	val	val	val	met
	258	glu	glu	glu	asp
	296	ala	ala	ala	val

Comparisons of the four CHIKV genomes revealed a total of 18 amino acids that were variable in any of the strains examined (Table 9) which were distributed across 55 regions of the genome. The nsp1 gene had 8 amino acids that were variable at 12 points of the gene. The nsp2 had 4 variable amino acids at 5 points of the gene while there were 11 variable amino acids across 12 points of the nsp3 gene. The nsp4 revealed 6 amino acids that were variable across 7 points of the gene. At position 410 of this gene, KPA 15 and Lamu 33 both had tyrosine while Comoros 25 and S27 had aspartic acid at this position. The capsid had no difference in amino acids while the E3 gene had only isoleucine being variable at position 17 of the gene. 10 amono acids were variable at 13 points of the E2 gene. Threonine appeared at position 142 of KPA 15 and Lamu 33 while isoleucine was found at this position of the Comoros 25 and S27. There appeared to be somewhat an interchange of amino acids valine and isoleucine at the 8^{th} and 54^{th} positions of the 6K gene. 3 amino acids were variable at 3 points of the E1 gene.

3.5 Amino acid sequences alignment of nsp4, E2 and 6K genes of CHIKV strains.

```
S27_nsp4     YIFSSDTGPG HLQQKSVRQS VLPVNTLEEV HEEKCYPPKL DEAKEQLLLK KLQESASMAN RSRYQSRKVE NMKATIIQRL
Com25_nsp4   .......... .......... .......... .......... .......... .......... .......... ....A.....
KPA15_nsp4   .......... .......... .......... .......... .......... .......... .......... ....A.....
Lamu33_nsp4  .......... .......... .......... .......... .......... .......... .......... ....A.....


S27_nsp4     KRGCRLYLMS ETPKVPTYRT TYPAPVYSPP INVRLSNPES AVAACNEFLA RNYPTVSSYQ ITDEYDAYLD MVDGSESCLD
Com25_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
KPA15_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
Lamu33_nsp4  .......... .......... .......... .......... .......... .......... .......... ..........


S27_nsp4     RATFNPSKLR SYPKQHAYHA PSIRSAVPSP FQNTLQNVLA AATKRNCNVT QMRELPTLDS AVFNVECFKK FACNQEYWEE
Com25_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
KPA15_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
Lamu33_nsp4  .......... .......... .......... .......... .......... .......... .......... ..........


S27_nsp4     FAASPIRITT ENLTTYVTKL KGPKAAALFA KTHNLLPLQE VPMDRFTVDM KRDVKVTPGT KHTEERPKVQ VIQAAEPLAT
Com25_nsp4   .......... ...A...... .......... .......... .......... .......... .......... ..........
KPA15_nsp4   .......... ...A...... .......... .......... .......... .......... .......... ..........
Lamu33_nsp4  .......... ...A...... .......... .......... .......... .......... .......... ..........


S27_nsp4     AYLCGIHREL VRRLNAVLLP NVHTLFDMSA EDFDAIIAAH FKPGDTVLET DIASFDKSQD DSLALTALML LEDLGVDHSL
Com25_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
KPA15_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
Lamu33_nsp4  .......... .......... .......... .......... .......... .......... .......... ..........


S27_nsp4     LDLIEAAFGE ISSCHLPTGT RFKFGAMMKS GMFLTLFVNT LLNITIASRV LEDRLTKSAC AAFIGDDNII HGVVSDELMA
Com25_nsp4   .......... .......... .......... .......... .......... .......... .......... ..........
KPA15_nsp4   .......... .......... .......... .......... .......... .......... ......Y... ..........
Lamu33_nsp4  .......... .......... .......... .......... .......... .......... ......Y... ..........


S27_nsp4     ARCATWMNME VKIIDAVVSQ KAPYFCGGFI LHDIVTGTAC RVADPLKRLF KLGKPLAAGD EQDEDRRRAL ADEVVRWQRT
Com25_nsp4   .......... .........L .......... ...T...... .......... .......... .......... ....I.....
KPA15_nsp4   .......... .........L .......... ...T...... .......... .......... .......... ....I.....
Lamu33_nsp4  .......... .........L .......... ...T...... .......... .......... .......... ....I.....


S27_nsp4     GLIDELEKAV YSRYEVQGIS VVVMSMATFA SSRSNFEKLR GPVVTLYGGP K
Com25_nsp4   .......... .......... .......... .......... ...I...... .
KPA15_nsp4   .......... .......... .......... .......... ...I...... .
Lamu33_nsp4  .......... .......... .......... .......... ...I...... .
```

FIG.6. Comparative amino acid sequences alignment of nsp4 gene of Chikungunya among isolates based on S27. Sequences aligned and amino acid differences denoted by the single letter code with the S27 sequence shown above in order to compare it with the rest of the strains.

Nsp4 gene had amino acids threonine (T), aspartic acid (D), tyrosine (Y), glutamine (Q), isoleucine (I), and valine (V) being variable at the 7 positions of the gene (Figure 6).

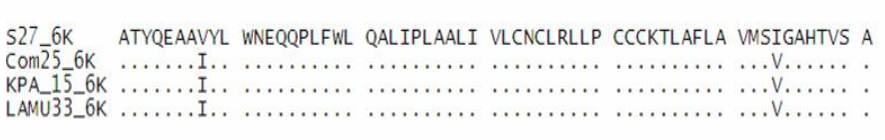

FIG.7. Comparative amino acid sequences alignment of 6K gene of CHIKV among isolates based on S27. Sequences aligned and amino acid differences denoted by the single letter code with the S27 sequence shown above in order to compare it with the rest of the strains.

The 6 K gene varied at positions 8 and 54 only where S27 had valine (V) and isoleucine (I) respectively and the rest having isoleucine (I) and valine (V) respectively. This appears as an interchange of the nucleotides.

3.6 Comparative analysis of nucleotide sequences among isolates.

TABLE 10: % Identity of genes of CHIKV strains compared with S27 African Prototype.

	Comoros 25	KPA 15	Lamu 33
Nsp1	97.7	97.7	97.7
Nsp2	97.7	97.7	97.7
Nsp3	96.2	96.2	96.1
Nsp4	97.4	97.3	97.4
Capsid	98.2	98.2	98.0
E3	97.9	97.9	97.9
E2	96.6	96.5	96.6
6K	97.3	97.8	97.8
E1	97.6	97.6	97.6

The variations of Comoros25, KPA15 and Lamu33 from the S27 African prototype were minimal across all the genes. That is the nucleotide sequences of each of the genomic strains appeared to be conserved. The comparison above showed that there was high level of homology across the genes with the capsid as would be expected (see analysis of capsid) having the highest percentage of similarity.

3.7 Comparison of Chikungunya virus whole genome sequences.

Table 11; Comparative Analysis of CHIKV whole genome sequences among Isolates

Pair Distances of ClustalW (weighted)

Percent Similarity in upper triangle

Percent Divergence in lower triangle

	S27	Comoros25	KPA15	Lamu33
S27		95.1	95.1	95.1
Comoros25	5.1		99.9	99.8
KPA15	5.0	0.1		99.9
Lamu33	5.0	0.2	0.1	
	S27	Comoros25	KPA15	Lamu33

The comparisons gave over 99% similarities among the isolates, KPA15, Lamu 33 and Comoros 25 while there was over 95 % similarity in comparison to the S27. Pairwise comparisons indicated that there was insignificant variability among the KPA 15, Lamu 33 and Comoros 25 isolates of CHIKV. This level of conservation indicates that the outbreak in Comoros islands originated from the 2004 outbreaks in coastal Kenya and may have subsequently moved to Reunion Island.

3.7.1 Comparison of 3′ and 5′ ends of Chikungunya virus whole genome sequences.

All of the 3 strains, Comoros25, KPA15 and Lamu33 compared to the S27 strain had a consensus 5′ sequence which was GAGATGTTATTTTGTTTTTAATATTTC ; and the 3′ sequence was ATGGCTGCGTGAGACACACGT AGCC just right before the polyA tail. This was further evidence to the genetic conservation of the viruses.

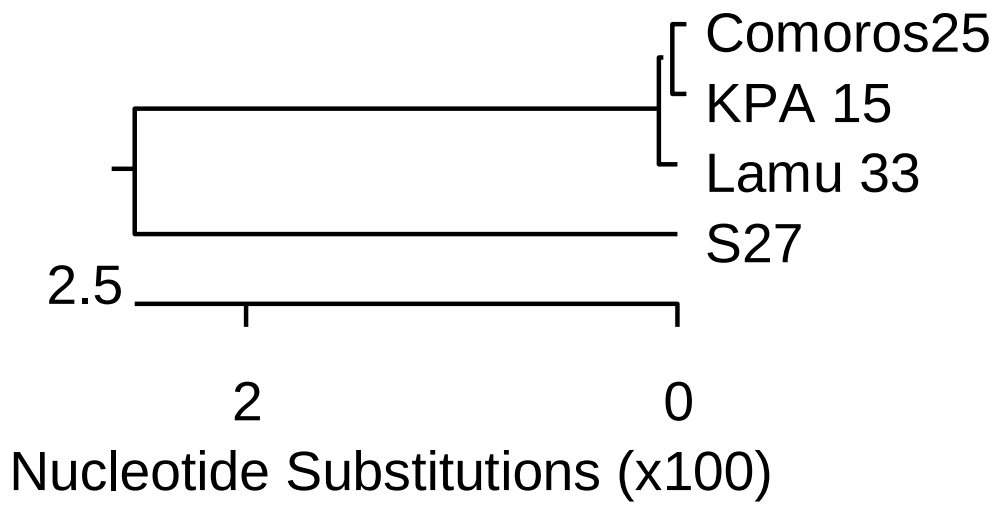

FIG.8. Phylogenetic relationships among CHIKV strains with S27 as reference. After alignment, a rooted phylogram was generated by Clustal W (weighted).

The phylogram showed the Comros25, KPA15 and Lamu33 clustered together on the tree supporting further their over 99% nucleotide sequence identity.

With over 95% nucleotide sequence identity, S27 was just at a little distance indicating that the isolates causing infection in the Indian Ocean islands and the S27 African prototype were relatively homogenous.

CHAPTER 4: DISCUSSION, CONCLUSION AND RECOMMENDAIONS

4.1 Discussion

The study revealed that there was a high level of nucleotide sequence homology among the isolates and also with the S27 African prototype strain. Comparison of each of the genes among the isolates showed over 96% nucleotide sequence identity and the number of amino acids predicted in each gene was equal across the strains e.g. the inferred number of amino acids from nsp1 gene sequence was 533 in all the strains under study. Such findings have also been documented about the haemagggglutinin in H1N1 swine influenza viruses where a high degree of conservation was reported with no amino acid substitutions observed relative to a reference strain (Noble *et al.*, 1993).

This genetic simlarity of CHIKV was also observed among Asian genotype viruses. Interesting to note also is the fact that sequences from viruses spanning a wide geographical range and isolated over a period close to 35 years showed less than 3% nucleotide sequence divergence (Powers *et al*., 2000). This contrasts the observation made of RNA viruses having a high error rate of RNA synthesis thus, a high mutation potential (Steinhauer and Holland, 1987). Genetic sequence conservation is also apparent in other in other groups of alphaviruses. The Eastern Equine Encephalomyelitis virus revealed remarkable conservation of the nucleotide sequence (Weaver *et al.*, 1991; Weaver *et al.*, 1993). Within a given year, the findings of (Weaver *et al*., 1993) were that genetic diversity was generally greater among geographically distant isolates than among those from the same transmission focus. Brault and others confirmed that there was indeed conservation of EEEV

strains from North America but consensus was lost among the South and Central America isolates (Brault *et al.*, 1999). In North America, the elaborate transmission of EEEV occurs among the songbirds in fresh water swamps by the mosquito vector while the virus is probably transmitted among small mammals and/or birds in enzootic foci in Central and North America (Weaver *et al.,* 1994). The limited mobility of mammalian reservoir hosts may be especially important in determining the degree of genetic isolation experienced by geographically isolated EEEV foci in the tropics while, the avian reservoir hosts in North America may provide for more efficient dispersal of EEEV (Brault *et al*., 1999).

