# WHOLE GENOME CHARACTERIZATION OF ROTAVIRUS GROUP A STRAINS ISOLATED FROM CHILDREN ADMITTED IN KILIFI COUNTY HOSPITAL FOLLOWING ROUTINE VACCINATION IN KENYA

**TIMOTHY MAKORI ONYINGE**

**MASTER OF SCIENCE**

**(Biochemistry)**

**JOMO KENYATTA UNIVERSITY**

**OF**

**AGRICULTURE AND TECHNOLOGY**

**2024**

# Whole Genome Characterization of Rotavirus Group A Strains Isolated from Children Admitted in Kilifi County Hospital Following Routine Vaccination in Kenya

**Timothy Makori Onyinge**

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

**2024**

## DECLARATION

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

Signature …… Date … ……..

**Timothy Onyinge Makori**

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

Signature …… Date … ……..

**Dr. Joel L. Bargul, PhD**

**JKUAT, Kenya**

Signature …… Date … ……..

**Dr. Charles N. Agoti, PhD**

**KEMRI, Kenya**

## DEDICATION

I dedicate this thesis to my dear parents, James Makori and Naomi Nyaboke, for your unconditional love and instilling in me the spirit of hard work, discipline, and desire to pursue higher studies. To my wife, Rhodah Mosoti, and Son, Nevins James, you are a big inspiration and source of strength to me.

## ACKNOWLEDGEMENT

I would like to extend my sincere gratitude and appreciation to everyone who contributed to the completion of my Msc thesis. First, I am immensely grateful to my supervisors Dr Charles Agoti (KWTRP) and Dr Joel Bargul (JKUAT), for their immense support, encouragement, invaluable insights, and mentorship. I am truly grateful to Dr Charles Agoti and Prof. D. James Nokes who conceived the project and gave me the opportunity to carry it out under your guidance in the Virus Epidemiology and Control (VEC) group of KEMRI-Wellcome Trust Programme (KWTRP). Your expertise, the scientific skills I have gained, and mentorship will go a long way in shaping my future career. Many thanks to Dr Agoti and Dr Bargul for the encouragements to keep me going and navigate through the challenges I encountered during the project.

I would also like to sincerely thank KWTRP staff including Arnold Lambisia, Mike Mwanga, Martin Mutunga, Clement Lewa, Robinson Cheruiyot and Zaydah R de Laurent who supported me in designing the laboratory assays and data analysis. I am deeply grateful to the KWTRP training department, Initiative Develop African Research Leaders (IDEAL) for funding my MSc studentship. Sincere thanks to Prof Sam Kinyanjui, Dr Dorcas Mbuvi, and Liz Murabu for the support and encouragement throughout my MSc studentship at KWTRP. Many thanks to IDEAL MSc cohort of 2017/2018 for the cheers to navigate through the challenges encountered during the project.

Many thanks to my parents, James Makori and Naomi Nyaboke for your love, prayers, and emotional support. Thanks to my siblings for the encouragements and emotional support. Special thanks to my wife Rhodah Mosoti and son Nevins James for the unwavering love, support, and motivation.

I would also acknowledge Jomo Kenyatta University of Agriculture and Technology for the opportunity to undertake my academic aspirations. Finally, I thank God for the strength and provision of opportunities and resources during this project.

Funding: “This work was supported by the Wellcome Trust (102975 and 203077). Timothy O. Makori and Dr. Charles N. Agoti. Were funded by the Initiative to Develop African Research Leaders (IDEAL) through The Developing Excellence in Leadership, Training and Science (DELTAS) Africa Initiative (DEL-15-003). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)’s Alliance for Accelerating Excellence in Science in Africa and supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (107769/Z/10/Z) and the UK government. The views expressed in this thesis are those of the author(s) and not necessarily those of the AAS, the NEPAD Agency, the Wellcome Trust, or the UK government.”

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF APPENDICES

## ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **AGE** | Acute Gastroenteritis |
| **cDNA** | Complimentary DNA |
| **DNA** | Deoxyribonucleic Acid |
| **ds-Cdna** | Double-stranded cDNA |
| **EIA** | Enzyme Immuno Assay |
| **GAVI** | Global Alliance for Vaccines and Immunization |
| **GoK** | Government of Kenya |
| **KCH** | Kilifi County Hospital |
| **KEMRI** | Kenya Medical Research Institute |
| **KHDSS** | Kilifi Health and Demographic Surveillance System |
| **KNH** | Kenyatta National Hospital |
| **KWTRP** | KEMRI Wellcome Trust Programme |
| **mRNA** | messenger-RNA |
| **NIPs** | National Immunization Programmes |
| **PCR** | Polymerase Chain Reaction |
| **RNA** | Ribonucleic Acid |
| **RV** | Rotavirus |
| **RVA** | Rotavirus group A |

| | |
|---|---|
| **SERU** | Scientific and Ethics Review Unit |
| **ViPR** | RVA Virus Pathogen Resource |
| **WHO** | World Health Organization |

## ABSTRACT

The introduction of rotavirus vaccines into national immunization programmes (NIPs) of many countries globally has considerably reduced the rotavirus group A (RVA) disease burden. Kenya introduced the monovalent Rotarix® vaccine (G1P[8] backbone) into its national immunization programme in July 2014 and a ~60% decrease in rotavirus group A-associated diarrhoea hospitalization was reported two years post-vaccine introduction. Despite this success, rotavirus group A continues to be among the leading causes of severe diarrhoea in young children hospitalized at Kilifi County Hospital (KCH) on the coast of Kenya. Coincidentally, studies elsewhere have shown an increase in Rotarix® vaccine partially or fully heterologous genotypes in countries implementing Rotarix® vaccine use. Therefore, this study aimed to study the genetic relatedness, origin, and evolution of rotavirus group A G2P[4] strains circulating in the Rotarix® pre- and post-vaccination period in Kilifi County and determine whether the introduction of the Rotarix® vaccine in Kenya impacted the antigenic diversity of G2P[4] RVA strains circulating in Kilifi County, coastal Kenya. Whole genome sequencing of 32 pre-(January 2012 to June 2014) and 31 post-vaccine -(July 2014- December 2018) periods rotavirus group A G2P[4] vaccine heterologous strains infecting children (<13 years old) admitted to KCH was done to understand the genetic diversity of these strains and determine their potential origins (source) and local transmission patterns. The results found that both pre- and post-vaccine G2P[4] strains had a typical DS-1-like genomic backbone (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2). The Kilifi G2P[4] strains classified into lineages II (for VP4 segment), lineage IV (for VP7, VP4, VP2, NSP1, and NSP5 sequences), lineage V (for VP6, VP1, VP3, NSP2, and NSP3 segments), lineage VI (for the NSP4 segment) and lineage VII (for the VP3 and NSP4 segments). Phylogenetically, the Kilifi pre- and post-vaccine period sequences majorly formed separate clusters on the global phylogeny across the 11 genome segments, implying distinct virus populations were in circulation pre- and post-vaccine periods. In addition, the Kilifi strains mainly clustered on the same branch separate from other global strains, suggesting observed local strains were most likely persisting in the Kilifi or surrounding regions with an accumulation of genetic changes over time. However, the Kilifi pre- and post-vaccine strains had conserved amino acid changes on the antigenic epitopes of the VP7 and VP4 surface proteins, suggesting that replacement of the pre-vaccine population was unlikely due to immune escape. In conclusion, the study indicated that the G2P[4] strains circulating in Kilifi pre-(January 2012 to June 2014) and post-(July 2014- December 2018) periods are genetically different but antigenically similar and constitute locally circulating strains. Further studies should be conducted to determine the genomic epidemiology of G2P[4] strains by extending the geographical sampling area by including various health facilities across Kenya and from non-hospitalized infected persons in the community to have a representative sample across the entire Kenyan population.

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background Information

This study investigated the genomic epidemiology of rotavirus group A (RVA) genotype G2P[4] circulating in the pre- (January 2012-June 2014) and post- (July 2014-December 2018) rotavirus vaccination periods in Kilifi, coastal Kenya. RVA is the leading cause of acute gastroenteritis (AGE) associated with approximately 122,000 to 215,000 deaths annually among under-five-year-old children globally (Barros De Arruda *et al.,* 2022). RVA is highly contagious and is transmitted through contact with faecal matter or contaminated surfaces, food, or water (Crawford *et al.,* 2017). Symptoms of RVA infection include watery diarrhoea, vomiting, fever, and abdominal pain, and can lead to severe dehydration, malnutrition, and death if left untreated (Crawford *et al.,* 2017). Children from low- and middle-income settings bear a greater RVA disease burden compared to high-income countries (Troeger *et al.,* 2018). A study on the global burden of RVA disease in 2016 indicated that 100 per 100, 000 children aged below five years die from RVA-related diarrhoea in Kenya (Tate *et al.,* 2016).

RVA-AGE disease can be effectively managed through providing affected children with oral rehydration therapy to prevent progression of the disease into severe form and prevent death which may arise from dehydration or through vaccination (Crawford *et al.,* 2017). For severe rehydration, initial intravenous rehydration therapy is provided followed by immediate provision of oral rehydration (Crawford *et al.,* 2017; Florez *et al.,* 2020). Improving hygiene in low-income settings may not significantly reduce the burden on RVA disease as the incidence of hospitalization due to RVA-related diarrhoea was about 40% in both the developed and developing countries prior to introduction of RVA vaccine (Crawford *et al.,* 2017). Thereby, vaccination against RVA is the most effective measure for management of RVA disease (Donato & Bines, 2021).

Since 2009, the World Health Organization (WHO) recommended inclusion of licenced RVA vaccines into national immunization programs (NIPs) of all countries globally with the aim of mitigating RVA-associated disease burden (WHO, 2021). In July 2014, Kenya implemented the use of a two-dose live-attenuated Rotarix® (GlaxoSmithKline Biologicals, Belgium), a monovalent vaccine based on the G1P[8] genotype into her NIP (Wandera *et al.,* 2018). Consequently, a significant decline in hospitalization associated with RVA-diarrhoea in children aged below five years of > 80% (95% CI, 46–93%) was reported in the second year and sustained to the third year post-vaccine introduction (Otieno *et al.,* 2020), a two- and one-dose Rotarix® vaccine effectiveness of 64% and 54%, respectively (Khagayi *et al.,* 2020), and a remarkable increase in non-Rotarix® vaccine type strains including G2P[4], G3, and G9 (Gikonyo *et al.,* 2020; Mwanga *et al.,* 2020)

Beginning January 2023, the Kenyan Ministry of Health (MoH) decided to switch from Rotarix® to Rotavac® (a live-attenuated monovalent) vaccine following transition from support by Global Alliance for Vaccines and Immunization (GAVI) to support by the Government of Kenya (GoK). Studies in India showed that Rotavac® vaccine efficacy against severe rotavirus gastroenteritis in the first year of life was 53.6% (95% CI, 35.0-66.9) (Bhandari *et al.,* 2014) and 55.1% (95% CI 39.9-66.4) in the second year of life (Bhandari *et al.,* 2014). A recent study conducted in Palestine revealed that following a two-year transition from Rotarix®, the epidemiological outcomes of the Rotavac vaccination were comparable to those of Rotarix®, where the RVA-positivity rate declined by 67.1 % (from 38.2 % to 13 %) following Rotarix introduction and maintained at 15% after transition to Rotavac vaccine (Rennert *et al.,* 2023). In addition, the results showed a measurable decline in the incidence of acute gastroenteritis among children under five years old, and a change in predominant genotypes from G9P[8] and G12P[8] during the Rotarix® vaccine period to G2P[4] two years post-Rotavac® vaccine introduction (Rennert *et al.,* 2023).

In 2016, an upsurge of G2P[4] strains among children admitted to Kilifi County Hospital (KCH), coastal Kenya in the post-Rotarix® vaccine period (Mwanga *et al.,* 2020) was observed (Figure 1.0) between June and November (post-vaccine period)

while only a small peak was reported previously between May and September 2012 (pre-vaccine period). Genotyping of the samples was done based on partial gene sequencing followed by the binary classification scheme (Mwanga *et al.*, 2020). The current study employed whole-genome analysis to understand the complete genotypic constellation of these pre- and post-vaccine G2P[4] strains, determine the genetic relatedness, establish the source and transmission pattern of the locally circulating strains, and establish whether the Rotarix® vaccine has a direct effect on the antigenic diversity of the local strains.

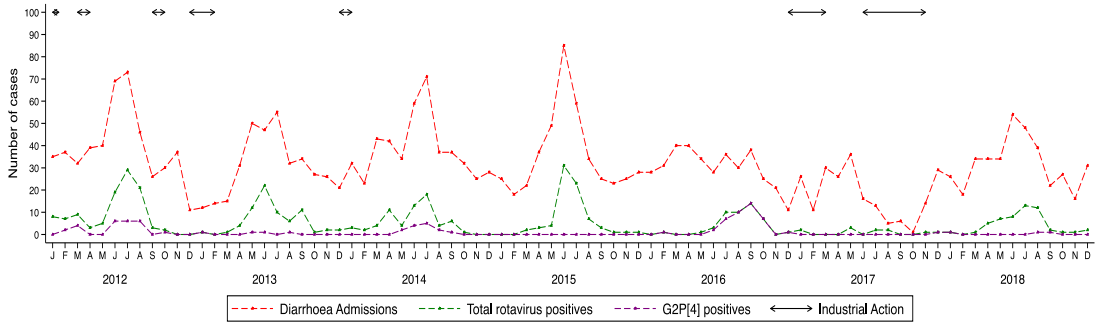


**Figure 1.1: Temporal plot showing total diarrhoea admissions at Kilifi County Hospital, total Rotavirus A detections, and G2P[4] detections from January 2012 to December 2018. Total diarrhoea admissions are shown by the red dotted line, total rotavirus A detections are shown by the green dotted line, while the purple dotted line shows total G2P[4] detections. The solid arrow lines at the top indicate periods of industrial actions of healthcare workers at KCH. Genotyping was done using the binary classification system via Sanger sequencing** (Mwanga et al., 2020).

### 1.2 Statement of the Problem

Group A Rotaviruses (RVAs) are ubiquitous and primarily cause life-threatening diarrhoea in children aged zero to five years (Barros De Arruda *et al.*, 2022). Despite the inclusion of RVA vaccines into the NIPs of 122 countries globally (IVAC 2022), the prevalence of RVA infection in hospitalized children is approximately 30-50 %, with > 80 % of children with fatal RVA infections living in developing countries

(Reiner *et al.,* 2020; Were *et al.,* 2022). In Kenya, in spite of RVA vaccination since July 2014, RVA strains have been reported in circulation albeit at lower frequencies, with a significant increase in vaccine heterotypic strains, including G2P[4] and G3 (Mwanga *et al.,* 2020). Genotype G2P[4] has been reported to increase significantly in countries using the Rotarix® vaccine (Bibera *et al.,* 2020; Leshem *et al.,* 2014). Therefore, whole genome sequencing of the G2P[4] circulating in Kilifi contributes to the discussion on the consequences of rotavirus vaccination on RVA genetic diversity.

### 1.3 Research Questions

(i) Does the agnostic sequencing method compare with the amplicon-based sequencing method in RVA G2P[4] genome recovery?

(ii) Does Rotarix® vaccine introduction contribute to the G2P[4] RVA genetic and antigenic diversity in Kilifi County?

(iii) What are the sources of G2P[4] RVA strains causing AGE epidemics in Kilifi County in the pre- and post-vaccination era?

### 1.4 Justification of the Study

The efficacy of the Rotarix® vaccine in developed countries is 83% while in Kenya, a low-economic setting it is 64% (Burnett *et al.,* 2020). Reductions in; vaccine immunogenicity and protection provided by natural RVA infection mostly contribute to reduced Rotarix® vaccination efficacy in underdeveloped settings (Lopman *et al.,* 2012). Surveillance of RVA by the KEMRI Wellcome Trust Programme (KWTRP) at Kilifi, Kenyan coast since 2009 (Khagayi *et al.,* 2020) indicates a significant increase in the incidence of G2P[4] strains among Kenyan children (Mwanga *et al.,* 2020) despite the inclusion of the WHO-prequalified Rotarix® vaccine into the NIP of Kenya in July 2014 (Wandera *et al.,* 2018). Genotyping of these strains was done through Sanger sequencing of the VP7 and VP4 surface proteins (Mwanga *et al.,* 2020). Studies involving whole genome sequencing of G2P[4] strains have revealed the antigenic diversity, and established the origin and overall genetic diversity of the strains circulating in a particular region (Donato *et al.,* 2021; Mwangi *et al.,* 2022). Therefore, whole genome sequencing of Kilifi G2P[4] strains reveals; whether

Rotarix® vaccine introduction has a direct effect on G2P[4] strains circulating post-vaccine period, the introduction sources of pre- and post-vaccination G2P[4] strains, their genetic diversity, and the antigenic differences between the pre- and post-vaccine period strains. Besides, this study contributed to the number of near-full-length G2P[4] genomes from Kenya.

## 1.5 Hypothesis

### 1.5.1 Null Hypotheses

The lower efficacy of the Rotarix® vaccine in Kilifi County (~64%) is not related to circulating G2P[4] antigenic diversity and that the G2P[4] RVA strains originating from regions outside Kilifi County are circulating in the County, during pre- and post-vaccine periods.

## 1.6 Objectives

### 1.6.1 General Objective

To determine whole genome characterization of RVA G2P[4] strains circulating in the Rotarix® pre- and post-vaccination period in Kilifi County and determine whether the introduction of the Rotarix® vaccine in Kenya impacted the antigenic diversity of G2P[4] RVA strains circulating in Kilifi County, coastal Kenya.

### 1.6.2 Specific Objectives

(i) To compare two whole genome sequencing methods for G2P[4] RVA strains and establish a WGS method for G2P[4] RVA strains at Kilifi, KWTRP laboratories and use it to obtain sequence data from RVA positive samples classified as G2P[4] by the binary system.

(ii) To determine whether Rotarix® vaccine introduction in Kenya has a direct or indirect impact on the antigenic diversity of G2P[4] strains circulating in Kilifi using the antigenic epitopes of the surface proteins (VP4 and VP7).

(iii) To determine the genetic diversity, genetic relatedness, origin and patterns of transmission of G2P[4] RVA strains circulating in Kilifi County, coastal

Kenya in the pre- (January 2012-June 2014) and post-vaccine (July 2014-December 2018) periods using whole genome sequencing.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 RVA Disease Burden

Globally, RVA is the principal pathogen of diarrhoea-associated morbidity and mortality in under five-year-old children (Troeger *et al.,* 2017). RVA causes an estimated 258,000,000 infections and 122,000 to 215,000 deaths among children aged below five years old globally (Barros De Arruda *et al.,* 2022). This is despite the inclusion of the WHO-prequalified RVA vaccines (WHO, 2021) in 122 countries globally (IVAC, 2022). A greater burden of RVA disease (80%; 104, 743 deaths) is experienced in sub-Saharan Africa (Troeger *et al.,* 2018). In Kenya, a Post-Rotarix® vaccine study involving children admitted with diarrhoea to KCH indicated no significant decline in RVA positivity rate post-vaccine introduction (27.4% vs 23.5%, $P = 0.253$) (Agoti *et al.,* 2022) coinciding with an increase in G2P[4] and G3P[8] non-Rotarix® vaccine genotypes (Mwanga *et al.,* 2020).

## 2.2 The Structure, Genome Organization, and Protein Composition of Rotavirus

Rotavirus (RV) has a wheel-shaped characteristic appearance as seen under an electron microscope, hence its name rota signifying a wheel (Figure 2.1A) (Prasad & Chiu, 1994). It was first discovered by Bishop Research group in Australia in children with gastroenteritis (Bishop *et al.,* 1973). RV is a non-enveloped double-stranded RNA virus that belongs to the Reoviridae family, Sedoreoviridae sub-family, genus Rotavirus (Sadiq *et al.,* 2018). A mature RV virion has a three-shelled icosahedral capsid of 70-100nm diameter consisting of the outer, middle, and inner layers as shown in figure 2.1A (King *et al.,* 2011). The outer layer is comprised of the viral proteins; VP7 and VP4, with the VP4 protein protruding in form of spikes (Figure 2.1A) (Sadiq *et al.,* 2018). The VP6 protein make up the middle layer, while the inner/core layer is comprised of the core shell VP2 protein enclosing the VP3 and VP1 proteins (Figure 2.1A) (Sadiq *et al.,* 2018). The double stranded RNA genome

(~18,555 base pairs) is made up of 11 segments, which is packaged within the inner layer (Figure 2.1A) (Desselberger, 2014). Each segment is a gene coding for six structural proteins (VP)-VP1, VP2, VP3, VP4, VP6 and VP7, and six or five non-structural proteins ((NSP)-NSP1, NSP2, NSP3, NSP4, NSP5, and NSP6) (Figure 2.1B) (Desselberger, 2014). Ten gene segments are monocistronic, while the eleventh genome segment is bicistronic i.e., in some RV strains encodes two genes (NSP5 and NSP6) (Sadiq *et al.*, 2018). These RV segments are numbered 1-11 based on sequence lengths (Desselberger, 2014; King *et al.*, 2011) as shown in figure 2.1C.

