MOLECULAR CHARACTERIZATION OF ROTAVIRUS STRAINS FROM PRE- AND POST-VACCINE INTRODUCTION IN SOUTH AFRICA

By

Peter Nthiga Mwangi

2017219839

A thesis submitted in fulfilment of the requirements in respect of the Doctoral Degree in Medical Virology in the Division of Virology in the Faculty of Health Sciences at the University of the Free State

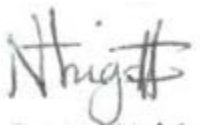
Promoter: Dr. Martin Nyaga
Next Generation Sequencing Unit, Office of the Dean: Health Sciences
and Division of Virology, Faculty of Health Sciences
University of the Free State
NyagaMM@ufs.ac.za

Co-promoter: Prof. Felicity Burt
Division of Virology
Faculty of Health Sciences
University of the Free State
BurtFJ@ufs.ac.za

25th November 2020

DECLARATION

I, Peter Nthiga Mwangi, declare that the thesis that I hereby submit for the Doctoral Degree in Medical Virology at the University of the Free State, is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education



Peter Nthiga Mwangi

25th November 2020

DEDICATION

This thesis is dedicated to every soul that has inspired me along the way.

ACKNOWLEDGEMENTS

My heartfelt appreciation goes to Dr. Martin Nyaga (Promoter), who has expertly guided me throughout this exciting challenge. Your readiness to assist (even during your busiest of schedules) and prompt feedback during this study has been truly encouraging. Your adeptness in research, extraordinary work-ethic, unwavering enthusiasm and overall prolific mentorship has greatly inspired me.

I cherish the outstanding guidance and invaluable insights of Prof. Felicity Burt (co-Promoter) during the entire period of this work. Your precious effort to ensure quality research work is highly appreciated.

I extend my appreciation to the African Enteric Viruses Genome Initiative (AEVGI) co-Principal Investigators (Dr. Francis Dennis, Dr. Jere Khuzwayo and Dr. Valantine Ndze) who together with the Principal Investigator, Dr. Martin Nyaga conceptualized the main project. I appreciate your valuable feedback in the course of reviewing the manuscripts.

Also, am greatly indebted to the fantastic team at the University of the Free State-Next Generation Sequencing (UFS-NGS) Unit for their tremendous technical and moral support: Mr. Milton Mogotsi, Mr. Emmanuel Ayodeji, Mr. Gilmore Pambuka, Mr. Teboho Mooko, Ms. Sebotsana Rasebotsa, Ms. Lesedi Mosime and Ms. Wairimu Maringa. You were all amazing in so many memorable ways. I wish you all the best.

Special gratitude goes to the team at the Diarrhoeal Pathogens Research Unit (DPRU), World Health Organization-Regional Rotavirus Reference Laboratory (WHO-RRL) based in Pretoria: I greatly thank Prof. Mapaseka Seheri for facilitating sample collection. Also, Mrs. Ina Peenze, Dr. Khutso Mothapo, Dr. Fortunate Mokoena, Ms. Kebareng Rakau, Ms. Nonkululeko Magagula, Ms. Grace Thenjiwe and Ms. Dorah Makinita for the assistance during the rotavirus extraction process.

The assistance of Prof. Nicola Page in facilitating additional samples for this work from the Center of Enteric Diseases, National Institute for Communicable Diseases (NICD), Johannesburg, is greatly acknowledged.

I am sincerely grateful to Dr. Mathew Esona, Dr. Celeste Donato and Dr. Benjamin Kumwenda, who were instrumental in data analysis. Also, Mr. Stephanus Riekert for facilitating ICT support.

The financial assistance of the Bill and Melinda Gates Foundation (BMGF) (the principal funder of the AEVGI, through a supervisor's grant to the Principal Investigator, Dr. Martin Nyaga, who awarded me a full scholarship), the Poliomyelitis Research Foundation (PRF) and the University of the Free State Postgraduate bursary, is duly acknowledged. The views emanating from this study are primarily of the author and not principally of the sponsors.

I owe my deep sense of reverence to my mother (Melita Muthoni), brother (James Murimi), relatives, and friends for their faith and encouragement during this endeavour. Special gratitude to Morris Mugo for being such a wonderful friend.

Above all, I mostly thank God for the strength, good health, and sound mind to accomplish this project. Lastly, mentioned or not, may God bless you all mightily, for your contribution to making this work a success.

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS/MANUSCRIPTS

1. **Mwangi, P. N.**, Mogotsi, M. T., Rasebotsa, S. P., Seheri, M. L., Mphahlele, M. J., Ndze, V. N., Dennis, F. E., Jere, K. C. and Nyaga, M.M. Uncovering the first atypical DS-1-like G1P [8] rotavirus strains that circulated during pre-rotavirus vaccine Introduction era in South Africa. *Pathogens*, *9*(5), p.391. **Impact Factor 3.405**
   https://pubmed.ncbi.nlm.nih.gov/32443835/

2. **Mwangi, P. N**., Mogotsi, M. T., Seheri, M. L., Mphahlele, M. J., Peenze, I., Esona, M. D., Kumwenda, B., Steele, A. D., Kirkwood, C. D., Ndze, V. N., Dennis, F. E. and Nyaga M. M. Whole genome in-silico analysis of South African G1P[8] rotavirus strains before and after vaccine introduction over a period of 14 years. *Vaccines, 8(4),* p. 609. **Impact Factor 4.086**
   https://pubmed.ncbi.nlm.nih.gov/33066615/

3. **Mwangi, P. N**., Mogotsi, M. T., Seheri, M.L., Mphahlele, M .J., Peenze, I., Esona, M. D., Kumwenda, B., Steele, A. D., Kirkwood, C. D., Ndze, V. N., Dennis, F. E. and Nyaga M. M. Whole genome analyses of South African G2P[4] strains pre-and post-rotavirus vaccine introduction.

OTHER PUBLICATIONS DURING THE STUDY PERIOD

1. Rasebotsa, S. P., **Mwangi, P. N**., Mogotsi, M.T., Sabiu, S., Magagula, N. B., Rakau, K., Uwimana, J., Mutesa, L., Muganga, N., Murenzi, D., Tuyisenge, L., and Nyaga, M.M. Whole genome and in-silico analyses of G1P [8] rotavirus strains from pre-and post-vaccination periods in Rwanda. *Scientific Reports* 2020, *10*(1), pp.1-22. **Impact Factor 4.120**
   https://www.nature.com/articles/s41598-020-69973-1

2. Maringa, W. M., **Mwangi, P. N**., Simwaka, J., Mpabalwani, E. M., Mwenda, J. M., Peenze, I., Esona, M. D., Mphahlele, M. J., Seheri, M. L. and Nyaga, M. M. Molecular characterisation of a rare Reassortant porcine-like G5P[6] rotavirus strain detected in an unvaccinated child in Kasama, Zambia. *Pathogens* **2020**, *9*, 663.**Impact Factor 3.405**
   https://pubmed.ncbi.nlm.nih.gov/32824526/

3. Mogotsi, M. T., **Mwangi, P. N**., Bester, P. A., Mphahlele, M. J., Seheri, M. L., O' Neill, H. G., and Nyaga, M. M. Metagenomic analysis of the enteric RNA virome of infants from the Oukasie Clinic, North West Province, South Africa, Reveals Diverse Eukaryotic Viruses. *Viruses* 2020, *12*, 1260.
   https://pubmed.ncbi.nlm.nih.gov/33167516/

CONFERENCE BASED DISSEMINATION DURING THE COURSE OF THE STUDY

(a) Virology Africa 2020 conference, Radisson Blue, Waterfront, Cape Town, South Africa, 10th February-14th February, 2020 (International conference)

Mwangi, P. N., Mogotsi, M. T., Rasebotsa, S. P., Seheri, M. L, Mphahlele, M. J, Ndze, V. N., Dennis, F. E; Jere K. C. and Nyaga, M. M. Uncovering the first atypical DS-1-like G1P[8] strains in South Africa, pre-rotavirus vaccine introduction. (Oral presentation)
www.virologyafrica2020.com/files/viral%20conference%20booklet%20program.pdf

Maringa, W. M., **Mwangi, P. N.**, Mogotsi, M. T., Rasebotsa, S. P., Simwaka, J., Magagula, N. B., Rakau, K., Seheri, M. L., Mphahlele, M. J., Mwenda, J. M., and Nyaga, M. M. Whole genome sequencing identifies idiosyncratic changes post-rotavirus vaccine introduction in Zambia. (Oral presentation)
www.virologyafrica2020.com/files/viral%20conference%20booklet%20program.pdf

(b) 12th African Rotavirus Symposium, Emperor's Palace, Johannesburg, South Africa, 30th July-1st August, 2019 (International conference)

Mwangi P. N., Mogotsi, M. T., Rasebotsa, S. P., Sabiu, S., Simwaka, J., Monze, M., Mpabalwani, E. M., Matapo, B., Magagula, N. B., Rakau, K., Seheri, M. L ., Mphahlele, M. J., Mwenda, J. M., and Nyaga, M. M. Molecular characterization of rotavirus strains using whole genome sequencing reveals unique changes post-rotavirus vaccine introduction in Zambia. (Oral presentation)
http://ars.samrc.ac.za/Programme.pdf

Rasebotsa, S. P., **Mwangi, P. N.**, Mogotsi, M. T., Sabiu, S., Mosime, L. P., Magagula, N. B., Rakau, K., Uwimana, J., Mutesa, L., Muganga, N., Murenzi, D., Tusiyenge, L., Mphahlele, M. J., Seheri, M. L., Mwenda, J. M and Nyaga, M. M. Analysis of G1P[8] whole genome constellations identified a vaccine-derived strain in Rwanda. (Oral presentation)
http://ars.samrc.ac.za/ARS%202019/ARS%20Orals/Rasebotsa.pdf

Mugweru, J. N., **Mwangi, P. N**., Rasebotsa, S. P., Mogotsi, M. T., Sabiu, S., Uwimana, J., Mutesa, L., Muganga, N., Murenzi, D., Tusiyenge, L., Magagula, N. B., Rakau, K., Seheri, M. L., Mphahlele, M. J., Mwenda, J. M. and Nyaga, M. M. Whole genome constellation of five reassortant rotavirus strains detected during post-rotavirus vaccine introduction in Rwanda. (Poster presentation)
http://ars.samrc.ac.za/ARS%202019/ARS%20posters/Mugweru.pdf

Sabiu, S., **Mwangi, P. N.,** Rasebotsa, S. P., Mogotsi, M. T., Magagula, N. B., Rakau, K., Uwimana, J., Mutesa, L., Muganga, N., Murenzi, D., Tusiyenge, L., Seheri, M. L., Mphahlele, M. J., Mwenda, J. M., and Nyaga, M. M. Variability of the P-types in stable genotype constellations pre- and post-vaccine introduction in Rwanda. (Poster presentation)
http://ars.samrc.ac.za/ARS%202019/ARS%20posters/Saheed.pdf

(c) 13th International Rotavirus Symposium, Houffalize, Belgium, 24th-28th September, 2018 (International conference)

Mogotsi, M. T., **Mwangi, P. N.,** O'Neill, H. G. and Nyaga, M.M. Genomic investigation of faecal RNA virome in children from Oukasie Hospital, North West Province, South Africa (Poster presentation)
https://kuleuvencongres.be/dsRNA2018/documents/13th-international-dsrna-symposium-conftool-pro.pdf

(d) University of the Free State Faculty of Health Sciences 51st Research Forum, Bloemfontein, South Africa, 29th August to 30th August, 2019 (Local conference)

Mwangi, P. N., Mogotsi, M. T., Rasebotsa, S .P. Sabiu, S., Simwaka, J., Monze, M., Mpabalwani, E. M., Matapo, B., Magagula, N. B., Rakau, K., Seheri, M. L., Mphahlele, M.J., Mwenda, J.M., and Nyaga, M. M. Whole genome characterization of Zambian rotavirus strains reveals remarkable changes post-rotavirus vaccine introduction. (Oral presentation)
https://www.ufs.ac.za/docs/default-source/all-documents/(website)-2019-frf-summary-programme.pdf?sfvrsn=9df68621_0

Rasebotsa, S. P., **Mwangi, P. N**., Mogotsi, M. T., Sabiu, S., Mosime, L. P., Magagula N. B., Rakau, K., Uwimana, J., Mutesa, L., Muganga, N., Mutesa, L., Tusiyenge, L., Mphahlele, M. J., Seheri, M. L., Mwenda, J. M and Nyaga M. M. Whole genome analyses of rotavirus G1P[8] strains circulating pre- and post- RotaTeq™ vaccine introduction in Rwanda (Oral presentation)
https://www.ufs.ac.za/docs/default-source/all-documents/(website)-2019-frf-summary-programme.pdf?sfvrsn=9df68621_0

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

aa	Amino acid
AEVGI	African Enteric Viruses Genome Initiative
AGE	Acute gastroenteritis
BLAST	Basic Alignment Search Tool
BMGF	Bill and Melinda Gates Foundation
Ca^{2+}	Calcium ions
CTL	Cytotoxic T lymphocyte
DLP	Double layered particle
dsRNA	Double stranded RNA
EED	Environmental enteric dysfunction
EPI	Expanded programme on Immunization
ER	Endoplasmic reticulum
eIF4G	Eukaryotic translation initiation factor 4G
FDA	Federal Drug Agency
FEL	Fixed Effect Likelihood
FUBAR	Fast Unconstrained Bayesian Approximation
HBGA	Histo blood group antigens
HSREC	Health Sciences Research Ethics Committee
HIT	Histidine triad
IFN-ʎ	Interferon gamma
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IL	Interleukin
$LiCl_2$	Lithium chloride
LRR	Leucine rich receptor
MAVs	Mitochondria antiviral signaling
MEGA	Molecular Evolutionary Genetics Analysis
MEME	Mixed Effects Model of Evolution
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NSP	Non-structural protein
nt	Nucleotide
NTD	N-terminal domain
NTP	Nucleoside triphosphate
OAS	Oligoadenylate synthetase
OPV	Oral polio vaccine
ORF	Open reading frame
ORS	Oral rehydration salts
ORT	Oral rehydration therapy
PABp	Poly A binding protein
Pdb	Protein data bank

PRR	Pattern recognition receptor
RdRp	RNA-dependent RNA polymerase
RIG-I	Retinoic acid-inducible gene I
RMSD	Root Mean Square Deviation
RNaseL	Latent ribonuclease
RVA	Group A rotavirus
RVB	Group B rotavirus
RVC	Group C rotavirus
RVD	Group D rotavirus
RVE	Group E rotavirus
RVG	Group G rotavirus
RVH	Group H rotavirus
RVI	Group I rotavirus
SLAC	Single Likelihood Ancestor Counting
ssRNA	Single-stranded RNA
TLP	Triple layered particle
TNF	Tumor necrosis factor
TLR	Toll-like receptor
TRAF2	Tumor necrosis factor receptor associated factor 2
UFS	University of the Free State
UTR	Untranslated regions
ViPR	Virus Pathogen Resource
VP	Viral protein
WHO	World Health Organization
WHO-RRL	World Health Organization-Regional Reference Laboratory

ABSTRACT

Introduction

Group A rotavirus (RVA) persists as the infectious etiological agent for acute gastroenteritis (AGE) in children under the age of five years globally, despite the introduction of RVA vaccines. A major concern during the post-RVA vaccination period is the potential for vaccine-induced strain changes, which may lead to putative vaccine-escape strains that may affect the efficacy of RVA vaccines in the long-term. There is a paucity of longitudinal whole-genome studies to assess the impact of RVA vaccine introduction on the circulating RVA strains in South Africa.

Aim

The study aimed to elucidate the impact of RVA vaccine introduction on the whole-genome of G1P[8] and G2P[4] RVA strains that circulated in South Africa, seven years before and seven years after RVA vaccine (Rotarix®) introduction.

Materials and Methods

Rotavirus positive stool samples genotyped conventionally as G1P[8] (*n*=103) and G2P[4] (*n*=98) as part of the routine World Health Organization (WHO) RVA surveillance were retrieved from the archival storage of Diarrhoea Pathogens Research Unit (DPRU) based in Pretoria and Center of Enteric Diseases, National Institute for Communicable Diseases (NICD), Johannesburg, South Africa. Rotavirus dsRNA was extracted from the stool samples. cDNA was synthesized from the purified extracted RNA using the Maxima Kit. The DNA libraries were prepared using the Nextera XT Kit and paired-end sequencing for 600 cycles (301 x 2) was performed on a MiSeq Illumina® platform.

Results and Discussion

Firstly, the study investigated the genotype variation and relative emergence of potential vaccine-escape strains. Two atypical DS-1-like G1P[8] strains were identified. The two atypical G1P[8] strains circulated prior to vaccine introduction in South Africa compared to reports in recent studies that detected this strain during the post-vaccination period. Phylogenetic analysis

showed that these atypical G1P[8] strains emerged through reassortment mechanism involving locally circulating South African G1P[8] strains and the DS-1-like backbone of G2P[6] strains.

Secondly, the study investigated the evolutionary dynamics of South African pre-and post-G1P[8] and G2P[4] strains. South African G1 strains clustered in G1 lineage-I and II while majority (84.2%) of P[8] strains grouped in P[8] lineage-III, highlighting prevalent co-circulation of the G1 and P[8] strains in these lineages before and after vaccine introduction. All the South African G2 strains segregated into the predominant G2-lineage IV, while 99.7% of the P[4] strains clustered in P[4]-lineage IV, underscoring the continued prevalence of these lineages over time.

Thirdly, the study investigated the genetic changes between pre and post-vaccine strains, whereby, for G1P[8] strains, we identified amino acid (aa) differences that potentially explained the yearly clustering of post-vaccine strains into sub-lineages. However, we did not observe aa differences that were consistently conserved throughout the post-vaccine period. For G2P[4] strains, we identified aa residues that were consistently prevalent and appeared at a significantly higher frequency in 60-80-% of post-vaccine strains at various sites in the gene segments compared to their negligible proportions of 1-2% during the pre-vaccine period.

Fourthly, we investigated the aa profiles within the mapped antigenic regions of the neutralization proteins. The aa profile in the antigenic profiles of pre- and post-vaccine G1 and P[8] strains was nearly the same except for the N147D substitution observed in eight post-vaccine G1 strains. Similarly, the overall aa profile in G2 and P[4] antigenic domains remained mostly unchanged between pre and post-vaccine strains.

Conclusion

Rotarix® did not appear to impact the antigenic and non-antigenic profile of South African post-vaccine G1P[8] and G2P[4] strains. Furthermore, the atypical DS-1-like strains identified in this study appeared to have emerged through natural RVA evolutionary processes and not necessarily vaccine-induced. Continued long-term whole-genome surveillance remains necessary to assess the impact of RVA vaccination on circulating RVA strains.

Keywords: atypical strains, DS-1-like constellation, evolution, genome constellation, lineages rotavirus strains, Wa-like constellation, whole-genome

CHAPTER ONE: INTRODUCTION

1.1 Background information

Diarrhoeal diseases are among the leading causes of mortality, only after pneumonia, in children below the age of five, globally (WHO, 2020). There are about 1.7 billion childhood diarrhoea cases worldwide every year (WHO, 2020). Hospital and community-based studies before 2006, when two rotavirus vaccines were licensed for use, indicated that RVA was the main cause of AGE, accounting for an estimated half a million mortality cases annually (Parashar *et al.*, 2006). During the pre-vaccine era, estimates showed that by the age of five years, every child would be infected, and one in five RVA positive cases would require an outpatient medical visit, one in 65 would be hospitalized, and one in 293 would die, particularly children from low-income countries (Parashar *et al.*, 2006). Based on 2016 post-RVA vaccine estimates, there was a substantial reduction in mortality to 128,500 deaths caused by RVA annually in the same age group and 82% of these deaths were reported in sub-Saharan Africa (Troeger *et al.*, 2018).

The burden of RVA-induced diarrhoeal disease can be managed and prevented by improved sanitation, treatment with Oral Rehydration Therapy (ORT), and most importantly, by immunization (Crawford *et al*., 2017). While ORT is deemed cost-effective, it is beset with significant setbacks of specialized skills and conditions for its distribution and oral administration instructions, especially in the developing world (Shillcutt *et al*., 2016). Therefore, RVA vaccines are considered the best long-term strategy to prevent RVA diarrhoea (Burke *et al.*, 2019). RVA is regarded as a 'democratic virus' due to its ability to infect hosts across the social-economic divide (WHO, 2013). This ability for the universal occurrence of RVA infections, even in settings with high hygiene standards, underscores the high transmissibility of the virus globally (WHO, 2013). RVA has a high shedding rate in faeces (in the order of up to 10^{11} viral particles/gram), low infectivity dose (as low as one infectious virus particle) and high resistance to frequently used disinfectants like 70% ethanol (Kampf, 2018).

The WHO recommends RVA epidemiological surveillance for countries intending to introduce and those that have implemented the RVA vaccine(s) (WHO, 2013). This is due to the varying geographical prevalence of circulating RVA strains, albeit all the WHO pre-qualified vaccines offering partial heterotypic protection (Dóró *et al.*, 2014). As a result, RVA surveillance networks were established in Africa (Mwenda *et al.*, 2014) and other regions, globally (Iturriza-Gomara and Gray, 2011; Agocs *et al.*, 2014). Four RVA vaccines: Rotarix® (GlaxoSmithKline Biologicals, Rixenstart, Belgium); RotaTeq™ (Merck and Co., West Point, PA. USA); ROTAVAC® (Bharat Biotech, Hyderabad, India) and Rotasiil®(Serum Institute of India, India), have been pre-qualified by WHO for global use after proving safe and highly efficacious in extensive safety and efficacy trials (Ruiz-Palacios *et al*., 2006; Vesikari *et al*., 2006; Bhandari *et al*., 2014a, 2014b, Desai *et a*l., 2018). The vaccines' efficacy is based on their ability to elicit serotype-specific and/or heterotypic immunity against the predominant RVA serotypes and/or genotypes (Kirkwood *et al.*, 2019).

However, there are concerns regarding RVA vaccines. One is that while high efficacy (82-90%) has been demonstrated in developed countries, it compares dismally with a sub-optimal efficacy (38.1-63%) reported in low-income countries (Cunliffe *et al*., 2012; Carvalho and Gill, 2019). Further, several studies have looked to the possibility that RVA vaccination induces emergence/re-emergence of new strains due to genetic changes as a result of vaccine pressure (Magagula *et al*., 2015; Jere *et al*., 2018; Santos *et al*., 2019; Hoa-Tran *et al*., 2020; Mishra *et al*., 2020; Mwangi *et al.,* 2020; Rasebotsa *et al*., 2020). However, there is still uncertainty and thus the capacity for the vaccine candidates to protect against newly emerging or unusual strains remains unknown for now (Velasquez and Jiang, 2019).

Since vaccine-induced selective pressure on RVA strains is subtle to describe within one or two years after vaccine introduction, it is suggested that it may take several years (at least seven years) before vaccine-induced genetic changes are apparent (Patton, 2012). Therefore, whole-genome longitudinal studies are required to comprehensively assess the impact of RVA vaccines on the evolution of RVA. There is a vast diversity of RVA strains in Africa (Seheri *et al*., 2018) and a suboptimal efficacy of RVA vaccines compared to developed countries (Steele *et al*., 2019). Therefore, whole-genome surveillance studies to decipher the impact of RVA vaccines on the diversity of RVA strains in Africa are imperative. In this regard, the African Enteric Viruses

Genome Initiative (AEVGI) funded by Bill and Melinda Gates Foundation (BMGF) is spearheading whole-genome characterisation of RVA strains in four African countries (Cameroon, Ghana, Malawi and South Africa). South Africa was the leading African nation to introduce the Rotarix® vaccine in 2009 (WHO, 2009) and presented an opportunity to evaluate the impact of Rotarix® vaccine on circulating RVA strains. This research focused on the long-term evolutionary dynamics of G1P[8] and G2P[4] (the most prevalent RVA strains circulating globally), seven years before and seven years after the introduction of the Rotarix® vaccine in South Africa.

1.2 Problem statement and rationale of the study

Although RVA vaccines are expected to impart immune pressure against all RVA strains (heterotypic immunity), there is the possibility of the emergence of novel or rare strains that may circulate and spread rapidly across human populations globally, affecting vaccine efficacy (Matthijnssens *et al*., 2009). Surveillance of the RVA molecular epidemiology has been a high priority pre- and post-RVA introduction (Nyaga *et al*., 2014a,b, 2015, 2016; Seheri *et al*., 2018; Simwaka *et al*., 2018; Lartey *et al*., 2018; Mwenda *et al*., 2019; Mwangi *et al*., 2020). Despite some significant decreases in RVA gastroenteritis in the countries that have incorporated RVA vaccines in their national immunization programs, rare/unusual RVA strains have been reported in the post-licensure RVA vaccination period (Fujii *et al*., 2014; Zeller *et al*., 2016; Jere *et al*., 2018; Guerra *et al*., 2019; Luchs *et al*., 2019; Mishra *et al*., 2020).

Based on available data regarding circulating RVA genotypes, it is highly probable to infer that selection pressure by the RVA vaccines is driving the emergence of putative vaccine-escape mutants. Yet, reports documenting comparable fluctuations during the pre-vaccine introduction period imply that in addition to vaccines, natural evolution factors of wild-type RVA might also be involved (Rivera *et al*., 2013, Moyo *et al*., 2014; Jere *et al*., 2018; Seheri *et al*., 2018). As of yet, there is inconclusive evidence to ascertain that RVA vaccination selects the emergence of vaccine-escape mutants. In addition, there is limited data in South Africa describing RVA evolution of G1P[8] and G2P[4] strains pre- and post-RVA vaccination introduction at the whole-genome level. Therefore, it is imperative to understand the potential effect of vaccine-induced

selective pressure on the emergence of mutant vaccine-escape strains and understand the natural evolutionary mechanisms involved in the most circulating RVA genotypes.

In this regard, the proposed study aimed to elucidate molecular dynamics of the most prevalent circulating RVA strains globally (G1P[8] and G2P[4]), collected from children in South Africa over 14 years period (seven years pre- and seven years post-RVA vaccination). Seven or more years of strain surveillance at the whole-genome level are necessary to report on vaccine-escape mutants conclusively (Patton, 2012). Availability of pre- and post-vaccination samples for this study hinged on the RVA strain surveillance that has been ongoing before and after RVA vaccine introduction in the South African national immunization program in 2009. The findings of this study contribute to information that may be important in understanding the evolutionary dynamics that shape the antigenic and epidemiological landscape of RVA strains, particularly in Africa.

1.3 Aims and objectives

1.3.1 Aim

To elucidate the impact of RVA vaccine introduction on the evolution of the whole-genome of circulating RVA strains (seven years pre- and seven years post-rotavirus vaccine introduction) and correlate potential vaccine-mutant strains through pairwise sequence analysis, phylogenetic analysis and computational protein modelling.

1.3.2 Objectives

i. To investigate genotype variation and relative emergence of potential vaccine-escape strains pre- and post-vaccination introduction in South Africa.
ii. To investigate evolutionary dynamics of human rotaviruses pre- and post-vaccination introduction with emphasis on G1P[8] and G2P[4] strains in South Africa.
iii. To identify genetic changes that impact structural and/or non-structural proteins not present prior to vaccine usage.
iv. To investigate the relationship between amino acid (aa) changes within the defined rotavirus antigenic regions.

1.4 Thesis organization

The thesis is structured into seven chapters. The introduction as based on the research proposal, is chapter one. Chapter two is the detailed literature review, while chapter three is the general materials and methods. Chapter's four to six are structured around the study's four objectives, presented in research article format, each with a brief independent background, materials and methods, results, and discussion.

In brief, chapter four mainly addresses the study's first objective, which investigated genotype variation and the emergence of vaccine escape strains pre-and post-vaccination introduction. The second, third, and fourth objectives of the study were addressed overlapping between chapters four and six. These chapters touch on the long-term phylodynamics of South African G1P[8] strains before and after the introduction of the vaccine and the phylodynamics of South African G2P[4] strains before and after implementation of the vaccine. Chapter seven provides a general summary of the results as a discussion, with recommendations and limitations. Chapter seven is then followed by the references and appendixes, respectively.

CHAPTER TWO: LITERATURE REVIEW

2.1 Rotavirus discovery and morphology

Ruth Bishop discovered RVA in 1973 in the cytoplasm of duodenal biopsies of children admitted with gastroenteritis (Bishop *et al*., 1973). The virus closely resembled viruses previously observed to cause diarrhoea in calves (Edwards and Sier, 1960). The virus derives its name from Latin word *rota,* which means wheel due to its characteristic wheel-like structure (Flewett *et al*., 1974).

2.2 Rotavirus gastroenteritis and burden of disease

After the discovery of RVA, it was determined as the primary etiological agent of AGE in infants resulting in 35-50% of all hospitalisations during the first two years of life (Estes and Kapikian 2007). RVA is primarily transmitted through the faecal-oral route, although transmission might still occur via respiratory droplets (Schellack *et al*., 2017). The infection by RVA is associated with symptoms such as: fever, mild short-lived watery diarrhoea to frequent profuse diarrhoea, and vomiting, which can lead to shock dehydration, electrolyte imbalance and dehydrating gastroenteritis (Crawford *et al*., 2017). Although the RVA disease incidence is similar in children in low- and high-income countries, poor access to health care and comorbidities such as malnutrition in low-income countries exacerbates these symptoms leading to higher mortality rates (Markkula *et al*., 2017).

The global RVA mortality rate for children <5 years of age based on 2016 estimates was 20.3/100,000 population with that of sub-Saharan Africa significantly higher at 66.9/100,000 population (Troeger *et al*., 2018) (Figure 2.1). In South Africa, the hospitalization estimates of RVA disease in children under the age of 2 years before the introduction of the RVA vaccine were approximately 30,000 annually (Mapaseka *et al*., 2010). The RVA infections occurred all year round peaking during cooler winter months and detection with at least one additional enteropathogen was common (Steele *et al*., 2003; Mapaseka *et al*., 2010). The introduction of RVA vaccination in South Africa had a substantial impact in reducing these RVA hospitalisations by approximately 60%, as indicated by laboratory-confirmed results (Msimang *et al*., 2013). However, despite the significant reduction in RVA mortality rate due to the impact of the RVA

vaccine(s), RVA-induced diarrhoea persists as the leading infectious etiological cause of diarrhoeal mortality in under 5 year old children (Troeger *et al.*, 2018).

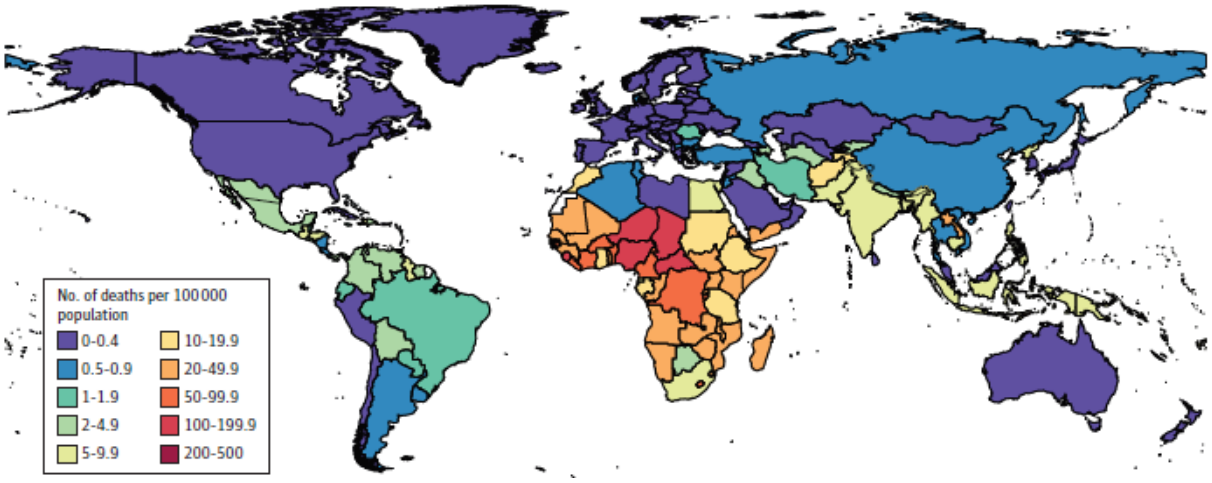


Figure 2.1: Geographical distribution of rotavirus mortality rates
The mortality rates are based on 2016 global estimates. The color codes indicate the severity of the RVA disease burden. Red color indicate intensity of severity. The sub-Saharan Africa exhibited approximately 80% of the global RVA mortality rates with a rate of 66.9/100,000 population cases.
Adapted from Troeger *et al.*, 2018 with permission granted from the publisher under the terms of the Creative Commons non-commercial license for scholarly articles (Appendix 1).

2.3 Structure and taxonomy of rotavirus

The genus rotavirus is classified within the *Reoviridae* family and is a dsRNA genome constituting of 11 gene segments (Estes and Kapikian, 2007). Rotavirus genome is encapsulated in a triple-layered non-enveloped protein capsid consisting of six structural proteins (viral proteins 1, 2, 3, 4, 6 and 7) (Estes and Greenberg, 2013) (Figure 2.2). The VP4 protein conspicuously protrudes from the smooth outer surface in 60 protein spike-like projections (Estes and Greenberg, 2013) (Figure 2.2). The non-structural proteins (NSP) include NSP1, NSP2, NSP3, NSP4 and NSP5/6 (Estes and Greenberg, 2013). The virus is icosahedral (Pesavento *et al*., 2006) and image reconstruction studies by Jayaram *et al.* (2004), suggest that the genome segments form conical cylinders around replication complexes.

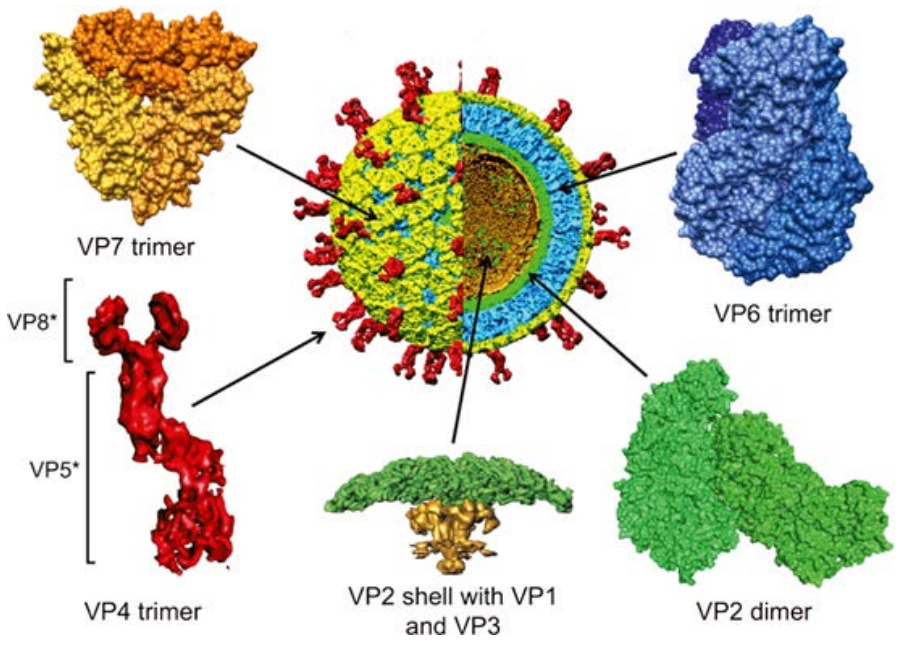


Figure 2.2:Features of rotavirus structure and location of protein structures
The figure shows the structural proteins of rotavirus (VP1, VP2, VP3, VP4, VP6 and VP7). The protein structures are showed in different colors. The VP1, VP2 and VP3 in complex together are indicated. The VP2 dimer is shown in green. The VP4 spikes (with the VP5*and VP8* cleaved subunits) are coloured in red. The VP6 structure is indicated in blue while the VP7 trimer is shown in yellow. Image used with permission from BVV Prasad (Appendix 2).

2.4 Rotavirus classification

The VP6 determines rotavirus classification into nine groups (A-I) (Matthijnssens *et al.,* 2011, Marthaler *et al*., 2014; Mihalov-Kovacs *et al*., 2015). Groups A-C primarily infects humans with RVA being of most medical importance, as it accounts for most of RVA human infections (Matthijnssens *et al.,* 2008). Group B (RVB) infects swine, cattle, sheep and rats and infects both adults and children, causing outbreaks and sporadic infections (Marthaler *et al*., 2014). Group C (RVC) infects swine and dogs and can cause epidemics in the human population while Groups D-G mostly infect avian species (Matthijnssens *et al*., 2011). Group H (RVH) has been identified in swine (Marthaler *et al*., 2014) while group I (RVI) has been described in dogs (Mihalov-Kovacs *et al*., 2015).

The outer capsid proteins, VP4 (denoted P due to its protease-sensitive property) and VP7 (designated G for glycoprotein), elicit neutralizing antibodies during host infection and are utilized in dual classification into P and G types, respectively (Estes and Kapikian, 2007). To fully

describe the rotavirus strains, the dual classification system was expanded to include the other nine gene segments (Matthijnssens *et al.,* 2008). In the whole-genome classification system, each gene segment is designated a distinct genotype underpinned by established nucleotide percent cut-off values (Matthijnssens *et al*., 2008). The RVA genome is described as follows: Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx (where x represents the specific genotype), while the alphabets describe the property of the 11 genes of RV strains (Matthijnssens *et al*., 2011). Most of the human RVA strains exhibit the Wa-like genotype backbone constellation (I1-R1-C1-M1-A1-T1-E1-HI) and the DS-1-like genotype backbone constellation (I2-R2-C2-M2-A2-N2-T2-E2-H2). There is also a relatively minor genotype constellation AU-1-like, (I3-R3-C3-M3-A3-N3-T3-E3-H3) (Matthijnssens and Van Ranst, 2012). GIP[8], G3P[8], G4P[8], G9P[8] and G12P[8] RVA are usually associated with the Wa-like genotype constellation, whereas G2P[4], G8P[4] or G8P[6] typically have a DS-1-like backbone constellation (Matthijnssens and Van Ranst, 2012). The G3P[9] genotype has the characteristic AU-1-like genotype constellation (Tsugawa *et al*., 2015). Genotypes identified in human and non-human hosts to date are: 36 G, 51 P, 26 I, 22 R, 20 C, 20 M, 31 A, 22 N, 22 T, 27 E and 22 H (https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification).