This relative homogeneity of the CHIKV strains studied indicates that there could be an actual spread of the virus. With Kenya's Lamu islands being the first to have an outbreak occurring then Mombasa and subsequently the Comoros islands, it is possible that the same virus was spread from Lamu to the rest of these places. The spread could have been by the vectors which thrive in the wet and humid climates of Tropical Africa and/or people traveling. There has been evidence to suggest that Chikungunya virus spread from Tropical Africa (where it was first identified) to Asia (Powers *et al.,* 2000). The spread of the virus in Africa is owed chiefly to the sylvatic transmission cycle involving wild primates and forest dwelling mosquitoes. CHIKV has been isolated from forest mosquitoes, with most of them isolated from *Ae. furcifer-taylori, Ae. luteocephalus,* and *Ae. Dalzieli* (Diallo *et al.,* 1999).

The isolation of the virus from non-human primates, other vertebrates such as squirrels and bats, and zoophilic mosquito species (that feed on animals) supports the existence of sylvatic transmission cycles in Africa, which may maintain the virus in the wild during inter-epidemic years. In Asia, transmission appears to be mainly from *Aedes aegypti* and *Ae. albopictus* to human in urban settings (Sam and AbuBakar, 2006). In Africa, dispersal of some arboviruses could now be aided by the well spread Asian tiger mosquito *Aedes* (stegomyia) *albopictus* (Skuse) (Pages *et al.*, 2006). This vector could be responsible for the spread of CHIKV in many parts of the globe due to its high adaptability.

Another substantiating fact to this argument is that the three virus strains from each of these places had most of their substitutions at similar positions and with the same nucleotides. The explanation to this is likely that as the virus spread, it was subjected to similar geographic pressures causing the strains to undergo selective but similar mutations so as to adapt to the regions. It has been observed that despite the fact that RNA virus populations are capable of rapid evolution in response to new cellular environments, there exists different selective pressures for virus replication in vertebrate and invertebrate cells, and that transfer in a two-host cycle selects for virus populations that are well adapted for replication in both hosts (Lynn and Thomas, 2001). Subjected to the same selective pressures, viruses from a common ancestry appear to follow the same path of growth, independently showing identical traits in response to common selective pressures (BULL *et al.*, 1997). The findings were supported by the analysis of CHIKV genomes.

There is evidence from the results to suggest the possibility of CHIKV, originally from Tanganyika, to have spread across the continent, showing similar rates of substitution and similar rates of fitness improvement across corresponding times of adaptation. This finding is in agreement with the hypothesis that the same virus could have been responsible for the outbreaks of disease, from whence isolates were obtained and consistent with the observation that arboviruses are conserved over time and space (Lynn and Thomas, 2001; Weaver *et al*., 1991; Weaver *et al*., 1993; Brault *et al*., 1999).

4.2 Conclusion and Recommendations

CHIKV outbreaks are increasing in both frequency and magnitude. Its symptoms are shifting from mere limited mobility to hemorrhage and fatality. Cost implications in managing outbreaks are usually staggering to governments, health bodies and to the local community. It is therefore important that a vaccine for Chikungunya virus is developed so that people are immunized thus preventing outbursts of epidemics. The nucleotide sequence data generated in this study is useful in that it shows areas of high tendancy of mutation and those that are relatively conserved e.g. the capsid gene. The conserved sequence in the capsid gene could be used to design peptides which could be tested for possible vaccine development. The nucleotide sequences could also be useful in designing RNA based drugs to block the replication of CHIKV. This would be helpful in treatment of the disease. The sequence data is also helpful in predicting of circulating sequences. Also highly recommended is the proper use of treated bed nets and control of mosquitoes.

Governments and all stakeholders in policy making should continue relentlessly with the campaign to control the mosquito.

In case of an outbreak, travels should be closely monitored in the areas to minimize chances of spread of the virus.

LIST OF REFERENCES

ABI Prism 3100 Genetic Analyzer, Perkin-Elmer/Applied Biosystems.

Afjal Hossain Khan[1], Kouichi Morita[1], Maria del Carmen Parquet[1], Futoshi Hasebe[1],Edward G. M. Mathenge[1] and Akira Igarashi . **Complete nucleotide sequence of Chikungunya virus and evidence for an internal polyadenylation site.** *Journal of General Virology.* 83: 3075-3084, 2002.

Alexis Carrel and Montrose T. Burrows.**Cultivation of Tissues in Vitro and its Techniques**. *Journal of Experimental Medicine.* 387-406, 1911.

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman"**Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**", *Nucleic Acids Research.* 25:3389-3402, 1997.

Ann M. Powers, Aaron C. Brault, Robert B. Tesh and Scott C. Weaver. **Re-emergence of Chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships**. *Journal of General Virology.* 81:471-479, 2000.

Aphis Veterinary services; **Venezuelan Equine Encephalomyelitis:** *aphis.usda.gov*,2002.

B. Lee Ligon, **Reemergence of an Unusual Disease: The Chikungunya Epidemic** *Seminars in Pediatric Infectious Diseases*. 17: 299-104, 2006.

Boutin J.P. **Chikungunya fever in La Reunion Island—2006**. *Medecine Tropicale (Mars): revue du corps de sante colonial*. 66(3):221-5, 2006. [Article in French]

Brault AC, AM Powers, CL Chavez, RN Lopez, MF Cachon, LF Gutierrez, W Kang, RB Tesh, RE Shope, and SC Weaver.**Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America.** *American Journal of Tropical Medicine and Hygiene*. 61(4): 579-586, 1999.

Büchen-Osmond, C. (Ed). 00.073.0.01.007. **Chikungunya virus In: ICTVdB - The Universal Virus Database, version 3.** *Büchen-Osmond, C. (Ed), ICTVdB Management, Columbia University, New York, USA*, 2004.

Bull J. J. , M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. **Exceptional Convergent Evolution in a Virus.** *Genetics*. 147(4): 1497–1507, 1997.

Chastel Claude. **Chikungunya virus: its recent spread to the southern Indian Ocean and Reunion Island (2005-2006)** *Bulletin de I'Academie Nationale de Medecine*. 189(8):1827-35, 2005. [Article in French]

Chomczynski P, Sacchi N. **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.** *Analalytical Biochemistry*. 162(1):156-9, 1987.

Courtney Peterson (2001). **How It Works: The Charged-Coupled Device, or CCD.** *Journal of Young Investigators.* Issue 1, March 2001.

Cunningham Ann, Sonja M. Buckley and J. Casals. **Isolation of Chikungunya Virus Contaminating an *Aedes albopictus* Cell Line.** *Journal of General Virology*. 27: 97-100, 1975.

Dana L. Vanlandingham, Chao Hong, Kimberly Klingler, Konstantin Tsetsarkin, Kate L. Mcelroy, Ann M. Powers, Michael J. Lehane, And Stephen Higgs. **Differential Infectivities Of O'nyong-Nyong And Chikungunya Virus Isolates In *Anopheles Gambiae* And *Aedes Aegypti* Mosquitoes.** *The American Journal of Tropical Medicine and Hygiene.* 72(5): 616-621, 2005.

DNASTAR Inc., MadisonWI.

Eagle H. **Nutritional Needs of Mammalian Cells in Tissue Culture**. *Science*. 122:501-504, 1955.

Fourcade S, Simon F, Morand J. **Chikungunya: a case of pain rash and fever in a patient returning from the Indian Ocean.** *Annales de Dermatologie et de Venereologie*. 133(6-7):549-5, 2006. [Article in French]

Hundekar SL, Thakare JP, Gokhale MD,Barde PV,Argade SV, Mourya DT. **Development of monoclonal antibody based antigen capture ELISA to detect Chikungunya virus antigen in mosquitoes.** *Indian Journal of Medical Research*. 115:144-8, 2002.

Hongbao Ma. **ELISA Technique.** *Nature and Science*. 4(2), 2006.

Invitrogen, life technologies, Carlsbad California.

Isabelle Schuffenecker, Isabelle Iteman, Alain Michault, Séverine Murri, Lionel Frangeul, Marie-Christine Vaney, Rachel Lavenir, Nathalie Pardigon, Jean-Marc Reynes, François Pettinelli, Leon Biscornet, Laure Diancourt, Stéphanie Michel, Stéphane Duquerroy, Ghislaine Guigon, Marie-Pascale Frenkiel, Anne-Claire Bréhin, Nadège Cubito, Philippe Desprès, Frank Kunst, Félix A Rey, Hervé Zeller, and Sylvain Brisse.**Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak**. *Public Library of Science (PLoS) Medicine*. 3(7): e263, 2006.

J. H. Strauss and E. G. Strauss. **The alphaviruses: gene expression, replication and evolution.** *Microbiology and Molecular Biology Reviews*. 58(3): 491-562, 1994.

Kanti Laras, Nono C. Sukri Ria P. Larasati, Michael J. Bangs, Rizal Kosim, Djauzi, Tony Wandra, John Master, Herman Kosasih, Sri Hartati, Charmagne Beckett, Endang R. Sedyaningsih, H. James Beecham III and Andrew L. Corwin. **Tracking the re-emergence of epidemic Chikungunya virus in Indonesia.** *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 99(2):128-41, 2005.

Kit L.S.; **Emerging and re-emerging diseases in Malaysia.** *Asia-Pacific Journal of Public Health*. 14(1):6-8, 2002.

Kumarasamy V, Prathapa S, Zuridah H, Chem YK, Norizah I, Chua KB. **Re-emergence of Chikungunya virus in Malaysia.** *Medical Journal of Malaysia*. 61(2):221-5, 2006.

Laine M, Luukkainen R, Toivanen A (Keuruu Health Center, Keuruu; Satakunta Central Hospital, Rauma; and Turku University, Turku, Finland). **Sindbis viruses and other alphaviruses as cause of human arthritic disease (Review)**. *Journal of Internal Medicine*. 256: 457–471, 2004.

Lenglet Y, Barau G, Robillard PY, Randrianaivo H, Michault A, Bouveret A, Gerardin P, Boumahni B, Touret Y, Kauffmann E, Schuffenecker I, Gabriele M, Fourmaintraux A. **Chikungunya infection in pregnancy: evidence for intrauterine infection in pregnant women and vertical transmission in the parturient. Survey**

of the Reunion Island outbreak. *Journal de Gynecologie, Obstetrique et Biologie de la Reproduction (Paris).* 35(6):578-583, 2006. [Article in French]

Levinson RS, Strauss JH, Strauss EG. **Complete sequence of the genomic RNA of O'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees.** *Virology*. 175(1):110-23, 1990.

Logue, C.H. and Atkins, G.J. **Microbiology**. *Trinity College Dublin, College green, Dublin Dublin 2, Ireland*, 2002.

Lumsden WHR. **"An Epidemic of Virus Disease in Southern Province, Tanganyika Territory, in 1952-53; II. General Description and Epidemiology".** *Transactions of The Royal Society of Tropical Medicine and Hygiene.* 49 (1): 33-57, 1955.

Lynn A. Cooper and Thomas W. Scott. **Differential Evolution of Eastern Equine Encephalitis Virus Populations in Response to Host Cell Type.** *Genetics.* 157: 1403-1412, 2001.