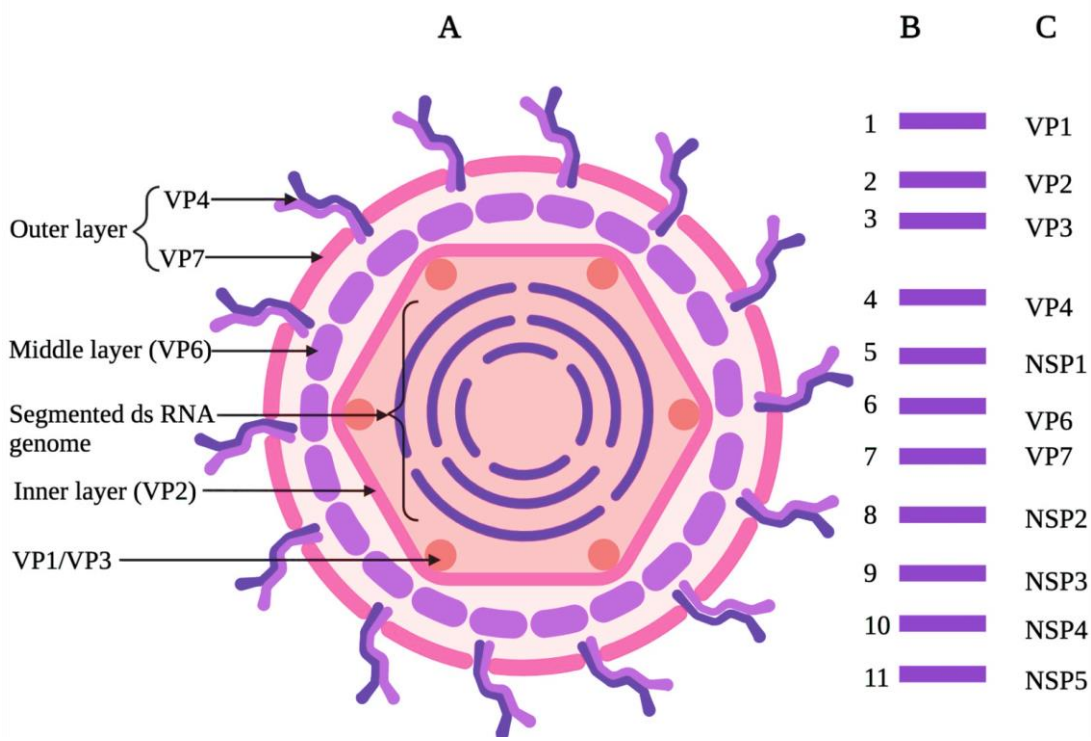


**Figure 2.1: The general structure of Rotavirus virion. A: Virion Schematic: An overview of the various structural proteins of the RV virion. The virion is organized into three concentric capsid layers, outer shell (capsid proteins VP4 and VP7), the middle shell (VP6 protein) and the inner shell (VP2 protein). C: Encoded proteins; shows the respective protein encoded by each segment.**

The VP1 protein serves as the RNA-dependent RNA polymerase of RV, while VP3 protein functions as a guanylyl transferase for the mRNA capping enzyme (Desselberger, 2014). These two proteins constitute the replication enzyme of RV and are in intense contact with the genomic double-stranded RNA segments via the VP1 protein (Desselberger, 2014). The VP2 protein functions as a co-factor for the replicase activity of VP1 RNA polymerase by recruiting the VP1/3 complex along with a double-strand RNA template (King *et al.,* 2011). VP4 (protease sensitive) protein facilitates RV attachment and entry into host enterocyte cells, virulence, and P serotype and host range determination, while the VP7 glycoprotein is implicated in $Ca^{2+}$ binding to facilitate RV entry into host cells and G serotype determination (Desselberger, 2014). In addition, the VP7 and VP4 proteins harbour various antigenic epitopes that trigger production of host-neutralization antibodies upon RV

infection (Desselberger, 2014). The VP6 protein is required for A-J grouping and sub-classifying into sub-groups (I, II, I + II, and "non-I, non-II") on the basis of their reactions to monoclonal antibodies (Sadiq *et al.,* 2018).

The five or six non-structural proteins (NSPs) are only expressed in cells infected with RVs (Desselberger, 2014). During infection with RV, NSP1 docks the interferon regulatory factor-3 to inhibit interferon response (Desselberger, 2014). NSP2 and NSP5 work together in synthesis and packaging of RV RNA, and replication of the genome, while NSP3 is implicated in the synthesis of viral proteins and shutting off synthesis of host proteins (Desselberger, 2014; King *et al.*, 2011). NSP4, the enterotoxin protein, is involved in diarrhoea development and serves as a viroporin facilitating viral replication, while NSP6 protein interacts with NSP5 facilitating RNA binding (Desselberger, 2014).

### 2.3 Classification of Rotaviruses

There are three classification schemes for RVs, namely VP6 antigen-based grouping, and the dual classification and full-genome based classification systems for group A RVs (RVA) (Sadiq *et al.,* 2018), as discussed below.

#### 2.3.1 VP6 Protein-Based Classification

On the basis of antigenic reactivity to VP6 protein, genus Rotavirus is classified into ten serological groups (RVA-RVJ) defined as RV species (Bányai *et al.,* 2017; Sadiq *et al.,* 2018). RVA-RVH have been accepted by the International Committee on Taxonomy of Viruses, while RVI and RVJ are pending confirmation (Sadiq *et al.,* 2018). Species RVA, RVB, RVC, and RVH cause acute gastroenteritis in humans and animals, while species RVD, RVE, RVF, RVG, RVJ, and RVI infect animals only (Bányai *et al.,* 2017; Mihalov-Kovács *et al.,* 2015; Sadiq *et al.,* 2018). RVA is the most important epidemiologically in both humans and animals (Leshem & Lopman, 2018). RVAs are further classified into four subgroups, namely SGI, SGII, SGI + II, and SG nonI/noni, based on VP6 sequence data and reactivity of VP6 protein to specific monoclonal antibodies (Sadiq *et al.,* 2018).

### 2.3.2 Binary Classification System for RVA

The two outer capsid proteins, VP7 (glycoprotein) and VP4 (protease sensitive) are used to classify RVAs into G and P genotypes, respectively (Sadiq *et al.,* 2018). Currently, up to 42G and 58 P RVA genotypes have been identified in humans (RCWG 2021).

### 2.3.3 Full Genome-Based Classification System

Matthijnssens and colleagues introduced a more detailed sequence-based classification and nomenclature scheme for classification of RVAs in 2008 (Matthijnssens, *et al.,* 2008). This classification system assigns RVA genes (VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6) into genotypes (Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, respectively; where x is an integer indicating the number of corresponding genotypes) on the basis of calculated nucleotide sequence identity cut-off values (Table 2.0) (Matthijnssens, *et al.,* 2008, 2011). Up to date (December 2022), 42 G, 58 P, 32 I, 28 R, 24 C, 24 M, 39, 28 A, 28 N, 28 T, 32 E, and 28 H genotypes have been described in humans (Table 2.1) (RCWG, 2021). The full genome-based classification scheme coupled with advancements in second-generation sequencing technologies allows researchers to determine genetic changes and relatedness in circulating RVA strains (Svensson *et al.,* 2016).

**Table 2.1: Percentage Identity Cut-Off Values for Genotyping (As of December 2022) and Functions for Each RVA Segment**

| RVA segment | Size (bp) | Genome protein | Nucleotide identity cut-off value (%) | Number of genotypes | Name of protein |
|---|---|---|---|---|---|
| 1 | 3302 | VP1 | 83 | 28 R | RNA dependent RNA polymerase |
| 2 | 2687 | VP2 | 84 | 24 C | Core shell protein |
| 3 | 2592 | VP3 | 81 | 24 M | Guanyl transferase |
| 4 | 2362 | VP4 | 80 | 58 P | Protease sensitive |
| 5 | 1356 | VP6 | 85 | 32 I | Middle capsid protein |
| 6 | 1062 | VP7 | 80 | 42 G | Glycoprotein |
| 7 | 1581 | NSP1 | 79 | 28 A | Interferon antagonist |
| 8 | 1059 | NSP2 | 85 | 28 N | NTPase |
| 9 | 1074 | NSP3 | 85 | 28 T | Translation enhancer |
| 10 | 751 | NSP4 | 85 | 32 E | Enterotoxin |
| 11 | 667 | NSP5 | 91 | 28 H | Phosphoprotein |

### 2.3.4 RVA Genotype Constellations

Based on the full-genome classification system, RVAs are further classified into genogroups (Matthijnssens, *et al.,* 2008, 2011). Two main RVA genogroup constellations have been identified in humans, Wa-like genogroup with a genomic backbone of Gx-P[x]-I1-R1-C1-M1-A1-N1-T1-E1-H1 and the DS-1 like genogroup with a genetic backbone of Gx-P[x]-I2-R2-C2-M2-A2-N2-T2-E2-H2 (Matthijnssens, *et al.,* 2011). A third and minor genogroup, the AU-1-like genotype with a genetic constellation of Gx-P[x]-I3-R3-C3-M3-A3-N3-T3-E3-H3 that shares ancestry with canine and feline RVs has been described in humans (Matthijnssens & Van Ranst, 2012; Wang *et al.,* 2013). The Wa-like RVA genogroup have a common ancestry with porcine RVs found in combination with P[8], while DS-1 human RVAs share

most of their genome segments with bovine RVs found in combination with P[4] (Matthijnssens & Van Ranst, 2012). Typical Wa-like genetic backbone is found in G1P[8], G3P[8], G4P[8], G9P[8], and G12P[8] RVA strains, while typical DS-l like constellation is found in G2P[4], G8P[4], and G8P[6] strains (Matthijnssens & Van Ranst, 2012). G3P[9] strains possess a typical AU-like genetic backbone (Matthijnssens & Van Ranst, 2012; Wang *et al.,* 2013). Therefore, whole genome-based genotyping of RVA offers a more robust method for investing the evolutionary patterns and genetic relationships between circulating RVA strains (Matthijnssens *et al.,* 2011).

### 2.4 RVA Vaccines

The use of RVA vaccines is the most effective intervention of combating severity of RVA disease (Donato & Bines, 2021). The large burden of RVA-associated disease led the Diarrhoeal Disease Control Programme of WHO, the National Academy of medicine previously called Institute of Medicine (1985-1986), the Bill and Melinda Gates Foundation Programme, and the Global Alliance for Vaccines and Immunization (GAVI) (2002) to select RVA as a high priority pathogen requiring accelerated vaccine development (Glass *et al.,* 2006).

Development of oral RIT vaccine from bovine RV strain (RIT 4237) by the Vesikari research group in 1983 was stopped due to inconsistent efficacies in clinical trials in developing countries (Glass *et al.,* 2006). The RotaShield® vaccine (Wyeth-Lederle, Pearl River, NY, USA) showed a two-dose efficacy of 70-90% in the US and Venezuela in children aged <5 years and was licensed by the Food and Drug Administration (FDA) of the USA in 1998 for routine immunisation of children. However, RotaShield® vaccine was withdrawn from national immunisation programms after reports of intestinal intussusception in 1999 (O'Ryan, 2017).

Currently, four RVA vaccines are pre-qualified by WHO for use globally: RotaTeq® (Merck & Co, Pennsylvania, USA; pre-qualified in 2008), Rotarix® (GlaxoSmithKline, Rixenstart, Belgium: pre-qualified in 2009), Rotasiil® (Serum Institute of India, Pune, India; pre-qualified in 2018), and Rotavac® (Bharat Biotech, Hyderabad, India; pre-qualified in 2018) (Table 2.1) (Burke *et al.,* 2019; Lee, 2021).

In addition, two RVA vaccines including Lanzhou lamb rotavirus vaccine (LLR-85) (Lanzhou Institute of Biological Products, Lanzhou, China) and Rotavin-M1®, (Polyvac, Hanoi, Vietnam) are available for private market use in China, and Vietnam, respectively (Table 2.1) (Burke *et al.*, 2019; Lee, 2021).

**Table 2.2: Characteristics of the Currently Licensed Oral RVA Vaccines**

| Vaccine | Year of WHO pre-qualification | Doses | Vaccine composition | Efficacy | | Vaccine Formulation, Dosage Volume, and Storage |
|---|---|---|---|---|---|---|
| | | | | Developed countries | Developing countries | |
| RotaTeq® (Merck) | 2009 | 3 | 5 attenuated human-bovine reassortant rotaviruses (G1, G2, G3 G4, P[8]) | 898–100% | 43–64% | Liquid suspended in a buffer<br>Dosage volume: 2.5 mL<br>Storage: Refrigerated (2–8°C) for 24 months |
| Rotarix® (GSK) | 2009 | 2 | Single, attenuated human rotavirus RIX4414 strain (G1P[8]) | 85–96% | 49–77% | Liquid or lyophilized vaccine reconstituted with buffer<br>Dosage volume: 1 mL (liquid) or 1.5mL (freeze-dried)<br>Storage: Refrigerated (2–8°C) for 36 months |
| Rotasiil (Serum Institute of India) | 2018 | 3 | 5 human-bovine (UK) reassortant rotaviruses (G1, G2, G3, G4, G9) | No data | 33-67% | Lyophilized vaccine in glass vial and reconstituted with antacid diluent from separate vial<br>Dosage volume: 2.5 mL<br>Storage: Refrigerated (2–8°C) for 30 months |
| Rotavac® (Bharat Biotech) | 2018 | 3 | Single, attenuated human rotavirus strain (G9P[11]) | No data | 48-56% | Liquid in glass vial<br>Dosage volume: 0.5 mL<br>Storage: refrigerated at 2–8°C |
| **Nationally Licenced** | | | | | | |
| LLR-85 (Lanzhou Institute of Biological Products | 2000 (Nationally in China) | 1 annually for 3 years (age 2 months to 3 years) | Live attenuated lamb rotavirus strain (G10P[15]) | | 35% | Liquid suspended in buffer<br>Dosage volume: 2.5 mL<br>Storage: Refrigerated at 2–8°C |
| Rotavin-M1®, (Polyvac) | 2012 (Nationally in Vietnam) | 2 | Single attenuated human rotavirus KH0118-2003 strain (G1P[8]) | | 73% | Liquid vaccine in glass vial<br>Dosage volume: 2.0 mL<br>Storage: refrigerated at 2–8°C |

Rotarix® (GSK) and RotaTeq (Merck) are the most widely used WHO-prequalified RVA vaccines, while Rotasiil (Serum Institute of India) and Rotavac® (Bharat Biotech) are primarily used in India (Lee, 2021). Globally, 122 countries had introduced WHO-prequalified RVA vaccines into their NIPs by the end of 2022 (Figure 2.2) (IVAC, 2022). Most of the lower- and middle-income countries are supported by the Global Alliance for Vaccines and Immunisation (GAVI), of which 42/47 GAVI supported countries had introduced WHO pre-qualified RVA vaccines by the end 2022 (IVAC, 2022). Regionally, 79% countries have introduced WHO pre-qualified RVA vaccines in Africa (Figure 2.3) (ROTA, 2022).

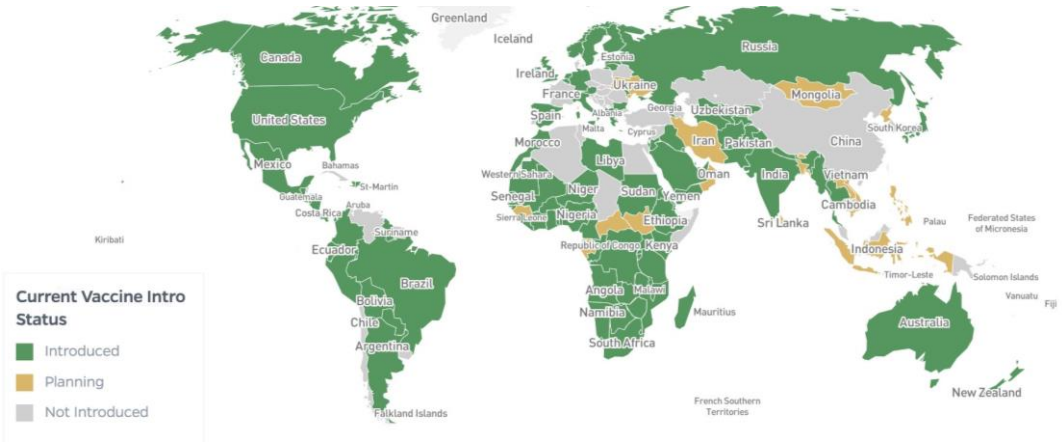


**Figure 2.2: Global Map Showing the Status of WHO Pre-Qualified RVA Vaccine Introductions by May 2023 (IVAC, 2022).**

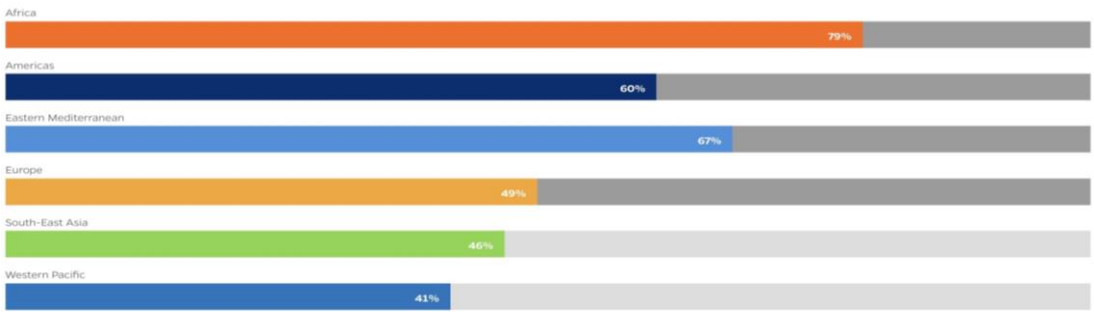


**Figure 2.1: Proportion of Countries that Have Introduced WHO Pre-Qualified RVA Vaccines in Each WHO Region by May 2023** (IVAC, 2022)**.**

## 2.5 RVA Genetic Diversity and Evolution

RVA can undergo evolution and gain genetic diversity through four approaches (Kirkwood, 2010). The first and most common mechanism involves the reassortment of genome segments (gene shift), in which the exchange of cognate gene segments occurs in progeny RVAs when one cell is infected with two or more homologous or heterologous RVA strains (Svensson *et al.,* 2016). This can result in a novel constellation of RVA gene segments, while the reassortment of surface proteins could give rise to a progeny of RVAs with different antigenic properties (Kirkwood, 2010). Secondly, RVAs can undergo genetic drift through accumulating point mutations, giving rise to novel lineages that can escape vaccine-induced immunity (Kirkwood, 2010). Thirdly, NSP genes can undergo gene rearrangement involving deletions, insertions, or duplications into coding and non-coding regions (Kirkwood, 2010). Lastly, it is through interspecies transfer involving the transmission of a genome segment between host species (Kirkwood, 2010; Mijatovic-Rustempasic *et al.,* 2016).

Higher rates of co-infection in low-resource settings estimated at 20% explain the high frequency of reassortants generation in these regions compared to developed economies with co-infection rates of less than 5% (Patton, 2012). Studies have shown that uncommon RVA genotypes resulting from reassortment events may become epidemiologically important (Luchs & Timenetsky, 2016). For instance, a recent study indicated that genotype G9P[8] is gaining dominance among human RVA genotypes circulating globally (Bibera *et al.,* 2020). In addition, introduction of RVA vaccines has increased immune pressure on circulating strains in a local set-up, which may change the factors of RVA evolution (Luchs and Timenetsky 2016; Zeller *et al.,* 2017). Therefore, full-genome surveillance is required to reveal gene-reassortment events, transmission patterns, and genetic drifts of circulating RVA strains in local settings and globally.

## 2.6 RVA Molecular Epidemiology

Studies have shown that both common and unusual RVA genotypes infect humans (Matthijnssens *et al.,* 2016; Matthijnssens *et al.,* 2008; Matthijnssens & Van Ranst,

2012). The most predominant G genotypes of RVA include G1, G2, G3, G4, G9, and G12 while the most predominant P genotypes are P[8], P[4], and P[6] (Donato & Bines, 2021; Dóró *et al.,* 2014; Ouermi *et al.,* 2017). Globally, the most common G/P genotype combinations include G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8] (Dóró *et al.,* 2014). These genotypes account for an estimated 90% of the total human RVA infections globally (Sadiq *et al.,* 2018). However, studies have reported regional variability of RVA genotypes (Bányai *et al.,* 2012). In Africa, the most dominant G/P genotype combinations include G1P[8] (23%), G2P[4] (8%), G9P[8] (7%), and G2P[6] (5%) (Ouermi *et al.,* 2017). In addition, genotypes G1P[6], G8P[6], G9P[6], G8P[8], and G8P[4] are relatively common in Africa, however rare elsewhere (Ouermi *et al.,* 2017). In Eastern and Southern Africa, a wide strain diversity was documented to have been in circulation in between 2010 and 2015, predominantly G1P[8] (26.9%), G2P[4] (15%), G2P[6] (6.6%),G9[8] (6.5%), G12[8] (4.5%), and G3P[6] (3.2%), with the G1P[8] and G2P[4] being the most predominant (Seheri *et al.,* 2018). A surveillance study of RVA genotype distribution in Kenya revealed a wide genotype diversity in circulation, predominantly G1P[8] (45.8%), G8P[4] (15.8%), G9P[8] (13.2%), G2P[4] (7.0%) and G3P[6] (3.1%) in the pre-vaccine period (January 2010-June 2014) and G1P[8] (52.1%), G2P[4] (20.7%) and G3P[8] (16.1%) in the post-vaccine period (July 2014 –December 2018) (Mwanga *et al.,* 2020).

### 2.7 RVA Genotype G2P[4]

RVA G2P[4] genotype is among the most predominant genotype infecting humans worldwide (Bibera *et al.,* 2020). During the pre-vaccine period, G2P[4] was the second most common genotype (10%) circulating globally after G1P[8], however exhibited region-specific temporal variations (Bányai *et al.,* 2012). In North Africa, G2P[4] was second most predominant from 1999 to 2009 (Khoury *et al.,* 2011). However, there have been significant fluctuations in the prevalence of G2P[4] genotype in the post-RVA vaccine period globally (Bibera *et al.,* 2020). Some studies indicated a significant increase in G2P[4] strains in countries using the monovalent Rotarix® vaccine compared with countries using the pentavalent RotaTeq® vaccine (Dóró *et al.,* 2014; Leshem *et al.,* 2014). However, a recent

systematic analysis showed no evidence of significant differences in G2P[4] genotype over time in regions or countries using the Rotarix® or RotaTeq® vaccines (Bibera *et al.,* 2020). Data collected in Kenya spanning the pre- and post-Rotarix® vaccine period shows a significant increase in G2P[4] genotype (Figure 1.1) (Gikonyo *et al.,* 2020; Mwanga *et al.,* 2020).

An evolutionary model established for the global whole genomes of G2P[4] postulates that they evolved in a stepwise manner from lineage I to IVa in the NSP5, NSP1, VP2, VP1, and VP7 genome segments, and from lineage I to V in the VP1, VP3, VP6, NSP2, NSP3, and NSP4 genome segments, with some of the strains undergoing intragenotype reassortments in the VP7, VP3, and NSP4 genes after 2004 giving rise to emergent lineages of V in the VP7 segment, lineages VI and VII in the VP3 gene, and VI, VII, VIII, IX and X lineages in the NSP4 gene (Agbemabiese *et al.,* 2016; Doan *et al.,* 2015). In Kenya, lineages of only two G2P[4] genome sequences that circulated in 1982 and 1989 have been described (Ghosh *et al.,* 2011).