2.5 Rotavirus molecular epidemiology

The medical importance of six predomiant RVA strains: G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8] and particularly, the global prevalence of G1P[8] is well documented (Doro *et al*., 2014). Certain strains have regional or local importance, for example, the genotype P[6] strains are markedly reported in Africa underlying their sustained circulation (Nyaga *et al*., 2018). Other strains like G2P[4] are associated with cyclic peaks (Linhares and Justino, 2014). In South Africa, there have been outbreaks of G2P[4] infections every three to four years and significant epidemics every ten years (Page and Steele, 2004). The epidemiological relevance of this genotypical cycling is not well understood. The hypothesis among researchers is that cycling may be a strategy for RVA to evade group immunity acquired from previous infections enabling the RVA to prevail in circulation (Parra, 2009).

The segmented RVA genome facilitates genetic reassortment and the inherent error-prone RNA-dependent polymerase (RdRp) leads to the occurrence of point mutations contributing to the vast diversity of RVA strains (Estes and Greenberg, 2013). It is challenging to predict which RVA strain may become epidemiologically important (Kirkwood, 2010). G9P[8] and G12P[8] are examples of RVA strains that circulated globally between the periods of the 1990s and 2000s (Doro *et al*., 2014). Therefore, RVA surveillance studies are crucial in monitoring RVA strains that can become epidemiologically fit and spread in the human population. RVA surveillance studies conducted for over three decades in South Africa indicate that at any given season, there is a vast diversity of RVA strains in circulation (Steele *et al*., 1995; Potgieter *et al*., 2010; Page *et al*., 2016; Page *et al*., 2018). Most of the RVA surveillance studies have focused on characterizing the outer capsid proteins VP4 and VP7. However, since RVA evolutionary processes can occur in other gene segments (Maringa *et al*., 2020; Mwangi *et al*., 2020), whole-genome based surveillance studies capacitated by next-generation sequencing (NGS) technology provide the potential for the comprehensive assessment of the RVA evolutionary dynamics (Nyaga *et al*., 2020). In particular, providing information such as: the origin of strains, reassortment events, zoonosis and genetic linkage of RVA gene segments as well as the potential to fully decipher the impact of RVA vaccines on the genetic and antigenic landscape of circulating RVA (Ghosh and Kobayashi, 2011).

2.6 Rotavirus genome

The 11 gene segments of RVA vary in size from 667 base pairs (bps) (NSP5) to 3302bps (VP1), with the total genome being approximately 18,522 bps (Estes and Greenberg, 2013). The plus-sense strand of the genomic dsRNA has a capped 5’ sequence. Each RNA segment features a 5’ guanidine, the 5’ part of conserved noncoding sequences, an open reading frame (ORF), and a set of conserved noncoding sequences of the 3’ part (Estes and Greenberg, 2013). Each of the ten segments is monocistronic, coding for one of the six structural and five non-structural proteins. Segment 11 encodes two proteins (bicistronic), namely NSP5 and NSP6 (Estes and Greenberg, 2013).

Table 2.1: Rotavirus genome segments sizes, proteins encoded and genotypes

Genome segment	Protein encoded	Percent identity[a]	Number of genotypes[b]	Genotype(acronym underlined)	Size (bp)	ORF (nucleotide number)	Size of protein encoded	Number of molecules per virion
1	VP1	83	22R	**R**dRp	3302	18-3282	1088 aa (125.0 kDa)	12
2	VP2	84	20C	**C**ore protein	2690	17-2659	881 aa (102.4 kDa)	120
3	VP3	81	20M	**M**ethyl transferase	2591	50-2554	835 aa (98.1 kDa)	12
4	VP4	80	51P	**P**rotease-sensitive	2362	10-2337	776 aa (86.7 kDa)	120
5	NSP1	79	31A	Interferon **A**ntagonist	1611	31-1515	495 aa (58.6 kDa)	NA
6	VP6	85	26I	**I**ntermediate	1356	24-1214	397 aa (48.1 kDa)	780
7	NSP3	85	22T	**T**ranslation **E**nhancer	1074	26-970	315 aa (34.6 kDa)	NA
8	NSP2	85	22N	**N**TPase	1059	47-997	317 aa(36.7 kDa)	NA
9	VP7	80	36G	**G**lycosylated	1062	49-1026	326 aa (37.3 kDa)	780
10	NSP4	85	27E	**E**nterotoxin	751	41-569	175 aa (20.3 kDa)	NA
11	NSP5/6	91	22H	P**H**osphoprotein	667/821	22-615/80-355	198 aa (21.7 kDa)/ 92 aa (11.0 kDa)	NA

The gene segments are numbered 1-11 based on their migration pattern on SDS-PAGE gel. [a] Nucleotide percent identity cut-off value defining genotypes. [b] The number of currently available genotypes. Acronym in bold and underlined indicate the letter utilized to describe the functional property of the genotype. NA refers to not applicable.
Modified from: http://www.reoviridae.org/dsRNA_virus_proteins/Rotavirus.htm

2.6.1 The rotavirus structural proteins

2.6.1.1 Viral protein 1 (VP1)

Viral protein 1 serves as an RNA dependent RNA polymerase (RdRp), designated 'R' in the whole genome classification (Matthijnssens *et al*., 2008). The VP1 is a compact, globular protein, with three distinct domains: (a) N-terminal domain (aa1-332) (b) a polymerase domain consisting of fingers, palm and thumb subdomains (aa333-778) and (c) a C-terminal "bracelet" domain (aa 779-1089) (Tao *et al*., 2002). The catalytic centre of the polymerase has a higher degree of sequence conservation than the N-terminal and C-terminal domains and active asparate residues D631 and D632 are implicated in catalysis of RNA synthesis (McDonald *et al*., 2009). Despite having all the necessary structural features for catalysis, VP1 is inactive as an apoenzyme and is dependent upon its interactions with VP2 and the VP2 contact site on VP1 occurs at high conserved residues 267 to 270 (Steger *et al*., 2019).

2.6.1.2 Viral protein 2 (VP2)

Viral protein 2 is located in the inner core of the virion and is denoted as 'C' for core in the whole-genome classification (Matthijnssens *et al*., 2008). The VP2 has 120 molecules with 60 dimers arranged in a T=1 symmetry (McClain *et al*., 2010). The VP2 has a replicase role that is involved in the formation of replication intermediates (Mansell *et al*., 1990). According to Buttafuoco *et al*. (2020), mutations occurring in residues L124, V865, and I878 are implicated in non-structural protein viroplasm formation and are conserved among RVA to H species.

2.6.1.3 Viral protein 3 (VP3)

Viral protein 3 is designated 'M' due to its methyltransferase property in the whole-genome classification (Matthijnssens *et al*., 2008). In addition to its capping activities, RVA VP3 has been shown to cleave 2',5'-oligoadenylates signalling molecules produced by the cytoplasmic dsRNA sensor oligoadenylate synthetase (OAS) which activate the latent ribonuclease (RNaseL) to cleave single-stranded viral and cellular RNAs (Silverman and Weiss, 2014). The VP3 is predicted to contain five structurally distinct domains: N-terminal domain (NTD) of unknown function, a guanine-N7-methyltransferase with an inserted 2'-O-methyltransferase, a combined RNA 5'-triphosphatase and a C-terminal 2'5'-phosophodiesterase (Ogden *et al*., 2014). The VP3 mediates

the suppression of Interferon-gamma (IFN-ʎ) production through phosphorylation of a mitochondria antiviral signalling (MAVS) proline-rich region (Ding *et al*., 2018). In addition, VP3 exhibits helicase activity by unwinding the RNA duplex formed transiently during endogenous transcription (Kumar *et al*., 2020).

2.6.1.4 Viral protein 4 (VP4)

Viral protein 4 is designated 'P' due to its protease sensitivity property (Matthijnssens *et al*., 2008). The VP4 is the key cell attachment and viral entry agent and contains 60 surface projections that are 10-12nm long and enables the attachment of the virion onto the host cell (Ludert *et al*., 1996). In infected cells, VP4 is expressed in different forms: classically as spike on the outer layer of virus particles and as free soluble protein in the cytosol (Condemine *et al*., 2019). Trypsin cleavage of VP4 alters its conformation such that the VP4 spikes project with two-fold symmetry exhibiting a trimeric base (Pesavento *et al*., 2006). The excision occurs on residue aa232-247 and cleavage rearrangement results in two fragments: VP5* and VP8*(Estes and Greenberg, 2013). The cleavage facilitates virus entry into the host cell (Kaljot *et al.,* 1988). Attachment of RVA on the host cell is regarded as either dependent on or independent of sialic acid presence, such as, N-acetylneuraminic acid in cell surface glyconjugates (Isa *et al*., 2006). This relationship has offered the platform to classify RVA into sialidase-sensitive and sialidase-insensitive strains (Graham *et al*., 2003).

The VP8* domain, a lectin-like globular protein, is situated at the spike (Ciaret and Estes, 2002) and is important in binding sialic acid (Fiore *et al*., 1991). Based on the VP8* sequences, the diverse P genotypes have been grouped into five genogroups PI-PV (Huang *et al*., 2012). The P[4], P[6], P[8], and P[19] genogroups are grouped into PII genogroup, among which P[8] RVs are the most widely circulating human RVA (Sun *et al*., 2018). The VP5* domain characterizes the other parts of the spike and is suggested to be the membrane active component (Dormitzer *et al*., 2004). The VP8* and VP5* regions contain epitopes that have been mapped as neutralization escape mutation sites (Dormitzer *et al*., 2002; Dormitzer *et al*., 2004; Monnier *et al*., 2006). Three potential VP4 sites (385, 393 and 471) have been reported to play a role in RVA attenuation and can be targeted by site-specific mutagenesis to develop attenuated vaccines (Guo *et al*., 2020).

2.6.1.5 Viral protein 6 (VP6)

Viral protein 6 constitutes the intermediate capsid protein, thus designated 'I' in the whole-genome classification (Matthijnssens *et al*., 2008). The VP6 is a trimer stabilized by non-covalent bonds (Prasad *et al*., 1988) and exhibits T=13 icosahedral symmetry with 260 trimeric arrangement (Mathieu *et al*., 2001). It is the most immunogenic and abundant RVA structural protein (Estes and Greenberg, 2013). The aa region 376-384 has been identified as a cytotoxic T lymphocyte (CTL) epitope (Franco *et al*., 1994). Additionally, a 14 aa region 289-302 of a VP6 peptide fragment containing a putative CD4+ T cell epitope was reported to confer nearly complete protection against RVA in an adult mouse model (Choi *et al*., 2000). Antibodies targeting VP6 have been reported to interfere with RVA replication cycle, making VP6 a potentially viable vaccine candidate (Corthesy *et al.,* 2006). The VP6-specific antibodies, notably Immunoglobulin A (IgA), serve as markers to indicate protection after RVA infection or vaccination (Franco *et al*., 2006).

2.6.1.6 Viral protein 7 (VP7)

The VP7 possesses an endoplasmic reticulum (ER)-synthesized glycoprotein hence denoted 'G' due to this property (Matthijnssens *et al*., 2008). It is a trimer stabilized by Ca^{2+} (Dormitzer *et al*., 2000). The VP7 crystal structure of rhesus rotavirus, serotype G3 reveals that the core of the trimer contains subunits that fold into two compact domains of disordered N and C terminal arms and that calcium ions (Ca^{2+}) are bound at each subunit interface of the trimer (Aoki *et al*., 2009). The virion entry into the host cell relies on calcium-dependent uncoating of VP7 (Dormitzer *et al*., 2000). Uncoating of the VP7 from the virion is facilitated by the loss of free Ca^{2+} enabling its dissociation into monomers therefore inducing conformational changes to the VP4 protein (Chen *et al*., 2009).

The VP7 is a neutralization antigen targeted by neutralization antibodies (Ciarlet and Estes, 2002). The immunogenic and structural properties of VP7 together with VP4 have greatly dominated focus in vaccine development (Hoshino and Kapikian, 1994). Structurally, the VP7 gene contains two epitope regions identified as 7-1 and 7-2 (Aoki *et al*., 2009). Four regions (aa87-101, aa145-152, aa208-221 and aa235-242) have also been described as cross-reactive and serotype-specific

neutralization epitopes (Taniguchi *et al*., 1988; Coulson and Kirkwood, 1991). There are also two CTL epitopes at aa positions 16-28 and 40-52 (Morozova *et al*., 2015).

2.6.2 The rotavirus non-structural proteins

2.6.2.1 Non-structural protein 1 (NSP1)

Non-structural protein 1 is denoted 'A' in the whole-genome classification due to its interferon antagonist property (Matthijnssens *et al*., 2008). The N terminal sequence of NSP1 has cysteine residues which are highly conserved (Hua *et al*., 1994). A conserved zinc-binding domain that spans residues 42-72 is well-defined (Morelli *et al*., 2015). The NSP1 suppresses innate immune response by inhibiting expression of type 1 interferon (Barro and Patton, 2005; Arnold *et al.*, 2013). The NSP1 has also been reported to degrade other host proteins such as retinoic acid-inducible gene-I (RIG-I), tumor necrosis factor receptor-associated factor 2 (TRAF2), mitochondrial antiviral signalling protein (MAVS) and poly (A) specific RNase subunit (Pan 3) (Morelli *et al*., 2015).

2.6.2.2 Non-structural protein 2 (NSP2)

Non-structural protein 2 is an octameric protein possessing Mg^{2+} dependent NTPase activities (Schuck *et al*., 2001) hence designated 'N' in the whole-genome classification due to this property (Matthijnssens *et al*., 2008). The NSP2 monomer has two structural domains. The C terminal domain exhibits antiparallel β-sheet flanked by α helices that display a HIT (histidine triad) like a motif, which is proposed to be the binding site of nucleoside triphosphate (NTP). The N terminal domain is α helical and within it, a 24 residue highly basic loop lines four prominent grooves, and it is proposed that these grooves are binding sites for ssRNA (Jayaram *et al*., 2002). Helix destabilising activity and capacity to recruit ssRNA templates for genomic dsRNA synthesis through the NTPase activity has led NSP2 to be referred to as a molecular motor that facilitates the packaging of template mRNAs into pre-capsid structures of replication intermediates (Borodavka *et al*., 2017). Antibody interactions at antibody-binding epitope region (aa244-252) are suggested to block several rotavirus functions including viroplasm formation and packaging (Donker *et al*., 2011a).

2.6.2.3 Non-structural protein 3 (NSP3)

Non-structural protein 3 serves as a translation enhancer and is designated 'T' in the whole-genome classification (Matthijnssens *et al*., 2008). It is a homodimer with RNA binding properties (Mattion *et al*., 1992) and with two distinct domains: The N terminal domain is bi-lobed, asymmetric homodimer that binds 3'GACC sequence at the 3' end of rotaviral mRNAs (Deo *et al*., 2002). The C terminal domain competes with poly (A) binding protein (PABP) for the same segment of eukaryotic translation initiation factor 4G (eIF4G), a ribosome recruiting factor (Piron *et al*., 1998). The PABP enhances translation initiation by stimulating mRNA circularization (Kuhn and Wahle, 2004). Contreras-Treviño *et al*. (2017) identified aa sites 170, 171,173 and 187 as critical in stabilization of NSP3 dimers and influencing the yield of NSP3.

2.6.2.4 Non-structural protein 4 (NSP4)

Non-structural protein 4 serves as an enterotoxin (the first described virus-encoded enterotoxin) and due to this property is designated 'E' in the whole-genome classification (Matthijnssens *et al*., 2008). It is an ER-resident glycoprotein whose transmembrane domain (aa24-44) traverses the ER bilayer in such a manner that the N terminal (24 aa residues) is retained in the ER lumen and an extended cytoplasmic domain is formed by 45-175 aa residues (Bergmann *et al*., 1989). The NSP4 disrupts Ca^{2+} homeostasis either via depletion of ER stores or elevating the levels of cytosolic Ca^{2+} (Tian *et al*., 1995). The peptide residues 114-135 are implicated in diarrhoea inducing activity (Hyser *et al.*, 2008). In addition, four antigenic sites in NSP4 designated antigenic site I (aa151-169), II (aa136-150), III (aa112-133) and IV (aa1-24), have been described in literature (Borgan *et al.*, 2003; Ball *et al.*, 2005).

2.6.2.5 Non-structural protein 5 (NSP5)

Non-structural protein 5 designated 'H' in the whole-genome classification due to its phosphorylation property (Matthijnssens *et al*., 2008). The NSP5 is O-glycosylated (Gonzalez and Burrone, 1991), acidic phosphoprotein with the lysine-rich region at the C terminus (Eichwald *et al.,* 2004). It recruits VP1, VP2 and, NSP2 inside the viroplasm to form the viral replication complex (Berois *et al*., 2003). It is also rich in serine and threonine residues and self assembles to form dimers (Torres-Vega *et al*., 2000). NSP5 also binds RNA non-specifically (Vende *et al.,* 2003).

The NSP2 and NSP5 are known to interact with one another (Jiang *et al*., 2006) and the co-expression of the two proteins leads to the upregulation of NSP5 hyper-phosphorylation resulting in the formation of viroplasm like structures (VLS) (Arnoldi *et al*., 2007). The signals that induce the formation of viroplasm like structures are embedded in the C and N terminal regions of NSP5 (Eichwald *et al*., 2004).

2.6.2.6 Non-structural protein 6 (NSP6)

Non-structural protein 6 is sometimes encoded by an alternative ORF of segment 11 of RVA and interacts with NSP5 in a manner that changes the three dimensional structure of NSP5 to enhance the activity of NSP5 (Torres-Vega *et al*., 2000). The NSP6 protein non-specifically binds to ssRNA and dsRNA (Rainsford and McCrae, 2007) although it may not be required for RVA replication since several RVA with defective NSP6 have been isolated (Ahmed *et al*., 2005). A study by Holloway and colleagues (2014), found NSP6 to contain highly conserved N terminal sequences that may be involved in directing it to mitochondria. The precise reasons for this localization have not been elucidated but it is well known that many viral proteins localize to mitochondria modulating mitochondria-related functions for the virus benefit (Castanier and Arnoult, 2011). Such virus-targeted activities include the release of cytochrome complex (c) factors from mitochondrial membrane space resulting in apoptotic cell death (Tait and Green, 2013). Since RVA has been found to promote apoptosis in certain enterocytes *in vivo*, NSP6 might be involved in modulating apoptosis (Halasz *et al*., 2010).

2.7 Replication cycle of rotavirus

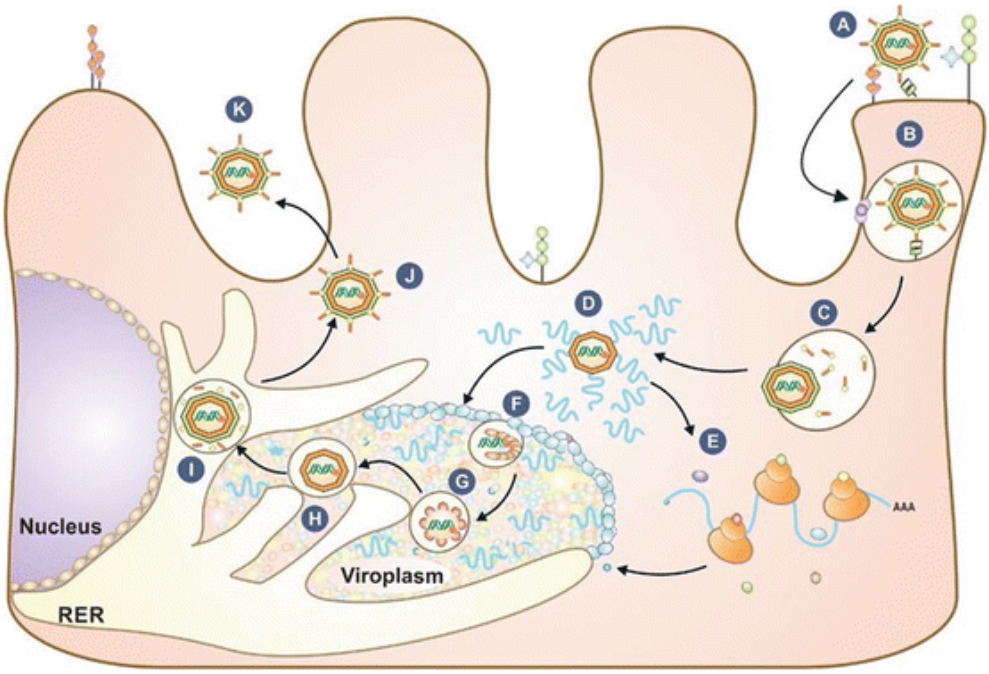


Figure 2.3: The rotavirus replication cycle
The rotavirus triple-layered particle (TLP) binds onto the host cellular receptors such as sialoglycans (**A**). Receptor-mediated endocytosis allows the TLP to enter into the host cell (**B**). The low Ca^{2+} levels in the endosome result in the loss of the outer capsid protein layer (**C**) releasing transcriptionally active double-layered particle (DLP) (**D**). Transcription of mRNA is then followed by mRNA translation resulting in viral proteins (**E**). Genome replication occurs after sufficient viral proteins are generated, and the synthesized DLP are packaged in viroplasm structures (**F-H**). Viroplasm then binds to the NSP4, which is an ER receptor. Transiently enveloped particles bud in the ER (**I**). As the transiently enveloped particles leave the ER, they assemble the VP4 and VP7 outer capsid proteins to form mature TLP (**J**) which are released from host cell via cell lysis (**K**).
From Lopez and Arias (2017) with permission granted from the publisher through the RightsLink® Copyright Clearance Center services (Appendix 3).

Rotavirus replication process occurs as follows:

(i) **Virus attachment** onto the host occurs initially through VP8*, one of the cleavage products of the protease-sensitive VP4, the other being VP5* (Settembre *et al*., 2011) (Figure 2.3). The VP5* and VP7 are involved in post-attachment interactions (Rodríguez and Luque, 2019). Several binding elements have been suggested such as cation-dependent mannose-6-phosphate (Diaz-Salinas *et al*., 2018), heat shock cognate protein 70 (Zarate *et al*., 2004), histo-blood group antigens (Arias *et al*., 2015),

integrins (Gutierrez *et al*., 2010), sialoglycans (Dormitzer *et al*., 2002) and Toll-like receptors (TLR) (Ge *et al*., 2013).

(ii) **Virus entry into the host cell** is mediated by clathrin-mediated endocytosis (Gutierrez *et al*., 2010) (Figure 2.3), although direct cell fusion entry model has also been suggested (Ludert *et al*., 1987). Studies have shown that there is a high level of plasticity for RVA entry pathways and possible existence of yet undescribed highly relevant entry factor (Arias *et al*., 2015).

(iii) **Uncoating** of the virion through the loss of the VP4 and VP7 is elicted by the low Ca^{2+} environment in the endosomes causing the transcriptionally active DLP to be released to the cell cytoplasm (Yoder *et al*., 2009) (Figure 2.3).

(iv) **Transcription and translation** occurs in the cytoplasm whereby the VP1-VP3 complex transcribes the capped positive (+) strand RNAs (+RNAs) of the 11 dsRNA segments (Silvestri *et al*., 2004). The +RNA serves as template for RNA synthesis during genome replication and can also be directly translated into viral proteins which occurs in viroplasms that are formed by the hyperphosphorlyation of NSP5 by NSP2 (Viskovska *et al*., 2014) (Figure 2.3). The cellular kinase CK1α phosphorylates NSP2 triggering NSP2 octamers to form a lattice structure required to build viroplasms (Criglar *et al*., 2018). Interaction of cytoplasmic NSP2 with VP1, VP2 and tubulin forms complexes (Criglar *et a*l., 2014) whereby tubulin presence induces stabilization of the viroplasm by acetylation (Eichwald *et al*., 2012).

(v) **Assembly** of subviral particles takes place in viroplasms. The assembly of outer capsid proteins is not well elucidated but it is suggested that NSP4 mediates the movement of VP4 as well as of DLPs through its affinity for VP6 to the cytosolic end of ER (Taylor *et al.*, 1996). The resultant complex of DLP, NSP4 and VP4 then buds into ER. The interaction with ER-bound VP7 results in the loss of transitory envelope and leads to formation of TLP (Desselberger, 2014).

(vi) **Virus release** from the infected cell after maturation occurs through non-classical vesicular transport or cell lysis (Gardet *et al*., 2006). The NSP4 is involved in the

rotavirus particle maturation (Silvestri *et al*., 2005). Upon release, the VP4 is cleaved by the proteases resulting in an infectious virion (Settembre *et al*., 2011).

2.8 Mechanisms of rotavirus evolution

Several mechanisms of genetic diversity like point mutations, gene reassortment, gene rearrangements and gene recombination contribute towards RVA evolution (Kirkwood, 2010).

2.8.1 Point mutations

The low-fidelity of viral RdRp results in high mutation rates (in the order of 10^{-4}) and is crucial to the evolution strategy of RNA viruses (Campagnola *et al.,* 2015). Base substitutions are the most prevalent of the point mutations (Turnpenny and Ellard, 2016). Several *in vitro* studies have enabled estimation of base substitution rates (Kitamoto *et al*., 1993; Blackhall *et al*., 1996), although the apparent mutation frequency is relatively higher *in vivo* due to the various selection pressure for viruses (e.g. neutralizing antibody) (Kobayashi *et al.,* 2003). Accumulation of point mutations contributes to the emergence of new genetic lineages/sub-lineages (Iturriza-Gomara *et al*., 2000). The occurrence of aa substitution in the neutralization epitopes of VP4 and VP7 may result in the generation of neutralizing antibody escape mutants (Rasebotsa *et al*., 2020). Amino acid substitutions in VP1 have been found to lead to changes in secondary structure that may alter functions such as viral replication/transcription (Abid *et al.,* 2020). Mutations at positions 385 and 393 in VP4 are involved in RVA attenuation and cell culture adaptation (Guo *et al*., 2020).

2.8.2 Gene reassortment

Gene reassortment in RVA occurs when RNA segments between two or more different strains recombine during co-infection in cells (Kobayashi *et al.,* 2003). Genetic reassortment contributes significantly to the overall diversity of RVA strains (Jere *et al*., 2018; Strydom *et al*., 2019; Maringa *et al*., 2020; Mwangi *et al*., 2020). Different RVA strains of the different genogroup may co-infect cells resulting in the progeny bearing a genome constellation that contains genome segments of parent strains (Nyaga *et al*., 2015). Although gene reassortment is not a rare event in RVA, there is inconclusive evidence that naturally co-circulating strains can generate predominant variants with a competitive advantage against existing strains in infecting host populations (Gentsch *et al*., 2005). Therefore, the infrequency of novel significant reassortants establishment in nature

may imply a degree of genetic fitness in the genome constellation of each virus that is not easily conferred to the reassortants from two different strains (Patton *et al*., 2006).

2.8.3 Gene rearrangements

Gene rearrangements take part in the evolution of RVA and contribute to their diversity (Tanigushi and Urasawa, 1995). It is suggested that during plus-strand synthesis the viral RdRp together with the associated synthesized RNA disengage from the RNA template and then rebinds to another upstream or downstream region along the same template where synthesis of nascent RNA is reinitiated, rearranging the sequence (Desselberger, 2014). Gene rearrangements in RVA frequently occur in gene segment 11 and less often involves segments five to ten. It is unclear whether there is some selective advantage for rearrangement in segment 11 (Estes and Greenberg, 2013). Rearranged RVA genome rarely occurs in nature and therefore, the evolutionary significance of rearrangement is not evident (Kobayashi *et al*., 2003).

2.8.4 Genetic recombination

Genetic recombination is assumed to be rare in RVA (Desselberger, 1996) due to the segmented nature of dsRNA genomes and the mechanism for transcription and replication (McDonald *et al.*, 2016). Unlike the +ssRNA and DNA viruses, dsRNA viruses are replicated within nucleocapsids (Patton *et al*., 2006), reducing chances for template switching (Hoxie and Dennehy, 2020). Despite this limitation, numerous studies show recombination plays a significant role in contributing to RVA diversity (Martinez-Laso *et al*., 2009; Donker *et al.*, 2011b; Jere *et al*., 2011; Esona *et al*., 2017; Jing *et al*., 2018). Recombination analysis of approximately 24,000 whole RVA genomes by Hoxie and Dennehy (2020) identified 109 occurrences of homologous recombination in all the gene segments except gene segment 7 and 11.

2.9 Rotavirus pathogenesis and pathophysiology

RVA infection stimulates the production of glycans, which acts as receptors aiding in RVA virulence (Engevik *et al*., 2020). To counter this glycan induction, *in vitro* microbiome studies demonstrate that mucin-degrading microbes such as *Bacteroides* and *Akkermanisa* bacterial members are stimulated to degrade the glycans and to reduce the binding capacity of RVA (Engevik *et al*., 2020). Viral-centric mechanisms are initiated to counter the early innate immune

response, with NSP1 as a key mediator (Arnoldi *et al*., 2013). The NSP1 suppresses interferon immune response by mediating degradation of IRF 3, 5 and 7 (Barro and Patton, 2007). Further, NSP1 downregulates RIG-I-like receptors, which are part of the pattern recognition receptors (PRRs) (Qin *et al*., 2011). The NSP3 accentuates viral translation by modifying the location of host PABp, minimizing its interaction with eIF4GI thus repressing host translation process (Lopez and Arias, 2012). The RVA-induced diarrhoea is primarily linked to the viral enterotoxin NSP4, which is a key viral protein of RVA pathogenesis (Ball *et al*., 1996). The NSP4 peptide region aa114-135 has been identified as critical for this enterotoxin activity (Tian *et a*l., 1995) and mediates the activation of the secretory component of RVA diarrhoea (Estes and Greenberg, 2013). It infects enterchromaffine gut cells releasing 5-hydroxytryptamin, a neurotransmitter that regulates gastrointestinal motility and triggers vagal afferent nerves leading to the stimulation of brain stem structures that control nausea and vomiting (Hagbom *et al*., 2012). These events mediated by interaction of the NSP4 with phospholipase C receptor elicit a cascade of reactions due to activation of cellular channels (such as store-operated calcium entry channel Orail1) (Chang-Graham *et al*., 2019) that results in the exit of chloride ions as well as water molecules across the plasma membrane (Sastri *et al*., 2016). During RVA infection, transport proteins usually present on brush-border intestinal membranes get intracellularly located, decreasing their activity (Zachos *et al*., 2019). Rotavirus infection is a systemic disease with clinical and pathophysiological implications beyond the gut, such as the liver and pancreas. However, the mechanisms involved are not well defined (Gomez-Rial *et al.*, 2019).

2.10 **Immune response and clearance**

Upon infection of the host cell, RVA is recognized by PRR or immune system cells such as dendritic cells, macrophages or adaptive B and T cells (Crawford *et al*., 2017). One of the best elucidated PRR is the RIG-I (Broquet *et al*., 2011). Other PRRs such as Toll-like receptor 3 (TLR3) and leucine rich repeats (LRR) are associated with age-dependent resistance to RVA as they are detected in elevated levels in adult animals (Zhu *et al*., 2017). Recognition of RVA by these PRRs results in induction of type I and III interferon response (Lin *et al*., 2016). Although Innate cytokines, interleukin 8 (IL-8) and interleukin 22 (IL-22), mediate RVA clearance as demonstrated in mice

studies (Zhang *et al*., 2014), the suppression of IFN by NSP1 underlies the weak response of the innate immunity (Holloway *et al*., 2014).

Therefore, RVA clearance is mainly mediated by the adaptive immune system cells such as CD8+ cytotoxic T cells (Mesa *et al*., 2010). Infants in the recuperative state have also been found to exhibit circulating RV specific T-helper (Th) cells in their blood samples (Parra *et al*., 2014). However, these T-helper cells express low efficiency in presenting antigens in infants and young children than in adults (Jaimes *et al*., 2002) and regulatory T cells do not appear to mediate or modulate RVA immunity during primary infection (Kim *et al*., 2008). Circulating RVA-specific CD8+ T cells do not mount a substantial immune response and it is suggested that B cells and antibodies are primarily essential in RVA clearance (Franco and Greenberg, 1995).

While the precise RVA immunity mechanisms are yet to be fully resolved, immunological factors are essential in susceptibility to RVA infection and systemic spread (Gomez-Rial *et al*., 2019). Genetic polymorphisms in genes encoding immunological factors, such as IFN-λ, have been shown to influence host immune response to RVA infection (Egli *et al*., 2014; Syedbasha and Egli, 2017). Serum IgA is the commonly used seroconversion measure (Paul *et al*., 2016). However, there is a need to consider VP6-specific immunoglobulin G (IgG) as a crucial correlate of protection as recent studies show that it is highly involved in intracellular antibody-mediated protection (Caddy *et al*., 2020).

2.11 **Management of rotavirus gastroenteritis**

Rotavirus gastroenteritis can be managed through oral rehydration therapy, diet and use of probiotics (Crawford *et al*., 2017). Oral rehydration salts (ORS) is a cornerstone therapy administered to treat RVA-induced dehydration (Centers for Disease Control and Prevention, 2004). Alternative homemade solutions of salt, sugar and water can be prepared in cases where ORS is not available (Churgay and Aftab, 2012). Management of diet is essential in caring for children with acute diarrhoea (Gaffey *et al*., 2013). Withdrawal of food for up to 6 hours after administering rehydration therapy is advised (Guarino *et al*., 2014). Adequate protein calorie intake (Gaffey *et al*., 2013) and zinc supplementation, especially in malnourished children, are regarded as essential measures to minimise symptoms of acute diarrhoea (Bzik *et al*., 2012).

Acute diarrhoea can be further managed by the use of probiotics (O'Ryan *et al*., 2010). Although probiotics are not commonly used to treat RVA-induced diarrhoea, they exhibit immunomodulatory effects and decrease repeat episodes of RVA diarrhoea (Vlasova *et al.*, 2016).

2.12 Prevention of rotavirus gastroenteritis using rotavirus vaccines

2.12.1 Prequalified rotavirus vaccines for global use

2.12.1.1 Rotarix®

Rotarix® is a monovalent, live attenuated oral vaccine developed by tissue culture passaging of wild type G1P[8] human RVA strain and licensed to GlaxoSmithKline Biologicals (Bernstein *et al*., 1999). Phase I trials in adults, children and infants demonstrated safety of the vaccine, albeit with mild fever incidences (De Vos *et al*., 2004). Evaluation of intussusception involving 63,000 enrolled infants from 11 South American countries and Finland found Rotarix® to be relatively free of intussusception risks (Ruiz-Palacios *et al*., 2006). Vaccine efficacy of Rotarix® involving 20,000 infants from this large clinical trial exhibited appreciable protection of 85% against severe RVA gastroenteritis. Vaccine efficacy against G1 serotype was 92% and 88% against serotypes G3, G4 or G9 (Ruiz-Palacios *et al*., 2006). Meta-analysis efficacy trials have demonstrated significant cross-protection by Rotarix® against non-G1P[8] strains whereby vaccine efficacy of 81% has been documented against G2P[4] strains (Vesikari *et al*., 2006).

2.12.1.1.1 Rotarix® vaccine trials and introduction in South Africa

Rotarix® was evaluated for interference with the Expanded Programme on Immunization (EPI) vaccines administered in South Africa by conducting trials comparing a 6-10 week schedule with a 10-14 week schedule (Steele *et al*., 2010a). Co-administration of Rotarix® with Oral Polio Vaccine (OPV) did not interfere with OPV seroprotection rates, although RVA IgA antibodies were significantly higher in the 10-14 weeks' schedule. The higher rate in the 10-14 weeks highlighted the potential role of maternal antibodies in diminishing the efficacy of RVA vaccines in low-income settings (Steele *et al*., 2010).

Safety, reactogenicity and immunogenicity of three Rotarix® doses at 6, 10 and 14 weeks was administered alongside routine EPI vaccines and was assessed in 100 HIV positive South African infants (Steele and Glass, 2011). The seroconversion rates in the Rotarix® group was 57.1% while

in the placebo groups it was at 18.2% and was not significantly different with the earlier immunogenicity studies (Steele and Glass, 2011). There was no adverse effects of the Rotarix® vaccine on the HIV positive infants (Steele and Glass, 2011). In this second phase of phase II trials, HIV status of the mothers was determined before enrolment of the infants. The necessity to know the HIV status was driven by the findings in the first phase of the phase II trials whereby seven deaths were reported in 271 subjects, which were linked to the HIV infection and disease (Steele *et al*., 2009). The immune response of the two dose regimen did not differ significantly (Steele *et al*., 2010).

Phase III clinical safety and efficacy trial was conducted in Malawi and South Africa (Madhi *et al*., 2010). From the 4417 enrolled infants, the vaccine efficacy in South Africa was 76.9% while that in Malawi was lower at 49.4% (Madhi *et al*., 2010). A total of 3168 infants aged between 5-10 weeks were enrolled at the South African trial site and randomized to receive the vaccine or placebo (Madhi *et al*., 2010). The national launch of the vaccine in South Africa occurred in September 2009 (Steele and Glass, 2011). Rotarix® is administered as two doses to infants beginning at 4 weeks and completion, preferably by age 16 weeks and no later than 24 weeks (Poelaert *et al.*, 2018).

2.12.1.2 Human-bovine rotavirus reassortant vaccine, RotaTeq®

RotaTeq™ is a pentavalent live, oral attenuated vaccine comprising five human-bovine reassortant RVA developed by Merck Research Company (Ciarlet and Schodel, 2009). Four reassortant viruses consist of human-derived VP7 proteins (G1, G2, G3 and G4) associated with bovine-derived VP4 protein (P7[5]). The fifth reassortant virus comprises a human-derived VP4 protein (P1A[8]) and bovine-derived VP7 protein (G6) (Ciarlet and Schodel, 2009). RotaTeq™ evaluation in Ghana, Kenya and Mali demonstrated vaccine efficacy of 64% against severe RVA diarrhoea during the first year of life (Armah *et al*., 2010). When co-administered with other routine childhood vaccines, RotaTeq™ immunogenicity is unaffected (Tanaka *et al*., 2017). RotaTeq™ administration is recommended in three oral doses at two, four and six months of age (beginning at 6 weeks of age with completion not later than 32 weeks) (Parashar *et al*., 2006).

2.12.1.3 Rotavac™

Rotavac™ is a monovalent, G9P[11] naturally attenuated oral vaccine derived from an Indian asymptomatic child and developed by Bharat Biotech (Das *et al.*, 1994). Rotavac™ is significantly efficacious against RVA-induced diarrhoea (53.6% vaccine efficacy in the first year of life and 48.9% in the second year) (Bhandari *et al.*, 2014a, 2014b). There is a non-interference of the immune response of other childhood vaccines when co-administered with Rotavac™(Chandola *et al.*, 2017). Rotavac™ was pre-qualified by WHO in January 2018 for global use in other developing country populations through GAVI-funded support (Kirkwood *et al*., 2019). Rotavac™ is administered alongside EPI vaccines at 6, 10, and 14 weeks (Ella *et al.*, 2018).

