EPIDEMIOLOGICAL INVESTIGATION AND ANTIGENIC CHARACTERIZATION OF DENGUE VIRUS FOR THE DEVELOPMENT OF NANOBODIES

GASPARY MWANYIKA

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF SOKOINE UNIVERSITY OF AGRICULTURE. MOROGORO, TANZANIA.

2022

EXTENDED ABSTRACT

Dengue is an important mosquito-borne viral disease of global health concerns. Vector control and early diagnosis remain the main interventions. Lack of vaccines against Dengue viruses (DENV) complicates the control of the disease hence increasing risk of DENV transmission. In this thesis, a systematic review was undertaken to analyse the prevalence of DENV infection and associated risk factors in Africa. A literature search was done using PubMed/MEDLINE, Scopus and Embase databases to identify articles published between 1960 and 2020. Meta-analysis was performed using a random-effect model at a 95% confidence interval. A study on seroprevalence and risk factors of dengue was carried out in Buhigwe, Kalambo, Kilindi, Kinondoni, Kondoa, Kyela, Mvomero, and Ukerewe districts representing five ecological zones in Tanzania. A total of 1,818 blood samples were collected from humans recruited from households and healthcare facilities and tested using an enzyme-linked immunosorbent assay specific for DENV immunoglobulin G (IgG) antibodies. During the 2019 outbreak in Tanzania, serum samples were collected from the outpatients seeking care from health facilities in Kinondoni and Ilala districts and tested to confirm the presence of DENV. DENV genome sequences were generated using Nanopore MinION and analysed using NanoGalaxy tool. The phylogenies and spatial distributions were reconstructed using maximum likelihood method and Bayesian Evolutionary Analysis. Multi immunoinformatics tools were used to predict B-cell epitopes on envelope protein domain III (EDIII) of DENV and the final construct was cloned into pET-22b (+) vector and expressed in *Escherichia coli.*

The findings of the systematic review showed that between 1960 and 2020, 45 outbreaks were reported in Africa, of which 17 and 16 occurred in East and West Africa, respectively. Overall, the prevalence of DENV was 29% (95% CI: 20–39%) and 3% (95%

CI: 1–5%) during the outbreak and non-outbreak periods, respectively. Old age (6/21 studies), lack of mosquito control interventions (6/21), urban residence (4/21), climate change (3/21), and recent history of travel (3/21) were the leading risks factors. Serological evidence from this study showed that the overall prevalence of DENV IgG antibodies was 16.1% (n= 292) in Tanzania. The prevalence was highest in Kinondoni district (43.8%, n= 103). Increasing age (> 28 years, $p < 0.001$), stagnant water bodies around homes ($p < 0.01$), and piped water at home ($p < 0.01$) were significantly associated with DENV seropositivity. This study reports for the first time DENV serotype 1 (DENV-1) genotype V in Tanzania that was responsible for the 2019 outbreak in Dar es Salaam. Spatial analysis suggests that the genotype was possibly introduced into Tanzania from a single source in Asia. Furthermore, epitope prediction analysis revealed five continuous epitopes at 305─320, 334─348, 320─335, 378─392, 389─400 and three conformational epitopes at 310─312, 336─337 and 341─342 residues of envelope protein domain III (EDIII) that can be potential targets for development of diagnostic nanobodies targeting DENV. In conclusion, the results from a systematic review highlights and increased risk of DENV serotype circulation in Africa during the past decade (2010─2020). The national-wide seroprevalence study conducted in Tanzania reveals a wide circulation of DENV in diverse ecological zones with the highest prevalence in the north-eastern zone. This study reports for the first time DENV-1 serotype in Tanzania during the 2019 outbreak in Dar es Salaam that was likely imported from Asia. The possibility of future introduction of new serotypesof the virus from multiple geographic origins due to globalization is still inevitable. It is important that surveillance and diagnostic systems of DENV strengthened nationwide to improve early detection, clinical management and outcomes. In addition, novel biomarkers on envelop protein domains

should be explored for development of vaccines that could possibly be used for the control of emerging variants of DENV-1 and DENV-2 strains.

DECLARATION

I, GASPARY MWANYIKA, do hereby declare to the Senate of Sokoine University of Agriculture that, this thesis is my own original work and that it has neither been submitted nor concurrently submitted for a degree award in any other institution.

_____ _____

Gaspary Mwanyika Date

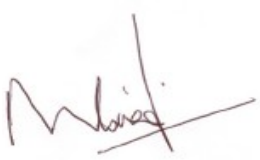
(PhD Candidate)

The above declaration is confirmed;

_____ _____

Professor Gerald Misinzo Date

(Supervisor)



5th May, 2022

_____ _____

Dr Leonard E. G. Mboera Date

(Supervisor)

COPYRIGHT

ACKNOWLEDGEMENTS

All praise and glory to the Almighty God for his blessings all the time since the start to the completion of my PhD research journey. My sincere gratitude to the Government of the United Republic of Tanzania and the World Bank for financial support [WB-ACE II Grant PAD1436, IDA credit 5799-TZ] through the SACIDS Africa Centre of Excellence for Infectious Diseases of Humans and Animals in East and Southern Africa and Pan-African Network for Rapid Research, Response, Relief and Preparedness for Infectious Disease Epidemics (PANDORA-ID-NET) Consortium Grant (EDCTP Reg/Grant RIA2016E-1609) funded by the European and Developing Countries Clinical Trials Partnership (EDCTP2) Programme.

This PhD research would not be possible without professional guidance and constant supervision from my main supervisor Prof. Gerald Misinzo (SUA) and Co-supervisors: Dr. Leonard Mboera (SACIDS Foundation for One Health), Prof. Julius Lutwama (Uganda Virus Research Institute in Uganda) and Prof. Janusz Paweska (National Institute for Communicable Disease in South Africa). I am grateful to Prof. Dr. Serge Muyldermans, Dr. Christopher Kariuki of Vrije Unirversitet Brussel in Belgium, and Mr. Edgar Kigozi from Makerere University, Kampala in Uganda for their technical guidance on recombinant protein engineering and production.

I highly acknowledge technical assistance from Ms. Mariam Makange and Mr. Charles Kayuki from SACIDS Molecular Biology Laboratory. I am also grateful for Dr. Calvin Sindato of the National Institute for Medical Research and Dr. Susan Rumisha of the Malaria Atlas Project, Geospatial Health and Development, Telethon Kids Institute in Australia for their enthusiasm and great support during the field work.

DEDICATION

This work is dedicated to my lovely parents, Eugene Mwanyika and Mama Josepha Mkongwa, my wife Teckla and my daughters Glory, Gerladina and Gladies for their prayers and moral support throughout the period of my PhD studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ORGANISATION OF THE THESIS

This thesis is organized in the "published papers format" and consists of Three Chapters as follows:

a. Chapter One: General Introduction

b. Chapter Two: Paper-based chapter consists of published papers and manuscripts from this study;

   i. Paper 1: Dengue virus infection and associated risk factors in Africa: a systematic review and meta-analysis. Published in *Viruses*, 13(4), 536: 394, April, 2021.doi.org/10.3390/v13040536.

   ii. Paper 2: Seroprevalence and associated risk factors of chikungunya, dengue and Zika in eight districts in Tanzania. Published in *International Journal of Infect Diseases*, S1201-9712(21)00674-3, August 2021. doi: 10.1016/j.ijid.2021.08.040

   iii. Paper 3: Circulation of dengue serotype 1 viruses during the 2019 outbreak in Dar es salaam, Tanzania. Published in *Pathogens and Global Health* 1-9, April 2021.doi.org/10.1080/20477724.2021.1905302.

   iv. Manuscript 4: Nanopore genome sequencing and phylogeographic analysis of dengue serotype 1 virus responsible for the 2019 outbreak in Dar es salaam, Tanzania (in preparation).

   v. Manuscript 5: *In silico* epitope prediction and expression of DENV-2 envelope protein domain III in *Escherichia coli*: a target design approach for nanobody binding (in preparation)

c. Chapter Three: General discussion, Conclusions and Recommendations

   i. Summary of the findings from specific objectives and recommendations

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
ºC	Degree celsius
DENV	Dengue virus
DNA	Deoxyribonucleic acid
DF	Dengue fever
DHF	Dengue haemorrhagic fever
DSS	Dengue shock syndrome
ELISA	Enzyme- Linked Immunosorbent Assay
EDCTP	European and Developing Countries Clinical Trials Partnership
EDIII	Envelope Protein Domain III
HPD	High Posterior Density
IPTG	Isopropyl-ß-D-1-Thiogalactopyranoside
NCBI	National Centre for Biotechnology Information
OR	Odds Ratio
ORF	Open Reading Frame
OD	Optic Density
PFU	Plaque forming units
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RDT	Rapid diagnostic test
RNA	Ribonucleic acid
SDS-PAGE	Sodium Dodecyl Polyacrylamide Gel

CHAPTER ONE

1.0 GENERAL INTRODUCTION

1.1 Global burden of dengue

Dengue is the most important mosquito-borne viral disease affecting nearly 390 million people annually, of which 96 million develop clinical signs that lead to approximately 500,000 hospitalizations and 25,000 deaths (Gubler, 1998). Dengue is endemic in more than 100 countries with 70% of the burden reported in Asia (Bhatt *et al*., 2013). In Africa, dengue is endemic in 34 countries with 15.2% of the continent having tropical climate that favour dengue virus transmission and spread (Were *et al*., 2012; Sintayehu *et al*., 2020).

1.2 The virus

Dengue virus (DENV) is spherical lipid-enveloped virus of approximately 50nm diameter, single-stranded positive sense RNA genome of nearly 11kb size. DENV belongs to family *Flaviviridae* in the genus flavivirus. DENV RNA genome is a single open reading frame (ORF) flanked by untranslated 5’ and 3’ regions (Figure 1.1). The ORF comprises of a polyprotein of 3400 amino acids that encodes for three structural proteins, the capsid protein (C), pre-membrane (prM) and envelope (E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Gebhard *et al*., 2011). Structural proteins play major biological functions for viral particle assembly. Glycoprotein E is responsible for virus attachment and membrane fusion in acidic endosomes as well as producing neutralizing antibodies. Non-structural glycoproteins are involved in replication and modulation of host immune responses. Non-structural glycoprotein 1 (NS1) is vital for virus viability, NS3 for protease, helicase and RNA triphosphatase activities, NS5 exhibits RNA-dependent RNA polymerase (RdRp) activity (Morrison *et al*., 2012).

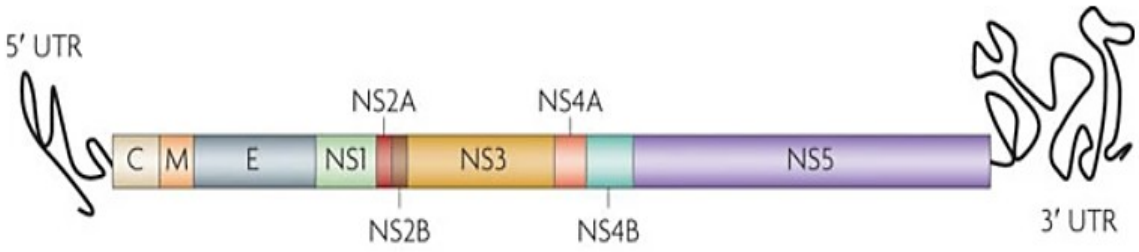


Figure 1.1: Schematic diagram of dengue virus genome
(Source: Guzman *et al.,* 2010).

DENV exists as four different serotypes (DENV-1-4). Each serotype can cause an infection of varying severity ranging from mild dengue fever (DF), severe dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Bhatt *et al*., 2013). Primary infection with any serotype induces lifelong immunity to that serotype, but a short-lived immunity following secondary infection by a different serotype due to antibody-dependent enhancement (Chareonsirisuthigul *et al*., 2007).

1.2.1 Vectors and transmission cycle

Dengue virus is transmitted mainly by *Aedes aegypti* and to less extent by *Ae. albopictus* mosquitoes. It is maintained in sylvatic cycles between non-human primates and Aedes mosquitoes that spill over into transmission cycles involving humans and peri domestic mosquitoes as shown in Figure 1.2 (Whitehead *et al*., 2007; Vasilakis *et al*., 2011). The distribution of dengue mosquito vectors is affected by climate change and human-modified environment resulting from human activities such as urbanization and deforestation (Higa *et al*., 2011; Ebi *et al*., 2016). These ecological changes may contribute to the geographic expansion of mosquito vectors and dengue transmission into non-affected areas.

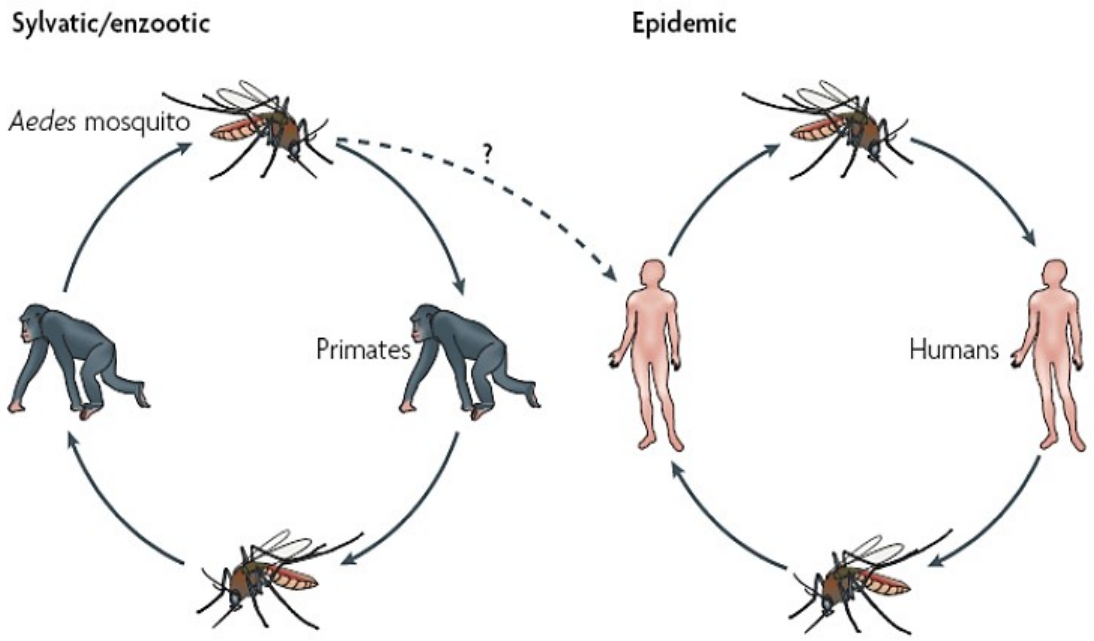


Figure 1.2: Transmission of dengue viruses
(Source: Whitehead *et al*., 2007)

1.2.2 Clinical presentation

Dengue virus infections show a spectrum of clinical manifestations ranging from mild febrile illness to severe disease with a short incubation period of 3─7 days (WHO, 2009). Approximately 60─80% of DENV infections are asymptomatic (Martinez *et al*., 2019). The early stage of infection is characterized by fever (≥ 38.5°C) and at least two of the following clinical symptoms: headache, vomiting, muscle, joint pain and a degree of transient rashes. If not recovered, the haemorrhagic fevers can occur after this stage characterized by significant mucosal bleeding, some plasma leakage, leukopenia and thrombocytopenia elevation. Shock and organ failure may develop following persistent bleeding (Thomas *et al*., 2010; Simmons *et al*., 2012).

1.2.3 Diagnosis of dengue

Dengue virus viremia occurs typically 2─3 days prior to onset of fever and can last up to 7 days. During this early stage of infection, the virus, its nucleic acid and antigens can be detected in the serum or plasma samples. At the end of the acute phase, specific anti-dengue antibodies can be detected from the blood of infected individuals. Generally, serum is the primary biological specimen for dengue diagnosis, other specimen types including plasma, cerebrospinal fluid, pleural fluids and reticuloendothelial tissues from liver, lungs or lymph nodes (WHO, 2009). The laboratory test sensitivity depends largely on the timing of specimen collection and proper storage since DENV particles are heat-sensitive (Kao *et al*., 2005). Therefore, appropriate handling of biological samples is usually required to maintain the accuracy of the test results.

1.2.3.1 Virus isolation

Isolation of dengue virus from acute serum or plasma samples within 1─5 days after onset of fever is done in mosquito cell lines, the C6/36 or AP64, CLA-1 cells and mammalian cells; Vero cells (Green monkey kidney cells), Lilly laboratories cell monkey kidney 2 (LLCMK2) and Baby hamster kidney 21 (WHO, 2009; Guzman *et al*., 2016). However, culture technique is constrained by the need for Biosafety level 3 facilities (Bhat *et al*., 2015).

1.2.3.2 Detection of viral nucleic acid

Detection of DENV ribonucleic acid (RNA) in clinical samples is done by conventional reverse-transcription polymerase chain (RT-PCR) with primers that anneal to specific regions of target RNA templates (Muller *et al*., 2017). RT-PCR method is more sensitive than culture methods. Several variations of RT-PCR method for the molecular detection of

DENV are available. They include, nested RT-PCR that involves two rounds of amplification with the first PCR using universal primers to generate the products as templates for the second PCR. In the second PCR more specific products are generated using more specific primers that target different regions of the DENV genome specific to a particular serotype (Lanciotti *et al*., 1992). Real time RT-PCR method uses serotype-specific primers and fluorescent probes such as TaqMan and SYBR green to improve detection sensitivity (Santiago *et al*., 2013). The main limitation of molecular detection methods is the high cost of reagents, equipment and requirement of well-trained personnel.

1.2.3.3 Detection of non-structural protein 1 (NS1)

The existence of closely related viruses in the genus Flavivirus such as yellow fever virus (YFV) and Zika virus (ZIKV) compromise the sensitivity of most antigen detection tests. Non-structural protein 1 (NS1) are highly conserved glycoproteins secreted into the blood circulation by the infected host cells and can be detected using enzyme linked immunosorbent (ELISA) or commercial rapid diagnostic test (RDTs) up to 9 days (Duyen *et al*., 2011; Kassim *et al*., 2011). The limitation of NS1 detection include its low sensitivity in secondary dengue infection (Mardekian and Roberts, 2015).

1.2.3.4 Detection of antibodies

Antibodies against dengue virus are detectable 3─5 days after the onset of fever. In this period anti-dengue immunoglobulin M (IgM) antibodies can be detected by IgM capture enzyme linked immunosorbent assay (MAC-ELISA) or IgM rapid diagnostic tests (RDTs) in 50% of victims and over 90% within 10 days of infection (Blacksell *et al.*, 2012). In the primary infection, IgM antibodies normally are detectable much earlier than

immunoglobulin G (IgG) antibodies indicatingrecent dengue infection. The use of IgG ELISA is limited due to wide cross-reactivity among flavivirus and its long persistence in the circulation that affects the ability to differentiate an acute infection from past flavivirus exposure (Hunsperger *et al*., 2016).

1.2.3.5 Hemagglutination inhibition (HI) assay

The principle of hemagglutination inhibition assay is based on the ability of dengue envelop proteins to agglutinate red blood cells. This reaction is inhibited by anti-dengue antibodies present in the serum of infected individuals and thus the level of inhibition can be quantified. Haemaglutination inhibition (HI) assay test can distinguish primary and secondary infection by measuring a 2-fold increase of antibody titre during secondary infection (1:640 to 1:1280). The main disadvantage of HI assay is the cross-reactions with other flaviviruses such as Zika and yellow fever (Peeling *et al*., 2010).

1.2.3.6 Plaque reduction neutralization test (PRNT)

Plaque reduction neutralization (PRNT) assay is the gold standard test for measuring anti-DENV specific neutralizing antibodies in the serum of infected individuals. In PRNT, the neutralizing antibodies inactivate viruses and block their replication inside the target host cells such as Vero and BHK21. The positive samples are defined by a 50% reduction or more of the viral particles known as plaque forming units (PFU). This method is labour intensive and requires special laboratory biosafety level 3 (WHO, 2009).

1.3 Nanobodies

1.3.1 Structure

Nanobodies (VHHs) are single-domain antigen binding fragments from camelid heavy chain only antibodies (HCAbs). The heavy chain only antibodies are unique immunoglobulin G isotypes devoid of light chains (VL and CL) and the first constant region (CH1). The HCAbs are found in the sera of camelids such as camels, llamas and in the certain cartilaginous fish such as sharks (Hamers-Casterman *et al*., 1993; Muyldermans, 2013). Nanobody gene segment is organized into four conserved frame work regions (FR1, FR2, FR3 and FR4) interspaced with three complementary determining regions (CDR1, CDR2, CDR3 and CDR4) (Figure 1.3) which are vital for antigen diversity and specificity (Sroga *et al*., 2020).

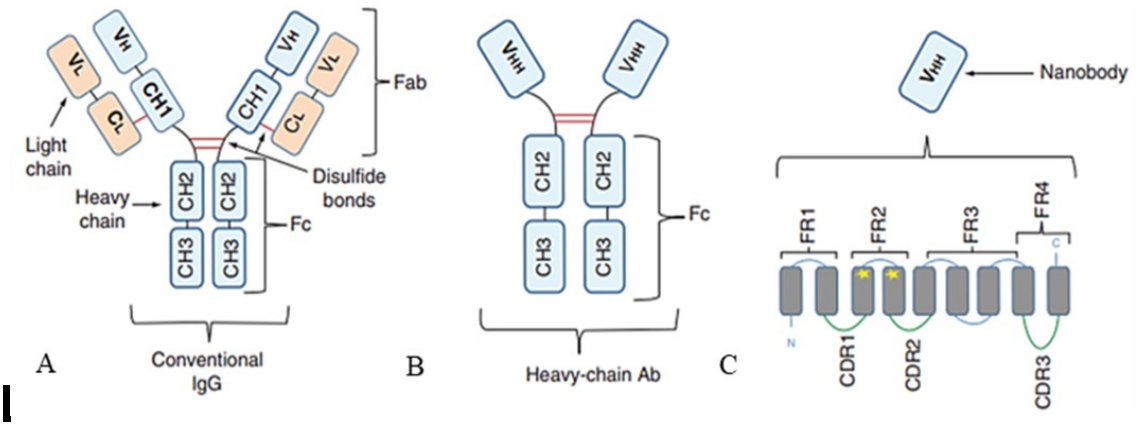


Figure 1.3: The structure of Camelidae nanobodies. (A) Conventional camelid antibody, (B) Heavy chain only antibody and (C) a nanobody with the organization of complementary determining regions (CDRs) and frame work regions (FRs)

(Source: Sroga *et al.*, 2020)

1.3 2 Nanobody features for diagnostic application

Nanobodies have unique properties potential for development of diagnostic tools; (i) high specificity and affinities for diverse antigens; (ii) single domain nature originating from a single V (D) J exon which encodes nanobody gene (VHH) of~450 bp without post-transcriptional splicing that makes genetic manipulation easy; (iii) Small in size approximately ~15kDa compared with traditional recombinant antibody fragments, Fab (~55kDa) and scFv (~28); (iv) ability to bind hidden epitopes due to small size and long CDR3 loops that expand antigen recognition repertoire of the epitopes located in protein cavities; (v) better solubility due to four hydrophilic amino-acid substitutions (V37F, G44E, L45R and W47G) in the framework region 2 that replace the classical hydrophobic VH-VL interface and (vi) chemical and thermal stable (Muyldermans, 2013; Salvador *et al.*, 2019).

1.4 Dengue in Tanzania

Dengue has a historical relationship with Tanzania. The disease is thought to be first described by the Spanish sailors in the southern coast of Tanzania during the 15th Century. The word "dengue" is a Spanish homonymy for the Kiswahili phrase "*Ki denga pepo*" meaning a sudden cramp like seizure caused by an evil spirit (Christie, 1872). First dengue virus outbreak was reported on the islands of Zanzibar in 1870 and later the virus was reported throughout the Indian Ocean Islands (Amarasinghe *et al*., 2011). Before 2010, very little was known regarding the epidemiology of dengue in Tanzania and its impact has been underestimated or underreported possibly due to inadequate diagnostic capacities. Periodic outbreaks and increasing serological evidence for the circulation of DENV in different parts of Tanzania (Hertz *et al*., 2012; Vairo *et al*., 2012; Chipwaza *et al*., 2014; 2016; Faustine *et al*., 2017; Ward *et al*., 2017; Chipwaza *et al*., 2020) prompted the need to study the epidemiology of the virus (Figure 1.4).

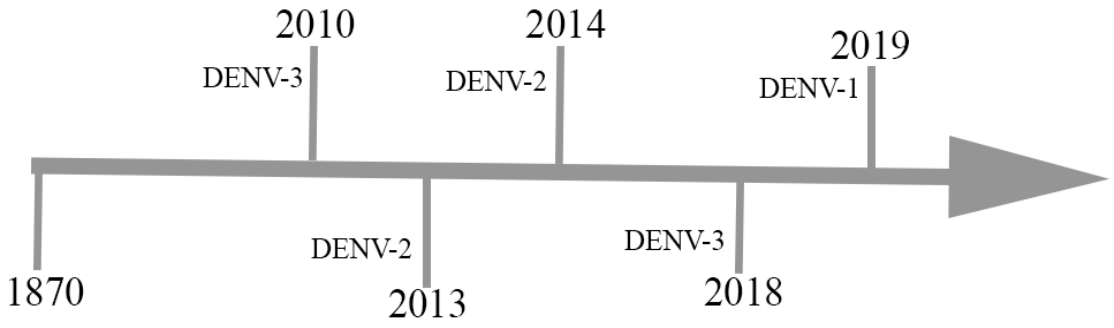


Figure 1.4: Main outbreak milestones of dengue virus in Tanzania
(Source: Authors' own)

1.5 Problem statement, Study Justification and Objectives

1.5.1 Problem statement and Justification of the study

Dengue is the disease of global health concern that has significant health and socioeconomic implication. In Tanzania, there is limited population-based evidence for the burden, drivers and vulnerability for dengue. Most data on the disease are from health facility-based studies (Ward *et al*., 2017), thus the magnitude, distribution and molecular epidemiology of dengue virus (DENV) in the country remains not known. Furthermore, routine serological rapid tests for DENV have inherent cross-reactivity and the use of more specific genomic-based assays is limited in Tanzania, therefore the application of nanobodies that have high specificity beyond the reach of conventional antibodies can help to improve the point of care diagnosis of DENV infection.

1.5.2 Main objective

To examine the epidemiological and antigenic characteristics of dengue virus in Tanzania for the development of nanobodies focusing at the rational diagnosis and control of dengue in the region.