Previously, there were 446 global G2P[4] genome sequences (>80% coverage) in Genbank, of which 155 were from Africa, and only five from Kenya (NCBI, 2022). This study has contributed an additional 63 G2P[4] genome sequence with a genome coverage of > 85% (accession numbers OP677569 to OP677754 and MZ093788 to MZ097268).

### 2.8 RVA Whole Genome Sequencing Approaches

Genomic analysis of RVA reveals the lineages, evolutionary events involving reassortment events and establishes the origin and overall genetic diversity of the strains circulating in a particular region (Ghosh *et al.,* 2011). However, whole genome sequencing of G2P[4] and RVs in general is limited by the complexity of the eleven-segmented nature of their genome (Dung *et al.,* 2017; Fujii *et al.,* 2012; Magagula *et al.,* 2015; Phan *et al.,* 2016). Hence, obtaining full/near-complete G2P[4] genome sequences (>80% coverage) requires a protocol that generates sequences of all the 11 segments (Dung *et al.,* 2017).

In the current study, we compared the performance of an agnostic sequencing method with an amplicon based WGS approaches in recovery of all the 11 segments and genome coverage of G2P[4] RVA. The agnostic WGS method allows isolation of total nucleic particles from clinical samples enabling detection and characterization of important pathogens in the sample (Gauthier *et al.,* 2023). In addition, the method does not require prior knowledge of the targeted genome, can reveal multiple viral pathogens within a sample, and can reveal co-infections. However, this method requires more sample volume and a pre-treatment step to deplete host genomic material and enrichment viral genomic material (Gauthier *et al.,* 2023). In viral agnostic sequencing, isolation of total nucleic acids is done through an initial bead-beating process that allows rapturing all the cells in the stool sample to expose the nucleic contents (Boom *et al.,* 1990; Phan *et al.,* 2016). Since all the 11 RVA segments lack the poly(A) tail, the agnostic genome sequencing approach employs non-ribosomal hexanucleotides targeting enrichment of viral RNAs while excluding contaminating ribosomal RNAs in faecal specimens (Endoh *et al.,* 2005; Nguyen *et al.,* 2016; Phan *et al.,* 2016).

Amplicon-based sequencing methods that employ RVA segment specific primers have been used to enhance coverage of all the 11 RVA segments (Dung *et al.,* 2017; Fujii *et al.,* 2012, 2019; Magagula *et al.,* 2015). A universal RVA genome sequencing approach that is amplicon-based uses highly sensitive and specific Enzyme Immuno Assay (EIA) plates to capture and purify RVA particles from faecal particulate (Dung *et al.,* 2017). This method uses primers specific for all 11 RVA segments to allow amplification of the entire RVA genome in a multiplex PCR reaction (Dung *et al.,* 2017). However, this method does not consider the cross-hybridization of the primers with each other, hence may miss amplifying the target segment(s) (Sint *et al.,* 2012). In addition, the concentration of antibodies in the EIA plates limits the number of bound RVA particles (Izzo *et al.,* 2012). In this study, we adopted a singleplex PCR amplification of each RVA segment that employs segment-specific primers that has been used to sequence RVA genomes, thus eliminating the possibility of primer hybridization (Fujii *et al.,* 2012; Magagula *et al.,* 2015). This method allows for lower sequencing depth per sample therefore permitting multiplexing of extensive samples, which lowers sequencing costs

(Gauthier *et al.*, 2023). However, this method is labour intensive (Moss *et al.*, 2011). It is therefore important to note that an effective sequencing method should recover all the 11 RVA segments and >80 % of genome to allow genomic characterization of the strains circulating in a given region (Matthijnssens *et al.*, 2011).

### 2.9 G2P[4] Whole Genome Studies

Recent studies have investigated genomic epidemiology of G2P[4] strains in the pre- and post-vaccine periods. The table in appendix I summarizes the findings of studies on whole genome of G2P[4] strains.

# CHAPTER THREE

# MATERIALS AND METHODS

## 3.1 Study Site and Population

The samples analysed in this project were from a longitudinal surveillance study monitoring rotavirus infections established at Kilifi County Hospital (KCH) paediatric ward recruiting children aged below 13 years old admitted with diarrhoea as one of their illness symptoms (Khagayi *et al.,* 2020; Otieno *et al.,* 2020). KCH is located at the North Coast of Kenya within Kilifi Health and Demographic Surveillance System (KHDSS) (Figure 3.1). KCH serves an urban, rural, and semi-rural population and is the only Kenyan government health facility that offers paediatric inpatient services in the KHDSS area (891 $km^2$) (Scott *et al.,* 2012). The KHDSS was established in 2000 to capture surveillance data majorly for patients admitted to KCH (Scott *et al.,* 2012). RVA surveillance at KCH was implemented in 2009 by the Kenya Medical Research Institute (KEMRI)-Wellcome Trust Programme, which is a partnership between KEMRI, Wellcome Trust, and the University of Oxford (Nokes *et al.,* 2008; Scott *et al.,* 2012). The study utilized stool samples collected in the pre-vaccine introduction (January 2012 to July 2014) and the post-vaccine (August 2014 to December 2018) periods. The stool samples had been collected and archived at the -80°C freezers at the KWTRP BIOBANK facility (NSW, n.d.). The industrial actions experienced in the ministry of health in Kenya were responsible for the low sample collections experienced in January to May 2012, September 2012, December 2012 to March 2013, December 2013, December 2016 to March 2017, and June 2017 to November 2017, as indicated in figure 1.1.

The stool samples were screened to detect RVA by using the enzyme-linked immunoassay kit (ProSPect™; Oxoid, Basingstoke, UK). All the RVA-positive samples (n=429) (Figure 1.1) were initially genotyped by a partial segment sequencing approach (Mwanga *et al.,* 2020). VP7 and VP4 genes were sequenced, and G and P genotypes were inferred using the Virus Pathogen Resource (ViPR) tool for RVA (Pickett *et al.,* 2012). Those samples that classified as G2P[4] (n= 87) based on the outer capsid proteins were selected for this study (Mwanga *et al.,* 2020). Only

samples (n=69) with sufficient stool sample available in our biobank (>100 µL or 200 mg) were processed for whole genome sequencing (35 pre- and 34 post-vaccine periods), of which 63 (32 pre-vaccination, 31 post-vaccination) samples yielded near-complete G2P[4] genome sequences (>85% coverage) in this study.

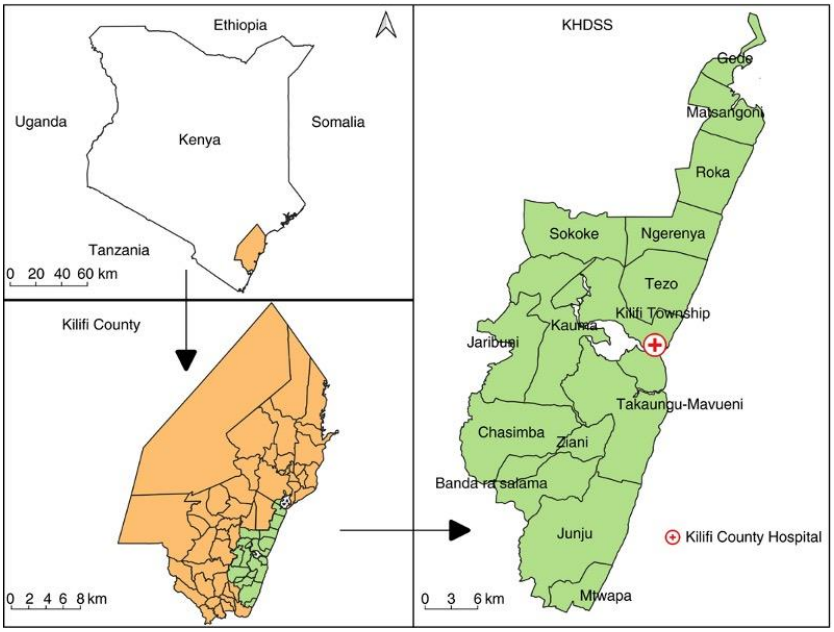


**Figure 3.1: Map Showing the Location of Kilifi Health and Demographic Surveillance System (KHDSS) and the Kilifi County Hospital (KCH)** (This map was adapted from Abuga et al., 2022).

### 3.1.1 Inclusion Criteria

Children <13 years old who were admitted with diarrhoea; defined as more than three episodes of watery stool within the past 24 hours. Availability of stool specimen as an archived faecal sample at KWTRP Biobank. Selected samples must have previously been genotyped as G2P[4] through Sanger Dideoxy sequencing of the surface proteins VP7 and VP4 (Mwanga *et al.*, 2020). Although severe rotavirus-associated diarrhoea tends to occur in children aged under 5 years, rotavirus disease burden may continue well beyond this age group, and infections/reinfections are known to occur throughout a lifetime. In addition, there are reports of shifting the burden of severe rotavirus infection to older age groups because of the introduction of rotavirus vaccination into NIPs (Kyo *et al.*, 2021). Thus, this study included a

broad age range to be able to examine such recent claims if they have also occurred in the local population

### 3.1.2 Exclusion Criteria

The parent/guardian declined consent and there was inadequate stool sample from the patient.

## 3.2 Ethical Consideration

The Scientific and Ethics Review Unit (SERU) of KEMRI approved the study (SERU protocols (#3049 and #2861). The parents or guardians of the children provided written informed consent. Collected samples were archived in the KWTRP Biobank using unique identification codes for confidentiality and anonymity.

## 3.3 Laboratory Methods

### 3.3.1 RNA Extraction and cDNA Synthesis

Two different extraction methods were used owing to the limited availability of nucleic acid extraction kits. The Boom protocol (Boom *et al.*, 1990) was used to extract total nucleic acids from pre-vaccine period (January 2012-July 2014) stool samples. The samples were suspended in 110µl of phosphate buffered saline (PBS) and centrifuged at 10,000 × g for 10 minutes, followed by addition of 20µl of TURBO DNase (Life Technologies, Carlsbad, USA). According to the Boom method, diatom suspension was prepared by adding 50 ml of $H_2O$ and 500 µl (32% HCL) to celite (10 g; Janssen Chemicals (Beerse, Belgium). Next, the reaction vessels were prepared by adding 900 µl of lysis buffer (120 g of GuSC dissolved in 100 ml of Tris HCL and 22 ml of EDTA) to 40 µl of diatom suspension in 1.5 Eppendorf tubes. To extract total nucleic acids (NAs), 200 µl of stool samples were added to vortexed reaction vessels and immediately vortexed for five seconds, then incubated at room temperature for 10 minutes. Thereafter, the vessels were vortexed for five seconds followed by centrifugation (12,000 x g for five seconds). Then, the supernatant was discarded, and the diatom-NA washed twice with wash buffer L2

(120 g of GuSCN dissolved in 100 ml Tris-HCL), twice with freshly prepared 70% ethanol, and once with acetone. After that, the pellets were air-dried in a heating block, then eluted in 60 µl of elution buffer through heating at 56°C for 10 minutes and centrifuging at 12,000 x g for two minutes and aliquoting the supernatant containing NAs into nuclease-free Eppendorf tubes. DNA was degraded from the samples using TURBO DNase (ThermoFisher Scientific). RNA quality was checked on the Bioanalyzer system (Agilent) using the Agilent RNA 6000 Nano Kit (Agilent) as described by the manufacturer. Synthesis of first-strand cDNA was performed using the SuperScript II Reverse Transcriptase (Life Technologies, Carlsbad, USA) with the non-ribosomal Endoh-hexamer primers that exclude transcription of bacterial and host rRNAs. Second strand-cDNA synthesis was done with the 5U of Klenow fragment 3' – 5' exo- (New England Biolabs, Ipswich, USA).

Extraction of RNA from the post-vaccine period (August 2014 to December 2018) was performed using the QIAamp Fast DNA Stool Mini Kit protocol. The samples (200 µl/sample) were first subjected to bead beating (Liu *et al.,* 2016), and then total nucleic acids were extracted using the QIAamp Fast DNA Stool Mini kit (Qiagen, Manchester, UK) as detailed in the manufacturer' protocol. DNA was depleted then from the samples using TURBO DNase (ThermoFisher Scientific), and the RNA quality checked using the Agilent RNA 6000 Nano Kit (Agilent) on the Bioanalyzer system (Agilent) as described by the manufacturer. Next, a one-step PCR was performed separate for each genome segment using segment-specific primers (Table 3.1) (Fujii *et al.*, 2012; Magagula *et al.,* 2015). In brief, a reaction mixture of RNA (2 µl) and reverse primer (2 µl for VPs or 3 µl for NSPs) was incubated in a VeritiPro Thermal Cycler (Applied Biosystems, Foster City, USA) at 95°C for five minutes. Next, a RT-PCR master mix was prepared using the SuperScript IV One-Step RT-PCR System (Thermo Fisher Scientific Inc, Waltham, USA) and added to the first reaction mixture. The RT-PCR conditions for the VP segments (VP1-VP4, VP6 and VP7) were; incubation at 50°C for 30 minutes to synthesise cDNA and then at 95°C for 15 minutes to deactivate RT-enzyme, followed by PCR of 40 cycles (denaturation at 90°C for 30 seconds, primer annealing at 61°C for one minute and an extension at 68°C for six minutes). Amplification conditions for NSP segments (NSP1-NSP5) consisted of cDNA synthesis (50°C for 30 minutes and 95°C for 15 minutes)

followed PCR of 40 cycles (30 seconds at 90°C, one minute at 55°C, and four minutes at 68°C). A final extension at 72°C for 10 minutes was included for both NSPs and VPs. The PCR amplicons were resolved on 2% agarose gel. The amplicons were then purified with Exonuclease I (Thermo Fisher Scientific, Waltham, USA) as per the protocol provided by the manufacturer, and the 11 amplicons pooled for each sample.

**Table 3.1: Primers for Each RVA Genome Segment Used in One-Step RT-PCR.**

| Genome Segment | Primer | 5' sequence 3' | Product Size | Full segment size | Source Publication |
|---|---|---|---|---|---|
| VP7 | VP7-1F | TGTAAAACGACGGCCAGTGGCTTTAAAAGAGAGAATTTC | 1023 | 1059 | (Magagula *et al.*, 2015) |
| | VP7-1063R | CAGGAAACAGCTATGACCGGTCACATCRWACAATTC | | | |
| VP4 | VP4-1F | CAGGAAACAGCTATGACCGGTCACATCCTCAATAG | 2333 | 2359 | (Magagula *et al.*, 2015) |
| | VP4-2359R | TGTAAAACGACGGCCAGTGGCTTTWAAACGAAGTCTTC | | | |
| VP6 | VP6-1F | TGTAAAACGACGGCCAGTGGCTTTWAAACGAAGTCTTC | 1320 | 1356 | (Magagula *et al.*, 2015) |
| | VP6-1364R | CAGGAAACAGCTATGACCGGTCACATCCTCTCAC | | | |
| VP1 | VP1-1F | TGTAAAACGACGGCCAGTGGCTATTAAAGCTGTAC | 3269 | 3302 | (Magagula *et al.*, 2015) |
| | VP1- 3302-R | CAGGAAACAGCTATGACCGGTCACATCTAAGCAC | | | |
| VP2 | VP2-1F | TGTAAAACGACGGCCAGTGGCTATTAAAGGCTCAATG | 2695 | 2727 | (Magagula *et al.*, 2015) |
| | VP2-1R | CAGGAAACAGCTATGACCGGTCATATCTCCACAGTG | | | |
| VP3 | VP3-1F | TGTAAAACGACGGCCAGTGTTTTACCTCTGATGGTG | 2530 | 2591 | (Magagula *et al.*, 2015) |
| | VP3-1R | CAGGAAACAGCTATGACCGGTCACATCATGACTAG | | | |
| NSP1 | NSP1 F | GGCTTTTTTTATGAAAAGTCTTGTG | 1547 | 1564 | (Fujii *et al.*, 2012) |
| | NSP R | CTAGGCGCTACTCTAGT | | | |
| NSP2 | NSP2 F | GGCTTTTAAAGCGTCTCAGTC | 1058 | 1058 | (Fujii *et al.*, 2012) |
| | NSP2 R | GGTCACATAAGCGCTTTCTATTC | | | |
| NSP3 | NSP3 F | GGCTTTTAATGCTTTTCAGTGGTTG | 1050 | 1050 | (Fujii *et al.*, 2012) |
| | NSP3 R | GGTCACATAACGCCCCTATAG | | | |
| NSP4 | NSP4 F | CTTTTAAAAGTTCTGTTCCGAGAG | 739 | 750 | (Fujii *et al.*, 2012) |
| | NSP4 R | AAGACCATTCCTTCCATTAAC | | | |
| NSP5 | NSP5 F | GGCTTTTAAAGCGCTACAGT | 633 | 633 | (Fujii *et al.*, 2012) |
| | NSP5 R | GGTCACAAAACGGGAGTGGGGA | | | |

### 3.3.1 Library Preparation and Sequencing

Standard Illumina libraries of pre-vaccine period samples were prepared following a published protocol (Phan *et al.,* 2016). In brief, second strand-cDNAs of pre-vaccine samples were sheared into fragments of 400-500 nucleotides long. Next, each sample was indexed, pooled and sequenced on the Illumina Hiseq 2500 machine at the Sanger Institute (UK) to generate 250 bp pair-end reads for each sample.

The pooled amplicons for each post-vaccine period samples were purified with Agencourt AMPure XP beads (Beckman Coulter, Brea, USA) as per the manufacturer's instructions. Preparation of standard Illumina libraries was done using the Illumina DNA flex Kit (Illumina, San Diego, USA) as specified by the manufacturer. In brief, tagmentantion of amplicons for each sample was performed using bead-linked transposomes. A limited PCR was performed to add adapters to the tagmented amplicons, and the adapter-bound DNA purified using the tagment wash buffer. A second limited cycle-PCR program was performed to add i7, i5 adapters and sequences (required for cluster generation during Illumina sequencing) to the adapter-bound amplicons. Thereafter, purification of our amplified libraries was performed using double-sided bead purification. Quantification of each library was done using a Qubit Analyzer Kit (Life Technologies, Waltham, USA) and confirmation of correct insert sizes was performed with the Agilent high sensitivity DNA kit (Agilent, Santa Clara, USA), normalized and pooled at equimolar concentrations. The multiplexed Illumina libraries were denatured, and sequencing was performed using the MiSeq Reagent Kits v2 (Illumina Biotechnology; San Diego, USA) on the Illumina MiSeq Machine (Illumina, San Diego, USA) at KWTRP to generate 150 paired end reads.

The processing of samples was done based on availability of reagents and sequencing platforms. The pre-vaccine samples were processed and cDNAs sent to Sanger Institute and sequenced on the Hi-Seq platform, while the post-vaccine samples were sequenced on the MiseQ platform at the KWTRP labs. Studies have shown that the sequences generated from HiSeq2000 and MiSeq are highly

comparable and consistent (Caporaso *et al.*, 2012), therefore the sequencing platform does not influence the sequence data.

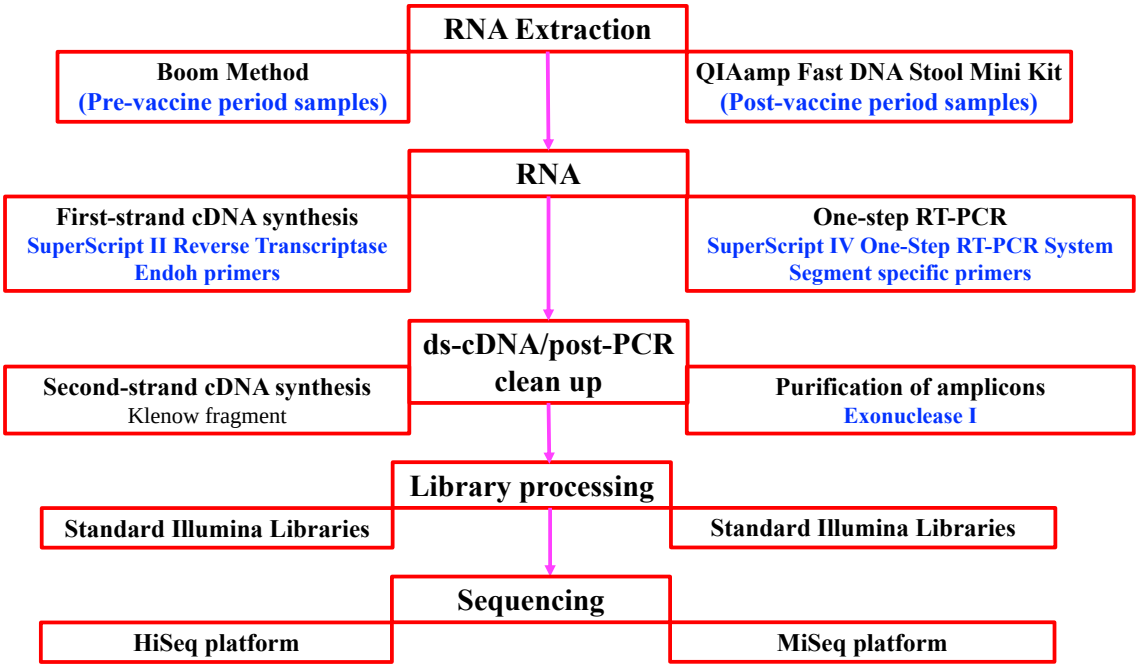


**Figure 3.2: Sample Processing Workflow in the Lab.**

### 3.4 Bioinformatic Analysis.

#### 3.4.1 Genome Assembly

The quality of the raw Illumina FASTQ reads was checked using FastQC (v0.12.0) (Andrews, 2010), and low-quality reads and adapters were trimmed using the Trimmomatic software (Phred score >30) (Bolger *et al.*, 2014). The quality trimmed reads were de-novo assembled using Spades (Bankevich *et al.*, 2012). Identification of RVA-specific contigs of the pre-vaccine samples was performed using USEARCH (Edgar & Bateman, 2010) and a slim algorithm (Cotten *et al.*, 2014). Joining of partial but overlapping RVA contigs was then performed using Sequencher to obtain complete RVA segments (Gene Codes, 2023). The quality of assembled contigs of post-vaccine sequences was assessed using Quast (Gurevich *et al.*, 2013) and identification of ORFs of each genome segment was done using

artemis (Carver *et al.,* 2012). Determination of full-genome genotypes was performed with the RVA Virus Pathogen Resource tool (Pickett *et al.,* 2012).