2.12.1.4 Rotasiil®

Rotasiil® is a pentavalent bovine-human reassortant live attenuated oral vaccine produced by Serum Institute of India (Zade *et al*., 2014). It is a distinctive vaccine owing to its thermostability of up to 18 months at 40^{o}C in its lyophilized form (Naik *et al*., 2017). The vaccine efficacy against severe RVA- induced gastroenteritis in a clinical trial in Niger was 67% and the data of adverse effects between the placebo and vaccine group was not statistically significant although the association of the vaccine with intussusception risk was not powered for evaluation (Coldiron *et al*., 2018). The vaccine efficacy against severe RVA-induced gastroenteritis in another clinical trial in India was 33% in the first year of life. The adverse effects reported between the placebo and vaccination group were similar although the study was not powered to evaluate association with intussusception risk (Kulkarni *et al*., 2017). Rotasiil® was found not to interfere with routinely administered childhood vaccines (Desai *et al*., 2018).

2.12.2 Rotavirus vaccines licensed for national use

2.12.2.1 Lanzhou lamb rotavirus (LLR) vaccine

Lanzhou lamb rotavirus vaccine is a monovalent live attenuated oral vaccine, developed by isolation of G10P[12] lamb strain by Lanzhou Institute of Biological Products (Ling-Qiao, 2001). A 44% vaccine effectiveness of LLR vaccine was reported against laboratory-confirmed RVA gastroenteritis in children 9–11 months of age and a vaccine effectiveness of 77% against RVA related hospitalizations (Fu, 2010; Fu *et al.*, 2012). The LLR vaccine is recommended in China and

is administered from 2 months to 3 years of age annually as a single dose and between 3 and 5 years of age as a booster dose (Fu *et al*., 2012).

2.12.2.2 Rotavin-M1-currently licensed in Vietnam

Rotavin-M1 is a monovalent, oral live attenuated human rotavirus vaccine, based on a multi tissue-culture passaged G1P[8] strain as identified at the Centre for Research and Production of Vaccines (POLYVAC) through the support of Ministry of Health, Vietnam (Le *et al.*, 2009). Immunogenicity assessment of Rotavin™ indicated a 4-fold increase in serum IgA responses with a range of 51% to 63% comparable to 58% seroconversion of Rotarix® (Anh *et al.*, 2012). Based on the appreciable immune responses, Rotavin-M1 was licensed in Vietnam in 2014 (Kirkwood *et al*., 2019).

2.13 Rotavirus vaccine efficacy in low-income countries

Sub-optimal vaccine efficacy has been reported in low-income countries. In Malawi and South Africa, trials on Rotarix® demonstrated approximate vaccine efficacy of 40% and 64%, respectively (Groome *et al.*, 2014; Madhi *et al.*, 2016). RotaTeq® trials in Ghana, Kenya, and Mali documented an average of 64% vaccine efficacy (Armah *et al.*, 2010), 51% in Vietnam and Bangladesh (Zaman *et al.*, 2010). The precise reasons for the suboptimal vaccine efficacy, especially in low-income countries have not been fully resolved but several factors listed, herein, have been suggested:

(i) ***Higher RVA genetic diversity and relative emergence of novel/unusual RVA strains*** (Todd *et al.*, 2010). While both Rotarix® and RotaTeq® have demonstrated effective heterotypic cross-protection against RVA strains (Correia *et al*., 2010; Steele *et al*., 2012), there has been considerable concern of potential strain replacement due to evolutionary selective pressure exerted by widespread use of RVA vaccines. Probable hints of strain replacements have been reported in Australia, Belgium, Latin America and the USA after Rotarix® and RotaTeq® vaccine introduction (Leshem *et al*., 2014). For instance, G2P[4] has been found to emerge with prolonged circulation in the post-vaccine era which is converse to earlier findings that G2P[4] occurs cyclically every 2-3 years (Linhares and Justino, 2014). In comparison to its comparatively infrequent occurrence during pre-vaccine period, G3P[8] strains emerged as the most common genotype during post-vaccine period in some regions like Australia and USA (Hull *et al*., 2011;

Kirkwood *et al*., 2011). Additionally, unusual genotypes such as G3P[14] (Cowley *et al.*, 2016; Matthijnssens *et al*., 2009) have been identified in post-vaccine era. Such strains might contribute to the suboptimal efficacy of RVA vaccine. Identification of unique sub-clusters during post-RVA vaccination period in a detailed whole-genome analysis of Belgian and Australian G1P[8] strains is an indication of potential vaccine-induced evolutionary pressure (Zeller *et al*., 2017). It's still unresolved whether the shift in circulating RVA strains is primarily due to the natural annual RVA fluctuations as documented in literature (Banyai *et al*., 2012) or vaccine-induced selective pressure is impacting the distinctive RVA genotype distribution patterns being reported during post-RVA vaccine era and this warrants whole-genome characterization studies.

(ii) ***Variations in gut microbiota likely due to environmental enteropathy*** (Korpe and Petri, 2012). In a Bangladesh study, environmental enteric dysfunction (EED) was linked to the suboptimal response to OPV or Rotarix® vaccine, even though parenteral immune response was not impaired (Naylor *et al*., 2015). The response of the microbiome to orally administered RVA vaccine was investigated by Harris and colleagues (2017). The first phase of the clinical study in Ghana matched vaccine responders with non-responders (Armah *et al*., 2016). *Streptococcus bovis* bacteria was substantially linked to vaccine immune responses and it was suggested that this bacterial family may have triggered immune response by acting as an adjuvant (Harris *et al*., 2017). These results indicate microbiota may give answers to the suboptimal RVA vaccine efficacy reported in developing countries and thus microbiome, including virome studies, are areas worth exploring (Steele *et al*., 2019).

(iii) ***Nutritional deficiency, especially insufficient nutritional factors such as zinc and vitamins A may affect vaccine uptake*** (Babji and Kang, 2012). Zinc deficiency is linked with EED and consequent suboptimal effect of oral RVA vaccine (Naylor *et al*., 2015). Clinical trials on an oral inactivated cholera vaccine showed that daily 20mg supplementation of elemental zinc resulted in a substantial (four-fold rise) in vibriocidal antibody titre in children from Bangladesh (Ahmed *et al*., 2009).

(iv) ***Possible neutralizing effect by the pre-existent maternal antibodies*** (Appaiahgari *et al.*, 2014). The immunogenicity of Rotarix® and RotaTeq® is inhibited by high levels of pre-existent

infant serum IgG as well as transplacentally-acquired IgG (Becker-Drepps *et al.*, 2015). There is a substantially higher IgA and IgG titer in women from low-income settings supposedly due to constant exposure to wild type RVA strains consequently suppressing vaccine replication (Chan *et al*., 2011; Moon *et al*., 2016). However, research evaluating the impact of withholding breastfeeding prior to and after vaccine administration indicated no adverse effect on the "uptake" of RVA vaccine (Groome *et al*., 2014) and one study reported a marginally improved IgA immune responses in infants during concomitant breastfeeding (Ali *et al*., 2015).

(v*) **Interference with co-administered OPV*** (Steele *et al*., 2010)**.** Five studies showed a substantial decrease of RVA-IgA seroconversion during co-administration OPV with RVA vaccines (Steele *et al*., 2010; Tregnaghi *et al*., 2011; Emperador *et al*., 2016, Li *et al*., 2016; Ramani *et al.*, 2016). There is still a need to resolve the interference mechanism involved (Emperador *et al*., 2016). There was no decrease in the efficacy of Rotarix® during co-administration in one efficacy study in middle income Latin American countries (Ruiz-Palacios *et al.*, 2006; Tregnaghi *et al*., 2011).

(vi) ***Histo-blood group antigen types*** (Nordgren *et al*., 2013). RVA strains bind differentially to human histo-blood group antigens (HBGAs) (Nordgren *et al.*, 2013). The P[8] genotypes are specifically associated with infection to Lewis-positive and secretor-positive phenotypes (Nordgren *et al.*, 2014). In contrast, the P[6] RVA strains mainly infect Lewis-negative individuals and these findings to a great extent answer the question of the suboptimal efficacy reported in African population (Atherly *et al*., 2012) whose majority of the individuals are Lewis-negative and the fact that P[6] genotypes have been reported to circulate at substantially higher levels in Africa (Banyai *et al*., 2012). However, a study by Lee *et al.* (2018) presented evidence that RVA vaccine susceptibility is unaffected by secretor status and that HBGA-mediated resistance to oral RVA vaccines is thus an unlikely mechanism of vaccine underperformance. To date, the cause of sub-optimal efficacy of oral RVA vaccines in low-income settings is not definitively resolved (Glass *et al*., 2018). Studies on withdrawal of breast-milk during immunization period (Groome *et al*., 2014), increasing a booster dose (Zaman *et al*., 2016), alteration of microbiome using probiotics or different vaccine schedule strategies have demonstrated only insignificant outcomes (Velasquez *et al*., 2018).

2.14 Rotavirus strains in the pre and post-vaccine era from a global perspective

Over time, RVA surveillance has shown the ever-changing dynamics of RVA strains in circulation (Iturriza-Gomara *et al*., 2001; Doro *et al*., 2014; Mwenda *et al*., 2014; Nyaga *et al*., 2018; Seheri *et al*., 2018). During the pre-vaccine era, different strains have been shown to circulate annually as well as reports of unusual strains occurring in low frequency attributed to the natural RVA evolutionary mechanisms. However, there is a keen interest in understanding the influence of RVA vaccine(s) for such dynamics observed during post-RVA vaccination era. Herein, an exploration of RVA strains during pre and post-vaccination era from a global perspective.

2.14.1 Americas

Following the introduction of RotaTeq® in USA in 2006, the subsequent year saw G3P[8] strains predominate over G1P[8] strains (Hull *et al*., 2011). RotaTeq® has the lowest (between 21% - 71%) serotype-specific immune response to G3P[8] strains (Vesikari *et al*., 2013) and this hypothetically explains the high identification of G3 strains in RotaTeq® jurisdictions. In Brazil, the first two years after the Rotarix® introduction, G2P[4] was the predominant strain (Gurgel *et al*., 2007). In subsequent years, other non-G1P[8] strains such as G12P[8] in 2014-2015 and G3P[8] in 2017, have predominated. Atypical DS-1-like G1P[8] strains in settings of Sao Paulo and Goias in Brazil have been reported with phylogenetic analysis reports indicating their importation from Asia (Luchs *et al*., 2019). Longitudinal 27-year full-genome surveillance on Brazilian G1P[8] strains did not find any evidence of vaccine-induced selective pressure with unchanged aa epitope profiles observed during pre- and post-vaccination periods (da Silva *et al*., 2015). However, there has been an evident incident of Rotarix®-reassortment with a community strain in Brazil (Rose *et al*., 2013). Equine-like G3P[8] RVA strains have also been reported in the Northern region of Brazil with a phylogenetic profile related to equine-like G3P[8] RVA reported in Asia, Europe and Oceania (Guerra *et al.,* 2016).

2.14.2 Asia

After the introduction of Rotarix® and RotaTeq® vaccines in Japan, the detection rate of some pre-vaccine era strains such as G1P[8] and G3P[8] decreased while strains that were found to be sporadic before vaccine introduction such as G2P[4], G2P[8], G9P[8] and G8P[8] emerged/re-

emerged (Hoque *et a*l., 2020). Atypical G9P[8] with I1-R1-C1-M1-A1-N1-T1-E2-H1 constellation has emerged recently in Japan (involving reassortment with local Japanese G9P[8] and G2P[4]) with detections in several locations in Tokyo (Fujii *et al*., 2019). DS-1-like G3P[8] RVA strains have also been sporadically detected in Japan (Komoto *et al.*, 2018). Post-RVA surveillance in Japan, reported two rare strains: G8P[14] with an I2-R2-C2-M2-A3-N2-T9-E2-H3 constellation and G3P[3] with an I3-R3-C3-M3-A9-N2-T3-E3-H6 constellation that circulated sporadically in 2014 and that emerged from interspecies transmission events (Okistu *et al*., 2018). Vietnam has also reported feline-like G3P[8] strains exhibiting a DS-1-like backbone that potentially emerged through reassortment events involving the DS-1-like backbone of G8P[8] strains acquiring feline like VP7 gene (Hoa-Tran *et al*., 2016). DS-1-like G8P[8] strains with bovine-like G8 genotype have emerged in Thailand, Vietnam and Japan (Guntapong *et al.*, 2017, Hoa-Tran *et al.,* 2016, Tacharoenmuang *et al*., 2016, Kondo *et al.,* 2017,Yodmeeklinen *et al.,*2018).

2.14.3 Europe

After introduction of Rotarix® vaccine in Belgium, G2P[4] immediately emerged as a predominant strain (Zeller *et al*. 2010). According to post-licensure epidemiological studies, Rotarix® exhibits the lowest (85%) serotype-specific immune response against G2P[4] strains (Matthijnssens *et al*., 2014). Thus, explaining the potential hints of strain shifts inclining towards emergence of G2P[4] in most Rotarix® jurisdictions. Identification of unique sub-clusters during post-RVA vaccination period in a detailed whole-genome analysis of Belgian and Australian G1P[8] strains is an indication of potential vaccine-induced evolutionary pressure (Zeller *et al*., 2017). A pre and post-RVA surveillance in Slovenia identified decrease in G1P[8] genotype from 74.1% to 8.7% from 2007 to 2008 immediately after Rotarix® introduction in 2006, with non-G1P[8] strains such as G4P[8] and G2P[4] prevailing in circulation in subsequent years (Steyer *et al*., 2014). The aa difference profile in the VP7 and VP8* antigenic epitopes was also observed in relation to the Rotarix® vaccine (Steyer *et al*., 2014). Recent study in Russia identified G9P[8] and G4P[8] as the most common genotypes circulating in Russia during the 2018/2019 post-RVA season (Ivashechekin *et al.,* 2020).

2.14.4 Oceania

Australia is among the countries that introduced both RotaTeq® and Rotarix® in the EPI which presented opportunity to study the impact of introduction of both vaccines in different states in Australia (Donato, 2012). After the two vaccines introduction, G2P[4] strain which circulated in low frequency (2-17%) during pre-vaccination years emerged as a dominant strain (occurring at 50%) during post-vaccine period (2008-2009 and 2010-2011) throughout Australia (Donato, 2012). The type of the vaccine used and the dominance of the strain observed was found not to correlate during the first two years of vaccine introduction (Donato, 2012). Further surveillance studies showed that these strain shifts were transient as G1P[8] which is covered by both vaccines was found to predominate in some seasons (Donato, 2012).

2.14.5 Africa

Post-vaccine Rotarix® introduction in Botswana saw a notable decline in G1P[8] and G9P[8] that had predominated during the pre-vaccine era (2011-2012), with G2P[4] emerging as a dominant strain post-vaccine (2013-2017) with peak levels during the 2013 season and with G3P[8] predominating in the 2018 season (Mokomane *et al*., 2019). Post-Rotarix® surveillance studies in Malawi reported atypical DS-1-like G1P[8] strains in a local setting in Blantyre which had emerged through reassortment events involving local Malawian strains (Jere *et al*., 2018). From a regional standpoint, based on a six year (2010-2015) surveillance of RVA diversity in 15 Eastern and Southern African countries, insignificant differences between pre and post circulating RVA genotypes, with yearly fluctuations demonstrated over time were reported (Seheri *et al*., 2018).

2.14.5.1 South Africa

South Africa phased in Rotarix® vaccine in its EPI in 2009 (WHO, 2009). After introduction of Rotarix®, non-G1P[8] strains which are not incorporated in the monovalent G1P[8] vaccine, increased dramatically (Page *et al*., 2018). Notably, at a key surveillance site (Dr George Mukhari Academic Hospital) no G1P[8] strains were reported in 2012 and 2013 seasons. The G2P[4] strains peaked and predominated in all surveillance sites in South Africa in 2013 (Page *et al*., 2018). While atypical G1P[8] rotavirus strains have been reported in several countries during post-vaccination period, a recent study by Mwangi *et al*. (2020) uncovered atypical G1P[8] strains that circulated

during pre-RVA vaccination period highlighting that some of the observed atypical strains reported during post-vaccination era are not necessarily due to vaccine pressure. There is a complex diversity of RVA strains in Africa and while most strain surveillance studies have been conducted by focusing on the outer capsid proteins VP4 and VP7, there is need for whole-genome characterization to fully decipher potential vaccine-induced strain changes (Nyaga *et al*., 2020).

2.15 Whole genome characterization using next generation sequencing (NGS) technology

Next generation sequencing (NGS) refers to different techniques such as sequencing by synthesis, sequencing by ligation and semi-conductor sequencing (Mardis, 2013). NGS generates billions of reads as compared to Sangers' single DNA sequence (Kulski *et al*., 2016). As Illumina® sequencing platforms dominates more than 90% of the NGS market share (Bronner and Quail, 2019), herein an exploration of its basic workflow.

The Illumina sequencing workflow entails the fragmentation of DNA fragments primarily by use of enzymes such as the transposase enzyme (Figure 2.4A), or mechanically using Covaris Focused-ultrasonicators. Adapters are ligated at both ends of the fragmented DNA, which are then PCR amplified to generate DNA libraries (Alvarez-Cubero *et al*., 2017). Adapter sequences have the following essential functions in Illumina® sequencing: barcoding/indexing sequences, forward/reverse primers for paired-end sequences and binding sequences for immobilizing the fragments to the flow cells (Mardis, 2013).

AMPureXP magnetic beads are utilized to size select DNA by removing low-quality fragments. Quantitative assessment of the DNA libraries is determined using Qubit or qPCR while the qualitative assessment of the DNA libraries is then performed using Agilent Bioanalyzer or Tape Station. The DNA library is then further normalized to appropriate Pico Molar concentrations, followed by chemical denaturation using sodium hydroxide (NaOH) and subsequent heat-denaturation at 96^{o}C before eventual pooling of the library for loading onto an Illumina® flow cell. Upon loading, the DNA libraries are covalently attached to the complementary oligos that are pre-affixed onto the flow cell. The free opposite end of a DNA library fragment hybridizes in a "bridge" like manner with a complementary oligo on the flow cell. DNA polymerase and the dNTP's then initiate bridge amplification (Caporaso *et al*., 2012) (Figure 2.4B).

The newly synthesized complementary strand and template strand are denatured and another amplification cycle is initiated. In the flow cell, millions of DNA clusters are generated and the generated clones of template DNA are then sequenced (Shedure and Ji, 2008) (Figure 2.4C). Sequencing entails a chain-terminating principle involving the use of reversible terminator (RT) nucleotide that is fluorescently tagged (Berglund *et al*., 2011). The DNA polymerase adds nucleotides onto the strand being synthesized. Each nucleotide incorporation is imaged with subsequent removal of terminating bases and fluorophores to allow another nucleotide incorporation cycle. The generated NGS reads are then de-multiplexed based on the index run results. The paired-end reads are then reference mapped or *de novo* assembled using bioinformatics tools (Figure 2.4D). The MiSeq™ and HiSeq™ platforms can provide the output of sequencing data at a relatively low cost in one to three days (Liu *et al*., 2012).

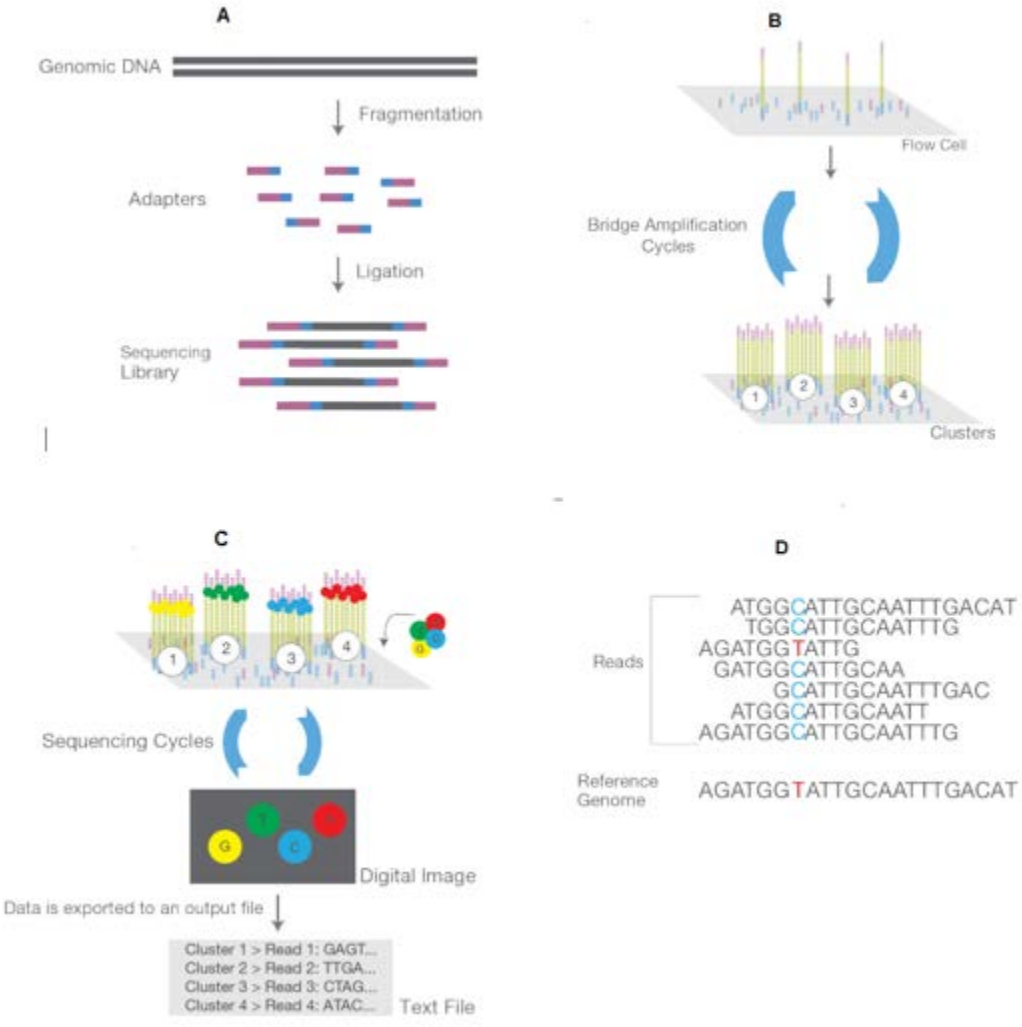


Figure 2.4:Illumina® next-generation sequencing chemistry overview
a) NGS library preparation. b) Cluster amplification of library c) Sequencing process d) Analysis of reads using bioinformatics
Image Courtesy of Illumina, Inc and adapted from https://emea.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf (Appendix 4)

2.16 Bioinformatics tools for analysis of rotavirus whole-genome data

2.16.1 Rotavirus genome assembly and generation of whole-genotype constellations

Rotavirus whole genome data generated from the Illumina platforms in FASTQ raw format can be *de novo* assembled or assembled using reference alignment to generate a consensus sequence. The genome assembly can be performed in multiple softwares, including the Geneious Prime software (Kearse *et al*., 2012) and/or CLC Genomics Workbench (https://www.qiagenbioinfomatics.com/). Alternative assembly tools can be developed from customized python scripts. The contigs generated after reference mapping or *de novo* assembly can be genotyped using the rVAgenotyper available at (https://www.viprbrc.org/brc/rvaGenotyper.spg?decorator=reo&method=RetrieveResults&ticketNumber=RG_594431877987). Multiple alignments can be performed in Molecular Evolutionary Genetics Analysis (MEGA) (Tamura *et al*., 2013) and/or Geneious Prime software packages (Kearse *et al*., 2012).

2.16.2 Tools to perform phylogenetic analysis

Phylogenetic analysis in RVA research is essential to enrich on the knowledge of evolution of RVA genes/ strains. Phylogenetic analysis resources abound and one of the most commonly used phylogenetic analysis tools is the MEGA and the MEGA Tree Explorer for its relative ease in editing the trees, selecting of sub-trees and tree image exports in different formats such as newick, phylip and nexus (Tamura *et al*., 2013). Other phylogenetic analysis tools include Dendroscope (Huson and Scornavacca, 2012) and ggtree (Yu *et al*., 2018).

2.16.3 Tools for analysis of selective pressure

Natural selection is a key evolution mechanism for any organism and DataMonkey, a webserver for the HyPhy package available at https://www.datamonkey.org/, is a valuable resource. DataMonkey is a collection of bioinformatics tools used to detect selection across branches,

individual sites or an entire genome. The tools include: Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), Mixed Effects Model of Evolution (MEME) and Single-likelihood Ancestor Counting (SLAC). FEL is based on the principle that each aa site in the phylogeny is under constant selective pressure. It is based on a maximum-likelihood approach for inferring non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny (Kosakovsky *et al.,* 2005). FUBAR is based on a Bayesian approach for the analysis of the non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny. The method is based on the assumption that selection pressure is constant for each site along the entire phylogeny (Murrell *et al*., 2013). MEME employs a mixed-effects maximum likelihood approach to test the hypothesis that individual sites have been subject to diversifying selection (Murrell *et al*., 2012). Single-Likelihood Ancestor Counting (SLAC) combines counting and maximum-likelihood approaches to infer non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny (Kosakovsky *et al*., 2005).

2.16.4 Tools to perform reassortment and recombination analysis

Genetic reassortment and recombination are some of the evolution mechanisms of RVA (Estes and Greenberg, 2013) and they can be assessed using several tools. One of the tools is the mVISTA, a suite of tools for comparative genomics enabling visualization of DNA alignments with annotation information (Frazer *et al*., 2004). It is implemented in an on-line webserver that provides access to global pairwise and multiple alignment tools. The mVISTA produces a clean output which can be used to visualize reassortment occurring in rotavirus DNA alignments at different resolution levels. Other tools that can be utilized for reassortment and recombination analysis include: Genetic Algorithm for Recombination Detection (GARD) (Kosakovsky *et al*., 2006), Recombination Detection Progragram (Martin and Rybicki, 2000) and Recombination Identification Program (RIP) (Siepel *et al*., 1995).

2.16.5 Tools to perform protein modelling and analysis

The understanding of RVA evolution can be enriched by assessing changes occurring at structural level through analysis of protein structures. One of the protein modelling resources is the SWISS-MODEL (available at https://swissmodel.expasy.org/), an ExPAsy webserver for protein structure homology-modelling (Waterhouse *et al*., 2018). The SWISS-MODEL webserver employs an intuitive user-interface that is easy to use. The aa target sequence file in FASTA format is uploaded or pasted. SWISS-MODEL algorithm then searches for the templates from the SWISS-MODEL Template Library (SMTL) that best fits the queried sequence. Upon display of the available templates, the user selects the optimal template based on preferences such as the resolution values and the method that was used to resolve the protein structure. The user then proceeds to build the protein model. Model results are generated and the quality of the protein structure can be assessed using the provided Structure Assessment feature. Other complementary protein modelling tools include YASARA (Krieger and Vriend, 2014), Modeller (Webb and Sali, 2016) and PyMOL (DeLano, 2009).

The effect of mutation(s) on the stability of the protein structure is essential to predict whether particular mutation(s) can be favoured in the longterm. FoldX is a structural protein model software used to predict energy differences in protein-structures with values that come close to experimental values (Van Durme *et al.,* 2011). Other tools that can be utilized include: ELASPIC (Strokach *et al*., 2019) and PROVEAN (Choi and Chan, 2015).

CHAPTER THREE: MATERIALS AND METHODS

3.1 Study location

The study was performed at the University of the Free State-Next Generation Sequencing (UFS-NGS) Unit, Division of Virology, Faculty of Health Sciences, Bloemfontein, South Africa.

3.2 Study design

The project was an experimental retrospective study to investigate the potential effect of vaccine pressure on the most prevalent circulating RVA genotypes (G1P[8] and G2P[4]) globally, over a 14 year period (seven years pre- and seven years post-vaccination introduction) and the evolution dynamics involved, therein. The study formed a subset of a broader ongoing project initiative entitled "Genomic investigation of Enteric Viruses from Human Diarrheal Samples Collected in Africa". Ethical approval for this extensive study is under ethics number: HSREC130/2016(UFS-HSD2016/1082). The Health Sciences Research Ethics Committee (HSREC) of the University of the Free State approved the study under ethics number: UFS-HSD2018/0510/3107 (Appendix 5).

3.3 Sampling method

The study involved archived RVA positive fecal samples from children under five years of age during pre- and post RVA vaccination introduction in South Africa. The samples were retrieved from the archival storage of the WHO regional rotavirus reference laboratory at the Diarrhoeal Pathogens Research Unit (DPRU) based at Sefako Makgatho Health Science University (SMU), Pretoria and also from Centre for Enteric Diseases, National Institute for Communicable Diseases (NICD), Johannesburg. The samples had been screened previously for rotavirus dsRNA by utilizing the Prospect™ Rotavirus Microplate Assay kit (Oxoid, Basingstoke, UK) and genotyped conventionally for genes encoding the VP4 and VP7, commonly used for binary classification of RVA into P- and G-types, respectively, as part of the WHO surveillance work for Africa. For this study, archived fecal samples collected from South Africa and previously genotyped as G1P[8] and G2P[4] were selected. Ideally, the study targeted a maximum of 420 faecal samples (15 of each genotype (G1P[8] and G2P[4]) per year for 14 years - seven years pre- and post- RVA vaccination) although collection was dependent on availability from archival storage.

3.4 Extraction of rotavirus dsRNA

A fecal suspension was prepared by adding a pea size (~ 100mg) stool sample into 200µl phosphate buffered saline (PBS) solution, 0.01M, pH 7.2 (Sigma-Aldrich®, St Louis, MO, USA) as RNA is more stable in PBS (Perchetti *et al*., 2020). The fecal suspension was vortexed (Chiltern Scientific, UK) briefly for 10 seconds and let to stand at room temperature for 10 minutes to allow the supernatant to form above the debris. A volume of 300µl supernatant was added to 900µl TRI-Reagent®-LS (Molecular Research Center, Cincinnati, OH, USA) in an RNase-free 2.0ml Eppendorf® tube (Eppendorf AG, Hamburg, Germany). The solution was vortexed briefly for 10 seconds and let to stand for 10 minutes at room temperature to ensure homogenization/lysis of the stool sample. A volume of 300µl chloroform (Sigma-Aldrich®, St Louis, MO, USA) was added to the solution containing Tri-Reagent®-LS (Molecular Research Center, Cincinnati, OH, USA) and fecal supernatant. The solution was vortexed briefly for 10 seconds and let to stand 5 minutes at room temperature. Then centrifugation was performed for 18000 x *g* for 20 minutes at 4^{o}C in a temperature-controlled micro-centrifuge (Eppendorf micro-centrifuge 5472R, Hamburg, Germany). The 4^{o}C temperature was ideal for phase separation as phenol is more soluble in the aqueous phase at room temperature. Three phases were produced, an upper aqueous phase containing the nucleic material and salts, interphase comprising of the proteins and a lower organic phase consisting of tissue debris and hydrophobic lipids (Chomczynski and Sacchi, 2006). After centrifugation, the aqueous supernatant was removed and added to 1.5ml Eppendorf® tubes (Eppendorf AG, Hamburg, Germany) containing 700µl of ice-cold isopropanol (Sigma-Aldrich®, St Louis, MO, USA). The ice-cold low temperature ensured minimal activity of RNA degrading enzymes. The solution was gently mixed by tilting the tubes up and down and then incubated at room temperature for 20 minutes to allow precipitation by isopropanol. Further, centrifugation was performed at 18000 *x g* for 30 minutes at room temperature in a microcentrifuge (Eppendorf micro-centrifuge 5472R, Hamburg, Germany). Electrophoresis was performed on 5µl aliquot of dsRNA in 1% 0.5 X TBE agarose (Bioline, UK) gel stained with Pronasafe (Condalab, UK) for 1 hour at 95 volts. The dsRNA bands were then visualized on a UV transilluminator (Syngene, Cambridge, UK) (Figure 3.1).

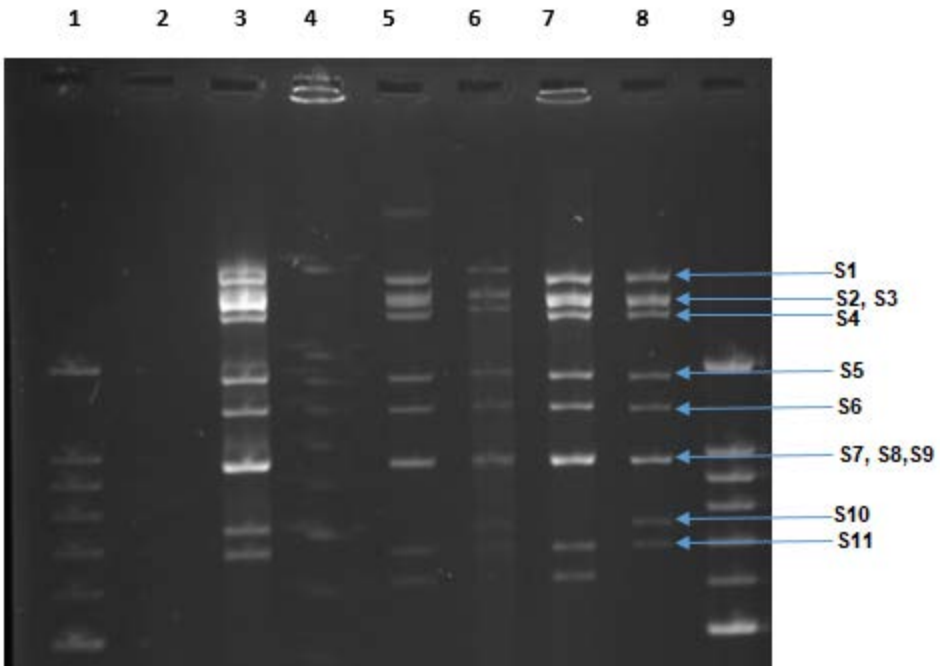


Figure 3.1:Verification of extracted RNA on agarose gel
A schematic representation of 1% 0.5X TBE agarose gel stained with Pronasafe (Condalab, UK) showing the bands/segments of the extracted RNA. In lane 1 and 9 the 100 bp ladder (Bioline, UK) is shown. Lane 2 indicates negative control, lane 3 indicates positive control, lane 4-8 indicate extracted RNA of study samples.

3.4.1 Enrichment of rotavirus dsRNA

An 8M $LiCl_2$ (Sigma-Aldrich®, St Louis, MO, USA) solution was prepared by dissolving 1.69g $LiCl_2$ in a 15ml Falcon conical tube (SPL Life Sciences, Korea) containing 5ml nuclease-free water (Sigma-Aldrich®, St Louis, MO, USA). A volume of 30µl $LiCl_2$ solution, while still hot (immediately after dissolving in water solvent), was then added to the tubes containing the extracted RNA isolate to enrich for dsRNA. The solution was further precipitated for 16 hours at 4°C in an even temperature water-bath incubation process (Potgieter *et al*., 2009).

3.5 Purification of rotavirus dsRNA

A volume of 330µl of PB buffer (Qiagen, Hilden, Germany), a solution containing guanidine hydrochloride and isopropanol, was added to the tubes containing dsRNA and lithium chloride. The solution was mixed gently by pipetting up and down three times. All the reaction mixture (450µl) was then transferred to labeled spin columns (Qiagen, Hilden, Germany) and centrifuged at 7000 *x g* at 4°C for 1 minute. Nucleic acids adsorbed to the spin column's silica membrane in

the high salt conditions provided by the PB buffer. The flow-through, containing the contaminants, was discarded. A 750µl volume of PE Wash buffer (Qiagen, Hilden, Germany), which is a 20% ethanol containing solution, was added to the spin columns and let to stand for 2 minutes. Centrifugation was then performed at 11,000 x g at 4^{o}C for 1 minute. The flow-through was discarded and centrifugation at 18,000 *x g* for 1 min at 4^{o}C was performed to get rid of the residual ethanol. The spin columns were then replaced with newly labeled sterile 1.5ml Eppendorf® tubes. A volume of 30µl EB elution resuspension buffer (Qiagen, Hilden, Germany) was added and let to stand for 1 minute and centrifuged at 7000 *x g* at 4^{o}C for 1 minute. The dsRNA was eluted in a much smaller volume (30µl) for RNA that is more concentrated. The column was then discarded and 2µl of purified RNA was run on 1% 0.5 X TBE agarose gel prepared at 100 volts for 90 minutes.

3.6 Quantification of dsRNA

The RNA concentration was measured on BioDrop™µLITE spectrophotometer (BioDrop, Cambridge, UK). Firstly, the BioDrop™µLITE was switched on to allow it to start up calibration. Then the 'Life Sciences' option was selected, followed by 'Nucleic Acids' and then the 'RNA' option. A volume of 2µl nuclease-free water (Sigma-Aldrich®, St Louis, MO, USA) was then used to clean the BioDrop uLite sample port. Then 2µl of EB buffer (Qiagen, Hilden, Germany) was loaded as a reference for calibration. The sample port was then wiped with a clean paper towel and 1µl of sample was loaded onto the port to measure the concentration. The RNA concentrations that fell within the A260/A280 absorbance ratio of 1.8-2.2 were considered to be of ideal purity range for further processing.

3.7 Complementary DNA (cDNA) synthesis

cDNA was synthesized from the extracted viral RNA using the Maxima H Minus Double Stranded Synthesis Protocol (Thermo Fischer Scientific, Waltham, MA), albeit with some modifications captured in the UFS-NGS Unit SOP. The modified Maxima H Minus Double Stranded Synthesis Protocol entailed two synthesis steps, first strand and second strand cDNA synthesis.

3.7.1 First strand cDNA synthesis

Firstly, a 13µl volume of RNA was denatured at 95^{o}C for 5 minutes in a Multigene Optimax thermocyler (Labnet, Edison, NJ, USA) with the lid pre-heated at 105^{o}C. The pre-heated lid allowed the samples to initiate action at the exact recommended temperature. The denaturing step allowed the dsRNA to unwind and separate into single strands. The tubes containing the denatured dsRNA were briefly spun down for 10 seconds to collect the condensation and consolidate the tube's contents and then they were again chilled on ice. The immediate chilling on ice minimized the renaturing of the strands. A volume of 1µl of 100µM random hexamer primer (Thermo Fischer Scientific, Waltham, MA) was then added into the tubes containing the denatured dsRNA on ice. The tubes were then centrifuged briefly for 10 seconds and incubated in the thermocyler (Labnet, Edison, NJ, USA) at 65^{o}C for 5 minutes to allow the random hexamer primer to anneal to the RNA template. After the annealing step, the tubes were immediately chilled on ice, briefly centrifuged for 10 seconds and then placed on ice. A volume of 5µl of the 4X first-strand reaction mix (Thermo Fischer Scientific, Waltham, MA) which is a blend of reverse transcriptase and RNase inhibitor was then added. The reverse transcriptase used is a proprietary enzyme that lacks RNase activity while the RNase inhibitor protected DNA degradation by RNases A, B and C at a temperature of up to 55^{o}C. The reaction was incubated in a thermocyler for 10 minutes at 25^{o}C followed by 2 hours at 50^{o}C, the optimal temperature for reverse transcription reaction. The strands were denatured at 85^{o}C for 5 minutes and hold at 4^{o}C in the thermocyler (Labnet, Edison, NJ, USA). The tubes were then placed on ice and then immediately proceeded with the second strand synthesis step.