1.5.3 Specific objectives

i. To analyse the magnitude and risk factors of DENV infection in Africa
ii. To determine seroprevalence and risk factors of DENV among humans in selected areas of Tanzania
iii. To determine the genetic diversity spanning the complete genome sequences of DENV recovered from humans in Tanzania
iv. To examine DENV envelope protein domain III for development of nanobodies

1.5.4 Key research questions

i. What is the magnitude, distribution and risk of DENV infection in Africa?
ii. What is the magnitude, distribution and risk of DENV in Tanzania?
iii. What is the genetic diversity of DENV serotypes/genotypes circulating in human population in Tanzania?
iv. What characteristics of DENV envelope protein domain III enable it to be used as target antigen for development of nanobodies?

CHAPTER TWO

Paper One

Dengue virus infection and associated risk factors in Africa: A systematic review and meta-analysis

Gaspary O. Mwanyika[1,2,3], Leonard E.G. Mboera[1], Sima Rugarabamu[1,2,4], Baraka Ngingo[1], Calvin Sindato[1,5], Julius Lutwama[6], Janusz T. Paweska[7] and Gerald Misinzo[1,2]

[1]SACIDS Africa Centre of Excellence for Infectious Diseases, Sokoine University of Agriculture, Morogoro, Tanzania

[2]Department of Veterinary Microbiology, Parasitology and Biotechnology, Sokoine University of Agriculture, Morogoro, Tanzania

[3]Department of Medical Sciences and Technology, Mbeya University of Science and Technology, Mbeya, Tanzania

[4]Department of Microbiology and Immunology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

[5]National Institute for Medical Research, Tabora Research Centre, Tabora, Tanzania;

[6]Department of Arbovirology, Emerging and Re-emerging Infectious Diseases, Uganda Virus Research Institute, Entebbe, Uganda

[7]National Institute for Communicable Diseases, National Health Laboratory Service, Sandringham, Johannesburg, South Africa

Viruses, 13(4), 536. https://doi.org/10.3390/v13040536 (April, 2021)

Paper Two

Seroprevalence and associated risk factors of chikungunya, dengue and Zika in eight districts in Tanzania

Gaspary O. Mwanyika[1,2], Calvin Sindato[1], Sima Rugarabamua,[2], Susan F. Rumisha,[3,4], Esron D. Karimuribo[1,5], Gerald Misinzo[1,2], Mark M. Rweyemamu[1], Muzamil M. Abdel Hamid[6], Najmul Haider[6], Francesco Vairo[8], Richard Kock[7] and Leonard E.G. Mboera[1]

[1]SACIDS Foundation for One Health, Sokoine University of Agriculture, Morogoro, Tanzania

[2]Department of Veterinary Microbiology, Parasitology and Biotechnology, Sokoine University of Agriculture, Morogoro, Tanzania

[3]National Institute for Medical Research, Headquarters, Dar es Salaam, Tanzania

[4]Malaria Atlas Project, Geospatial Health and Development, Telethon Kids Institute, West Perth, Western Australia

[5]Department of Veterinary Medicine and Public Health, Sokoine University of Agriculture, Morogoro, Tanzania

[6]Institute of Endemic Diseases, Medical Campus, University of Khartoum, Sudan

[7]The Royal Veterinary College, University of London, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire, United Kingdom

[8]National Institute for Infectious Diseases L. Spallanzani, Rome, Italy.

Int J Infect Dis, S1201-9712(21)00674-3. doi: 10.1016/j.ijid.2021.08.040 (August, 2021)

Paper Three

Circulation of dengue serotype 1 viruses during the 2019 outbreak in Dar es Salaam, Tanzania

Gaspary O. Mwanyika[1,2,3], Leonard E.G. Mboera1, Sima Rugarabamu[1,2,4], Mariam Makange[1], Calvin Sindato[1,5], Julius J. Lutwama[6], Janusz T. Paweska[7], and Gerald Misinzo[1,2]

[1]SACIDS Africa Centre of Excellence for Infectious Diseases, Sokoine University of Agriculture, Morogoro, Tanzania

[2]Department of Veterinary Microbiology, Parasitology and Biotechnology, Sokoine University of Agriculture, Morogoro, Tanzania

[3]Department of Health Science and Technology, Mbeya University of Science and Technology, Mbeya, Tanzania

[4]Department of Microbiology and Immunology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

[5]National Institute for Medical Research, Tabora Research Centre, Tabora, Tanzania;

[6]Department of Arbovirology, Emerging and Re-emerging Infectious Diseases, Uganda Virus Research Institute, Entebbe, Uganda

[7]National Institute for Communicable Diseases, National Health Laboratory Service, Sandringham, Johannesburg, South Africa

Pathogens and Global Health 1-9. doi.org/10.1080/20477724.2021.1905302 (April 2021)

Manuscript Four

Nanopore genome sequencing and phylogeographic analysis of dengue serotype 1 virus responsible for the 2019 outbreak in Dar es Salaam, Tanzania

Gaspary O. Mwanyika[1,2,3], Leonard E.G. Mboera[1], Sima Rugarabamu[1,2,4], Calvin Sindato[5], Julius J. Lutwama[5], Janusz T. Paweska[7], and Gerald Misinzo[1,2]

[1]SACIDS Africa Centre of Excellence for Infectious Diseases, Sokoine University of Agriculture, Morogoro, Tanzania

[2]Department of Veterinary Microbiology, Parasitology and Biotechnology, Sokoine University of Agriculture, Morogoro, Tanzania

[3]Department of Medical Sciences and Technology, Mbeya University of Science and Technology, Mbeya, Tanzania

[4]Department of Microbiology and Immunology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

[5]National Institute for Medical Research, Tabora Research Centre, Tabora, Tanzania;

[6]Department of Arbovirology, Emerging and Re-emerging Infectious Diseases, Uganda Virus Research Institute, Entebbe, Uganda

[7]National Institute for Communicable Diseases, National Health Laboratory Service, Sandringham, Johannesburg, South Africa

Manuscript in preparation

Abstract

Background: Dengue is the most important mosquito-borne viral disease in tropical and subtopic regions of the world. The genetic variations of dengue virus (DENV) serotypes present differences in virulence and epidemic potential. This phenomenon probed the need for genomic characterization of DENV serotypes. The objective of this study was to determine the evolutionary origin and geographical dispersal of DENV-1 strains detected during the 2019 outbreak in Dar es Salaam, Tanzania**.**

Methods: Nanopore MinION technology was used to generate genome sequences of DENV-1 strains and performed sequence analysis using NanoGalaxy tool kit. The phylogenies were reconstructed by maximum likelihood method with GTR+G+I model. Divergence time and evolutionary rates were estimated by Bayesian Evolutionary Analysis by Sampling Trees program with Markov Chain Monte Carlo posterior probabilities. Phylogeographic reconstruction was conducted in PastML with maximum likelihood marginal posterior probabilities approximation.

Results: The Tanzanian DENV-1 strains were found to be clustered with genotype V strains reported in China and Japan in 2019. The genotype originated from the most recent common ancestor approximately 132.9 years ago with an evolutionary rate of 5.15×10^{-4} (95% HPD interval (4.04×10^{-4}, 6.26×10^{-4}) substitution rates per site per year. Phylogeographic reconstruction showed that there was an introduction of DENV-1 genotype V into Tanzania through a single source in Asia during the 2019 outbreak in Dar es Salaam**.**

Conclusions: This study provides evidence of DENV-1 genotype V occurrence in Tanzania that was possibly introduced through a single source in Asia. These findings emphasize the need to enhance genomic surveillance of dengue in Tanzania.

Keywords: nanopore; genome; dengue virus; serotype 1; phylogeography; Tanzania

Introduction

Dengue is an important mosquito-borne viral disease in tropical and subtropical countries. Globally, the disease is responsible for approximately 500 000 hospitalizations and 25 000 deaths each year (Bhatt *et al*., 2013). In the past 50 years, the incidence of dengue has increased by a 30-fold affecting over 100 countries worldwide (Stanaway *et al*., 2016). The dengue virus (DENV) is a positive sense single-stranded RNA virus of the family *Flaviviridae* and genus flavivirus. DENV genome is 10.7kb long with single open reading frame (ORF) containing 3,391 amino acids. The ORF comprises of three structural proteins: capsid (C), precursor membrane (prM) and envelope (E) and seven non-structural proteins, NS1, NS2A, NS2B, NS3, NS4, NS4B and NS5 (Gebhard *et al*., 2011). DENV exist in four different serotypes 1 to 4 (DENV 1─4) that share approximately 60─70% amino acid homology. Each serotype exists in different genetic distinct subtypes known as genotypes that may present differences in virulence and epidemic potential (Guzman *et al*., 2010).

The phylogenetic studies of DENV-1 based on complete envelope gene group DENV-1 strains into five genotypes: (i) genotype I (GI) representing strains from South East Asia, China and East Africa; (i) genotype II (GII) representing strains reported in Thailand between 1950─1960s; (iii) genotype III (GIII) representing the sylvatic strain detected in

Malaysia (GIII); (iv) genotype IV (GIV) representing strains from west pacific Islands and Australia; and (v) genotype V (GV) that represent all the strains reported in Americas and Africa and limited strains from Asia (Holmes and Twiddy, 2003; Weaver and Vasilakis, 2009). Several authors have demonstrated the practicability of complete envelope gene for inferring the evolutionary relationship of DENV strains and their assignment into correct genotypes (Klungthong *et al*., 2008; Yu *et al*., 2019).

Since 1964, dengue serotype 1 virus epidemics have been reported in more than 15 African countries (Amarasinghe *et al*., 2011; Mwanyika *et al*., 2021a). Although DENV-1 strains cause frequent epidemics in Africa, the genomic characterization and evolution studies are limited due to high cost of sequencing technologies and lack of adequate skilled personnel (Sharma *et al.,* 2015). In 2019, DENV-1 outbreak was reported for the first time in Tanzania that affected nine regions (Arusha, Dar-es-Salaam, Dodoma, Kagera, Lindi, Morogoro, Pwani, Ruvuma and Tanga). During the outbreak, a total of 6,917 confirmed cases and 13 deaths were reported with Dar es Salaam city accounting for more than one-third of the cases (Mwanyika *et al*., 2021b). The genetic variations within DENV serotype give rise to genotypes and distinct lineages that exhibit differences in virulence, transmission potential and disease severity (OhAinle *et al*., 2011; Santiago *et al*., 2019).

There are limited studies that have utilized complete genome sequences to establish the evolutionary relationships of DENV in resource-limited settings. A recent study in Guangzhou, China based on complete genome sequences of DENV-1 has revealed new DENV-1 genotypes that caused unusual outbreaks during the 2002─2014 period (Ma *et al*., 2021). Further, it was shown that the new invading DENV-1 genotypes had the

lowest nucleotide similarity in E and NS3 gene sequences (Ma *et al*., 2021). Thus, it is useful to understand the distinct genetic subtypes of circulating DENV strains using both individual genes and complete genome sequences. Oxford nanopore MinION is a low-cost, portable and long-read third generation sequencing technology that can serve this purpose in resource-limited settings (Hill *et al*., 2019). Nanopore can sequence long-read genomes that are easy to assemble and call structural variants in complex repetitive genome regions. In Tanzania, there is limited information on the evolutionary origin and geographical dispersal of circulating DENV serotypes/genotypes (Gautret *et al*., 2010; Mboera *et al*., 2016; Vairo *et al*., 2016). The objective of this study was to determine the evolutionary origin and transmission dispersal of DENV-1 strains that were detected during the 2019 outbreak in Dar es Salaam, Tanzania.

Methods

Clinical samples

A total of four RT-PCR positive serum samples obtained from our previous study were selected for complete genome sequencing. The quality of extracted RNA was evaluated using NanoDrop ND1000 spectrophotometer at 260 and 280 absorbance units (GE Healthcare, Buckinghamshire, UK) and the resulting cDNA starting materials required for library preparation and nanopore ligation sequencing with Qubit dsDNA HS Assay kit. One sample was finally selected for genome sequencing.

Viral RNA extraction and genome amplification

Dengue virus ribonucleic acid (RNA) was purified from 140 µL of serum using the QIAmp Viral RNA Mini kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. The extracted RNA samples were used to synthesize cDNA templates by the

SuperScript III First Strand Synthesis system (InvitrogenTM, CA, USA) using 10 µM of DENV gene-specific reverse primer F5-10605R1. The cDNA templates were synthesized using 8 µL of RNA mixed with 1 µL of reverse primer and 1 µL of dNTPs mix (10 mM) and incubated at 65ºC for 5 min to denature RNA secondary structures and thereafter cooled on ice for 5 minutes. The components of reverse transcription reaction mixture were prepared in a separate 0.2 mL RNAase-free tube to a total of a 10 µL reaction volume containing 2 µL 10x RT buffer, 4 µL of 25 mM $MgCl_2$, 2 µL of 0.1 M DDT (Dithiothreitol), 1 µL of RNAseOUT (40 U/ µL) and 1 µL of SuperScript III RT enzyme (200 U/ µL). The cDNA synthesis mix was added to 10 µL RNA/primer mix on ice to bring the volume to 20 µL and incubated at 25ºC for 5 minutes to activate the enzyme followed by incubation at 50ºC for 50 min to synthesize cDNA and at 85ºC for 5 minutes to terminate the reactions in a Mastercyler nexus gradient thermocycler (Eppendorf, Hamburg, Germany). The cDNA products were placed on ice to cool and 1 µL of RNase H was added and incubated at 37ºC for 20 minutes to degrade viral RNA and career RNA. The final cDNA products were stored at -80ºC until analysed. A five-fragment strategy using the primers previously described was used for genome amplification (Sessions *et al*., 2013) (Appendix 5; Table S1). PCR was performed in a 12.5 µL reaction volume containing 6.5 µL of 2x LongAmp master mix, 0.5 µL of sense primer (10 µM), 0.5 µL of antisense primer (10 µM), 2 µL of cDNA template and 3 µL of Nuclease-free water. The DENV-1 genome amplification was performed with LongAmp *Taq* DNA polymerase (NEB Inc, USA) and the annealing temperatures for each set of primers were optimized in a Mastercyler nexus gradient thermocycler at the temperature range of 45ºC to 65ºC.

Library preparation and nanopore sequencing

The PCR amplicons were gel purified prior to library preparation using GeneJET extraction kit according to the manufacturer' instructions (ThermalFisher K0691). The purified DNA fragments were cleaned using 0.4x Agencourt AmPure XP beads and the concentration of DNA normalized using Qubit dsDNA HS Assay kit (ThermalFisher Q32851) on a Qubit fluorimeter (ThermalFisher). The sequencing library was prepared using Ligation Sequencing (SQK-LSK109, ONT, Oxford, UK) and Native Barcoding Expansion kits (EXP-FLP002, ONT, Oxford, UK) according to manufacturer' protocol. The end-prepped DNA was prepared on ice using NEBNext Ultra II End repair/dA-tailing Module (E7546) followed by barcoding using native barcode, NEBNext Ultra II Ligation Master Mix and NEBNext Ligation Enhancer following the manufacturer's instructions. Thereafter, the adaptors were ligated to pooled barcoded amplicons using NEBNext Quick Ligation Module reagents (E7595) and the concentration of adaptor-ligated amplicons quantified using Qubit dsDNA HS Assay kit. The MinION Mk 1C (MC-110368) flow cell R9.4.1 (FLO-MIN106, ONT, Oxford, UK) was primed using Flow Cell Priming buffer (EXP-FLP002, ONT, Oxford, UK) according to the manufacturers' instructions. The final library mix containing 37.5 µL Sequencing Buffer (SQK-LSK109, ONT, Oxford, UK), 25.5 µL Loading Beads (SQK-LSK109, ONT, Oxford, UK) and 12 µL of DNA library was loaded into MinION Mk 1C flow cell for sequencing run.

Bioinformatics analysis pipeline

The bioinformatics analyses were performed using NanoGalaxy tool kit (https://galaxyproject.org/use/nanogalaxy/). Base calling was done with nanopore guppy base caller v3.4.4 and the quality of reads was evaluated using nanoplot v1.28.2. Porechop v0.2.4 was used to trim adaptors and barcodes and the trimmed reads were assembled using canu v2.1.1 (Koren *et al*., 2017) with a corrected error rate of 0.144 and 40x target

coverage. The quality of assembly was evaluated using QUAST v5.0.2 (Gurevich *et al*., 2013). The assembled reads were mapped to NCBI reference dengue virus 1 complete genome (NC_001477.1) using GraphMap v0.5.2 (Sović *et al*., 2016) and the alignments were visualized in integrative genomics viewer v2.9.4. The unmapped and reads with similar external coordinates or low mapping quality (< 40) were filtered using samtools v1.9 (Li *et al*., 2009). The quality reads were subjected to genome detective virus tool version v1.132 https://www.genomedetective.com/app/typingtool/virus/) to generate consensus sequence alignment. Online open reading frame (ORF) finder (https://www. ncbi.nlm.nih.gov/orffinder/) and genome annotation transfer utility (GATU) (https://4virology.net/virology-ca-tools/gatu/) were used to identify and annotate the ORF respectively. Nucleotide and amino acid identities were determined using sequence identity and similarity bioinformatics (SIAS) tool (http://imed.med.ucm.es/Tools/sias.html).

Phylogenetic and phylogeographic analysis

The complete envelope (E) gene containing 1,485 nucleotides provide a better resolution to characterize the genetics and evolution of the DENV (Wu *et al.,* 2010), and thus it was selected for phylogenetic and phylogeographic analysis. The target Tanzanian DENV-1 E gene sequence plus other 87 reference DENV-1 E gene sequences available in the GenBank (https://www.ncbi.nlm.nih.gov/) that were reported from different countries between 1960 and 2019 representing the main genotypes of DENV-1 strains were used as input sequence data. The trimming and quality check of the sequences was conducted using Bioedit software v7.2 and multiple sequence alignments created with Clustral W implemented in MEGA X (Kumar *et al*., 2018). Phylogenetic analysis was performed by maximum likelihood method using general time reversal model (Rodriguez *et al*., 1990)

with gamma distribution rate of variation and a proportion of invariant sites (GTR+G+I) in 1000 bootstraps. Maximum likelihood fit method based on Akaike information criterion was used to select the best fitting model. TempEst v1.5.3 (Rambaut *et al*., 2016) was used to evaluate whether our sequence data had sufficient temporal signal for molecular clock analysis. Ancestral divergence time and reconstruction of maximum clade credibility (MCC) trees were inferred using the Bayesian Evolutionary Analysis by Sampling Trees software (BEAST) v1.10.4 with the Markov Chain Monte Carlo (MCMC) posterior probabilities based on strict molecular clock model. The MCMC chain was performed in 20 million steps with sub-sampling at every 5000 steps to obtain an effective sample size > 200 required for the accurate estimation of divergence time and evolutionary rate. The convergence of Markov chain was evaluated using Tracer v1.7.2 and MCC trees were reconstructed using TreeAnnotator v1.10.4 and visualized with FigTree v1.4.2 (Suchard *et al*., 2018). The Phylogeographic reconstruction was performed in PastML program using maximum likelihood marginal posterior probabilities approximation and Felsenstein 1981 (F81) model (Ishikawa *et al*., 2019). The final MCC tree was visualized and edited using iTOL (Letunic and Bork, 2019).

Ethical consideration

This study received ethical approval from Medical Research Coordinating Committee of the National Institute for Medical Research in Tanzania (Ref No. NIMR/HQ/R.8a/ Vol.IX/2974).

Results

DENV-1 genome sequences, quality and annotation

In this study, the amplification of four dengue virus 1 genome fragments (F1, F2, F3 AND F4) were optimized at annealing temperatures of 51 ºC (F1), 53 ºC (F2 and F3) and 47 ºC (F4) as shown in Figure 2.1.1 and 2.1.2. An attempt to amplify the last fragment (F5) was not successful with the present protocol using the temperature gradient from 46─65 ºC.

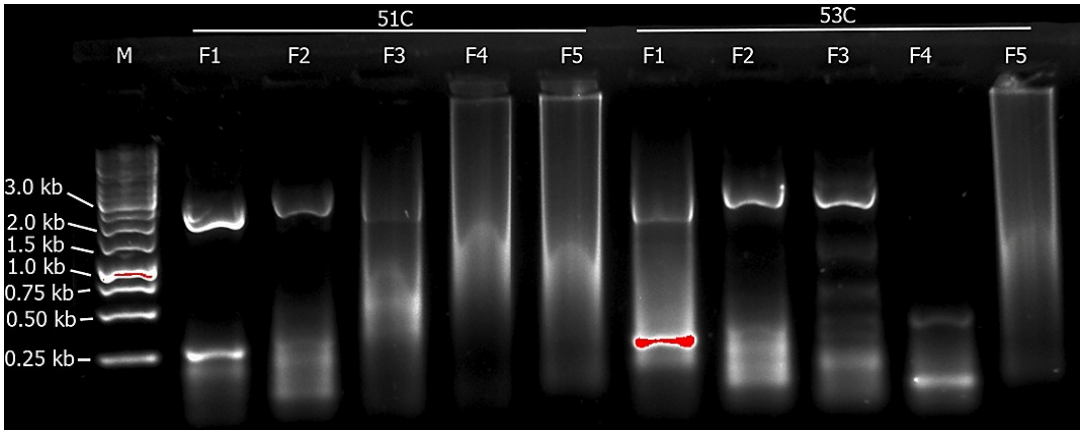


Figure 2.1.1: Gel electrophoresis image of PCR amplicons for DENV-1 genome fragments F1 (2205 bp), F2 (2490 bp) and F3 (2363 bp), F4 and F5 at 51˚C and 53˚C annealing temperatures. M, GeneralRuler 1 kb DNA ladder

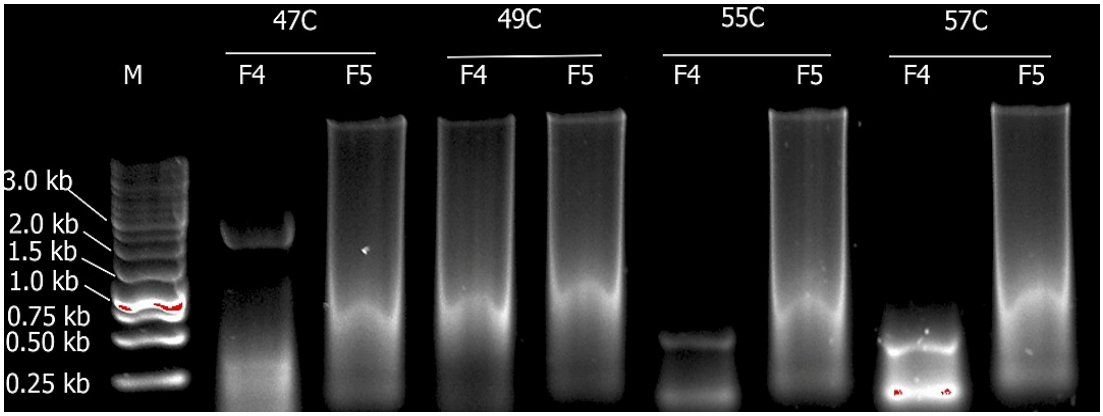


Figure 2.1.2: Gel electrophoresis image of PCR amplicon for DENV-1 genome fragment F4 (2216 bp) and F5 at 47─57˚C gradient temperature. M, GeneralRuler 1 kb DNA ladder

Nanopore amplicon sequencing generated a total of 2,756 reads and 2,529,696 nucleotide bases with mean read length of 917.9 bp. The summary quality control (QC), fastq data and genome consensus sequence are summarized in sequencing report (Appendix 6). The distribution of read lengths is shown in Figure 2.3.

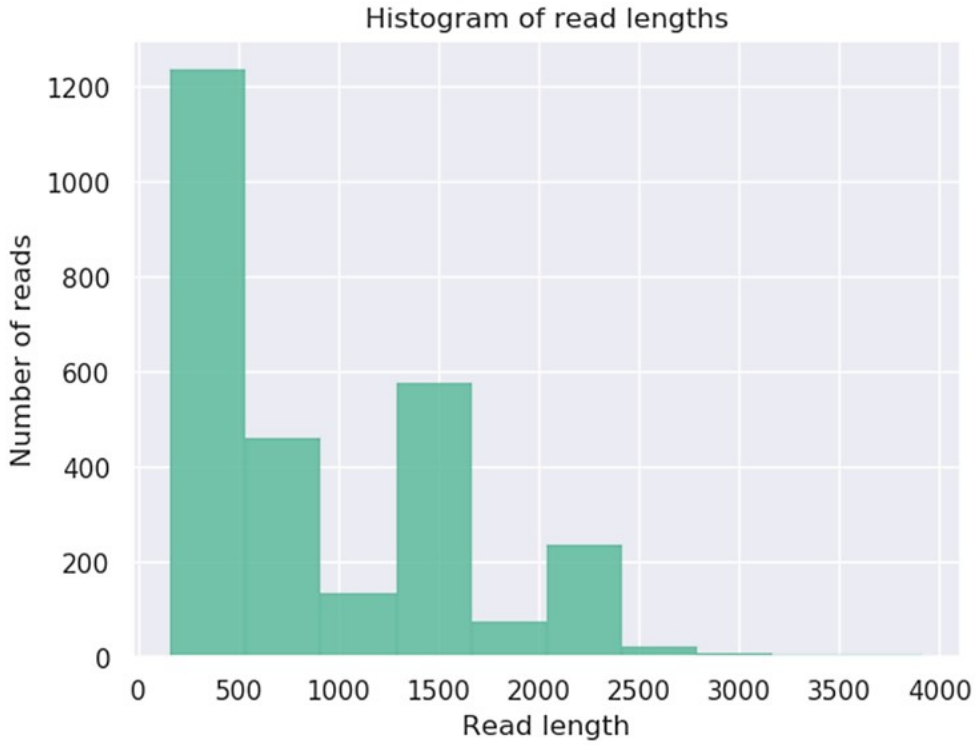


Figure 2.1.3: Distribution of nanopore reads by lengths

Dengue virus serotype 1 genome sequences with 8,541 nucleotides (nt) that correspond to 79.5% coverage of complete genome (10,700 nt) were obtained in this study. The coding region comprised of 8,445 nucleotides that corresponds to 2,815 amino acids with an open reading frame spanning between 95 and > 8439 nucleotides. A total of 13/15 (86.7%) annotations of the complete polyprotein gene fragments were identified from the coding region with three annotations showing 100% similarity and 10 annotations showing < 100─95% (Table 2.1.1).