### 3.4.2 Collection and Processing of Global Sequences for G2P[4]

All contemporary G2P[4] sequences along with the respective metadata consisting of collection year and originating country were downloaded from the RVA Virus Pathogen Resource (ViPR) for all the 11 genome segments (Pickett *et al.,* 2012). The nucleotide sequences whose metadata were missing were searched manually and information on collection year and country of origin that could be found in the primary publication were included in the corresponding sequence data. More sequences were downloaded from GenBank. Duplicates from the two datasets were removed using the seqkit linux command (seqkit rmdup -s < in.fa > out.fa). Datasets of each genome segment were obtained by sub-setting the downloaded global dataset. The datasets were further filtered to exclude samples without all 11 RVA segments and with a coverage of < 80% of the ORF from analysis. Therefore, 350 contemporary G2P[4] genome sequences were included in phylogenetic analyses.

## 3.5 Phylogenetic Analysis

The fasta files of all 11 genome segments were separately aligned using MAFFT v7.487 with the script "mafft --auto --reorder --preservecase input_file.fasta > output_file.fasta" (Katoh *et al.,* 2002), and manually edited using AliView (Larsson, 2014). Phylogenetic trees were inferred using IQ-TREE v2.1.3 (Minh *et al.,* 2020) using the best fit models of evolution selection in IQ-TREE "iqtree -s input_file.fasta -m TEST -bb 1000" (Subha *et al.,* 2017). Maximum likelihood tree sampling was accomplished by 1,000 bootstraps in IQ-TREE (Hoang *et al.,* 2018) and the trees were linked to the corresponding metadata in R v4.1.0 and visualized and plotted in "ggtree" R package (Posit team, 2022).

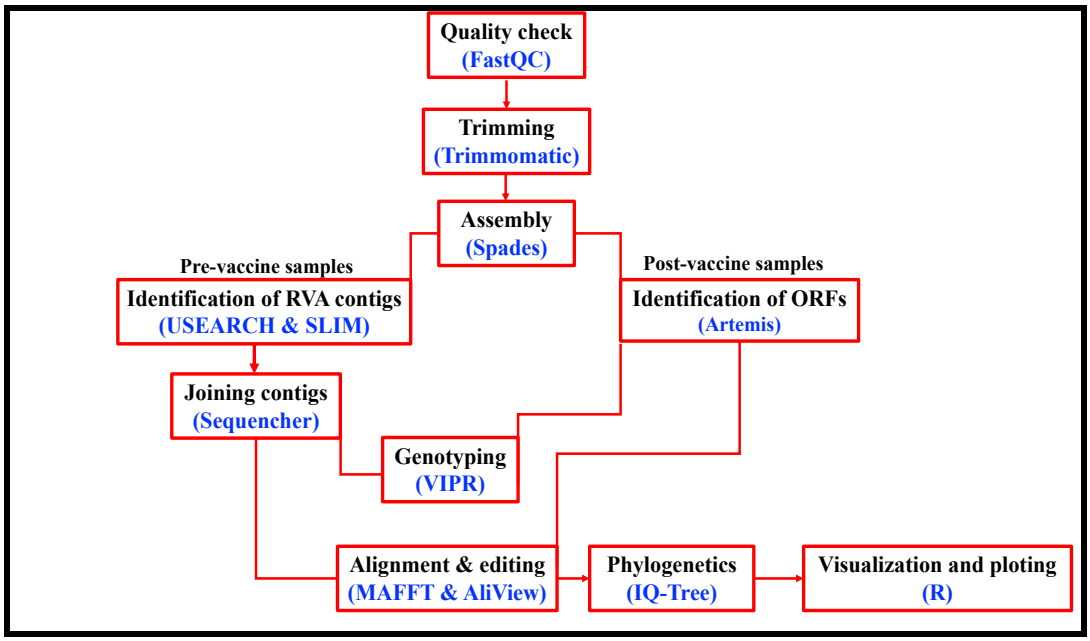


**Figure 3.3: A summary of Bioinformatics Workflow for Analysis of Next-Generation Sequencing Data.**

### 3.6 Selection Pressure Analysis

Analysis of selection pressure for the sixty-three Kilifi G2P[4] strains was performed using the tools in the DataMonkey webserver (fixed effects likelihood and Fast Unconstrained Bayesian AppRoximation (FUBAR) (Murrell *et al.*, 2013; Weaver *et al.*, 2018). A site was considered to be under positive selection if it was detected by the two methods. In addition, synonymous and non-synonymous amino acid changes were analysed between the study strains and the Rotarix vaccine strain through multiple sequence alignment in Aliview and manual inspection.

### 3.7 Data Availability and Scripts Used in Analysis

Nucleotide sequence data generated in this study have been deposited in GenBank and can be accessed using accession numbers: OP677569 to OP677754 and MZ093788 to MZ097268. The epidemiological data analysed in this study are available in the Harvard dataverse (https://doi.org/10.7910/DVN/P4MRVF). The scripts used in analysis have been deposited in Github

(https://github.com/tmakori/Genomic-epidemiology-of-G2P-4-in-Kilifi-Kenya-pre--and-post-vaccine-introduction.git).

# CHAPTER FOUR

# RESULTS

## 4.1 General Characteristics of the Study Participants

Overall, there were 429 RVA positive cases recruited during the study period (January 2012-December 2018) of which 87 (20.3%) were genotyped using the VP7/VP4 binary classification system as G2P[4] (Table 4.1). Sixty-nine samples had sufficient stool sample available in our biobank (>100 μL or 200 mg) and were processed for whole genome sequencing (35 pre- and 34 post-vaccine periods) (Figure 4.1A). Sixty-three near complete genome sequences (>85% coverage) from the G2P[4] samples were recovered in this study (32 pre-vaccination, 31 post-vaccination) (Figure 4.1B & Table 4.1). Segment sequencing failure happened in six samples (Figure 4.1A & B). The baseline characteristics (i.e., age, gender, and vaccination status) of the study participants are provided in Table 4.1. Out of the 31 post-vaccine period samples, 13 (42%) were obtained from children who were given two doses of the Rotarix® vaccine (fully vaccinated) (Table 4.1).

**Table 4.1: Baseline Characteristics of the Study Participants**

| Characteristic | Total RVA cases (%) | Successfully Sequenced G2P[4] (%) | Cases of all G2P[4] cases (%) | P-value: all G2P[4] vs non-G2P[4] |
|---|---|---|---|---|
| Total case numbers | 429 | 63 (14.7) | 87 (20.3) | |
| Vaccination period | | | | 0.04 |
| Pre-vaccine | 215 (49.9) | 32 (50.7) | 35 (40.2) | |
| Post-vaccine | 214 (50.1) | 31 (49.3) | 52 (59.8) | |
| Age (in months) | | | | 0.95 |
| Mean (SD) | 14.9 (13.1) | 17.6 (16.4) | 16.9 (15.4) | |
| Median (IQR) | 11.7 (8.3-17.9) | 11.2 (8.6-20.7) | 10.6 (8.4-19.7) | |
| Age group (in months) | | | | 0.10 |
| 0-11 months | 216 (50.4) | 33 (52.4) | 47 (54.0) | |
| 12– 23 months | 169 (39.4) | 20 (31.2) | 26 (29.9) | |
| 24– 59 months | 35 (8.2) | 7 (11.1) | 11 (12.6) | |
| >=60 months | 9 (2.1) | 3 (4.8) | 3 (3.5) | |
| Gender | | | | 0.79 |
| Male | 251 (51.5) | 37 (58.7) | 52 (59.7) | |
| Female | 178 (41.5) | 26 (41.3) | 35 (40.3) | |
| Vaccination status | | | | 0.13 |
| Vaccinated | 88 (20.6) | 14 (22.2) | 22 (25.6) | |
| Not vaccinated | 215 (50.4) | 33 (52.4) | 35 (40.7) | |
| Unknown | 124 (29.0) | 16 (25.4) | 29 (33.7) | |
| Two doses (full vaccination) | 79 (18.4) | 13 (20.6) | 21 (24.1) | 0.26 |

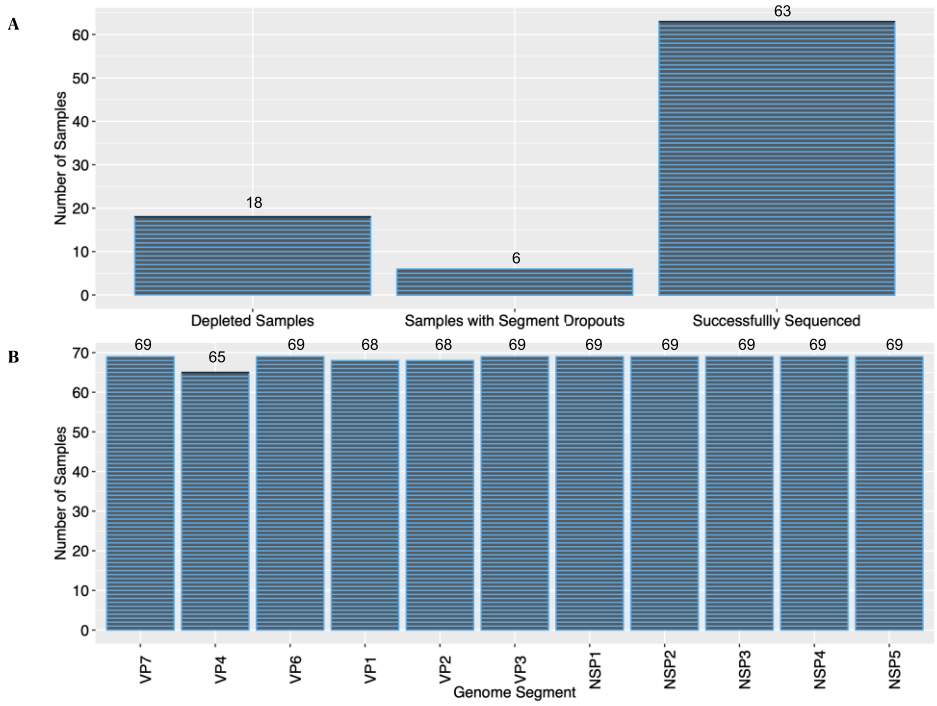


**Figure 4.1: Bar Plots Showing (A) Number of G2P[4] Samples Processed for Whole Genome Sequencing (B) Genome Segment Recovery Across the 69 Processed Samples**

### 4.2 G2P[4] Genome Recovery

The ability of the amplicon based (single-plex PCR method targeting amplification of each of the 11 RVA genome segment separately) and agnostic sequencing approaches to generate high coverage genomes was evaluated. In the amplicon-based sequencing approach, >85% (IQR, 90.0%-90.6%) coverage of G2P[4] genomes were produced for all the samples, with slight differences in segment coverage (Figure 4.2A). The VP7, VP1, VP2, VP3, NSP1, and NSP5 segments exhibited the highest coverage (>90%, IQR 90%-99%), while the VP6, NSP3, and NSP4 segments had a coverage ranging between 70% and <90% (IQR, 70%-89.7%) (Figure 4.2B). The VP4 segment had the lowest coverage (median=68.4%, IQR, 68.7%-69.2%) (Figure 4.2B). In the agnostic sequencing method, genome coverage of > 92% (IQR, 96.1%-97.9%) was recovered across all the samples (Figure 4.2C & D). All the genome segments exhibited good coverage of >92% (IQR, 93.8%-100%) in the agnostic sequencing approach (Figure 4.2D). This study has contributed 63 G2P[4] genome

sequence with a genome coverage of > 85% (accession numbers OP677569 to OP677754 and MZ093788 to MZ097268) to NCBI.

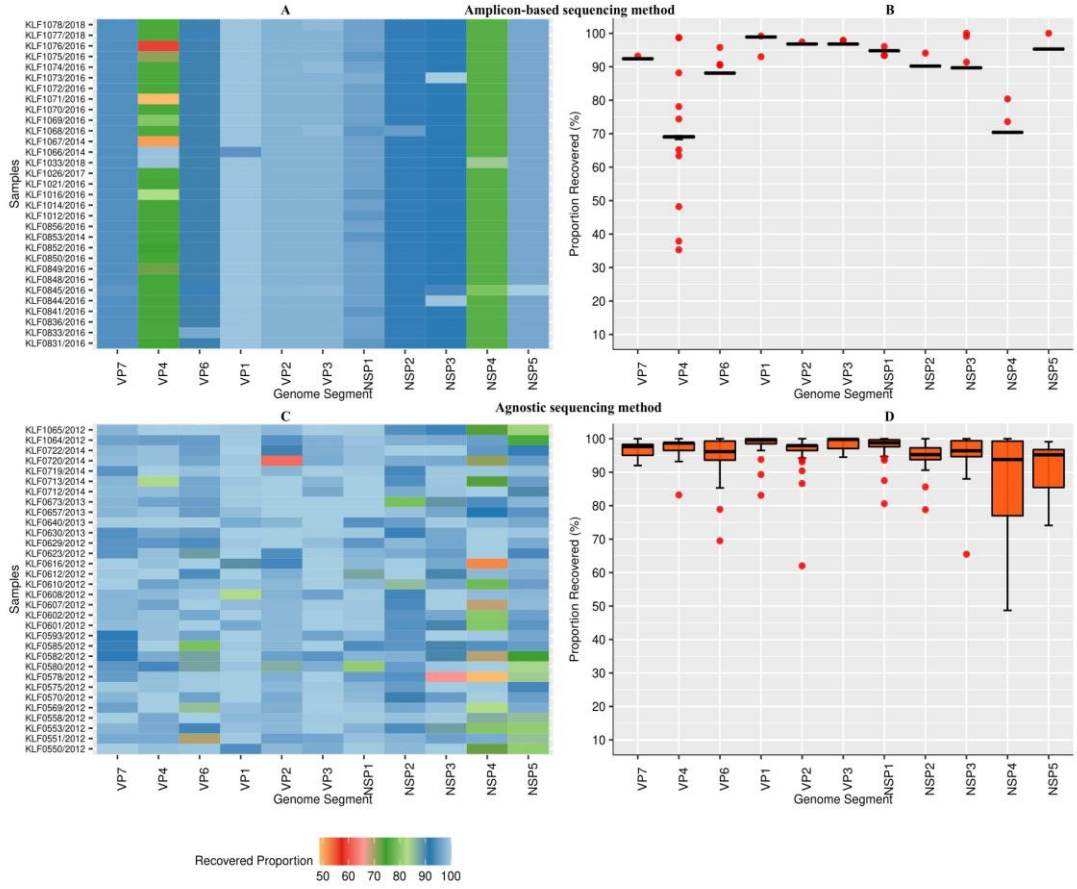


**Figure 4.2: Heatmaps Showing Percentage Recovery of the 11 RVA Segments Across the Samples Sequenced in the Amplicon-Based and Agnostic Sequencing Approaches. (B) Box Plot Showing Segment Coverage for All 11 RVA Segments Across all the Samples Sequenced in the Amplicon-Based and Agnostic Sequencing Methods.**

### 4.3 Experience with Agnostic Vs Amplicon Sequencing Approaches for RVA

The pre-vaccine samples were shipped and processed from the Wellcome Trust Sanger Institute (WTSI) and sequenced on the Hi-Seq platform using the agnostic sequencing approach, while the post-vaccine samples were sequenced using an amplicon-based approach locally at the KWTRP once the next generation sequencing

facility was set up locally and the MiSeq platform available. The snipit plots for the VP7 and VP4 segments confirm the absence of remnants of the primer sequences in the post-vaccine period samples hence results from the two sequencing approaches and platforms adopted in this study are comparable (Figure 4.3). Studies elsewhere have shown that the sequence data generated from HiSeq2000 and MiSeq are comparable (Caporaso *et al.*, 2012), therefore sequencing on either platform should not influence the sequence data. For instance, in the COVID-19 pandemic different sequencing platforms and sequencing methods of SARS-CoV-2 have been adopted to generate SARS-CoV-2 genomes which are highly comparable and have used to inform public policy. In addition, the global G2P[4] dataset we retrieved from GenBank for phylogenetic reconstruction was sequenced using different sequencing methods and platforms, but the results are comparable.

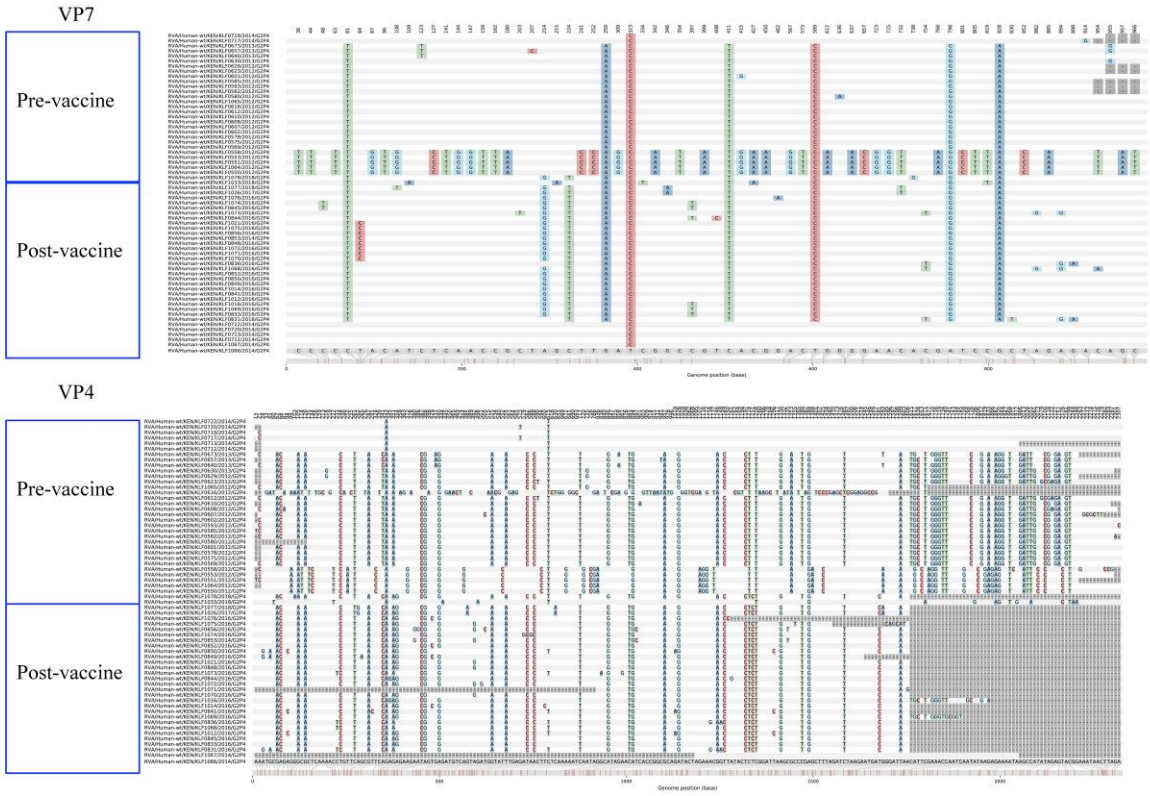


**Figure 4.3: Snipit Plots for the VP7 and VP4 Segments for the Kilifi Strains.**

## 4.4 Genotypic Constellations of the Kilifi Strains

All the 63 Kilifi G2P[4] strains classified as typical DS-1 like genotype (G2-P[4]-I2-R2-C2-M1-A2-N2-T2-E2-H1) using the ViPR tool (Table 4.2). Therefore, no evidence of intergenotype reassortment was reported in the G2P[4] strains circulating in Kilifi, pre- and post-vaccine period introduction.

**Table 4.2: Genotypic Characterization of the Kilifi G2P[4] Strains**

| RVA segment | VP7 | VP4 | VP1 | VP2 | VP3 | VP4 | NSP1 | NSP2 | NSP3 | NSP4 | NSP5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Genotype (n=63) | G2 | P[4] | I2 | R2 | C2 | M1 | A2 | N2 | T2 | E2 | H1 |

### 4.5 Genetic Relatedness of Kilifi G2P[4] Strains

Maximum likelihood phylogenetic trees were inferred to determine the phylogenetic relatedness of the Kilifi strains with global strains for all the 11 genome segments. The genetic diversity of the Kilifi strains was determined in MEGA v11. The VP7, VP4, VP6, VP1, VP2, NSP1, NSP2, NSP3, and NSP5 nucleotide sequences had high percentage similarities (93%-100%), while the VP3 and NSP4 nucleotide sequences exhibited middle to high percentage similarities (85%-100%) (Table 4.3) based on a previously defined nucleotide sequence criteria for G2P[4] (Thanh *et al.,* 2018). All the 11 genome segments had high percentage amino identities (91.5%-100%) (Table 4.3).

**Table 4.3: Percentage Nucleotide and Amino Acid Identity for the 63 Kilifi G2P[4] Strains**

| Genome segment | Percentage nucleotide similarity | Percentage amino acid similarity |
|---|---|---|
| VP7 | 94.3-100 | 96.3-100 |
| VP4 | 94.3-100 | 96.8-100 |
| VP6 | 94.0-100 | 96.0-100 |
| VP3 | 87.1-100 | 92.1-100 |
| VP2 | 97.2-100 | 99.2-100 |
| VP1 | 93.9-100 | 98.0-100 |
| NSP1 | 95.7-100 | 96.0-100 |
| NSP2 | 97.3-100 | 97.5-100 |
| NSP3 | 97.1-100 | 98.4-100 |
| NSP4 | 85.0-100 | 91.5-100 |
| NSP5 | 94.5-100 | 94.5-100 |

### 4.6 Global Phylogenetic Context of Kilifi Strains

The 63 Kilifi G2P[4] strains were phylogenetically compared with 350 global G2P[4] sequences (retrieved from ViPR and GenBank; Table 4.4) for each genome segment. This was done by constructing a maximum likelihood tree for each genome segment in IQTREE V.1.6.12. The spatial temporal distribution of the global sequences retrieved from GenBank is provided in Table 4.4. All the available sequences that made the inclusion criteria were added in the analysis.