3.7.2 Second strand cDNA synthesis

A volume of 55µl of nuclease-free water (Thermo Fischer Scientific, Waltham, MA) was added to the reaction mixture from the first strand reaction step. Then it was followed by addition of 20µl of 5X second strand reaction mix (Thermo Fischer Scientific, Waltham, MA) and 5µl of second strand enzyme mix (Thermo Fischer Scientific, Waltham, MA). The reaction mixture was briefly centrifuged for 10 seconds and incubated in the thermocyler (Labnet, Edison, NJ, USA) at 16^{o}C for 60 minutes. A 6µl volume of Ethylenediaminetetraacetic acid (EDTA), pH 8.0, (Thermo Fischer Scientific, Waltham, MA), a chelating agent that binds divalent metal ions such as Mg^{2+} that are

required for enzyme activity, was added to stop the reaction. Then a volume of 10µl of RNase I (Thermo Fischer Scientific, Waltham, MA) was added to the reaction mixture to remove residual RNA. The reaction mixture was then stored at -20^{o}C in readiness for cDNA purification.

3.8 Complementary DNA (cDNA) purification

The synthesized cDNA was purified using MSB® Spin PCRapace Kit (Stratec Molecular, Berlin Germany). A 50µl volume of cDNA sample was transferred to labelled DNase-free 1.5ml Eppendorf's® tubes (Eppendorf AG, Hamburg, Germany). A volume of 250µl of the binding buffer (Stratec Molecular, Berlin, Germany) was added to the 1.5ml Eppendorf® tubes (Eppendorf AG, Hamburg, Germany) containing 50µl of cDNA sample. The reaction mixture was transferred onto spin filter (Stratec Molecular, Berlin Germany) pre-placed in 2.0ml receiver tubes (Stratec Molecular, Berlin Germany). The reaction mixture was centrifuged for 4 minutes at 18,000 *x g* in a Z 216 Hermle MK microcentrifuge (Hermle Labortechnik, Wehingen, Germany). The spin filters were then placed onto labelled 1.5ml receiver tubes (Stratec Molecular, Berlin Germany). A 12µl volume of elution buffer (Stratec Molecular, Berlin Germany) was added directly onto the center of the spin filter. The reaction mixture was then centrifuged for 1 minute at 13,000 *x g*. The spin filter was then discarded. The purified cDNA was stored at -20^{o}C until the quantification step.

3.9 Purified complementary DNA (cDNA) quantification

The purified cDNA was quantified using a Qubit™ 3.0 (Invitrogen, Carlsbad, CA, USA), a fluorometric-based assay instrument. Qubit™ Working Solution was prepared by making a 1:200 dilution of the Qubit™ ds DNA HS reagent and Qubit™ ds DNA HS buffer from the Invitrogen™ Qubit™ ds DNA HS Assay Kit (ThermoFisher, Oregon, USA). Assay standards were prepared by adding 10µl of HS DNA Standards 1 and 2 (ThermoFisher, Oregon, USA) in separate Qubit™ assay tubes (Invitrogen, ThermoFisher, Oregon, USA). The HS DNA standard 1 and 2 were used to calibrate the Qubit™ instrument. A 2µl volume of the rotavirus cDNA sample was added to a Qubit™ assay tube containing 198µl of the Qubit™ working solution. The tubes were pulse-vortexed for 3 seconds and let to stand for 2 minutes at room temperature to allow the fluorescent dye in the Qubit™ working solution to bind the DNA. The tubes were then inserted in the reading port of the Qubit™ 3.0 instrument and the DNA concentration readings recorded in

ng/µl. The concentration range of 0.2ng/µl – 0.3ng/µl was considered ideal to proceed with DNA library preparations.

3.10 DNA library preparation

The DNA libraries were prepared using the Nextera® XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). The DNA library preparation entailed: (a) genomic DNA tagmentation (b) tagmented DNA amplification and (c) post-tagmented DNA clean up.

3.10.1 Genomic DNA tagmentation

This step involved fragmentation of genomic DNA by using the Nextera® transposome, a proprietary enzyme that also subsequently tags the fragmented DNA with adapter sequences. The following reagents were utilized during the tagmentation of genomic DNA: Tagment DNA (TD) buffer, Ampilicon Tagment Mix (ATM) and Neutralize Tagment (NT) buffer from the Nextera XT Library Preparation Kit. A 5µl volume of purified cDNA within the 0.2-0.3ng/µl concentration range was pipetted to each well of UltraFlux® 96-well PCR plate (Scientific Specialties Inc, Lodi, CA, USA). Tagment DNA buffer (10µl), was added to each well of the PCR plate followed by 5µl of ATM. The reaction mixture was mixed up and down with a pipette and the PCR plate was centrifuged at 280 *x g* at room temperature for 1 minute. The PCR plate was placed in a thermocyler programmed with pre-heated lid option at 55^{o}C for 5 minutes with a 10^{o}C hold temperature. A 5µl volume of NT buffer was added to neutralize the activity of the transposome enzyme. The PCR plate was sealed and centrifuged in Z306 Hermle microcentrifuge (Hermle Labortechnik, Wehingen, Germany) at 280 *x g* at room temperature for 1 minute followed by 5 minutes incubation at room temperature and then proceeded immediately with the amplification step.

3.10.2 Tagmented DNA amplification

The tagmented DNA was amplified using a limited-cycle PCR program. This PCR step also added indices, index 1 (i7) and index 2 (i5) required for cluster generation and sequencing. The reagents utilized were Nextera PCR Mastermix (NPM) from Nextera®XT DNA Library Preparation Kit and index 1 and 2 primers from the Nextera® XT Index Kit. Firstly, an index sample sheet was generated using the Illumina® Experimental Manager (IEM) software v.1.18.1 (Illumina

Experimental Manager (IEM) Software Downloads, 2020). The index sample sheet enabled to associate each sample with its unique dual index combinations. A volume of 5µl of each index 1 primers was added to each of the sample wells followed by an addition of 5µl of index 2 into the sample wells based on the unique combinations allotted on the sample sheet. A 15µl volume of Nextera PCR Mastermix was added to each well and then the plate was covered with Microseal®A film (Bio-Rad, Hercules, CA, USA) and centrifuged at 280 *x g* at 20^{o}C for 1 minute. The PCR program was performed on a thermocyler programmed with pre-heated lid option, 72^{o}C for 3 minutes and 12 cycles of 95^{o}C for 10 seconds, 55^{o}C for 30 seconds and 72^{o}C for 30 seconds then 72^{o}C for 5 minutes and hold at 10^{o}C. The amplified tagmented DNA library was stored at 2^{o}C-8^{o}C for a day before proceeding with library clean-up step.

3.10.3 Post-tagmented and amplified DNA clean up

The post-tagmented and amplified DNA was purified to remove unwanted PCR contaminants. AMPure XP beads (Beckman Coulter, Pasadena, CA, USA) were let to acclimatize to room temperature for 30 minutes after getting removed from 4^{o}C storage. The indexed libraries in the PCR plate were centrifuged at 280 *x g* for 1 minute at room temperature. The AMPure XP beads were vortexed for 30 seconds using a VX-200 Vortex Mixer (Labnet, USA) for even distribution of the beads in the solution. A sufficient volume of beads was added to a trough (SPL Life Sciences, Korea) from which 30µl of beads was aspirated and added to each sample well of the PCR plate.

The plate was sealed using Microseal 'A' film (Bio-Rad, USA) and shook for 2 minutes at 1800 rpm using a Micro Multiple Genie Plate Shaker (Scientific Industries, Bohemia, New York, USA). The plate was then incubated for 5 minutes without shaking. This step enabled the genomic material to bind to the magnetic beads through their negatively charged DNA molecules. The plate was placed on a magnetic stand until the supernatant was clear. A 200µl volume of freshly prepared 80% ethanol (Sigma-Aldrich®, St Louis, MO, USA) was used to wash off any impurities bound to the beads and the step was repeated twice. The beads, while on the magnetic stand, were air-dried for 10 minutes followed by removal of the plate from the stand. Resuspension Buffer (RSB) (52.5µl) was collected from a trough and added to each sample well. The plate was sealed by Microseal 'A' film (Bio-Rad, Hercules, CA, USA) and shook at 400 *x g* for 2 minutes and then

incubated for 2 minutes at room temperature. The PCR plate was then placed on a magnetic stand and when the supernatant was clear, 50µl of supernatant was removed from each of the sample and transferred to a newly-labelled 96 well PCR plate, sealed and stored at -20^{o}C awaiting next step.

3.11 Quantitative assessment of DNA libraries

The indexed and cleaned DNA libraries were quantified on the Qubit™ 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), following the procedures as earlier described in section 3.9.

3.12 Qualitative assessment of DNA libraries

The fragment size distribution and quality of the DNA libraries was assessed on an Agilent 2100 BioAnalyzer® (Agilent Technologies, Waldbronn, Germany). Briefly, the gel-dye mix was prepared by letting the DNA dye concentrate and DNA gel matrix to normalize for 30 minutes at room temperature. The DNA dye concentrate was vortexed for 10 second followed by pipetting 15 µl the DNA dye concentrate into the DNA gel matrix and stored at 4^{o}C in the dark. The tube was then capped, vortexed for 10 seconds and the gel and dye were visually inspected. The complete gel-dye mix was transferred onto a spin filter. The spin filter was spun for 10 minutes at room temperature for 4000 *x g*. The filter was then discarded. A volume of 9.0µl of the gel-dye mix was inserted to the bottom of well indicated by capital 'G' (in the high sensitivity DNA chip). The plunger of the DNA chip priming station was positioned at 1ml. After locking the chip priming station in place, the syringe was immediately pushed down and held by the clip. The plunger was released after 60 seconds and visually inspected to ensure it was was past the 0.3ml mark before pulling it slowly back to the 1ml position. After opening the chip priming station, a 9.0 µl volume of the gel-dye mixture was added to the wells indicated by letter 'G'. A 5µl volume of the DNA marker was added to the well indicated by the ladder symbol and then also added to all the 11 sample wells. A volume of 1µl of the DNA ladder was added into the well indicated by the ladder symbol. Then 1µl of the sample was pipetted into the 11 sample wells. The chip was the vortexed on a IKA® vortex mixer (IKA®-Werke, Staufen, Germany) for 60 seconds at 700 *x g* and then placed into the Agilent 2100 Bionalyzer (Agilent Technologies, Waldbronn, Germany) to perform a run.

A full report was generated for each run showing the electrophoregram with fragment size distribution (Figure 3.2).

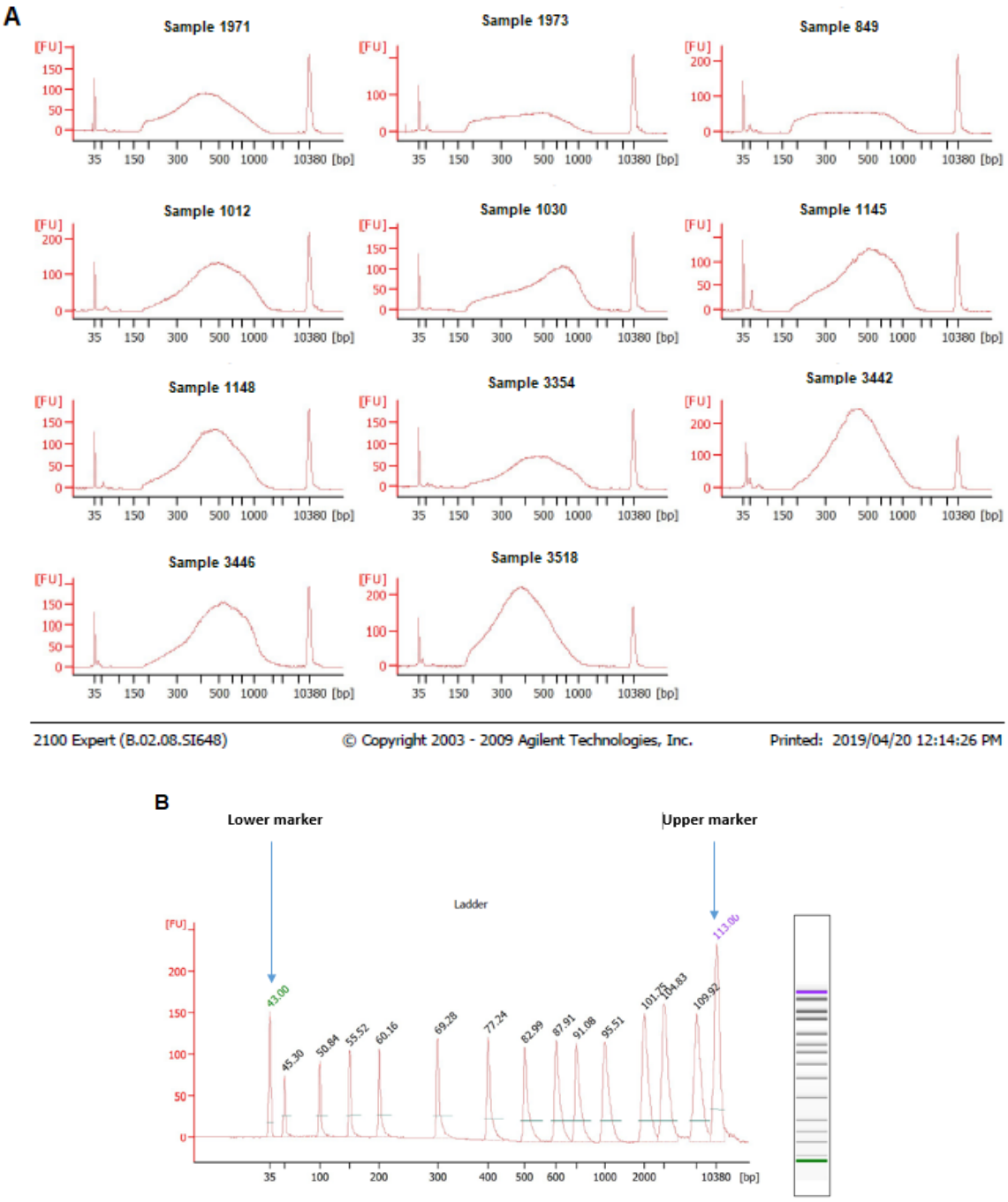


Figure 3.2: BioAnalyzer® electrophoregram report

A: An electrophoregram report showing fragment size distribution in base pairs (bp) and library concentration in arbitrary fluorescence units (FU) for 11 randomly selected representative DNA libraries. Sample IDS 1971 and 1973 are featured in manuscript 1 (chapter 4), sample IDs 849, 1012,

1030, 1145 and 1148 are representative samples processed for manuscript 2 (chapter 5) while sample IDs 3354,3442, 3466 and 3518 are a representation of samples processed for manuscript 3 (Chapter 6).**B**:An electrophoregram ladder report showing peaks that were utilized to validate the assay. The arrows indicate the lower marker (in green) and upper marker (in purple).

3.13 Normalization of DNA libraries

The DNA libraries concentrations in ng/µl was converted into nM using the formula below:

$$\frac{Concentration\ in\ ng/\mu l}{(660g/mol\ x\ Average\ library\ size)}\ X\ 10^6 = Concentration\ in\ nM$$

After determining the library concentrations in nM units, all the libraries were normalized to 4 nM as recommended in the Nextera XT protocol ((Illumina, San Diego, CA, USA). This was done by diluting the library with Elution Buffer (Qiagen, Hilden, Germany) using the formula below:

$$C1 \times V1 = C2 \times V2$$

Whereby

C_1=Concentration of the DNA library
V_1=Volume of the DNA library
C_2=Concentration of the DNA library targeted
V_2=Total volume

3.14 Pooling of the DNA libraries

All the DNA libraries (at 4nM concentration) were pooled together into a single 1.5ml Eppendorf® tube (Eppendorf AG, Hamburg, Germany) by combining 5 µL of each DNA library.

3.14.1 Library denaturation and dilution

A fresh solution of 0.2N NaOH (Sigma-Aldrich®, St Louis, MO, USA) was prepared from a stock solution of 50% w/v by pipetting 11µl of stock solution to 250µl molecular-grade water (Sigma-Aldrich®, St Louis, MO, USA) with a final volume of 1ml. The mixture was vortexed for 10 seconds and then incubated at room temperature for 5 minutes to allow denaturation of DNA libraries into single strands. The 4nM DNA library was diluted to 20pM by addition of 990 µl of hybridization buffer.The reaction mixture was further diluted to 8pM by addition of 240µl of the 20pM DNA library to 360µl pre-chilled HT1 and mixed by gently inverting five times.

3.14.2 PhiX library denaturation and dilution

PhiX (Illumina, San Diego, CA, USA), was used as a positive control during Illumina® sequencing. The 10nM Phix control was diluted to 20pM concentration.Briefly, 2 µl of the PhiX was added to 3 µl of Elution Buffer (Qiagen, Hilden, Germany). The resultant mixture was chemically denatured by addition of 5µl of freshly prepared 0.2N NaOH. The mixture was incubated for 5 minutes at room temperature and 990µl of hybridization buffer was then added to dilute the PhiX to 20pM.

3.14.3 Combining DNA library and PhiX V3 Library

The DNA library (480µl at 8pM) was combined with 20% of the PhiX (120µl at 20pM) in preparation for loading onto the Illumina MiSeq V3 reagent cartridge (Illumina, San Diego, CA, USA).

3.15 Preparation of the MiSeq v3 reagent cartridge

The MiSeq V3 cartridge, already thawed in de-ionized water for approximately 90 minutes was removed and dried completely by gently tapping it on a paper towel. The reagent cartridge was gently titled up and down ten times to ensure the thawed reagents inside the cartridge were thoroughly mixed and free of precipitates.

3.16 Performing a run on the Illumina MiSeq®

3.16.1 Cleaning the flow cell

The flow cell was removed from its storage container and lightly rinsed with molecular-grade water (Sigma-Aldrich®, St Louis, MO, USA) to get rid of the excess salts. The flow cell was then wiped using a paper towel. The area of the gasket and adjacent glass was gently pat dry.

3.16.2 Loading sample libraries

The combined DNA library and PhiX control library was incubated at 96^{o}C for two minutes on an Accublock™ Digital Dry Bath (Labnet, NJ,USA). After incubation, the tube was gently tilted up and down two times and then placed in the ice-water bath. The foil seal covering the reservoir labeled ‘Load Samples’ on the MiSeq cartridge was cleaned with a low-lint laboratory tissue. The foil seal was pierced using a clean 1000µl pipette tip. A volume of 600µl of the prepared libraries (DNA library and PhiX v3 control library) was used at a percentage value of 90% for the DNA library

versus 10% of the PhiX control. Then proceeded immediately to run the setup steps using the MiSeq® Control Software (MCS) interface and the run performed for 600 cycles (301 x 2 paired end) on a MiSeq® benchtop sequencer (Illumina, San Diego, CA, USA) at the UFS-NGS Unit.

3.17 Data analysis

3.17.1 Genome assembly

The FASTQ raw data from the Illumina MiSeq® was subjected to both *de novo* and reference-based rotavirus genome assembly by using Geneious Prime® 2020.1.1 (Kearse *et al*., 2012). The reference mapping and *de novo* tools integrated into the Geneious Prime were utilized for subsequent assembly process. An in-house data assembly pipeline and CLC Genomics Workbench (https://www.qiagenbioinfomatics.com/), were utilized as complementary genome assembly resources.

3.17.2 Generation of whole-genome constellations

The generated assembled contigs were genotyped using rvaGenotyper, an annotation pipeline for genotyping RVA. rvaGenotyper is an optimized reimplementation of RotaC$^{2.0}$ (Maes *et al*., 2009).

3.17.3 Phylogenetic analysis

Full-length sequences for each gene segment were aligned using MAFFT package in Geneious Prime. Duplicated sequences were identified using ElimDupes, a duplicate sequence removal tool (https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html). Upon alignment, the optimum evolutionary model befitting each dataset was calculated using Model Test program in MEGA 6 (Tamura et *al*., 2013). The models were applied to construct phylogenetic trees (Maximum-likelihood) using MEGA 6 supported by a 1000 bootstrap replicates to estimate branch support. The number of duplicate sequences were indicated alongside the sequences in the phylogenetic tree. Identity genetic matrices were performed using the p-distance algorithm in MEGA 6 (Tamura et *al*., 2013).

3.17.4 Selection pressure and recombination analysis

Analysis of natural selection in RVA gene segments was done using the suite from the DataMonkey Webserver (Weaver *et al*., 2018): Fixed-effects Likelihood (FEL) (Kosakovsky *et al*., 2005), Fast Unconstrained Bayesian Approximation for Inferring Selection (FUBAR) (Murrell *et al*., 2013), Mixed-effects model of episodic selection (MEME)(Murrell *et al*., 2012) and Single-Likelihood Ancestor Counting (SLAC) (Kosakovsky *et al*., 2005). Due to possible variations in the results from the methods, sites under positive selection that were identified by at least three of the aforementioned methods were considered. Genetic Algorithm for Recombination Detection (GARD) was used to perform analysis of genetic recombination (Kosakovsky *et al*., 2005).

3.17.5 Reassortment analysis

Reassortment analysys was assessed using the mVISTA web platform (Frazer *et al*., 2004). Briefly, concatenated sequences were uploaded in fasta format into the mVISTA web interface and the 'Submit' button was clicked. Three viewing options were generated: (a) the Text Browser, which provided all the detailed information such as sequences, alignments and conserved sequence statistics (b) VISTA Broswer, which is a dynamic visualization tool that provided options to adjust VISTA curve and conserved sequence parameters and (c)PDF file, which provided a static VISTA plot of alignment.

3.17.6 Protein-modelling

The rotavirus protein structures were modeled using SWISS-MODEL (Waterhouse *et al*., 2018). First, the aa fasta sequence was programed to perform a template search. The template was selected from the SWISS-MODEL Template Library (SMTL). The protein structure templates and respective resolution values for the gene segments VP7 and VP4 were (3fmg.1, 3.40Å) and (2dwr.1, 2.50Å), respectively. Upon selection of the template, the modelling was then performed. The evaluation of stereo chemical quality parameters of the generated structures was performed using the structure assessment feature in the SWISS-MODEL server (Waterhouse *et al*., 2018) (Figure 3.3).

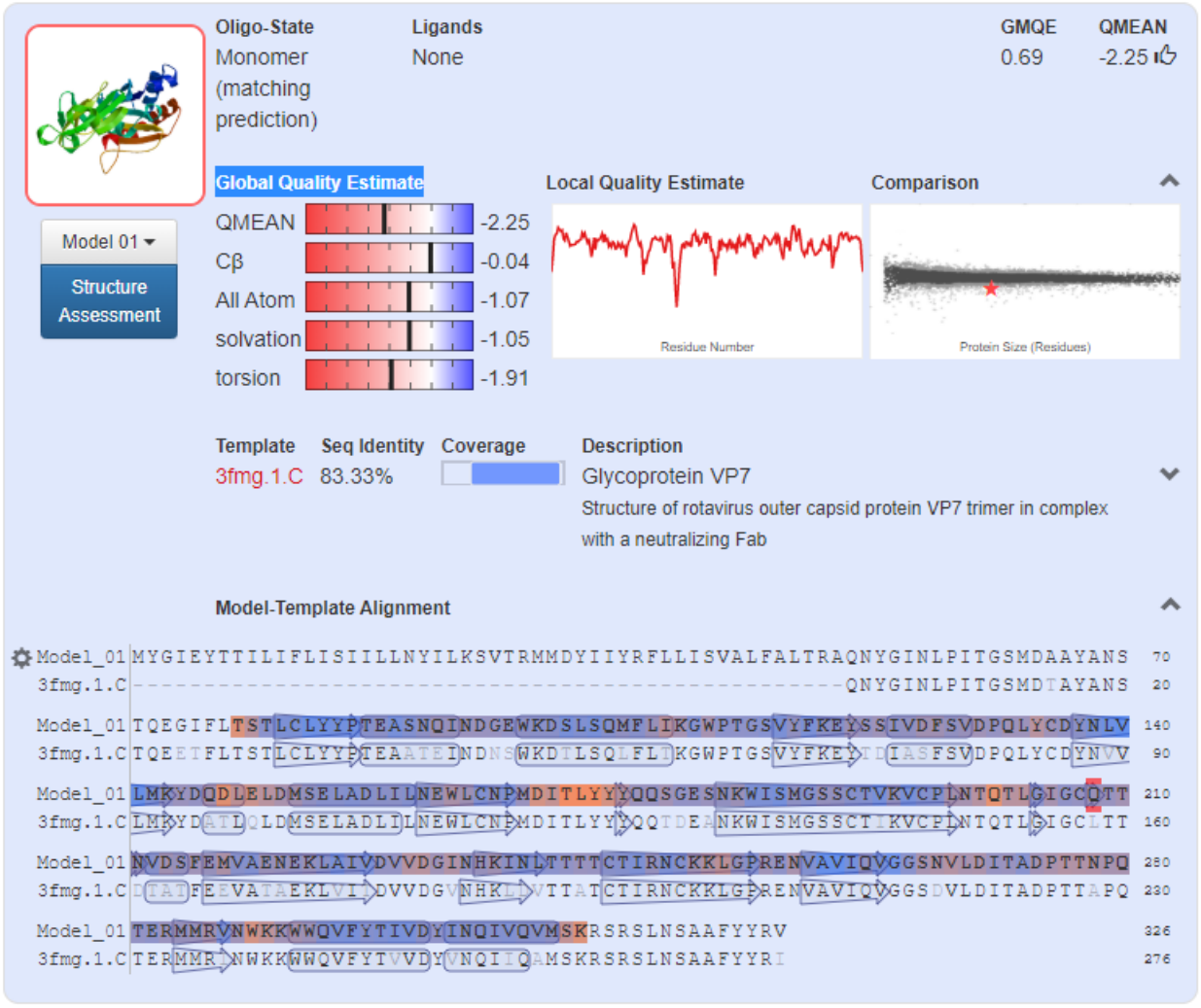


Figure 3.3: Swiss-Model Structural Assessment Report
A Swiss-Model Structural Assessment Report for a representative protein structure of VP7 study strain 4357 analysed in manuscript 2 (chapter 5) Indicating Global Model Quality Estimate (GMQE). GMQE is expressed as a number between 0 and 1, reflecting the accuracy of a model built with that alignment and template and the coverage of the target.

The 3D quality of the structure was further assessed using VERIFY 3D which generates a pass or fail result based on the quality of the structure (Figure 3.4). In this study, protein structures that did not pass the quality check were discarded. Visualization and production of images were performed in PyMol software (DeLano, 2009). In order to assess structural conformation, superposition of two protein structures was performed in PyMol and the alignment value generated in Root Mean Square Deviation (RMSD), whereby a value of 0 indicates absolute structural homology (Maiorov and Crippen, 1994).

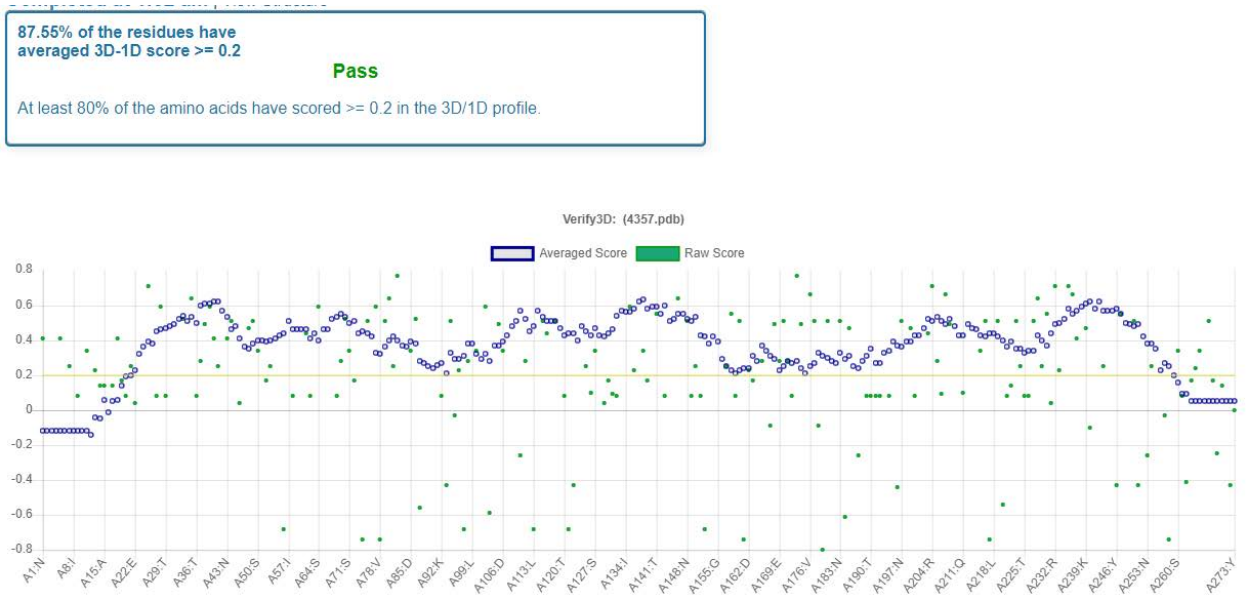


Figure 3.4: VERIFY 3D protein structure quality assessment
A VERIFY 3D quality assessment result for a representative VP7 study strain 4357 in manuscript 2 (chapter 5) Indicating a Pass result for the analysed structure. VERIFY 3D generates a pass or fail result based on the quality of the structure.

3.17.7 In-silico analysis of effect of mutation(s) on protein stability

The FoldX plugin (Van Durme *et al*., 2011) integrated in the YASARA platform (Krieger *et al*., 2002) was used to predict the stability effect of mutation(s) in a 3D structure. First, the pdb (protein data bank) file was loaded into YASARA. The FoldX RepairPDB command was utilized to minimize the energy of the protein structure by rearranging the aa side chains in order to get a better free energy of the protein. This was necessary in order to get accurate results when subsequently calculating the stability effects. The mutate residue command was then applied to mutate one or multiple aa residues to one or more new aa residues. The FoldX Stability command was used to calculate the free energy of unfolding. FoldX uses an empirical method to estimate the stability effect of a mutation whereby the stability (ΔG) of a protein was defined by the free energy, which was expressed in kcal/mol. Herein, G is the difference of free energy between the Rotarix® (vaccine) strain and mutant strain. The energy with positive value was regarded to destabilize the structure, while a mutation with a negative value was regarded to stabilize the structure. Folding energy change of ±0.5 kcal/mol was regarded to be statistically significant for stabilizing/destabilizing effect (van Durme *et al.,* 2011).

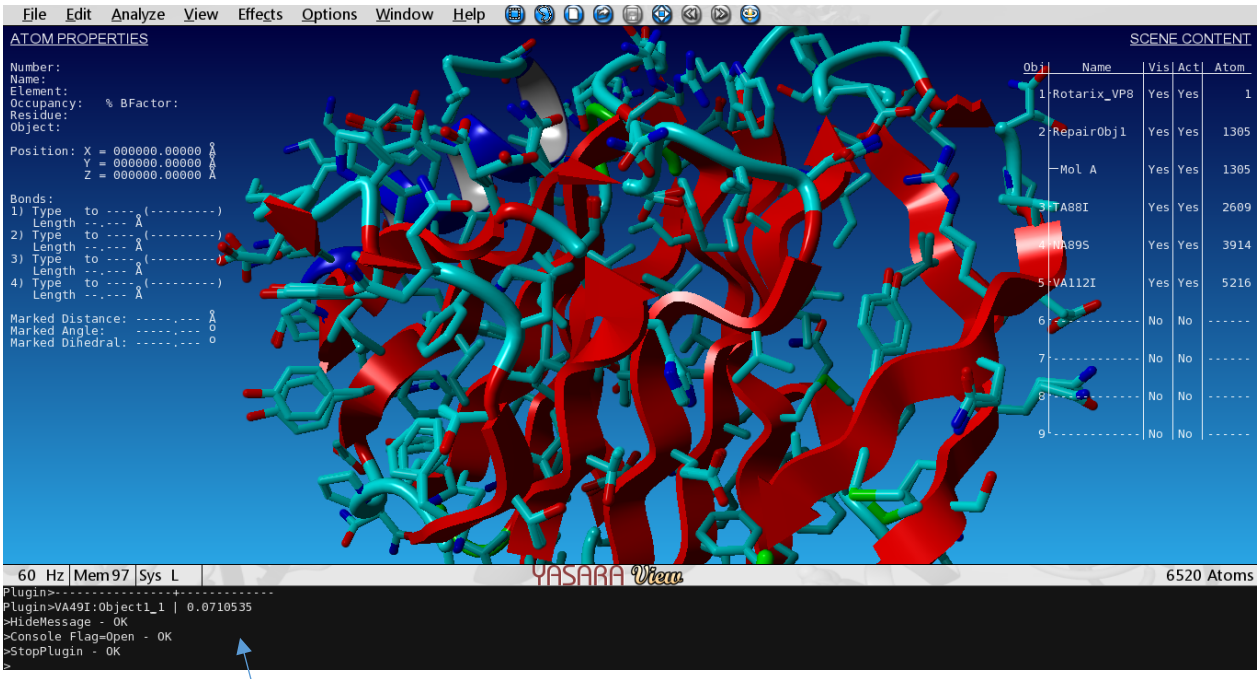


Figure 3.5: FoldX YASARA free energy change analysis

A generated visual report of the free energy change analysis of a mutated residue in kcalmol of a representative VP7 study protein of strain 1544 analysed in manuscript 2 (chapter 5) performed in YASARA using the FoldX plug in. The arrow indicates the generated result. FoldX calculates the free energy of the wild type and the mutant and generates the difference.

CHAPTER FOUR

INVESTIGATION OF GENOTYPE VARIATION AND THE EMERGENCE OF POTENTIAL VACCINE ESCAPE STRAINS PRE-AND POST-VACCINATION INTRODUCTION.

Chapter four is based on the publication "Uncovering the first atypical DS-1-like G1P[8] rotavirus strains that circulated during pre-rotavirus vaccine introduction era in South Africa"

Mwangi, P.N., Mogotsi, M.T., Rasebotsa, S.P., Seheri, M.L., Mphahlele, M.J., Ndze, V.N., Dennis, F.E., Jere, K.C. and Nyaga, M.M., 2020. Uncovering the First Atypical DS-1-like G1P [8] Rotavirus Strains That Circulated during Pre-Rotavirus Vaccine Introduction Era in South Africa. *Pathogens*, *9*(5), p.391. Impact Factor 3.405

https://pubmed.ncbi.nlm.nih.gov/32443835/

4.1 Introduction

Diarrhea persists as a leading infectious mortality cause in children under the age of five worldwide (WHO, 2020). Group A rotavirus (RVA) is the primary viral etiologic agent for AGE in children under five years of age (Estes and Greenberg, 2013), resulting in annual mortality cases ranging from 122,322 to 215,757 with an estimated 81% reported in sub-Saharan Africa and Southeast Asia (Tate *et al*., 2016; Troeger *et al*., 2018). To combat RVA diarrhea especially in countries with high RVA disease burden, the WHO recommends incorporation of RVA vaccines into the national immunization programs alongside other childhood vaccines (WHO, 2009). The WHO has prequalified four vaccines (Rotarix®, GlaxoSmithKline, Rixenstart, Belgium; RotaTeq®, Merck & Co, USA; Rotavac®, Bharat Biotech, Hyderabad, India and Rotasiil®, Serum Institute of India, Pune, India) for global use (WHO, 2020). Two vaccines (Rotavin-M1®, POLYVAC, Hanoi, Vietnam and Lanzhou lamb rotavirus, Lanzhou Institute of Biological Products, Lanzhou, China) have been approved for national use in Vietnam and China, respectively (Fu *et al*., 2010; Anh *et al*., 2012). Human neonatal RVA vaccine (RV3-BB) and bovine human reassortant RVA vaccine candidates, as well as neonatal and non-replicating injectable vaccines, are in the pipeline (Kirkwood and Steele, 2018). South Africa was the first African country to adopt the monovalent RVA vaccine (Rotarix®) in September 2009, into its Expanded Program on Immunization (EPI) (WHO, 2009) which culminated in a 77% reduction in RVA disease during the first year that the vaccine was introduced (Madhi *et al*., 2010; Madhi *et al*., 2012).

Rotaviruses belong to the *Reoviridae* family. The rotavirus genome is composed of 11 segments of double-stranded RNA (dsRNA) encapsulated in a three-layered protein capsid. Six structural proteins (VP1 to VP4, VP6 and VP7) and five or sometimes six, non-structural proteins (NSP1 to NSP5/NSP6) that encode the rotavirus genome (Estes and Greenberg, 2013). The outer capsid proteins, VP7 and VP4, which act as neutralizing agents, are universally applied in the binary classification of rotavirus strains into G and P types, respectively (Estes and Greenberg, 2013). The contemporary classification of RVA strains is based on whole-genome composition underpinned by the nucleotide homology cutoff values that have been determined for the open reading frame (ORF) of each gene segment (Maes *et al*., 2009; Matthijnssens *et al*., 2011). The number of currently described genotypes are 36 G (VP7), 51 P (VP4), 26 I (VP6), 22 R (VP1), 20 C (VP2), 20 M (VP3), 31 A (NSP1), 22 N (NSP2), 22T (NSP3), 27 E (NSP4), and 22 H (NSP5) (http://rega.kuleuven.be/ cev/viralmetagenomics/virus-classification).

The globally predominant RVA genotypes are G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8] (Doro *et al*., 2014). However, RVA strains variability by region is well documented (Banyai *et al*., 2011). In Africa, RVA genotypes such as G1P[6], G8P[4], G8P[6], G8P[8] and G9P[6] are substantially prevalent but uncommon elsewhere (Banyai et al., 2011, Doro *et al*., 2014; Mwenda *et al*., 2010; Seheri *et al*., 2018). Additionally, the G3P[8] and G4P[8] genotypes have been on the decline in Africa and have not been detected in many African countries for almost a decade, albeit an impromptu emergence of equine-like G3P[6] and G3P[8] in Botswana, Eswatini and Malawi (Seheri *et al*., 2019). RVA are classified further into three genogroups: Wa-like which bears a genotype 1 constellation (I1-R1-C1-M1-A1-N1-T1-E1-H1), DS-1-like which bears genotype 2 constellation (I2-R2-C2-M2-A2-N2-T2-E2-H2) and a relatively minor AU-1-like characterized by genotype 3 constellation (I3-R3-C3-M3-A3-N3-T3-E3-H3) (Matthijnssens and Van Ranst, 2012). Typically, G1P[8], G3P[8], G4P[8], G9P[8] and G12P[8] RVA have a Wa-like genotype constellation, whereas G2P[4], G8P[4] and G8P[6] strains usually have a DS-1-like genotype constellation (Matthijnssens and Van Ranst, 2012). G1P[8] is the world's most prevalent genotype accountable for an estimated 50% of RVA infections (Do *et al*., 2016). The vast antigenic and genetic

heterogeneity of G1P[8] strains contributes to the persistent recurrence of the VP4 and VP7 protein variants and the epidemiological fitness of some of these variants might be accountable for their global prevalence (Santos *et al*., 2019).