Table 2.1.1: The annotation features of dengue virus serotype 1 genome sequence obtained from this study with respect to NCBI reference genome strain 45AZ5 (NC_001477.1)

Reference gene	Product	Start	Stop	Size	P. Size	Score	% Identity
Poly protein gene	polyprotein	1	8445	8445	10179	5712.11	82.1
Poly_1	anchored capsid protein ancC	1	342	342	342	220.3	98.2
Poly_2	capsid protein C	1	300	300	300	193.4	98
Poly_3	membrane glycoprotein precursor prM	343	840	498	498	345.5	99.4
Poly_4	protein pr	343	615	273	273	193.7	98.9
Poly_5	membrane glycoprotein M	616	840	225	225	153.7	100
Poly_6	envelope protein E	841	2325	1485	1485	995.3	99
Poly_7	nonstructural protein NS1	2326	3381	1056	1056	731.5	98.9
Poly_8	nonstructural protein NS2A	3382	4035	654	654	410.6	96.8
Poly_9	nonstructural protein NS2B	4036	4425	390	390	262.7	100
Poly_10	nonstructural protein NS3	4426	6282	1857	1857	1276.15	99.4
Poly_11	nonstructural protein NS4A	6283	6663	381	381	248.8	98.4
Poly_12	protein 2K	6664	6732	69	69	48.1	100
Poly_13	nonstructural protein NS4B	6733	7479	747	747	497.3	98.4
Poly_14	RNA-dependent RNA polymerase	7480	8445	966	2697	676.8	35.8
Poly_15	Not covered	-	-	-	-	-	-

Phylogeny of DENV-1 genotypes

Phylogenetic analysis of 87 DENV-1 E gene sequences revealed five DENV-1 genotypes; I, II and IV that represented strains mainly from Asia and Southern Pacific, genotype III represented a sylvatic strain from Malaysia and genotype V represented different DENV-1 strains reported in Africa, Asia and the Americas (Appendix 7). The Tanzanian DENV-1 strains clustered with genotype V strains reported in China (GenBank accession numbers MN923085.1, MN923096.1, MN923101.1, MN923102.1 and MW261832.1) and Japan (GenBank accession number LC485151) in 2019 with a nucleotide sequence divergence of 0.34─2.2% (Table 2.1.2). The phylogenetic tree representing different DENV-1 genotypes is shown Figure 2.1.4.

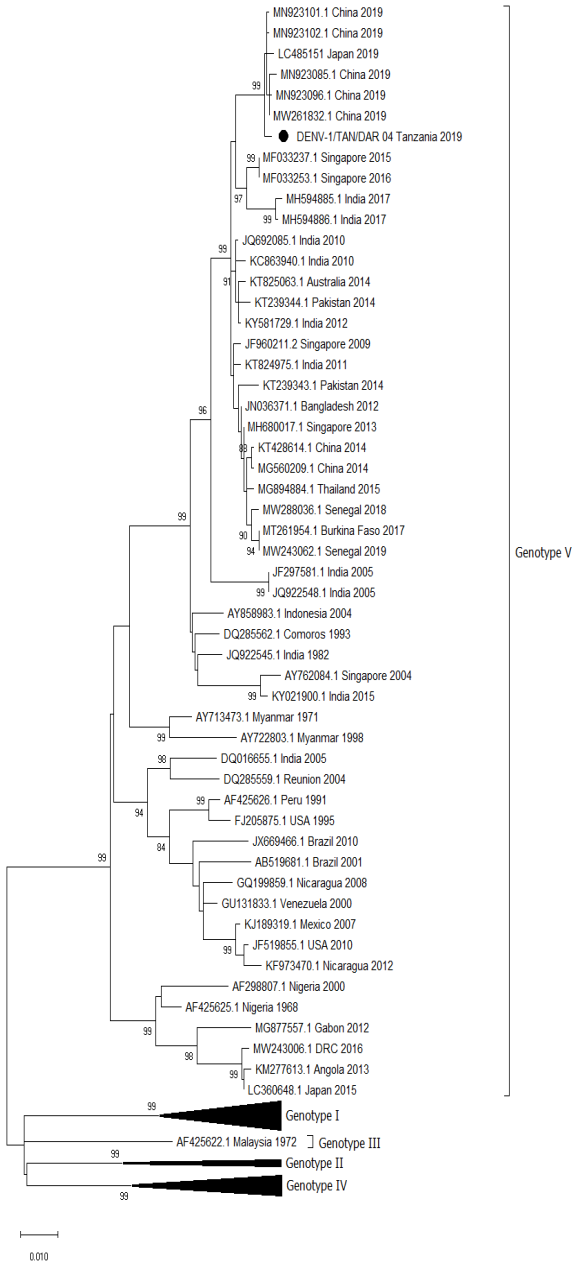


Figure 2.1.4: The phylogenetic analysis of DENV-1 complete envelope gene (1485 nt).

The analysis involved E gene sequences from the Tanzania strain and 86 reference strains from the GenBank. The evolutionary history was inferred using Maximum Likelihood method with GTR+G+I model and the analyses were conducted in MEGA X (Kumar *et al*., 2018). The Tanzanian DENV-1 strain from this study is shown in black cycle.

Table 2.1.2: Nucleotide (NT) and amino acid (AA) sequence identity of Tanzanian DENV-1 E gene compared to the most similar genotype V strains

S/n	Strain	NT identity (%)	AA identity (%)
1	MF033237.1/Singapore/2015	98.45	99.59
2	MF033253.1/Singapore/2015	98.45	99.59
3	MH594885.1/India/2017	97.84	99.39
4	MH594886.1/India/2017	97.97	99.79
5	LC485151/Japan/2019	99.52	100
6	MN923085.1/China/2019	99.59	99.59
7	MN923096.1/China/2019	99.59	99.79
8	MN923101.1/China/2019	99.66	100
9	MN923102.1/China/2019	99.66	99.79
10	MW261832.1/China/2019	99.66	99.79

Molecular clock and phylogeography of DENV-1 genotype V

The subset of envelope gene sequences of DENV-1 genotype V (n =51) containing the Tanzanian E gene sequence plus 50 E reference sequences available in the GenBank that were reported from 24 different countries in America (n = 6), Asia (n = 10) and Africa (n = 8) were used for phylogeographic analysis. The sampling periods of the sequences dated between 1968─2019. A positive temporal signal from this data with a correlation coefficient of 0.996 and R^2 of 0.997 was obtained. An effective sample size of 1492.6 was obtained in Markov chain to estimate the time to the most recent common ancestor (tMRCA). The results showed that the tMRCA for DENV-1 genotype V lineages were dated back to 1886.96 (1813.7, 1926.9) approximately 132.9 years ago. Bayesian evolutionary analysis showed that the genotype V lineages found in Asia originated from an ancestor that diverged approximately 77.9 years ago (1941.1) while the lineages reported in the Americas and Africa emerged from ancestors that diverged approximately 83.5 (1935.5) and 86.3 (1932.7) years ago, respectively (Figure 2.1.5).

Molecular clock analysis showed that DENV-1 genotype V lineages showed an evolutionary rate of 5.15x 10 $^{-4}$ [95% highest posterior density (HPD) interval (4.04 x 10 $^{-4}$, 6.26 x 10 $^{-4}$)] substitution rates per site per year. The phylogeographic reconstruction revealed the wide spread of genotype V lineages in Asia, the Americas and Africa. The results of this study showed that there is a possibility that the DENV-1 genotype V was introduced into Tanzania from a single source in Asia during the 2019 outbreak in Dar es Salaam (Figure 2.1.6).

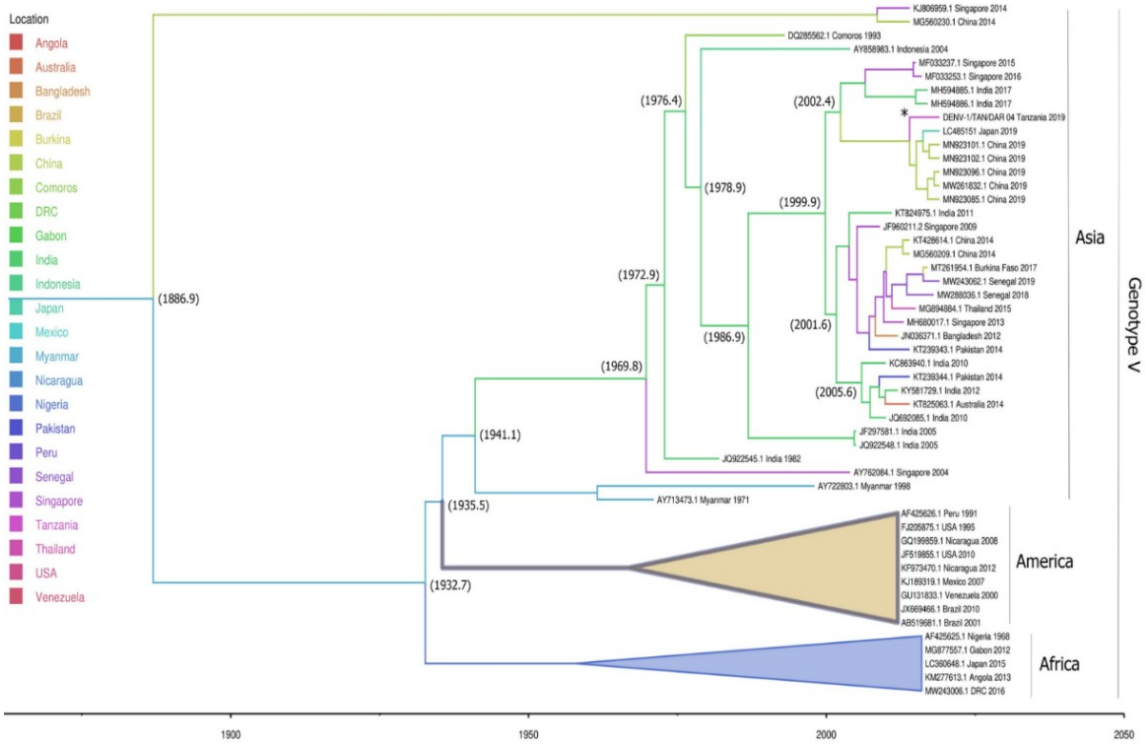


Figure 2.1.5: Bayesian evolutionary analysis of DENV-1 genotype V envelope gene.

The analysis involved E gene sequences from the Tanzanian DENV-1 strain and 50 reference strains available in the GenBank. The maximum clade credibility (MCC.) tree was reconstructed by maximum likelihood with GTR+G+I model. The estimated

years of ancestral divergence are indicated at the nodes and the Tanzanian DENV-1 strain from this study is marked with*.

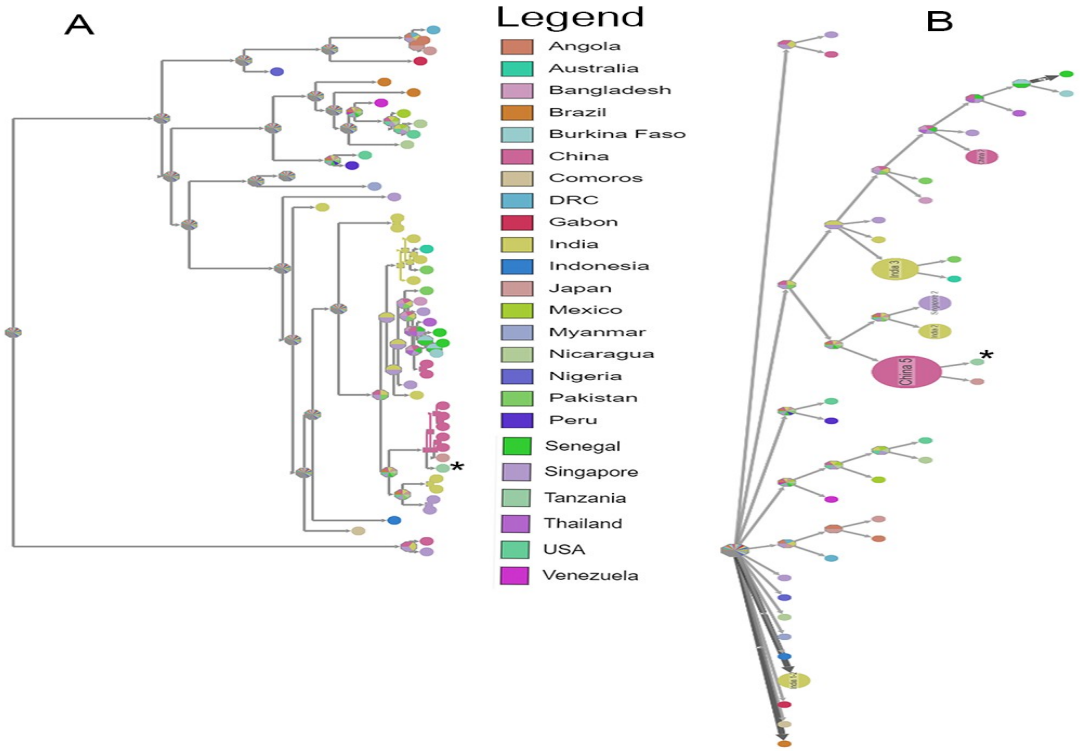


Figure 2.1.6: Phylogeographic analysis of DENV-1 genotype V strains based on complete envelope gene from the Tanzanian strain and reference strains reported from 23 different countries in Americas, Asia and Africa. The figure shows the full tree visualization (A) and compressed tree visualization (B) that was generated by PastML algorithm with maximum likelihood marginal posterior probabilities approximation and Felsenstein 1981 model. The Tanzanian DENV-1 strain is marked with*

Discussion

Over the past decade, Tanzania has reported five dengue outbreaks with the commercial coastal city of Dar es Salaam being the epicentre of these outbreaks. During this period, the country has experienced a shift in dengue virus serotypes from dengue virus 3 (DENV-3) in 2010 (Gautret *et al*., 2010) to DENV-2 in 2013 through 2014 (Vairo *et al*., 2016), DENV-3 in 2018 (Chipwaza *et al*., 2021) and DENV-1 in 2019 (Mwanyika *et al*., 2021b). In the current study for the first time we have determined the genome sequences of DENV-1 strains detected during the 2019 outbreak in Dar es Salaam with nanopore MinION technology. In addition, we established their evolutionary relationship using Bayesian Evolutionary Analysis by Sampling Trees program with Markov Chain Monte Carlo posterior probabilities and phylogeography in PastML program using maximum likelihood marginal posterior probabilities approximation.

Phylogenetic analysis results revealed that the Tanzanian DENV-1 strains clustered with genotype V strains reported in China and Japan in 2019 with 99.52─99.66% sequence homology. The DENV-1 genotype V is known to be endemic in several countries in Asia including, Singapore, Japan, India and China (Sun *et al*., 2016) while in the Americas, the genotype has been the most prevalent in the past 40 years (de Bruycker-Nogueira *et al*., 2016). Similarly, the occurrence of DENV-1 genotype V in Africa has been previously reported in several countries (Caron *et al*., 2013; Dieng *et al*., 2021). In 2019, DENV-1 serotype was reported for the first time in Tanzania and was likely to have been introduced from endemic countries in Asia (Mwanyika *et al*., 2021b). Globally, genotype V is the most prevalent DENV-1 genotype responsible for most DENV-1 infections associated with large-scale epidemics in Asia, Americas, Africa and South Pacific (Villabona-Arenas and Zanotto, 2013; de Bruycker-Nogueira *et al*., 2016). Furthermore, findings from several studies revealed that genotype replacements among DENV-1 strains are associated

with expansion of local dengue transmission and disease severity (Teoh *et al*., 2013; Inizan *et al*., 2021; Ma *et al*., 2021).

The Phylogeographic analysis of Tanzanian DENV-1 genotype V E gene and reference E sequence data reported from different countries between 1968 and 2019, this suggests that DENV-1 genotype V lineages originated from the most recent common ancestor approximately 132.9 years ago (1813.7─1926.9). These results concur with 1825─1925 estimates reported previously (Patil *et al.,* 2011). Our analysis revealed that the DENV-1 genotype V lineages showed an evolutionary rate of 5.15×10^{-4} [95% highest posterior density (HPD) interval (4.04×10^{-4}, 6.26×10^{-4})] substitution rates per site per year. The results are similar to 5.58×10^{-4} substitution rates of DENV-1 strains detected from Korean travelers (Hwang *et al*., 2021), 5.79×10^{-4} substitution rates of DENV-1 strains isolated during epidemics in Brazil between 1994 and 2011 (Carneiro *et al*., 2012) and 6.5×10^{-4} substitution rates among DENV-1 strains reported in India from 1962─2005 (Patil *et al*., 2011). These findings may suggest that DENV-1 genotype V lineages experience similar replication fidelity throughout the phylogeny with the possibility that external selection pressures could be largely responsible for shaping the evolution and adaptation of DENV-1 genotype V lineages. In agreement with these findings, a study by Koo et al. (2018) demonstrated that the sustained transmission of DENV-1 lineages is shaped by stochastic forces that are likely to be influenced by external forces.

Reconstruction of geographic spread inferred a single introduction of DENV-1 genotype V into Tanzania from China in 2019. This genotype was also introduced into Japan during the same year (Okada *et al*., 2019). Among DENV-1 genotypes, genotype V is the major imported genotype in China (Sun *et al*., 2016). This finding could be partly supported by

the results of this study that show DENV-1 genotype V lineages reported in China originated from multiple countries in South East Asia region. These observations highlight the possibility of intercontinental spread of DENV-1 genotype V lineages through multiple introductions from different geographical areas.

It is worth noting that inadequate availability of quality clinical samples and limited laboratory resources to optimize the protocol for full genome sequencing using nanopore MinION technology are likely to be the limitation of this study. With the limited laboratory resources, it was not possible to generate full genome sequences of DENV-1 strain using the current protocol. Despite these limitations, the findings of this study provide comprehensive and useful information to inform public health interventions of dengue virus infection in Tanzania.

Conclusions

This study provides evidence of DENV-1 genotype V occurrence in Tanzania that possibly was introduced through a single source in Asia. There is a possibility of future multiple introductions of DENV-1 serotypes into Tanzania that could lead to genotype/lineage replacements and increased risk of severe disease. The phylogeographic information of DENV-1 genotype V from this study can guide intervention strategies that could help to limit the possible introduction of new or multiple DENV-1 lineages from endemic countries into Tanzania. These findings emphasize the need to enhance genomic surveillance of dengue virus serotypes/genotypes in Tanzania through rapid genome sequencing using nanopore MinION technology.

Acknowledgements

This study was funded by the Government of the United Republic of Tanzania through the World Bank (WB-ACE II Grant PAD1436, IDA credit 5799-TZ] to the SACIDS Africa Centre of Excellence for Infectious Diseases at the Sokoine University of Agriculture. The authors acknowledge the contributions and laboratory technical assistance from Ms. Mariam Makange and Mr. Charles Kayuki at the SACIDS Molecular Biology Laboratory at Sokoine University of Agriculture in Morogoro, Tanzania.

References

Amarasinghe, A., Kuritsky, J. N., Letson, G. W. and Margolis, H. S. (2011). Dengue virus infection in Africa. *Emerging Infectious Diseases* 17(8): 1349–1354.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J. and Hay, S. I. (2013). The global distribution and burden of dengue. *Nature* 496(746): 504–507.

Carneiro, A. R., Cruz, A. C. R., Vallinoto, M., Melo, D. de V., Ramos, R. T. J., Medeiros, D. B. A., Silva, E. V. P. da and Vasconcelos, P. F. da C. (2012). Molecular characterisation of dengue virus type 1 reveals lineage replacement during circulation in Brazilian territory. *Memorias Do Instituto Oswaldo Cruz* 107(6): 805–812.

Caron, M., Grard, G., Paupy, C., Mombo, I. M., Nso, B. B. B., Kassa, F. R. K., Nkoghe, D.and Leroy, E. M. (2013). First evidence of simultaneous circulation of three different dengue virus serotypes in Africa. *PloS ONE* 8(10): e78030.

Chipwaza, B., Sumaye, R. D., Weisser, M., Gingo, W., Yeo, N. K.-W., Amrun, S. N., Okumu, F. O. and Ng, L. F. P. (2021). Occurrence of 4 dengue virus serotypes and Chikungunya Virus in Kilombero Valley, Tanzania, during the dengue outbreak in 2018. *Open Forum Infectious Diseases* 8(1): 1–6.

de Bruycker-Nogueira, F., Mir, D., dos Santos, F. B. and Bello, G. (2016). Evolutionary history and spatiotemporal dynamics of DENV-1 genotype V in the Americas. *Infection, Genetics and Evolution* 45: 454–460.

Dieng, I., Cunha, M. dos P., Diagne, M. M., Sembène, P. M., Zanotto, P. M. de A., Faye, O. and Sall, A. A. (2021). Origin and spread of the dengue virus type 1, genotype V in Senegal, 2015–2019. *Viruses* 13(57): 1–12.

Gautret, P., Simon, F., Hervius Askling, H., Bouchaud, O., Leparc-Goffart, I., Ninove, L., Parola, P. and Euro, T. N. (2010). Dengue type 3 virus infections in European travellers returning from the Comoros and Zanzibar, February-April 2010. Euro Surveillance: Bulletin European sur les maladies transmissible *European Communicable Disease Bulletin* 15(15): 1–3.

Gebhard, L. G., Filomatori, C. V. and Gamarnik, A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses* 3(9): 1739–1756.

Gurevich, A., Saveliev, V., Vyahhi, N.and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29(8): 1072–1075.

Hill, S. C., Neto de Vasconcelos, J., Granja, B. G., Thézé, J., Jandondo, D., Neto, Z., Mirandela, M., Sebastião, C. dos S., Cândido, A. L. M., Clemente, C., Pereira da Silva, S., de Oliveira, T., Pybus, O. G., Faria, N. R.and Afonso, J. M. (2019). Early genomic detection of cosmopolitan genotype of dengue virus serotype 2, Angola, 2018. *Emerging Infectious Diseases* 25(4): 784–787.

Holmes, E. C. and Twiddy, S. S. (2003). The origin, emergence and evolutionary genetics of dengue virus. Infection, Genetics and Evolution: *Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 3(1): 19–28.

Hwang, E.-H., Kim, G., Chung, H., Oh, H., Park, J.-H., Hur, G. H., Hong, J.and Koo, B.-S. (2021). Molecular evolution of dengue virus types 1 and 4 in Korean travelers. *Archives of Virology* 166(4): 1103–1112.

Inizan, C., Minier, M., Prot, M., O'Connor, O., Forfait, C., Laumond, S., Marois, I., Biron, A., Gourinat, A.-C., Goujart, M.-A., Descloux, E., Sakuntabhai, A., Tarantola, A., Simon-Lorière, E. and Dupont-Rouzeyrol, M. (2021). Viral evolution sustains a dengue outbreak of enhanced severity. *Emerging Microbes and Infections* 10(1): 536–544.

Ishikawa, S. A., Zhukova, A., Iwasaki, W. and Gascuel, O. (2019). A fast likelihood method to reconstruct and visualize ancestral scenarios. *Molecular Biology and Evolution* 36(9): 2069–2085.

Klungthong, C., Putnak, R., Mammen, M. P., Li, T. and Zhang, C. (2008). Molecular genotyping of dengue viruses by phylogenetic analysis of the sequences of individual genes. *Journal of Virological Methods* 154(1): 175–181.

Koo, C., Tien, W. P., Xu, H., Ong, J., Rajarethinam, J., Lai, Y. L., Ng, L.-C. and Hapuarachchi, H. C. (2018). Highly selective transmission success of dengue virus type 1 lineages in a dynamic virus population: an evolutionary and fitness perspective. *IScience* 6: 38–51.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27(5): 722–736.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6): 1547–1549.

Letunic, I. and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research* 47(1): 256–259.

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

Ma, M., Wu, S., He, Z., Yuan, L., Bai, Z., Jiang, L., Marshall, J., Lu, J., Yang, Z. and Jing, Q. (2021b). New genotype invasion of dengue virus serotype 1 drove massive outbreak in Guangzhou, China. *Parasites and Vectors* 14(1): 1–12.

Mboera, L. E. G., Mweya, C. N., Rumisha, S. F., Tungu, P. K., Stanley, G., Makange, M. R., Misinzo, G., De Nardo, P., Vairo, F. and Oriyo, N. M. (2016). The Risk of Dengue Virus Transmission in Dar es Salaam, Tanzania during an Epidemic Period of 2014. *PLoS Neglected Tropical Diseases* 10(1): 1–15.

Mwanyika, G. O., Mboera, L. E., Rugarabamu, S., Ngingo, B., Sindato, C., Lutwama, J. J., and Misinzo, G. (2021a). Dengue virus infection and associated risk factors in Africa: a systematic review and meta-analysis. *Viruses* 13(4): 536.

Mwanyika, G. O., Mboera, L. E., Rugarabamu, S., Makange, M., Sindato, C., Lutwama, J. J., Misinzo, G. (2021b). Circulation of dengue serotype 1 viruses during the 2019 outbreak in Dar es Salaam, Tanzania. *Pathogens and global health* 115(7-8): 467-475.

OhAinle, M., Balmaseda, A., Macalalad, A. R., Tellez, Y., Zody, M. C., Saborio, S., Nunez, A., Lennon, N. J., Birren, B. W., Gordon, A., Henn, M. R. and Harris, E. (2011). Dynamics of dengue disease severity determined by the interplay between viral genetics and serotype-specific immunity. *Science Translational Medicine* 3(114): 114–128.

Okada, K., Morita, R., Egawa, K., Hirai, Y., Kaida, A., Shirano, M., Kubo, H., Goto, T. and Yamamoto, S. P. (2019). Dengue Virus Type 1 Infection in Traveler Returning from Tanzania to Japan, 2019. *Emerging Infectious Diseases* 25(9):1782.

Patil, J. A., Cherian, S., Walimbe, A. M., Patil, B. R., Sathe, P. S., Shah, P. S. and Cecilia, D. (2011). Evolutionary dynamics of the American African genotype of dengue type 1 virus in India (1962–2005). *Infection Genetics and Evolution* 11(6): 1443–1448.

Rambaut, A., Lam, T. T., Max Carvalho, L. and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst. *Virus Evolution* 2(1): 007.

Rodriguez, F., Oliver, J. L., Marín, A. and Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* 142(4): 485–501.

Santiago, G. A., González, G. L., Cruz-López, F. and Muñoz-Jordan, J. L. (2019). Development of a standardized sanger-based method for partial sequencing and genotyping of dengue viruses. Journal of Clinical Microbiology 57(4): 1–11.