**Table 4.4: Spatial Temporal Distribution of the 350 Global Sequences Obtained from GenBank**

| Region | Country | Period | Number of sequences |
|---|---|---|---|
| **Oceania** | Australia | 1975-1999 | 2 |
| | | 2000-2010 | 56 |
| | | 2011-2018 | 6 |
| **Asia** | Bangladesh | 2005-2010 | 17 |
| | | 2011-2018 | 8 |
| | China | 2015-2017 | 11 |
| | Japan | 1980-1999 | 2 |
| | | 2010-2018 | 46 |
| | Vietnam | 2008 | 2 |
| | | 2010-2013 | 3 |
| **Europe** | Belgium | 1999 | 1 |
| | | 2000-2010 | 20 |
| | | 2010-2013 | 7 |
| | Russia | 2017 | 1 |
| | Hungary | 2012 | 22 |
| | Italy | 1996 | 1 |
| | | 2004-2010 | 6 |
| | | 2011 | 1 |
| **Americas** | Brazil | 2005-2010 | 8 |
| | | 2011 | 2 |
| | Canada | 2005-2010 | 8 |
| | Dominican Republic | 2016 | 2 |
| | Paraguay | 2005-2008 | 9 |
| | USA | 1976 | 1 |
| | | 2000-2010 | 12 |
| | | 2011-2017 | 21 |
| **Africa** | Cameroon | 2010 | 2 |
| | Gambia | 2010 | 1 |
| | Ghana | 2005-2010 | 15 |
| | | 2011-2013 | 2 |
| | Kenya | 1982-1999 | 2 |
| | | 2017 | 3 |
| | Malawi | 2012-2013 | 11 |
| | Mauritania | 2012 | 2 |
| | Mozambique | 2012-2013 | 5 |
| | Senegal | 2008 | 1 |
| | South Africa | 2003-2010 | 7 |
| | | 2011-2018 | 23 |
| | Uganda | 2009 | 1 |

### 4.6.1 Phylogenetic Analysis of the VP7 and VP4 Genes

The VP7 and VP4 genes are highly variable and encode the outer capsid immune response glycoprotein and protease sensitive proteins, respectively (Liu, 2014). The Kilifi sequences clustered separate from the global sequences in the VP7 and VP4 global phylogenetic trees (Figure 4.4). Furthermore, the Kilifi sequences formed clusters based on the vaccination period in both the VP7 and VP4 phylogenetic trees (Figure 4.4). However, in both the VP7 and VP4 trees, the Kilifi post-vaccine sequences collected in 2014 clustered with pre-vaccine sequences (Figure 4.4). The Kilifi VP7 and VP4 post-vaccine sequences were interspersed with GenBank sequences collected in 2017 from children admitted to Kenyatta National Hospital (KNH), Nairobi (Figure 4.4A). Only one sequence from Kilifi (KLF1033/2018) clustered closely to sequences that circulated in Mozambique in the VP7 and VP4 phylogenies (Figure 4.4).

G2 sequences of typical DS-1 like strains have been postulated to have evolved in a stepwise pattern from lineage I to IVa by the year 2000 and further underwent intragenotype reassortments after 2004 resulting in an emergent lineage V (Agbemabiese *et al.*, 2016; Doan *et al.*, 2015). The Kilifi G2 sequences classified into lineage IVa-1 and IVa-3 (Figure 4.5A). Lineage IVa-3 co-circulated with low numbers of lineage IVa-1 strains in 2012 and thereafter lineage IVa-3 strains pre-dominantly circulated until 2018 (Figure 4.5A). However, lineages IVa-1, IVa-3, IVnon-a and V circulated in the global context during the study period (Figure 4.5A). The Kilifi P[4] sequences classified as lineage II and IVa, with lineages IVa and II co-circulating in 2012 (Figure 4.5B). However, in 2013, P[4] lineage IVa strains replaced all lineage II sequences and circulated pre-dominantly until 2018 (Figure 4.5B). A similar trend was seen in the global context, where P[4] lineage IVa strains dominated in circulation from 2012 to 2018, with few lineage II sequences circulating in 2012 and 2017 (Figure 4.5B). No lineage shift was reported pre- and post-vaccine periods both for the VP7 and VP4 sequences (Figure 4.5B).

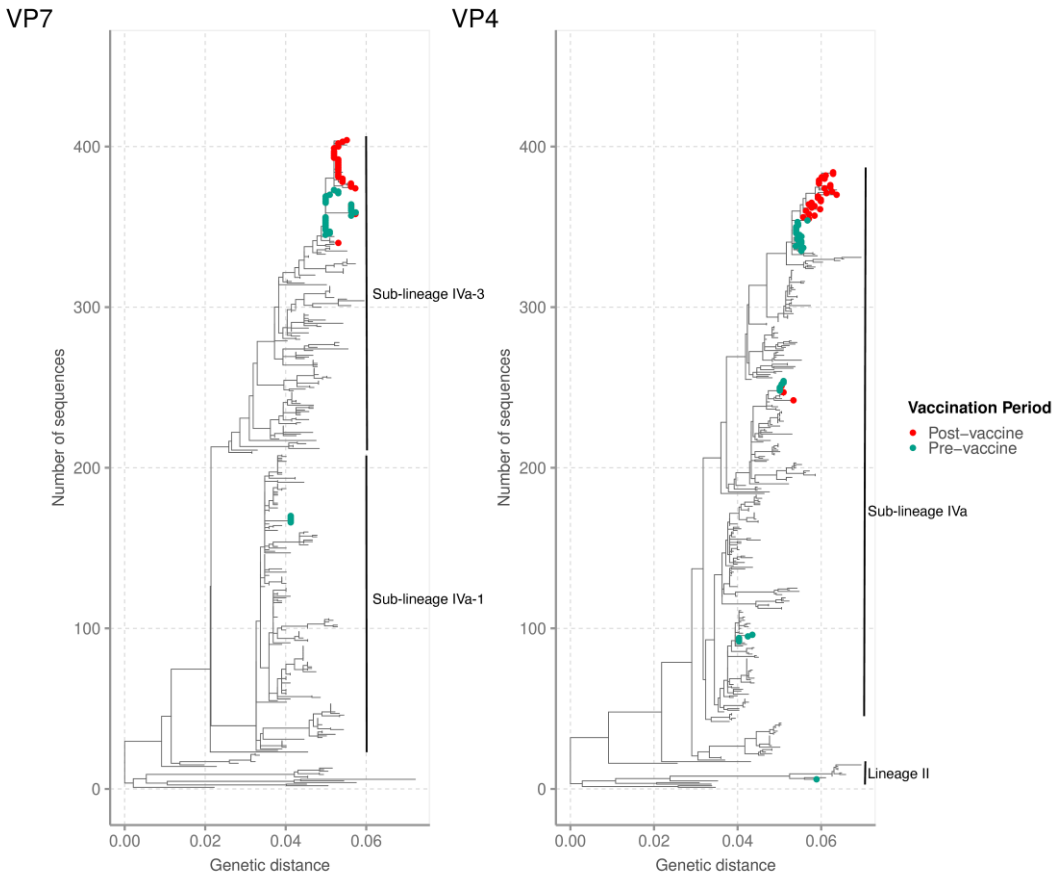


**Figure 4.4: Phylogenetic Reconstruction of the Sixty-Three Kilifi G2P[4] Sequences against A Backdrop of 350 Global Sequences for the VP7 and VP4 RVA Genome Segments Using ML Methods. The Kilifi Sequences are Coloured by the Period of the Sample Collection (Either before Or After Vaccine Introduction in Kenya).**

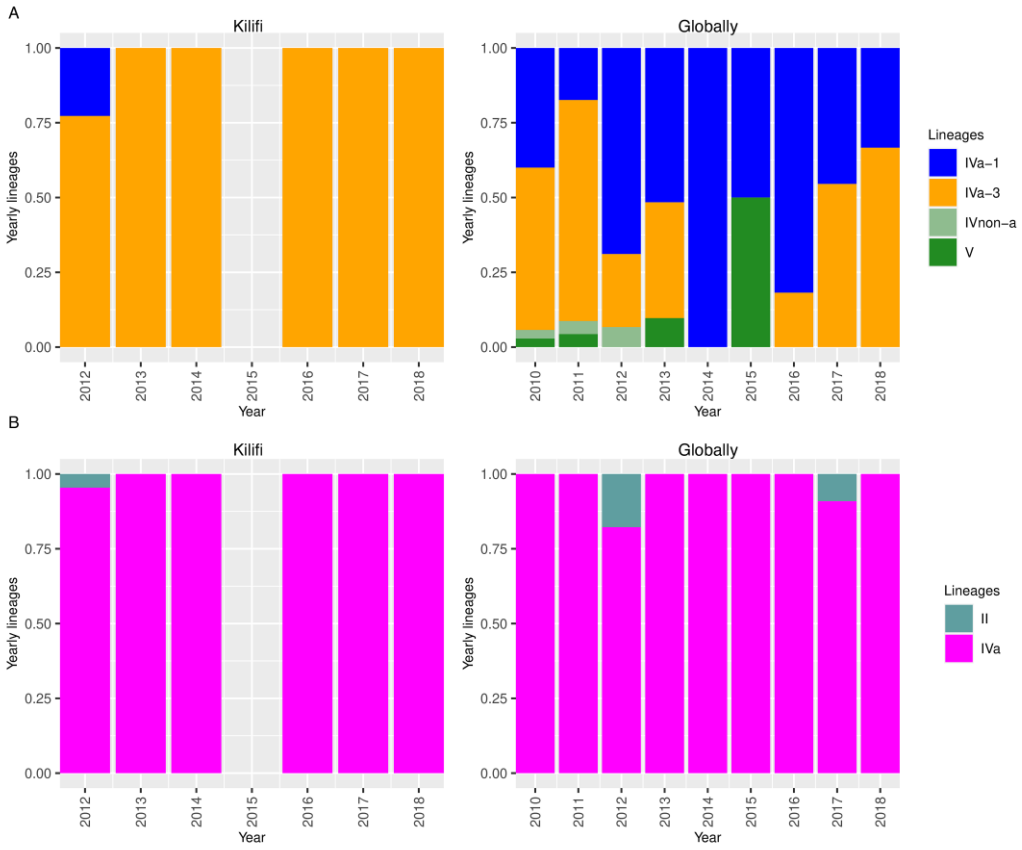


**Figure 4.5: Temporal Pattern of the G2 and P[4] Lineages Observed in Kilifi and Globally. (A) Temporal Pattern of the Kilifi G2 Lineages from 2012 to 2018 and Temporal Pattern of the Global G2 Lineages from 2010 to 2018. (B) Temporal Pattern of the Kilifi P[4] Lineages from 2012 To 2018 and Temporal Pattern of the Global P[4] Lineages from 2010 To 2018.**

### 4.6.2 Analysis of the Backbone G2P[4] Sequences

Phylogenetic analysis revealed that the Kilifi strains clustered into lineages IVa (for the VP2, NSP1, NSP5 sequences), V (for the VP6, VP1, VP3, NSP2, and NSP3 sequences), VI (for the NSP4 sequences) and VII (for the VP3 and NSP4 sequences) (Figures 4.6 and 4.7). Majority of Kilifi sequences in the VP6, VP1, VP2, VP3, NSP1, and NSP2 phylogenetic trees clustered into one major clade which further segregated into two sub-clades separated by vaccination period (Figure 4.6). The sequences sampled in the post-vaccine period in these gene segments were interspersed with GenBank sequences that circulated in 2017 in children admitted to

KNH (Figure 4.6). A further clade of Kilifi sequences that circulated in 2014 (pre- and post-vaccine periods) was observed in the VP3, VP6, NSP1, and NSP2 phylogenetic trees (Figure 4.6). In the NSP4 phylogenetic tree, four strains (KLF0831, KLF0836, KLF1068, and KLF1078) that circulated post-vaccine introduction were interspersed with sequences that circulated in the pre-vaccine period, with majority of post-vaccine sequences forming a distinct clade (Figure 4.7). Similarly, the KLF1078 post-vaccine sequence clustered with pre-vaccine sequences in the NSP3 phylogenetic tree (Figure 4.7). In the NSP5 phylogeny, the Kilifi post-vaccine sequences formed one clade, except for the 2014 post-vaccine sequences that were interspersed with pre-vaccine sequences (Figure 4.7). However, the Kilifi pre-vaccine sequences in the NSP5 phylogeny clustered into patterns of three clusters ($n \geq 2$), and three singletons (Figure 4.7). A distinct clade of five strains (KLF1064, KLF0550, KLF0551, KLF0553, and KLF0558) that circulated in 2012 (pre-vaccine period) was observed in the VP1, VP2, VP3, NSP2, NSP4, and NSP5 phylogenies (Figures 4.6 and 4.7). However, in the NSP3 phylogeny this clade was interspersed with global sequences collected from Japan, Belgium, Australia, and Hungary between 2012 and 2017 (Figure 4.7). One Kilifi post-vaccine sequence (KLF1033/2018) formed a distinct cluster with sequences collected from Mozambique across all the backbone sequences and consistent with the VP7 and VP4 phylogenies (Figures 4.4, 4.6, and 4.7). In the NSP5 phylogeny, two sequences (KLF1066/2014 and KLF0722/2014) clustered as singletons, while one sequence (KLF0601/2012) clustered with a sequence from KNH (Figure 4.7).

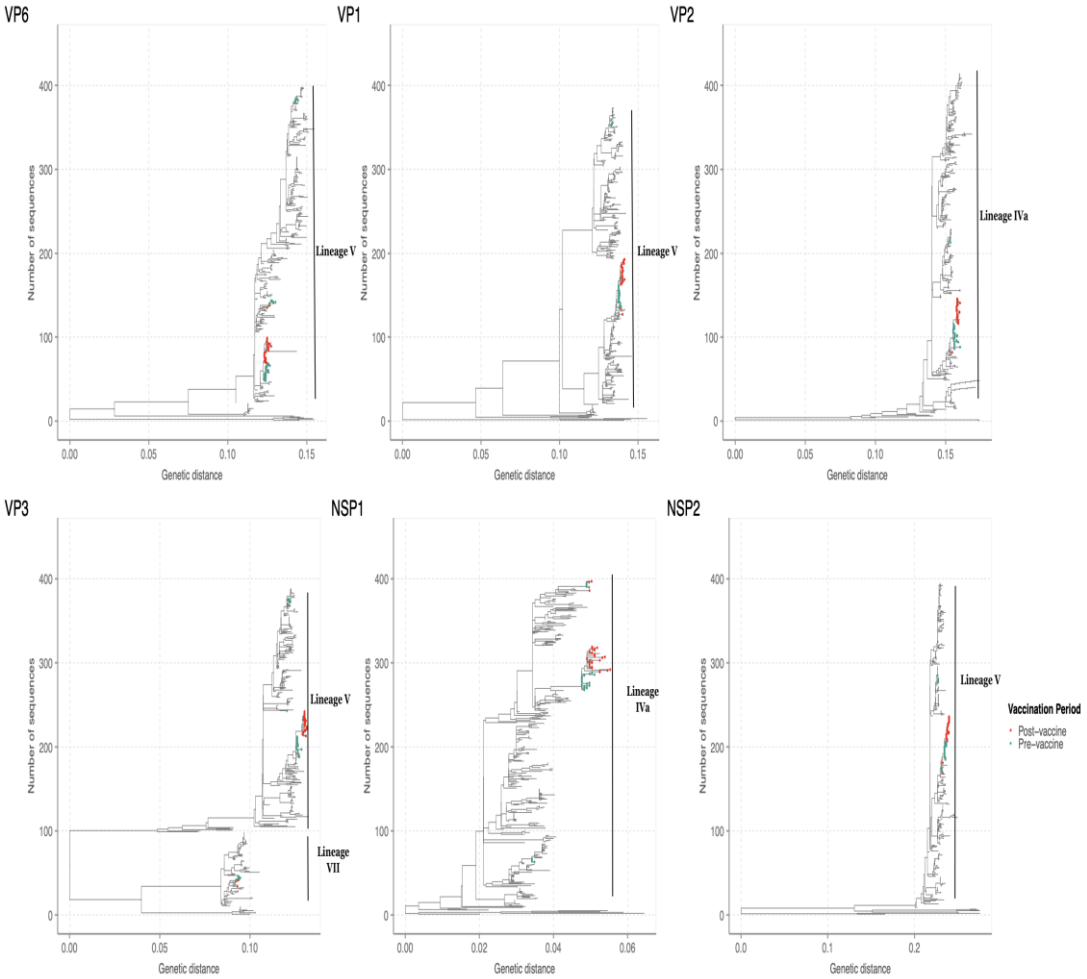


**Figure 4.6: Maximum Likelihood Phylogenetic Trees for the VP6, VP1, VP2, VP3, NSP1, and NSP2 Segments for the 63 Kilifi G2P[4] Strains with A Backdrop of 350 Global Sequences for each Segment Obtained from Genbank. the Kilifi Pre-Vaccine Sequences are Shown in Green and Post-Vaccine Sequences in Red.**

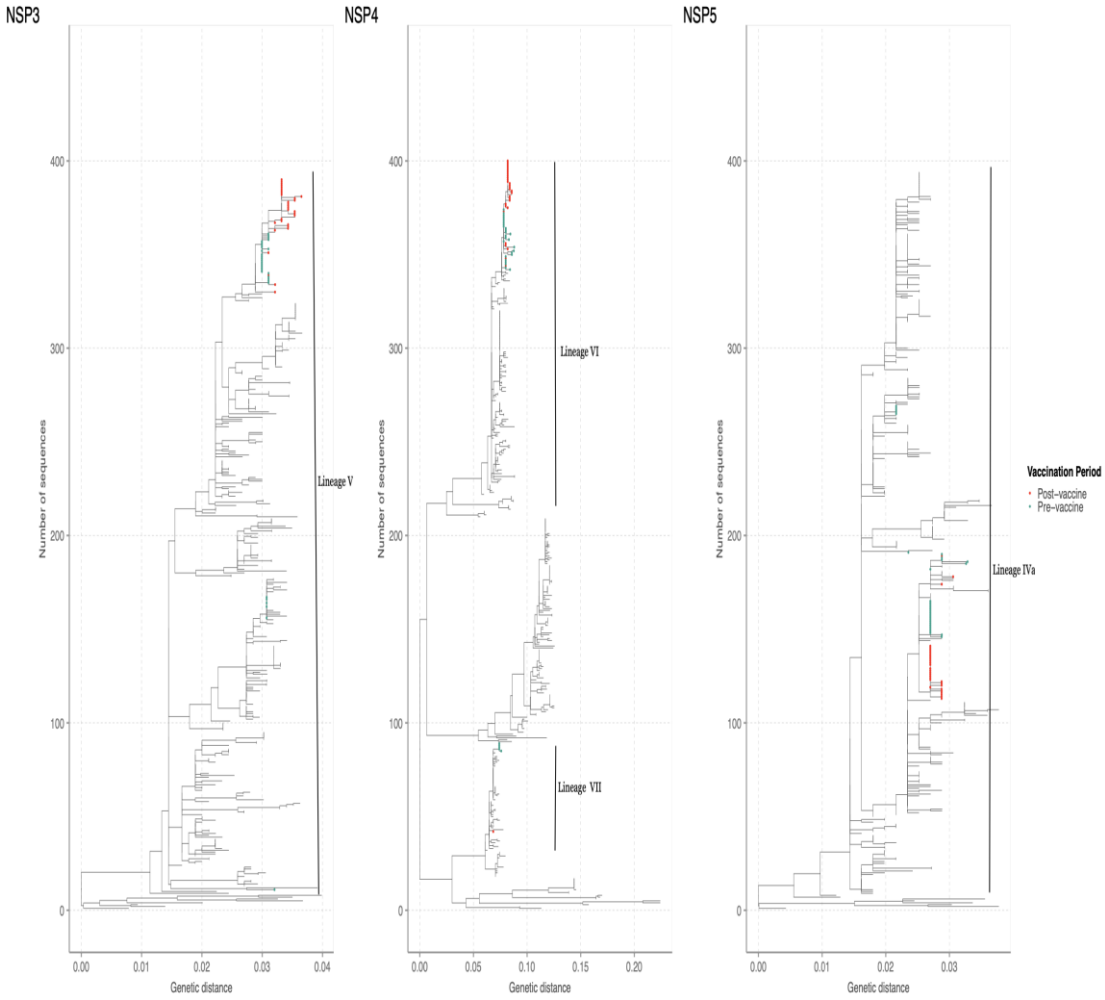


**Figure 4.7: Maximum Likelihood Phylogenetic Trees for the NSP3, NSP4 and NSP5 Segments for the 63 Kilifi G2P[4] Strains with a Backdrop of 350 Global Sequences for Each Segment Obtained from Genbank. The Kilifi Pre-Vaccine Sequences are Shown in Green and Post-Vaccine Sequences in Red.**

### 4.7 Amino Acid Changes in the VP7 Glycoprotein (G) and VP4 Protease Sensitive (P) Proteins

Amino acid changes in the VP7 and VP4 proteins were analysed with respect to the DS-1 like ancestral strain. The VP7 protein (glycoprotein) has two important antigenic epitopes (7-1 (7-1a and 7-1b) and 7-2) and amino acid changes in these epitopes can potentially alter the ability of neutralizing antibodies to neutralize virus infectivity and reduce the effectiveness of vaccines (Aoki *et al.*, 2009). Both pre- and post-vaccine period proteins sequences from Kilifi contained conserved amino acid substitutions (A87T, D96N, N125T, and V129M) in the 7-1a antigenic epitope with respect to the DS-1 like ancestral (Table 4.5). All Kilifi lineage IVa-1 sequences harboured the N242S aa change relative to the DS-1 strain, while lineage IVa-3 had the N213D aa substitution (Table 4.5). Other amino acid changes were observed outside the antigenic epitopes as shown in Table 4.5. Notably, the I144M aa substitution was observed in the T-lymphocyte epitope in the VP7 protein of all the Kilifi strains (Table 4.5).