M Diallo, J Thonnon, M Traore-Lamizana, and D Fontenille. **Vectors of Chikungunya virus in Senegal: current data and transmission cycles.** *American Journal of Tropical Medicine and Hygiene*. 60(2): 281-286, 1999.

Maël Bessaud, Christophe N. Peyrefitte, Boris A.M. Pastorino, Patrick Gravier, Fabienne Tock, Fabrice Boete, Hugues J. Tolou, and Marc Grandadam. **O'nyong-nyong Virus, Chad;** *CDC Emerging Infectious Diseases Journal Home*. 12; 8, 2006.

Mallet F, Oriol G, Mary C, Verrier B, Mandrand B.**Continuous RT-PCR using AMV-RT and Taq DNA Polymerase: characterization and comparison to uncoupled procedures.** *Biotechniques*. 18(4):678-87, 1995.

Maurice R. Atkinson, Murray P. Deutscher, Arthur Kornberg, Alan F. Russell, and J.G. Moffatt. **Enzymatic Synthesis of Deoxyribonucleic Acid. XXXIV. Termination of Chain Growth by a 2',3'-Dideoxyribonucleotide.** *Biochemistry*. 8: 4897, 1969.

Minna T. Tuittila, Maria G. Santagati, Matias Röyttä, Jorma A. Määttä, and Ari E. Hinkkanen. **Replicase Complex Genes of Semliki Forest Virus Confer Lethal Neurovirulence.** *Journal of Virology*. 74(10): 4579-89, 2000.

N.K. Blackburn, T.G. Besselaar and G. Gibson. **Antigenic relationship between** Chikungunya **virus strains and o'nyong nyong virus using monoclonal antibodies.** Research in Virology. 146(1): 69-73, 1995.

Noble Simon, Martha S. McGregor, David E. Went worth, Virginia S., Hinshaw. **Antigenic and genetic conservation of the haemagglutinin in H1N1 swine influenza viruses**. *Journal of virology*. 74: 1197-1200, 1993.

Paganin F, Borgherini G, Staikowsky F, Arvin-Berod C, Poubeau P. **Chikungunya on Reunion Island: chronicle of an epidemic foretold.** *Presse Medicale (Paris, France: 1983)*. 35(4 Pt 2):641-6, 2006. [Article in French]

Pages F., Corbel V., Paupy C. **Aedes albopictus: chronical of a spreading vector.** *Medecine Tropicale (Mars): revue du corps de sante colonial.* 66(3):226-8, 2006. [Article in French]

Pastorino B, Muyembe-Tamfum JJ, Bessaud M, Tock F, Tolou H, Durand JP, Peyrefitte CN.**Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain.** *Journal of Medical Virology*. 74(2):277-82, 2004.

Pekka Kujala, Anne Ikäheimonen, Neda Ehsani, Helena Vihinen, Petri Auvinen, and Leevi Kääriäinen. **Biogenesis of the Semliki Forest Virus RNA Replication Complex.** *Journal of Virology*. 75(8): 3873-3884, 2001.

Pialoux G, Gauzere BA, Strobel M. **Chikungunya virus infection: review through an epidemic.** *Medecine et Maladies Infectieuses*. 36(5):253-63, 2006. [Article in French]

Prescott, L. **Microbiology**, *Wm. C. Brown Publishers*, ISBN 0-697-01372-3, 1993.

QIAgen Inc, Valencia, CA

Ravi V. **Re-emergence of Chikungunya virus in India.** *Indian Journal of Medical Microbiology*. 24:83-84, 2006.

Robert Breiman. Personal communication, 2005.

Robillard PY, Boumahni B, Gerardin P, Michault A, Fourmaintraux A, Schuffenecker I, Carbonnier M, Djemili S, Choker G, Roge-Wolter M, Barau G. **Vertical maternal fetal transmission of the Chikungunya virus. Ten cases among 84 pregnant women.** *Presse Medicale (Paris, France: 1983)*. 35(5 Pt 1):785-8, 2006. [Article in French]

Robinson Marion. "**An Epidemic of Virus Disease in Southern Province, Tanganyika Territory, in 1952-53; I. Clinical Features**". *Transactions of The Royal Society of Tropical Medicine and Hygiene*. 49 (1): 28-32, 1955.

S.K. Lam and K.B. Chua. **Alpha viruses**. WHO *Collaborating Centre for Arbovirus Reference and Research (DF/DHF), Department of Medical Microbiology, Faculty of Medicine, University of Malaya*, 2005.

Sam I.C. and AbuBakar. **Chikungunya Virus Infection;** http://www.mma.org.my/info/2_chikungunya_06.htm

Sam I.C. and AbuBakar S. **Chikungunya Virus Infection.** *Medical Journal of Malaysia*. 61(2):264-9, 2006.

Sambrook Joseph and David W.Russel. **Molecular Cloning: A Laboratory Manual**. **2nd Edition**. *Cold Spring Harbor Press: New York*, 1989.

Sambrook, Joseph and David W. Russell,. **Molecular Cloning: A Laboratory Manual** 3rd Edition *Cold Spring Harbor Press: New York.* 370-379, 2001.

Sanger F., S. Nicklen, and A. R. Coulson. **DNA sequencing with chain-terminating inhibitors.** *Proceedings of the National Academy of Sciences of U S A.* 74(12): 5463–5467, 1977.

Severini A, Morgan AR. **An assay for proteinases and their inhibitors based on DNA/ ethidium bromide fluorescence.** *Analytical Biochemistry.* 193(1); 83-9, 1991.

Smithsonian Institution Archives. The History of PCR (RU 9577).

Steinhauer DA, Holland JJ. **Rapid evolution of RNA viruses.** *Annual Review of Microbiology*. 41:409-33, 1987.

Steve Rozen and Helen J. Skaletsky (2000) **Primer3** on the WWW

Thaikruea L, Charearnsook O, Reanphumkarnkit S, Dissomboon P, Phonjan R, Ratchbud S, Kounsang Y, Buranapiyawong D. **Chikungunya in Thailand: a re-emerging disease?** *Southeast Asian Journal of Tropical Medicine and Public Health.* 28(2):359-64, 1997.

Volk *et al*. **Entails of Medical Microbiology**, **fifth edition**. *Lippincott-Raven publishers, Philadelphia,* 1996.

Warfield KL, Swenson DL, Olinger GG, Nichols DK, Pratt WD. **Gene-specific countermeasures against Ebola virus based on antisense phosphorodiamidate morpholino oligomers**. *PLoS Pathog* 2 (1): el, 2006.

Weaver SC, Bellew LA, Gousset L, Repik PM, Scott TW, Holland JJ. **Diversity within natural populations of eastern equine encephalomyelitis virus.** *Virology*. 195(2):700-9, 1993.

Weaver SC, Scott TW, Rico-Hesse R. **Molecular evolution of eastern equine encephalomyelitis virus in North America.** *Virology*. 182(2):774-84, 1991.

Weston J, Villoing S, Bremont M, Castric J, Pfeffer M, Jewhurst V, McLoughlin M, Rodseth O, Christie KE, Koumans J, Todd D.**Comparison of two aquatic alphaviruses, salmon pancreas disease virus and sleeping disease virus, by using genome sequence analysis, monoclonal reactivity, and cross-infection.** *Journal of Virology*. 76(12):6155-63, 2002.

www.geographyiq.com/countries/ke/Kenya_climate_f.htm

www.phac-aspc.gc.ca Public Health Agency of Canada: Material Safety Data Sheets- Infectious Substances.

Yasamura Y.and Y. Kawakita. **A Line of Cells Derived from African Green Monkey Kidney**. *Nippon Rinsho*. 21:1209-1210, 1963.

APPENDICES

Appendix 1; Lamu 33- whole genome sequence

atggctgcgt gagacacacg tagcctacca gtttcttact gctctactctgcaaagcaag agattaataa cccatcatgg atcctgtgta cgtggacatagacgctgaca gcgccttttt gaaggccctg caacgtgcgt accccatgtttgaggtggaa ccaaggcagg tcacaccgaa tgaccatgct aatgctagagcgttctcgca tctagctata aaactaatag agcaggaaat tgaccccgactcaaccatcc tggatatcgg cagtgcgcca gcaaggagga tgatgtcggacaggaagtac cactgcgtct gcccgatgcg cagtgcggaa gatcccgagagactcgccaa ttatgcgaga aagctagcat ctgccgcagg aaaagtcctggacagaaaca tctctggaaa gatcggggac ttacaagcag taatggccgt

gccagacacg gagacgccaa cattctgctt acacacagac gtctcatgtagacagagagc agacgtcgct atataccaag acgtctatgc tgtacacgcacccacgtcgc tataccacca ggcgattaaa ggggtccgag tggcgtactgggttgggttc gacacaaccc cgttcatgta caatgccatg gcgggtgcctacccctcata ctcgacaaac tgggcagatg agcaggtact gaaggctaagaacataggat tatgttcaac agacctgacg gaaggtagac gaggcaagttgtctattatg agagggaaaa agctaaaacc gtgcgaccgt gtgctgttctcagtagggtc aacgctctac ccggaaagcc gcaagctact taagagctggcacctgccat cggtgttcca tttaaagggc aaactcagct tcacatgccg

ctgtgataca gtggtttcgt gtgagggcta cgtcgttaag agaataacgatgagcccagg cctttatgga aaaaccacag ggtatgcggt aacccaccacgcagacggat tcctgatgtg caagactacc gacacggttg acggcgaaagagtgtcattc tcggtgtgca catacgtgcc ggcgaccatt tgtgatcaaatgaccggcat ccttgctaca gaagtcacgc cggaggatgc acagaagctgttggtggggc tgaaccagag aatagtggtt aacggcagaa cgcaacggaatatgaacacc atgaaaaatt atctgcttcc cgtggtcgcc caagccttcagtaagtgggc aaaggagtgc cggaaagaca tggaagatga aaaactcctgggggtcagag aaagaacact gacctgctgc tgtctatggg attcaagaa

gcagaaaaca cacacggtct acaagaggcc tgatacccag tcaattcagaaggttcaggc cgagtttgac agctttgtgg taccgagtct gtggtcgtccgggttgtcaa tccctttgag gactagaatc aaatggttgt taagcaaggtgccaaaaacc gacctgatcc catacagcgg agacgcccga gaagcccgggacgcagaaaa agaagcagag gaagaacgag aagcagaact gactcgcgaagccctaccac ctctacaggc agcacaggaa gatgttcagg tcgaaatcgacgtggaacag cttgaggaca gagcgggcgc aggaataata gagactccgagaggagctat caaagttact gcccaaccaa cagaccacgt cgtgggagagtacctggtac tctccccgca gaccgtacta cgtagccaga agctcagtctgattcacgct ttggcggagc aagtgaagac gtgcacgcac aacggacgagcagggaggta tgcggtcgaa gcgtacgacg gccgagtcct agtgccctcaggctatgcaa tctcgcctga agacttccag agtctaagcg aaagcgcaac

gatggtgtat aacgaaagag agttcgtaaa cagaaagcta caccatattgcgatgcacgg accagccctg aacaccgacg aagagtcgta tgagctggtgagggcagaga ggacagaaca cgagtacgtc tacgacgtgg atcagagaagatgctgtaag aaggaagaag ccgcaggact ggtactggtg ggcgacttgactaatccgcc ctaccacgaa ttcgcatatg aagggctaaa aatccgccct

gcctgcccat acaaaattgc agtcatagga gtcttcggag taccgggatctggcaagtca gctattatca agaacctagt taccaggcag gacctggtgactagcggaaa gaaagaaaac tgccaagaaa tcaccaccga cgtgatgagacagagaggtc tagagatatc tgcacgtacg gttgactcgc tgctcttgaatggatgcaac agaccagtcg acgtgttgta cgtagacgag gcgtttgcgt