Sessions, O. M., Khan, K., Hou, Y. A., Meltzer, E., Quam, M., Schwartz, E. and Wilder-Smith, A. (2013). Exploring the origin and potential for spread of the 2013 dengue outbreak in Luanda, Angola. *Global health action* 6(1): 21822.

Sharma, S., Zapatero-Rodríguez, J., Estrela, P. and O'Kennedy, R. (2015). Point-of-care diagnostics in low resource settings: Present Status and Future Role of Microfluidics. Biosensors 5(3): 577–601.

Sović, I., Šikić, M., Wilm, A., Fenlon, S. N., Chen, S. and Nagarajan, N. (2016). Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nature Communications* 7(1): 11307.

Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., Hay, S. I., Bedi, N., Bensenor, I. M. and Castañeda-Orjuela, C. A. (2016).

The global burden of dengue: An analysis from the Global Burden of Disease Study 2013. *Lancet Infectious Diseases* 16(6): 712–723.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J. and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *Virus Evolution* 4(16): 1–5.

Sun, J., Wu, D., Zhou, H., Zhang, H., Guan, D., He, X., Cai, S., Ke, C. and Lin, J. (2016). The epidemiological characteristics and genetic diversity of dengue virus during the third largest historical outbreak of dengue in Guangdong, China, in 2014. *Journal of Infection* 72(1): 80–90.

Teoh, B. T., Sam, S. S., Tan, K. K., Johari, J., Shu, M. H., Danlami, M., Abd-Jamil, J., MatRahim, N., Mahadi, N. and AbuBakar, S. (2013). Dengue virus type 1 clade replacement in recurring homotypic outbreaks. *BMC Evolutionary Biology* 13(213): 1–10.

Vairo, F., Mboera, L. E. G., De Nardo, P., Oriyo, N. M., Meschi, S., Rumisha, S. F., Colavita, F., Mhina, A., Carletti, F., Mwakapeje, E., Capobianchi, M. R., Castilletti, C., Di Caro, A., Nicastri, E., Malecela, M. N. and Ippolito, G. (2016). Clinical, virologic, and epidemiologic characteristics of dengue outbreak, Dar es Salaam, Tanzania. *Emerging Infectious Diseases* 22(5): 895–899.

Villabona-Arenas, C. J. and Zanotto, P. M. de A. (2013). Worldwide spread of dengue virus type 1. PloS ONE 8(5): e62649.Weaver, S. C. and Vasilakis, N. (2009). Molecular evolution of dengue viruses: Contributions of phylogenetics to

understanding the history and epidemiology of the preeminent arboviral disease. *Infection, Genetics and Evolution* 9(4): 523–540.

Wu, J. Y., Lun, Z. R., James, A. A. and Chen, X. G. (2010). Dengue Fever in Mainland China. *American Journal of Tropical Medicine and Hygiene* 83(3): 664–671.

Yu, J., Li, X., He, X., Liu, X., Zhong, Z., Xie, Q., Zhu, L., Jia, F., Mao, Y., Chen, Z., Wen, Y., Ma, D., Yu, L., Zhang, B., Zhao, W. and Xiao, W. (2019). Epidemiological and Evolutionary Analysis of Dengue-1 Virus Detected in Guangdong during 2014: Recycling of Old and Formation of New Lineages. *American Journal of Tropical Medicine and Hygiene* 101(4): 870–883.

Manuscript Five

In silico epitope prediction and expression of DENV-2 envelope protein domain III in *Escherichia coli*: A target design approach for nanobody binding

Gaspary O. Mwanyika[1,2,3], Edgar Kigozi[4], Leonard E.G. Mboera[1], Julius J. Lutwama[5] Janusz T. Paweska[6], and Gerald Misinzo[1,2]

[1]SACIDS Africa Centre of Excellence for Infectious Diseases, Sokoine University of Agriculture, Morogoro, Tanzania

[2]Department of Veterinary Microbiology, Parasitology and Biotechnology, Sokoine University of Agriculture, Morogoro, Tanzania

[3]Department of Medical Sciences and Technology, Mbeya University of Science and Technology, Mbeya, Tanzania

[4]Department of Medical Microbiology, Makerere University College of Health Sciences, Kampala, Uganda

[5]Department of Arbovirology, Emerging and Re-emerging Infectious Diseases, Uganda Virus Research Institute, Entebbe, Uganda

[6]National Institute for Communicable Diseases, National Health Laboratory Service, Sandringham, Johannesburg, South Africa

Manuscript in preparation

Abstract

Background: Dengue is a major mosquito-borne viral infection in tropical and subtropical regions. Routine serological tests exhibit poor specificity due to cross-reactivity of the dengue viruses (DENV). Envelope protein domain III (EDIII) of DENV harbour many conserved subcomplex virus and serotype-specific epitopes. The objective of this study was to identify antigenic epitopes of DENV-2 EDIII protein using bioinformatics approaches and optimize its expression in *Escherichia coli* BL21 (DE3) cells.

Methods: Three bioinformatics tools, namely ABCpred, Bepipred and CBTOPE were used to predict continuous and conformational B-cell epitopes. The Hydrophilicity and antigenicity of selected epitopes were evaluated using Parker hydrophilicity model and VaxiJen v2.0 program, respectively. Surface accessibility and secondary structure were evaluated with NetsurfP v2.0 (Threshold 25% exposure). The EDIII was cloned on pET-22b (+) vector and transformed into competent *E. coli* BL21 (DE3) grown on Luria-Bertani agar supplemented with 100µg/mL of ampicillin. The expression was induced with 1 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) at 30ºC and analysed on 12% SDS-PAGE.

Results: Five continuous epitopes were identified at 305─320, 334─348, 320─335, 378─392, 389─400 and three conformational epitopes at 310─312, 336─337 and 341─342 amino acid residues of envelope protein of DENV-2. High-level expression of EDIII protein was observed four hours after induction with 1 mM of IPTG at 30 ºC.

Conclusion: Overall, the findings of this study show that EDIII protein of DENV-2 harbor antigenic epitopes that can be expressed in *Escherichia coli* BL21 (ED3) as potential targets for development of nanobodies.

Keywords: dengue virus; serotype 2; envelope; domain III; epitopes; expression; *E. coli*

Introduction

Dengue virus (DENV) is an important mosquito-borne infection in tropical and subtropical regions. DENV is a single-stranded RNA virus of family *Flaviviridae* that is transmitted to humans by infected Aedes mosquitoes. The virus is endemic in more than 125 countries causing approximately 340 million infections and 20,000 deaths each year (Bhatt *et al*., 2013). The current prevention and control strategies mainly rely on early detection and vector control (Stanaway *et al*., 2016). DENV exists in four serotypes (DENV1─4) that share 60% amino acid homology and each having the ability to cause a distinct infection ranging from mild fever to severe disease (Gubler, 1997). Globally, DENV-2 is frequently associated with more severe infections and epidemics compared with other serotypes (Messina *et al*., 2014). In Africa, all the four DENV serotypes have been reported with frequent epidemics caused by DENV-1 and DENV-2 serotypes (Amarasinghe *et al*., 2011).

DENV RNA genome encodes three structural proteins: Capsid (C), precursor membrane (prM) and envelope (E) and seven non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Gebhard *et al*., 2011). The E protein contains 495 amino acid residues and is the major structural protein involved in host cell binding and fusion that make it good target for diagnostic and vaccine development (Modis *et al*., 2004). E protein has three structural domains (EDI, EDII, and EDIII). The immunoglobulin-like domain III (EDIII) which span amino acids 300 to 400 of E protein C-terminus is the center of host receptor binding and induction of strong neutralizing antibodies (Tripathi *et al*., 2011). The domain contains many conserved subcomplex virus and serotype-specific epitopes

(Guzman *et al*., 2010; Hapugoda *et al*., 2007), thus the EDIII could be utilized as useful immunodiagnostic reagents.

B-cell epitopes are clusters of accessible amino acid residues in the protein which can be recognized by specific antibodies. On the basis of spatial structure, the B-cell epitopes are grouped into two main classes; continuous or linear epitopes and discontinuous or conformational epitopes (Regenmortel, 2009). The continuous epitopes are sequential peptide fragments of a protein capable of binding to antibodies while discontinuous epitopes are form of a cluster of amino acid residues in the protein that are brought together by folding of the polypeptide chain and their antigenic reactivity relies on the native conformation of the protein (Potocnakova *et al*., 2016).

Identification of these epitopes is a crucial step for designing diagnostic antibodies. There are several methods of epitope mapping that include X-ray crystallography that defines the structure of the epitopes by measuring the strength of antigen-antibody complex binding (Caliandro *et al*., 2013); nuclear magnetic resonance (NMR) that maps the epitopes based on the binding energy of antigen-antibody complex (Kwan *et al*., 2011); site-directed mutagenesis that maps epitopes based on substitution of individual amino acid residues that forms a functional epitope leading to loss of antibody binding (Benjamin and Perdue, 1996); and phage display technique based on testing the binding affinity of a variety of peptides displayed on the surface of phages or other display platforms to the monoclonal antibodies of interest through biopanning method (Huang *et al*., 2011). Since, the majority of these methods are expensive, labor-intensive and time consuming there are several computational methods for prediction of B-cell epitopes that utilizes trained neural network and support vector algorithms to predict epitopes based on

either protein sequence or structural data (Saha and Raghava, 2007; Ferdous *et al*., 2019).

Several authors have investigated the potential of EDIII protein in serodiagnosis of DENV infection and revealed that EDIII can be an attractive alternative diagnostic biomarker (Gaylord, *et al*., 2015; Nguyen *et al*., 2019). In recent years, recombinant EDIII protein has been expressed in *Escherichia coli* that normally result into insoluble products and/or low yield. The purification of insoluble proteins is cumbersome because the insoluble aggregates known as inclusion bodies must be purified under denaturation conditions followed by refolding (Jaiswal *et al*., 2004). The denaturation process often lead to improperly folded protein and loss of bioactivity (Jaiswal *et al*., 2004; San-Miguel *et al*., 2013). The objective of this study was to identify antigenic epitopes on envelope protein domain III of DENV-2 using bioinformatics approaches and optimize its expression in *E. coli* BL21 (ED3) cells.

Methods

Creation of the consensus envelope protein domain III (EDIII)

Amino acid sequences of EDIII protein representing dengue virus serotype 2 were retrieved from the National Centre for Biotechnology Information protein database (Table 2.2.1). The consensus sequence was created using Bioedit software v.7.2 and the conservation confirmed in sequence manipulation suit tool (https://www.bioinformatics.org/sms2/ color_align_cons.html). The physicochemical properties of target EDIII protein were predicted using ProtParam program implemented in ExPASy Bioinformatics Resource Portal (https://web.expasy.org/protparam/).

Table 2.2.1: The representative dengue virus serotype 2 strains used in this study

Strain/isolate	Accession No
16681-PDK53	P2999.1
China/D2-04	P30026.1
Jamaica/1409/1983	P07564.2
Malaysia M2	P14338.1
Peru/IQT2912/1996	Q9WDA6.1
Thailand/16681/84	P29990.1

Prediction of linear and conformational B-cell epitopes

The following aligorithms ABCpred (Threshold 0.7), and Bepipred Linear Epitope Prediction-2.0 (Threshold 0.7) were used to predict linear B-cell epitopes. The ABCpred uses scores acquired from the trained recurrent neural network of virus epitope data to predict the epitopes (Potocnakova *et al*., 2016) while BepiPred-2.0 predicts epitopes from a random forest algorithm derived from antibody-antigen constructions (Jespersen *et al*., 2017). On the other hand, CBTOPE (Threshold 0.7) was used to predict conformational B-cell epitopes. The tool utilizes amino acids composition as input feature for support vector machine with a probability scale of 1─9 and accuracy of > 85% (https://webs.iiitd. edu.in/ raghava/cbtope/index.php). The score above 4 was considered as an epitope residue (Ansari *et al*., 2010).

Evaluation of predicted epitopes

The antigenic propensity was evaluated with VaxiJen v2.0 (Threshold 0.4) (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). The higher the score of the antigenic determinant, the higher the likelihood to induce immune response (Doytchinova *et al*., 2007). Hydrophilic regions were predicted using Parker hydrophilicity model implemented in Immune Epitope Database (http://tools.immuneepitope.org/bcell/result/) using default values. The surface accessibility and secondary structures were predicted using NetsurfP-2.0 (threshold 25% exposure). To enhance the expression efficiency, the presence of codons that are rarely used by *E. coli* was analysed with rare codon analysis tool (https://www.genscript.com/tools/rare-codon-analysis).

Bacterial strains and plasmids

The BL21 (DE3) cells (Novagen, USA), the *E. coli* B strain which lack Ion and ompT proteases were used for expression of recombinant EDIII protein. The strain is lysogenic to lambda phage DE3 (λ) and thus contains a copy of the T7 RNA polymerase gene under the control of T7 promoter that is inducible with isopropyl β-D-1 thiogalactopyranoside (IPTG) inducer. The construct pET-22b (+) (Novagen, USA) containing target D2EDIII insert in the open reading frame was used as an expression vector. The vector is 5,802 base pairs long containing T7/lac promoter, pelB leader signal sequence at N-terminus to direct the secreted target proteins into the periplasm of *E. coli*, and 6x histidine tag at C-terminus for purification of expressed protein using metal affinity columns and ampicillin resistance genes as the selectable makers (Figure 2.2.2).

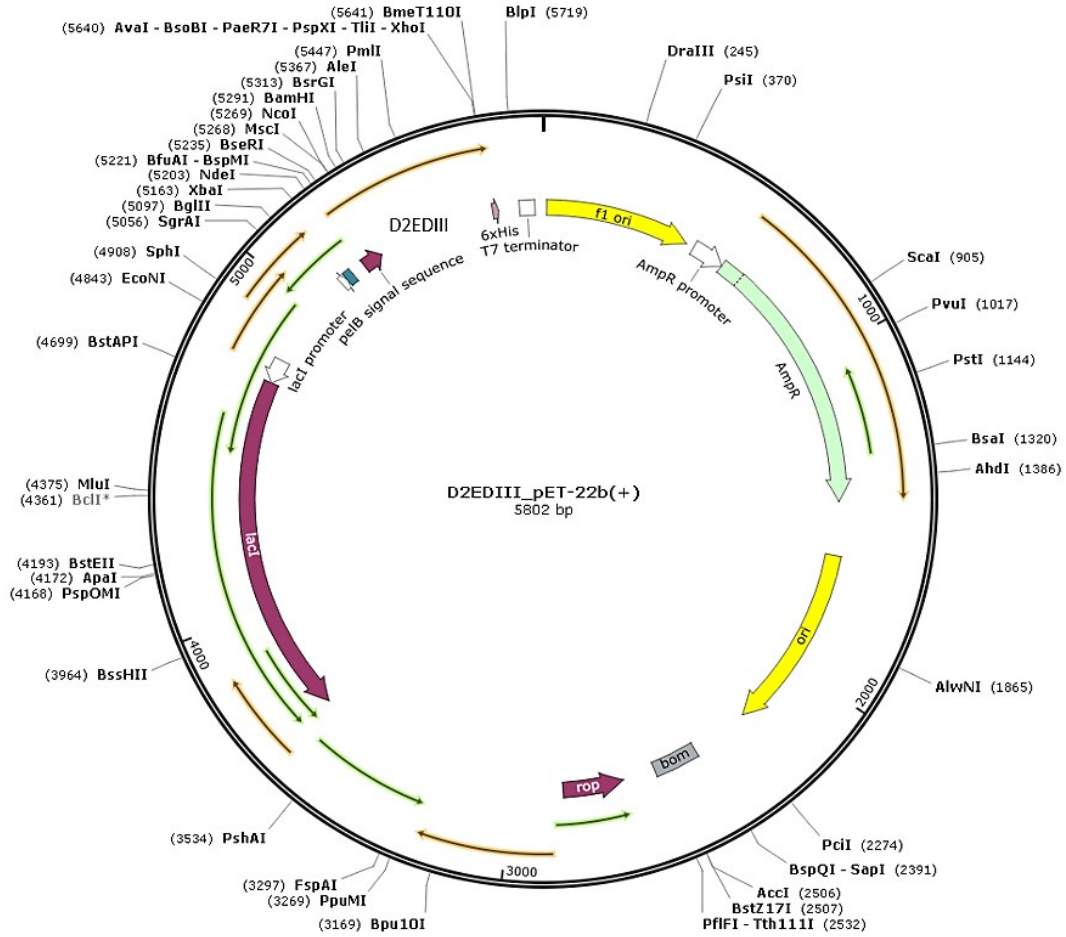


Figure 2.2.2: A schematic map of D2EDIII_pET22b (+) expression construct
(Source: Authors' own)

Culture and induction conditions

The competent cells were prepared using Calcium chloride protocol as previously described by Sambrook and Russell (2006). The D2EDIII_pET22b plasmid was transformed into competent *E. coli* BL21(DE3) and a single colony containing the expression plasmid was selected and inoculated into 50 mL of Luria-Bertani (LB) medium supplemented with 100µg/mL of ampicillin (Sigma, USA) and 0.5% glucose and grown overnight at 37ºC.

The overnight saturated starter culture was resuspended into 200 mL fresh LB medium supplemented with 100µg/mL of ampicillin and incubated at 37ºC with shaking at 200 rpm until mid-logarithmic phase when OD_{600} is approximately 0.6 usually after 1.5─2 hours. After then, the culture was cooled to 30ºC and 1 mL was collected as pre-induced culture. The remaining portion was induced with 1 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) and monitored every 2 hours up to24 hrs while shaking at 225 rpm. The expression was repeated with induction at normal bacterial growth temperature of 37ºC to investigate the effect on expression level.

SDS-PAGE analysis

The cells were harvested by centrifugation at 5,000 x g for 10 minutes and the supernatant discarded and the remaining cell pellet was resuspended in lysis buffer (4x Laemi buffer). A 12% Sodium Dodecyl Polyacrylamide Gel (SDS-PAGE) was prepared and placed into an electrophoresis chamber containing 1x Running buffer (25 Mm Tris base, 192 Mm Glycine and 0.1% SDS, pH=8.3). Thereafter, a 10µL of sample was mixed with 4x Laemmli sample buffer (Bio-Rad) containing 100 µL of 2-mercaptoethanol per 900 µL of sample buffer. The protein samples were loaded on the pre-casted wells along with a pre-stained protein ladder and ran for 20 minutes at 75 V for stacking followed by 70 minutes at 150 V for resolution. Thereafter the gel was rinsed two times with double distilled water for 5 minutes and treated with a fixation solution containing 40% methanol and 10% acetic acid for 5 minutes. The fixed gel was treated with 0.025% (w/v) Coomassie Blue G-250 stain containing 10% acetic acid for approximately 20 minutes while shaking on Orbital Platform shaker at 100 rpm followed by detaining with fixation solution. The gel was imaged on Bio-Rad ChemiDoc Imager to detect the band size (14 kDa) of target proteins.

Results

D2EDIII sequence and theoretical physicochemical properties

The consensus envelope protein domain III containing 114 amino acid residues was created from the amino acid sequences obtained from NCBI database representing the DENV-2 strains (Figure 2.2.3) and the predicted physical and chemical properties of D2EDIII are shown in Table 2.2.4.

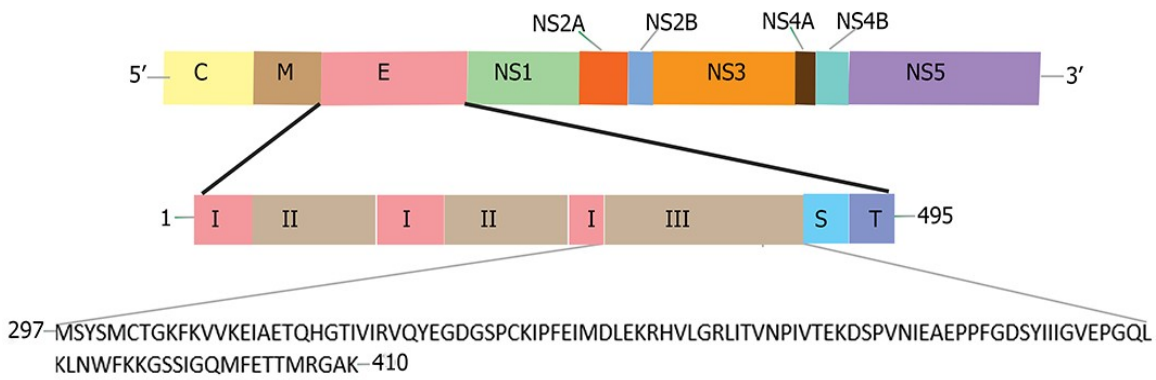


Figure 2.2.3: DENV-2 envelope protein domain III span between 297─410aa

(Source: Authors' own)

Table 2.2.4: Theoretical physical chemical properties of D2EDIII protein

Parameter	Value
Molecular weight (MW)	13.6 kDa
Theoretical isoelectric point (pI)	7.1
Instability index*	37.2
Gravy**	-0.35

*Protein with instability index < 40 is considered stable; **Gravy index indicates the solubility of proteins +gravy refers to hydrophobicity and -gravy refer to hydrophilicity.

Predicted D2EDIII epitopes

Two linear epitopes were identified by ABCpred at 320─335 and 378─392 residues and three epitopes at 305─320, 334─348 and 389─400 with Bepipred linear Epitope Prediction-2.0. Two conformational epitopes were predicted by CBTOPE at 310-312 and 336-337 amino acid residues of E protein. The hydrophilic regions and antigenicity profiles of the predicted epitopes are shown in Table 2.5 and predicted surface accessibility and secondary structure are shown in Figure 2.2.4.

Table 2.2.5: The distribution of linear and conformation B-cell epitopes and their hydrophilicity and antigenicity profiles

ABC pred	**Linear epitopes**
N320─335	MSYSMCTGKFKVVKEIAETQ**HGTIVIRVQYEGDGSP**CKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Hydrophilicity	MSYSMCTGKFKVVKEIAETQHGTIVIRV**QYEGD**GSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Antigenicity	0.9363
N378─392	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFG**DSYIIIGVEPGQLKL**NWFKKGSSIGQMFETTMRGAK
Hydrophilicity	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGV**EPGQLKL**NWFKKGSSIGQMFETTMRGAK
Antigenicity	1.0727
Bepipred-2.0	
N305─320	MSYSM**CTGKFKVVKEIAETQH**GTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Hydrophilicity	MSYSMCTGKFKVVK**EIAETQH**GTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Antigenicity	0.1051
N334─348	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGS**PCKIPFEIMDLEKR**HVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Hydrophilicity	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPF**EIMDLEK**RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
Antigenicity	1.0142
N389─400	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPG**QLKLNWFKKGSS**IGQMFETTMRGAK
Hydrophilicity	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKL**NWFKKGSS**IGQMFETTMRGAK
	0.6801
CBTOPE	**Conformational epitopes**
N310-312	MSYSMCTGKF**KVV**KEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
N336-337	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGS**PC**KIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK
N341-342	MSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIP**FE**IMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAK

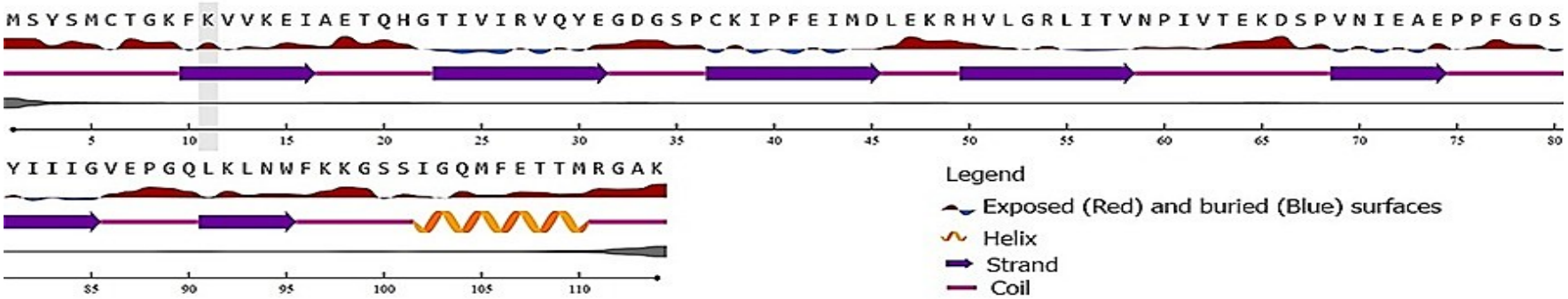


Figure 2.2.4: The predicted surface accessibility and secondary structure of D2EDIII protein

(Source: Authors' own)

Expression of recombinant D2EDIII protein

The transformed *E. coli* BL21 (DE3) cells were selected on LB agar plates containing 100µg/mL of ampicillin (Figure 2.2.5). Recombinant D2EDIII protein was optimally expressed using 200 mL of bacteria culture four hours after induction with 1 mM IPTG at 30ºC with incubator shaking at 225 rpm. Identification of expression by SDS-PAGE is shown in Figure 5. To investigate the effect of temperature on the expression level, results showed that higher *E. coli* culture optical density (OD) was observed at 30ºC compared to the common bacterial growth temperature of 37ºC (Figure 2.2.6).

Figure 2.2.5: Selection of transformants on LB-ampicillin plates; A: Untransformed *E. coli* cells did not grow on LB-ampicillin agar (-ve); B: Transformed *E. coli* cells contained ampicillin resistant markers grew on LB agar (+ve).

(Source: Authors' own)

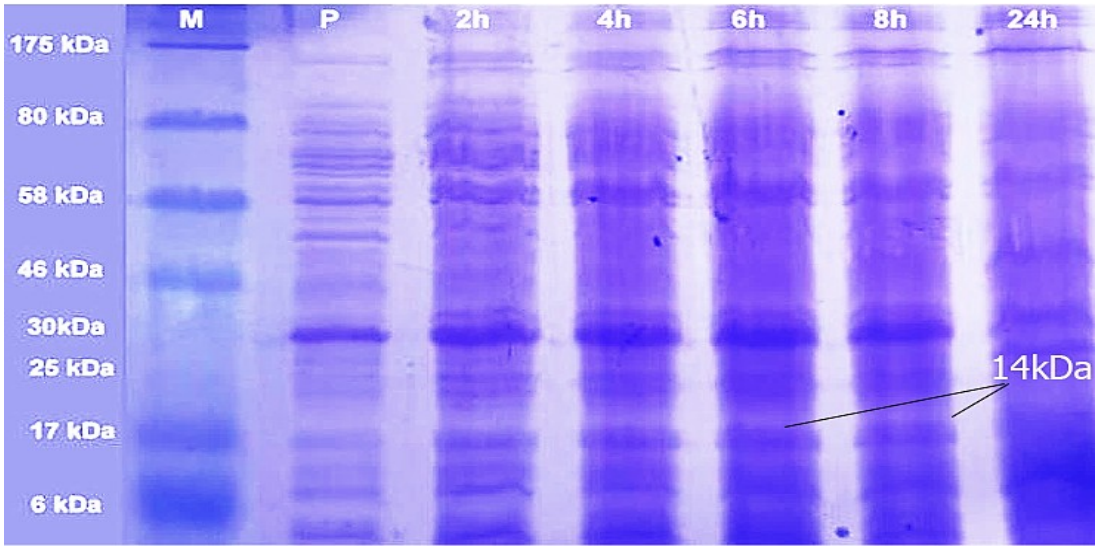


Figure 2.2.6: SDS-PAGE analysis of recombinant envelope protein domain III of DENV-2 (D2EDII). M: broad range protein ladder, P, pre-induced culture, 2 h, 4 h, 6 h, 8 h and 24 h post-induction cultures at 30°C.