The segmented RNA genome of RVs facilitates reassortment and recombination events and the error-prone RNA-dependent RNA polymerase promotes high mutation rates (Estes and Greenberg, 2013). These evolutionary mechanisms lead to the emergence of novel strains and distinct lineages (Kirkwood, 2010). Intergenogroup reassortment of G1P[8] gene segments has been reported in Africa, Asia and the Americas (Fujii *et al*., 2014; Kuzuya *et al*., 2014; Yamamoto *et al*., 2014; Jere *et al*., 2018; Luchs *et al*., 2019). These atypical G1P[8] strains were first reported in Okayama Prefecture, Japan during 2012-2013 post-RVA vaccine surveillance of acute gastroenteritis ,and then in other prefectures, including Aichi, Akita, Kyoto and Osaka (Fujii *et al*., 2014; Kuzuya *et al*., 2014; Yamamoto *et al*., 2014). Subsequent incidences were then reported during 2013 post-RVA vaccine surveillance in Phetchabun and Sukhothai provinces in Thailand (Komoto *et al*., 2015; Komoto *et al*., 2016) and in 2012-2013 during the pre-RVA vaccine period in Hanoi, Vietnam (Nakagomi *et al*., 2017). Although unpublished, sequence data of G1P[8] DS-1-like sequence strains isolated during pre-vaccine period between August-November 2012 in Palawan, Southwestern region of Philippines, have been deposited in the GenBank database. Recently, for the first time in the Americas, G1P[8] DS-1-like strains were reported in 2013 during post-RVA vaccination period from the states of Sao Paulo and Goias in Brazil (Luchs *et al*., 2019). In Africa, Jere and colleagues reported the emergence of atypical G1P[8] strains during the post-RVA vaccination period in Blantyre, Malawi (Jere *et al*., 2018). It is not definitively resolved whether these atypical G1P[8] strains are widespread. In addition, there is a paucity of information on whole-genome sequences of G1P[8] strains post-vaccine era with only a few countries performing full-genome characterization of the strain (Rahman *et al*., 2010; Arora and Chitamber, 2011; Banyai *et al*., 2011; Shintani *et al*., 2012; Magagula *et al*., 2015; Zeller *et al*., 2017; Santos *et al*., 2019). The AEVGI is conducting whole-genome characterization of country-specific pre- and post-vaccine RVA strains in Africa and has identified, for the first time in South Africa, atypical G1P[8] strains that were circulating before vaccine introduction. This study aimed

to determine the genetic relationship and evolutionary origin of these pre-vaccine atypical G1P[8] RVA strains.

4. 2 Materials and Methods

4.2.1 Ethics approval

The study was approved under ethics number, UFS-HSD2018/0510/3107, by the Health Sciences Research Ethics Committee (HSREC) of the UFS, Bloemfontein, South Africa. The patient identities and demographics were de-linked from their unique laboratory identifiers to ensure confidentiality.

4.2.2 Sample collection

Rotavirus positive stool samples from children under five years of age treated for gastroenteritis at Dr George Mukhari Hospital, Pretoria North, South Africa and conventionally genotyped as G1P[8] were sourced from archival storage (2002 to 2017) of DPRU, a WHO-RRL in Pretoria, South Africa. The two stool samples that were later genotyped as DS-1-like G1P[8] strains had been collected from a 6-month old female and a 12-month old male on 15th and 16th May 2008, respectively, from Soshanguve, Pretoria.

4.2.3 Extraction and purification of double-stranded RNA

The extraction of RVA ds-RNA was conducted by utilizing a previously described method albeit with modifications (Potgieter *et al*., 2009). Briefly, a pea size (~100mg) sample of stool was added to 200 µl of Phosphate-Buffered Saline (PBS) solution, pH 7.2 (Sigma-Aldrich®, St Louis, Missouri, US). The solution was mixed by pulse- vortexing for five seconds. A 1 ml volume of TRI-Reagent®-LS (Molecular Research Center, Inc, Cincinnati, OH, USA) was added and let to stand for five minutes. Phase separation was achieved by addition of 270µl of chloroform (Sigma-Aldrich®, St Louis, Missouri, US). Afterward, centrifugation for 18,000 x *g* was performed for 20 minutes at 4^{o}C in a temperature-controlled microcentrifuge (Eppendorf microcentrifuge 5427R, Hamburg, Germany). A volume of 1ml isopropanol (Sigma-Aldrich®, St Louis, Missouri, US) was added to the supernatant and centrifugation was performed at 18,000 x *g* for 30 minutes at room temperature. The supernatant was poured off and the tubes were let to dry for 10 minutes after

which 95µl of Elution Buffer (EB) from the MinElute Gel extraction kit (Qiagen, Hilden, Germany) was added. A 30µl volume of 8M $LiCl_2$ (Sigma, St. Louis, MO) was added and the solution was precipitated for 16 hours at 4°C in a water bath in a Tupperware box. The MinElute gel extraction kit (Qiagen, Hilden, Germany) was used to purify the extracted RNA according to manufacturer's instructions and 1% 0.5 X TBE agarose gel stained with Pronasafe (Condalab, UK) electrophoresis was used to verify the integrity and enrichment of dsRNA which was visualized on a G:Box Syngene UV transilluminator (Syngene, Cambridge, UK).

4.2.4 Synthesis and purification of complementary DNA(cDNA)

The Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fischer Scientific, Waltham, MA) was utilized to synthesize cDNA from the extracted viral RNA. Briefly, denaturation at 95°C for 5 minutes of the extracted RNA was performed followed by addition of 1µl of 100µM Random Hexamer primer. Incubation was performed in a thermocyler at 65°C for five minutes. The First-Strand Reaction mix (5µl) and First Strand Enzyme Mix (1µl) was added and the solution was incubated at 25°C for 10 minutes followed by 2 hours at 50°C and then the reaction was terminated by heating at 85°C for 5 minutes. A volume of 55µl of nuclease-free water, 20µl of 5X Second Strand Reaction Mix, 5µl of Second Strand Reaction Mix was then added. The solution was then incubated at 16°C for 60 minutes after which the reaction was stopped by adding 6µl 0.5M EDTA. A volume of 10µl RNAse I was then added and the synthesized cDNA was incubated for five minutes at room temperature. Subsequently, the MSB® Spin PCRapace (Stratec) Purification Kit was used to purify the synthesized cDNA.

4.2.5 DNA library preparation and whole-genome sequencing

The Nextera® XT DNA Library Preparation Kit (Illumina, San Diego, California, US) was utilized to prepare DNA libraries by following manufacturer's instructions. Briefly, the genomic DNA was tagmented by using the Nextera® transposome enzyme and the tagmented DNA was subsequently amplified using a limited-cycle PCR program. The DNA libraries were cleaned-up using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA) and 80% freshly prepared ethanol. The quantity of the DNA was determined using Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and the quality of the libraries and fragment sizes was assessed

using Agilent 2100 BioAnalyzer® (Agilent Technologies, Waldbronn, Germany) by following the manufacturer's specified protocol. The Illumina MiSeq® sequencer (Illumina, San Diego, CA, USA) was utilized to perform paired-end nucleotide sequencing (301 x2) for 600 cycles by using a MiSeq Reagent Kit v3 at the University of the Free State-Next Generation Sequencing (UFS-NGS) Unit, Bloemfontein, South Africa.

4.2.6 Genome assembly

Geneious Prime® software, version 2019.1.1 (Biomatters, https:www.geneious.com/; Kearse *et al*., 2012) was used for genome assembly. Briefly, for use with the reference mapping tools integrated in Geneious Prime version 2019.1.1, the default medium sensitivity parameter was selected to generate contigs from the FASTQ files data generated by the Illumina MiSeq® instrument. Complementary RVA genome assembly was also performed using an in-house genome assembly pipeline and CLC Genomics Workbench 12 (https://www.qiagenbioinformatics.com/).

4.2.7 Determination of rotavirus whole-genotype constellations

The genotype of each gene segment was determined using Rota C, v 2.0 (Maes *et al*., 2009), an online server for genotyping RVA strains. This was used to generate the full genotype constellations for each RVA strain.

4.2.8 Phylogenetic analyses

Complete sequences for each gene segment were aligned and sequence comparisons performed as described previously (Ward *et al*., 2016; Esona *et al*., 2017; Esona *et al*., 2018). Multiple sequence alignments were implemented utilizing the MUSCLE package in Molecular Evolutionary Genetics Analysis (MEGA) 6 software (Tamura *et al*., 2013; http://www.megasoftware.net/). Upon alignment, the DNA Model Test program in MEGA 6 was used to determine the evolutionary model that best fits each gene sequence datasets. The models identified as best fitting with the sequence data for the indicated genes using the Corrected Akaike Information Criterion (AICc) were as follows: GTR+G+I (VP7, VP4, VP6,VP1,VP2,VP3,NSP1,NSP2,NSP3) and HKY+G+I (NSP4 and NSP5). These models were utilized in Maximum-likelihood trees' construction using MEGA 6 with 1000 bootstrap replicates to estimate branch support. Genetic

distance matrices were prepared using the *p*-distance algorithm of MEGA 6 software (Tamura *et al*., 2013). In addition to the two whole-genome sequences of the strains in this study, other cognate sequences were acquired from GenBank (http://www.ncbi.nlm.gov/genbank). Further phylogenetic analysis by geographical regions Africa (Eastern Africa, Southern Africa and West Africa), Asia, Americas, Europe and Oceania was also performed. mVISTA software was used to visualize the comparative sequence similarities of concatenated whole-genome of genetically related strains (Frazer *et al*., 2004).

4.3 Results

4.3.1 Nucleotide sequencing

Illumina® MiSeq sequencing yielded 14.7 x10^5 reads (379 bp fragment size) and 11.3 x10^5 reads (364 bp fragment size) for strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/G1P[8], respectively. All the sequences had a phred score of Q≥30 (99.9% base calling accuracy).

4.3.2 Full-genome constellation analysis

Whole-gene sequences of the 11 genes of strains, RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/G1P[8], were determined and their genotype constellations were revealed as G1-P[8]-I2-R2-C2-M2-A2-N2-T2-E2-H2 (Table 4.1). The sizes of full-length segments 1 to 11 and their respective open reading frames (ORFs) for the two study strains were determined (Table 4.1). The ORF sequences for all the 11 genes of these two South African atypical G1P[8] strains were deposited in GenBank under accession numbers MT163245-MT163266.

Table 4.1. Whole genotype constellations of the South African DS-1-like G1P[8] RVA strains.

Gene segment	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
Base pair size for full length sequences	1062	2359	1355	3302	2684	2591	1566	1059	1066	751	810
Base pair size for the complete study strain ORF	978	2325	1191	3264	2637	2505	1458	951	939	525	600
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8]	G1	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8]	G1	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2

Color codes indicate genogroup attribution. Green color represents the genotype associated with the Wa-like genogroup while red color represents the genotype belonging to the DS-1-like genogroup. The nomenclature of the rotavirus strains indicates the rotavirus group, species where the strain was isolated, name of the country where the strain was originally isolated, the common name, year of isolation and the genotypes for genome segment 4 and 9 as proposed by the Rotavirus Classification Working Group (RCWG) (Matthijnssens *et al*., 2011)

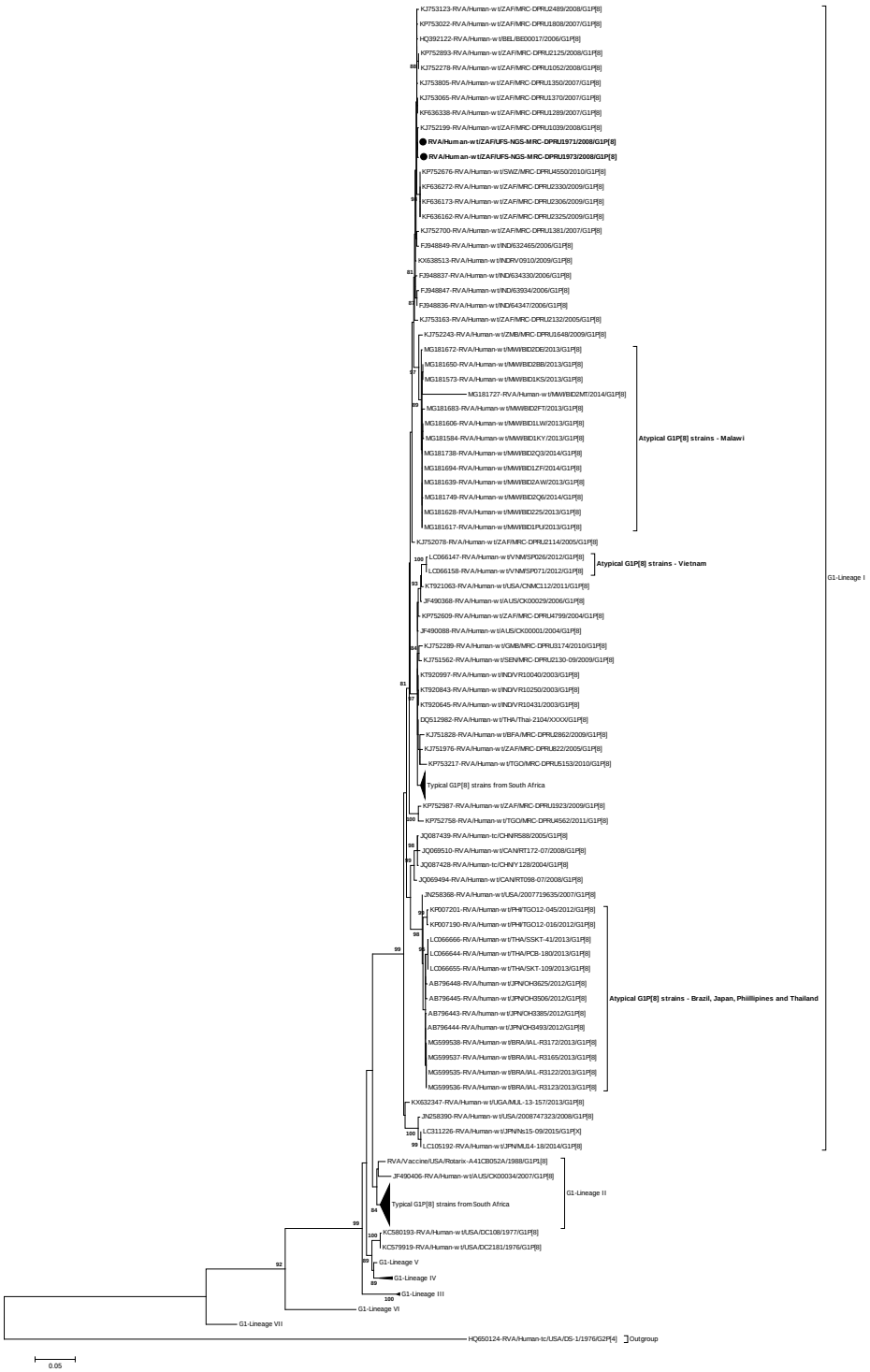
4.3.3 Sequence and phylogenetic analysis

4.3.3.1 Phylogenetic analysis of VP7

Phylogenetically, the diversity of the VP7 G1 genes has been established through seven known lineages (I-VII) (Arista *et al*., 2006) (Figure 4.1). The VP7 genes of the atypical G1P[8] study strains, RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] clustered in genetic lineage I, which consisted of a global collection of G1 strains that circulated from 2002 to 2015 (Figure 4.1). In this lineage I, the two G1 study strains clustered closely together and shared almost absolute gene identities amongst themselves; nt (aa) 99.9% (100%) (Figure 4.1; Supplementary data available at https://www.mdpi.com/2076-0817/9/5/391/s1). Analysis of the G1 study strains with locally circulating South African strains retrieved from the GenBank identified the highest sequence identities of 99.8-99.9% (100%) with strain RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8] and clustered closely with this strain which was isolated the same year, 2008, as the two study strains (Figure 4.1). However, within the same lineage, the VP7 genes of the two G1 study strains from South Africa clustered distinctly away from the atypical G1P[8] strains reported in Brazil, Japan, Malawi, Philippines, Thailand and Vietnam (Fujii *et al*., 2014; Kuzuya *et al*., 2014;

Yamamoto *et al*., 2014; Jere *et al*., 2018; Luchs *et al*., 2019) and displayed overall nt (aa) similarities that ranged from 87.0-98.5% (87.1-98.8%). Specifically, the nt (aa) similarities ranged from 96.9-97.0% (98.5%), 96.9-97.1% (98.5-98.8%), 87.0-98.5% (87.1-98.8%), 96.5-96.8% (98.2-98.5%), 96.9-97.0% (98.8%) and 97.0-97.1% (96.9%) to the post-vaccination G1 atypical strains reported in Brazil, Japan, Philippines, Malawi, Thailand and Vietnam, respectively.

When the two South African G1 study strains were compared to the typical G1 strains selected globally, they displayed the highest nt (aa) similarities of 99.7-99.8% (100%) with a European strain, RVA/Human-wt/BEL/BEL00017/2006/G1P[8]. The nt (aa) similarities comparison to representative strains from Africa (Eastern Africa, Southern Africa and West Africa), America, Asia, Europe and Oceania ranged from 93.9-97.6% (94.2-98.2%), 97.1-99.5% (98.2-100%), 97.1-97.9% (96.9-98.2%), 93.1.5-97.6% (94.8%-98.5%), 96.4-99.6% (97.8-99.7%), 99.7-99.8% (100%) and 93.7-98.4% (94.5-98.8%), respectively. In addition, comparison of the VP7 genes of the two study strains to cognate gene sequence of the Rotarix® and RotaTeq® vaccine strains, displayed nt (aa) identities that ranged from 94.2-94.3% (95.7%) and 91.0-91.1% (93.2%), respectively.



Figures 4.1: VP7 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Japan, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

4.3.3.1.1 Analysis of the VP7 neutralization epitopes

The VP7 genes contain three established neutralization epitopes: 7-1a, 7-1b and 7-2. Twenty-nine amino acids (14 residues in 7-1a, 6 residues in 7-1b and 9 residues in 7-2) define the three VP7 antigenic epitopes (Aoki *et al*., 2009). The VP7 neutralization epitope sites of the two South African study strains were aligned and mapped against cognate neutralization sites of the two RVA vaccines, Rotarix® and RotaTeq®. Four aa differences (N94S, S123N, K291R and M217T) in the VP7 genes of the two South African study strains were identified relative to Rotarix® VP7 neutralization sites while five aa acid differences (D97E, S123N, K291R, S147N and M217T) were identified with comparison to RotaTeq® G1 antigenic sites (Figure 4.2). Antigenically, similar aa residues in the VP7 epitopes of the study strains were observed in the corresponding VP7 epitopes of the multiple atypical G1P[8] strains (Figure 4.2). The VP7 epitopes of the two South African G1 strains were compared with those of globally selected lineage I G1 strains. The analysis showed ten aa differences (T91N, S94N, D100N, D100E, N123S, R291K, T242A, N147D, L148F and T217M) (Figure 4.2).

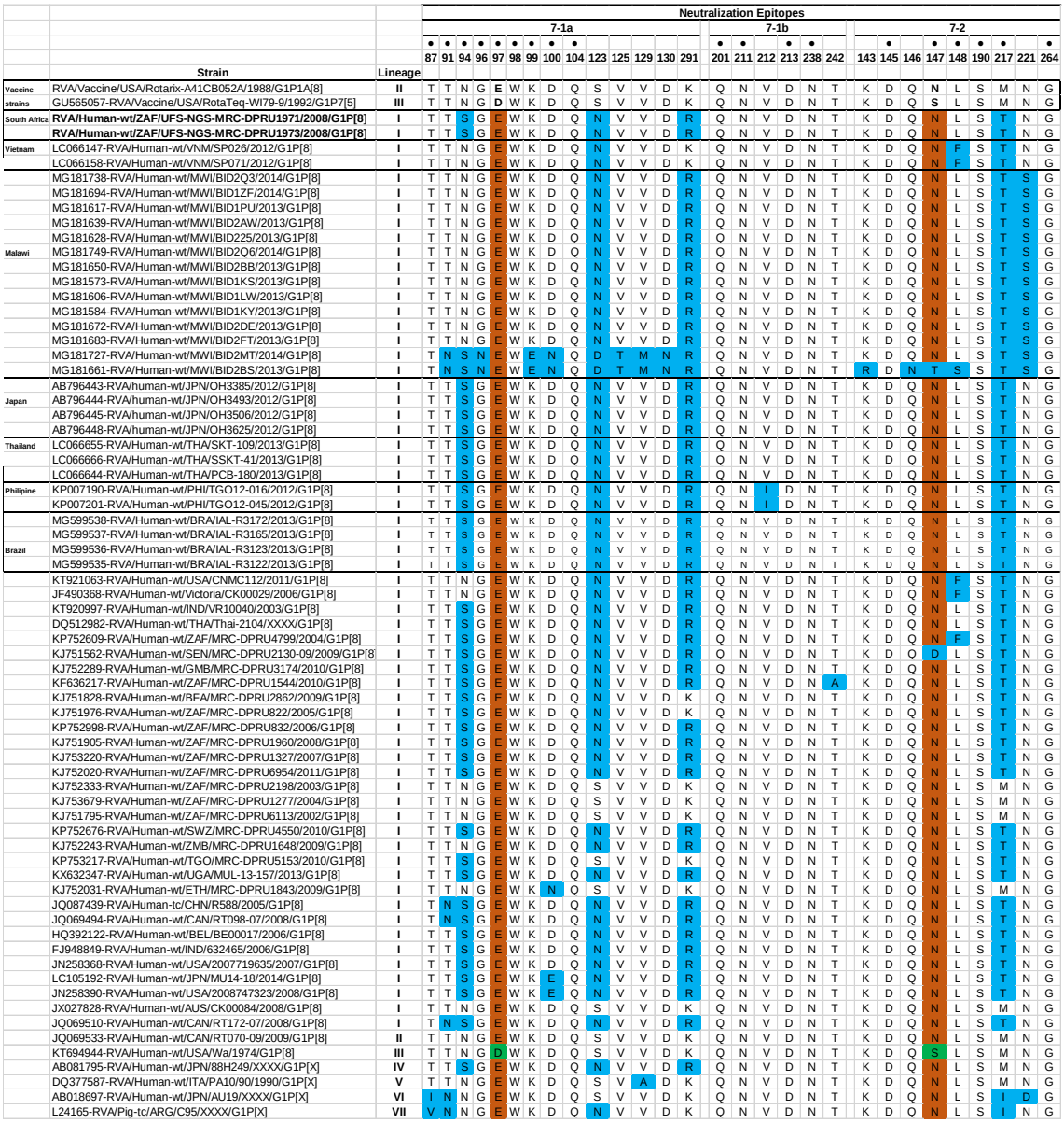
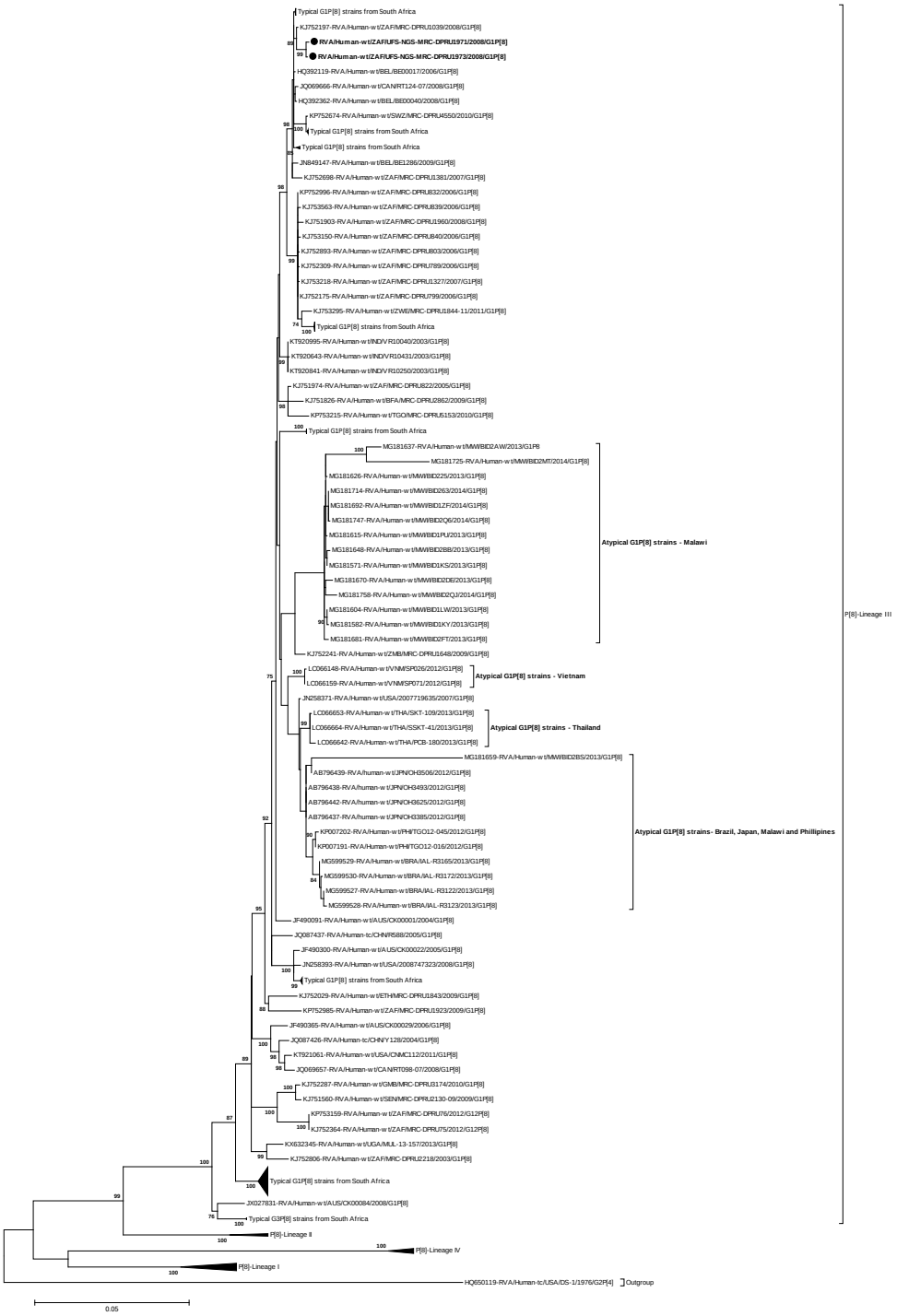


Figure 4.2: Alignment of antigenic residues in VP7 between vaccine strains and wild type strains.

Antigenic residues are divided in three epitopes (7-1a, 7-1b, and 7-2). Amino acids that differ between Rotarix® and RotaTeq® are indicated in boldface. Sky blue colored residues are residues that are different from both Rotarix® and RotaTeq®, green colored residues are different from Rotarix® and brown colored residues are different from RotaTeq®. Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. Atypical G1P[8] and countries of detection are indicated on the left side of the figure. South Africa atypical G1P[8] strains are in boldface characters.

4.3.3.2 Phylogenetic analysis of VP4

The VP4 genes of the two atypical G1P[8] study strains were phylogenetically compared to the four established lineages (I-IV) of the P[8] genotypes (Le *et al*., 2010) (Figure 4.3). The P[8] genes of the South African strains, RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8], clustered in lineage III which consisted of a global collection of P[8] strains that circulated from 2002 to 2014 (Figure 4.3). Within the P[8]-lineage-III, the two atypical G1P[8] study strains clustered closely together and shared nt (aa) identities of 99.9% (99.7%) amongst themselves (Figure 4.3). Homology analysis of the P[8] sequences of the two South African strains with sequences of South African strains retrieved from the GenBank demonstrated the highest nt (aa) sequence identities of 99.6% (99.1-99.4%) with strain RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8] (Figure 4.3). However, within the same lineage, the VP4 genes of the two atypical strains from South Africa segregated distinctly away from the atypical strains that have been detected in Brazil, Japan, Malawi, Philippines, Thailand and Vietnam. They exhibited overall nt (aa) similarities that ranged from 95.2-98.0% (95.1-98.6%). Specifically, the nt (aa) similarities ranged from 97.6-97.8% (98.1-98.6), 98.2-98.5% (98.5%), 95.2-98.0% (95.1-98.2%), 97.7-97.8% (98.1-98.5%), 97.8-98.0% (97.7-98.2%) and 98.1-98.2% (97.9-98.3%) to the post-vaccination atypical strains reported in Brazil, Japan, Malawi, Philippines, Thailand and Vietnam, respectively. A comparison of the South African P[8] study strains characterized in this study with a global collection of P[8] strains showed their closeness and the study strains shared the highest nt (aa) similarity of 99.5% (99.1-99.4%) to a Belgian strain, RVA/Human-wt/BEL/BEL00017/2006/G1P[8]. Overall, the nt (aa) similarities in comparison to representative strains from Africa (Eastern Africa, Southern Africa, Western Africa), America, Asia, Europe and Oceania ranged from 97.4-97.8% (97.8-98.2%), 86.7-99.1% (91.1-99.2%), 86.7-99.1% (91.1-99.2%), 86.6-99.4% (91.0-99.2%), 86.5.-98.9% (91.1-98.7%), 86.8-99.5% (91.2-99.5%) and 86.5-98.5% (91.0%-98.6%), respectively. In addition, the comparison of the atypical VP4 genes to the P[8] genes of the Rotarix® and RotaTeq® vaccine strains, displayed nt (aa) identities that ranged from 90.3-90.4% (93.9-94.2%) and 92.3% (95.2%), respectively.



Figures. 4.3: VP4 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Japan, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

4.3.3.2.1 Analysis of the VP4 neutralization epitopes

The VP4 spike protein is cleaved by trypsin into two distinct structural proteins, VP5* and VP8* (Estes and Greenberg, 2013). Analysis of the two South African study strains' VP4 sequences showed a conserved trypsin cleavage site (arginine) at positions; 230, 240 and 581 (Ciarlet *et al*., 2002). Furthermore, the neutralization epitopes in the VP8* and VP5* regions were analyzed. The VP8* region has four (8-1 to 8-4) neutralization epitopes while VP5* has five (5-1 to 5-5) (Figure 4.4) (Zeller *et al*., 2012). Comparison of the two South African P[8] strains relative to the Rotarix® and RotaTeq® P[8] sequences displayed 32 and 35 identical amino acid residues, respectively, spanning the VP4 antigenic epitopes (Figure 4.4). Amino acid differences between the two P[8] study strains and P[8] component of vaccine strains were only identified in 8-1, 8-2 and 8-3 VP8* epitopes. Five aa differences (E150D, N195G, S125N, S131R and N135D) were identified in the study strains in relation to Rotarix® P[8] strain while two aa differences (E150D and D195G) were identified relative to P[8] strain of RotaTeq® (Figure 4.4). Analysis with VP4 epitopes of other atypical G1P[8] strains identified similar aa residues with the exception of position 113 in the 8-3 epitope whereby asparagine was observed in the study strains while other atypical strains had either an aspartate or serine at this position (Figure 4.4). Further analysis of the study strain's VP4 neutralization epitopes with corresponding VP4 neutralization epitopes of globally selected P[8]-lineage-III strains identified two aa differences (S146G and N113D) (Figure 4.4).

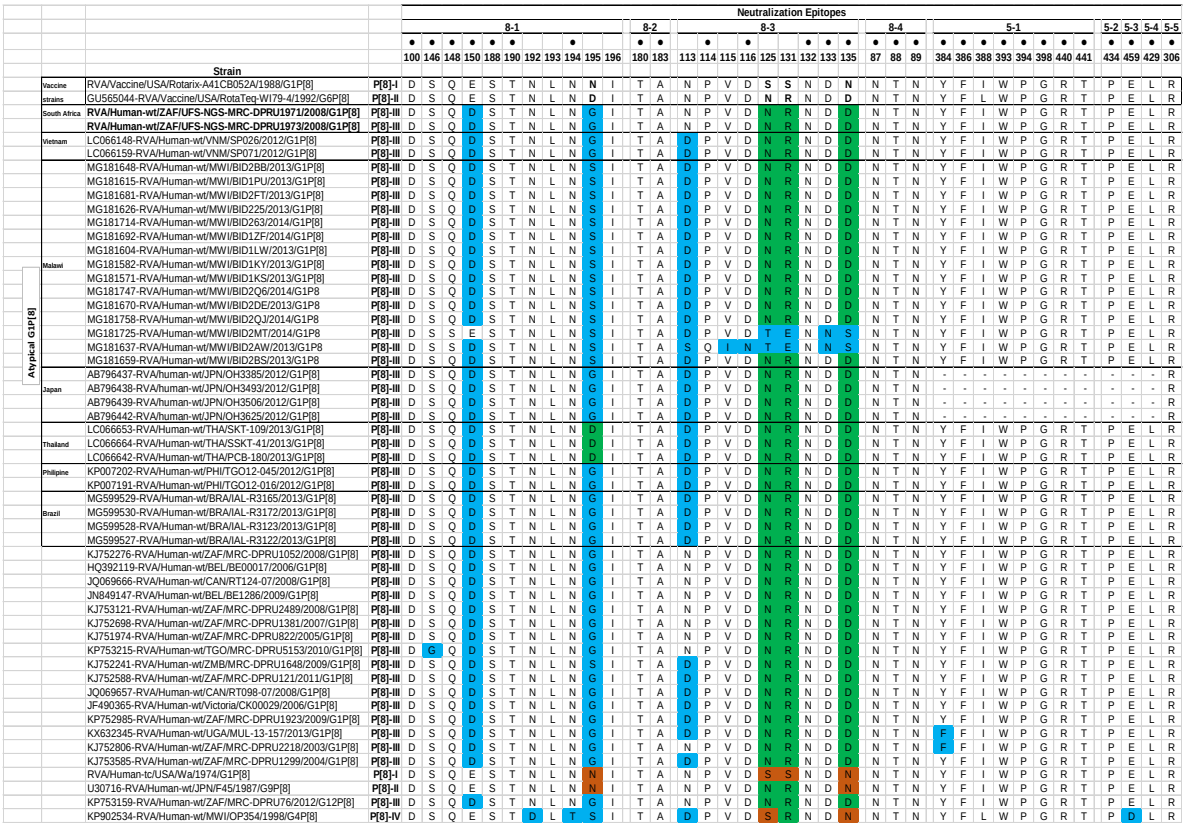


Figure 4.4: Alignment of antigenic residues in VP4 between vaccine and and wild type strains. Antigenic residues are divided in four antigenic epitopes in VP8* (8-1, 8-2, 8-3, and 8-4) and five antigenic epitopes in VP5* (5-1, 5-2, 5-3, 5-4, and 5-5). Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. Amino acids that differ between Rotarix® and RotaTeq® are indicated in boldface. Green colored residues are residues that are different from Rotarix®, Brown colored residues are different from RotaTeq® and residues colored in sky blue are different from both Rotarix® and RotaTeq®. Dashes (-) indicates no amino acid sequence.

4.3.3.3 Phylogenetic analysis of VP1-VP3 and VP6

The evolutionary relationship of the VP1-VP3 and VP6 genes of the two South African study strains with a selection of global RVA strains was performed. The VP1-VP3 and VP6 genes of the two South African study strains clustered closely and displayed nearly absolute gene identities (≥ 99.9%) amongst each other (Figures 4.5-4.8). The VP1-VP3 and VP6 genes of the two South African study strains clustered closely in sub-lineage comprising mainly of locally circulating South African DS-1-like strains and were all found to cluster closely with cognate genes of strain RVA/Human-wt/ZAF/MRC-DPRU2344/2008/G2P[6] which was co-circulating in the population in the same year, 2008, as the study strains (Figures 4.5-4.8). The VP1-VP3 and VP6 genes of the two study strains were closely related to cognate genes of strain RVA/Human-wt/ZAF/MRC-

DPRU2344/2008/G2P[6] with nt (aa) identities ranging from ≥ 99.8-99.9% (≥99.9-100%) for VP1-VP3 and VP6 genes. Phylogenetic relationship of the VP1-VP3 and VP6 genes of the atypical study strains with the cognate genes of the atypical G1P[8] strains reported in Brazil, Japan, Philippines, Thailand, Vietnam and Malawi exhibited distinct clustering albeit belonging within the same lineage and displayed overall nt (aa) identities that ranged from ≥ 94.4-98.0% (≥ 97.1-100%) (Figures 4.5-4.8). When VP1-VP3 and VP6 genes of the study strains were compared with selected global strains, highest genetic similarities ranging from ≥ 99.4-100% were identified with cognate genes of G3P[6] strains; RVA/Human-RVA/Human-wt/CMR/ES293/2011/G3P[6], RVA/Human-wt/TGO/MRC-DPRU2206/2009/G3G9P[6], RVA/Human-wt/BEL/F01498/2009/G3P[6], and RVA/Human-wt/UGA/MUL-13-166/2013/G3P[6] for VP1-VP3 and VP6, respectively.