gccactctgg aacgctactt gctttgatcg ccttggtgag accaaggcagaaagttgtac tttgtggtga cccgaagcag tgcggcttct tcaatatgatgcagatgaaa gtcaactata atcacaacat ctgcacccaa gtgtaccacaaaagtatctc caggcggtgt acactgcctg tgaccgccat tgtgtcatcgttgcattacg aaggcaaaat gcgcactacg aatgagtaca acaagccgattgtagtggac actacaggct caacaaaacc tgaccctgga gacctcgtgttaacgtgctt cagagggtgg gttaaacaac tgcaaattga ctatcgtggatacgaggtca tgacagcagc cgcatcccaa gggttaacca gaaaaggagtttacgcagtt agacaaaaag ttaatgaaaa cccgctctat gcatcaacgtcagagcacgt caacgtactc ctaacgcgta cggaaggtaa actggtatggaagacacttt ccggcgaccc gtggataaag acgctgcaga acccaccgaaaggaaacttc aaagcaacta ttaaggagtg ggaggtggag catgcatcaataatggcggg catctgcagt caccaaatga ccttcgatac attccaaaataaagccaacg tttgttgggc taagagcttg gtccctatcc tcgaaacagcggggataaaa ctaaatgata ggcagtggtc tcagataatt caagccttcaaagaagacaa agcatactca cctgaagtag ccctgaatga aatatgtacg

cgcatgtatg gggtggatct agacagcggg ctattttcta aaccgttggtgtctgtgtat tacgcggata accactggga taataggcct ggagggaaaatgttcggatt taaccccgag gcagcatcca ttctagaaag aaagtatccattcacaaaag ggaagtggaa catcaacaag cagatctgcg tgactaccaggaggatagaa gactttaacc ctaccaccaa catcataccg gccaacaggagactaccaca ctcattagtg gccgaacacc gcccagtaaa aggggaaagaatggaatggc tggttaacaa gataaacggc caccacgtgc tcctggtcag

tggctataac cttgcactgc ctactaagag agtcacttgg gtagcgccgttaggtgtccg cggagcggac tacacataca acctagagtt gggtctgccagcaacgcttg gtaggtatga cctagtggtc ataaacatcc acacacctttcgcatacac cattaccaac agtgcgtcga ccacgcaatg aaactgcaaatgctcggggg tgactcattg agactgctca aaccgggcgg ctctctattgatcagagcat atggttacgc agatagaacc agtgaacgag tcatctgcgtattgggacgc aagtttagat cgtctagagc gttgaaacca ccatgtgtcaccagcaacac tgagatgttt ttcctattca gcaactttga caatggcagaaggaatttca caactcatgt catgaacaat caactgaatg cagccttcgt

aggacaggtc acccgagcag gatgtgcacc gtcgtaccgg gtaaaacgcatggacatcgc gaagaacgat gaagagtgcg tagtcaacgc cgctaaccctcgcgggttac cgggtgacgg tgtttgcaag gcagtataca aaaaatggccggagtccttt aagaacagtg caacaccagt gggaaccgca aaaacagttatgtgcggtac gtatccagta atccacgctg ttggaccaaa cttctctaat
tattcggagt ctgaagggga ccgggaattg gcagctgcct atcgagaagtcgcaaaggaa gtaactaggc tgggagtaaa tagtgtagct atacctctcctctccacagg tgtatactca ggagggaaag acaggctgac ccagtcactgaaccacctct ttacagccat ggactcgacg gatgcagacg tggtcatctactgccgcgac aaagaatggg agaagaaaat atctgaggcc atacagatgcggacccaagt agagctgctg gatgagcaca tctccataga ctgcgatattgttcgcgtgc accctgacag cagcttggca ggcagaaaag gatacagcaccacggaaggc gcactgtact catatctaga agggacccgt tttcatcagacggctgtgga tatggcggag atacatacta tgtggccaaa gcaaacagag
gccaatgagc aagtctgcca atatgccctg ggggaaagta ttgaatcgatcaggcagaaa tgcccggtgg atgatgcaga cgcatcatct ccccccaaaactgtcccgtg cctttgccgt tacgctatga ctccagaacg cgtcacccggcttcgcatga accacgtcac aagcataatt gtgtgttctt cgtttcccctcccaaagtac aaaatagaag gagtgcaaaa agtcaaatgc tctaaggtaatgctatttga ccacaacgtg ccatcgcgcg taagtccaag ggagtatagatcttcccagg agtctgcaca ggaggcgagt acaatcacgt cactgacgcatagtcaattc gacctaagcg ttgatggcga gatactgccc gtcccgtcagacctggatgc tgacgcccca gccctagaac cagcactaga cgacggggcg acacacacgc tgccatccac aaccggaaac cttgcggccg tgtctgactgggtaatgagc accgtacctg tcgcgccgcc cagaagaagg cgagggagaaacctgactgt gacatgtgac gagagagaag ggaatataac acccatggctagcgtccgat tctttagggc agagctgtgt ccggtcgtac aagaaacagcggagacgcgt gacacagcaa tgtctcttca ggcaccaccg agtaccgccacggaaccgaa tcatccgccg atctccttcg gagcatcaag cgagacgttccccattacat ttggggactt caacgaagga gaaatcgaaa gcttgtcttctgagctacta actttcggag acttcttacc aggagaagtg gatgacttgacagacagcga ctggtccacg tgctcagaca cggacgacga gttatgacta
gacagggcag gtgggtatat attctcgtcg gacaccggtc caggtcatttacaacagaag tcagtacgcc agtcagtgct gccggtgaac accctggaggaagtccacga ggagaagtgt tacccaccta agctggatga agcaaaggagcaactattac ttaagaaact ccaggagagt gcatccatgg ccaacagaagcaggtatcag tcgcgcaaag tagaaaacat gaaagcagca atcatccagagactaaagag aggctgtaga ctatacttaa tgtcagagac cccaaaagtccctacttacc ggactacata tccggcgcct gtgtactcgc ctccgatcaa
cgtccgattg tccaatcccg agtccgcagt ggcagcatgc aatgagttcttagctagaaa ctatccaact gtctcatcat accaaattac cgacgagtatgatgcatatc tagacatggt ggacgggtcg gagagttgcc tggaccgagcgacattcaat ccgtcaaaac tcaggagcta cccgaaacag cacgcttaccacgcgccctc catcagaagc gctgtaccgt ccccattcca gaacacactacagaatgtac tggcagcagc cacgaaaaga aactgcaacg tcacacagatgagggaatta cccactttgg actcagcagt attcaacgtg gagtgtttcaaaaaattcgc atgcaaccaa gaatactggg aagaatttgc tgccagccctattaggataa caactgagaa tttagcaacc tatgttacta aactaaaagg
gccaaaagca gcagcgctat tcgcaaaaac ccataatcta ctgccactacaggaagtacc aatggatagg ttcacagtag atatgaaaag ggacgtgaaggtgactcctg gtacaaagca tacagaggaa agacctaagg tgcaggttatacaggcggct gaacccttgg cgacagcata cctatgtggg attcacagagagctggttag gaggctgaac gccgtcctcc tacccaatgt acatacactatttgacatgt ctgccgagga tttcgatgcc atcatagccg cacactttaagccaggagac actgttttgg aaacggacat agcctccttt gataagagccaagatgattc acttgcgctt actgctttga tgctgttaga ggatttaggggtggatcact ccctgctgga cttgatagag gctgctttcg gagagatttc
cagctgtcac ctaccgacag gtacgcgctt caagttcggc gccatgatgaaatcaggtat gttcctaact ctgttcgtca acacattgtt aaacatcaccatcgccagcc gagtgctgga agatcgtctg acaaaatccg cgtgcgcggccttcatcggc gactacaaca taatacatgg agtcgtctcc gatgaattgatggcagccag atgtgccact tggatgaaca tggaagtgaa gatcatagatgcagttgtat ccttgaaagc cccttacttt tgtggagggt ttatactgcacgatactgtg acaggaacag cttgcagagt ggcagacccg ctaaaaaggctttttaaact gggcaaaccg ctagcggcag gtgacgaaca agatgaagatagaagacgag cgctggctga cgaagtgatc agatggcaac gaacagggctaattgatgag ctggagaaag cggtatactc taggtacgaa gtgcagggtatatcagttgt ggtaatgtcc atggccacct ttgcaagctc cagatccaacttcgagaagc tcagaggacc cgtcataact ttgtacggcg gtcctaaataggtacacact acagctacct attttgcaga agccgacagc aagtatctaaacactaatca gctacaatgg agttcatccc aacccaaact ttttacaataggaggtacca gcctcgaccc tggactccgc gccctactat ccaagtcatcaggcccagac cgcgccctca gaggcaagct gggcaacttg cccagctgat
ctcagcagtc aataaactga caatgcgcgc ggtaccccaa cagaagccacgcaggaatcg gaagaataag aagcaaaagc aaaaacaaca ggcgccacaaaacaacacaa atcaaaagaa gcagccacct aaaaagaaac cggctcaaaagaaaaagaag ccgggccgca gagagaggat gtgcatgaaa atcgaaaatg attgtatttt cgaagtcaag cacgaaggta aggtaacagg ttacgcgtgc ctggtggggg acaaagtaat gaaaccagca cacgtaaagg ggaccatcgataacgcggac ctggccaaac tggcctttaa gcggtcatct aagtatgaccttgaatgcgc gcagataccc gtgcacatga agtccgacgc ttcgaagttcacccatgaga aaccggaggg gtactacaac tggcaccacg gagcagtacagtactcagga ggccggttca ccatccctac aggtgctggc aaaccaggggatagcggcag accgatcttc gacaacaagg gacgcgtggt ggccatagtcttaggaggag ctaatgaagg agcccgtaca gccctctcgg tggtgacctggaataaagac attgtcacta aaatcacccc cgaggggggcc gaagagtggagtcttgccat cccagttatg tgcctgttgg caaacaccac gttcccctgc
tcccagcccc cttgcacgcc ctgctgctac gaaaaggaac cggaggaaaccctacgcatg cttgaggaca acgtcatgag acctgggtac tatcagctgctacaagcatc cttaacatgt tctccccacc gccagcgacg cagcaccaaggacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcccgactgtgga gaagggcact cgtgccatag tcccgtagca ctagaacgcatcagaaatga agcgacagac gggacgctga aaatccaggt ctccttgcaaatcggaataa agacggatga cagccacgat tggaccaagc tgcgttatatggacaaccac atgccagcag acgcagagag ggcggggcta tttgtaagaacatcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcc
cgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaagattagtcac tcatgtacgc acccatttca ccacgaccct cctgtgataggtcgggaaaa attccattcc cgaccgcagc acggtaaaga gctaccttgcagcacgtacg tgcagagcac cgccgcaact accgaggaga tagaggtacacatgccccca gacacccctg atcgcacatt aatgtcacaa cagtccggcaacgtaaagat cacagtcaat