(Source: Authors' own)

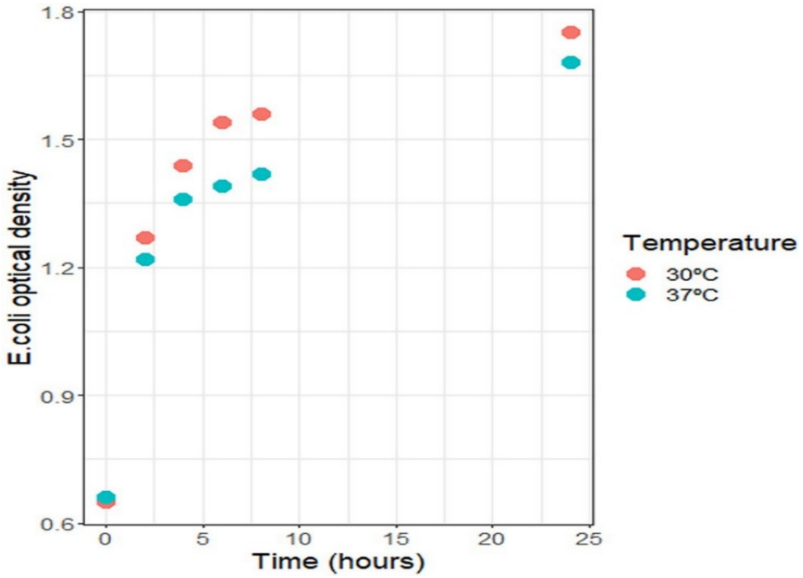


Figure 2.2.7: Effect of induction temperature on expression level of recombinant envelope protein domain III of DENV-2 (D2EDIII) after IPTG induction at 30°C and 37°C temperatures

(Source: Authors' own)

Discussion

The prediction analysis of D2EDIII protein showed that there were five linear epitopes; “HGTIVIRVQYEGDGSP” at 320─335, “DSYIIIGVEPGQLKLN” at 378─394, “CTGKFKVVKEIAETQH” at 305─320, “SPCKIPFEIMDLEKR” at 334─348 and “QLKLNWFKKGSS” at 389─400 and two conformational epitopes were identified “KVV” at 310─312, “PC” at 336─337 and “FE” at 341─342 positions and they formed part of continuous epitopes “CTGKFKVVKEIAETQH” and “SPCKIPFEIMDLEKR” respectively.

In comparison, the epitopes “EGDGSP” and “KGSSIGQMFETTMRGAK” that were previously identified by Nguyen et al. (2019) agree with the current predicted linear epitopes “HGTIVIRVQYEGDGSP” at 320─335 and “QLKLNWFKKGSS” at 389─400. In contrast, the conformation epitopes predicted in this study are different from previous report by Nguyen et al. (2019) who identified “LEK” at 345–374, and “EKD” at 363–365 positions. This difference could be explained by the fact that in present study, we used CBTOPE algorithm that utilizes amino acid composition and properties as input features to predict the epitopes on the primary sequence of the protein compared with DiscoTope algorithm used in the previous study that uses a 3D structure of protein available in the protein database.

Secondary structure analysis showed that the conformational epitopes; “KVV”, “PC” and “FE” are located on buried surfaces of envelope protein compared to “LEK” and “EKD” that are located on the exposed surface suggesting that “KVV”, “PC” and “FE” epitopes

could be fairly conserved antigenic determinants and tend to vary less than outer surface epitopes and thus they could be suitable diagnostic reagents (Regenmortel *et al.,* 2009). The single-domain antibodies (nanobodies) exhibit the ability to penetrate protein cavities and recognize hidden epitopes that normally cannot be accessed by classical antibodies (Baral *et al.,* 2013; Muyldermans, 2013). These findings show that the identified conformational epitopes could be useful targets for single-domain antibody binding.

The expression of envelope protein domain III of DENV-2 has been previously documented (Libraty *et al*., 2015; Nguyen *et al*., 2019). Nevertheless, its production in *E. coli* has not been easy due to its origin from a membrane protein. The membrane proteins are difficult to express and can be toxic to in *E. coli* cells. The expression of membrane proteins normally results into aggregated products known as inclusion bodies (Schlegel *et al*., 2010; Klepsch *et al*., 2011). In attempt to improve the expression of soluble and biologically active recombinant membrane proteins, several strategies have been adapted namely, co-expression with solubility tags like maltose-binding protein (MBP) and thioredoxin (TrxA) (di Guana *et al*., 1988; LaVallie *et al*., 1993), targeting their expression into bacterial periplasm and inducing the expression at low temperatures below the normal bacterial growth condition (Chin *et al*., 2007; Yang *et al*., 2012). However, the use of solubility fusion tags may not be beneficial due to the steric hindrance interference which can destruct the structure and biological activity of target proteins, in addition the removal of fusion tags is relatively expensive and requires tedious cleavage optimization conditions (Bell *et al*., 2013).

In this study, we used combined approaches to enhance the expression of envelope protein domain III through improving the physicochemical properties of the target, optimizing the

codons that are infrequently used by *E. coli* to enhance the expression efficiency. The presence of rare codons in *E. coli* tRNA may halt transcription and stop translation (Kane, 1995; Samatova *et al*., 2021). Targeting the expression into the *E. coli* periplasm allows the correct disulphide bond formation of the proteins for attaining its native conformation, solubility and biological activity (Malik, 2016). The proteins containing cystine bridges requires oxidation environment present in the periplasm for disulfide bond formation. The pET22b (+) vector used in the present study carries pectate lyase B (pelB) signal sequence from *Erwinia carotovora* for directing the expressed proteins into periplasm (Choi and Lee, 2004). In comparison, the expression of protein that requires cysteine bridges for stability into cytoplasm could lead improper folding and aggregation of expressed proteins because the cytoplasm is a reducing environment that does not support the formation of disulfide bonds (Makrides, 1996; Klint *et al*., 2013).

The results of this study indicate that high level expression of envelope protein domain III of DENV-2 in *E. coli* BL21 (DE3) cells was observed four hours after IPTG induction at a reduced temperature of 30ºC. Low temperatures help to slow down the rate of protein synthesis and allow proper folding of the target protein. These observations suggest that it is possible to improve the expression of recombinant of EDIII protein in *E. coli* by improving the physicochemical properties of the target protein and culture conditions.

In this study there were several limitations. It was not possible to measure the concentration of recombinant proteins and establish immune assay to confirm their biological activity and specificity due to unavailability of the required laboratory reagents. Despite the limitations the findings of this study highlight benefits of combined approaches to improve the expression of difficult protein in *Escherichia coli*.

Conclusions

The findings of this study show that EDIII protein of DENV-2 contains antigenic epitopes that can be expressed in *Escherichia coli* BL21 (ED3) cells as potential targets for development of nanobodies targeting dengue viruses.

Acknowledgments

This study was funded by the Government of the United Republic of Tanzania through the World Bank (WB-ACE II Grant PAD1436, IDA credit 5799-TZ] to the SACIDS Africa Centre of Excellence for Infectious Diseases at the Sokoine University of Agriculture. The authors thank the African-German Network of Excellence in Science (AGNES) for a small grant award sponsored by German Federal Ministry of Education and Research and supported by the Alexander von Humboldt Foundation that supported part of this work. We appreciate Prof Douglas Scott and Dr Jeannette Whitmire of Uniformed Services University (USU), Maryland, United States of America who kindly donated *Escherichia coli* BL21 (DE3) cells used in this study.

References

Amarasinghe, A., Kuritsky, J. N., Letson, G. W. and Margolis, H. S. (2011). Dengue virus infection in Africa. *Emerging Infectious Diseases* 17(8): 1349–1354.

Baral, T. N., MacKenzie, R., and Ghahroudi, M. A. (2013). Single-domain antibodies and their utility. *Current Protocols in Immunology* 103(1): 17–57.

Bell, M. R., Engleka, M. J., Malik, A. and Strickler, J. E. (2013). To fuse or not to fuse: What is your purpose? *Protein Science* 22(11): 1466–1477.

Benjamin, D. C. and Perdue, S. S. (1996). Site-directed mutagenesis in epitope mapping. *Methods* 9(3): 508–515.

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

Caliandro, R., Belviso, D. B., Aresta, B. M., de Candia, M., and Altomare, C. D. (2013). Protein crystallography and fragment-based drug design. *Future Medicinal Chemistry* 5(10): 1121–1140.

Chin, J. F. L., Chu, J. J. H. and Ng, M. L. (2007). The envelope glycoprotein domain III of dengue virus serotypes 1 and 2 inhibit virus entry. *Microbes and Infection* 9(1): 1–6.

Choi, J. H. and Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Applied Microbiology and Biotechnology* 64(5): 625–635.

di Guana, C., Lib, P., Riggsa, P. D. and Inouyeb, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* 67(1): 21–30.

Ferdous, S., Kelm, S., Baker, T. S., Shi, J. and Martin, A. C. R. (2019). B-cell epitopes: Discontinuity and conformational analysis. *Molecular Immunology* 114: 643–650.

Gaylord, S. T., Abdul-Aziz, S. and Walt, D. R. (2015). Single-molecule arrays for ultrasensitive detection of host immune response to dengue virus infection. *Journal of Clinical Microbiology* 53(5): 1722–1724.

Gebhard, L. G., Filomatori, C. V. and Gamarnik, A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses* 3(9): 1739–1756.

Gubler, D. J. (1997). Epidemic dengue/dengue haemorrhagic fever: A global public health problem in the 21st century. *Dengue Bulletin* 21: 1–14.

Guzman, M. G., Jaenisch, T., Gaczkowski, R., Ty Hang, V. T., Sekaran, S. D., Kroeger, A., Vazquez, S., Ruiz, D., Martinez, E. and Mercado, J. C. (2010). Multi-country evaluation of the sensitivity and specificity of two commercially-available NS1

ELISA assays for dengue diagnosis. *PLoS Neglected Tropical Diseases* 4(8): e811.

Hapugoda, M. D., Batra, G., Abeyewickreme, W., Swaminathan, S. and Khanna, N. (2007). Single antigen detects both immunoglobulin M (IgM) and IgG antibodies elicited by all four dengue virus serotypes. *Clinical and Vaccine Immunology* 14(11): 1505–1514.

Huang, J., Ru, B. and Dai, P. (2011). Bioinformatics resources and tools for phage display. *Molecules* 16(1): 694–709.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* 6(5): 494–500.

Klepsch, M. M., Persson, J. O. and De Gier, J.-W. L. (2011). Consequences of the overexpression of a eukaryotic membrane protein, the human KDEL receptor, in Escherichia coli. *Journal of Molecular Biology* 407(4): 532–542.

Klint, J. K., Senff, S., Saez, N. J., Seshadri, R., Lau, H. Y., Bende, N. S., Undheim, E. A. B., Rash, L. D., Mobli, M. and King, G. F. (2013). Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the Periplasm of E. coli. *PLoS ONE* 8(5): e63865.

Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. and Mackay, J. P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist: Macromolecular NMR for the non-spectroscopists I. *FEBS Journal* 278(5): 687–703.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F.and McCoy, J. M. (1993). A Thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Bio/Technology* 11(2): 187–193.

Libraty, D. H., Zhang, L., Obcena, A., Brion, J. D. and Capeding, R. Z. (2015). Anti-dengue virus envelope protein domain III IgG ELISA among infants with primary dengue virus infections. *Acta Tropica* 142: 103–107.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews* 60(3): 512–538.

Malik, A. (2016). Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of *E. coli*. *Biotech* 6(1): 1–44.

Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E. and Battle, K. E. (2014). Global spread of dengue virus types: Mapping the 70-year history. *Trends in Microbiology* 22(3): 138–146.

Modis, Y., Ogata, S., Clements, D. and Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427(6972): 313–319.

Muyldermans, S. (2013). Nanobodies: Natural single-domain antibodies. *Annual Review of Biochemistry* 82(1): 775–797.

Nguyen, N. M., Duong, B. T., Azam, M., Phuong, T. T., Park, H., Thuy, P. T. B. and Yeo, S.J. (2019). Diagnostic performance of dengue virus envelope domain III in acute dengue infection. *International Journal of Molecular Sciences* 20(3424): 1–16.

Potocnakova, L., Bhide, M. and Pulzova, L. B. (2016). An Introduction to B-Cell epitope mapping and in silico epitope prediction. *Journal of Immunology Research* 2016: 1–11.

Regenmortel, M. H. V. (2009). What Is a B-Cell Epitope? *Epitope Mapping Protocols* 524: 3–20.

Saha, S. and Raghava, G. P. (2007). Prediction Methods for B-Cell Epitopes. *Methods Molecular Biology* 409: 387–394.

Samatova, E., Daberger, J., Liutkute, M. and Rodnina, M. V. (2021). Translational control by ribosome pausing in Bacteria: How a non-uniform pace of translation affects protein production and folding. *Frontiers in Microbiology* 11: 619–430.

Sambrook, J. and Russell, D. W. (2006). Preparation and transformation of competent E. coli using calcium chloride. *Cold Spring Harbor Protocols* 2006(1): 3932.

San-Miguel, T., Pérez-Bermúdez, P. and Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *Springer Plus* 2(89): 1–4.

Schlegel, S., Klepsch, M., Gialama, D., Wickström, D., Slotboom, D. J. and De Gier, J.W. (2010). Revolutionizing membrane protein overexpression in bacteria. *Microbial Biotechnology* 3(4): 403–411.

Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., Hay, S. I., Bedi, N., Bensenor, I. M. and Castañeda-Orjuela, C. A. (2016). The global burden of dengue: An analysis from the global burden of disease study 2013. *Lancet Infectious Diseases* 16(6): 712–723.

Tripathi, N. K., Shrivastava, A., Biswal, K. C. and Rao, P. V. L. (2011). Recombinant dengue virus type 3 envelope domain III protein from *Escherichia coli*. *Biotechnology Journal* 6(5): 604–608.

Yang, J., Zhang, J., Chen, W., Hu, Z., Zhu, J., Fang, X., Yuan, W., Li, M., Hu, X., Tan, Y., Hu, F. and Rao, X. (2012). Eliciting cross-neutralizing antibodies in mice challenged with a dengue virus envelope domain III expressed in Escherichia coli. *Canadian Journal of Microbiology* 58(4): 369–380.

CHAPTER THREE

3.0 GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

3.1 General Discussion

Generally, the results from a systematic review and meta-analysis showed that the overall prevalence of dengue virus (DENV) in Africa was 14%. The findings of a recent study conducted countrywide in Tanzania showed a similar (16.1%) seroprevalence of immunoglobulin G (IgG) antibodies. The findings from this study highlights the presence of DENV among human population in almost all ecological zones suggesting a wide circulation of the viruses. The individuals are unlikely to be diagnosed through routine health service system in Tanzania. Previous studies in Tanzania have reported higher prevalence of the virus among febrile patients in Kilosa, Kinondoni and Ilala (Vairo *et al*., 2016; Chipwaza *et al*., 2020) and relatively lower in Temeke, Moshi, Iringa, Kilombero, Pemba and Babati (Hertz *et al*., 2012; Vairo *et al*., 2012; Faustine *et al*., 2017). These observations may indicate a wide distribution of DENV in Tanzania. The current study confirms the findings that active dengue transmission and circulation are occurring in the country and that the coastal areas (Kinondoni) are highly affected. Furthermore, the current study reveals the presence of conformational B-cell antigenic determinants on envelope protein domain III of DENV that could possibly be used as target antigens for development of diagnostic nanobodies.

Systematic review and meta-analysis revealed an expansion of multiple DENV serotypes in Africa during the past decade (2010-2020), with a greater proportion of serotypes reported in East and West Africa (Mwanyika *et al*., 2021a). Malaria+dengue co-infections were the most prevalent followed by dengue+chikungunya co-infections. In Africa, co-

infections of arboviral infections are usually not diagnosed due to lack of differential diagnostic capacity (Petit *et al*., 2006; Bebell and Muiru, 2014). These findings underscore the need to enhance differential diagnosis of febrile illnesses in Africa. Results from this review and the seroprevalence study showed that increasing old age and lack of mosquito control measures were leading risk factors of dengue in tropical Africa.

Serological evidence from this study reveals the highest DENV seroprevalence in the north-eastern zone of Tanzania. High seroprevalence in north eastern zone may be attributed to frequent exposure following periodic DENV outbreaks that occur mostly in Dar es Salaam (Gautret *et al*., 2010; Mboera *et al*., 2016; Vairo *et al*., 2016). The lowest prevalence of DENV IgG antibodies was recorded in the central zone. This is due to the fact that the ecological conditions of central zone are characterized by semi-arid climate and a short wet-season. The presence and abundance of mosquitoes is highly influenced by high humidity and warm temperature conditions (Costa *et al.*, 2010). Several studies have reported seroprevalence of the dengue virus in different parts of Tanzania (Vairo *et al*., 2012; Hertz *et al*., 2012; Vairo *et al*., 2016; Chipwaza *et al*., 2020), however most of these studies were health facility based focusing on febrile patients (Ward *et al*., 2017). In comparison, the current study included both individuals seeking care at health facilities and those at homes and the results indicate that the prevalence of DENV antibodies was higher among individuals sampled from households than those at health facilities suggesting that there is wide circulation of DENV in the population.

Increasing age (> 28 years old) was significantly associated with DENV seropositivity in both the seroprevalence and dengue outbreak studies. Similarly, higher DENV

seroprevalence rates among older individuals compared with the younger counter parts has been reported by others in Malaysia (Dhanoa *et al*., 2018) and elsewhere in Africa (Nasir *et al*., 2017; Sawadogo *et al*., 2020). In the adults, the age-related decline of immune functions could be partly explained by prolonged exposure to viral infection (Weiskopf *et al*., 2009). Moreover, the fact that Aedes mosquitoes are day-biting, adults are likely to counter these mosquitoes when attending to their daily livelihood activities (Mboera *et al*., 2016). It is speculated that active involvement of this age-group in socio-economic activities was a possible reason for high incidence in this age group. Moreover, the high DENV prevalence among adults may suggests low endemicity of the virus in the study area. In endemic settings, high prevalence of DENV fever is mostly reported among children as adults have acquired immunity following long term exposure (Bhatia *et al*., 2013; Dayan *et al*., 2015). Similar observations have been previously reported elsewhere (Low *et al*., 2015; Du *et al*., 2021).

Phylogenetic analysis showed that the 2019 outbreak DENV strains detected in Dar es Salaam were genetically identical to DENV-1 strains reported from China (GenBank accession MN923102.1, MN923101.1, MN923096.1, MN923085.1) in 2019 and Indian (GenBank accession MK796420.1, MT126440.1, MT126438.1, MT126436.1) in 2018, 2017, 2016 and 2015 indicating the possibility of intercontinental spread of DENV-1 through globalization. Over the last decade, five DENV outbreaks in 2010, 2013, 2014, 2018 and 2019 involving different serotypes have been reported in Tanzania (Gautret *et al*., 2010; Vairo *et al*., 2016; Chipwaza *et al.*, 2020). The epidemiological change of DENV serotype circulation in Tanzania prompted to the need of studying the evolutionary relationship and geographic origin of the viruses through complete genome sequences. In this current study, the genome sequences of the DENV-1 strains were determined for

the first time using nanopore MinION sequencer. Nanopore MinION is a cost-effective portable third generation sequencer that identifies nucleotide bases by measuring an interruption of current flow generated as DNA bases pass through a small biological pore. The sequencer can generate long reads of the viral genomes in real-time that can be easily assembled (Lu *et al*., 2016; Jain *et al*., 2018). Therefore, the application of this field-deployable and cost-effective sequencing technomolgy in Tanzania can facilitate timely identification of specific DENV genotypes causing outbreaks for appropriate public health response.

Bayesian evolution analysis based on complete envelope gene (1,485 nt) indicated that the genetic identity of the Tanzanian DENV-1 strains was genotype V (of Asian origin). The genotype was closely related to strains reported in China and Japan in 2019 with > 99% nucleotide sequence homology. Globally, genotype V is the most prevalent and is known to be endemic in several countries in South East Asia (Villabona-Arenas and de Andrade Zanotto, 2013; Sun *et al*., 2016). In Africa, the presence of genotype V has also been reported in Gabon (Caron *et al*., 2013) and Senegal (Dieng *et al*., 2021).

Spatial distribution analysis from this study suggests that there was a possible introduction of DENV-1 genotype V into Tanzania during the 2019 outbreak that most likely originated from a single source in Asia. Introduction of new serotypes and/or genotypes into Tanzania may increase the risk of severe disease outbreaks in the future (Guzman *et al*., 2010; Wilder-Smith *et al*., 2017). Furthermore, evidence from two studies show that DENV genotype shift is associated with unusual outbreaks and disease severity (Shrivastava *et al*., 2015; Ma *et al*., 2021).

In the current situation, vector control and early diagnosis remain the main interventions available for dengue (Katzelnick *et al*., 2017). Routine serological tests show cross-reactivity caused by the ability of antibodies to recognize a variety of antigenic determinants (epitopes) that exhibit structural similarity among closely related viruses. The use of more specific molecular assays like reverse-transcription polymerase chain reaction is limited in resource-poor settings due to high-cost of reagents, equipment, laboratory infrastructure and the need for skilled personnel (Bhat *et al*., 2015).

In the previous studies, it was revealed that envelope protein domain III (EDIII) of DENV comprised of many conserved serotype-specific antibody-binding regions (Modis *et al*., 2004; Chávez *et al*., 2010). In this study, it was hypothesized that EDIII protein could be an alternative diagnostic marker for DENV. In addition, Dengue virus serotype 2 (DENV-2) was selected in this study to explore the diagnostic potential of EDIII protein as it is the most prevalent serotype it is associated with severe infections worldwide (Vaughn *et al*., 2000; Messina *et al*., 2014). Although, non-structural protein 1 (NS1) is the major diagnostic marker for dengue virus infection due to its presence in the blood of infected patients mostly from days 1─6 after onset of infection (Xu *et al*., 2006; Amorim *et al*., 2014), this protein has a high molecular weight approxumatley 40-50 kDa, contains sites for N-linked glycosylation and disulfide bonds making it difficult to express in normal bacterial cells that do not have the ability to execute post-translation modification of proteins (Flamand *et al*., 1999; Wallis *et al*., 2004). Furthermore, inability to discriminate dengue virus serotypes due to cross-reactive linear and conformational epitopes common to the NS1 glycoproteins of all four dengue virus serotypes and within flaviviruses (Falconar *et al*., 1991). This could be the possible source of cross-reactivity of serological diagnostic tests that lead to misdiagnosis of dengue virus infection.

Epitope prediction results from this study identified five linear epitopes at 305─320, 334─348, 320─335, 378─392, 389─400 and three conformational epitopes at 310─312, 336─337 and 341─342 amino acid residues of envelope protein domain III. The current results concur with the findings from previous reports (Nguyen *et al*., 2019). The protein secondary structure analysis from this study demonstrated that conformational epitopes were located on the buried surface of EDIII protein suggesting that the epitopes could be more conserved than those located on the outer surface. It is presumption that these epitopes can be appropriate targets for nanobody binding. The nanobodies exhibit high specificity for the cognate antigens due to their ability to penetrate protein cavities and recognize hidden epitopes that normally are not recognized by classical antibodies (Stijlemans *et al*., 2004; Baral *et al*., 2013). Results from protein expression of EDIII showed that the protein can be expressed from *E. coli* four hours after induction at a reduced temperature of 30°C. This observation sggests that optimal expression of EDIII protein in *E. coli* cells can be achieved at a lower temperature below the normal bacterial growth temperature. Successful expression of heterologous proteins at lower temperatures using *E. coli* system has been described previously (Chin *et al*., 2007; Tan *et al*., 2010).

The findings eminating from this study show that there is a large proportion of individuals who were previously exposed to DENV indicating that these individuals were unlikely to be diagnosed through the routine health delivery system because arboviruses frequently cause asymptomatic infections (Endy *et al*., 2011; Chastel, 2012). On the other hand, the symptomatic cases may develop mild symptoms that overlap with other diseases and requires that a differential diagnosis to be considered (Moreli and Costa, 2013). Therefore, it is important that diagnostic and surveillance platforms of DENV are strengthened in Tanzania in order to improve early case detection, clinical management and to monitor

their circulation for the timely and appropriate response. The findings from this study, further demonstrated that envelope protein domain III consists of antigenic B-cell linear and conformational epitopes that can be expressed from *E. coli* as potential targets for development of nanobodies targeting DENV.

3.2 Conclusions

This study highlights an increased risk of severe dengue in tropical Africa due to expanding circulation of multiple DENV serotypes. The current national wide study conducted in Tanzania reveals a wide circulation of DENV in almost all ecological zones of Tanzania with the highest seroprevalence in a coastal ecological zone. For example, in this study the highest seroprevalence was found in Kinondoni district. It was observed that aging, stagnant water bodies and piped water at home are significant risk factors for DENV exposure in Tanzania. Less availability of piped water in low income settings encourage people to store water in containers that are usually open. This provides conducive environment for mosquito breeding and larvae development resulting in higher risk of DENV exposure.

As in 2019, the DENV serotype 1 (DENV-1) genotype V (of Asian origin) was responsible for outbreak in Dar es Salaam and it is highly possible that the virus was introduced into Tanzania from Asia. The possibility of future introduction of new serotypes and/or genotypes of the viruses from multiple geographic origins due to globalization is still inevitable. The international connectivity through human travels and business with endemic countries countries is likely to increase the risk of DENV transmission and dispersation into Tanzania.