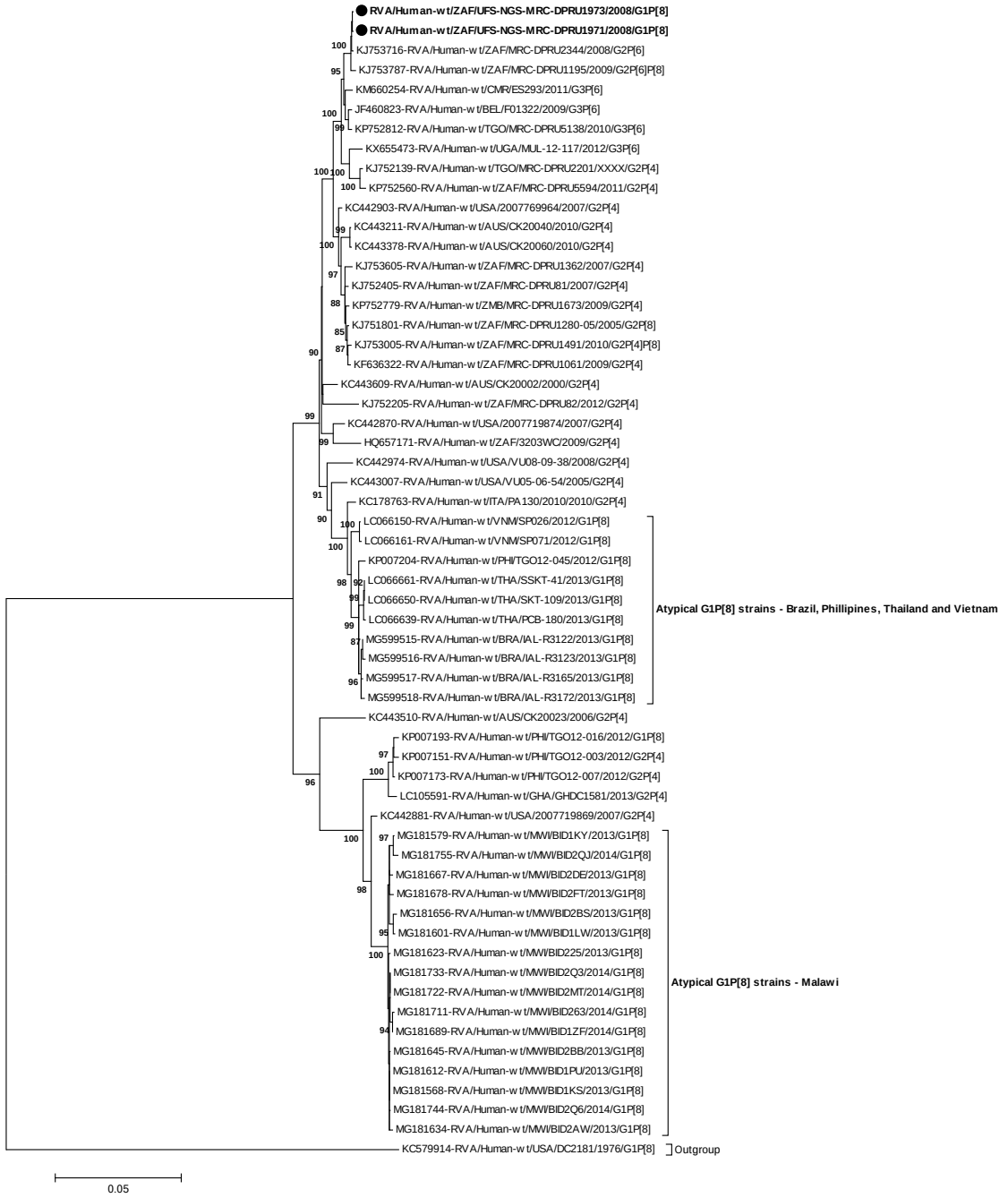


Figure 4.5: VP1 phylogenetic tree based on the full-length nucleotide sequences. Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

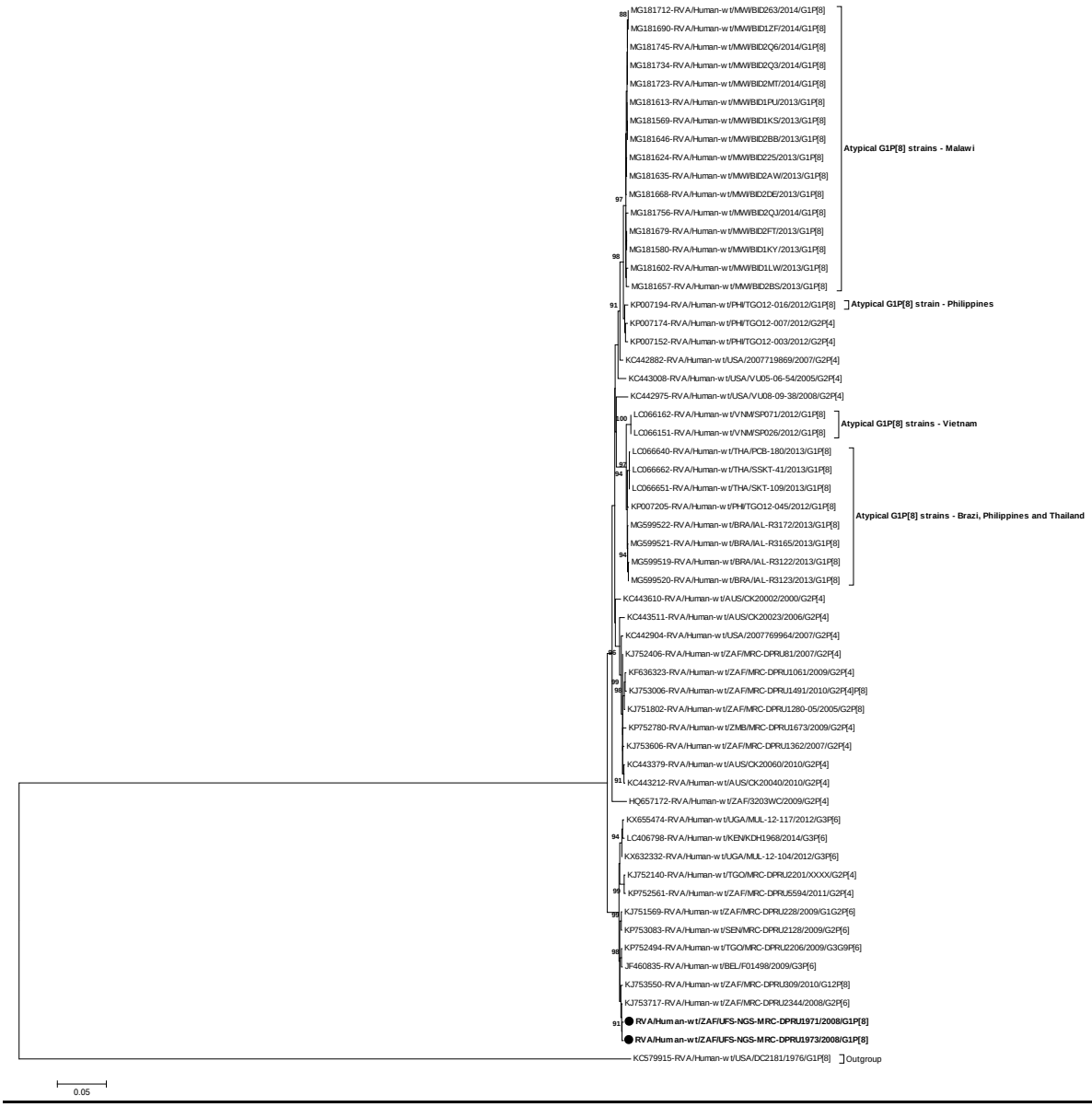


Figure 4.6: VP2 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

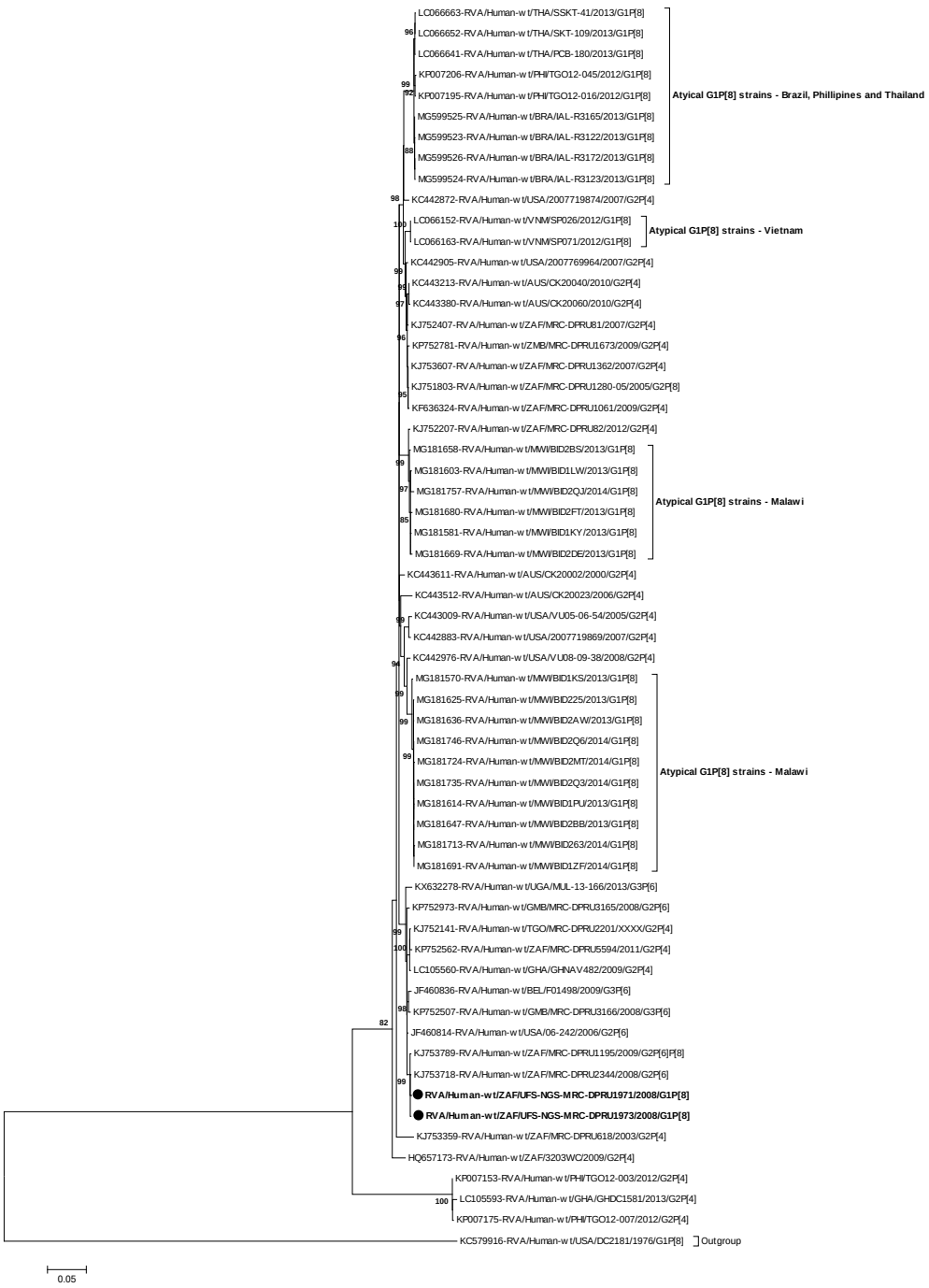


Figure 4.7: VP3 phylogenetic tree based on the full-length nucleotide sequences. Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

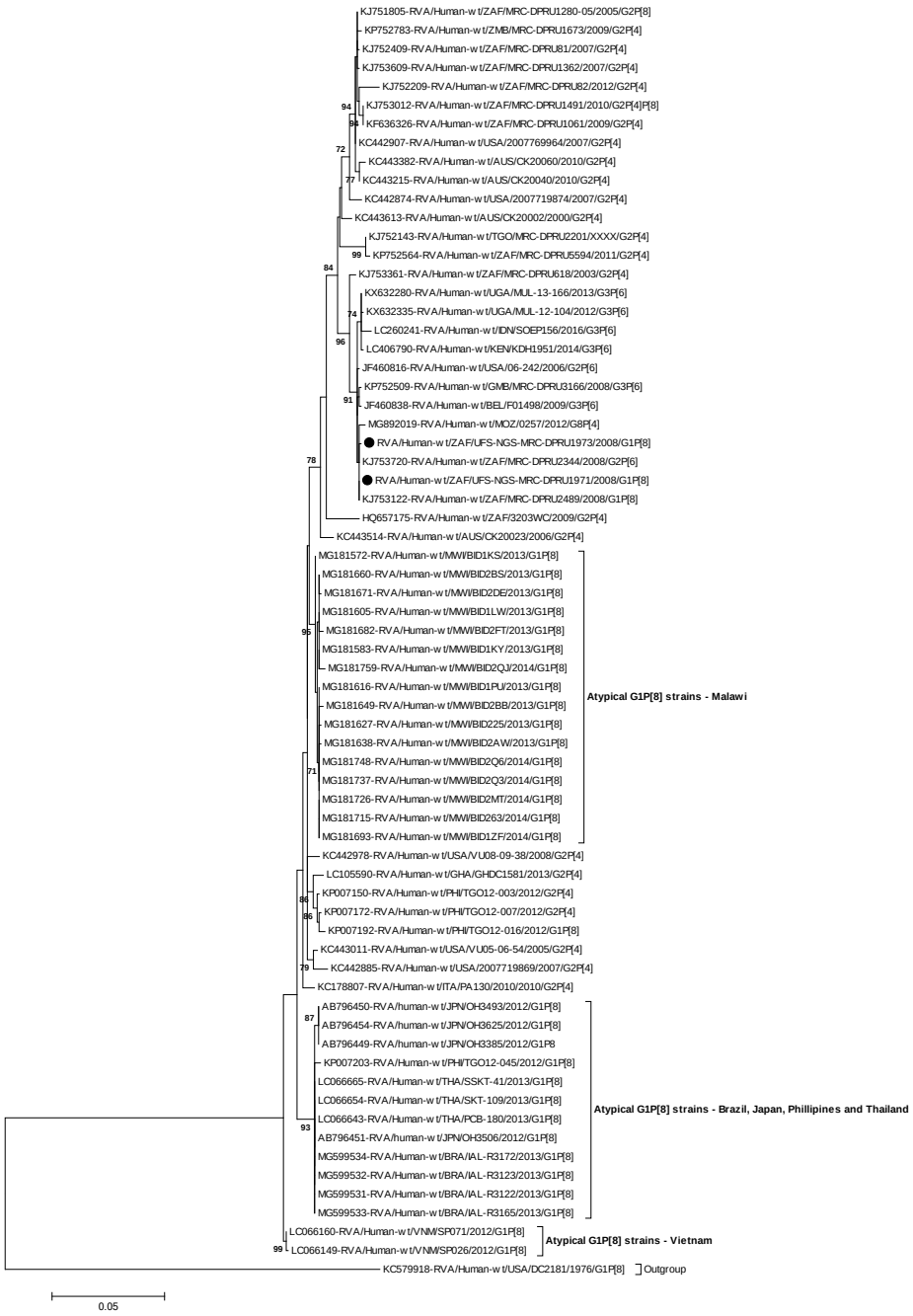


Figure 4.8: VP6 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Japan, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

4.3.3.4 Phlylogenetic analysis of NSP1-NSP5

The NSP1-NSP5 genes of the two South African study strains were highly identical amongst each other with nt (aa) identity value of ≥ 99.8%, and clustered closely (Figures 4.9-4.13). Close clustering with cognate genes of a locally circulating strain, RVA/Human-wt/ZAF/MRC-DPRU2344/2008/G2P[6], was observed for the all the NSP1-NSP5 genes and they shared the highest nt (aa) similarities that ranged from ≥ 99.5-100% (99.4-100%). In contrast, the NSP1-NSP5 genes of the atypical study strains grouped distinctly away from cognate genes of atypical strains reported in Brazil, Japan, Philippines, Thailand, Vietnam and Malawi and shared an overall nt (aa) similarities that ranged from ≥ 89.0-99.8% (≥ 94.3-100%) (Figures 4.9-4.13). Comparison of NSP1-NSP5 genes of the two South African study strains with corresponding selected reference strains collected globally demonstrated nt (aa) identities in the range of 98.6-100% (99.4-100%) with cognate A2, N2, T2, E2 and H2 genes of strains; RVA/Human-wt/BEL/F01498/2009/G3P[6], RVA/Human-wt/KEN/KDH1968/2014/G3P[6], RVA/Human-wt/GHA/GH018-08/2008/G8P[6], RVA/Human-wt/ZMB/MRC-DPRU1673/2009/G2P[4] and RVA/Human-wt/USA/2007769964/2007/G2P[4], respectively.

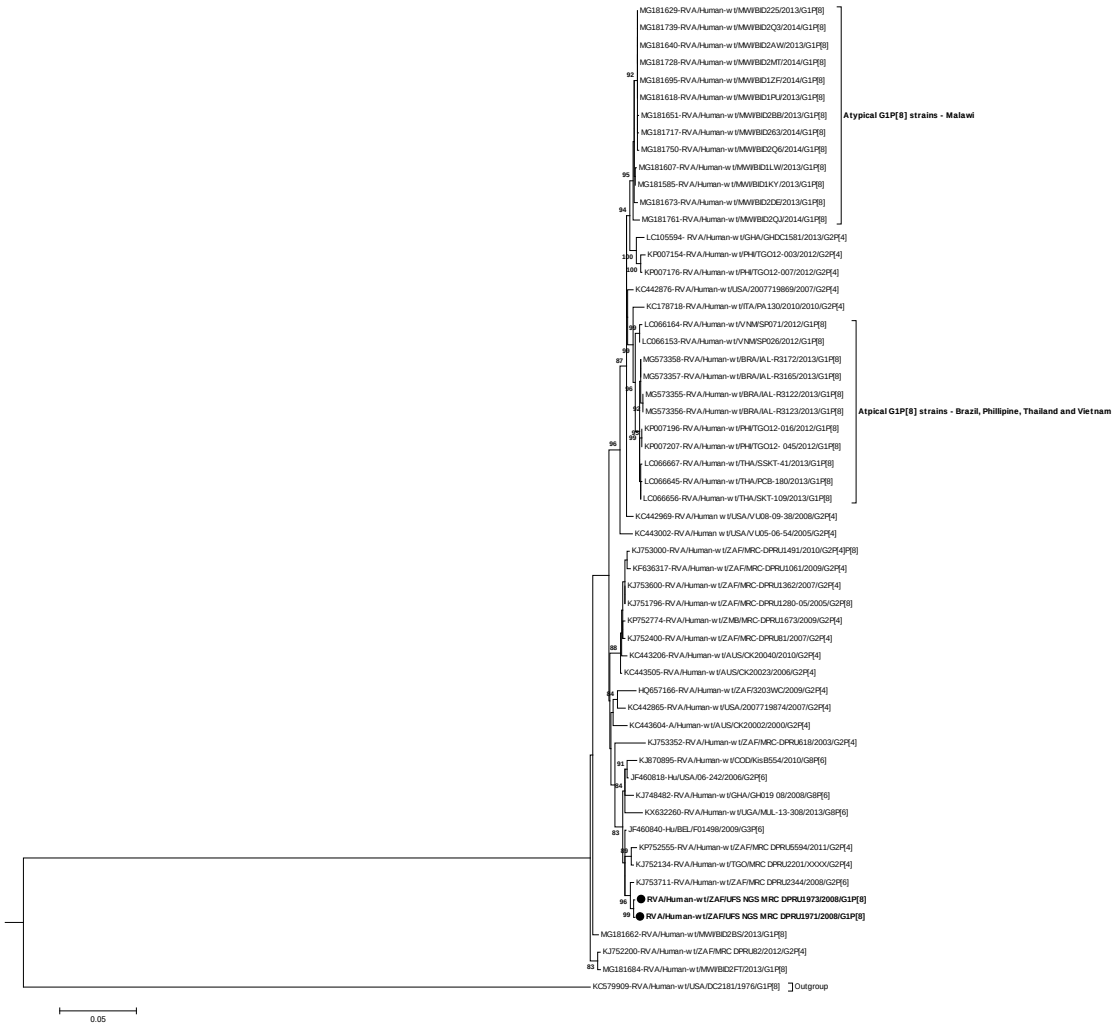


Figure 4.9: NSP1 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

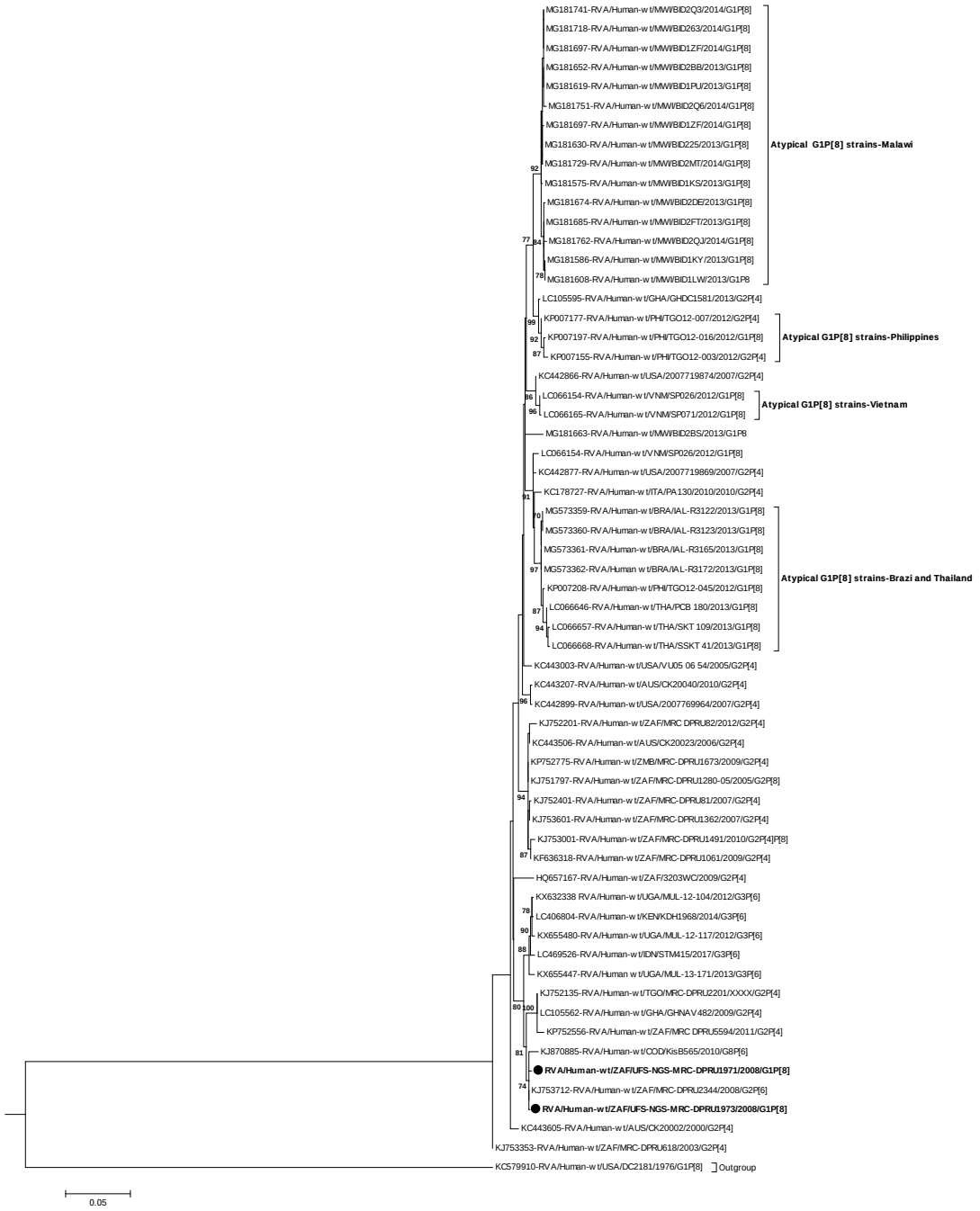


Figure 4.10: NSP2 phylogenetic tree based on the full-length nucleotide sequences. Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

Figure 4.11: NSP3 phylogenetic tree based on the full-length nucleotide sequences. Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

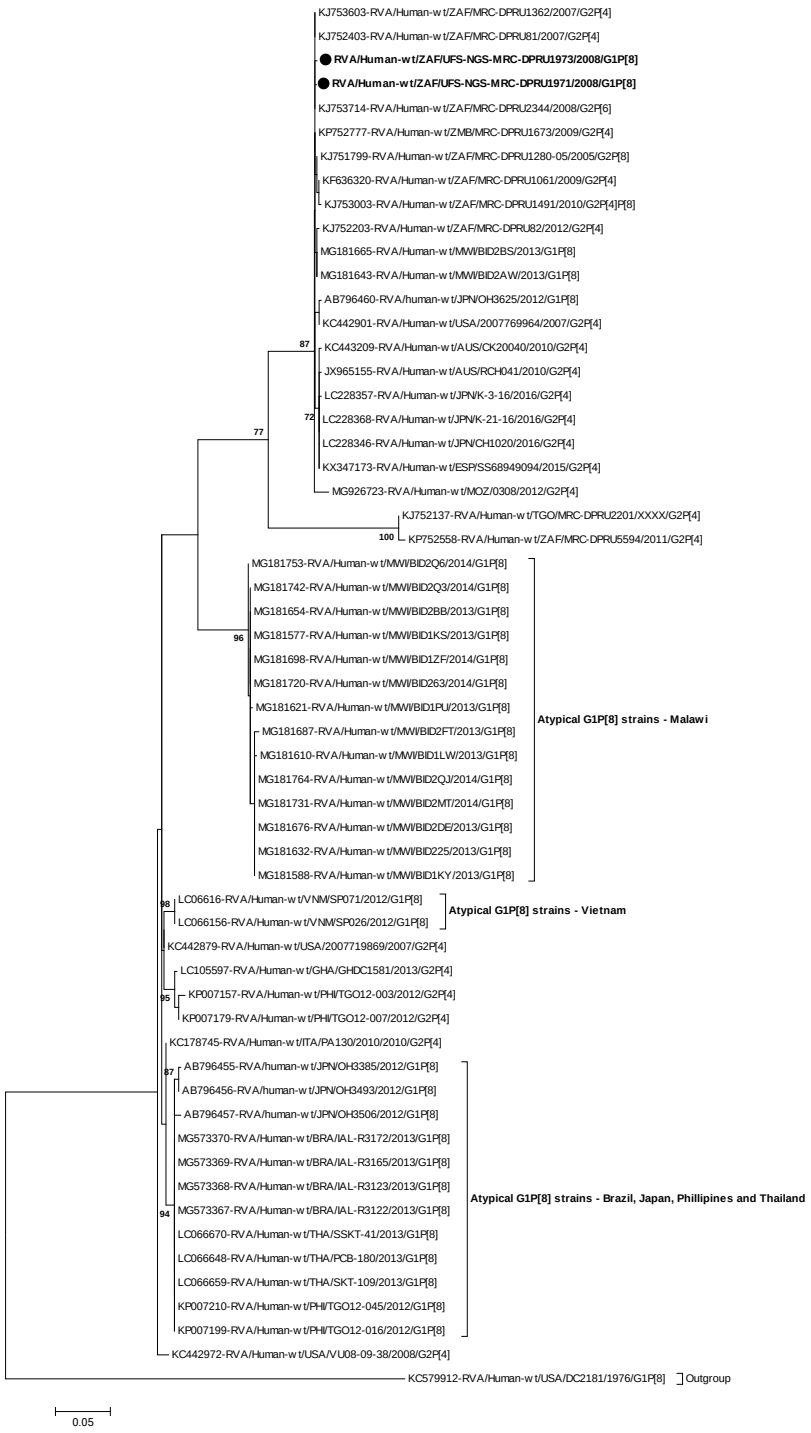


Figure 4.12: NSP4 phylogenetic tree based on the full-length nucleotide sequences.
Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Japan, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

Figure 4.13: NSP5 phylogenetic tree based on the full-length nucleotide sequences. Strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1491/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] are identified by the black filled circular dots (●). Unusual G1P[8] strains from Malawi, Thailand, Vietnam, Brazil and Philippine are indicated. Bootstrap values ≥70% are shown adjacent to each branch node. Each scale bar indicates the number of nucleotide substitutions per site.

4.3.4 Reassortment analysis

The concatenated genomes of strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] were compared with two South African strains RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8] and RVA/Human-wt/ZAF/MRC-DPRU2344/G2P[6] (Figure 4.14). The two atypical South African study strains shared a highly conserved backbone with all genes exhibiting > 99.8% nucleotide similarity. The VP7 and VP4 genes of the two atypical strains shared the highest genetic similarities to RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8]. However, the internal backbone genes were extremely diverse. The internal backbone genes of the atypical strains exhibited highest genetic similarity to RVA/Human-wt/ZAF/MRC-DPRU2344/G2P[6]. The results of this analysis suggests that the atypical G1P[8] strains were likely derived via reassortment events between contemporary, endemic, South African strains.

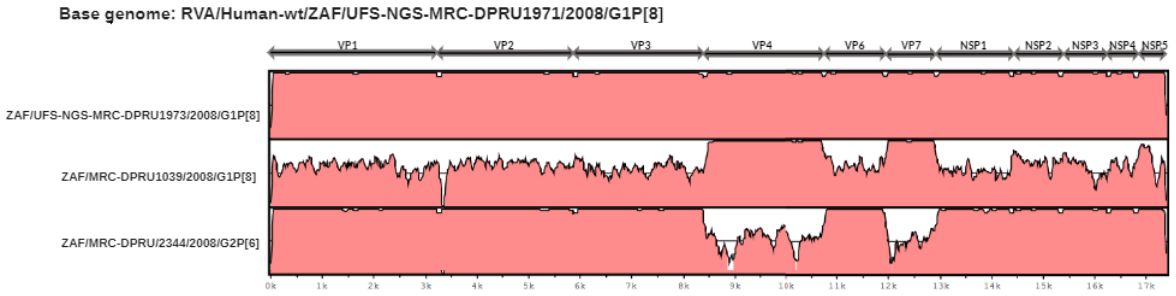


Figure 4.14. Reassortment analysis

Nucleotide sequence similarities of the concatenated genome of RVA//Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] were compared with South African strains RVA//Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8], RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8] and RVA/Human-wt/ZAF/MRC-DPRU2344/G2P[6]. The left axis displays the strains, rotavirus genome segment is included in the top scale. The bottom scale shows distance in kb.

4.4. Discussion

This study has described the first pre-vaccine era atypical reassortant G1P[8] strains, whose outer gene segments (VP7 and VP4) expressed a Wa-like genotype, whereas the backbone genes expressed a DS-1-like genotype constellation. Analysis of the whole-genome constellation showed that genetic reassortment mechanism generated the DS-1-like G1P[8] strains. Rotavirus reassortment events are mainly facilitated by the segmentation inherent in the rotavirus genome (Estes and Greenberg, 2013) which can generate rare or novel rotavirus strains and hence

contribute to the vast RVA diversity (Kirkwood, 2010). Wa-like and DS-1-like intergenogroup reassortment events involving G1P[8] and DS-1-like genotype constellation have been described recently in six countries: Brazil, Japan, Philippines, Thailand, Vietnam and Malawi (Fujii *et al*.,2014; Kuzuya *et al*., 2014; Yamamoto *et al*., 2014; Komoto *et al*., 2015; Komoto *et al*., 2016; Nakagomi *et al*., 2017; Jere *et al*., 2018; Luchs *et al*., 2019). According to literature, viable atypical reassortant strains can occur under natural conditions involving Wa-like G1P[8] or G3P[8] outer capsid genes expressing DS-1-like genetic background (Arana *et al*., 2016; Cowley *et al*., 2016; Guerra *et al*., 2016; Guntapong *et al*., 2017). This study identified two DS-1-like G1P[8] strains in the course of the ongoing AEVGI whole-genome characterization of South African RVA strains. While reported in low frequencies and limited settings in Brazil (1.6% during 2013-2017 seasons) (Luchs *et al*., 2019), Thailand (0.4% during 2012-2014 seasons) (Komoto *et al*., 2015) and Vietnam (14% during 2012/2013 season) (Nakagomi *et al*., 2017), 31-62% of these DS-1-like G1P[8] strains accounted for RVA positive strains circulating across selected regions in Japan (Kuzuya *et al*., 2014) and 40% of randomly sampled post-vaccine samples were reported in Malawi (Jere *et al*., 2018). Such atypical reassortant strains have the potential to predominate in circulation. G1P[4] strains suggested to have emerged from intergenogroup reassortment events accounted for 41% of RVA strains circulating in the peak months of the 2001 RVA season in Detroit, USA (Abdel-Haq *et al*., 2003) whereas a surge in G3P[4] strains also presumed to have emerged from intergenogroup reassortment events were detected in Brazil at 36% (Rosa *et al*., 2002) and in Ghana at 64% (Asmah *et al*., 2001). The two South African study strains were identified during the pre-RVA vaccination period in South Africa in contrast to the previously reported atypical strains found during the post-RVA vaccination period. These implies that the reassortment events that led to the emergence of the South African atypical G1P[8] strains may not necessarily be driven by vaccine-induced selective pressure but by natural evolutionary processes of of RVA genome.

In order to identify the ancestral origin of these G1P[8] strains, assessment of the whole-gene sequences and phylogenetic analysis showed that the outer capsid genes, VP7 and VP4 of strains RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8] were 99.6-99.9% (99.1-100%) identical with South African

strain RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8] and clustered together in the same clade. For the internal genes, the highest nucleotide identities were identified with cognate genes of another South African strain, RVA/Human-wt/ZAF/MRC-DPRU2344/2008/G2P[6], detected in the 2008 RVA season. Put together, it is probable that a locally circulating G2P[6] strain such as RVA/Human-wt/ZAF/MRC-DPRU2344/2008/G2P[6] with a DS-1-like backbone derived the VP7 and VP4 genes from a locally co-circulating G1P[8] strain such as RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8], generating the double-gene reassortants. Consequently, the results obtained in this study indicate that the two atypical South African G1P[8] strains were generated locally through genetic reassortment events. The generation of these reassortant double-gene strains tend to be independent of the events in which the Brazilian, Japanese, Thai, Vietnamese and Malawian DS-1-like G1P[8] strains were generated. The nine internal genes of the South African DS-1-like G1P[8] strains always clustered together with cognate genes of a locally circulating G2P[6] strain distinctly away from the cluster comprising the Japanese, Thai, Philippines and Malawi DS-1-like G1P[8] strains. Therefore, the South African DS-1-like G1P[8] strains emerged clonally from independent events, a phenomenon observed for the Malawian (Jere *et al*., 2018) and Vietnamese [30] DS-1-like G1P[8] strains. In contrast, the Brazilian, Japanese and Thai G1P[8] DS-1-like strains were established to have been derived from a common ancestor (Kuzuya *et al*., 2014; Nakagomi *et al*., 2017; Luchs *et al*., 2019).

Vaccine escape mutants can result due to mutations occurring in well-known VP7 neutralization epitope regions (Coulson and Kirkwood, 1991). Host-antigen binding interactions involving human G1 strains are significantly impacted by mutations occurring at positions 94, 97, 147 and 291 (Coulson and Kirkwood, 1991). The identified N94S substitution involving polar non-charged amino acid residues (Betts and Russell, 2003), may not significantly alter the overall morphology of the protein surface. However, since asparagine is usually N-glycosylated, there is a likely loss of glycosylation site which could have a wide-ranging impact on the immunogenicity of the 7-1a epitope (Caust *et al*., 1987). The D97E amino acid substitution involving polar negatively charged residues, aspartate (D) and glutamate (E), is likely to be a silent nucleotide change (Betts and Russell, 2003). Similarly, a K291R substitution involving lysine (K), an amphipathic polar amino

acid and arginine (R), a positively charged amino acid, is unlikely to have a far-reaching structural effect on the VP7 protein surfaces. However, M217T substitution was identified resulting in substitution of methionine, a non-polar residue to threonine, a polar amino acid residue, that could likely result in significant changes in biochemical properties of VP7 (Betts and Russell, 2003). This M217T substitution was also present in earlier strains as well as some post-vaccine strains that were included for analysis and the role it plays in driving epidemiological fitness of G1 strains is not fully resolved. In the host cell, trypsin-like proteases cleave the VP4 spike protein into two structural domains (VP8* and VP5*) (Estes and Greenberg, 2013). Four antigenic epitopes (8-1 to 8-4) have been described in the VP8* region while five antigenic epitopes (5-1 to 5-5) in the VP5* region have been documented (Zeller *et al*., 2012). The amino acid changes E150D, N195G, S125N and N135D that were observed relative to the vaccine strains were conservative. However, a S131R substitution that resulted in a change in polarity might play a role in escape of host immunity (Betts and Russell, 2003).

4.5. Conclusion

Whole-genome analyses demonstrated that South African DS-1-like G1P[8] strains were generated by genetic reassortment mechanism between comtemporary endemic, South African strains (VP7 and VP4 outer capsid proteins from G1P[8] strains and DS-1-like backbone from G2P[6] strains). Similar to their pre-vaccine era detection in Vietnam and Philippines, the identification of these atypical DS-1-like G1P[8] strains during the pre-vaccine period in South Africa, as opposed to their detection during post-vaccination era in selected settings in Brazil (Sao Paulo and Goias in 2013), Japan (Okayama, Aichi, Akita, Kyoto and Osaka Prefectures in 2012), Thailand (Phetchabun and Sukhothai in 2013) and Malawi (Blantyre in 2013/2014), suggests that they originated from natural evolutionary processes of RVA genome. Whole-genome surveillance of RVA genotypes is imperative to understand the occurrence rate as well as mechanisms that drive emergence of such atypical strains, their epidemiological fitness and also to assess the effect of vaccine selective pressure in shaping the antigenic landscape of RVA strains.

CHAPTER FIVE

INVESTIGATION OF WHOLE-GENOME AND PHYLODYNAMICS OF SOUTH AFRICAN G1P[8] ROTAVIRUS STRAINS PRE- AND POST-VACCINATION INTRODUCTION

Chapter five is based on the publication "Whole genome in-silico analysis of South African G1P[8] rotavirus strains before and after vaccine introduction over a period of 14 years"

Mwangi, P. N., Mogotsi, M. T., Seheri, M. L., Mphahlele, M. J., Peenze, I., Esona, M. D., Kumwenda, B., Steele, A. D., Kirkwood, C. D., Ndze, V. N., Dennis, F. E. and Nyaga M. M. Whole genome in-silico analysis of South African G1P[8] rotavirus strains before and after vaccine introduction over a period of 14 years. *Vaccines,* 8*(4),* p. 609. Impact Factor 4.086

https://pubmed.ncbi.nlm.nih.gov/33066615/

5.1 Introduction

RVA is the major causative agent of AGE in children under five years (Estes and Greenberg, 2013). RVA-induced diarrhea is responsible for approximately 125,000 childhood mortality cases worldwide (Troeger *et al*., 2018). RVA is classified in the *Reoviridae* family and comprises 11 dsRNA gene segments that encode six structural proteins (VP1-VP4, VP6, and VP7) and five/six non-structural proteins (NSP1-NSP5/6) (Estes and Greenberg, 2013). A binary classification scheme underpinned by the neutralizing proteins, VP7 and VP4, has been universally used for classification of RVA strains (Estes and Kapikian, 2007). However, to fully describe rotavirus strains, the binary classification system was expanded by incorporating the other nine genome segments (Matthijnssens *et al*., 2011).

In the whole-genome classification scheme, nucleotide percentage similarity cut-off values of all the eleven viral gene segments, as recommended by the Rotavirus Classification Working Group (RCWG), are used to determine a genotypic scheme Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx designating VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6, respectively (Matthijnssens *et al*., 2008). The majority of RVA strains are assigned into three genogroups: Wa-

like (I1-R1-C1-M1-A1-N1-T1-E1-H1), DS-1-like (I2-R2-C2-M2-A2-N2-T2-E2-H2) and a relatively minor group, AU-1-like (I3-R3-C3-M3-A3-N3-T3-E3-H3) (Matthijnssens and van Rast, 2012).