ggccagacgg tgcggtacaa gtgtaattgcggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactgcaaggttgat caatgtcatg ccgcggtcac caatcacaaa aagtggcagtataactcccc tctggtcccg cgtaatgctg aacttgggga ccgaaaagga aaaattcaca tcccgtttcc gctggcaaat gtaacatgca gggtgcctaaagcaaggaac cccaccgtga cgtacgggaa aaaccaagtc atcatgctactgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaaccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaaccgtgccgact gaagggctcg aggtcacgtg gggcaacaac gagccgtataagtattggcc gcagttatct acaaacggta cagcccatgg ccacccgcatgagataattc tgtattatta tgagctgtac cctactatga ctgtagtagttgtgtcagtg gccacgttca tactcctgtc gatggtgggt atggcagcggggatgtgcat gtgtgcacga cgcagatgca tcacaccgta tgaactgaca ccaggagcta ccgtcccttt cctgcttagc ctaatatgct gcatcagaacagctaaagcg gccacatacc aagaggctgc gatatacctg tggaacgagcagcaaccttt gttttggcta caagccctta ttccgctggc agccctgattgttctatgca actgtctgag actcttacca tgctgctgta aaacgttggctttttagcc gtaatgagcg tcggtgccca cactgtgagc gcgtacgaacacgtaacagt gatcccgaac acggtgggag taccgtataa gactctagtcaatagacctg gctacagccc catggtattg gagatggaac tactgtcagtcactttggag ccaacactat cgcttgatta catcacgtgc gagtacaaaaccgtcatccc gtctccgtac gtgaagtgct gcggtacagc agagtgcaag gacaaaaacc tacctgacta cagctgtaag gtcttcaccg gcgtctacccatttatgtgg ggcggcgcct actgcttctg cgacgctgaa aacacgcagttgagcgaagc acatgtggag aagtccgaat catgcaaaac agaatttgcatcagcataca gggctcatac cgcatctgca tcagctaagc tccgcgtcctttaccaagga aataacatca ctgtaactgc ctatgcaaac ggcgaccatgccgtcacagt taaggacgcc aaattcattg tggggccaat gtcttcagcctggacacctt tcgacaacaa aattgtggtg tacaaaggtg acgtctataacatggactac ccgccctttg gcgcaggaag accaggacaa tttggcgatatccaaagtcg cacacctgag agtaaagacg tctatgctaa tacacaactg gtactgcaga gaccggctgc gggtacggta cacgtgccat actctcaggcaccatctggc tttaagtatt ggctaaaaga acgcggggcg tcactgcagcacacagcacc atttggctgc caaatagcaa caaacccggt aagagcggtgaactgcgccg tagggaacat gcccatctcc atcgacatac cggaagcggccttcactagg gtcgtcgacg cgccctcttt aacggacatg tcgtgcgaggtaccagcctg cacccattcc tcagactttg ggggcgtcgc cattattaaatatgcagcca gcaagaaagg caagtgtgcg gtgcattcga tgactaacgccgtcactatt cgggaagctg agatagaagt tgaagggaat tctcagctgcaaatctcttt ctcgacggcc ttagccagcg ccgaattccg cgtacaagtc tgttctacac aagtacactg tgcagccgag tgccaccccc cgaaggaccacatagtcaac tacccggcgt cacataccac cctcggggtc caggacatctccgctacggc gatgtcatgg gtgcagaaga tcacgggagg tgtgggactggttgttgctg ttgccgcact gattctaatc gtggtgctat gcgtgtcgttcagcaggcac taacttgaca attaagtatg aaggtatatg tgtcccctaagagacacact gtacatagca aataatctat agatcaaagg gctacgcaaccccctgaatag taacaaaata caaaatcact aaaaattata aaaacagaaaaatacataaa taggtatacg tgtcccctaa gagacacatt gtatgtaggtgataagtata gatcaaaggg ccgaataacc cctgaatagt aacaaaatat gaaaatcaat aaaaatcata aaatagaaaa accataaaca gaagtagttcaaagggctat aaaaccccctg aatagtaaca aacataaaa ttaataaaaatcaaatgaat accataattg gcaaacggaa gagatgtagg tacttaagcttcctaaaagc agccgaactc actttgagaa gtaggcatag cataccgaactcttccacga ttctccgaac ccacagggac gtaggagatg ttattttgtttttaatattt caaaaaaaaa aaaaaaaaaa aaaaaag

Appendix 2; KPA15- whole genome sequence

atggctgcgt gagacacacg tagcctacca gtttcttact gctctactctgcaaagcaag agattaataa cccatcatgg atcctgtgta cgtggacatagacgctgaca gcgccttttt gaaggccctg caacgtgcgt accccatgtttgaggtggaa ccaaggcagg tcacaccgaa tgaccatgct aatgctagagcgttctcgca tctagctata aaactaatag agcaggaaat tgaccccgactcaaccatcc tggatatcgg cagtgcgcca gcaaggagga tgatgtcggacaggaagtac cactgcgtct gcccgatgcg cagtgcggaa gatcccgagagactcgccaa ttatgcgaga aagctagcat ctgccgcagg aaaagtcctggacagaaaca tctctggaaa gatcggggac ttacaagcag taatggccgt
gccagacacg gagacgccaa cattctgctt acacacagac gtctcatgtagacagagagc agacgtcgct atataccaag acgtctatgc tgtacacgcacccacgtcgc tataccacca ggcgattaaa ggggtccgag tggcgtactgggttgggttc gacacaaccc cgttcatgta caatgccatg gcgggtgcctacccctcata ctcgacaaac tgggcagatg agcaggtact gaaggctaagaacataggat tatgttcaac agacctgacg gaaggtagac gaggcaagttgtctattatg agagggaaaa agctaaaacc gtgcgaccgt gtgctgttctcagtagggtc aacgctctac ccggaaagcc gcaagctact taagagctggcacctgccat cggtgttcca tttaaagggc aaactcagct tcacatgccg
ctgtgataca gtggtttcgt gtgagggcta cgtcgttaag agaataacgatgagcccagg cctttatgga aaaaccacag ggtatgcggt aacccaccacgcagacggat tcctgatgtg caagactacc gacacggttg acggcgaaagagtgtcattc tcggtgtgca catacgtgcc ggcgaccatt tgtgatcaaatgaccggcat ccttgctaca gaagtcacgc cggaggatgc acagaagctgttggtggggc tgaaccagag aatagtggtt aacggcagaa cgcaacggaatacgaacacc atgaaaaatt atctgcttcc cgtggtcgcc caagccttcagtaagtgggc aaaggagtgc cggaaagaca tggaagatga aaaactcctgggggtcagag aaagaacact gacctgctgc tgtctatgggcattcaagaa
gcagaaaaca cacacggtct acaagaggcc tgatacccag tcaattcagaaggttcaggc cgagtttgac agctttgtgg taccgagtct gtggtcgaccgggttgtcaa tccctttgag gactagaatc aaatggttgt taagcaaggtgccaaaaacc gacctgatcc catacagcgg agacgcccga gaagcccgggacgcagaaaa agaagcagag gaagaacgag aagcagaact gactcgcgaagccctaccac ctctacaggc agcacaggaa gatgttcagg tcgaaatcgacgtggaacag cttgaggaca gagcgggcgc aggaataata gagactccgagaggagctat caaagttact gcccaaccaa cagaccacgt cgtgggagagtacctggtac tctccccgca gaccgtacta cgtagccaga agctcagtctgattcacgct ttggcggagc aagtgaagac gtgcacgcac aacggacgagcagggaggta tgcggtcgaa gcgtacgacg gccgagtcct agtgccctcaggctatgcaa tctcgcctga agacttccag agtctaagcg aaagcgcaacgatggtgtat aacgaaagag agttcgtaaa cagaaagcta caccatattgcgatgcacgg accagccctg aacaccgacg aagagtcgta tgagctggtgagggcagaga ggacagaaca cgagtacgtc tacgacgtgg atcagagaagatgctgtaag aaggaagaag ccgcaggact ggtactggcg ggcgacttgactaatccgcc ctaccacgaa ttcgcatatg aagggctaaa aatccgccct
gcctgcccat acaaaattgc agtcatagga gtcttcggag taccgggatctggcaagtca gctattatca agaacctagt taccaggcag gacctggtgactagcggaaa gaaagaaaac tgccaagaaa tcaccaccga cgtgatgagacagagaggtc tagagatatctgcacgtacg gttgactcgc tgctcttgaatggatgcaac agaccagtcg acgtgttgta cgtagacgag gcgtttgcgtgccactctgg aacgctactt gctttgatcg ccttggtgag accaaggcagaaagttgtac tttgtggtga cccgaagcag tgcggcttct tcaatatgatgcagatgaaa gtcaactata atcacaacat ctgcacccaa gtgtaccacaaaagtatctc caggcggtgt acactgcctg tgaccgccat tgtgtcatcg
ttgcattacg aaggcaaaat gcgcactacg aatgagtaca acaagccgattgtagtggac actacaggct caacaaaacc tgaccctgga gacctcgtgttaacgtgctt cagagggtgg gttaaacaac tgcaaattga ctatcgtggatacgaggtca tgacagcagc cgcatcccaa gggttaacca gaaaaggagtttacgcagtt agacaaaaag ttaatgaaaa cccgctctat gcatcaacgtcagagcacgt caacgtactc ctaacgcgta cggaaggtaa actggtatggaagacacttt ccggcgaccc gtggataaag acgctgcaga acccaccgaaaggaaacttc aaagcaacta ttaaggagtg ggaggtggag catgcatcaataatggcggg catctgcagt caccaaatga ccttcgatac attccaaaat
aaagccaacg tttgttgggc taagagcttg gtccctatcc tcgaaacagcggggataaaa ctaaatgata ggcagtggtc tcagataatt caagccttcaaagaagacaa agcatactca cctgaagtag ccctgaatga aatatgtacgcgcatgtatg gggtggatct agacagcggg ctattttcta aaccgttggtgtctgtgtat tacgcggata accactggga taataggcct ggagggaaaatgttcggatt taaccccgag gcagcatcca ttctagaaag aaagtatccattcacaaaag ggaagtggaa catcaacaag cagatctgcg tgactaccaggaggatagaa gactttaacc ctaccaccaa catcataccg gccaacaggagactaccaca ctcattagtg gccgaacacc gcccagtaaa aggggaaaga
atggaatggc tggttaacaa gataaacggc caccacgtgc tcctggtcagtggctataac cttgcactgc ctactaagag agtcacttgg gtagcgccgttaggtgtccg cggagcggac tacacataca acctagagtt gggtctgccagcaacgcttg gtaggtatga cctagtggtc ataaacatcc acacacctttcgcatacac cattaccaac agtgcgtcga ccacgcaatg aaactgcaaatgctcggggg tgactcattg agactgctca aaccgggcgg ctctctattgatcagagcat atggttacgc agatagaacc agtgaacgag tcatctgcgtattgggacgc aagtttagat cgtctagagc gttgaaacca ccatgtgtcaccagcaacac tgagatgttt ttcctattca gcaactttga caatggcaga
aggaatttca caactcatgt catgaacaat caactgaatg cagccttcgtaggacaggtc acccgagcag gatgtgcacc gtcgtaccgg gtaaaacgcatggacatcgc gaagaacgat gaagagtgcg tagtcaacgc cgctaaccctcgcgggttac cgggtgacgg tgtttgcaag gcagtataca aaaaatggccggagtccttt aagaacagtg caacaccagt gggaaccgca aaaacagttatgtgcggtac gtatccagta atccacgctg ttggaccaaa cttctctaattattcggagt ctgaagggga ccgggaattg gcagctgcct atcgagaagtcgcaaaggaa gtaactaggc tgggagtaaa tagtgtagct atacctctcctctccacagg tgtatactca ggagggaaag acaggctgac ccagtcactg
aaccacctct ttacagccat ggactcgacg gatgcagacg tggtcatctactgccgcgac aaagaatggg agaagaaaat atctgaggcc atacagatgcggacccaagt agagctgctg gatgagcaca tctccataga ctgcgatattgttcgcgtgc accctgacag cagcttggca ggcagaaaag gatacagcaccacggaaggc gcactgtact catatctaga agggacccgt tttcatcagacggctgtgga tatggcggag