3.3 Recommendations

It is recommended that:

i. Active and effective field surveillance including genomic surveillance should be implemented to monitor the occurence for early detection of DENV in Tanzania.

ii. The government should establish a holistic mosquito-borne disease vector control programme.

iii. Dengue, malaria and other febrile illnesses share similar clinical manifestations, thus laboratory diagnosis should be strengthened.

iv. Education promotion and behavioural change awareness to take action against all mosquito-borne vectors.

v. There is a need to establish transmission indices of DENV in rural settings where little information is available.

References

Abi Thomas, E., John, M. and Kanish, B. (2010). Mucocutaneous manifestations of dengue fever. *Indian Journal of Dermatology* 55(1): 79─85.

Amarasinghe, A., Kuritsky, J. N., Letson, G. W. and Margolis, H. S. (2011). Dengue virus infection in Africa. *Emerging Infectious Diseases* 17(8): 1349–1354.

Amorim, J. H., dos Santos Alves, R. P., Boscardin, S. B. and de Souza Ferreira, L. C. (2014). The dengue virus non-structural 1 protein: risks and benefits. *Virus Research* 181: 53-60.

Baral, T. N., MacKenzie, R. and Ghahroudi, M. A. (2013). Single-domain antibodies and their utility. *Current Protocols in Immunology* 103(1): 17–57.

Bebell, L. M. and Muiru, A. N. (2014). Antibiotic use and emerging resistance: how can resource-limited countries turn the tide? *Global Heart* 9(3): 347–358.

Bhat, V. G., Chavan, P., Ojha, S. and Nair, P. K. (2015). Challenges in the laboratory diagnosis and management of dengue infections. *The Open Microbiology Journal* 9: 33–37.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J. and Hay, S. I. (2013). The global distribution and burden of dengue. *Nature* 496(746): 504–507.

Blacksell, S. D., Jarman, R. G., Gibbons, R. V., Tanganuchitcharnchai, A., Mammen Jr, M. P., Nisalak, A., Kalayanarooj, S., Bailey, M. S., Premaratna, R. and de Silva, H. J. (2012). Comparison of seven commercial antigen and antibody enzyme-linked immunosorbent assays for detection of acute dengue infection. *Clinical and Vaccine Immunology* 19(5): 804–810.

Bhatia, R., Dash, A. P. and Sunyoto, T. (2013). Changing epidemiology of dengue in South-East Asia. WHO South-East Asia. *Journal of Public Health* 2(1): 1–23.

Caron, M., Grard, G., Paupy, C., Mombo, I. M., Bikie Bi Nso, B., Kassa Kassa, F. R. and Leroy, E. M. (2013). First evidence of simultaneous circulation of three different dengue virus serotypes in Africa. *PLoS ONE* 8(10): e78030.

Chastel, C. (2012). Eventual role of asymptomatic cases of dengue for the introduction and spread of dengue viruses in non-endemic regions. *Frontiers in Physiology* 3: 1–70.

Chareonsirisuthigul, T., Kalayanarooj, S. and Ubol, S. (2007). Dengue virus (DENV) antibody-dependent enhancement of infection upregulates the production of anti-inflammatory cytokines, but suppresses anti-DENV free radical and pro-inflammatory cytokine production, in THP-1 cells. *Journal of General Virology* 88(2): 365–375.

Chastel, C. (2012). Eventual role of asymptomatic cases of dengue for the introduction and spread of dengue viruses in non-endemic regions. *Frontiers in Physiology* 3: 70.

Chávez, J. H., Silva, J. R., Amarilla, A. A. and Figueiredo, L. T. M. (2010). Domain III peptides from flavivirus envelope protein are useful antigens for serologic diagnosis and targets for immunization. *Biologicals* 38(6): 613–618.

Chin, J. F. L., Chu, J. J. H. and Ng, M. L. (2007). The envelope glycoprotein domain III of dengue virus serotypes 1 and 2 inhibit virus entry. *Microbes and Infection* 9(1): 1–6.

Chipwaza, B., Mugasa, J. P., Selemani, M., Amuri, M., Mosha, F., Ngatunga, S. D. and Gwakisa, P. S. (2014). Dengue and Chikungunya fever among viral diseases in outpatient febrile children in Kilosa district hospital, Tanzania. *PLoS Neglected Tropical Diseases* 8(11): e3335.

Chipwaza, B., Sumaye, R. D., Weisser, M., Gingo, W., Yeo, N. K.W., Amrun, S. N., Okumu, F. O. and Ng, L. F. P. (2020). Occurrence of 4 dengue virus serotypes and Chikungunya Virus in Kilombero Valley, Tanzania, during the dengue outbreak in 2018. *Open Forum Infectious Diseases* 8(1): 1–6.

Christie, J. (1872). Remarks on “Kidinga Pepo”: A peculiar form of exanthematous disease. *British Medical Journal* 1(596): 577–579.

Dayan, G., Arredondo, J. L., Carrasquilla, G., Deseda, C. C., Dietze, R., Luz, K., Costa, M. S. N., Cunha, R. V., Rey, L. C. and Morales, J. (2015). Prospective cohort study with active surveillance for fever in four dengue endemic countries in Latin America. *American Journal of Tropical Medicine and Hygiene* 93(1): 1–18.

Dieng, I., Cunha, M. D. P., Diagne, M. M., Sembène, P. M., Znotto, P. M. D. A., Faye, O. and Faye, O. (2021). Origin and spread of the dengue virus type 1, genotype V in senegal, 2015–2019. *Viruses* 13(1):57.

Duyen, H. T., Ngoc, T. V., Ha, D. T., Hang, V. T., Kieu, N. T., Young, P. R., Farrar, J. J., Simmons, C. P., Wolbers, M. and Wills, B. A. (2011). Kinetics of plasma viremia and soluble nonstructural protein 1 concentration in dengue: Differential effects according to serotype and immune status. *Journal of Infectious Diseases* 203(9): 1292–1300.

Dhanoa, A., Hassan, S. S., Jahan, N. K., Reidpath, D. D., Fatt, Q. K., Ahmad, M. P. and Allotey, P. (2018). Seroprevalence of dengue among healthy adults in a rural community in Southern Malaysia: a pilot study. *Infectious Diseases of Poverty* 7(1): 1–13.

Du, M., Jing, W., Liu, M. and Liu, J. (2021). The global trends and regional differences in incidence of dengue infection from 1990 to 2019: An analysis from the global burden of disease study 2019. *Infectious Diseases and Therapy* 10(3): 1625–1643.

Ebi, K. L. and Nealon, J. (2016). Dengue in a changing climate. *Environmental Research* 151: 115-123.

Endy, T. P., Anderson, K. B., Nisalak, A., Yoon, I.K., Green, S., Rothman, A. L., Thomas, S. J., Jarman, R. G., Libraty, D. H. and Gibbons, R. V. (2011). Determinants of inapparent and symptomatic dengue infection in a prospective study of primary school children in Kamphaeng Phet, Thailand. *PLoS Neglected Tropical Diseases* 5(3): e975.

Faustine, N. L., Sabuni, E. J., Ndaro, A. J., Paul, E. and Chilongola, J. O. (2017). Chikungunya, dengue and west Nile virus infections in Northern Tanzania. *Journal of Advances in Medicine and Medical Research* 24(4): 1–7.

Falconar, A. K. I. and Young, P. R. (1991). Production of dimer-specific and dengue virus group cross-reactive mouse monoclonal antibodies to the dengue 2 virus non-structural glycoprotein NS1. *Journal of General Virology* 72(4): 961–965.

Gautret, P., Simon, F., Askling, H. H., Bouchaud, O., Leparc-Goffart, I., Ninove, L. and Parola, P. (2010). Dengue type 3 virus infections in European travellers returning from the Comoros and Zanzibar, February. *Eurosurveillance* 15(15): 19541.

Gebhard, L. G., Filomatori, C. V. and Gamarnik, A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses* 3(9): 1739–1756.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews* 11(3): 480-496.

Guzman, M. G., Jaenisch, T., Gaczkowski, R., Ty Hang, V. T., Sekaran, S. D., Kroeger, A., Vazquez, S., Ruiz, D., Martinez, E. and Mercado, J. C. (2010). Multi-country evaluation of the sensitivity and specificity of two commercially-available NS1 ELISA assays for dengue diagnosis. *PloS Neglected Tropical Diseases* 4(8): e811.

Guzman, M. G., Gubler, D. J., Izquierdo, A., Martinez, E. and Halstead, S. B. (2016). Dengue infection. *Nature Reviews Disease Primers* 2(1): 1–25.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S. al, Robinson, G., Hammers, C., Songa, E. B., Bendahman, N. and Hammers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363(6428): 446–448.

Hertz, J. T., Munishi, O. M., Ooi, E. E., Howe, S., Lim, W. Y., Chow, A., Morrissey, A. B., Bartlett, J. A., Onyango, J. J., Maro, V. P., Kinabo, G. D., Saganda, W., Gubler, D. J. and Crump, J. A. (2012). Chikungunya and dengue fever among hospitalized febrile patients in Northern Tanzania. *American Journal of Tropical Medicine and Hygiene* 86(1): 171–177.

Higa, Y. (2011). Dengue vectors and their spatial distribution. *Tropical medicine and health* 39(4): 17-27.

Hunsperger, E. A., Sharp, T. M., Lalita, P., Tikomaidraubuta, K., Cardoso, Y. R., Naivalu, T., Khan, A. S., Marfel, M., Hancock, W. T., Tomashek, K. M. and Margolis, H. S. (2016). Use of a rapid test for diagnosis of dengue during suspected dengue outbreaks in resource-limited regions. *Journal of Clinical Microbiology* 54(8): 2090–2095.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T. and Fiddes, I. T. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 36(4): 338–345.

Kassim, F. M., Izati, M. N., TgRogayah, T. A., Apandi, Y. M. and Saat, Z. (2011). Use of dengue NS1 antigen for early diagnosis of dengue virus infection. *Southeast Asian Journal of Tropical Medicine and Public Health* 42(3): 562–569.

Kao, C. L., King, C. C., Chao, D. Y., Wu, H. L. and Chang, G. J. (2005). Laboratory diagnosis of dengue virus infection: current and future perspectives in clinical diagnosis and public health. *Journal of Microbiology Immunology and Infection* 38(1): 5-16.

Katzelnick, L. C., Coloma, J. and Harris, E. (2017). Dengue: Knowledge gaps, unmet needs and research priorities. *Lancet Infectious Diseases* 17(3): 88–100.

Lanciotti, R. S., Calisher, C. H., Gubler, D. J., Chang, G.-J. and Vorndam, A. V. (1992). Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *Journal of Clinical Microbiology* 30(3): 545–551.

Low, S.L., Lam, S., Wong, W.Y., Teo, D., Ng, L.C. and Tan, L.K. (2015). Dengue sero-prevalence of healthy adults in Singapore: sero-survey among blood donors, 2009. *American Journal of Tropical Medicine and Hygiene* 93(1): 40–45.

Lu, H., Giordano, F. and Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics and Bioinformatics* 14(5): 265–279.

Ma, M., Wu, S., He, Z., Yuan, L., Bai, Z., Jiang, L., Marshall, J., Lu, J., Yang, Z. and Jing, Q. (2021). New genotype invasion of dengue virus serotype 1 drove massive outbreak in Guangzhou, China. *Parasites and Vectors* 14(1): 1–12.

Mardekian, S. K. and Roberts, A. L. (2015). Diagnostic options and challenges for dengue and chikungunya viruses. *BioMed Research International* 2015: 834371.

Martinez, J. D., Cardenas-de la Garza, J. A. and Cuellar-Barboza, A. (2019). Going viral 2019: Zika, chikungunya, and dengue. *Dermatologic Clinics* 37(1): 95–105.

Mboera, L. E. G., Mweya, C. N., Rumisha, S. F., Tungu, P. K., Stanley, G., Makange, M. R., Misinzo, G., De Nardo, P., Vairo, F. and Oriyo, N. M. (2016). The Risk of Dengue Virus Transmission in Dar es Salaam, Tanzania during an Epidemic Period of 2014. *PLoS Neglected Tropical Diseases* 10(1): 1–15.

Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E. and Battle, K. E. (2014). Global spread of dengue virus types: Mapping the 70-year history. *Trends in Microbiology* 22(3): 138–146.

Modis, Y., Ogata, S., Clements, D. and Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427(6972): 313–319.

Moreli, M. L. and Costa, V. G. da. (2013). A systematic review of molecular diagnostic methods for the detection of arboviruses in clinical specimens in Brazil and the importance of a differential diagnosis. *Virology Discovery* 1(1): 1–8.

Morrison, J., Aguirre, S. and Fernandez-Sesma, A. (2012). Innate Immunity Evasion by Dengue Virus. *Viruses* 4(3): 397–413.

Muller, D. A., Depelsenaire, A. C. and Young, P. R. (2017). Clinical and laboratory diagnosis of dengue virus infection. *The Journal of Infectious Diseases* 215(2): 89–95.

Muyldermans, S. (2013). Nanobodies: Natural single-domain antibodies. *Annual Review of Biochemistry* 82(1): 775–797.

Mwanyika, G. O., Mboera, L. E., Rugarabamu, S., Ngingo, B., Sindato, C., Lutwama, J. J., Paweska, J. T. and Misinzo, G. (2021a). Dengue virus infection and associated risk factors in Africa: a systematic review and meta-analysis. *Viruses* 13(536): 1–7.

Nasir, I. A., Agbede, O. O., Dangana, A., Baba, M. and Haruna, A. S. (2017). Dengue virus non-structural Protein-1 expression and associated risk factors among febrile Patients attending University of Abuja Teaching Hospital, Nigeria. *Virus research* 230: 7-12.

Nguyen, N. M., Duong, B. T., Azam, M., Phuong, T. T., Park, H., Thuy, P. T. B. and Yeo, S.-J. (2019). Diagnostic performance of dengue virus envelope domain iii in acute dengue infection. *International Journal of Molecular Sciences* 20(3424): 1–16.

Peeling, R. W., Artsob, H., Pelegrino, J. L., Buchy, P., Cardosa, M. J., Devi, S., Enria, D. A., Farrar, J., Gubler, D. J. and Guzman, M. G. (2010). Evaluation of diagnostic tests: Dengue. *Nature Reviews Microbiology* 8(12): 30–37.

Petti, C. A., Polage, C. R., Quinn, T. C., Ronald, A. R. and Sande, M. A. (2006). Laboratory medicine in Africa: A barrier to effective health care. *Clinical Infectious Diseases* 42(3): 377–382.

Salvador, J. P., Vilaplana, L. and Marco, M. P. (2019). Nanobody: Outstanding features for diagnostic and therapeutic applications. *Analytical and Bioanalytical Chemistry* 411(9): 1703–1713.

Santiago, G. A., Vergne, E., Quiles, Y., Cosme, J., Vazquez, J., Medina, J. F., Medina, F., Colón, C., Margolis, H. and Muñoz-Jordán, J. L. (2013). Analytical and Clinical Performance of the CDC Real Time RT-PCR assay for detection and typing of dengue virus. *PLoS Neglected Tropical Diseases* 7(7): e2311.

Sawadogo, S., Baguiya, A., Yougbare, F., Bicaba, B. W., Nebie, K., Millogo, T. and Deneys, V. (2020). Seroprevalence and factors associated with IgG anti-DENV positivity in blood donors in Burkina Faso during the 2016 dengue outbreak and implications for blood supply. *Transfusion Medicine* 30(1): 37-45.

Sintayehu, D. W., Tassie, N. and De Boer, W. F. (2020). Present and future climatic suitability for dengue fever in Africa. *Infection Ecology and Epidemiology* 10(1): 1782042.

Simmons, C. P., Farrar, J. J., van Vinh Chau, N. and Wills, B. (2012). Dengue. New England Journal of Medicine 366(15): 1423-1432.

Shrivastava, A., Soni, M., Shrivastava, S., Sharma, S., Dash, P. K., Gopalan, N., Behera, P. K. and Parida, M. M. (2015). Lineage shift of dengue virus in Eastern India: An increased implication for DHF/DSS. *Epidemiology and Infection* 143(8): 1599–1605.

Sroga, P., Safronetz, D. and Stein, D. R. (2020). Nanobodies: A new approach for the diagnosis and treatment of viral infectious diseases. *Future Virology* 15(3): 195–205.

Stijlemans, B., Conrath, K., Cortez-Retamozo, V., Van Xong, H., Wyns, L., Senter, P., Revets, H., De Baetselier, P., Muyldermans, S. and Magez, S. (2004). Efficient targeting of conserved cryptic epitopes of infectious agents by single domain antibodies: African trypanosomes as paradigm. *Journal of Biological Chemistry* 279(2): 1256–1261.

Sun, J., Wu, D., Zhou, H., Zhang, H., Guan, D., He, X., Cai, S., Ke, C. and Lin, J. (2016). The epidemiological characteristics and genetic diversity of dengue virus during the third largest historical outbreak of dengue in Guangdong, China, in 2014. *Journal of Infection* 72(1): 80–90.

Tan, L. C. M., Chua, A. J. S., Goh, L. S. L., Pua, S. M., Cheong, Y. K. and Ng, M. L. (2010). Rapid purification of recombinant dengue and West Nile virus envelope

Domain III proteins by metal affinity membrane chromatography. *Protein Expression and Purification* 74(1): 129–137.

Vairo, F., Mboera, L. E. G., de Nardo, P., Oriyo, N. M., Meschi, S., Rumisha, S. F., Colavita, F., Mhina, A., Carletti, F., Mwakapeje, E., Capobianchi, M. R., Castilletti, C., di Caro, A., Nicastri, E., Malecela, M. N. and Ippolito, G. (2016). Clinical, virologic, and epidemiologic characteristics of dengue outbreak, Dar es Salaam, Tanzania, 2014. *Emerging Infectious Diseases* 22(5): 895–899.

Vairo, F., Nicastri, E., Meschi, S., Schepisi, M. S., Paglia, M. G., Bevilacqua, N., Mangi, S., Sciarrone, M. R., Chiappini, R. and Mohamed, J. (2012). Seroprevalence of dengue infection: A cross-sectional survey in mainland Tanzania and on Pemba Island, Zanzibar. *International Journal of Infectious Diseases* 16(1): 44–46.

Vaughn, D. W., Green, S., Kalayanarooj, S., Innis, B. L., Nimmannitya, S., Suntayakorn, S., Endy, T. P., Raengsakulrach, B., Rothman, A. L. and Ennis, F. A. (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *Journal of Infectious Diseases* 181(1): 2–9.

Villabona-Arenas, C. J. and de Andrade Zanotto, P. M. (2013). Worldwide spread of dengue virus type 1. *PloS ONE* 8(5): e62649.

Vasilakis, N., Cardosa, J., Hanley, K. A., Holmes, E. C. and Weaver, S. C. (2011). Fever from the forest: Prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nature Reviews Microbiology* 9(7):532–541.

Ward, T., Samuel, M., Maoz, D., Runge-Ranzinger, S., Boyce, R., Toledo, J. and Horstick, O. (2017). Dengue data and surveillance in Tanzania: a systematic literature review. *Medicine and International Health* 22(8): 960-970.

Weiskopf, D., Weinberger, B. and Grubeck-Loebenstein, B. (2009). The aging of the immune system. *Transplant International* 22(11): 1041–1050.

Whitehead, S. S., Blaney, J. E., Durbin, A. P. and Murphy, B. R. (2007). Prospects for a dengue virus vaccine. *Nature Reviews Microbiology* 5(7): 518–528.

Wilder-Smith, A., Gubler, D. J., Weaver, S. C., Monath, T. P., Heymann, D. L. and Scott, T. W. (2017). Epidemic arboviral diseases: Priorities for research and public health. *Lancet Infectious Diseases* 17(3): 101–106.

World Health Organization (2009). *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control. In Special Programme for Research and Training in Tropical Diseases*. World Health Organization, Geneva. 147pp.

Xu, H., Di, B., Pan, Y. X., Qiu, L. W., Wang, Y. D., Hao, W. and Che, X. Y. (2006). Serotype 1-specific monoclonal antibody-based antigen capture immunoassay for detection of circulating nonstructural protein NS1: implications for early diagnosis and serotyping of dengue virus infections. *Journal of Clinical Microbiology* 44(8): 2872–2878.

APPENDICES

Appendix 1: Sample size estimation

Chapter Two Paper Two: Materials and Methods

Sample size determination

Sample size of study participants recruited from 8 districts in Tanzania was estimated using a formula for estimating sample size for prevalence studies (Naing *et al*., 2016);

$$n_c = [z^2_{\alpha 0.05} \times (p \times (1\text{-}p))/e^2]/ ((z^2_{\alpha 0.05} \times p(1\text{-}p)/ e^2 \times N))] \times DE \quad \text{(i)}$$

$$no = [(nc \times Nr) + nc] \quad \text{(ii)}$$

n_c = cluster sample size

Z = Z statistic confidence level of 95% (1.96)

p = assumed virus prevalence, zone specific based on Simon *et al* 2019 estimate prevalence for healthy and febrile population

1-p = probability of having no attribute of interest

e = marginal error (5%)

N=district population size according to TBS 2018 census

DE= design effect (1.5) to account for clustering effect

Nr= non-response rate/missing values (10%)

n_o = desired sample size

Appendix 2: Ethical clearance certificate

THE UNITED REPUBLIC
OF TANZANIA

National Institute for Medical Research
3 Barack Obama Drive
P.O. Box 9653
11101 Dar es Salaam
Tel: 255 22 2121400
Fax: 255 22 2121360
E-mail: ethics@nimr.or.tz

Ministry of Health, Community
Development, Gender, Elderly & Children
University of Dodoma, Faculty of Arts
and Social Sciences
Building No 11
P.O. Box 743
40478 Dodoma

NIMR/HQ/R.8a/Vol. IX/2724

09th March 2018

Dr. Leonard E. G. Mboera
Southern African Center for Infectious Disease Surveillance
P.O. Box 3297
Morogoro

RE: ETHICAL CLEARANCE CERTIFICATE FOR CONDUCTING MEDICAL RESEARCH IN TANZANIA

This is to certify that the research entitled: Seroprevalence of viral haemorragic fevers in Tanzania: Strengthening scientific capacity for surveillance and response. (Mboera L *et al.)* has been granted ethical clearance to be conducted in Tanzania.

The Principal Investigator of the study must ensure that the following conditions are fulfilled:

1. Progress report is submitted to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research, Regional and District Medical Officers after every six months.
2. Permission to publish the results is obtained from National Institute for Medical Research.
3. Copies of final publications are made available to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research.
4. Any researcher, who contravenes or fails to comply with these conditions, shall be guilty of an offence and shall be liable on conviction to a fine as per NIMR Act No. 23 of 1979, PART III Section 10(2).
5. Site: Kigoma, Rukwa, Mbeya, Dar es Salaam, Tanga, Morogoro, Dodoma and Mwanza

Approval is valid for one year: 09th March 2018 to 08th March 2019.

Name: Prof. Yunus Daud Mgaya

Signature
CHAIRPERSON
MEDICAL RESEARCH
COORDINATING COMMITTEE

Name: Prof. Muhammad Bakari Kambi

Signature
CHIEF MEDICAL OFFICER
MINISTRY OF HEALTH, COMMUNITY
DEVELOPMENT, GENDER, ELDERLY &
CHILDREN

CC: RMOs of: Kigoma, Rukwa, Mbeya, Dar es Salaam, Tanga, Morogoro, Dodoma and Mwanza

DMOs/DEDs of selected districts.

Appendix 3: Informed consent form

INFORMED CONSENT

SEROPREVALENCE OF DENGUE VIRUSES IN SELECTED DISTRICTS IN TANZANIA

SOKOINE UNIVERSITY OF AGRICULTURE (SUA)

SACIDS Africa Centre of Excellence for Infectious Diseases of Humans and Animals in Eastern and Southern Africa College of Veterinary Medicine and Biomedical Sciences

P.O Box 3015, Chuo Kikuu, Morogoro, Tanzania

Tel: +255 23 264 0037; +255 787 011 677

The following statement will be read to all individuals asked to participate in the survey.

My name is ……, and I am from Sokoine University of Agriculture. I am carrying out a study to establish the seroprevalence of dengue and chikungunya antibodies among participants in selected districts of Tanzania. It is your free choice to be part of this study. The results of the study will lead to improvement of surveillance and diagnosis of dengue viruses in our country which have been often misdiagnosed with other febrile illnesses such as malaria, flu and yellow fever.

A qualified technician will describe to you about the collection of blood and He/she will seek your permission to collect small amount of blood (approximately 3─5mL) which he/she will take from hand using special syringe and needle. This procedure is less painful and there is the possibility of little discomfort at the site where the blood is taken. The

blood specimen will be taken to a laboratory at Sokoine University of Agriculture in Morogoro and will be tested for antibodies against dengue viruses.

There are no direct benefits for participating in this study, but it will help the government of Tanzania to plan for appropriate intervention strategies against dengue. You are free to choose to be part of this survey. However, if you accept to take part in this study, there will be no payment to you. The facts about you from this survey will be kept confidential as directed by the Laws of the United Republic of Tanzania. No names will be used on any of the survey reports, publications or presentations. Only we, the researchers, will ever see the surveys with people's names. If you choose not to participate in this study, that is fine too. You will not be treated differently by the health personnel in this area. You may ask the researchers any questions you have at any time.