**Table 4.5: Amino Acid Changes in the VP7 Glycoprotein Sequences of Kilifi Strains Relative to the DS-1 Ancestral Strain.**

| VP7 Protein | Epitope 7-1a | Epitope 7-1b | Epitope 7-2 | Outside antigenic epitopes |
|---|---|---|---|---|
| Lineages IVa-1 | A87T, D96N, N125T, and V129M | N242S | None | S15F, I44M, P75S, I113T, I241M, I287V, V306I, A319T. |
| Lineage IVa-3 | A87T, D96N, N125T, and V129M | N213D | None | I44M, S72G (post-vaccine sequences), P75L/P75S, I113T, V139I, S178N, I287V, V306I, A319T. |

The VP4 spike protein is proteolytically cleaved into VP8* and VP5* domains for its activation. In the activated form the VP8* domain forms a globular head which sits on VP5* stalk. The VP5* stalk and VP8* head contain five (5-1 to 5-5) and four (8-1 to 8-4) antigenic epitopes which are exposed to the surface and have been shown to

contain 37 aa (Dormitzer *et al.,* 2002). The Kilifi strains harboured three aa substitutions in the 8-3 (Q114P or L114P and N133S) and 8-4 (N89D) antigenic epitopes with respect to the DS-1 ancestral strain (Table 4.6).

**Table 4.1: Amino Acid Changes in the Antigenic Epitopes of VP4 Protease Sensitive Proteins of Kilifi Strains Relative to the DS-1 Ancestral Strain.**

| **VP4 protein** | **8-1** | **8-2** | **8-3** | **8-4** | **5-1** | **5-2** | **5-3** | **5-4** | **5-5** |
|---|---|---|---|---|---|---|---|---|---|
| Lineage II | None | None | Q114P/L114P | N89D | None | None | None | None | None |
| Lineage IV | None | None | Q114P/L114P | N89D | None | None | None | None | None |

## 4.8 Amino Acid Changes between Pre-and Post-Vaccine Strains Across All 11 Proteins

The amino acid changes between the pre- and post-vaccine strains were analysed. Non-synonymous amino acid changes were observed in 10 proteins except for the NSP3 protein. All the 11 genome segments were under negative selection (Table 4.7). However no positive selection was detected across the 11 genes (Table 4.7).

**Table 4.2: Amino Acid Changes between the Kilifi Pre- and Post-Vaccine Strains**

| Protein | Non-synonymous aa changes | Synonymous aa changes | Analysis of selection pressure in FER and FUBAR | |
|---|---|---|---|---|
| | | | Positive selection | Negative selection |
| VP1 | Q959R | A157, S389, R932 | 0 | 57 |
| VP2 | S233N, S416L, L658P, Y792C | R831 | 0 | 52 |
| VP3 | N301D, P347S | T110, S132, Y622, V764, S819 | 0 | 104 |
| VP4 | S7R | F417, A451, F527 | 0 | 26 |
| VP6 | G13D | None | 0 | 18 |
| VP7 | S72G, S75L | None | 0 | 12 |
| NSP1 | D403N | I90 | 0 | 47 |
| NSP2 | K92R | None | 0 | 9 |
| NSP3 | None | N16, L223 | 0 | 21 |
| NSP4 | K59R | P34 | 0 | 19 |
| NSP5 | N141S | None | 0 | 5 |

## 4.9 Comparison of Antigenic Epitopes of Kilifi G2P[4] Strains with the Rotarix® G1P[8] Strain

The VP7 and VP4 (VP8* and VP5*) antigenic epitopes of the Kilifi G2P[4] strains were compared with those of the Rotarix®-A41CB052A vaccine strain. The percentage nucleotide and amino acid identities between the Kilifi strains and the Rotarix®-vaccine strain were lowest in the VP7 segment (73% and 72-74%, respectively) and highest in the VP4 segment (84-87% and 87-90%) (Table 4.8).

Comparison of the VP7 antigenic epitopes of Kilifi strains with the Rotarix®-A41CB052A strain revealed amino acid differences at seven positions in the

7-1a region, at two positions in the 7-1b region, and at eight positions in the 7-2 region (Figure 4.6A).

**Table 4.8: Percentage Nucleotide and Amino Acid Identity for the Kilifi G2P[4] Strains Relative to the Rotarix® -G1P[8] Strain**

| Segment | Percentage nucleotide similarity (%) | Percentage amino acid similarity (%) |
|---|---|---|
| VP7 | 73 | 72-74 |
| VP4 | 84-87 | 87-89 |
| VP6 | 78-80 | 91-92 |
| VP1 | 79 | 89-90 |
| VP2 | 81-82 | 89-91 |
| VP3 | 75-76 | 53-55 |
| NSP1 | 75 | 68-69 |
| NSP2 | 80-82 | 89-90 |
| NSP3 | 77-78 | 82-83 |
| NSP4 | 78-80 | 80-84 |
| NSP5 | 84-85 | 80-83 |

Comparing VP4 antigenic epitopes of the Kilifi strains with those of the Rotarix®-A41CB052A vaccine strain showed differences in amino acid residues at six positions in region 8-1, at two positions in region 8-2, at seven positions in region 8-3, and at one position in region 8-4 of the VP8* antigenic epitope (Figure 4.8B). Amino acid differences were observed in entire VP5* antigenic epitope across all the Kilifi strains relative to the Rotarix®-A41CB052A vaccine strain (Figure 4.8B).

A

| Strain | Vaccine period | Lineage | Epitope 7-1a | | | | | | | | | | | | | | Epitope 7-1b | | | | | | Epitope 7-2 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 87 | 91 | 94 | 96 | 97 | 98 | 99 | 100 | 104 | 123 | 125 | 129 | 130 | 291 | 201 | 211 | 212 | 213 | 238 | 242 | 143 | 145 | 146 | 147 | 148 | 190 | 217 | 221 | 264 |
| Rotarix G1P1 A[8] | Vaccine strain | | T | T | N | G | E | W | K | D | Q | S | V | V | D | K | Q | N | V | D | N | T | K | D | Q | N | L | S | M | N | G |
| KLF0831/2016 | Post-vaccine | IVa-3 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | D | N | N | R | D | N | T | S | D | I | S | G |
| KLF0833/2016 | Post-vaccine | IVa-3 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | D | N | N | R | D | N | T | S | D | I | S | G |
| KLF1012/2016 | Post-vaccine | IVa-3 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | D | N | N | R | D | N | T | S | D | I | S | G |
| KLF0558/2012 | Pre-vaccine period | IVa-1 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | N | N | S | R | D | N | T | S | D | I | S | G |
| KLF0569/2012 | Pre-vaccine period | IVa-3 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | D | N | N | R | D | N | T | S | D | I | S | G |
| KLF0570/2012 | Pre-vaccine period | IVa-3 | T | N | S | N | E | W | E | N | Q | D | T | M | N | K | Q | D | V | D | N | N | R | D | N | T | S | D | I | S | G |

B

| Strain | Vaccine period | Lineage | 8-1 | | | | | | | | | | | 8-2 | | 8-3 | | | | | | | | | | 8-4 | | | 5-1 | | | | | | | | | 5-2 | 5-3 | 5-4 | 5-5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 100 | 146 | 148 | 150 | 188 | 190 | 192 | 193 | 194 | 195 | 196 | 180 | 183 | 113 | 114 | 115 | 116 | 125 | 131 | 132 | 133 | 135 | 87 | 88 | 89 | 384 | 386 | 388 | 393 | 394 | 398 | 440 | 441 | 434 | 439 | 429 | 306 |
| Rotarix-A41CB052AG1P1 A[8] | | Rotarix strain | D | S | S | N | S | S | A | N | L | N | N | E | R | N | P | V | D | S | S | N | D | N | N | T | N | S | Y | S | A | W | N | L | R | E | S | S | L |
| KLF0831/2016 | Post-vaccine | IVa | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | P | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |
| KLF0833/2016 | Post-vaccine | IVa | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | P | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |
| KLF1012/2016 | Post-vaccine | IVa | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | P | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |
| KLF1064/2012 | Pre-vaccine period | II | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | Q | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |
| KLF0558/2012 | Pre-vaccine period | IVb | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | Q | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |
| KLF0569/2012 | Pre-vaccine period | IVa | D | S | Q | D | S | T | D | L | N | N | I | T | A | S | L | T | N | N | E | N | S | D | N | T | D | F | F | L | W | P | G | R | T | P | I | L | R |

**Figure 4.8: Alignments of (A) VP7 (7-1a, 7-1b, and 7-2) and (B) VP4 (VP8* and VP5*) Antigenic Epitopes of Representative Kilifi Strains with Those of the Rotarix®-A41CB052A Vaccine Strain. Amino Acid Residues Shaded in Purple Are those Different from the Rotarix® Strain.**

# CHAPTER FIVE

# DISCUSSION

## 5.1 Discussion

The current study performed genomic epidemiology of RVA G2P[4] strains circulating pre- and post-rotavirus vaccine introduction in Kilifi, coastal Kenya. All the 63 genomes generated in this study classified as typical human DS-1 like genotype (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) indicating no evidence of interspecies transmission. These findings are consistent with G2P[4] whole genome studies in several other countries where the reported G2P[4] strains evolved devoid of reassortment with animal RVs (Agbemabiese *et al.,* 2016; Aida *et al.*, 2016; Do *et al.,* 2015; Doan *et al.,* 2011, 2015; Donato *et al.,* 2014, 2021; Mwangi *et al.,* 2022, 2023).

The study sequences formed separate clusters based on the vaccination period in the VP7, VP4, VP1-VP3, NSP1, NSP2, and NSP5 genome segments. Besides, the G2P[4] strains that circulated in the early post-vaccine period (i.e., July 2014) were interspersed with strains circulating in the post-vaccine period across all the 11 genome segments possibly because vaccine coverage was low, hence the vaccine had no effect yet on the circulating strains. Whole genome studies characterising RVA G2P[4] strains circulating in South Africa (Mwangi *et al.,* 2022) and Zambia (Mwangi *et al.,* 2023) following vaccine introduction showed separate clustering of the pre- and post-vaccine period strains consistent with our findings. Similar findings have been reported in whole genome studies of G1P[8] strains in Rwanda (Rasebotsa *et al.,* 2020), and Belgium (Zeller *et al.,* 2017). However, all the strains (both G2P[4] and G1P[8]) reported in these studies (Magagula *et al.,* 2015; Mwangi *et al.,* 2022, 2023; Rasebotsa *et al.,* 2020; Zeller *et al.,* 2017) were shown to be under negative purification pressure and not positive selection pressure and therefore the observed genetic changes post-vaccine period were attributed to natural fluctuations possibly a genetic drift (Kirkwood, 2010) and not vaccine-induced evolution.

In the global context, the Kilifi G2P[4] strains formed clusters separate from the global sequences both in the pre- and post-vaccine periods in the 11 genome segments. This suggests that the genetic diversity of the strains was locally restricted in the pre- and post-vaccine periods. The Kilifi G2P[4] strains circulating in the post-vaccine period clustered with sequences sampled from children admitted to Kenyatta National Hospital, Nairobi further suggestive of restricted local diversity of Kenyan G2P[4] strains. The limited number of G2P[4] sequences from Kenya and East Africa available in GenBank may have contributed to the uncertainty regarding the regional context of Kilifi G2P[4] diversity. In the NSP3 gene, five strains clustered with global strains from Japan, Belgium, Australia, and Hungary reflective of circulation of G2P[4] strains with a similar gene elsewhere. In all the 11 gene segments, one Kilifi strain (KLF1033/2018) clustered with strains that circulated in Mozambique indicating there was limited importation of global strains into Kilifi.

Temporal lineage distribution coupled with phylogenetic analyses of Kilifi strains together with global sequences showed that lineages II (for VP4 segments), IV (for VP7, VP4, VP2, NSP1, and NSP5 sequences), V (for VP6, VP1, VP3, NSP2, and NSP3 segments), VI (for the NSP4 segment) and VII (for the VP3 and NSP4 segments) were circulating in Kilifi during the study period. The VP7 lineages further classified into sub-lineages IVa-1 and IVa-3, while all VP4 sequences classified into sub-lineage IVa. In contrast to findings in South Africa (Mwangi *et al.*, 2022) and Zambia (Mwangi *et al.*, 2023) where the introduction of RVA vaccine was linked to VP7 and VP4 lineage shifts, there was no VP7 and VP4 lineage shifts in Kilifi during the pre- and post-vaccination periods. In addition, temporal lineage analysis showed that few VP7 and VP4 lineages were in circulation in Kilifi compared to the global context during the study period. This further supported our hypothesis that the evolution of G2P[4] strains circulating in Kilifi was driven by local drivers. However, it is important to note that regardless of the locally restricted genetic diversity of G2P[4] strains in Kilifi, the observed lineages in Kilifi (II (for VP4 segments), IV (for VP7, VP4, VP2, NSP1, and NSP5 sequences), V (for VP6, VP1, VP3, NSP2, and NSP3 segments), VI (for the NSP4 segment) and VII (for the VP3 and NSP4 segments)) have been reported in numerous regions worldwide, including Zambia (Mwangi *et al.*, 2023), South Korea (Thanh *et al.*, 2018), Japan

(Doan *et al.,* 2015), South Africa (Mwangi *et al.,* 2022), Ghana (Agbemabiese *et al.,* 2016), Australia (Donato *et al.,* 2021), and USA (Dennis *et al.,* 2014) regardless of the period of vaccination.

Six conserved amino acid changes (A87T, D96N, N125T, V129M, N242S, and N213D) were observed in the 7-1a and 7-1b VP7 antigenic epitopes of both the pre- and post-vaccine period strains circulating in Kilifi relative to the ancestral DS-1 strain. The aa changes (A87T, D96N, and N213D) have been implicated in escape of the virus from host neutralization antibodies (Dyall-Smith *et al.,* 1986). In addition, the I44M aa change observed in the T lymphocyte epitope (40-52) of all the Kilifi strains may potentially lead to loss of virus recognition by T cells resulting in escape from host immune responses (Morozova *et al.,* 2015; Wei *et al.,* 2009). The VP4 antigenic epitopes of the Kilifi strains (both pre- and post-vaccine period) harboured amino acid changes; N89D in epitope 8-4 and Q114P or L114P and N133S in the 8-3 epitope, which have been linked to the escape of the virus from host neutralising monoclonal antibodies (Monnier *et al.,* 2006). These aa changes in the VP7 and VP4 antigenic epitopes were present in both the Kilifi pre- and post-vaccine period strains suggesting they were not caused by vaccine usage.

The presence of few aa acid matches between the Rotarix® vaccine strain and Kilifi pre- and post-vaccine period strains in the VP7 and VP4 antigenic epitopes indicates that vaccination pressure did not generate significant antigenic changes. Besides, purifying pressure, which could serve as a strategy to get rid of any deleterious mutations resulting from the error-prone RNA-dependent RNA polymerase (Liu, 2014) to maintain the functional elements in the genome (Hoxie & Dennehy, 2021), was the consensus selection pressure in the coding regions of the 11 segments of all Kilifi strains. The absence of positive selection in the coding regions across the 11 genes of all Kilifi strains further supports that they were not under vaccine-induced pressure consistent with study findings on the G2P[4] strains circulating in South Africa following vaccination (Mwangi *et al.,* 2022).

The two whole genome sequencing methods (agnostic and amplicon-based approach) adopted in this study yielded near complete genomes of RVA G2P[4] genotype

(>85% and >92% coverage, respectively). However, significant variations in segment recovery were observed in the single-plex amplicon-based method implemented on the MiSeq platform. The VP1 segment had the highest coverage (>99%) while the NSP4 and VP4 exhibited the lowest coverages (70% and 68%, respectively). Single-plex amplicon-based sequencing methods have been used to generate RVA genomes globally (Dennis *et al.*, 2014; Fujii *et al.*, 2012; Magagula *et al.*, 2015; Mokoena *et al.*, 2021). Amplicon based sequencing enriches the target pathogen for low sample titres and is more specific allowing high genome coverage (Lee *et al.*, 2019). In this study, lower coverage for the VP4 and NSP4 segment may be attributed to primer mismatches arising from mutations in the primer annealing sites resulting in amplicon drop-offs during amplification (Bei *et al.*, 2022). This can be overcome through updating the primers used in this study taking to consideration the mutations (Bei *et al.*, 2022).

The agnostic sequencing method implemented on the HiSeq platform yielded G2P[4] genomes of high coverage (>92%) and exhibited high and uniform coverage in segment recovery across the 11 genome segments (IQR, 93.8%-100%). The HiSeq platform generates higher read yield (1.2 billion paired-end reads) compared to the MiSeq platform (24-30 million paired-end reads) (Caporaso *et al.*, 2012) which could explain the higher genome coverage in the samples sequenced using the agnostic approach on the HiSeq platform. Furthermore, agnostic sequencing is unbiased hence overcomes the issue of primers mismatches unlike the amplicon-based method that relies on pathogen-specific primers (Chen *et al.*, 2022). Agnostic sequencing has been deployed in whole genome sequencing of RVs in previous studies yielding high genome coverage (>70%) and remains an important component of surveillance of infectious diseases especially during outbreaks to inform rapid responses (Fujii *et al.*, 2019; Phan *et al.*, 2016).

This study had limitations. First, we analysed only sequences sampled from children admitted to a paediatric ward at KCH, therefore may not fully representative of the genetic diversity of G2P[4] RVA strains circulating at the Kenyan coast. Second, near-complete genomes were recovered during sequencing, with only 68% coverage generated for the VP4 segment. Finally, the global sequence dataset obtained from

GenBank and used in the phylogenetic comparison was a collection from numerous countries (Table 4.3). However, the genomes from those countries that may have introduced RVA vaccines into their NIPs lacked data regarding host vaccination status. Therefore, it was not possible to compare our study sequences from vaccinated children with global sequences obtained from vaccinated children.

# CHAPTER SIX

# CONCLUSION AND RECOMMENDATION

## 6.1 Conclusion

Both the agnostic and amplicon-based whole genome sequencing methods exhibited high genome recovery and coverage for the G2P[4] RVA strains. Our findings emphasize the importance of genomic surveillance to track the impact of vaccination on circulating RVA strains in Kenya. The G2P[4] strains that circulated in Kilifi 2012-2018, were distinct by vaccine period and clustered separate from the global strains. In addition, the observed amino acid sequences in the VP7 and VP4 antigenic epitopes were conserved across the Kilifi pre- and post-vaccine strains indicating that Rotarix® vaccine usage did result in new immune escape mutations in G2P[4] strains circulating in Kilifi.

## 6.2 Recommendations

i. A new study should design new primer set, particularly for the VP4 and NSP4 segments for the G2P[4] RVA strains for the amplicon-based whole genome sequencing method to improve on whole genome segment recovery.

ii. Further studies on the genomic epidemiology of G2P[4] strains should be conducted by extending the geographical sampling area through including health facilities across Kenya and from non-hospitalized infected children in the community to reveal the genomic epidemiology of G2P[4] circulating in Kenya.

iii. Studies should continue to conduct the genomic surveillance of RVA strains circulating in Kenya to monitor the evolving genetic diversity and transmission patterns of RVA strains circulating in the post-vaccine period, especially now that Kenya has transitioned from the Rotarix® vaccine to the Rotavac® vaccine (G9P[11]).

## REFERENCES


Abuga, J. A., Kariuki, S. M., Abubakar, A., Nyundo, C., Kinyanjui, S. M., Van Hensbroek, M. B., & Newton, C. R. J. C. (2022). Neurological impairment and disability in children in rural Kenya. *Developmental Medicine and Child Neurology*, *64*(3), 347–356. https://doi.org/10.1111/DMCN.15059

Agbemabiese, C. A., Nakagomi, T., Doan, Y. H., Do, L. P., Damanka, S., Armah, G. E., & Nakagomi, O. (2016). Genomic constellation and evolution of Ghanaian G2P[4] rotavirus strains from a global perspective. *Infection, Genetics and Evolution*, *45*, 122–131. https://doi.org/10.1016/j.meegid.2016.08.024

Agoti, C. N., Curran, M. D., Murunga, N., Ngari, M., Muthumbi, E., Lambisia, A. W., Frost, S. D. W., … & Drumright, L. N. (2022). Differences in epidemiology of enteropathogens in children pre- and post-rotavirus vaccine introduction in Kilifi, coastal Kenya. *Gut Pathogens*, *14*(1), 1–12. https://doi.org/10.1186/S13099-022-00506-Z/FIGURES/3

Aida, S., Nahar, S., Paul, S. K., Hossain, M. A., Kabir, M. R., Sarkar, S. R., Ahmed, S., … & Kobayashi, N. (2016). Whole genomic analysis of G2P[4] human Rotaviruses in Mymensingh, north-central Bangladesh. *Heliyon*, *2*(9), e00168. https://doi.org/10.1016/J.HELIYON.2016.E00168

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*, United Kingdom: Cambridge.

Aoki, S. T., Settembre, E. C., Trask, S. D., Greenberg, H. B., Harrison, S. C., & Dormitzer, P. R. (2009). Structure of rotavirus outer-layer protein

VP7 bound with a neutralizing Fab. *Science*, *324*(5933), 1444. https://doi.org/10.1126/SCIENCE.1170481

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., … & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5). http://www.ncbi.nlm.nih.gov/pubmed/22506599

Bányai, K., Kemenesi, G., Budinski, I., Földes, F., Zana, B., Marton, S., Varga-Kugler, R., … & Jakab, F. (2017). Candidate new rotavirus species in Schreiber's bats, Serbia. *Infection, Genetics and Evolution*, *48*, 19–26. https://doi.org/10.1016/J.MEEGID.2016.12.002

Bányai, K., László, B., Duque, J., Steele, A. D., Nelson, E. A. S., Gentsch, J. R., & Parashar, U. D. (2012). Systematic review of regional and temporal trends in global rotavirus strain diversity in the pre rotavirus vaccine era: insights for understanding the impact of rotavirus vaccination programs. *Vaccine*, *30 Suppl 1*(SUPPL. 1). https://doi.org/10.1016/J.VACCINE.2011.09.111

Barros De Arruda, L., Campos, F. S., Guimaraes Da Fonseca, F., Omatola, C. A., & Olaniran, A. O. (2022). Rotaviruses: From pathogenesis to disease control—A critical review. *Viruses*, *14*(5), 875.