atacatacta tgtggccaaa gcaaacagaggccaatgagc aagtctgcct atatgccctg ggggaaagta ttgaatcgatcaggcagaaa
tgcccggtgg atgatgcaga cgcatcatct ccccccaaaactgtcccgtg cctttgccgt tacgctatga ctccagaacg cgtcacccgg
cttcgcatga accacgtcac aagcataatt gtgtgttctt cgtttcccctcccaaagtac aaaatagaag gagtgcaaaa agtcaaatgc
tctaaggtaatgctatttga ccacaacgtg ccatcgcgcg taagtccaag ggaatatagatcttcccagg agtctgcaca ggaggcgagt
acaatcacgt cactgacgcatagtcaattc gacctaagcg ttgatggcga gatactgccc gtcccgtcagacctggatgc tgacgcccca
gccctagaac cagcactaga cgacggggcgacacacacgc tgccatccac aaccggaaac cttgcggccg tgtctgattgggtaatgagc
accgtacctg tcgcgccgcc cagaagaagg cgagggagaaacctgactgt gacatgtgac gagagagaag ggaatataac
acccatggctagcgtccgat tctttagggc agagctgtgt ccggtcgtac aagaaacagcggagacgcgt gacacagcaa tgtctcttca
ggcaccaccg agtaccgccacggaaccgaa tcatccgccg atctccttcg gagcatcaag cgagacgttccccattacat ttggggactt
caacgaagga gaaatcgaaa gcttgtcttctgagctacta actttcggag acttcttacc aggagaagtg gatgacttgacagacagcga
ctggtccacg tgctcagaca cggacgacga gttatgactagacagggcag gtgggtatat attctcgtcg gacaccggtc caggtcattt
acaacagaag tcagtacgcc agtcagtgct gccggtgaac accctggaggaagtccacga ggagaagtgt tacccaccta agctggatga
agcaaaggagcaactattac ttaagaaact ccaggagagt gcatccatgg ccaacagaagcaggtatcag tcgcgcaaag tagaaaacat
gaaagcagca atcatccagagactaaagag aggctgtaga ctatacttaa tgtcagagac cccaaaagtccctacttacc ggactacata
tccggcgcct gtgtactcgc ctccgatcaacgtccgattg tccaatcccg agtccgcagt ggcagcatgc aatgagttcttagctagaaa
ctatccaact gtctcatcat accaaattac cgacgagtatgatgcatatc tagacatggt ggacgggtcg gagagttgcc tggaccgagc
gacattcaat ccgtcaaaac tcaggagcta cccgaaacag cacgcttaccacgcgccctc catcagaagc gctgtaccgt ccccattcca
gaacacactacagaatgtac tggcagcagc cacgaaaaga aactgcaacg tcacacagatgagggaatta cccactttgg actcagcagt
attcaacgtg gagtgtttcaaaaaattcgc atgcaaccaa gaatactggg aagaatttgc tgccagccctattaggataa caactgagaa
tttagcaacc tatgttacta aactaaaagggccaaaagca gcagcgctat tcgcaaaaac ccataatcta ctgccactacaggaagtacc
aatggatagg ttcacagtag atatgaaaag ggacgtgaaggtgactcctg gtacaaagca tacagaggaa agaccaaagg gcaggttat
acaggcggct gaacccttgg cgacagcata cctatgtggg attcacagagagctggttag gaggctgaac gccgtcctcc tacccaatgt
acatacactatttgacatgt ctgccgagga tttcgatgcc atcatagccg cacactttaagccaggagac actgttttgg aaacggacat
agcctccttt gataagagccaagatgattc acttgcgctt actgctttga tgctgttaga ggatttaggggtggatcact ccctgctgga
cttgatagag gctgctttcg gagagatttccagctgtcac ctaccgacag gtacgcgctt caagttcggc gccatgatgaaatcaggtat
gttcctaact ctgttcgtca acacattgtt aaacatcaccatcgccagcc gagtgctgga agatcgtctg acaaaatccg cgtgcgcggc
cttcatcggc gactacaaca taatacatgg agtcgtctcc gatgaattgatggcagccag atgtgccact tggatgaaca tggaagtgaa
gatcatagatgcagttgtat ccttgaaagc cccttacttt tgtggagggt ttatactgcacgatactgtg acaggaacag cttgcagagt
ggcagacccg ctaaaaaggctttttaaact gggcaaaccg ctagcggcag gtgacgaaca agatgaagatagaagacgag gctggctga
cgaagtgatc agatggcaac gaacagggctaattgatgag ctggagaaag cggtatactc taggtacgaa gtgcagggtatatcagttgt
ggtaatgtcc atggccacct ttgcaagctc cagatccaacttcgagaagc tcagaggacc cgtcataact ttgtacggcg gtcctaaata
ggtacgcact acagctacct attttgcaga agccgacagc aagtatctaaacactaatca gctacaatgg agttcatccc aacccaaact
ttttacaataggaggtacca gcctcgaccc tggactccgc gccctactat ccaagtcatcaggcccagac cgcgccctca gaggcaagct
gggcaacttg cccagctgatctcagcagtt aataaactga caatgcgcgc ggtaccccaa cagaagccacgcaggaatcg gaagaataag
aagcaaaagc aaaaacaaca ggcgccacaaaacaacacaa atcaaaagaa gcagccacct aaaaagaaac cggctcaaaa
gaaaaagaag ccgggccgca gagagaggat gtgcatgaaa atcgaaaatgattgtatttt cgaagtcaag cacgaaggta ggtaacagg
ttacgcgtgcctggtggggg acaaagtaat gaaaccagca cacgtaaagg ggaccatcgataacgcggac ctggccaaac tggcctttaa
gcggtcatct aagtatgaccttgaatgcgc gcagataccc gtgcacatga agtccgacgc ttcgaagttcacccatgaga aaccggaggg
gtactacaac tggcaccacg gagcagtacagtactcagga ggccggttca ccatccctac aggtgctggc aaaccaggggacagcggcag
accgatcttc gacaacaagg gacgcgtggt ggccatagtcttaggaggag ctaatgaagg agcccgtaca gccctctcgg tggtgacctg
gaataaagac attgtcacta aaatcacccc cgagggggcc gaagagtggagtcttgccat cccagttatg tgcctgttgg caaacaccac
gttcccctgctcccagcccc cttgcacgcc ctgctgctac gaaaaggaac cggaggaaaccctacgcatg cttgaggaca acgtcatgag
acctgggtac tatcagctgctacaagcatc cttaacatgt tctccccacc gccagcgacg cagcaccaaggacaacttca atgtctataa
agccacaaga ccatacttag ctcactgtcccgactgtgga gaagggcact cgtgccatag tcccgtagca ctagaacgcatcagaaatga
agcgacagac gggacgctga aaatccaggt ctccttgcaaatcggaataa agacggatga cagccacgat tggaccaagc tgcgttatat
ggacaaccac atgccagcag acgcagagag ggcggggcta tttgtaagaacatcagcacc gtgtacgatt actggaacaa tgggacactt
catcctggcccgatgtccaa aaggggaaac tctgacggtg ggattcactg acagtaggaagattagtcac tcatgtacgc acccatttca
ccacgaccct cctgtgataggtcgggaaaa attccattcc cgaccgcagc acggtaaaga gctaccttgcagcacgtacg tgcagagcac
cgccgcaact accgaggaga tagaggtacacatgccccca gacacccctg atcgcacatt aatgtcacaa cagtccggcaacgtaaagat
cacagtcaat ggccagacgg tgcggtacaa gtgtaattgcggtggctcaa atgaaggact aacaactaca gacaaagtga ttaataactg
caaggttgat caatgtcatg ccgcggtcac caatcacaaa aagtggcagtataactcccc tctggtcccg cgtaatgctg aacttgggga
ccgaaaaggaaaaattcaca tcccgtttcc gctggcaaat gtaacatgca gggtgcctaaagcaaggaac cccaccgtga cgtacgggaa
aaaccaagtc atcatgctactgtatcctga ccacccaaca ctcctgtcct accggaatat gggagaagaaccaaactatc aagaagagtg
ggtgatgcat aagaaggaag tcgtgctaaccgtgccgact gaagggctcg aggtcacgtg gggcaacaac gagccgtataagtattggcc
gcagttatct acaaacggta cagcccatgg ccacccgcatgagataattc tgtattatta tgagctgtac cccactatga ctgtagtagt
tgtgtcagtg gccacgttca tactcctgtc gatggtgggt atggcagcggggatgtgcat gtgtgcacga cgcagatgca tcacaccgta
tgaactgacaccaggagcta ccgtcccttt cctgcttagc ctaatatgct gcatcagaacagctaaagcg gccacatacc aagaggctgc

gatatacctg tggaacgagcagcaaccttt gttttggcta caagccctta ttccgctggc agccctgattgttctatgca actgtctgag actcttacca tgctgctgta aaacgttggctttttttagcc gtaatgagcg tcggtgccca cactgtgagc gcgtacgaacacgtaacagt gatcccgaac acggtgggag taccgtataa gactctagtcaatagacctg gctacagccc catggtattg gagatggaac tactgtcagt cactttggag ccaacactat cgcttgatta catcacgtgc gagtacaaaaccgtcatccc gtctccgtac gtgaagtgct gcggtacagc agagtgcaaggacaaaaacc tacctgacta cagctgtaag gtcttcaccg gcgtctacccatttatgtgg ggcggcgcct actgcttctg cgacgctgaa aacacgcagttgagcgaagc acacgtggag aagtccgaat catgcaaaac agaatttgcatcagcataca gggctcatac cgcatctgca tcagctaagc tccgcgtcctttaccaagga aataacatca ctgtaactgc ctatgcaaac ggcgaccatgccgtcacagt taaggacgcc aaattcattg tggggccaat gtcttcagcctggacacctt tcgacaacaa aattgtggtg tacaaaggtg acgtctataa catggactac ccgcccttg gcgcaggaag accaggacaa tttggcgatatccaaagtcg cacacctgag agtaaagacg tctatgctaa tacacaactggtactgcaga gaccggctgc gggtacggta cacgtgccat actctcaggcaccatctggc tttaagtatt ggctaaaaga acgcggggcg tcgctgcagcacacagcacc atttggctgc caaatagcaa caaacccggt aagagcggtgaactgcgccg tagggaacat gcccatctcc atcgacatac cggaagcggccttcactagg gtcgtcgacg cgccctcttt aacggacatg tcgtgcgaggtaccagcctg cacccattcc tcagactttg ggggcgtcgc cattattaaatatgcagcca gcaagaaagg caagtgtgcg gtgcattcga tgactaacgc cgtcactatt cgggaagctg agatagaagt tgaagggaat tctcagctgcaaatctcttt ctcgacggcc ttagccagcg ccgaattccg cgtacaagtctgttctacac aagtacactg tgcagccgag tgccacccccc cgaaggaccacatagtcaac tacccggcgt cacataccac cctcggggtc caggacatctccgctacggc gatgtcatgg gtgcagaaga tcacgggagg tgtgggactggttgttgctg ttgccgcact gattctaatc gtggtgctat gcgtgtcgttcagcaggcac taacttgaca attaagtatg aaggtatatg tgtcccctaagagacacact gtacatagca aataatctat agatcaaagg gctacgcaacccctgaatag taacaaaata caaaatcact aaaaattata aaaacagaaa aatacataaa taggtatacg tgtcccctaa gagacacatt gtatgtaggtgataagtata gatcaaaggg ccgaataacc cctgaatagt aacaaaatatgaaaatcaat aaaaatcata aaatagaaaa accataaaca gaagtagttcaaagggctat aaaacccctg aatagtaaca aaacataaaa ttaataaaaatcaaatgaat accataattg gcaaacggaa gagatgtagg tacttaagcttcctaaaagc agccgaactc actttgagaa gtaggcatag cataccgaactcttccacga ttctccgaac ccacagggac gtaggagatg ttattttgtttttaatattt caaaaaaaaa aaaaaaaaaa aaaaaag