Do you wish to participate? YES; NO (Please circle)

Age ≥18 years

Participant's signature (or thumb print): _____

Date: _____

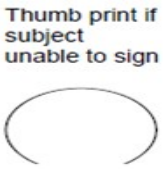
Age < 18 years

Name of the child: _____

Signature of the child (or thumb print): _____

Parent's/Guardian's signature (or thumb print):

_____Date: _____



Contacts: Telephone: +255 756 205 005; E-mail: mwanyikag254@gmail.com

Appendix 4: Structured questionnaire

Assessment of risk factors associated with increased probability of dengue virus exposure in Tanzania

Interviewer names_____

A. Identification Information *(To be filled by Interviewer)*	
A.1 Region:	
A.2 District done	
A.3 Village/Street	
A.4 Date of interview/sample collection (dd/mm/yyyy):	[__][__]/[__][__]/[__][__][__][__]
A.5Place of sample/data collection	1. Household [___] 2. Facility [___]
A.6. If A.5 = "2"; mention the name and level of facility	6.0. Facility name_____ Level: 6.1. Dispensary [___], 6.2. Health Centre [__], 6.3. Hospital [__]
B. Participant details *(To be filled by Interviewer)*	
B.0 Participant ID (**Take from names sheet**; use it to label the sample). The first three letters represent the district name ***(Refer A.2)*** followed by three numbers representing the ID of participant).	[__][__][__] [__][__][__]
B.1 Age (years) / (months)	[__][__][__] [_]
B.2 Sex	1. Male [__] 2. Female [__]
B.3 Place of Living (Village/street)	
B.4 How long have you lived in this village/street?	Years [___] Months [____]
B.5 Occupation (multiple selection)	1=Livestock farming 2=Crop farming

	3=Fishing 4=Hunting 3=Informal sector (Daily labored/self-employed) 4=Formal employment (Public service, /private sector/NGOs) 6=Student (primary/secondary) 7=Business/trader 8=Mining 8=Others, specify _ __ _____
B.6 Highest Educational level of the respondent	1= Primary school; 2=Secondary school 3= College/University; 4=Vocational training school; 5= Adult education; 6=None
C. Travel history (*to be filled by Interviewer)*	
C.1 On average how many times have you been outside your region during the past 12 months?	1=none 2= weekly 2=monthly 3=quarterly
C.2 Please, mention the countries other than Tanzania where you have been to during the past 12 months	0. Never been outside Tanzania 1._____ 2._____ 3._____
D. Clinical Assessment	
Present Illness	**Write notes, if any**
For the past one months: D.1.Ever suffered from fever? Yes [__] No [__] NK[__] D.2.History of fever persisting after using antimalarials Yes [__] No [__] NK[__]	
D.3 Headache	1 [_] Yes 2 [_] No _____
D.4 Rash	1 [_] Yes 2 [_] No _____
D.5 Muscle, bone, back or joint pain	1 [_] Yes 2 [_] No

	_____
D.6 Bleeding from natural body openings	1 [_] Yes 2 [_] No _____
D.7 Nausea	1 [_] Yes 2 [_] No _____
D.8 Vomiting	1 [_] Yes 2[_] No _____
D.9 Bruising	1 [_] Yes 2 [_] No _____
D.10 Fatigue	1 [_] Yes 2 [_] No _____
D.11 Stomach ache	1 [_] Yes 2 [_] No _____
D.12 Red spots (skin/eyes/mucosa)	1 [_] Yes 2 [_] No _____
D.13. Jaundice	1 [_] Yes 2 [_] No _____
D.14 Others, specify	1 [_] Yes 2 [_] No _____
E. Exposure risk practices (*To be filled by interviewer***)**	
During the past three months; has participant:	
E.1 Have you Visited caves?	[_] Yes [_] No [_] NK
E.2 Have you Visited mines?	[_] Yes [_] No [_] NK
E.3 Do you have piped water at home?	[ ] Yes [ ] No [ ] NK
E.4 Do you frequently visit recreation facilities?	[ ] Yes [ ] No [ ] NK
E.5 Do you use mosquito net?	[ ] Yes [ ] No [ ] NK
E.6 Do you use mosquito repellants?	[ ] Yes [ ] No [ ] NK
E.7 Are there some stagnant water bodies at home?	[ ] yes [ ] No [ ]NK

Appendix 5: Primer sequences used for PCR genome amplification

Table S1: Primer sequences used to amplify DENV-1 genome

Fragment	Primer name	Sequence (‘5-3’)	Reference
F1	D1F1-5F	GTTAGTCTACGTGGACCGAC	(Sessions *et al*., 2013)
	D1F12084R	CACCTGCTCCTATCACGATG	
F2	D1F2-2201F	GGYTCTATAGGAGGRGTGTTCAC	(Sessions *et al*., 2013)
	D1F2-4561R	TTCCACTTCYGGAGGGCT	
F3	D1F3-4221F	CACTAATAGCTGGAGGCATGC	(Sessions *et al*., 2013)
	D1F3-6461R	CCAGGTTGTCCAAGGCATTC	
F4	D1F4-6442F	GAATGCCTTGGACAACCTGG	(Sessions *et al*., 2013)
	D1F48519R	CACCATTGACCATGGATGAGGC	
F5	D1F5-8540F	GCCTCATCCATGGTCAATGGTG	(Sessions *et al*., 2013)
	D1F5-10693R	CTGTGCCTGGAATGATGCTG	

Appendix 6: Oxford nanopore MinION sequencing report

NanoPlot quality control report (QC)

Median read length: 643.0

Median read quality: 9.7

Number of reads: 2,756.0

Read length N50: 1,487.0

Total bases: 2,529,696.0

Number, percentage and megabases of reads above quality cutoffs

>Q5: 2756 (100.0%) 2.5Mb

>Q7: 2756 (100.0%) 2.5Mb

>Q10: 1186 (43.0%) 1.1Mb

>Q12: 107 (3.9%) 0.1Mb

>Q15: 0 (0.0%) 0.0Mb

Top 5 highest mean basecall quality scores and their read lengths

1: 14.3 (254)

2: 14.3 (196)

3: 13.6 (205)

4: 13.6 (403)

5: 13.5 (816)

Top 5 longest reads and their mean basecall quality score

1: 3916 (8.0)

2: 3747 (8.3)

3: 3610 (8.0)

4: 3585 (10.0)

5: 3423 (7.7)

Nanopore fastq reads data

@195ef502-6a02-4199-85dc-a041321547f3
runid=e30d20a9344ea019ca91a11df5053149e2ec3aa5 read=18737 ch=199
start_time=2020-12-12T01:03:41Z flow_cell_id=FAN52084
protocol_group_id=Gaspary_Mwanyika sample_id= barcode=barcode21
CATTGTAGCGTTCAGTTACGTATTGCCTAAGGTTAAGAGCCTCTCATTGTCCGTTCTCT
ACAGCACCTCACCTGCTCCTATCGTCTTAGCTCTCACCAAAGGGTGGTTCTGTTTCAAT
GTTGACTGGTTTTTCTTTGTCAGTAACTATAGGGTTGGCTGTTATCAATCTCCATTCTA
GGTCACTGCACTTTTTTCTCATTGGTTGGGTTGAAATTGGAATCTTGCATGGTGCATCT
GTTCCTTCATATTTAACCTGCACCCCAGACGATTATTAATGCTGGGTCTCAGCCATATG
GACTTAAATGAACCTGTGCACATCACATATGACACCCCTTTTAAAGTCAATATTCAGT
TGGTCGGACATTTCAGGTGTCCTGCGAAGATTGTCGTTGTTCCTGACGTCTGGATTTCT
ATTTACCCCAGTCCAACGCAGTGTTACATTGCTCTTCTTGTGACCCCAGTACAACTACT
TCCTGCTTCTTTGCATGAGCCACATCTTAAATGTGACCAGCAGATCTTGTCTGTTCCAA
GTCTCTTGAGATGTTAAGCCCCGAGGTCCAGGGCAATGGTAAGTCTAAAAACCATTGT
TTGTGGACAAACCGCGATTTCCCTGCTGTTGTCCAACAGCACCATCTCATTAAAAGTC
CAGCCCTGTTCTAGGTGAGCAATCCAATGCAAGGGCTCCGTAATCAGTCAGCTGTATT
TCCGACAGTGGGAGTTTGAGGTGTTAGTAGTTTGCAATTGTTCCATGTTCTGTAGTTTC
ATTTCCTGCCTGGTGCTGGTCCCCAGTGTGGACAGTGAATGCCGAATATTTTAAGTTTT
CATATTCTTGCTAAGTTTGTACTTAAACCAGGACACGTTAGTAGCTTCCCTTCCCGAAT
AGTCCACAGCCATTACCCCAGCCTCTGTCTCCACGAATGTTCGTCGACACACAAAGTT
TGCGTCTTGTTCTTCCACCAGTGTAACTTCTCTTGTGTTGGACATCTTGAATCGGTAGT
GGTGTTTGATATTTTAGCTTCAATGCACAGTTTGCGCAAGACGGCAGGGTTCGTGACC
TCCGTCTTCAAGAGTTCAATGTCAATGTTGGTTTATTTTTGCCATGGTGATACATGGGC
ATGCCATCAGTACCACATCCACTACCCATGTTGCTCATGACGGTCGGGCAACGAAGTC
TCTGTTGCCTATTCCCACATCGCATGGCCATTGATAAGTGTCACAGCATCAGCAAAAT
ATGAAATGATCGGTCTGGGTAAATCTGGATGTTTCTGGATGGCATGTGCTAGAAAAAG
GGCTATCACCAGCTGAATCCTGGGTGTCTCAAGGCCCAAGTCTCCCTTTTGTATCTGTC
TTAGGCGCCAGGATTCAGAGGACATCCACGTTTCGGTTCTTGTTTCGAGACCAAGCCC
CACGTGTGGGAGCCAATGCGACAGAGCGTTTATCCCGTCGGTGTTCGCCAGTCACGAA
CACGTCCCGCAGGTCATAATATTGTCTGTGGCATTGCACCAGCAATCAACATCATCTG
GTTCCGCTTCCAGTGATCCAGGGGCATTTGTGAACATTGTATCCTCACACAACTCTCTC
TCTTTTTCTCTCTCTTTCTCTCTGCCCCAAATCTGGACAATGAGGGTGCACATGTTGGC
ACCTACAGGGTCGATAAAAACAAAGAGTGACTTTCCCTCTTTCCTGCTTGGTAACTAT
CAAATGCTCTCCCCTCGTGTGGTAAATTGGAGACAGGGCTGTGGGCAGCATGAGAGC
ATGGTCACGGATCTTTTCTCTACATTCATTATGTTCAGCATGCTTGAGATCTCTTTTGA
AACCTCGTAACACTTTTGATCATCATTCTTCTTGAATGATATCTAACAAGATCTTGCTA
TTTAGGGGTATGGCTAGAAATCTTAGGAATGCTATAGAAAGCCATCACCATTTTCATG
GGTCCTTGGCCTGAGAGCAATCCTTTTGAGAATCTCTTCGCCAACTGTGAACCAGTTG
ACGCGGTTTCTCGCGCGTTTTCAACATATTGAAAGACGGTCGACCCGTCTTTGGACCG
TTGGTTGTTCACAGGGAATCTGCTCTCTAATAAAAAACTGTTAGAGAACTACACGTTA
AGAAAAGAAGCTTCAGGCGAGAAACTGTTCTTGTCGGTC

@36528d90-0d28-4ebe-a6f1-c26ecdc82645
runid=e30d20a9344ea019ca91a11df5053149e2ec3aa5 read=3372 ch=73 start_time=2020-

12-11T22:30:21Z flow_cell_id=FAN52084 protocol_group_id=Gaspary_Mwanyika sample_id= barcode=barcode21
CGTGTACTTCGTTTCGGTTACGTATTGCTGCAAGGTTAAGAGCCTCTCATTGTCCCGTT
CTCTACAGCACCTGGTTCTATGGGAGGGGTATTTGGCTGTCAGTGGAAAAGTGGTACA
GGTCGACGGAACCGCATAGGGTCTGATGTTCGCGAGTGGACGGACCATGAAAATAGG
AATAGGGATTCTGCTGACATGGCTAAGGAGAGGTGGATCAGGGCACGTCTCTGGCAG
CGTGCATTGCAGCCCGGCGTGTGCTGTACCTGAGTCATGGTTCAAGCGGTGGCGGGGT
GTGTGATCAACTGGAAGGGCAGAAGGGCTTTCAATCATGGAAATGTGTCACTAATGG
AGTCCATACTTGGACAGAGCAATACAAATTCCAGGTAGCTCGCTTTAAGACTATCAGC
CATTGGGAAGGCATGGGAGGAGGCGTGTGTAGATTCAGTCATTACATCTTTGAGAAC
ATCATGCTGGAAGCAAATATCAAATGGAATTGAACCACATTCTGCTTGAAAATGACAT
GAAATTCCCCAGTGGTTGTAGGGAGATGCCAATGGAATCTTGACCCAGAAAAAGATG
ATTAGGCCACAACCTATGGAACCCCAAATACTCCTTGGAAAAGCTGAGGAAAAGCCA
AAATCATAGGGCAGATACACAGGGATACTACCCGGGTGCTCATCTGTTGGCCAGAAC
ACTCAGAATGCTGATGACCAGAGAGCATGGAACATTTGGGAAGTTGAGGACTATATG
GGTTTGGAGTTTTCACAAAACCAAACATATATATTTGAAATTGGCGGCTCCTATTTTA
AATGTGTGACCACCGGCTAATGTCAGCTGCCATCAAGGACAGCAGGGCAGTCCATTG
CCTGACATAGGGTACTGGATAAGTGAAAAGAACGAGACTTGGAAGCTAGCGAGCCTC
CTTAATAAAGTGAAGACGCCATCTGGCCCAGATCCCTCTATGGAGCAATGGAGTCTTG
GAAAGTGAGATGATAATCCCAAAAAGATATATGGAGGACCAATATCTCAGCACAACA
CAGACCAGGGTATTTCACACAGAAACAGCAGGGCCATGGCACTAGGCAAGTTGAAGA
GAGAAGGAAGAGAAGGACTGGTTCAATTTGTGCTTAAGAGGCACCACAGTTGTTGTG
GATGAACATTATTGAAATCGAGTCCATCTTAGGAGATTTCACAACAGTCACAGGAAG
ATAATCCATGAATGGTATTGCCAGATCTACGCTACCCCCTTACGCTTCAGAGAAGACA
GGTGGGACATGGAAATCAGACCAGTTAAGGAGAGAGAGAGAATCTAGTCAGGTCAAT
GGTCTGGGGTCAGGAGAAGTGGATAGTTTTCCATTGGGACTACTATGCGTATCAATAA
TGATTGGAAAAGTGATGAGATCCAGATGGGTAGAAGATGCTGATGACTGGAACACCT
GGCTGTTTCTCTTCCTCCGGAATGGGCTAATAGCTTGAATGATCTGATCAGGTTATGC
ATCATGGTTTAGACTAATGTCTCAGACAGAATGGGGATGGGAACAACGTACCTAGCC
TTAATGGCCACTTTTAAATGAGACCGATGTTTAAGCTGTTGGGCTGTTATTTCATGAAT
TACATAACATCAGAGGGTGCTCCTTCTCACAGATTGGATAAACACAGGCATCTCTCCC
CGATGTTACTGCACCCCTGTAATTCTGGAAGAGCTGGGGACGGACTTGCGATGAACTT
TCCTTGATGTTAAAATTGTTGACTGATTTTCCAGTCCTCTACCAGTTAATGGACTACCT
TACTGTATCTCTGACGTTGCTATCAAAAACTAGCCCTTTCACTGGATTGTAAAGCGGA
AGACAATGAGCCGTATTTATGGTGCGATGTATATATATTTTCCTCCTGCTATTCCTAAA
TGGATAAAAAGCAACATGGCTTCCGGTGCCTTTAGGATCATGGAGGATGTAAACCAC
TAACCATGTTTCTTATAACCAAGAAAACAGAATTTGGGGAAGAAAAGTTGGCCCCTC
AATGAGGAATTATGGCTATTGGAATAGTCAGCATTTTTACTAGGATTCACTCCTCAAA
AATGACGTGCCACTAGCCGGCCCACTAATGGCTATGAGCATGCTGATAGCATGTACGT
CATATCGGAAGCTCAGCCGATTTATCATTGGAGAGACGGCTGAGATTCTCCTGGGAAA
AAAGCAGAACACTCTGGTGCCTCTTACAACATACTAGTGGAGGTCCAAGATGATGGA
ACCATGAAATAAAAGATGAAGAGAGAGGTCAGGCACTCTTCTGCCATATAGAAAGCA
ACTTTTTGCTGGCAGAAAGTTTCAGGGTACCCAATGTCCAGCGCAACAACTCTTTTGG
CATGGTATTTTGGCGAAAACACAGAGATCAGGAGTCTTGTGGGACGCACAGCCTC

@e4065ee0-09bb-4527-9844-6340c61c2dca
runid=e30d20a9344ea019ca91a11df5053149e2ec3aa5 read=5886 ch=206

start_time=2020-12-11T23:04:14Z flow_cell_id=FAN52084
protocol_group_id=Gaspary_Mwanyika sample_id= barcode=barcode21
CAGTAGTACTTCCGTTCAGTTACATTGCTAAGGGTTAAAGAGCCTCTCATTGTCCGTTC
TCTGACACACCTCACTAATAGCTGGAGGCATGCTGATAGCATGCTACGTCATATCCAG
AGCTCAGCCGATTTGCTTCATTGGAGAGGCGCGGGGTCCTCTGGGAAGAAGAAGCAG
AACACTCACAGGTGCCTCACACACAACATACTAGTGGAGGTCAAGATGATGGAACCA
TGAAAATAAAAGATGAAGAGAGATACGCTCAAGGACCACTCTCAAAGCAGTTGGCGG
TCAGTGAGGGGTGTACCCAATGTCAATACCAGCAACCCTTTTTGTGTGGTATTTTTGG
CAGAAAAAGAAACAGAGATCAGAGTCTTGTGGGACACCAGCCCTCAAGTAGAAAAG
AGCAGTTCTCGATGATGGTATCTATAGAATCTTGCAAAGAGGACTGTTGGGTAGGTCC
CAGGTGGAGTGGGAGGTTTTCAAGACGGCGTGTTCCACACAATGGCATGTCACCAGG
GCTGTCCTCATGTACCAAGGAAGAGGCTGGAACCAAGTTGGGCCAGTGTCCAAAGAC
CTGATCTCCTTATGGAGGAGGTTGGAGGTTTCAGGATCATGGAACACAGGAGAAGAG
TACAGGTGATAGCTGTTGAACCAGGAAAAAACCAAAAGAATGTACAGACAACGCCGG
GCACCTTCAAGACCCCTGAAGGCGAAGTTGGAGCCATAGCCTAGATTTCAAACCAGG
CACATCTGGATCTCCCCATCGTGAACAGAGGGAAAAATAGTAGGTCTTTATGGAAAT
GGAGTGGTGACAACAAGTGGAACCTATGTTAGTGCCATTGCCCAAGCCAAAGCGTCA
CAAGAAGGGCCTCTACCAGAGATTGAGGACAGGTATTTAAGAAAGAAACTAACAATA
ATGGACCTACATCCAGGATCAGGAAAAACAAGAAGATATCTTCAGCCATAGTCCGTG
AGGCCATAAAGAAACTGCGCACGCTAATCTTGGCTCCCACAAGGTTGTCGCCTCTGAA
ATGGCAGAGGCGCTCAAGAGATACCAATAAGATATCAGACAACAGCAGTAAAGTGA
ACACACAGGAAAGGAGATAGTCGATCTCATGTGCCACGCCGCTGTTTCATGCGTCTCT
ATCTCCAGTGAGAGTTCCCAATTACAACACACCGTTGTATTATGGATGAAGCACACTT
TACTGATCAACTTTTTTTTTATAGCGACATCAGGGGTACATCTCAACAGTGGGCATGG
GTGAAGCAGCTGCGATCTTTATGACAACCACTCCCCCAGGATCAGTGAAGCCACCAC
AGAGAAATACAGTTGGCTAAAGATGGGGAAAGAGACATTCCTGAGGAGATCATGGA
ACTCGGGCCACGACTGGATCACTGATTTCCCAGGAAAAACAGTCTTGGTTGTTCAAGT
ATCAAATCAGGAAATGACATTGCCAACTGTTTAAGAAAAAACGGAAACAGGTGATCA
ATTGAACGAAAAACCTTTGACACTGAATACCCCAGAAAACAAAAAACAACAGATTGG
GACTATGTTGTCTAACAGACATTTCAGGAAATGGGAATGGGTCTTCCGGGCCGACAG
GGTAATAGACCCAAGGCGGTGCTTGAAAGCCAGTAATACTAAAAGATGGCCCAGAGC
GCGTCATTCTAGCCGGACCAATACCCAGTGACCGCAGCCAGTGCTACCAGAGAAGAG
GGAAGAATTGGAAGGAACCAAAACAAGGTGATCAGTATATTTATATGGGACAGCCTT
TAAATAATGATGAGGGCCACGCTCATTGGACAGAAGCAAAAATGCTCCTTGACAATA
TAAATACGCCAGAAGGGATTATCAGCCTGGAGGCCAGAGAGAGAAAAGAGTGCAGC
AATGAGCAGGAGTACAGACTGCGGGGAGAAGCAAGGAAAACGTTCGTGGAGCTAAT
GAGAAGAGGGGATCTACCAGTCTGGCTATCTTACAAAAGTTGCCTCAGAAGGCTTCC
AGTACTCCGACAGAAGATGGTGCTGACGGGGGAAAGGAACAACCAGGTGTTGGAGG
AGGAGAAACATGGACGTGGAGATCTGGACAAAGGAAGGAGAAGAAAGAAATTGCGA
CCTCGCTGGTTAGATGCCAGAACCTACTCTAGCAATGTACGCGAGTTTAAAGAGTTCC
TGGCAGGAAGAAGAAGAAGTGTCTCAGGTGACCTAATATTAGAAATGAGAAACTTCC
ACAGCATTTGACGCTAAGGGCCCGGAATGCCTTGGACAACCTGGAGGTGCTGTATGA
GAACAGACAATGAGAGGCTCTTAACCTTAGCAATACATGTGGC

@a48d748f-9c47-4240-9e10-a78f1b37f515
runid=e30d20a9344ea019ca91a11df5053149e2ec3aa5 read=468 ch=267 start_time=2020-

12-11T22:04:24Z flow_cell_id=FAN52084 protocol_group_id=Gaspary_Mwanyika sample_id= barcode=barcode21
CAGTAGCCTTCCGTTCAGTTACGTATTGCTAAGGTTAGAGCCTCTCATTGTCCGTTCTC
TACAGCACCTGAATGCTGTGGACAACCTGGCCATGTTTTTAATTCCGAACAAGGAAGC
CTATAGACATGCTGAGAGCTTGCCAGGCTATGCATTTGGTCAAGGACTGGCCTTGATG
GCTATGCTGACTGGTGGAGTTGTTTTCCGGGGAGAAAAGGTCTAGGGAAAACATCCA
TTGGCCTACTTTGTGTAATAATGTGAGTGTACTGTTATGGATGTAGTGTGGAGCCCCA
TTGGATAGCGGCCTCCATCATATTAGAGTTTTCTGATGGTGCTGCTCATTTCCAGAGCC
AGACAGACAGCAGTACTCCACAGGACAATCAACTAGCATATGTGGTGATAGGTTTTC
ATGATATTGACAGTAGCCAATGAGATGGGATTACTGGAAACCACAAAAAGAAAGGAA
AACCTAGAGGGTGTCATGTAGTCGCCGAAAACCTACCATAATATACAATGCTAGACA
TAGACTTACATCCAGCTTCAGCCTGGACCTCTATATATTATACAACAATTATCACTCCC
ATGATGAGACACGTAATTGAAAACACAACGGCAAACATTTCCCTGACAGCTATTGAG
CGGCTATATTAATGGGACTTGACAAGGGATGGCCAATATCGAAGATGGTGCCGGAGT
GCCATACTCGCTTTGGGGTGCTGTTTATTCCCAGGTGAACCCATTGACACTGACAGCG
GCGGTGTTGATGTTAGTCGGCTCATTATGTATAATCGGACCTGGACTGCAAGCAAAAG
CTACTAGAAGCCCAAGGACTGCGACCGGAATAATGAAAAATCAACTGTAGACAGGAT
TGTTAATAGATTTGGATCCTGGTTTATGATGCAAAATTTAGAAAAATCACTAGGCAAA
TAATGTTACTAATACTCTGCACATCACAGATTCTTGATGCCGGACTACATGGGCAGAG
AATGGCAGATCATCACTGGCTACTGGACCCTTGACCACTCTGGGAGAGGATCTCCCTG
GAAAGATTACCAGACACCGATAGCAGTGTCCATGGCAAACATCTTCAGGGGAAGTTA
TCTAGCAGGAGCAGGTCTGGCCTTCTCATTAATGAAATCTTTAGGAGGAATTTAAGGA
GAAGTGCAGGGCTCAAGGAAGTGGGAGAAAGATGGAAAAAGACAATTAAACCAACT
GAGCAGAGTCAGAATTTAACACCTAAAAGGAGTGGAATTATGGAGGTGTGGACAGAT
CTGGAAGCCCAAGGAGGACTGAAAAGAGGAGAAACAACCAAACATGCAGTGTCGAG
AGGAACAGCCAAACTTAGGTGGTTGGTTTGTGGAGAGGAACCTTGTGGAAACCAGAA
GGGAAAGTGATAGACCTCGGTTGTGGAAGAGGTGGCTGGTCTGTGCTGCGCTGGGCT
GAAAAGGTCACAGAAGTGAGGATGCAAGGAGGACCCAGTATGAGGAACCTATCCCA
ATGGCACATGGATGGAACATAATTGACAGCTATGGAAGGATGTTCTTTACACCACCTG
AGAAATGTGACACCCTTTTGTGTATGTGATATTGGTGGTCCTCTCGAATCAACTATAG
AAGAAGAAAGGAAGAACGTTACGTGTTCTAAGATGGTGGAACCATGGCTCAGAGAAA
TCAGTTTGCATAAAAGTCCTAAATCCTTACATGAGAATGGTAGAAACTCTGGGTAAAT
GCAAAGAAAACATGGGAACTAGTGCAAACTCCACTCTCAAGAAATTCTACCCATGAA
ATGTACTGGGTTTCATGTGGAACGGAAAACATTGTCCGGCAGTGAACATGACATCAG
AATGTTGCTGAATCGGTTCACAATGGCTCACAGGAAGCCAACGTATGAAAGAGGGCC
GGACTTGGGCGCTGGAACAAGACATGTGGCTTGTGGAACCAGAGATAGCCAGCCTAG
ACATCATTGGCCCAGAGGATAGAGAACATAAAATGAACACAAGTCAACATGGCATTA
CATGATGAGGACAATCCATACCACAAAACATGGGTATCATGTGGATCATATGAGGTC
AAGCCGTCAGGTGTACCTCATCCATGGTCAATGGTGACCTCACCCCCATGGTCAATGG
TGAATGCCTTGGACAACCTGGAGATTCGTTTTTATCTACATCTGAGCCCCCACCCAGC
C