Bei, Y., Pinet, K., Vrtis, K. B., Borgaro, J. G., Sun, L., Campbell, M., Apone, L., Langhorst, B. W., & Nichols, N. M. (2022). Overcoming variant mutation-related impacts on viral sequencing and detection methodologies. *Frontiers in Medicine*, *9*, 3159. https://doi.org/10.3389/FMED.2022.989913/BIBTEX

Bhandari, N., Rongsen-Chandola, T., Bavdekar, A., John, J., Antony, K., Taneja, S., Goyal, N., … & Bhan, M. K. (2014). Efficacy of a monovalent human-bovine (116E) rotavirus vaccine in Indian

children in the second year of life. *Vaccine*, *32 Suppl 1*(S1). https://doi.org/10.1016/J.VACCINE.2014.04.079

Bibera, G. L., Chen, J., Pereira, P., & Benninghoff, B. (2020). Dynamics of G2P[4] strain evolution and rotavirus vaccination: A review of evidence for Rotarix. *Vaccine*, *38*(35), 5591–5600. https://doi.org/10.1016/J.VACCINE.2020.06.059

Bishop, R. F., Davidson, G. P., Holmes, I. H., & Ruck, B. J. (1973). Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *The Lancet*, *302*(7841), 1281–1283. https://doi.org/10.1016/S0140-6736(73)92867-5

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114. https://doi.org/10.1093/ BIOINFORMATICS/BTU170

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Dillen, P. M. W., & Noordaa, J. van der. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, *28*(3), 495. /pmc/articles/PMC269651/ ?report=abstract

Burke, R. M., Tate, J. E., Kirkwood, C. D., Steele, A. D., & Parashar, U. D. (2019). Current and new rotavirus vaccines. *Current Opinion in Infectious Diseases*, *32*(5), 435–444. https://doi.org/10.1097/QCO.0000000000000572

Burnett, E., Parashar, U. D., & Tate, J. E. (2020). Real-world effectiveness of rotavirus vaccines, 2006–19: a literature review and meta-analysis. *The Lancet Global Health*, *8*(9), e1195–e1202. https://doi.org/10.1016/S2214-109X(20)30262-X

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., & & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq

and MiSeq platforms. *The ISME Journal 2012, 6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, *28*(4), 464. https://doi.org/ 10.1093/BIOINFORMATICS/BTR703

Chen, Y., Fan, L. C., Chai, Y. H., & Xu, J. F. (2022). Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. *The Clinical Respiratory Journal*, *16*(10), 646. https://doi.org/10.1111/CRJ.13538

Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., & Van Der Hoek, L. (2014). Full Genome Virus Detection in Fecal Samples Using Sensitive Nucleic Acid Preparation, Deep Sequencing, and a Novel Iterative Sequence Classification Algorithm. *PLOS ONE*, *9*(4), e93269. https://doi.org/10.1371/JOURNAL.PONE.0093269

Crawford, S. E., Ramani, S., Tate, J. E., Parashar, U. D., Svensson, L., Hagbom, M., Franco, M. A., Greenberg, H. B., O'Ryan, M., Kang, G., Desselberger, U., & Estes, M. K. (2017). Rotavirus infection. *Nature Reviews. Disease Primers*, *3*. https://doi.org/10.1038/NRDP.2017.83

Dennis, A. F., McDonald, S. M., Payne, D. C., Mijatovic-Rustempasic, S., Esona, M. D., Edwards, K. M., Chappell, J. D., & Patton, J. T. (2014). Molecular epidemiology of contemporary G2P[4] human rotaviruses cocirculating in a single U.S. community: footprints of a globally transitioning genotype. *Journal of Virology*, *88*(7), 3789–3801. https://doi.org/10.1128/JVI.03516-13

Dennis, F., Fujii, Y., Haga, K., Damanka, S., Lartey, B., Agbemabiese, C. A., Ohta, N., Armah, G. E., & Katayama, K. (2014). Identification of novel Ghanaian G8P[6] human-bovine reassortant rotavirus strain by next generation sequencing. *PloS One*, *9*(6), e100699. https://doi.org/10.1371/journal.pone.0100699

Desselberger, U. (2014). Rotaviruses. *Virus Research*, *190*, 75–96. https://doi.org/10.1016/J.VIRUSRES.2014.06.016

Do, L. P., Doan, Y. H., Nakagomi, T., Gauchan, P., Kaneko, M., Agbemabiese, C., Dang, A. D., & Nakagomi, O. (2015). Whole genome analysis of Vietnamese G2P[4] rotavirus strains possessing the NSP2 gene sharing an ancestral sequence with Chinese sheep and goat rotavirus strains. *Microbiology and Immunology*, *59*(10), 605–613. https://doi.org/10.1111/1348-0421.12323

Doan, Y. H., Nakagomi, T., Agbemabiese, C. A., & Nakagomi, O. (2015). Changes in the distribution of lineage constellations of G2P[4] Rotavirus A strains detected in Japan over 32years (1980-2011). *Infection, Genetics and Evolution*, *34*, 423–433. https://doi.org/10.1016/j.meegid.2015.05.026

Doan, Y. H., Nakagomi, T., & Nakagomi, O. (2012). Repeated circulation over 6 years of intergenogroup mono-reassortant G2P[4] rotavirus strains with genotype N1 of the NSP2 gene. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *12*(6), 1202–1212. https://doi.org/10.1016/J.MEEGID.2012.04.023

Doan, Y. H., T, N., NA, C., BD, P., JB, S., & O, N. (2011). The occurrence of amino acid substitutions D96N and S242N in VP7 of emergent G2P[4] rotaviruses in Nepal in 2004-2005: a global and evolutionary perspective. *Archives of Virology*, *156*(11), 1969–1978. https://doi.org/10.1007/S00705-011-1083-Z

Donato, C. M., & Bines, J. E. (2021). Rotaviruses and Rotavirus Vaccines. *Pathogens*, *10*(8). https://doi.org/10.3390/PATHOGENS10080959

Donato, C. M., Pingault, N., Demosthenous, E., Roczo-farkas, S., & Bines, J. E. (2021). Characterisation of a G2P[4] Rotavirus Outbreak in Western Australia, Predominantly Impacting Aboriginal Children. *Pathogens 2021, Vol. 10, Page 350*, *10*(3), 350. https://doi.org/10.3390/PATHOGENS10030350

Donato, C. M., Zhang, Z. A., Donker, N. C., & Kirkwood, C. D. (2014). Characterization of G2P[4] rotavirus strains associated with increased detection in Australian states using the RotaTeq® vaccine during the 2010-2011 surveillance period. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *28*, 398–412. https://doi.org/10.1016/j.meegid.2014.05.020

Dormitzer, P. R., Sun, Z. Y. J., Wagner, G., & Harrison, S. C. (2002). The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO Journal*, *21*(5), 885–897. https://doi.org/10.1093/EMBOJ/21.5.885

Dóró, R., László, B., Martella, V., Leshem, E., Gentsch, J., Parashar, U., & Bányai, K. (2014). Review of global rotavirus strain prevalence data from six years post vaccine licensure surveillance: is there evidence of strain selection from vaccine pressure? *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *28*, 446–461. https://doi.org/10.1016/J.MEEGID.2014.08.017

Dung, T. T. N., Duy, P. T., Sessions, O. M., Sangumathi, U. K., Phat, V. V., Tam, P. T. T., To, N. T. N., Phuc, T. M., … & Baker, S. (2017). A universal genome sequencing method for rotavirus A from human fecal samples which identifies segment reassortment and multi-

genotype mixed infection. *BMC Genomics*, *18*(1), 324. https://doi.org/10.1186/s12864-017-3714-6

Dyall-Smith, M. L., Lazdins, I., Tregear, G. W., & Holmes, I. H. (1986). Location of the major antigenic sites involved in rotavirus serotype-specific neutralization. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(10), 3465–3468. https://doi.org/10.1073/PNAS.83.10.3465

Edgar, R. C., & Bateman, A. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/ BIOINFORMATICS/BTQ461

Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S., & Hayashi, M. (2005). Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Research*, *33*(6), 1–11. https://doi.org/10.1093/NAR/GNI064

Florez, I. D., Niño-Serna, L. F., & Beltrán-Arroyave, C. P. (2020). Acute Infectious Diarrhea and Gastroenteritis in Children. *Current Infectious Disease Reports*, *22*(2), 1–12. https://doi.org/10.1007/S11908-020-0713-6/METRICS

Fujii, Y., Doan, Y. H., Suzuki, Y., Nakagomi, T., Nakagomi, O., & Katayama, K. (2019). Study of complete genome sequences of Rotavirus a epidemics and evolution in Japan in 2012-2014. *Frontiers in Microbiology*, *10*(JAN). https://doi.org/10.3389/FMICB.2019.00038/FULL

Fujii, Y., Shimoike, T., Takagi, H., Murakami, K., Todaka-Takai, R., Park, Y., & Katayama, K. (2012). Amplification of all 11 RNA segments of group A rotaviruses based on reverse transcription polymerase chain

reaction. *Microbiology and Immunology*, *56*(9), 630–638. https://doi.org/10.1111/J.1348-0421.2012.00479.X

Gauthier, N. P. G., Chorlton, S. D., Krajden, M., & Manges, A. R. (2023). Agnostic Sequencing for Detection of Viral Pathogens. *Clinical Microbiology Reviews*. https://doi.org/10.1128/CMR.00119-22

Gene Codes. (2023). *Sequencher DNA Sequence Analysis Software from Gene Codes Corporation*. http://www.genecodes.com/

Ghosh, S., Adachi, N., Gatheru, Z., Nyangao, J., Yamamoto, D., Ishino, M., Urushibara, N., & Kobayashi, N. (2011). Whole-genome analysis reveals the complex evolutionary dynamics of Kenyan G2P[4] human rotavirus strains. *The Journal of General Virology*, *92*(Pt 9), 2201–2208. https://doi.org/10.1099/vir.0.033001-0

Giammanco, G. M., Bonura, F., Zeller, M., Heylen, E., Van Ranst, M., Martella, V., Bányai, K., Matthijnssens, J., & De Grazia, S. (2014). Evolution of DS-1-like human G2P[4] rotaviruses assessed by complete genome analyses. *Journal of General Virology*, *95*(PART 1), 91–109. https://doi.org/10.1099/vir.0.056788-0

Gikonyo, J. N. u., Mbatia, B., Okanya, P. W., Obiero, G. F. O., Sang, C., Steele, D., & Nyangao, J. (2020). Post-vaccine rotavirus genotype distribution in Nairobi County, Kenya. *International Journal of Infectious Diseases*, *100*, 434. https://doi.org/10.1016/J.IJID.2020.09.005

Glass, R. I., Parashar, U. D., Bresee, J. S., Turcios, R., Fischer, T. K., Widdowson, M.-A., Jiang, B., & Gentsch, J. R. (2006). Rotavirus vaccines: current prospects and future challenges. *The Lancet*, *368*(9532), 323–332. https://doi.org/10.1016/S0140-6736(06)68815-6

Gómez, M. M., Carvalho-Costa, F. A., Volotão, E. de M., Rose, T. L., da Silva, M. F. M., Fialho, A. M., Assis, R. M. S., … & Leite, J. P. G. (2014). Prevalence and genomic characterization of G2P[4] group A rotavirus strains during monovalent vaccine introduction in Brazil. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *28*, 486–494. https://doi.org/10.1016/j.meegid.2014.09.012

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072. https://doi.org/10.1093/BIOINFORMATICS/BTT086

Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518. https://doi.org/10.1093/MOLBEV/MSX281

Hoxie, I., & Dennehy, J. J. (2021). Rotavirus a genome segments show distinct segregation and codon usage patterns. *Viruses*, *13*(8), 1460. https://doi.org/10.3390/V13081460/S1

IVAC. (2022). *International Vaccine Access Center (IVAC), Johns Hopkins Bloomberg School of Public Health.* https://view-hub.org/map/?set=current-vaccine-intro-status&group=vaccine-introduction&category=rv

Izzo, M. M., Kirkland, P. D., Gu, X., Lele, Y., Gunn, A. A., & House, J. K. (2012). Comparison of three diagnostic techniques for detection of rotavirus and coronavirus in calf faeces in Australia. *Australian Veterinary Journal*, *90*(4), 122–129. https://doi.org/10.1111/j.1751-0813.2011.00891.x

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier

transform. *Nucleic Acids Research*, *30*(14), 3059. https://doi.org/10.1093/NAR/GKF436

Khagayi, S., Omore, R., Otieno, G. P., Ogwel, B., Ochieng, J. B., Juma, J., Apondi, E., … & Verani, J. R. (2020). *Effectiveness of Monovalent Rotavirus Vaccine Against Hospitalization With Acute Rotavirus Gastroenteritis in Kenyan Children*. *70*(11), 2298–2305. https://academic.oup.com/cid/ article/70/11/ 2298/5536640

Khoury, H., Ogilvie, I., El Khoury, A. C., Duan, Y., & Goetghebeur, M. M. (2011). Burden of rotavirus gastroenteritis in the Middle Eastern and North African pediatric population. *BMC Infectious Diseases*, *11*. https://doi.org/10.1186/1471-2334-11-9

King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (2011). Virus taxonomy: Ninth report of the international committee on taxonomy of viruses. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier. http://www.sciencedirect.com:5070/ book/9780123846846/ virus-taxonomy

Kirkwood, C. D. (2010). Genetic and Antigenic Diversity of Human Rotaviruses: Potential Impact on Vaccination Programs. *The Journal of Infectious Diseases*, *202*(Supplement_1), S43–S48. https://doi.org/10.1086/653548

Kyo, K., Takano, C., Kasuga, Y., Ogawa, E., Ishige, M., Pham, N. T. K., Okitsu, S., … & Morioka, I. (2021). Severe rotavirus gastroenteritis in children older than 5 years after vaccine introduction. *Journal of Infection and Chemotherapy*, *27*(4), 598–603. https://doi.org/10.1016/J.JIAC.2020.11.018

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics (Oxford, England)*, *30*(22), 3276–3278. https://doi.org/10.1093/BIOINFORMATICS/BTU531

Lee, B. (2021). Update on rotavirus vaccine underperformance in low- to middle-income countries and next-generation vaccines. *Human Vaccines and Immunotherapeutics*, *17*(6), 1787–1802. https://doi.org/10.1080/ 21645515. 2020.1844525

Lee, S., Choi, S., Kim, J. S., Lee, E. J., Hyun, J., & Kim, H. S. (2019). Whole-genome analysis of rotavirus G4P[6] strains isolated from Korean neonates: association of Korean neonates and rotavirus P[6] genotypes. *Gut Pathogens*, *11*(1). https://doi.org/10.1186/S13099-019-0318-5

Leshem, E., & Lopman, B. A. (2018). Viral Gastroenteritis. In *Principles and Practice of Pediatric Infectious Diseases*. Elsevier. https://doi.org/10.1016/B978-0-323-40181-4.00056-6

Leshem, E., Lopman, B., Glass, R., Gentsch, J., Bányai, K., Parashar, U., & Patel, M. (2014). Distribution of rotavirus strains and strain-specific effectiveness of the rotavirus vaccine after its introduction: A systematic review and meta-analysis. *The Lancet Infectious Diseases*, *14*(9), 847–856. https://doi.org/10.1016/S1473-3099(14)70832-1

Liu. (2014). Fields Virology, 6th Edition. *Clinical Infectious Diseases*, *59*(4), 613–613. https://doi.org/10.1093/CID/CIU346

Liu, J., Gratz, J., Amour, C., Nshama, R., Walongo, T., Maro, A., Mduma, E., … & Houpt, E. R. (2016). Optimization of quantitative PCR methods for enteropathogen detection. *PLoS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0158199

Lopman, B. A., Pitzer, V. E., Sarkar, R., Gladstone, B., Patel, M., Glasser, J., Gambhir, M., Atchison, C., Grenfell, B. T., Edmunds, W. J., Kang, G., & Parashar, U. D. (2012). Understanding Reduced Rotavirus Vaccine Efficacy in Low Socio-Economic Settings. *PLOS ONE*, *7*(8), e41720. https://doi.org/10.1371/ JOURNAL.PONE.0041720

Luchs, A., & Timenetsky, M. do C. S. T. (2016). Group A rotavirus gastroenteritis: post-vaccine era, genotypes and zoonotic transmission. *Einstein*, *14*(2), 278. https://doi.org/10.1590/S1679-45082016RB3582

Magagula, N. B., Esona, M. D., Nyaga, M. M., Stucker, K. M., Halpin, R. A., Stockwell, T. B., Seheri, M. L., … & M Jeffrey. (2015). Whole Genome Analyses of G1P[8] Rotavirus Strains From Vaccinated and Non-Vaccinated South African Children Presenting With Diarrhea HHS Public Access. *J Med Virol*, *87*(1), 79–101. https://doi.org/10.1002/jmv.23971

Matthijnssens, J., Ciarlet, M., McDonald, S. M., Attoui, H., Bányai, K., Brister, J. R., Buesa, J., …& van Ranst, M. (2011). Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Archives of Virology*, *156*(8), 1397–1413. https://doi.org/10.1007/S00705-011-1006-Z

Matthijnssens, J., Ciarlet, M., Rahman, M., Attoui, H., Bányai, K., Estes, M. K., Gentsch, J. R., … & Van Ranst, M. (2008). Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. *Archives of Virology*, *153*(8), 1621–1629. https://doi.org/10.1007/s00705-008-0155-1

Matthijnssens, J., De Grazia, S., Piessens, J., Heylen, E., Zeller, M., Giammanco, G. M., Bányai, K., Buonavoglia, C., Ciarlet, M., Martella, V., & Van Ranst, M. (2011). Multiple reassortment and interspecies transmission events contribute to the diversity of feline,

canine and feline/canine-like human group A rotavirus strains. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *11*(6), 1396–1406. https://doi.org/10.1016/J.MEEGID.2011.05.007

Matthijnssens, J., Nuyts, V., Heylen, E., De Coster, S., Conceição-Neto, N., Van Ranst, M., & Matthijnssens, J. (2016). Emergence of human G2P[4] rotaviruses containing animal derived gene segments in the post-vaccine era. *Scientific Reports*, *6*(1), 36841. https://doi.org/10.1038/srep36841

Matthijnssens, J., Rahman, M., & Van Ranst, M. (2008). Two out of the 11 genes of an unusual human G6P[6] rotavirus isolate are of bovine origin. *The Journal of General Virology*, *89*(Pt 10), 2630–2635. https://doi.org/10.1099/vir. 0.2008/003780-0

Matthijnssens, J., & Van Ranst, M. (2012). Genotype constellation and evolution of group A rotaviruses infecting humans. *Current Opinion in Virology*, *2*(4), 426–433. https://doi.org/10.1016/J.COVIRO.2012.04.007

Mihalov-Kovács, E., Gellért, Á., Marton, S., Farkas, S. L., Fehér, E., Oldal, M., Jakab, F., Martella, V., & Bányai, K. (2015). Candidate New Rotavirus Species in Sheltered Dogs, Hungary - Volume 21, Number 4—April 2015 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, *21*(4), 660–663. https://doi.org/10.3201/EID2104.141370

Mijatovic-Rustempasic, S., Roy, S., Teel, E. N., Weinberg, G. A., Payne, D. C., Parashar, U. D., & Bowen, M. D. (2016). Full genome characterization of the first G3P[24] rotavirus strain detected in humans provides evidence of interspecies reassortment and mutational saturation in the VP7 gene. *Journal of General Virology*,

*97*(2), 389–402. https://doi.org/10.1099/JGV.0. 000349/ CITE/ REFWORKS

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/MOLBEV/MSAA015

Mokoena, F., Esona, M. D., Seheri, L. M., Nyaga, M. M., Magagula, N. B., Mukaratirwa, A., Mulindwa, A., … & Steele, A. D. (2021). Whole Genome Analysis of African G12P[6] and G12P[8] Rotaviruses Provides Evidence of Porcine-Human Reassortment at NSP2, NSP3, and NSP4. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/FMICB.2020.604444/FULL

Monnier, N., Higo-Moriguchi, K., Sun, Z.-Y. J., Prasad, B. V. V., Taniguchi, K., & Dormitzer, P. R. (2006). High-Resolution Molecular and Antigen Structure of the VP8* Core of a Sialic Acid-Independent Human Rotavirus Strain. *Journal of Virology*, *80*(3), 1513. https://doi.org/10.1128/JVI.80.3.1513-1523.2006

Morozova, O. V., Sashina, T. A., Fomina, S. G., & Novikova, N. A. (2015). Comparative characteristics of the VP7 and VP4 antigenic epitopes of the rotaviruses circulating in Russia (Nizhny Novgorod) and the Rotarix and RotaTeq vaccines. *Archives of Virology*, *160*(7), 1693–1703. https://doi.org/10.1007/S00705-015-2439-6

Moss, D. M., Priest, J. W., Boyd, A., Weinkopff, T., Kucerova, Z., Beach, M. J., & Lammie, P. J. (2011). Multiplex Bead Assay for Serum Samples from Children in Haiti Enrolled in a Drug Study for the Treatment of Lymphatic Filariasis. *The American Journal of Tropical Medicine and Hygiene*, *85*(2), 229. https://doi.org/10.4269/AJTMH.2011.11-0029

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Molecular biology and evolution, 30(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Mwanga, M. J., Owor, B. E., Ochieng, J. B., Ngama, M. H., Ogwel, B., Onyango, C., Juma, J., … & Nokes, D. J. (2020). Rotavirus group A genotype circulation patterns across Kenya before and after nationwide vaccine introduction, 2010-2018. *BMC Infectious Diseases*, *20*(1), 504. https://doi.org/10.1186/s12879-020-05230-0

Mwangi, P. N., Page, N. A., Seheri, M. L., Mphahlele, M. J., Nadan, S., Esona, M. D., Kumwenda, B., Kamng'ona, A. W., … & Nyaga, M. M. (2022). Evolutionary changes between pre- and post-vaccine South African group A G2P[4] rotavirus strains, 2003–2017. *Microbial Genomics*, *8*(4), 000809. https://doi.org/10.1099/MGEN.0.000809

Mwangi, P. N., Potgieter, R. L., Simwaka, J., Mpabalwani, E. M., Mwenda, J. M., Mogotsi, M. T., Magagula, N., … & Nyaga, M. M. (2023). Genomic Analysis of G2P[4] Group A Rotaviruses in Zambia Reveals Positive Selection in Amino Acid Site 7 of Viral Protein 3. *Viruses*, *15*(2). https://doi.org/10.3390/V15020501

NCBI. (2022). *NCBI: https://www.ncbi.nlm.nih.gov/nuccore/?term=(G2P4)+AND+%22cds%22AND+%22Rotavirus+A%22%5Bporgn%3A__txid28875%5D*. (G2P4) AND 'Cds'AND 'Rotavirus A'[Porgn:__txid28875] - Nucleotide - . https://www.ncbi.nlm.nih.gov/nuccore/?term=(G2P4)+AND+%22cds%22AND+%22Rotavirus+A%22%5Bporgn%3A__txid28875%5D

Nguyen, A. T., Tran, T. T., Hoang, V. M. T., Nghiem, N. M., Le, N. N. T., Le, T. T. M., Phan, Q. T., … & Le, T. Van. (2016). Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. *Virology Journal*, *13*(1). https://doi.org/10.1186/S12985-016-0580-9

Nokes, D. J., Abwao, J., Pamba, A., Peenze, I., Dewar, J., Maghenda, J. K., Gatakaa, H., … & Williams, T. N. (2008). Incidence and Clinical Characteristics of Group A Rotavirus Infections among Children Admitted to Hospital in Kilifi , Kenya. *PLoS MEDICINE*, *5*(7), 1154–1162. https://doi.org/10.1371/ journal. pmed.0050153

NSW. (n.d.). *KEMRI Wellcome Trust Biobank*.