Appendix 3; Comoros25- whole genome sequence

atggctgcgt gagacacacg tagcctacca gtttcttact gctctactctgcaaagcaag agattaataa cccatcatgg atcctgtgta
cgtggacatagacgctgaca gcgccttttt gaaggccctg caacgtgcgt accccatgtttgaggtggaa ccaaggcagg tcacaccgaa
tgaccatgct aatgctagagcgttctcgca tctagctata aaactaatag agcaggaaat tgaccccgactcaaccatcc tggatatcgg
cagtgcgcca gcaaggagga ggatgtcggacaggaagtac cactgcgtct ccccgatgcg cagtgcggaa gatcccgaga
gactcgccaa ttatgcgaga aagctagcat ctgccgcagg aaaagtcctggacagaaaca tctctggaaa gatcggggac ttacaagcag
taatggccgtgccagacacg gagacgccaa cattctgctt acacacagac gtctcatgtagacagagagc agacgtcgct atataccaag
acgtctatgc tgtacacgcacccacgtcgc tataccacca ggcgattaaa ggggtccaag tggcgtactgggttgggttc gacacaaccc
cgttcatgta caatgccatg gcgggtgcctacccctcata ctcgacaaac tgggcagatg agcaggtact gaaggctaagaacataggat
tatgttcaac agacctgacg gaaggtagac gaggcaagttgtctattatg agagggaaaa agctaaaacc gtgcgaccgt gtgctgttct
cagtagggtc aacgctctac ccggaaagcc gcaagctact taagagctgcacctgccat cggtgttcca tttaaagggc aaactcagct
tcacatgccgctgtgataca gtggtttcgt gtgagggcta cgtcgttaag agaataacgatgagcccagg cctttatgga aaaaccacag
ggtatgcggt aacccaccacgcagacggat tcctgatgtg caagactacc gacacggttg acggcgaaagagtgtcattc tcggtgtgca
catacgtgcc ggcgaccatt tgtgatcaaatgaccggcat ccttgctaca gaagtcacgc cggaggatgc acagaagctgttggtggggc
tgaaccagag aatagtggtt aacggcagaa cgcaacggaatacgaacacc atgaaaaatt atctgcttcc cgtggtcgcc caagccttca
gtaagtgggc aaaggagtgc cggaaagaca tggaagatga aaaactcctgggggtcagag aaagaacact gacctgctgctgtctatggg
cattcaagaagcagaaaaca cacacggtct acaagaggcc tgatacccag tcaattcagaaggttcaggc cgagtttgac agctttgtgg
taccgagtct gtggtcgtccgggttgtcaa tccctttgag gactagaatc aaatggttgt taagcaaggtgccaaaaacc gacctgatcc
catacagcgg agacgcccga gaagcccgggacgcagaaaa agaagcagag gaagaacgag aagcagaact gactcgcgaa
gccctaccac ctctacaggc agcacaggaa gatgttcagg tcgaaatcgacgtggaacag cttgaggaca gagcgggcgc aggaataata
gagactccgagaggagctat caaagttact gcccaaccaa cagaccacgt cgtgggagagtacctggtac tctccccgca gaccgtacta
cgtagccaga agctcagtctgattcacgct ttggcggagc aagtgaagac gtgcacgcac aacggacgag
cagggaggta tgcggtcgaa gcgtacgacg gccgagtcct agtgccctcaggctatgcaa tctcgcctga agacttccag agtctaagcg
aaagcgcaacgatggtgtat aacgaaagag agttcgtaaa cagaaagcta caccatattgcgatgcacgg accagccctg aacaccgacg
aagagtcgta tgagctggtgagggcagaga ggacagaaca cgagtacgtc tacgacgtgg atcagagaagatgctgtaag aggaagaag
ccgcaggact ggtactggtg ggcgacttgactaatccgcc ctaccacgaa ttcgcatatg aagggctaaa aatccgccctgcctgcccat
acaaaattgc agtcatagga gtcttcggag taccgggatctggcaagtca gctattatca agaacctagt taccaggcag gacctggtga
ctagcggaaa gaaagaaaac tgccaagaaa tcaccaccga cgtgatgagacagagaggtc tagagatatc tgcacgtacg ttgactcgc
tgctcttgaatggatgcaac agaccagtcg acgtgttgta cgtagacgag gcgtttgcgtgccactctgg aacgctactt gctttgatcg
ccttggtgag accaaggcagaaagttgtac tttgtggtga cccgaagcag tgcggcttct tcaatatgatgcagatgaaa gtcaactata
atcacaacat ctgcacccaa gtgtaccacaaaagtatctc caggcggtgt acactgcctg tgaccgccat tgtgtcatcgttgcattacg
aaggcaaaat gcgcactacg aatgagtaca acaagccgattgtagtggac actacaggct caacaaaacc tgaccctgga gacctcgtgt
taacgtgctt cagagggtgg gttaaacaac tgcaaattga ctatcgtggatacgaggtca tgacagcagc cgcatcccaa gggttaacca
gaaaaggagtttacgcagtt agacaaaaag ttaatgaaaa cccgctctat gcatcaacgtcagagcacgt caacgtactc ctaacgcgta
cggaaggtaa actggtatggaagacacttt ccggcgaccc gtggataaag acgctgcaga acccaccgaaaggaaacttc aaagcaacta
ttaaggagtg ggaggtggag catgcatcaataatggcggg catctgcagt caccaaatga ccttcgatac attccaaaataaagccaacg
tttgttgggc taagagcttg gtccctatcc tcgaaacagcggggataaaa ctaaatgata ggcagtggtc tcagataatt caagccttca
aagaagacaa agcatactca cctgaagtag ccctgaatga aatatgtacgcgcatgtatg gggtggatct agacagcggg ctattttcta
aaccgttggtgtctgtgtat tacgcggata accactggga taataggcct ggagggaaaatgttcggatt taaccccgag gcagcatcca
ttctagaaag aaagtatccattcacaaaag ggaagtggaa catcaacaag cagatctgcg tgactaccaggaggatagaa gactttaacc
ctaccaccaa catcataccg gccaacaggagactaccaca ctcattagtg gccgaacacc gcccagtaaa aggggaaagaatggaatggc
tggttaacaa gataaacggc caccacgtgc tcctggtcagtggctataac cttgcactgc ctactaagag agtcacttgg gtagcgccgt
taggtgtccg cggagcggac tacacataca acctagagtt gggtctgccagcaacgcttg gtaggtatga cctagtggtc ataaacatcc
acacacctttcgcatacac cattaccaac agtgcgtcga ccacgcaatg aaactgcaaatgctcggggg tgactcattg agactgctca
aaccgggcgg ctctctattgatcagagcat atggttacgc agatagaacc agtgaacgag tcatctgcgtattgggacgc aagtttagat
cgtctagagc gttgaaacca ccatgtgtcaccagcaacac tgagatgttt ttcctattca gcaactttga caatggcagaaggaatttca
caactcatgt catgaacaat caactgaatg cagccttcgtaggacaggtc acccgagcag gatgtgcacc gtcgtaccgg gtaaaacgca
tggacatcgc gaagaacgat gaagagtgcg tagtcaacgc cgctaaccctcgcgggttac cgggtgacgg tgtttgcaag gcagtataca
aaaaatggccggagtccttt aagaacagtg caacaccagt gggaaccgca aaaacagttatgtgcggtac gtatccagta atccacgctg
ttggaccaaa cttctctaattattcggagt ctgaagggga ccgggaattg gcagctgcct atcgagaagtcgcaaaggaa gtaactaggc
tgggagtaaa tagtgtagct atacctctcctctccacagg tgtatactca ggagggaaag acaggctgac ccagtcactgaaccacctct
ttacagccat ggactcgacg gatgcagacg tggtcatctactgccgcgac aaagaatggg agaagaaaat atctgaggcc atacagatgc
ggacccaagt agagctgctg gatgagcaca tctccataga ctgcgatattgttcgcgtgc accctgacag cagcttggca ggcagaaaag
gatacagcaccacggaaggc gcactgtact catatctaga agggacccgt tttcatcagacggctgtgga tatggcggag atacatacta
tgtggccaaa gcaaacagaggccaatgagc aagtctgcct atatgccctg ggggaaagta ttgaatcgatcaggcagaaa tgcccggtgg