Draft genome consensus sequence

>strain_1 DENV-1/TAN/DAR_04 Tanzania 2019

TGTTAGTCTACGTGGACCGACAAGAACAGTTTCGAATCGGAAGCTTGCTTAACGTAGT
TCTAACAGTTTTTTATTAGAGAGCAGATCTCTGATGAACAACCAACGGAAGAAGACG
GGTCGACCGTCTTTCAATATGCTGAAACGCGCGAGAAACCGCGTGTCAACTGGTTCAC
AGTTGGCGAAGAGATTCTCAAAAGGATTGCTCTCAGGCCAAGGACCCATGAAAATGG
TGATGGCTTTCATAGCATTCCTAAGATTTCTAGCCATACCCCCAACAGCAGGAATTTT
GGCTAGATGGAGCTCATTCAAGAAGAATGGAGCGATCAAAGTGTTACGAGGTTTCAA
AAAAGAGATCTCAAGCATGCTGAACATAATGAACAGGAGAAAAAGATCCGTGACCAT
GCTCCTCATGCTGCTGCCCACAGCCCTGGCGTTCCACTTGACCACACGAGGGGGAGAG
CCACACATGATAGTTACCAAGCAGGAAAGAGGAAAGTCACTCTTGTTTAAGACCTCT
GCAGGTGTCAACATGTGCACCCTCATTGCAATGGATTTGGGAGAGTTGTGTGAGGATA
CAATGACCTACAAATGCCCCCGGATCACTGAAGCGGAACCAGATGATGTTGATTGCT
GGTGCAATGCCACAGACACATGGGTGACCTATGGGACGTGTTCTCAGACTGGCGAAC
ACCGACGGGATAAACGCTCTGTCGCATTGGCCCCACACGTGGGGCTTGGTCTCGAAAC
AAGAACCGAAACGTGGATGTCCTCTGAAGGCGCCTGGAGACAGATACAAAAAGTGGA
GACTTGGGCCTTGAGACACCCAGGATTCACGGTGATAGCCCTTTTTCTAGCACATGCC
ATAGGAACATCCATTACCCAGAAAGGGATCATTTTCATTTTGCTGATGCTGGTGACAC
CATCAATGGCCATGCGATGCGTGGGAATAGGCAACAGAGACTTCGTTGAAGGACTGT
CAGGAGCAACATGGGTGGATGTGGTACTGGAGCATGGAAGCTGCGTCACCACCATGG
CAAAAAATAAACCAACATTGGACATTGAACTCTTGAAGACGGAGGTCACGAACCCTG
CCGTCTTGCGCAAACTGTGCATTGAAGCTAAAATATCAAACACCACTACCGATTCAAG
ATGTCCAACACAAGGAGAAGCTACACTGGTGGAAGAACAAGACGCAAACTTTGTGTG
TCGACGAACATTCGTGGACAGAGGCTGGGGTAATGGTTGTGGACTATTCGGGAAGGG
AAGCTTACTAACGTGTGCTAAGTTTAAGTGTGTGACAAAACTTGAAGGAAAGATAGTT
CAATATGAAAACTTAAAATATTCGGTGATAGTCACTGTCCACACTGGGGACCAGCACC
AGGTAGGAAATGAAACTACAGAACATGGAACAATTGCAACCATAACACCTCAAGCTC
CCACGTCGGAAATACAGCTGACTGACTACGGAGCCCTTACATTGGATTGCTCACCTAG
AACAGGGCTGGACTTTAATGAGATGGTGCTGTTGACAATGAAAGAGAAATCATGGCT
TGTCCACAAACAATGGTTTCTAGACTTACCATTGCCCTGGACCTCGGGGGCTTCAACA
TCTCAAGAGACTTGGAACAGACAAGATCTGCTGGTCACATTTAAGACAGCTCATGCA
AAGAAGCAGGAAGTAGTTGTACTGGGGTCACAAGAAGGAGCAATGCACACTGCGTTG
ACTGGGGCAACAGAAATCCAGACGTCAGGAACAACGACAATCTTCGCAGGACACCTG
AAATGTAGACTAAAAATGGATAAACTGACTTTAAAAGGGGTGTCATATGTGATGTGC
ACAGGTTCATTTAAGCTAGAGAAGGAAGTGGCTGAGACCCAGCATGGAACTGTCCTA
GTGCAGGTTAAATATGAAGGAACAGATGCACCATGCAAGATTCCAATTTCAACCCAA
GATGAGAAAGGAGTGACCCAGAATGGGAGATTGATAACAGCCAATCCCATAGTTACT
GACAAAGAAAAACCAGTCAACATTGAAACAGAACCACCCTTTGGTGAGAGCTACATC
GTGATAGGAGCAGGTGAAAAAGCTTTGAAACTAAGCTGGTTCAAGAAAGGAAGCAGC
ATAGGGAAAATGTTCGAAGCCACCGCCCGAGGAGCACGAAGGATGGCTATCCTGGGA
GACACAGCATGGGACTTCGGTTCCATAGGAGGAATGTTCACATCAGTGGGAAAATTG
GTACACCAGGTCTTTGGAACCGCATATGGGGTCTTGTTCAGCGGTGTTTCTTGGACCA
TGAAAATAGGAATAGGGATTCTGCTGACATGGCTAGGATTGAATTCAAGGAGCACGT
CACTCTCAATGACGTGCATTGCAGTTGGCATGGTCACACTGTACCTAGGAGTCATGGT
TCAAGCGGACTCAGGGTGTGTGATCAACTGGAAGGGCAGAGAACTCAAGTGTGGAAG
TGGCATCTTTGTCACTAATGAAGTCCATACTTGGACAGAGCAATACAAATTCCAGGCT
GACTCCCCCAAAAGACTATCAGCAGCCATTGGGAAGGCATGGGAGGAAGGCGTGTGT
GGAATTCGATCAGCCACACGTCTTGAGAACATCATGTGGAAGCAAATATCAAATGAA

TTGAACCACATTCTACTTGAAAATGACATGAAATTCACAGTGGTTGTAGGAGATGCCA
ATGGAATCTTGACCCAAGGAAAAAAGATGATTAGGCCACAACCTATGGAACACAAAT
ACTCATGGAAAAGCTGGGGAAAAGCCAAAATCATAGGAGCAGATACACAGAATACT
ACCTTTATCATCGATGGCCCAGACACTCCAGAATGCCCCGATGACCAGAGAGCATGG
AACATTTGGGAAGTTGAGGACTATGGGTTTGGAGTTTTCACGACAAACATATGGCTGA
AATTGCGTGACTCCTACACCCAAATGTGTGACCACCGGCTAATGTCAGCTGCCATCAA
GGACAGCAAGGCAGTCCATGCTGACATGGGGTACTGGATAGAAAGTGAAAAGAACG
AGACTTGGAAGCTAGCGAGAGCCTCCTTCATAGAAGTGAAGACATGCATCTGGCCAA
GATCCCACACTCTATGGAGCAATGGAGTCTTGGAAAGTGAGATGATAATCCCAAAGA
TATATGGAGGACCAATATCTCAGCACAACTACAGACCAGGGTATTTCACACAAACAG
CAGGGCCATGGCACCTAGGCAAGTTGGAACTAGACTTTAATTTGTGTGAAGGCACCA
CAGTTGTTGTGGATGAACATTGTGGAAATCGAGGTCCATCTCTTAGAACTACAACAGT
CACAGGAAAGATAATCCATGAATGGTGTTGCAGATCCTGCACGCTACCCCCCTTACGC
TTCAGAGGAGAAGACGGATGTTGGTATGGCATGGAAATCAGACCAGTTAAGGAGAAA
GAAGAGAATCTAGTCAGGTCAATGGTCTCTGCAGGGTCAGGAGAAGTGGACAGTTTT
TCATTGGGACTACTATGCGTATCAATAATGATTGAAGAAGTGATGAGATCCAGATGG
AGTAGAAAGATGCTGATGACTGGAACACTGGCTGTTTTCTTCCTCCTTATAATGGGAC
AACTGACATGGAATGATCTGATCAGGTTATGCATCATGGTTGGAGCTAATGTCTCAGA
CAGAATGGGGATGGGAACAACGTACCTAGCCTTAATGGCCACTTTTAAAATGAGACC
GATGTTCGCTGTTGGGCTATTATTTCGCAGATTAACATCCAGAGAGGTGCTCCTTCTCA
CAATTGGACTAAGCCTGGTGGCATCCGTGGAGCTACCAAATTCTTTAGAGGAGCTAGG
GGACGGACTTGCGATGGGCATCATGATGTTAAAATTGTTGACTGATTTTCAGTCACAC
CAGTTGTGGACTACCTTACTGTCTCTGACATTCATCAAAACAACCCTTTCACTGGATTA
TGCATGGAAGACAATGGCTATGGTATTGTCAATCGTATCTCTCTTTCCTCTATGTCTAT
CCACGACCTCTCAAAAAACAACATGGCTTCCGGTGCTGCTAGGATCCTTTGGATGTAA
ACCACTAACCATGTTTCTTATAACAGAAAACAAAATTTGGGAAAGGGAAAGTTGGCC
CCTCAATGAGGGAATTATGGCTGTTGGAATAGTCAGCATTTTACTAAGTTCACTCCTC
AAAAATGACGTGCCACTAGCCGGCCCACTAATAGCTGGAGGCATGCTGATAGCATGC
TACGTCATATCCGGAAGCTCAGCCGATTTATCATTGGAGAGAGCGGCTGAGGTCTCCT
GGGAAGAAGAAGCAGAACACTCTGGTGCCTCACACAACATACTAGTGGAGGTCCAAG
ATGATGGAACCATGAAAATAAAAGATGAAGAGAGAGATGATACGCTCACTATACTCC
TCAAAGCAACTTTGCTGGCAGTTTCAGGGGTGTACCCAATGTCAATACCAGCAACCCC
TTTTGTGTGGTATTTTTGGCAGAAAAAGAAACAGAGATCAGGAGTCTTGTGGGACACA
CCCAGCCCTCCAGAAGTAGAAAGAGCAGTTCTCGATGATGGTATCTATAGAATCTTGC
AAAGAGGACTGTTGGGTAGGTCCCAGGTAGGAGTGGGAGTTTTTCAAGACGGCGTGT
TCCACACAATGTGGCATGTCACCAGGGGGGCTGTCCTCATGTACCAAGGGAAGAGGC
TGGAACCAAGTTGGGCCAGTGTCAAAAAAGACCTGATCTCATATGGAGGAGGTTGGA
GGTTTCAAGGATCATGGAACACGGGAGAAGAAGTACAGGTGATAGCTGTTGAACCAG
GAAAAAACCCAAAGAATGTACAGACAACGCCGGGCACCTTCAAGACCCCTGAAGGCG
AAGTTGGAGCCATAGCCCTAGATTTCAAACCGGGCACATCTGGATCTCCCATCGTGAA
CAGAGAGGGAAAAATAGTAGGTCTTTATGGAAATGGAGTGGTGACAACAAGTGGAAC
CTATGTTAGTGCCATTGCCCAAGCCAAAGCGTCACAAGAAGGGCCTCTACCAGAGATT
GAGGACGAGGTATTTAAGAAAAGAAACCTAACAATAATGGACCTACATCCAGGATCA
GGAAAAACAAGAAGATATCTTCCAGCCATAGTCCGTGAGGCCATAAAAAGGAAACTG
CGCACGCTAATCTTGGCTCCCACAAGAGTTGTCGCCTCTGAAATGGCAGAGGCGCTCA
AAGGAATGCCAATAAGATATCAGACAACAGCAGTAAAGAGTGAACACACAGGAAAG

GAGATAGTCGATCTCATGTGCCACGCCACTTTCACCATGCGTCTCCTATCTCCAGTGA
GAGTTCCCAATTACAACATGATTATTATGGATGAAGCACACTTTACTGATCCAGCCAG
CATAGCGGCCAGAGGGTACATCTCAACCCGAGTGGGCATGGGTGAAGCAGCTGCGAT
CTTTATGACAGCCACTCCCCCAGGATCGGTAGAGGCCTTTCCACAGAGCAATGCAGTT
ATCCAAGATGAGGAAAGAGACATTCCTGAGAGATCATGGAACTCGGGCCACGACTGG
ATCACTGATTTCCCAGGAAAAACAGTCTGGTTTGTTCCAAGTATCAAATCAGGAAATG
ACATTGCCAACTGTTTAAGAAAAAACGGGAAACGGGTGATCCAATTGAGCAGAAAAA
CCTTTGACACTGAATACCAGAAAACAAAAAACAATGATTGGGACTATGTTGTCACAA
CAGACATTTCCGAAATGGGAGCAAATTTCCGGGCCGACAGGGTAATAGACCCAAGGC
GGTGCTTGAAGCCGGTAATACTAAAAGATGGCCCAGAGCGCGTCATTCTAGCCGGAC
CAATGCCAGTGACCGCGGCCAGTGCTGCCCAGAGGAGAGGAAGAATTGGAAGGAAC
CAAAACAAGGAAGGTGACCAGTATATTTATATGGGACAGCCTTTAAATAATGATGAG
GACCACGCTCATTGGACAGAAGCAAAAATGCTCCTTGACAATATAAATACGCCAGAA
GGGATTATCCCAGCCCTCTTTGAGCCAGAGAGAGAAAAGAGTGCAGCAATAGACGGG
GAGTACAGACTGCGGGGAGAAGCAAGGAAAACGTTCGTGGAGCTAATGAGAAGAGG
GGATCTACCAGTCTGGCTATCTTACAAAGTTGCCTCAGAAGGCTTCCAGTACTCCGAC
AGAAGATGGTGCTTTGACGGGGAAAGGAACAACCAGGTGTTGGAGGAGAACATGGA
CGTGGAGATCTGGACAAAGGAAGGAGAAAGAAAGAAATTGCGACCTCGCTGGTTAG
ATGCCAGAACCTACTCTGATCCATTGGCCCTGCGCGAGTTTAAAGAGTTCGCAGCGGG
AAGAAGAAGTGTCTCAGGTGACCTAATATTAGAAATAGGGAAACTTCCACAGCATTT
GACGCTAAGGGCCCAGAATGCCTTGGACAACCTGGTCATGTTGCACAATTCCGAACA
AGGAGGAAAAGCCTATAGACATGCTATGGAAGAACTACCAGACACCATAGAAACATT
GATGCTACTAGCCTTGATAGCTGTGCTGACTGGTGGAGTTACGTTGTTTTTCCTATCAG
GAAAAGGTCTAGGAAAAACATCCATTGGCCTACTTTGTGTAATGGCCTCAAGTGTACT
ATTATGGATGGCCAGTGTGGAGCCCCATTGGATAGCGGCCTCCATCATATTAGAGTTT
TTTCTGATGGTGCTGCTCATTCCAGAGCCAGACAGACAGCGTACTCCACAGGACAATC
AACTAGCATATGTGGTGATAGGTTTGTTATTCATGATATTGACAGTAGCAGCCAATGA
GATGGGATTACTGGAAACCACAAAGAAAGACCTGGGGATTGGCCATGTAGTCGCCGA
AAACCACCACCATGCCACAATGCTAGACATAGACTTACATCCAGCTTCAGCCTGGACC
CTCTATGCAGTAGCCACAACAATTATCACTCCCATGATGAGACACACAATTGAAAACA
CAACAGCAAACATTTCCCTGACAGCTATTGCAAATCAGGCGGCTATATTAATGGGACT
TGACAAGGGATGGCCAATATCGAAGATGGACATAGGAGTGCCACTTCTCGCTTTAGG
GTGCTATTCCCAGGTGAACCCATTGACACTGACAGCGGCGGTGTTGATGTTAGTGGCT
CATTATGCCATAATCGGACCTGGACTGCAAGCAAAAGCTACTAGAGAAGCCCAAAAA
AGGACTGCGGCCGGAATAATGAAAAATCCAACTGTAGACAGGATTGTTGCAATAGAT
TTGGATCCTGTGGTTTATGATGCAAAATTTGAAAAACAGCTAGGCCAAATAATGTTAC
TAATACTCTGCACATCACAGATTCTCTTGATGCGGACTACATGGGCCTTATGTGAATC
CATCACACTGGCTACTGGACCCTTGACCACTCTCTGGGAGGGATCTCCTGGAAGATTC
TGGAATACCACGATAGCAGTGTCCATGGCAAACATCTTCAGGGGAAGTTATCTAGCA
GGAGCAGGTCTGGCCTTCTCATTAATGAAATCTTTAGGAGGAGGTAGGAGAGGTACG
GGAGCTCAAGGGAAAACATTGGGAGAAAAATGGGAAAGACAATTAAACCAACTGAG
CAAGTCAGAATTTAACACCTACAAAAGGAGTGGAATTATGGAGGTGGACAGATCTGA
AGCCAAGGAGGGACTGAAAAGAGGAGAAACAACCAAACATGCAGTGTCGAGAGGAA
CAGCCAAACTTAGGTGGTTTGTGGAGAGGAACCTTGTGAAACCAGAAGGGAAAGTGA
TAGACCTCGGTTGTGGAAGAGGTGGCTGGTCATATTACTGCGCTGGGCTGAAGAAGG
TCACAGAAGTGAAAGGATACACAAAAGGAGGACCCGGTCATGAGGAACCTATCCCAA

TGGCGACTTATGGATGGAACATAGTGAAGCTATACTCTGGAAGGGATGTATTCTTTAC
ACCACCTGAGAAATGTGACACCCTTTTGTGTGATATTGGTGAGTCCTCTCGAATCCCA
ACTATAGAAGAAGGAAGAACGTTACGTGTTCTAAAGATGGTGGAACCATGGCTCAGA
GGAAATCAGTTTTGCATAAAAGTCCTAAATCCTTACATGCCAAGTGTGGTAGAAACTC
TGGAGCAAATGCAAAGAAAACATGGAGGGATGCTAGTGCGAAACCCACTCTCAAGAA
ATTCTACCCATGAAATGTACTGGGTTTCATGTGGAACAGGAAACATTGTGTCGGCAGT
GAACATGACATCCAGAATGTTGCTGAATCGGTTCACAATGGCTCACAGGAAGCCAAC
GTATGAAAGAGACGTGGACTTGGGCGCTGGAACAAGACATGTGGCAGTGGAACCAGA
GATAGCCAACCTAGACATCATTGGCCAGAGGATAGAGAACATAAAAAATGAACACAA
GTCAACATGGCATTATGATGAGGACAATCCATACAAAACATGGG

Appendix 7: Dengue virus serotype 1 genotypes

Table S2: Representative dengue virus 1 genotypes used in this study

s/n	Accession	Year collection	Country	Genotype
1	AF298808.1	2000	Djibouti	GI
2	AY726552.1	2002	Myanmar	GI
3	FR666924.1	2005	Malaysia	GI
4	EU359008.1	2007	China	GI
5	JF967803.1	2008	Thailand	GI
6	JF967877.1	2009	Indonesia	GI
7	JF960223.1	2010	Singapore	GI
8	KT824981.1	2011	Thailand	GI
9	KC848580.1	2011	Somalia	GI
10	AB915381.1	2012	Indonesia	GI
11	KT824985.1	2012	Cambodia	GI
12	KT825002.1	2013	Australia	GI
13	KT825008.1	2013	Papua New Guinea	GI
14	KJ806959.1	2014	Singapore	GI
15	KT825029.1	2014	Sri Lanka	GI
16	MG560209.1	2014	China	GI
17	JF297570.1	1960	Thailand	GII
18	AF425629.1	1963	Thailand	GII
19	AF180817.1	2000	USA	GII
20	MT076935.1	2015	Kenya	GII
21	AF425622.1	1972	Malaysia	GIII
22	AY722803.1	1998	Myanmar	GIII
23	EU863650.1	2002	Chile	GIV
24	FJ196842.1	2003	China	GIV
25	GQ868602.1	2004	Philippines	GIV
26	JN415516.1	2005	Philippines	GIV
27	EU448403.1	2006	Vietnam	GIV
28	EU448412.1	2006	Madagascar	GIV
29	KJ946238.1	2008	Philippines	GIV
30	KM216671.1	2010	Indonesia	GIV
31	JN029810.1	2010	China	GIV
32	KT824972.1	2011	Indonesia	GIV
33	MG894881.1	2015	Philippines	GIV
34	AF425625.1	1968	Nigeria	GV
35	AY713473.1	1971	Myanmar	GV
36	JQ922545.1	1982	India	GV
37	AF425626.1	1991	Peru	GV
38	DQ285562.1	1993	Comoros	GV
39	FJ205875.1	1995	USA	GV
40	GU131833.1	2000	Venezuela	GV
41	AF298807.1	2000	Nigeria	GV
42	AB519681.1	2001	Brazil	GV
43	AY762084.1	2004	Singapore	GV

s/n	Accession	Year collection	Country	Genotype
44	AY858983.1	2004	Indonesia	GV
45	DQ285559.1	2004	Reunion	GV
46	JF297581.1	2005	India	GV
47	JQ922548.1	2005	India	GV
48	DQ016655.1	2005	India	GV
49	EU448402.1	2006	Indonesia	GV
50	KJ189319.1	2007	Mexico	GV
51	GQ199859.1	2008	Nicaragua	GV
52	JF960211.2	2009	Singapore	GV
53	JF967897.1	2010	Malaysia	GV
54	KC863940.1	2010	India	GV
55	JX669466.1	2010	Brazil	GV
56	JF519855.1	2010	USA	GV
57	JQ692085.1	2010	India	GV
58	KT824975.1	2011	India	GV
59	MG877557.1	2012	Gabon	GV
60	KY581728.1	2012	India	GV
61	JN036371.1	2012	Bangladesh	GV
62	KF973470.1	2012	Nicaragua	GV
63	KM277613.1	2013	Angola	GV
64	MH680017.1	2013	Singapore	GV
65	MG560230.1	2014	China	GV
66	KT428614.1	2014	China	GV
67	KT239343.1	2014	Pakistan	GV
68	KT825063.1	2014	Australia	GV
69	KT239344.1	2014	Pakistan	GV
70	KY021900.1	2015	India	GV
71	LC360648.1	2015	Japan	GV
72	MG894884.1	2015	Thailand	GV
73	MF033237.1	2015	Singapore	GV
74	MF033253.1	2016	Singapore	GV
75	MW243006.1	2016	DRC	GV
76	MH594886.1	2017	India	GV
77	MH594885.1	2017	India	GV
78	MT261954.1	2017	Burkina Faso	GV
79	MW288036.1	2018	Senegal	GV
80	LC485151.1	2019	Japan	GV
81	DENV-1/TAN/DAR-04	2019	Tanzania	GV
82	MW261832.1	2019	China	GV
83	MN923102.1	2019	China	GV
84	MN923101.1	2019	China	GV
85	MN923096.1	2019	China	GV
86	MN923085.1	2019	China	GV
87	MW243062.1	2019	Senegal	GV

Appendix 8: Permissions to publish

UNITED REPUBLIC OF TANZANIA

MINISTRY OF HEALTH, COMMUNITY DEVELOPMENT, GENDER, ELDERLY AND CHILDREN

NATIONAL INSTITUTE FOR MEDICAL RESEARCH

In reply please quote:

Ref. No: NIMR/HQ/P.12 VOL XXXII/63 Date: 10th February 2021

Dr Leonard Mboera,

Sokoine University of Agriculture (SUA)

SACIDS Foundation for One Health

P.O. Box 3297, Chuo Kikuu

MOROGORO.

Dear Dr. Mboera,

RE: PERMISSION TO PUBLISH

Reference is made to your request for permission to publish dated 5th February with reference number MB/P2P/CEVBD/02/2021/01.

2. Permission has been granted to publish a manuscript titled: "**Seroprevalence and associated risk factors of mosquito-borne viral infections in diverse ecological zones of Tanzania**" by authors: Gaspary O. Mwanyika, Calvin Sindato, Sima Rugarabamu, Susan F. Rumisha, Esron D. Karimuribo, Gerald Misinzo, Mark M. Rweyemamu, Muzamil Abdeli Hamid, Najmul Haider, Francesco Vairo, Richard Kock, Leonard E.G. Mboera for PANDORA-ID-NET.

3. Please submit an electronic copy of the published manuscript to the National Institute for Medical Research through email publications@nimr.or.tz.

Dr. Ndekya Maria Oriyo

DIRECTOR OF RESEARCH INFORMATION, TECHNOLOGY AND COMMUNICATION

Headquarters: 3 Barack Obama Drive, P.O. Box 9653, 11101 Dar es Salaam, Tanzania,
Email: info@nimr.or.tz, Website: www.nimr.or.tz

NIMR Amani P.O. Box 81 Muheza, Tanga Tanzania Email: amani@nimr.or.tz

NIMR Dodoma Benjamin Mkapa Hospital P.O. Box 806 41119 Dodoma, Tanzania Email: dodoma@nimr.or.tz

NIMR Mbeya Hospital Hill P.O. Box 2410 Mbeya, Tanzania Email: nimr-mmrp@ihcu.tz

NIMR Muhimbili P.O. Box 3436 Dar es Salaam, Tanzania Email: muhimbili@nimr.or.tz

NIMR Mwanza P.O. Box 1462 Mwanza Tanzania Email: mwanza@nimr.or.tz

NIMR Ngorongoro P.O. Box 514 USA River, Arusha Tanzania Email: ngorongoro@nimr.or.tz

NIMR Tabora P.O. Box 482 Tabora Tanzania Email: tabora@nimr.or.tz

NIMR Tanga P.O. Box 5004 Tanga Tanzania Email: tanga@nimr.or.tz

NATIONAL INSTITUTE FOR MEDICAL RESEARCH
HEADQUARTERS

Telephone: +255-22-2121400
Telefax: +255-22-2121360
E-mail: hq@nimr.or.tz
Website: www.nimr.or.tz

3 Barack Obama Drive
P.O. Box 9653
11101 Dar es Salaam
Tanzania

Your Ref: PVM/D/2017/0019
Our Ref: NIMR/HQ/P.12 VOL XXXI/

23rd October 2020

Gaspary Mwanyika
Department of Veterinary Microbiology, Parasitology and Biotechnology,
College of Veterinary Medicine & Biomedical Sciences
P.O. Box 3019,
Chuo Kikuu, Morogoro
Sokoine University of Agriculture
MOROGORO.

Dear Gaspary Mwanyika,

RE: PERMISSION TO PUBLISH

Reference is made to your request for permission to publish.

2. Permission has been granted to publish a manuscript titled: "**Circulation of dengue serotype 1 viruses during the 2019 outbreak in Dar es Salaam, Tanzania**" by authors: Gaspary O. Mwanyika, Leonard E.G. Mboera, Sima Rugarabamu, Mariam Makange, Calvin Sindato, Julius J. Lutwama, Janusz T. Paweska and Gerald Misinzo.

3. Please submit an electronic copy of the published manuscript to the National Institute for Medical Research through email publications@nimr.or.tz.

Dr. Ndekya Maria Oriyo
DIRECTOR OF RESEARCH INFORMATION, TECHNOLOGY AND COMMUNICATION