G1P[8] is among the most predominant and medically important RVA strain, globally (Doro *et al*., 2014). In Africa, G1P[8] accounts for approximately 29% of all the circulating RVA strains (Doro *et al*., 2014). The G1 and P[8] gene segments sub-cluster into lineages whose emergence is attributed to various mechanisms of genetic diversity common in RNA viruses such as genetic mutation, recombination, and reassortment (Kirkwood, 2010). Distinct lineages for both genotypes G1 and P[8] collected from different geographical regions have been described in the literature (Arista *et al*., 2006; Banyai *et al*., 2009; Le *et al*., 2010; Bucardo *et al*., 2012; Cho *et al*., 2013).

In order to alleaviate RVA disease burden, four vaccines: RotaTeq® (Merck & Co., West Point, PA, USA); Rotarix® (GlaxoSmithKline, Rixenstart, Belgium); ROTAVAC® (Bharat Biotech, Hyderabad, India) and Rotasiil® (Serum Institute of India, Pune, India) have been pre-qualified by the World Health Organization (WHO) for global use to alleviate RVA disease burden (WHO, 2020). Rotarix® contains an attenuated human G1P[8] RVA strain (Bernstein *et al*., 1999) while RotaTeq® is composed of five human-bovine reassortant strains (G1P[5], G2P[5], G3P[5], G4P[5] and G6P[8]) (Ciarlet and Schodel, 2009). ROTAVAC® contains a G9P[11] strain (Bhandari *et al*., 2014a), while Rotasiil® is a pentavalent human-bovine reassortant vaccine comprising five reassortant strains containing human VP7 proteins representing G1, G2, G3, G4 and G9 genotypes (Zade *et al*., 2014).

South Africa introduced the Rotarix® vaccine into its Expanded Programme on Immunization (EPI) in September 2009 (WHO, 2009). In the first year, after the vaccine was introduced, RVA infections indicated by laboratory confirmed results and hospitalizations were reduced significantly by approximately 58% (Msimang *et al*., 2013). After the introduction of Rotarix® in South Africa, non-G1P[8] strains (such as G2P[4], G2P[6], G9P[8], G12P[8] and G8P[4]) which are not incorporated in the monovalent G1P[8] vaccine, increased significantly (Page *et al*., 2018). Notably, at Dr. George Mukhari Academic Hospital, a key RVA surveillance site, no G1P[8] strains were reported in 2012 (Page *et al*., 2018), whereas this was a predominant strain during the pre-

vaccine period (Steele *et al*., 2003). The enormous genetic and antigenic diversity within RVA (Kirkwood, 2010; Hoxie and Dennehy, 2020) and the recent emergence of novel strains (Esona *et al*., 2018; Hoa-Tran *et al*., 2020; Maringa *et al*., 2020) emphasize the need to monitor the impact of RVA vaccines on the genetic and antigenic landscape of RVA circulating in the population.

The impact of RVA vaccination in Sub-Saharan Africa has been substantial (Shah *et al*., 2017) and it is essential to continuously assess the long-term impact of vaccination on circulating RVA srains. Thus, there is a need for whole-genome longitudinal surveillance studies in South Africa to decipher potential RVA vaccine-induced strain changes. In this study, we investigated the impact of RVA vaccine introduction in South Africa on the most common human RVA strain, G1P[8]. This study is the first large-scale genomic analysis of human RVA collected seven years before and seven years after the introduction of the RVA vaccine in South Africa.

5.2 Materials and Methods

5.2.1 Ethics approval

The diarrheal stool samples were collected as a routine diagnostic clinical specimen when the parents brought their child to a health facility for clinical management, requiring no written informed consent. As part of the WHO-coordinated RVA surveillance network, the archived RVA-positive specimens were anonymized and utilized for strain characterization under a Technical Service Agreement and a Materials Transfer Agreement (MTA) to the WHO/AFRO Regional Reference Laboratory (WHO-RRL) based at Sefako Makgatho Health Sciences University (SMU), Pretoria, South Africa. The WHO Research Ethics Review Committee granted an exemption activity, noting that the study procedures were part of routine hospital-based RVA surveillance. The samples were transferred for whole-genome sequencing at the University of the Free State-Next Generation Sequencing (UFS-NGS) Unit through a MTA (SMU-UFS.1). The Health Sciences Research Ethics Committee (HSREC) of the UFS, Bloemfontein, South Africa, approved the study under ethics number UFS-HSD2018/0510/3107.

5.2.2 Strain description

Rotavirus positive stool samples (*n* = 103) previously characterized as G1P[8] were sourced from the archival storage of the Diarrheal Pathogens Research Unit (DPRU), WHO-RRL in Pretoria, South Africa. The samples were distributed as follows: 2002 (*n* = 14), 2003 (*n* = 7), 2004 (*n* = 14), 2005 (*n* = 6), 2006 (*n* = 22), 2007 (*n* = 7), 2008 (*n* = 9), 2009 (*n* = 13), 2010 (*n* = 1), 2013 (*n* = 1), 2014 (*n* = 5), 2015 (*n* = 3), and 2017 (*n* = 1). In addition to the 103 samples, an additional 68 whole-genome sequences for G1P[8] strains collected from South Africa were extracted from the GenBank database (Benson *et al*., 2015).

5.2.3 Extraction and purification of double-stranded RNA

A fecal suspension was prepared by adding ~ 100mg stool sample into 200µl Phosphate Buffered saline (PBS) solution, 0.01M, pH 7.2 (Sigma-Aldrich®, St Louis, MO, USA) with subsequent extraction of viral RNA as previously described (Potgieter *et al*., 2009) albeit, with some modification. Briefly, the modifications included the volume (900 µL TRIzol™-LS: 300 µL stool sample suspension), the incubation period of dsRNA enrichment (24 h), centrifugation speeds (20000 × *g*), and the staining reagent (PronaSafe, Condalab, Camberley, UK), as captured in the UFS-NGS unit extraction standard operating procedure (SOP).The extracted RNA was purified using the MinElute PCR purification Kit by following the manufacturers' instructions (Qiagen, Hilden, Germany).

5.2.4 Complementary DNA(cDNA) synthesis

cDNA was synthesized from the extracted viral RNA using the Maxima H Minus Double-Stranded Synthesis Kit and protocol (ThermoFisher, Waltham, MA) with some modifications (UFS-NGS Unit SOP). Briefly, the purified extracted RNA was denatured at 95^{o}C for 5 min in a thermocyler (Labnet, Edison, NJ, USA). A volume of 1µl of Random Hexamer primer, 100µM, was then added to the denatured DNA. The reaction mixture was incubated in a thermoycler at 65^{o}C for 5 minutes. Afterward, 5µl volume of the 4X First-Strand Reaction Mix and 1µl of First Strand Enzyme Mix was added and the reaction mixture incubated in a thermocyler pre-programmed as follows (10 minutes at 25^{o}C, 120 minutes at 50^{o}C, 5 minutes at 85^{o}C). For second-strand cDNA synthesis step, a 55µl volume of Nuclease-Free water, 20µl of 5X Second Strand Reaction Mix and 5µl of Second Strand Enzyme Mix was added and incubated in the thermocyler at 16^{o}C for 60 minutes. A 6µl volume of EDTA, pH 8.9 was added followed by 10µl of RNase I. The synthesized

cDNA was purified using MSB® Spin PCRapace Kit by following manufacturer's protocol (Stratec Molecular, Berlin, Germany).

5.2.5 DNA Library preparation and whole-genome sequencing

The DNA libraries were prepared by utilizing the Nextera® XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Quantitative and qualitative assessment of DNA was performed using Qubit 3.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and Agilent 2100 BioAnalyzer® (Agilent Technologies, Waldbronn, Germany), respectively, by following the manufacturer's instructions. The DNA library and the PhiX Control v3 library (Illumina, San Diego, CA, USA) were normalized to 8 pM and 20 pM concentrations, respectively. A volume of 600 µL pooled denatured DNA library spiked with 20% PhiX Control v3 library was loaded into a MiSeq® Reagent Kit V3 for paired-end nucleotide sequencing (301 × 2) on a MiSeq® sequencer (Illumina, San Diego, CA, USA) at the University of the Free State-Next Generation Sequencing (UFS-NGS) Unit, Bloemfontein, South Africa.

5.2.6 Genome assembly

The trimming of Illumina read ends and subsequent genome assembly was performed using a suite of tools embedded in Geneious Prime® software, version 2020.1.1 (Kearse *et al*., 2012). An inhouse data analysis pipeline and CLC Genomics Workbench 12 (https://www.qiagenbioinformatics.com), were utilized as complementary resources.

5.2.7 Generation of whole-genome constellations

The whole-genome constellation were determined using the Virus Pathogen Database and Analysis Resource (ViPR) (http://www.vipbrc.org)).

5.2.8 Phylogenetic analyses

Alignments and comparative analysis of the full length sequences for each gene segment was performed as described previously (Esona *et al*., 2018). Duplicated sequences in the alignments were identified utilizing ElimDupes. The best evolutionary models for each gene segment were estimated using the DNA Model Test program in MEGA 6 to guide in the construction of

Maximum-likelihood trees with 1000 bootstrap replicates. The number of duplicate sequences were indicated alongside the sequences in the phylogenetic tree.

5.2.9 Selection pressure and recombination analysis

Analysis of natural selection in RVA genome segments was done using the suite from the DataMonkey Webserver (Weaver *et al*., 2018): Fixed-effects Likelihood (FEL) (Kosakovsky *et al*., 2005), Fast Unconstrained Bayesian Approximation for Inferring Selection (FUBAR) (Murrell *et al*., 2013) and Mixed-effects model of episodic selection (MEME) (Murrell *et al*., 2012). Amino acid sites that were identified by all the three abovementioned methods were considered for positive selection. Analysis of genetic recombination was performed using Genetic Algorithm for Recombination Detection (GARD) (Kosakovsky *et al*., 2005).

5.2.10 Protein modelling

The RVA protein structures were modeled using SWISS-MODEL with an initial template search (Waterhouse *et al*., 2018). The templates were selected from the SWISS-MODEL Template Library (SMTL) and their respective resolution values for the analyzed genes were as follows: VP7 (3fmg.1, 3.40Å) and VP4 (2dwr.1, 2.50Å). The evaluation of stereochemical quality parameters of the generated structures was performed using the Structure Assessment feature in the Swiss-MODEL server (Waterhouse *et al*., 2018) and VERIFY3D (Eisenberg *et al*., 1997). Image visualization and analysis was performed in PyMol software (De Lano, 2009).

5.2.11 In silico analysis of effect of mutation(s) on protein stability

The FoldX plugin (Van Durme *et al*., 2011) integrated in the YASARA platform (Krieger *et al*., 2002) was used to predict the stability effect of mutation(s) in a 3D structure. FoldX estimates stability effect of a mutation empirically whereby the stability (ΔG) of a protein is defined by the free energy, which is expressed in kcal/mol. In this study, G is the difference of free energy between the Rotarix® (vaccine) strain and mutant strain. The energy with positive value is regarded to destabilize the structure, while a mutation with a negative value is regarded to stabilize the structure. Free energy change of ±0.5 kcal/mol is regarded to be statistically significant for stabilizing/destabilizing effect (Van Durme *et al*., 2011).

5.3 Results

5.3.1 Whole genome constellation determination

The genotype constellation for the 103 South African G1P[8] strains (92 pre- and 11 post-vaccine) sequenced in this study and corresponding 68 strains (56 pre- and 12 post-vaccine) acquired from GenBank database as reference strains) was typical G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1 (Appendix 6). The sizes of the full-length genome segments one to eleven and their respective open reading frames (ORFs) were determined (Appendix 6). All whole-gene sequences of the 103 G1P[8] strains sequenced in this study were submitted in the NCBI GenBank database under accession numbers MT854335-MT855467.

5.3.2 Phylogenetic analyses

5.3.2.1 Phylogenetic analyses of VP7 and VP4

Phylogenetic trees for each of the 11 gene segments were constructed. For VP7 and VP4, well-known lineage designations were utilized (Arista *et al*., 2006; Le *et al*., 2010). The VP7 phylogenetic tree comprised the VP7 gene sequences of the South African strains together with VP7 gene sequences of the reference strains from the seven established VP7 G1 genotype lineages (Arista *et al*., 2006). The 171 G1 South African sequences utilized in this study segregated into two main lineages: G1-lineage I and lineage II (Figure 5.1). G1-lineage I comprised 76 pre-vaccine G1 strains and 12 post-vaccine G1 strains, while G1-lineage II comprised 72 pre-vaccine G1 strains and 11 post-vaccine G1strains (Figure 5.1). Strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2250/2013/G1P[8] had ≥ 99.9% nucleotide identity with Rotarix® in all its structural and nonstructural genes and clustered alongside cognate Rotarix® genes in all the 11 phylogenetic trees. South African pre- and post-vaccine G1 strains were highly identical to each other and similarly distant from the Rotarix® strain (Table 5.1).

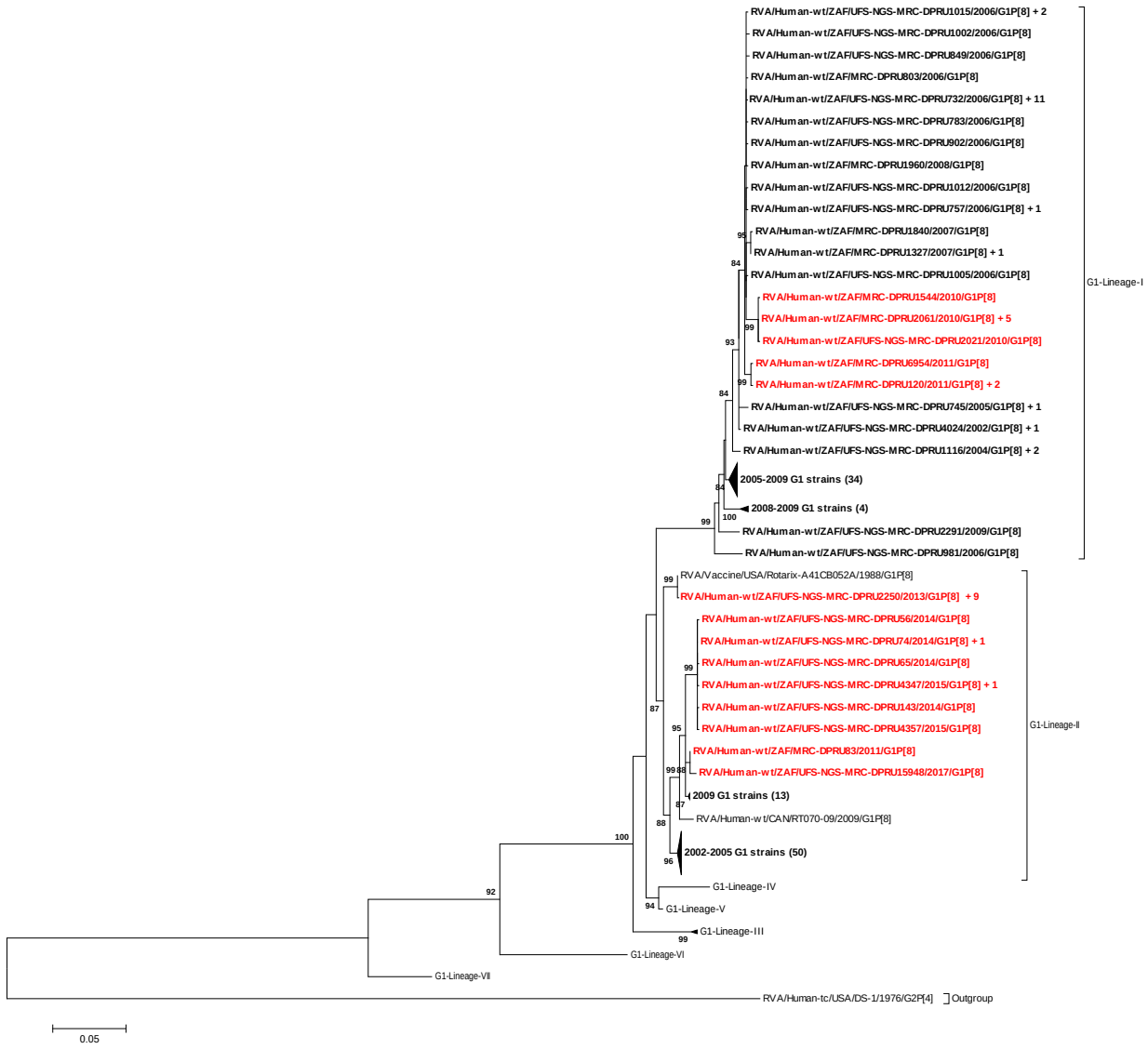


Figure 5.1: VP7 phylogenetic tree of G1 strains

Maximum likelihood phylogenetic tree based on the full-length nucleotide sequences of genome segment 9 encoding the VP7 protein. The T92 + G evolutionary model was used for phylogenetic inference. South African pre-vaccine G1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Only bootstrap values ≥ 70% are shown adjacent to each branch node. Scale bar indicates the number of nucleotide substitutions per site.

Table 5.1: Nucleotide identitiy analyses between South African pre- and post-vaccine G1P[8] strains, pre-vaccine G1P[8] strains with Rotarix® and post-vaccine G1P[8] strains with Rotarix® strain.

Gene segments/Nucleotide identity values in percentage	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
Comparison between pre- and post-vaccine G1P[8] strains	92.1–100	87.8–99.5	89.0–99.4	93.7–99.0	92.9–99.6	89.4–99.1	83.5–99.5	89.1–99.7	94.2–100	91.0–100	92.2–99.8
Comparison between pre-vaccine G1P[8] strains and Rotarix® strain	92.7–100	89.6–91.1	88.2–98.7	94.5–99.1	93.1–99.0	91.2–98.6	83.6–100	88.3–90.9	95.3–100	91.4–98.9	92.2–98.8
Comparison between post-vaccine strains and Rotarix® strain	93.3–100	89.9–99.9	88.9–100	94.5–100	93.0–100	90.9–99.9	84.0–99.9	89.7–100	97.1–100	91.0–100	92.9–100

South African pre- and post-vaccine G1P[8] strains were compared with each other and also compared with the Rotarix® strain. The nucleotide identity values were calculated using the p-distance algorithm in MEGA 6 software.

The VP4 phylogenetic tree comprised nucleotide sequences of the South African RVA strains and those of the reference strains from the four established VP4 P[8] genotype lineages (P[8]-lineage I to IV) (Le *et al*., 2010) (Figure 5.2). The P[8] strains for South African RVA segregated into three evolutionary lineages, P[8]-lineage I, III and IV (Figure 5.2). Lineage III comprised a mixture of 144 pre-vaccine and 21 post-vaccine P[8] strains, while lineage IV comprised only five pre-vaccine strains (Figure 5.2). The P[8] strains identified pre- and post-vaccination were highly identical to each other and similarly distant from the Rotarix strain (Table 5.1).

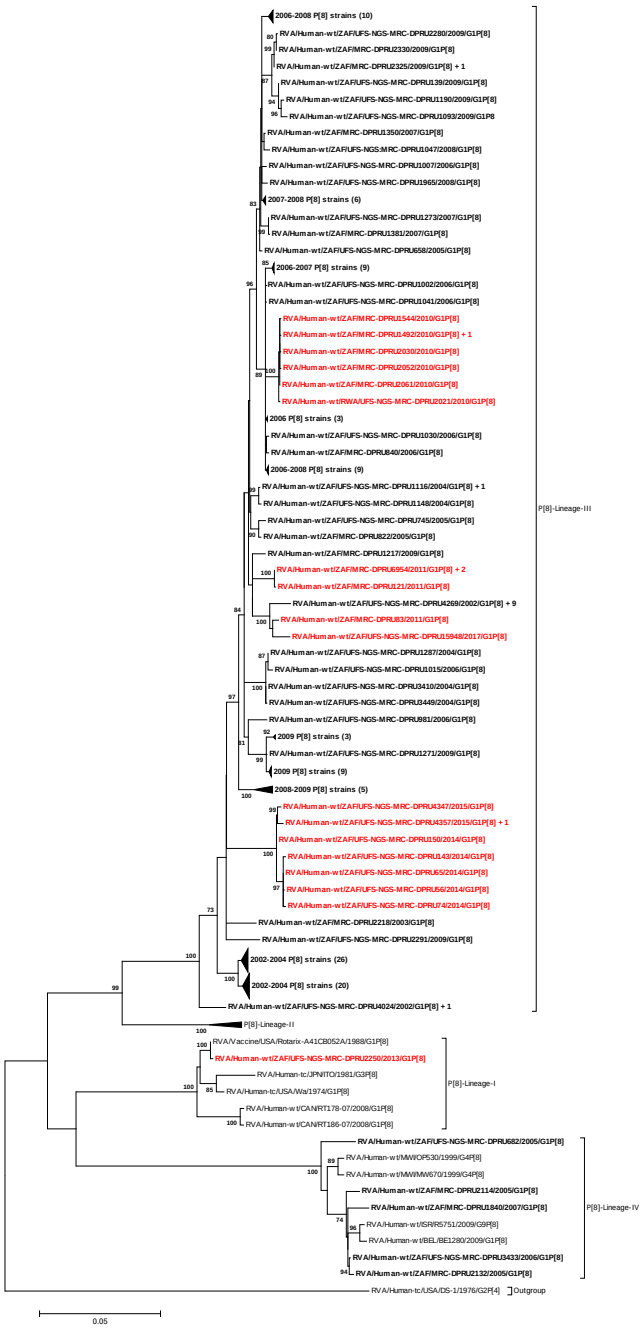


Figure 5.2: VP4 phylogenetic tree of P[8] strains

Maximum likelihood phylogenetic tree based on the full-length nucleotide sequences of genome segment 4 encoding VP4 protein. The GTR+G+I evolutionary model was used for phylogenetic inference. African pre-vaccine P[8] strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences (S3). The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Only bootstrap values ≥70% are shown adjacent to each branch node. Scale bar indicates the number of nucleotide substitutions per site.

5.3.2.2 Phylogenetic analysis of VP1-VP3, VP6 and NSP1-NSP5

South African VP1-VP3 and VP6 genes clustered into two main lineages, one of the lineages comprising of the Rotarix® genome segments (Figures 5.3-5.11). South African strains identified pre- and post-vaccination were highly identical to each other as well as similarly distant from the vaccine strain (Table 5.1).

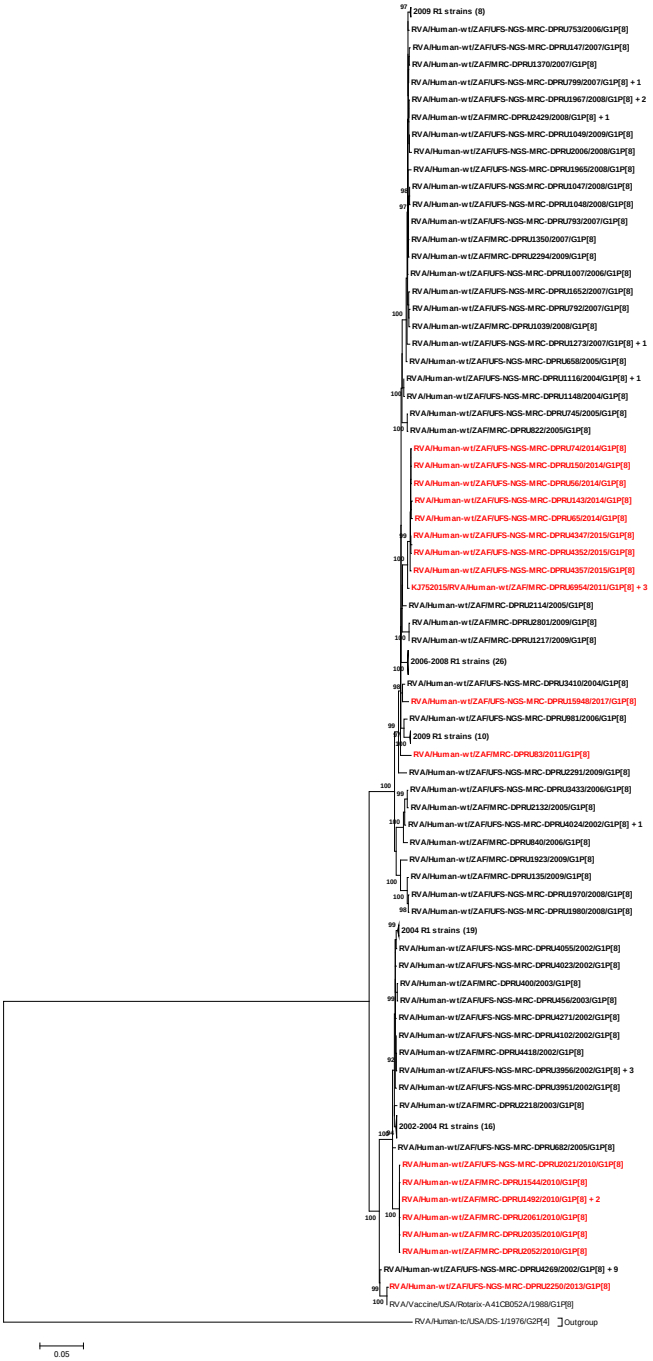


Figure 5.3: VP1 phylogenetic tree of R1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of VP1 gene segment. South African pre-vaccine R1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

Figure 5.4: VP2 phylogenetic tree of C1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP2 gene segment. South African pre-vaccine C1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

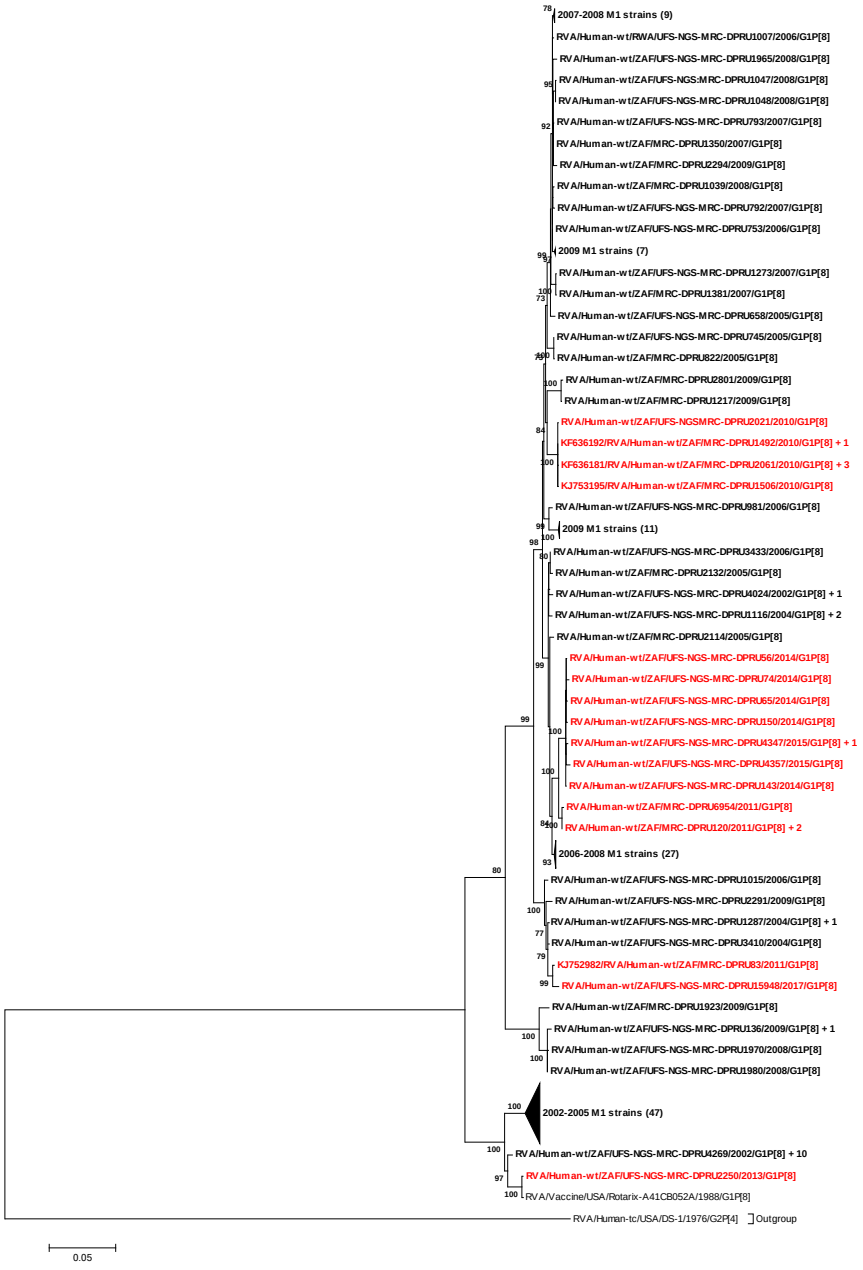


Figure 5.5: VP3 phylogenetic tree of M1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of VP3 gene segment. South African pre-vaccine M1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

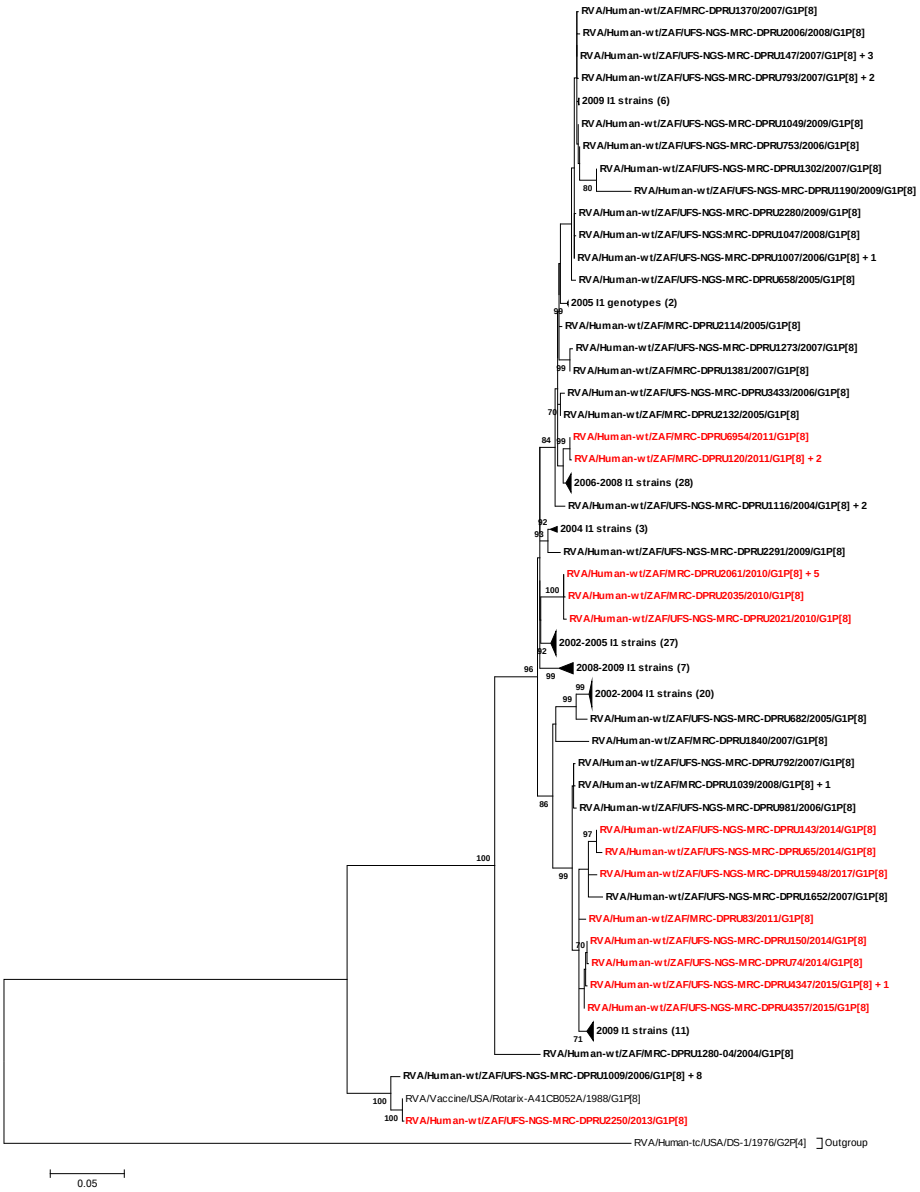


Figure 5.6: VP6 phylogenetic tree of I1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequence of VP6 gene segment. South African pre-vaccine I1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

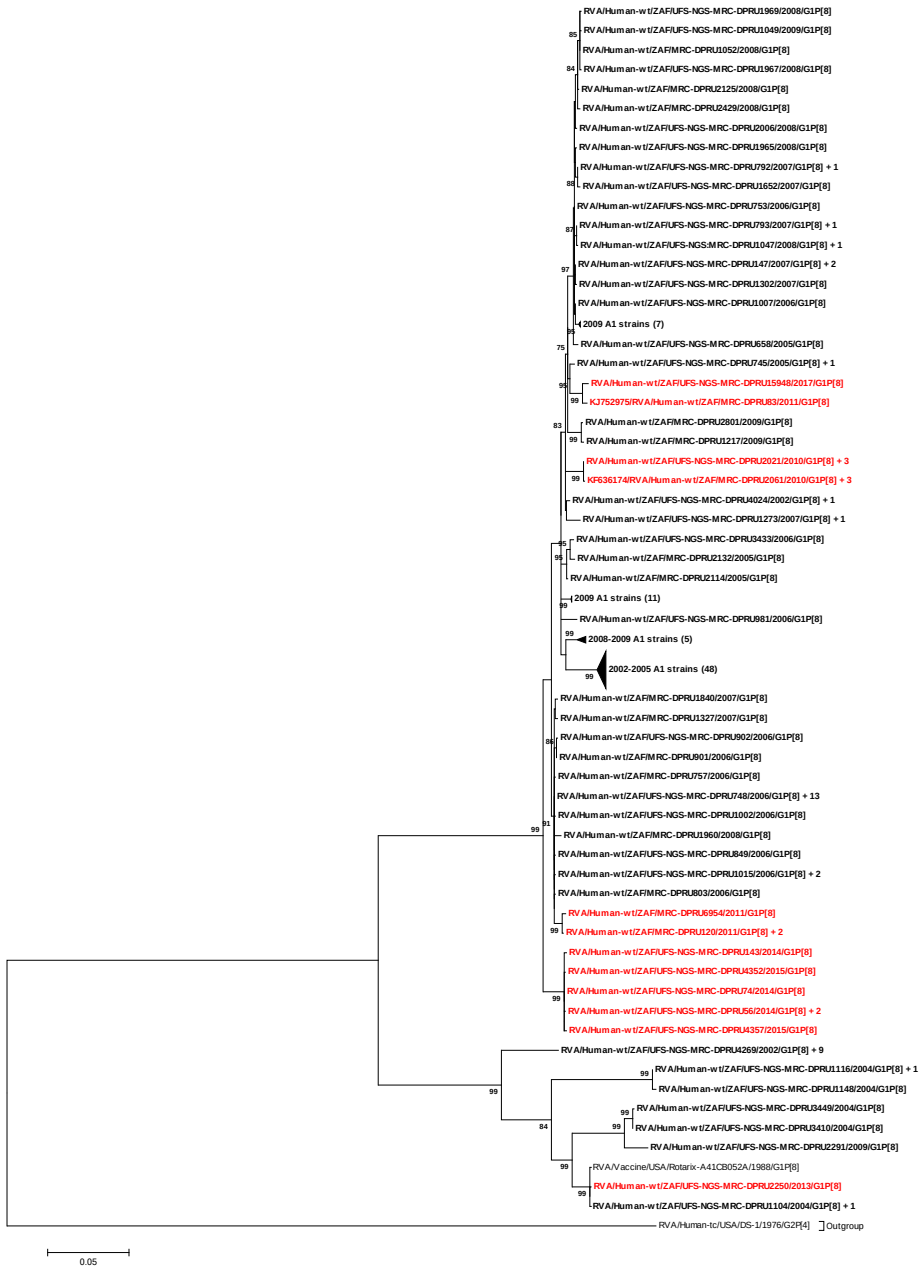


Figure 5.7: NSP1 phylogenetic tree of A1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP1 gene segment. South African pre-vaccine A1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

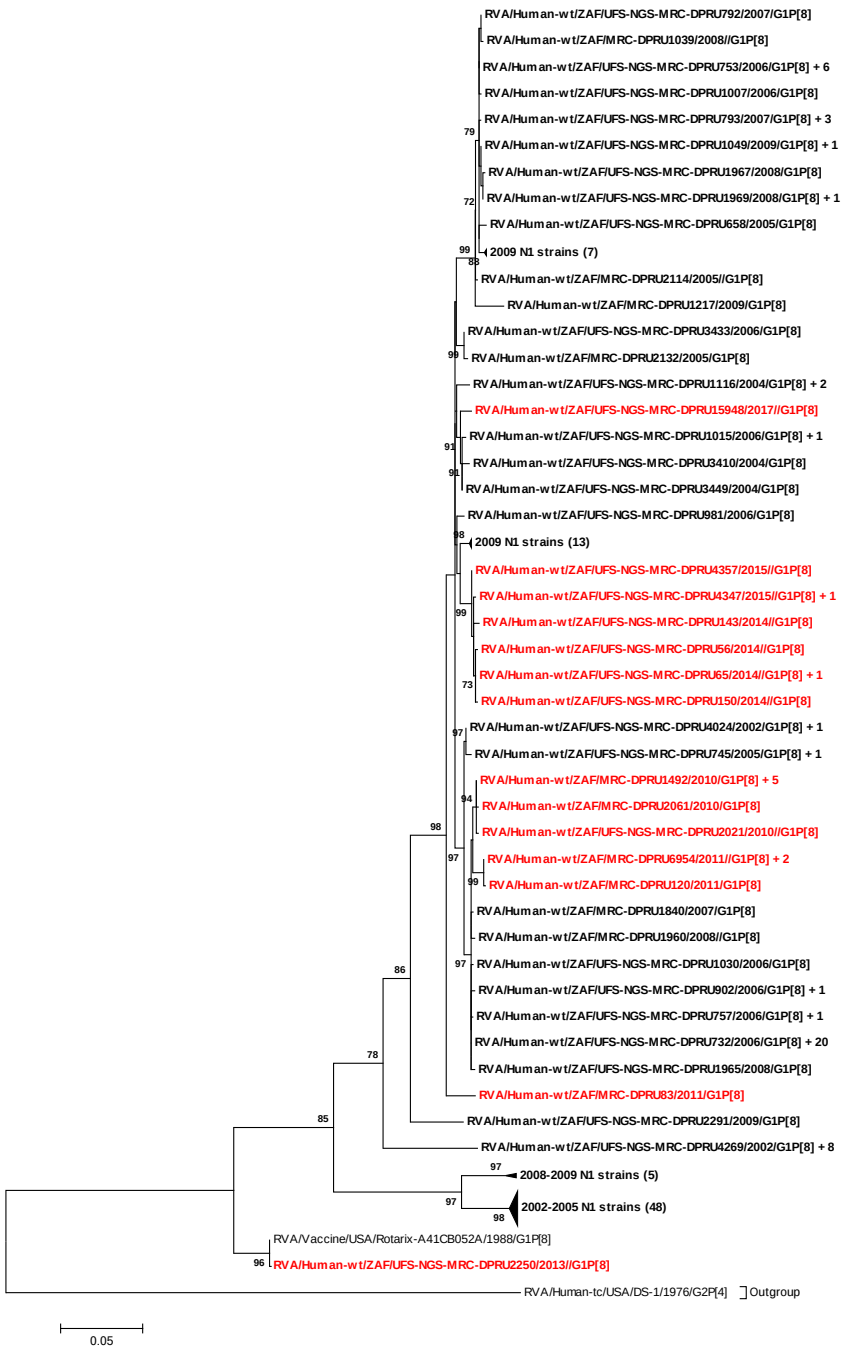


Figure 5.8: NSP2 phylogenetic tree of N1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP2 gene segment. South African pre-vaccine N1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