O'Ryan, M. (2017). Rotavirus Vaccines: a story of success with challenges ahead. *F1000Research*, *6*. https://doi.org/10.12688/F1000RESEARCH.11912.1

Otieno, G. P., Bottomley, C., Khagayi, S., Adetifa, I., Ngama, M., Omore, R., Ogwel, B., … & James Nokes, D. (2020). Impact of the introduction of rotavirus vaccine on hospital admissions for diarrhea among children in Kenya: A controlled interrupted time-series analysis. *Clinical Infectious Diseases*, *70*(11), 2306–2313. https://doi.org/10.1093/cid/ciz912

Ouermi, D., Soubeiga, D., Nadembèga, W. M. C., Sawadogo, P. M., Zohoncon, T. M., Obiri-Yeboah, D., Djigma, F. W., Nordgren, J., & Simpore, J. (2017). Molecular epidemiology of rotavirus in children under five in Africa (2006-2016): A systematic review. *Pakistan Journal of Biological Sciences*, *20*(2), 59–69. https://doi.org/10.3923/PJBS.2017.59.69

Patton, J. T. (2012). Rotavirus Diversity and Evolution in the Post-Vaccine World. *Discovery Medicine*, *13*(68), 85. /pmc/articles/PMC3738915/

Phan, M. V. T., Anh, P. H., Cuong, N. Van, Munnink, B. B. O., Hoek, L. van der, My, P. T., Tri, T. N., …& Consortium. (2016). Unbiased whole-genome deep sequencing of human and porcine stool samples reveals circulation of multiple groups of rotaviruses and a putative zoonotic infection. *Virus Evolution*, *2*(2), vew027. https://doi.org/10.1093/ve/vew027

Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., … & Scheuermann, R. H. (2012). ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, *40*(Database issue), D593. https://doi.org/10.1093/NAR/GKR859

Posit team. (2022). *RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA.* http://www.posit.co/.

Prasad, B. V. V., & Chiu, W. (1994). Structure of Rotavirus. *Current Topics In Microbiology And Immunology*, 9–29. https://doi.org/10.1007/978-3-642-78256-5_2

Rasebotsa, S., Mwangi, P. N., Mogotsi, M. T., Sabiu, S., Magagula, N. B., Rakau, K., Uwimana, J., … & Nyaga, M. M. (2020). Whole genome and in-silico analyses of G1P[8] rotavirus strains from pre- and post-vaccination periods in Rwanda. *Scientific Reports 2020 10:1*, *10*(1), 1–22. https://doi.org/10.1038/s41598-020-69973-1

RCWG. (2021). *Rotavirus Classification Working Group: RCWG - Laboratory of Viral Metagenomics*. Laboratory of Viral Metagenomics, KU Leuven. https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification/rcwg

Reiner, R. C., Wiens, K. E., Deshpande, A., Baumann, M. M., Lindstedt, P. A., Blacker, B. F., Troeger, C. E., … & Hay, S. I. (2020). Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000-17: Analysis for the Global Burden of Disease Study 2017. *The Lancet*, *395*(10239), 1779–1801. https://doi.org/10.1016/S0140-6736(20)30114-8

Rennert, W., Hindiyeh, M., Allahham, M., Mercer, L. D., Hamad, K. I., Ghuneim, N. I., A. M. Eljaro, Z., … & Marzouqa, H. (2023). Introducing ROTAVAC® to the occupied Palestinian Territories: Impact on diarrhea incidence, rotavirus prevalence and genotype composition. *Vaccine*, *41*(4), 945. https://doi.org/10.1016/J.VACCINE.2022.12.046

ROTA. (2022). *Global Introduction Status | Rota Council*. https://preventrotavirus.org/vaccine-introduction/global-introduction-status/

Sadiq, A., Bostan, N., Yinda, K. C., Naseem, S., & Sattar, S. (2018). Rotavirus: Genetics, pathogenesis and vaccine advances. *Reviews in Medical Virology*, *28*(6), e2003. https://doi.org/10.1002/RMV.2003

Scott, J. A. G., Bauni, E., Moisi, J. C., Ojal, J., Gatakaa, H., Nyundo, C., Molyneux, C. S., … & Williams, T. N. (2012). Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *International Journal of Epidemiology*, *41*(3), 650–657. https://doi.org/10.1093/IJE/DYS062

Seheri, L. M., Magagula, N. B., Peenze, I., Rakau, K., Ndadza, A., Mwenda, J. M., Weldegebriel, G., Steele, A. D., & Mphahlele, M. J. (2018). Rotavirus strain diversity in Eastern and Southern African countries before and after vaccine introduction. *Vaccine*, *36*(47), 7222–7230. https://doi.org/10.1016/J. VACCINE. 2017.11.068

Sint, D., Raso, L., & Traugott, M. (2012). Advances in multiplex PCR: balancing primer efficiencies and improving detection success. *Methods in Ecology and Evolution*, *3*(5), 898. https://doi.org/10.1111/J.2041-210X.2012.00215.X

Subha, K., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Svensson, L., Desselberger, U., Estes, M. K., & Greenberg, H. B. (2016). *Viral gastroenteritis : molecular epidemiology and pathogenesis*. http://www.sciencedirect.com: 5070/book/9780128022412/viral-gastroenteritis

Tate, J. E., Burton, A. H., Boschi-Pinto, C., & Parashar, U. D. (2016). Global, Regional, and National Estimates of Rotavirus Mortality in Children <5 Years of Age, 2000–2013. *Clinical Infectious Diseases*, *62*(2), 96–105. https://doi.org/10.1093/cid/civ1013

Thanh, H. D., Tran, V. T., Lim, I., & Kim, W. (2018). Emergence of Human G2P[4] Rotaviruses in the Post-vaccination Era in South Korea: Footprints of Multiple Interspecies Re-assortment Events. *Scientific Reports 2018 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-24511-y

Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner, R. C., Fullman, N., … & Mokdad, A. H. (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases*, *17*(9), 909–948. https://doi.org/10.1016/S1473-3099(17)30276-1

Troeger, C., Khalil, I. A., Rao, P. C., Cao, S., Blacker, B. F., Ahmed, T., Armah, G., … & R. C. (2018). Rotavirus Vaccination and the Global Burden of Rotavirus Diarrhea among Children Younger Than 5 Years. *JAMA Pediatrics*, *172*(10), 958–965. https://doi.org/10.1001/jamapediatrics.2018.1960

Wandera, E. A., Mohammad, S., Bundi, M., Nyangao, J., Galata, A., Kathiiko, C., Odoyo, E., … & Ichinose, Y. (2018). Impact of rotavirus vaccination on rotavirus hospitalisation rates among a resource-limited rural population in Mbita, Western Kenya. *Tropical Medicine & International Health*, *23*(4), 425–432. https://doi.org/10.1111/TMI.13040

Wang, Y. H., Pang, B. B., Zhou, X., Ghosh, S., Tang, W. F., Peng, J. S., Hu, Q., Zhou, D. J., & Kobayashi, N. (2013). Complex evolutionary patterns of two rare human G3P[9] rotavirus strains possessing a feline/canine-like H6 genotype on an AU-1-like genotype constellation. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *16*, 103–112. https://doi.org/10.1016/J.MEEGID.2013.01.016

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Molecular biology and evolution, 35(3), 773-777.

Wei, J., Li, J., Zhang, X., Tang, Y., Wang, J., & Wu, Y. (2009). A naturally processed epitope on rotavirus VP7 glycoprotein recognized by HLA-A2.1-restricted cytotoxic CD8+ T cells. *Viral Immunology*, *22*(3), 189–194. https://doi.org/10.1089/VIM.2008.0091

Were, F. N., Jere, K. C., Armah, G. E., Mphahlele, M. J., Mwenda, J. M., & Steele, A. D. (2022). Maintaining Momentum for Rotavirus

Immunization in Africa during the COVID-19 Era: Report of the 13th African Rotavirus Symposium. *Vaccines 2022, 10*(9), 1463. https://doi.org/10.3390/VACCINES10091463

WHO. (2021). *Rotavirus vaccines: WHO position paper - July 2021*. https://www.who.int/publications/i/item/WHO-WER9628

Zeller, M., Heylen, E., Tamim, S., McAllen, J. K., Kirkness, E. F., Akopov, A., De Coster, S., Van Ranst, M., & Matthijnssens, J. (2017). Comparative analysis of the Rotarix™ vaccine strain and G1P[8] rotaviruses detected before and after vaccine introduction in Belgium. *PeerJ, 5*, e2733. https://doi.org/10. 7717/peerj.2733

Zeller, M., Nuyts, V., Heylen, E., Coster, S. De, Conceição-Neto, N., Ranst, M. Van, & Matthijnssens, J. (2016). Emergence of human G2P[4] rotaviruses containing animal derived gene segments in the post-vaccine era. *Scientific Reports 2017 7:1*, *7*(1), 1–1. https://doi.org/10.1038/srep39436

## APPENDICES

### Appendix I: Summary of Studies on Genomic Characterisation and Evolution of G2P[4] Strains

| Authors | Country | Period of study | Number of genomes | Findings |
|---|---|---|---|---|
| (Ghosh *et al.*, 2011) | Kenya | 1982-1989 | 2 | • Strain D205/1989 had a typical DS-1 genotype constellation.<br>• Strain AK26/1982 was a mono-reassortant at NSP2 gene from a Wa-like genogroup of animal origin. |
| (Doan *et al.*, 2012) | Japan | 1981-1991 | 2 | • A mono-reassortant strain AU605/1986 with a Wa-like genogroup NSP2 gene of human origin circulated in epidemic seasons dominated by non-G2P[4] genotypes. |
| (Giammanco *et al.*, 2014) | Italy | 1996-2011 | 9 | • Strain PAI11/1996 clustered with reassortant Kenyan AK26/1982 strain and other similar global G2P[4] strains that circulated between 1982 and 1999. |

| | | | | |
|---|---|---|---|---|
| | | | | • Italian G2P[4] strains that circulated between 2004 and 2011 differed genetically from the PAI11/1996 strain in all 11 segments and clustered with global strains that circulated between 2000 and 2011. |
| (Dennis *et al.*, 2014) | USA | 2010-2011 | 12 | • Typical DS-1 genotype strains that formed three different clusters across 10 segments with no evidence of reassortment circulated during the study period.<br>• Study strains clustered with global G2P[4] strains that circulated from 2000 to 2010.<br>• VP7 antigenic epitopes of the three clusters were similar. |
| (Gómez *et al.*, 2014) | Brazil | 2005-2011 | 11 | • Brazilian G2P[4] strains from vaccinated and unvaccinated children clustered together for all the 11 segments.<br>• Study strains were interspersed with contemporary global strains.<br>• Similar amino acid mutations in the VP7 and VP8* antigenic epitopes were observed between pre- and post- Rotarix® vaccine introduction strains. |
| (Donato *et al.*, 2014) | Australia | 2019 | 4 | • The four G2P[4] strains that circulated in Australia possessed a typical DS-1 like genetic constellation and clustered with global G2P[4] strains that |

| | | | | |
|---|---|---|---|---|
| | | | | circulated from 2005 to 2011.<br>• All the strains harbored conserved amino acid changes (A87T, D96N and S213D) in the VP7 antigenic epitopes relative to the G2 component of the RotaTeq® strain. |
| (Do *et al.*, 2015) | Vietnam | 2008 | 2 | • The VP1-VP4, VP6, VP7, NSP1, NSP3, and NSP5 segments clustered with contemporaneous sequences that circulated from 2000 to 2008.<br>• The NSP2 genes of the study strains were interspersed with NSP2 genes of G10P[15] strains obtained from Chinese sheep.<br>• The NSP4 genes clustered into an emergent lineage VIII that diverged from lineage VI. |
| (Doan *et al.*, 2015) | Japan | 1983-2011 | 19 | • Eight G2P[4] strains possessed a typical DS-1 genetic backbone, while 11 strains were mono-reassortants of Wa-like NSP2 gene on the DS-1 genetic background.<br>• In the global context, G2P[4] strains have evolved in stepwise changes overtime from lineage I to V or from I to IVa in all gene segments.<br>• Some G2P[4] strains in the global context underwent intragenotype |

| | | | | |
|---|---|---|---|---|
| | | | | reassortments after 2004 in the VP7, VP3, and NSP4 genes to give rise to new lineages V (for VP7 segment), VI and/or VII (for VP3 and NSP4 segments). |
| (Aida *et al.*, 2016) | Bangladesh | 2010-2013 | 16 | • The G2P[4] strains that circulated in Bangladesh possessed a typical DS-1 backbone and clustered with contemporary G2P[4] global strains that circulated between 2005 and 2013 for all genome segments.<br>• Amino acid changes in antigenic epitopes of VP7 and VP4 genes of Bangladesh G2P[4] strains were similar to those of global strains circulating after 2000.<br>• However, the Bangladesh G2P[4] strains that circulated in 2013 harbored novel amino acid mutations; T88A and Q114P in the VP4 antigenic epitopes relative to the contemporary G2P[4] global strains. |
| (Agbemabiese *et al.*, 2016) | Ghana | 2008-2013 | 6 | • All the study G2P[4] strains had a typical DS-1 genetic backbone and clustered with contemporary G2P[4] global strains that circulated from 2005 to 2011.<br>• However, the NSP4 segment of G2P[4] strains that circulated in Ghana in |

| | | | | |
|---|---|---|---|---|
| | | | | 2009 along with some strains that circulated in Africa clustered into African specific lineages IX and X. |
| (Zeller *et al.*, 2016) | Belgium | 1999-2013 | 28 | • Two Belgian G2P[4] strains harbored an H3 (AU-1-like gene) NSP5 on the DS-1 genetic background, while 26 possessed a typical DS-1 genetic backbone.<br>• A wide variation in genetic composition was observed across the 11 genome segments of the Belgian G2P[4] strains; 12 strains both from pre- and post- Rotarix® vaccine periods clustered with human DS-1 strains; 10 strains were interspersed with bovine RVA strains in at least one genome segment; while two strains were closely interspersed with bovine RVAs in six segments. |
| (Thanh *et al.*, 2018) | South Korea | 2010-2015 | 12 | • The study G2P[4] strains had a typical DS-1 genetic backbone.<br>• Eight G2P[4] strains were reassortants with one or more gene segments (VP1, VP3 and/or NSP4) from animal infecting RVAs.<br>• Four study strains clustered with contemporary human G2P[4] strains that circulated globally between 2005 and 2015 for all the 11 genome segments. |

| | | | | |
|---|---|---|---|---|
| | | | | • Four conserved amino acid mutations (A87T, D96N, S213D, and S242N) were observed in the VP7 antigenic epitopes of the study strains relative to the G2 component of the RotaTeq® strain. |
| (Donato *et al.*, 2021) | Australia | 2017 | 3 | • The three G2P[4] study strains possessed a typical DS-1 genotype and were interspersed with contemporary G2P[4] strains that circulated in Japan in 2016 and 2017 across the 11 genome segments.<br>• The study strains harbored two conserved amino acid changes (D96N and S213D) in the VP7 antigenic epitopes relative to the G2 component of the RotaTeq® strain. |
| (Mwangi *et al.*, 2022) | South Africa | 2003-2017 | 103 | • The G2P[4] strains circulating in South Africa had a typical DS-1 genetic backbone and clustered separated by vaccination period.<br>• Both pre- and post-vaccine strains had conserved amino acid changes (A87T, D96N, N125T, V129M, N213D, N242S) in the VP7 antigenic epitopes relative to ancestral DS-1 strain. |

## Appendix II: Ethical Approval

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel: (254) 2722541, 2713349, 0722-205901,0733-400003, Fax: (254) (020) 2720030
Email: director@kemri.org, info@kemri.org, Website. www.kemri.org

**KEMRI/RES/7/3/1** **October 27, 2020**

**TO:** **PROF. JAMES NOKES**
**PRINCIPAL INVESTIGATOR**

**THROUGH:** **THE DEPUTY DIRECTOR, CGMR-C,**
**KILIFI**

Dear Sir,

**RE:** **SSC PROTOCOL NO. 2861 (REQUEST FOR ANNUAL RENEWAL): LONG TERM SURVEILLANCE OF ROTAVIRUS AND OTHER ENTERIC VIRUSES ASSOCIATED SEVERE DIARRHEA IN PEDIATRIC ADMISSIONS TO KILIFI COUNTY HOSPITAL**

Thank you for continuing review report for the period **24th September 2019 to 24th September 2020.**

This is to inform you that the Expedited Review Team of the KEMRI Scientific and Ethics Review Unit (SERU) was of the informed opinion that the progress made during the reported period is satisfactory. The study has therefore been granted **approval** for continuation.

This approval is valid from **November 08, 2020** through to **November 07, 2021.** Please note that authorization to conduct this study will automatically expire on **November 07, 2021.** If you plan to continue with data collection or analysis beyond this date please submit an application for continuing approval to the SERU by **September 26, 2021.**

You are required to submit any amendments to this protocol and any other information pertinent to human participation in this study to the SERU for review prior to initiation. You may continue with your study.

Yours faithfully,

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

In Search of Better Health

## Appendix III: Published Manuscript

VIRUS EVOLUTION

*Virus Evolution*, 2023, **9(1)**, 1–10
DOI: https://doi.org/10.1093/ve/vead025
Advance access publication date: 15 April 2023
**Research Article**

# Genomic epidemiology of the rotavirus G2P[4] strains in coastal Kenya pre- and post-rotavirus vaccine introduction, 2012–8

Timothy O. Makori,[1,2,†,*] Joel L. Bargul,[2,3,‡] Arnold W. Lambisia,[1,§] Mike J. Mwanga,[1] Nickson Murunga,[1] Zaydah R. de Laurent,[1] Clement S. Lewa,[1] Martin Mutunga,[1] Paul Kellam,[4,5] Matthew Cotten,[6,7,**] D. James Nokes,[1,8,††] My Phan,[6,7,‡‡] and Charles N. Agoti[1,9,§§]

[1]Epidemiology and Demography Department Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Off Hospital Road, P.O BOX 230-80108, Kilifi, Kenya, [2]Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Kalimoni, PO Box 62000-00200, Juja, Kenya, [3]International Centre of Insect Physiology and Ecology, Animal Health Theme, ICIPE Road Kasarani, P.O BOX 30772-00100, Nairobi, Kenya, [4]Department of Infectious Diseases, Faculty of Medicine, Imperial College London, Exhibition Road, London SW7 2AZ, UK, [5]Kymab Ltd, The Bennet Building (B930), Babraham Research Campus, Cambridge CB22 3AT, UK, [6]Medical Research Centre (MRC)/Uganda Virus Research Institute, Plot No: 51-59 Nakiwogo Road, P.O.Box 49, Entebbe, Uganda, [7]MRC-University of Glasgow, Centre for Virus Research Glasgow, 464 Bearsden Road, Glasgow G61 1QH UK, [8]School of Life Sciences and Zeeman Institute (SBIDER), The University of Warwick, Gibbet Hill Campus, Coventry CV4 7AL, UK and [9]School of Health and Human Sciences, Pwani University, Kilifi-Malindi Road, P.O BOX 195-80108, Kilifi, Kenya
[†]https://orcid.org/0000-0003-4482-9729
[‡]https://orcid.org/0000-0001-8573-6807
[§]https://orcid.org/0000-0001-5312-0960
[**]https://orcid.org/0000-0002-3361-3351
[††]https://orcid.org/0000-0001-5426-1984
[‡‡]https://orcid.org/0000-0002-6905-8513
[§§]https://orcid.org/0000-0002-2160-567X
*Corresponding author: E-mail: tmakori@kemri-wellcome.org

## Abstract

The introduction of rotavirus vaccines into the national immunization programme in many countries has led to a decline in childhood diarrhoea disease burden. Coincidentally, the incidence of some rotavirus group A (RVA) genotypes has increased, which may result from non-vaccine-type replacement. Here, we investigate the evolutionary genomics of rotavirus G2P[4] which has shown an increase in countries that introduced the monovalent Rotarix® vaccine. We examined sixty-three RVA G2P[4] strains sampled from children (aged below 13 years) admitted to Kilifi County Hospital, coastal Kenya, pre- (2012 to June 2014) and post-(July 2014 to 2018) rotavirus vaccine introduction. All the sixty-three genome sequences showed a typical DS-1-like genome constellation (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2). Pre-vaccine G2 sequences predominantly classified as sub-lineage IVa-3 and co-circulated with low numbers of sub-lineage IVa-1 strains, whereas post-vaccine G2 sequences mainly classified into sub-lineage IVa-3. In addition, in the pre-vaccine period, P[4] sub-lineage IVa strains co-circulated with low numbers of P[4] lineage II strains, but P[4] sub-lineage IVa strains predominated in the post-vaccine period. On the global phylogeny, the Kenyan pre- and post-vaccine G2P[4] strains clustered separately, suggesting that different virus populations circulated in the two periods. However, the strains from both periods exhibited conserved amino acid changes in the known antigenic epitopes, suggesting that replacement of the predominant G2P[4] cluster was unlikely a result of immune escape. Our findings demonstrate that the pre- and post-vaccine G2P[4] strains circulating in Kilifi, coastal Kenya, differed genetically but likely were antigenically similar. This information informs the discussion on the consequences of rotavirus vaccination on rotavirus diversity.

**Keywords:** rotavirus group A; G2P[4]; coastal Kenya.