atgatgcaga cgcatcatct ccccccaaaactgtcccgtg cctttgccgt tacgctatga ctccagaacg cgtcacccggcttcgcatga
accacgtcac aagcataatt gtgtgttctt cgtttcccctcccaaagtac aaaatagaag gagtgcaaaa agtcaaatgc tctaaggtaa
tgctatttga ccacaacgtg ccatcgcgcg taagtccaag ggaatatagatcttcccagg agtctgcaca ggaggcgagt acaatcacgt
cactgacgcatagtcaattc gacctaagcg ttgatggcga gatactgccc gtcccgtcagacctggatgc tgacgcccca gccctagaac
cagcactaga cgacggggcgacacacacgc tgccatccac aaccggaaac cttgcggccg tgtctgattgggtaatgagc accgtacctg
tcgcgccgcc cagaagaagg cgagggagaaacctgactgt gacatgtgac gagagagaag ggaatataac acccatggct
agcgtccgat tctttagggc agagctgtgt ccggtcgtac aagaaacagcggagacgcgt gacacagcaa tgtctcttca ggcaccaccg
agtaccgccacggaaccgaa tcatccgccg atctccttcg gagcatcaag cgagacgttccccattacat ttggggactt caacgaagga
gaaatcgaaa gcttgtcttctgagctacta actttcggag acttcttacc aggagaagtg gatgacttgacagacagcga ctggtccacg
tgctcagaca cggacgacga gttacgactagacagggcag gtgggtatat attctcgtcg gacaccggtc caggtcatttacaacagaag
tcagtacgcc agtcagtgct gccggtgaac accctggaggaagtccacga ggagaagtgt tacccaccta agctggatga agcaaaggag
caactattac ttaagaaact ccaggagagt gcatccatgg ccaacagaagcaggtatcag tcgcgcaaag tagaaaacat gaaagcagca
atcatccagagactaaagag aggctgtaga ctatacttaa tgtcagagac cccaaaagtccctacttacc ggactacata tccggcgcct
gtgtactcgc ctccgatcaacgtccgattg tccaatcccg agtccgcagt ggcagcatgc aatgagttcttagctagaaa ctatccaact
gtctcatcat accaaattac cgacgagtatgatgcatatc tagacatggt ggacgggtcg gagagttgcc tggaccgagc
gacattcaat ccgtcaaaac tcaggagcta cccgaaacag cacgcttaccacgcgccctc catcagaagc gctgtaccgt ccccattcca
gaacacactacagaatgtac tggcagcagc cacgaaaaga aactgcaacg tcacacagatgagggaatta cccactttgg actcagcagt
attcaacgtg gagtgtttcaaaaaattcgc atgcaaccaa gaatactggg aagaatttgc tgccagccctattaggataa caactgagaa
tttagcaacc tatgttacta aactaaaagggccaaaagca gcagcgctat tcgcaaaaac ccataatcta ctgccactacaggaagtacc
aatggatagg ttcacagtag atatgaaaag ggacgtgaaggtgactcctg gtacaaagca tacagaggaa agacctaagg tgcaggttat
acaggcggct gaacccttgg cgacagcata cctatgtggg attcacagagagctggttag gaggctgaac gccgtcctcc tacccaatgt
acatacactatttgacatgt ctgccgagga tttcgatgcc atcatagccg cacactttaagccaggagac actgttttgg aaacggacat
agcctccttt gataagagccaagatgattc acttgcgctt actgctttga tgctgttaga ggatttaggggtggatcact ccctgctgga
cttgatagag gctgctttcg gagagatttccagctgtcac ctaccgacag gtacgcgctt caagttcggc gccatgatgaaatcaggtat
gttcctaact ctgttcgtca acacattgtt aaacatcaccatcgccagcc gagtgctgga agatcgtctg acaaaatccg cgtgcgcggc
cttcatcggc gacgacaaca taatacatgg agtcgtctcc gatgaattgatggcagccag atgtgccact tggatgaaca tggaagtgaa
gatcatagatgcagttgtat ccttgaaagc cccttacttt tgtggagggt ttatactgcacgatactgtg acaggaacag cttgcagagt
ggcagacccg ctaaaaaggctttttaaact gggcaaaccg ctagcggcag gtgacgaaca agatgaagatagaagacgag cgctggctga
cgaagtgatc agatggcaac gaacagggctaattgatgag ctggagaaag cggtatactc taggtacgaa gtgcagggta
tatcagttgt ggtaatgtcc atggccacct ttgcaagctc cagatccaacttcgagaagc tcagaggacc cgtcataact ttgtacggcg
gtcctaaataggtacgcact acagctacct attttgcaga agccgacagc aagtatctaaacactaatca gctacaatgg agttcatccc
aacccaaact ttttacaataggaggtacca gcctcgaccc tggactccgc gccctactat ccaagtcatcaggcccagac cgcgccctca
gaggcaagct gggcaacttg cccagctgatctcagcagtt aataaactga caatgcgcgc ggtaccccaa cagaagccacgcaggaatcg
gaagaataag aagcaaaagc aaaaacaaca ggcgccacaaaacaacacaa atcaaaagaa gcagccacct aaaaagaaac
cggctcaaaagaaaaagaag ccgggccgca gagagaggat gtgcatgaaa atcgaaaatgattgtatttt cgaagtcaag cacgaaggta
aggtaacagg ttacgcgtgcctggtggggg acaaagtaat gaaaccagca cacgtaaagg ggaccatcga
taacgcggac ctggccaaac tggcctttaa gcggtcatct aagtatgaccttgaatgcgc gcagataccc gtgcacatga agtccgacgc
ttcgaagttcacccatgaga aaccggaggg gtactacaac tggcaccacg gagcagtacagtactcagga ggccggttca ccatccctac
aggtgctggc aaaccaggggacagcggcag accgatcttc gacaacaagg gacgcgtggt ggccatagtcttaggaggag
ctaatgaagg agcccgtaca gccctctcgg tggtgacctggaataaagac attgtcacta aaatcacccc cgagggggcc gaagagtgga
gtcttgccat cccagttatg tgcctgttgg caaacaccac gttcccctgctcccagcccc cttgcacgcc ctgctgctac gaaaaggaac
cggaggaaaccctacgcatg cttgaggaca acgtcatgag acctgggtac tatcagctgctacaagcatc cttaacatgt tctccccacc
gccagcgacg cagcaccaaggacaacttca atgtctataa agccacaaga ccatacttag ctcactgtcccgactgtgga gaagggcact
cgtgccatag tcccgtagca ctagaacgcatcagaaatga agcgacagac gggacgctga aaatccaggt ctccttgcaaatcggaataa
agacggatga cagccacgat tggaccaagc tgcgttatatggacaaccac atgccagcag acgcagagag ggcggggcta ttgtaagaa
catcagcacc gtgtacgatt actggaacaa tgggacactt catcctggcccgatgtccaa aaggggaaac tctgacggtg ggattcactg
acagtaggaagattagtcac tcatgtacgc acccatttca ccacgaccct cctgtgataggtcgggaaaa attccattcc cgaccgcagc
acggtaaaga gctaccttgcagcacgtacg tgcagagcac cgccgcaact accgaggaga tagaggtacacatgccccca gacacccctg
atcgcacatt aatgtcacaa cagtccggcaacgtaaagat cacagtcaat ggccagacgg tgcggtacaa gtgtaattgcggtggctcaa
atgaaggact aataactaca gacaaagtga ttaataactgcaaggttgat caatgtcatg ccgcggtcac caatcacaaa aagtggcagt
ataactcccc tctggtcccg cgtaatgctg aacttgggga ccgaaaaggaaaaattcaca tcccgtttcc gctggcaaat gtaacatgca
gggtgcctaaagcaaggaac cccaccgtga cgtacgggaa aaaccaagtc atcatgctactgtatcctga ccacccaaca ctcctgtcct
accggaatat gggagaagaaccaaactatc aagaagagtg ggtgatgcat aagaaggaag tcgtgctaaccgtgccgact aagggctcg
aggtcacgtg gggcaacaac gagccgtataagtattggcc gcagttatct acaaacggta cagcccatgg ccacccgcatgagataattc
tgtattatta tgagctgtac cccactatga ctgtagtagttgtgtcagtg gccacgttca tactcctgtc gatggtgggt atggcagcgg
ggatgtgcat gtgtgcacga cgcagatgca tcacaccgta tgaactgacaccaggagcta ccgtcccttt cctgcttagc ctaatatgct
gcatcagaacagctaaagcg gccacatacc aagaggctgc gatatacctg tggaacgagcagcaaccttt gttttggcta caagccctta

ttccgctggc agccctgattgttctatgca actgtctgag actcttacca tgctgctgta aaacgttggctttttagct gtaatgagcg tcggtgccca cactgtgagc gcgtacgaacacgtaacagt gatcccgaac acggtgggag taccgtataa gactctagtcaatagacctg gctacagccc catggtattg gagatggaac tactgtcagtcactttggag ccaacactat cgcttgatta catcacgtgc gagtacaaaa ccgtcatccc gtctccgtac gtgaagtgct gcggtacagc agagtgcaaggacaaaaacc tacctgacta cagctgtaag gtcttcaccg gcgtctacccatttatgtgg ggcggcgcct actgcttctg cgacgctgaa aacacgcagttgagcgaagc acacgtggag aagtccgaat catgcaaaac agaatttgcatcagcataca gggctcatac cgcatctgca tcagctaagc tccgcgtcctttaccaagga aataacatca ctgtaactgc ctatgcaaac ggcgaccatgccgtcacagt taaggacgcc aaattcattg tggggccaat gtcttcagcctggacacctt tcgacaacaa aattgtggtg tacaaaggtg acgtctataacatggactac ccgccctttg gcgcaggaag accaggacaa tttggcgata tccaaagtcg cacacctgag agtaaagacg tctatgctaa tacacaactggtactgcaga gaccggctgc gggtacggta cacgtgccat actctcaggcaccatctggc tttaagtatt ggctaaaaga acgcggggcg tcgctgcagcacacagcacc atttggctgc caaatagcaa caaacccggt aagagcggtgaactgcgccg tagggaacat gcccatctcc atcgacatac cggaagcgcttcactagg gtcgtcgacg cgccctcttt aacggacatg tcgtgcgaggtaccagcctg cacccattcc tcagactttg ggggcgtcgc cattattaaatatgcagcca gcaagaaagg caagtgtgcg gtgcattcga tgactaacgccgtcactatt cgggaagctg agatagaagt tgaagggaat tctcagctgc aaatctcttt ctcgacggcc ttagccagcg ccgaattccg cgtacaagtctgttctacac aagtacactg tgcagccgag tgccaccccc cgaaggaccacatagtcaac tacccggcgt cacataccac cctcggggtc caggacatctccgctacggc gatgtcatgg gtgcagaaga tcacgggagg tgtgggactggttgttgctg ttgccgcact gattctaatc gtggtgctat gcgtgtcgttcagcaggcac taacttgaca attaagtatg aaggtatatg tgtcccctaagagacacact gtacatagca aataatctat agatcaaagg gctacgcaacccctgaatag taacaaaata caaaatcact aaaaattata aaaacagaaaaatacataaa taggtatacg tgtcccctaa gagacacatt gtatgtaggt agtaagtata gatcaaaggg ccgaataacc cctgaatagt aacaaaatatgaaaatcaat aaaaatcata aaatagaaaa ccataaaca gaagtagttcaaagggctat aaaacccctg aatagtaaca aaacataaaa ttaataaaaatcaaatgaat accataattg gcaaacggaa gagatgtagg tacttaagcttcctaaaagc agccgaactc actttgagaa gtaggcatag cataccgaactcttccacga ttctccgaac ccacagggac gtaggagatg ttattttgtttttaatattt caaaaaaaaa aaaaaaaaaa aaaaaac

Appendix 4: List of other primers used in the study

Sequence Primer	Sequence	Gene to amplify
CHIKk 251 R	GTCGGGATCAATTTCCTGCTC	NSP1
CHIK 325 F	CTATGCGCAGCGCAGAAGAC	NSP1
CHIK 662 R	GTACGAGGGGTATGCACCTGC	NSP1
CHIK 1500 R	CTTTGCTTAGCAGCCACTTGATTCTG	NSP1
CHIK 1032 F	GATACGGTAGATGGCGAGAGAGTG	NSP1
CHIK 2048 R	CAGCTCGTACGACTCCTCGTCAGTG	NSP2
CHIK 1816 F	CGGAACAAGTGAAGACATGCAC	NSP2
CHIK 2723 R	CGAGCCTGTAGTATCCACTACAA	NSP2
CHIK 2625 F	CTGCCTGTGACTGCCATTGTGTCC	NSP2
CHIK 3499 R	CTGTTGGCAGGTATAATGTTGG	NSP2
CHIK 3159 F	AAACTAAATGATAGGCAGTGGTCCC	NSP2
CHIK 4870R	CGGCAAAGACACGGGACAGT	NSP3
CHIK 4274 F	CCACGCCGTAGGACCAAACTTCT	NSP3
CHIK 4130 F	TGTGTAGTCAACGCCGCAACC	NSP3
CHIK 5783 R	CTTAGGTGGGTAACATTTCTCC	NSP4
CHIK 5236 F	CGGCCGTGTCTGACTGGGTAAT	NSP4
CHIK 5640 F	GAGTTATTACTAGACAGGGCAGGCGG	NSP4
CHIK 6313 F	CCACTTTGGACTCAGCAGTATTCAAC	NSP4
CHIK 7045 R	GCCCACGCGGATTTTGTCAGA	NSP4
CHIK 7044 F	TGCGCGGCCTTCATCGGCGACTAC	C
CHIK 8139 F	CACGGAGCAGTGCAGTATTCAG	E3
CHIK 8433 R	TCGTAGCAGCAGGGTGTGCAAG	E3
CHIK 8599 F	GTCCYGACTGYGGAGAAGGGC	E3
CHIK 9641 R	GTCATAGTAGGGTACAGCTC	6K
CHIK 9199 F	GCCGCGGTCACCAATCACA	E1
CHIK 10407 R	TTGGTAAAGGACGCGGAGCTTCGC	E2
CHIK 10064 F	TGGAGCTACTGTCAGTCACTTTGGA	E1
CHIK 10771 F	CAAATAGCAACAAACCCGGTAAGAG	E1
CHIK 11535 R	AGCCCTTTGATCTATACTTACTTAC	E1
CHIK 11213 F	TGTCATGGGTGCAGAAGATCACGGG	3□ ‘end

Appendix 5; Amino acid codes.

Amino acid	Three letter code	One letter code
alanine	ala	A
arginine	arg	R
asparagine	asn	N
aspartic acid	asp	D
asparagine or aspartic acid	asx	B
cysteine	cys	C
glutamic acid	glu	E
glutamine	gln	Q
glutamine or glutamic acid	glx	Z
glycine	gly	G
histidine	his	H
isoleucine	ile	I
leucine	leu	L
lysine	lys	K
methionine	met	M
phenylalanine	phe	F
proline	pro	P
serine	ser	S
threonine	thr	T
tryptophan	try	W
tyrosine	tyr	Y
valine	val	V