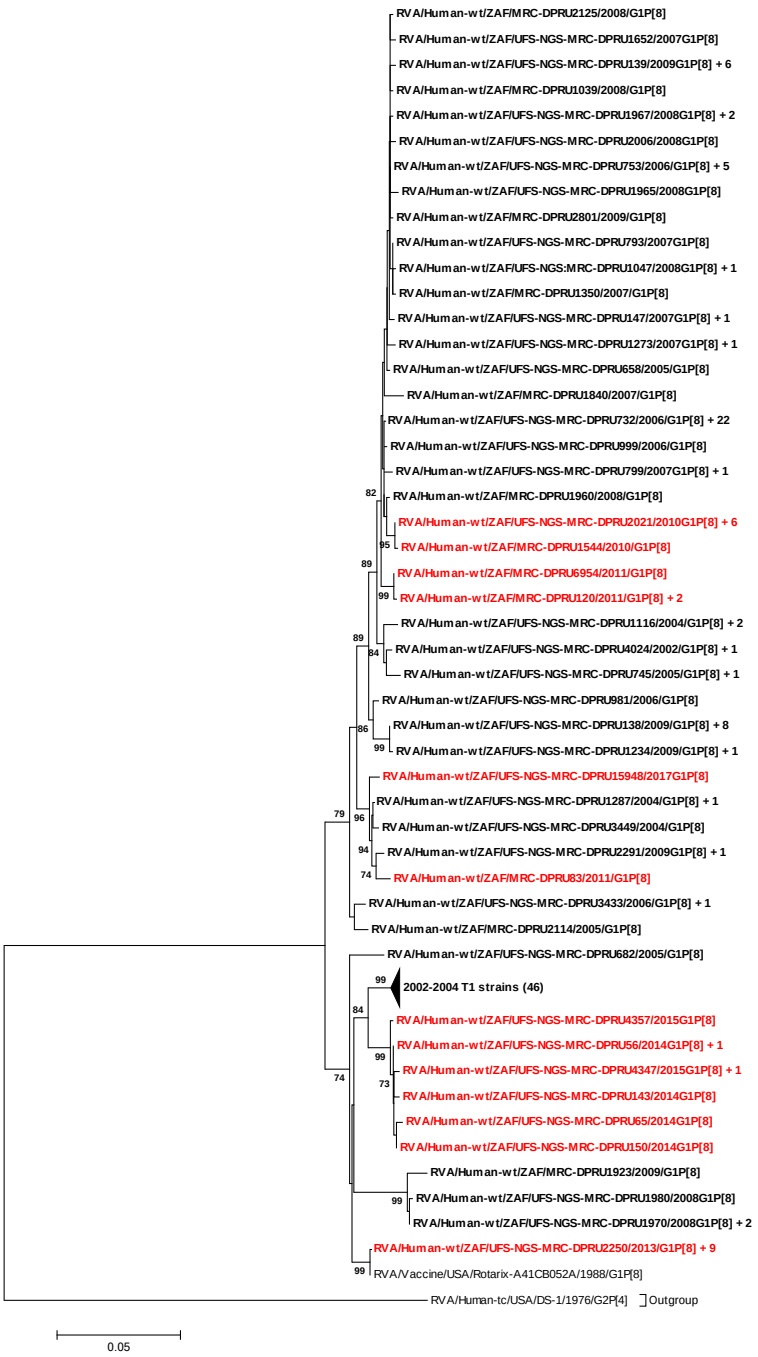


Figure 5.9: NSP3 phylogenetic tree of T1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP3 gene segment. South African pre-vaccine T1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

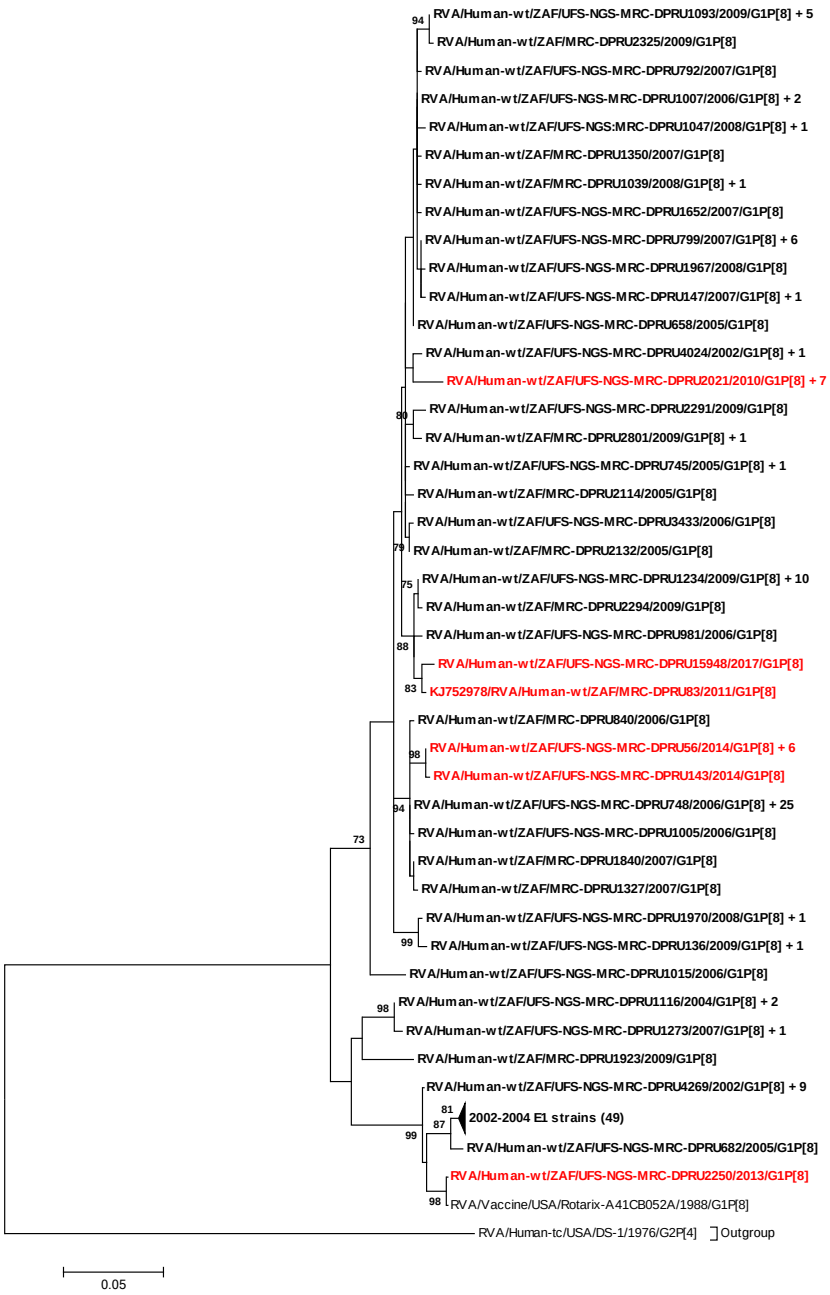


Figure 5.10: NSP4 phylogenetic tree of E1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP4 gene segment. South African pre-vaccine E1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

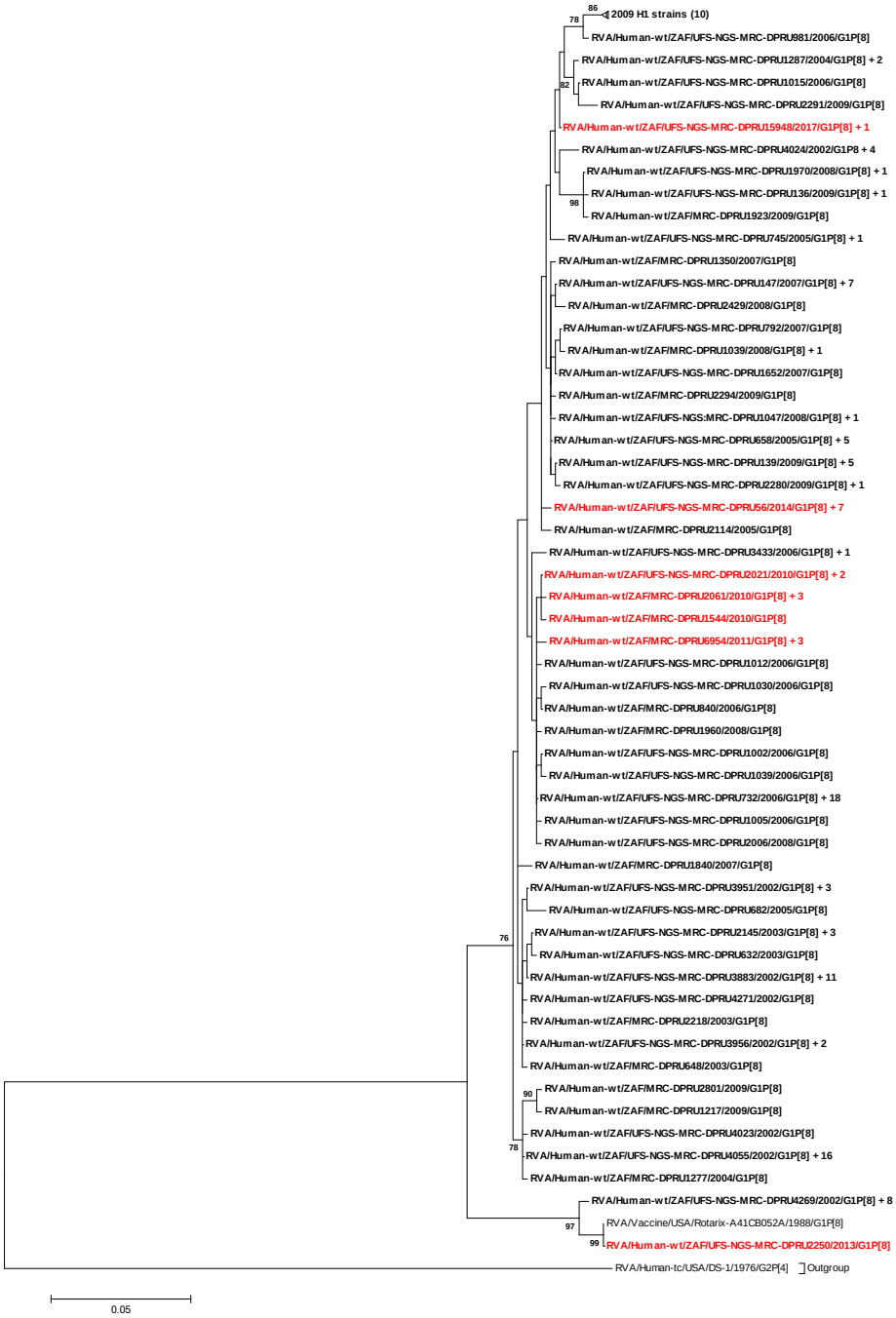


Figure 5.11: NSP5 phylogenetic tree of H1 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP5 gene segment. South African pre-vaccine H1 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. Adjacent to some sequences is indicated a plus (+) sign followed by the number of identical sequences. The number in brackets denotes the number of compressed strains. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

5.3.3 Protein modelling and amino acid analysis

5.3.3.1 Comparative analysis of neutralizing antigeni VP7 epitopes of South African G1 strains and Rotarix® strains

The VP7 protein contains two defined neutralization antigenic epitopes spanning across 29 amino acid residues: 7–1 (7–1a and 7–1b) and 7–2 regions (Aoki *et al*., 2009). A comparison of the amino acid residues in the antigenic epitopes of wild type South African G1 strains with cognate sites in the Rotarix® strain was performed (Appendix 7). Five amino acid differences appeared in both pre- and post-vaccine strains, four amino acids occurred before vaccine introduction while two amino acid differences, N147D (in eight strains) and T242A (in one strain), appeared only during the post-vaccine period (Figure 5.12; Table 5.2). We performed protein modeling analysis on the post-vaccine strains, which showed amino acid differences relative to both the pre-vaccine strains and Rotarix® strain. Therefore, for VP7, we selected strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4357/2015/G1P[8] as a representative strain of the eight G1 post-vaccine strains that had N147D amino acid substitution (Appendix 7). Additionally, we selected strain RVA/Human-wt/ZAF/MRC-DPRU1544/2010/G1P[8], the only post-vaccine strain exhibiting a T242A amino acid difference (Appendix 7). The VP7 protein structures of the selected G1 strains were superposed with the Rotarix® VP7 structure to assess differences in the structural conformation. The VP7 structure of strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4357/2015/G1P[8] when superposed with the Rotarix® VP7 had a root mean square deviation (RMSD) value of 0.020 Å (Table 5.2; Figure 5.12). The RMSD value closer to zero suggests high structural homology (Maiorov and Crippen, 1994). Replacement of asparagine with aspartate (N147D) significantly destabilized the protein structure as the free energy change was +0.527 kcal/mol (Van Durme et al., 2011) (Table 5.2; Figure 5.12). Apart from the VP7 protein structure of Rotarix®, we performed protein modeling analysis using the protein structures of five randomly selected pre-vaccine study strains. Similar significant free energy change trends for the destabilizing effect were observed and ranged from +0.506 to +0.579 kcal/mol. The VP7 structure of strain RVA/Human-wt/ZAF/MRC-DPRU1544/2010/G1P[8] had an RMSD value of 0.012 Å when superposed with Rotarix® VP7 (Table 5.2; Figure 5.12). Threonine, which is a neutral hydrophilic amino acid, was replaced with alanine, an aliphatic hydrophobic amino acid, showing a change in polarity.

However, the replacing amino acid (alanine) does not significantly alter the stability of the protein at this epitope (−0.076 kcal/mol) (Van Durme *et al*., 2002). Protein modeling analysis with VP7 protein structures from five randomly selected pre-vaccine study strains showed similar free energy change trends ranging from −0.070 to −0.078 kcal/mol.

Table 5.2: Possible effects of amino acid mutations in South African post-vaccine G1 neutralization epitopes

Strain used for the protein modelling	Mutation	No. of post-vaccine strain(s) with the mutation	Region	Amino acid property change	Superposition value (RMSD)	Free energy change (kcal/mol)	Possible effect
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4357/2015/G1P[8]	N147D	8	7-2 epitope	Hydrophilic to hydrophilic; Neutral to negative charge	0.020 Å	+0.527	The change in charge may alter the biochemical properties of the epitope. The mutation significantly destabilizes the structure of the protein.
RVA/Human-wt/ZAF/MRC-DPRU1544/2010/G1P[8]	T242A	1	7-1b epitope	Hydrophilic to hydrophobic; Neutral to Neutral charge	0.012 Å	-0.076	The change in polarity may alter the physicochemical property of the epitope

Relative Mean Square Deviation (RMSD) is the superposition value where value of zero indicates absolute similarity. The stability of the protein after mutation was measured in kcal/mol whereby free energy change of ±0.5 kcal/mol was regarded as significant for either stabilizing/destabilizing effect. Minus (-) value is indicative of stabilizing effect while positive (+) value is indicative of destabilizing effect.

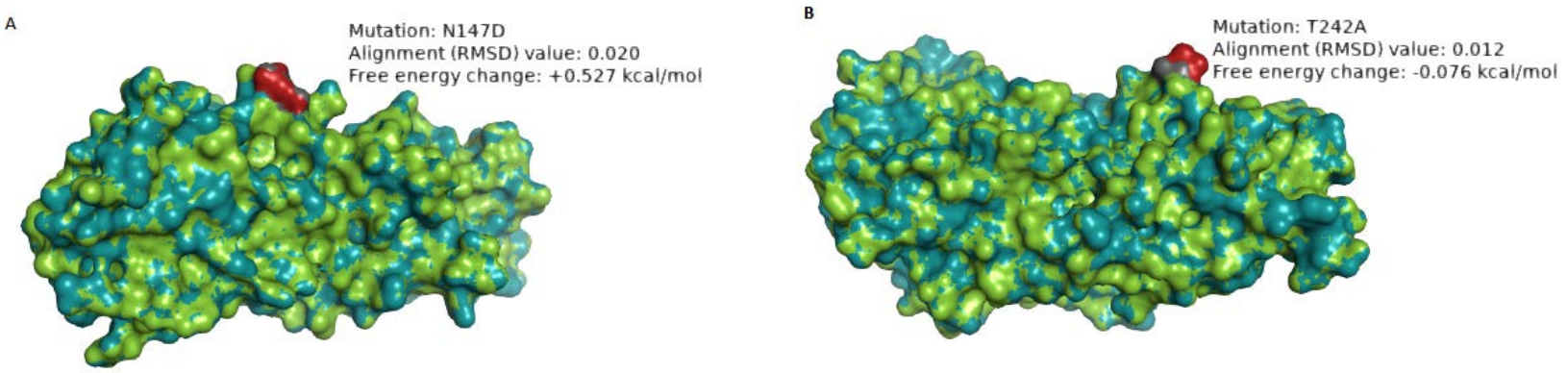


Figure 5.12: Protein modelling of VP7 protein structures

Structure models showing the mutation sites N147D (**A**) and T242A (**B**), respectively, that were only present in post-vaccine G1P[8] strains when superposed with VP7 of Rotarix®. The structure of the study strain is in green while Rotarix® structure is in deep-teal. The aa residues are represented in dark gray and red to examine whether the replacing aa alters the conformation of the protein structure. The aa residue highlighted in dark gray represent the study strain while that in firebrick represent the Rotarix® strain. RMSD is the superposition value where value of zero indicates absolute similarity. The stability of the protein after mutation was measured in kcal/mol whereby free energy change of ± 0.5 kcal/mol was regarded as significant for stabilizing/destabilizing effect.

5.3.3.2 Comparative analysis of VP7 cytotoxic T lymphocyte epitopes

The two established VP7 cytotoxic T cell epitopes at positions 16-28 and 40-52 (Morozova *et al*., 2015) of the VP7 of South African G1P[8] strains were compared with cognate regions of the Rotarix® VP7. The aa differences (I22M in two strains, R28Q in one strain, A46V in three strains, T48A in twenty strains and T48I in one strain) were only observed during pre-vaccine era while aa differences (A43V in two strains and A46T in four strains) were only present during post-vaccine era (Table 5.3; Appendix 7).

Table 5.3: Amino acid differences in South African post-vaccine G1 Cytotoxic T Lymphocyte epitope and non-epitope regions

No of strain(s)	Mutation	Region	Amino acid property change	Possible effect
All 2010 strains	S37I	Signal domain (H2 hydrophobic domain)	Hydrophilic to hydrophobic; Neutral charge to neutral charge	The change in polarity charge may alter the chemical properties of the protein
Four 2011 strains	A46T	Signal domain (H2 hydrophobic domain)/ CTL epitope region	Hydrophobic to hydrophilic Neutral charge to neutral charge	The change in polarity may alter the chemical properties of the protein
One 2011 strain	E222K	β-jelly roll domain/ antigenic site C	Hydrophilic to hydrophilic Negative charge to positive charge	The change in charge may alter the chemical properties of the protein. The folding energy is significant for stabilizing effect on the protein structure
One 2014 strain	A68D	Extension region of N-terminal	Hydrophobic to hydrophilic Neutral charge to negative charge	The change in charge may alter the physicochemical properties of the protein
Two 2015 strains	A43V	Signal domain (H2 hydrophobic domain)/CTL epitope region	Hydrophobic to hydrophobic; Neutral charge to Neutral charge	The biochemical property of the region may be conserved as the amino acid property is not altered.

The table indicates the aa differences that were observed in the VP7 cytotoxic T lymphocyte epitopes at positions 43 (in two post-vaccine G1 strains) and 46 (in four post-vaccine G1 strains) during post-vaccine period. Also the figure shows the aa differences observed in non-neutralizing regions of South African post-vaccine G1 strains that were not present during pre-vaccine period.

5.3.3.3 Comparative analysis of neutralizing antigenic epitopes in VP4 genes of South African P[8] strains and Rotarix®strain

The VP4 protein comprises the VP5* and VP8* regions. The VP8* region contains four (8–1 to 8–4) neutralizing antigenic epitopes, while VP5* has five (5–1 to 5–5) (Dormitzer *et al*., 2002)]. The differences between the P[8] study strains and Rotarix® P[8] component were mostly contained in the VP8* epitopes 8–1, 8–3, and 8–4 (Appendix 8). Seven amino acid differences were identified in both pre- and post-vaccine strains, seven amino acid substitutions during pre-vaccination introduction and two amino acid substitutions, T88I (in one strain) and N89S (in one strain) were identified after vaccine introduction (Appendix 8). We performed protein modeling analysis on the post-vaccine strains that showed amino acid differences relative to both the pre-vaccine strains and Rotarix® strain and these were RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU74/2014/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU83/2011/G1P[8] (Appendix 8). The VP4 structure of strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU74/2014/G1P[8] aligned with the RMSD value of RMSD 0.048 when superposed with the Rotarix® VP4 (Table 5.4; Figure 5.13). Replacement with the isoleucine at position T88I altered the polarity of the residue from polar to nonpolar without affecting the charge (Betts and Russell, 2003). This change in polarity may alter the physical properties of the protein. The mutation did not significantly impact the stability of the protein (−0.297 kcal/mol) (Table 5.4; Figure 5.13). Further free energy change analysis with VP4 structures from five randomly selected pre-vaccine strains demonstrated similar free energy change trends that ranged from −0.504 to −0.322 kcal/mol. The VP4 structure of strain RVA/Human-wt/ZAF/MRC-DPRU83/2011/G1P[8] had an RMSD value of 0.049 when superposed with Rotarix® VP4, indicating significant alignment. Asparagine is a polar, neutrally charged amino acid, and the replacing amino acid, serine, is also polar with a neutral charge (Betts and Russell, 2003). The replacing amino acid, serine, significantly destabilized the protein structure with +1.166 kcal/mol energy change (Table 5.4: Figure 5.13).When protein modeling analysis was performed with VP4 structures of five randomly selected pre-vaccine P[8] strains, significant free energy change values ranging from +1.128 to +1.352 kcal/mol were also observed, indicating the amino acid substitution had a destabilizing effect on the protein structure.

Table 5.4: Possible effects of amino acid mutations in South African post-vaccine VP4 epitopes

Strain used for protein modelling	Mutation	No. of strain(s) with the mutation	Region	Amino acid property change	Superimposition value (RMSD)	Free energy change (kcal/mol)	Possible effect
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU74/2014/G1P[8]	T88I	1	8-4 epitope	Hydrophilic to hydrophobic; Neutral charge to neutral charge	0.048 Å	-0.297	The change in polarity may alter the physicochemical properties of the protein. No significant impact on the stability of the protein structure
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU83/2011/G1P[8]	N89S	1	8-4 epitope	Hydrophilic to hydrophilic; Neutral charge to Neutral charge	0.049 Å	+1.166	The loss of the glycosylation site may alter the chemical properties of the protein. The mutation destabilized the protein structure.

Relative Mean Square Deviation (RMSD) is the superimposition value where value of zero indicates absolute similarity. The stability of the protein after mutation was measured in kcal/mol whereby folding energy change of ±0.5 kcal/mol is considered statistically significant for either stabilizing (-)/destabilizing (+) effect.

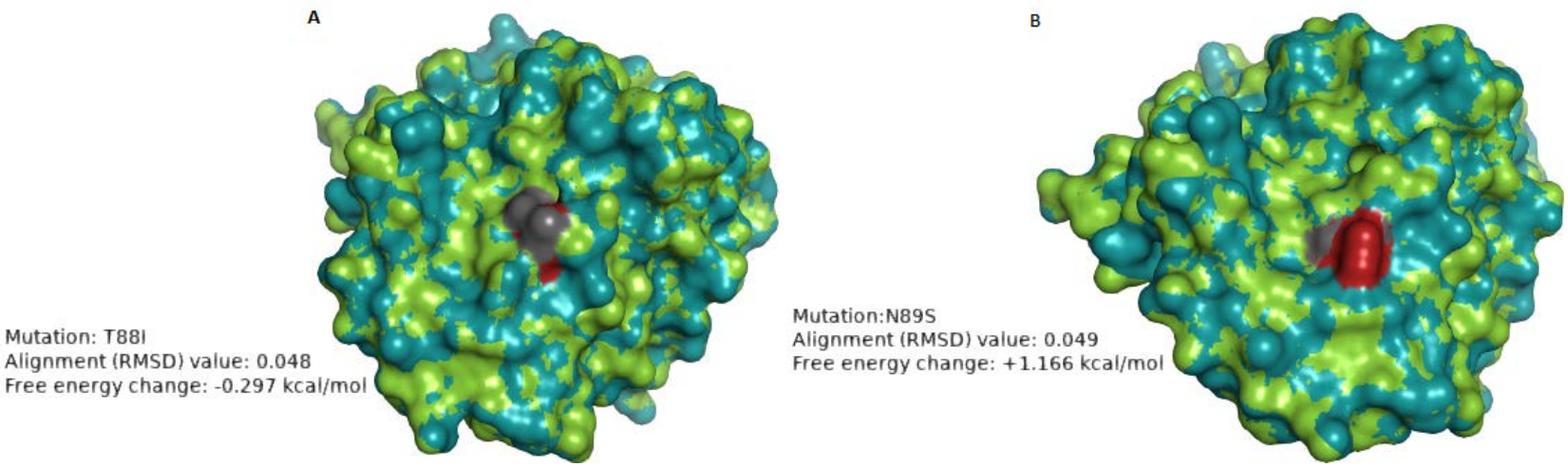


Figure 5.13: Protein modelling of VP4 structures

VP4 structure models showing mutations at sites T88I (**A**) and N89S (**B**), respectively, that were only seen in some post-vaccine strains when superposed with VP4 of Rotarix®. The aa residues are represented in dark gray and red to examine whether the replacing aa alters the conformation of the protein structure. The aa residue highlighted in dark gray represents the study strain while firebrick red represents the Rotarix® strain. The stability of the protein after mutation was measured in kcal/mol whereby folding energy change of ±0.5 kcal/mol is regarded as statistically significant for either stabilizing (-)/destabilizing (+) effect.

5.3.3.4 Analysis of the VP4 and VP6 non-neutralizing regions

Comparative analysis of the South African post-vaccine P[8] genotypes outside the known epitope regions with cognate regions in pre-vaccine strains identified five aa differences that were coservative in nature. (Table 5.4).

Table 5.5: Amino acid differences in South African post-vaccine P[8] RVA strains

No of strain(s)	Mutation	Region	Amino acid property change	Possible effect
All 2010 eight strains	V112I	VP8*	Hydrophobic to hydrophobic; Neutral to neutral charge	No change in the physicochemical properties
One 2015 strain	V444A	VP5*	Hydrophobic to hydrophobic Neutral charge to neutral charge	No change in the physicochemical properties
Two strains (one in 2010 and 2017)	T498I	VP5*	Hydrophilic to hydrophobic Neutral charge to neutral charge	Possible change in the chemical properties
All the 2014 and 2015 strains	V600L	VP5*	Hydrophobic to hydrophobic Neutral charge to neutral charge	No change in the physicochemical properties
	K616R		Hydrophilic to hydrophilic Positive charge to positive charge	No change in the physicochemical properties

The table indicates the aa differences that were observed in the VP4 non-neutralization epitope sites at positions 43 (in two post-vaccine G1 strains) and 46 (in four post-vaccine G1 strains) during post-vaccine period.

Analysis of the amino acid differences between the VP6 gene sequences of South African pre- and post-vaccine strains identified an aa difference Y353H in a 2017 post-vaccine strain, RVA/Human-wt/ZAF/MRC-DPRU15948/2017/G1P[8]. The aa mutation Y353H occurred in a conserved region of the VP6 that has been associated with trimerization and single-shelled particle formation (Tosser *et al*., 1992) (Table 5.5).

Table 5.6: Amino acid differences in South African VP6 post-vaccine RVA strains

Strain	Mutation	Region	Amino acid property change	Possible effect
Strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU15948/2017/G1P[8]	Y353H	Non-epitope	Hydrophilic to hydrophilic; Neutral charge to positive (10%)/neutral charge (90%)	The change in charge and destabilization effect might alter physicochemical properties of the protein

The table indicates the only aa difference that was observed in one post-vaccine I1 strain when comparison of South African pre and post-vaccine strains was performed.

5.3.3.5 Analysis of amino acid differences in VP1-VP3 and NSP1-NSP5 amino acid sequences

The number of amino acid differences that were identified after Rotarix® vaccine introduction in South African G1P[8] VP1–VP3 and NSP1–NSP5 amino acid sequences were 12, 6, 7, 11, 3, 4, 5, and 2, respectively (Tables 5.7-5.14). Most of the amino acid substitutions were conservative (Betts and Russell, 2003).

No of strains(s)	Mutation(s)	Region	Amino acid property change	Possible effect
All the 2010 strains	L167M	N-terminal domain	Hydrophobic to hydrophobic; Neutral charge to neutral charge	The polarity change may alter biochemical properties of the protein
	T352N	Polymerase domain (Finger subdomain)	Hydrophilic to hydrophilic; Neutral charge to Neutral charge	
	A1028S	C-terminal domain	Hydrophobic to hydrophilic Neutral charge to neutral charge	
Four of the 2011 strains	K164R	N-terminal domain	Hydrophilic to hydrophilic Positive charge to positive charge	The change in charge may alter biochemical properties of the protein
	I279V,V301A	N-terminal domain	Hydrophobic to hydrophobic Neutral charge to neutral charge	
	K898T	C-terminal domain	Hydrophilic to hydrophilic Positive charge to neutral charge	
All the 2014 and 2015 strains	K96R	N-terminal domain	Hydrophilic to hydrophobic Positive charge to positive charge	The change in charge and polarity may alter biochemical properties of the protein
	K164R			
	D286N		Hydrophilic to hydrophilic Negative charge to neutral charge	
Strain RVA/Human-wt/ZAF/MRC-DPRU15948/2017/G1P[8]	I231T	N-terminal domain	Hydrophobic to hydrophilic Neutral charge to neutral charge	The polarity change may alter biochemical properties of the protein
	P969L	C-terminal domain	Hydrophobic to hydrophobic Neutral charge to neutral charge	

Table 5.7: Amino acid differences in South African post-vaccine R1 RVA strains

The table indicates the aa differences that were observed during post-vaccine period in different South African VP1 RVA strains when compared to cognate regions in pre-vaccine strains.

No of strain (s)	Mutation(s)	Region	Amino acid property change	Possible effect
All the 2010 strains	R44K	Nucleic acid binding domain	Hydrophilic to hydrophilic; Positive charge to positive charge	No effect on the physicochemical properties of the protein
Four of the 2011 strains	N30E	Nucleic acid binding domain	Hydrophilic to hydrophilic Neutral charge to negative charge	The mutations may not alter the biochemical properties of the protein
	T32N		Hydrophilic to hydrophilic Neutral charge to neutral charge	
Strain RVA/Human-wt/ZAF/MRC-DPRU83/2011/G1P [8]	P14L	Nucleic acid binding domain	Hydrophobic to hydrophobic Neutral charge to neutral charge	The mutations may not alter the biochemical properties of the protein
	N510S	Principal domain	Hydrophilic to hydrophilic Neutral to neutral	
Two 2015 strains	P669S	Principal domain	Hydrophobic to hydrophilic Neutral charge to neutral charge	The change in polarity may alter the biochemical properties of the protein

Table 5.8: Amino acid differences in South African post-vaccine C1 RVA strains

The table indicates the aa differences that were observed during the post-vaccine period in different South African VP2 RVA strains when compared to cognate regions in pre-vaccine strains.

No of strain(s)	Mutation(s)	Region(s)	Amino acid property change	Possible effect
All the 2010 strains	I234V	N7-Methyltransferase	Hydrophobic to hydrophobic; Neutral charge to neutral charge	The mutation may not alter the biochemical properties of the protein
Four of the 2011 strains	A411V and M762	2'-O—Methyltransferase domain and phosphodiesterase domains, respectively	Hydrophobic to hydrophobic Neutral charge to neutral charge	The mutation may not alter the biochemical properties of the protein
All the 2014 and 2015 strains	V459I and M762	N7-Methyltransferase and phosphodiesterase domains, respectively	Hydrophobic to hydrophobic Neutral charge to neutral charge	The mutation may not alter the biochemical properties of the protein
Strain RVA/Human-wt/ZAF/MRC-DPRU15948/2017/ G1P[8]	K204R	N7-Methyltransferase domain	Hydrophilic to hydrophilic Positive charge to positive charge	The mutation may potentially not alter the biochemical properties of the protein
	N804S	Phosphodiesterase domain	Hydrophilic to hydrophilic Neutral charge to neutral charge	

Table 5.9: Amino acid differences in South African post-vaccine M1 RVA strains

The table indicates the aa acid differences that were observed during the post-vaccine period in different South African VP3 RVA strains when compared to cognate regions in pre-vaccine strains.

No. of strain(s)	Mutation(s)	Region	Amino acid property change	Possible effect
All the 2010 strains	T120I	Cytoskeleton binding domain	Hydrophilic to hydrophobic; Neutral charge to neutral charge	The change in polarity may alter the biochemical properties
	V266I	Undefined region	Hydrophobic to hydrophobic; Neutral charge to Neutral charge	
Four of the 2011 strains	V189I, V266I, I278V	Undefined region	Hydrophobic to hydrophobic Neutral charge to neutral charge	The mutation may not alter the biochemical properties
	I341V	IRF-binding domain		
	D381Y	IRF-binding domain	Hydrophilic to hydrophilic Negative charge to neutral charge	
	Y445H	IRF-binding domain	Hydrophilic to hydrophilic Neutral to positive (10%)/ neutral (90%)	
2013 Strain RVA/Human-wt/ZAF/MRC-DPRU2250/2013/ G1P[8]	Q132R	RING domain	Hydrophilic to hydrophilic Neutral charge to positive charge	The change in charge may alter the biochemical properties
All 2014 strain and one 2015 strain RVA/Human-wt/ZAF/MRC-DPRU4357/2015/ G1P[8]	R13K	RING domain	Hydrophobic to hydrophilic Positive charge to positive charge	The change in polarity may alter the biochemical properties
	A259T	Undefined region	Hydrophobic to hydrophilic Neutral charge to neutral charge	

Table 5.10: Amino acid differences in South African post-vaccine A1 RVA strains

The table indicates the aa differences that were observed during the post-vaccine period in different South African NSP1 RVA strains when compared to cognate regions in pre-vaccine strains.

Table 5.11: Amino acid differences in South African post-vaccine N1 RVA strains

No of strain(s)	Mutation(s)	Region	Amino acid property change	Possible effect
2010 strain RVA/Human-wt/ZAF/MRC-DPRU2061/2010/G1P[8]	E73K	N terminal domain	Hydrophilic to hydrophilic; Negative charge to positive charge	The change in charge may alter the biochemical properties
Four 2011 strains	V254P	C terminal domain	Hydrophobic to hydrophobic Neutral charge to neutral charge	The mutation may not alter the biochemical properties
2014 Strain RVA/Human-wt/ZAF/MRC-DPRU143/2014/G1P[8]	P299S	C terminal domain	Hydrophobic to hydrophilic Neutral charge to neutral charge	The change in charge may alter the functionality of the protein

The table indicates the aa differences that were observed during the post-vaccine period in different South African NSP2 RVA strains when compared to cognate regions in pre-vaccine strains.

Table 5.12: Amino acid differences in South African post-vaccine T1 RVA strains

No of strain(s)	Mutation (s)	Region	Amino acid property change	Possible effect
Two 2015 strains	S169F	Dimerization domain	Hydrophilic to hydrophobic; Neutral charge to neutral charge	The change in polarity may alter the biochemical properties
	N215S	C-terminal domain	Hydrophilic to hydrophilic; Neutral charge to neutral charge	
2014 strain RVA/Human-wt/ZAF/MRC-DPRU65/2014/G1P[8	Y204F	Dimerization domain	Hydrophilic to hydrophobic; Neutral charge to neutral charge	The change in polarity may alter the biochemical properties
	K256R	C-terminal domain	Hydrophilic to hydrophilic Positive charge to positive charge	

The table indicates the aa differences that were observed during the post-vaccine period in different South African NSP3 RVA strains when compared to cognate regions in pre-vaccine strains.

Table 5.13: Amino acid differences in South African post-vaccine E1 RVA strains

No of strain(s)	Mutation(s)	Region	Amino acid property change	Possible effect
All 2010 strains	V73I	N terminal hydrophobic domain 3 (H3)	Hydrophobic to hydrophobic; Neutral charge to neutral charge	The mutation may not alter the biochemical properties
Strain RVA/Human-wt/ZAF/MRC-DPRU83/2011/G1P[8]	D157E	Tubulin binding domain/AS-I	Hydrophilic to hydrophilic; Negative charge to negative charge	The mutation may not alter the biochemical properties
Strain RVA/Human-wt/ZAF/MRC-DPRU15948/2017/G1P[8]	V13I	N-terminal hydrophobic domain 1(H1)/AS-IV	Hydrophobic to hydrophobic; Neutral charge to neutral charge	The mutation may not alter the biochemical properties
	D157E	Tubulin binding domain/AS-I	Hydrophilic to hydrophilic; Negative charge to negative charge	The mutation may not alter the biochemical properties
	E167G	DLP-binding region/AS-I	Hydrophilic to hydrophobic Negative to neutral charge	The change in polarity may alter the biochemical properties

The table indicates the aa differences that were observed during the post-vaccine period in different South African NSP4 RVA strains when compared to cognate regions in pre-vaccine strains.

Table 5.14: Amino acid differences in South African post-vaccine H1 RVA strains

No of strain(s)	Mutation(s)	Region	Amino acid property change	Possible effect
Four 2010 strains	S22L	N-terminal binding domain	Hydrophilic to hydrophobic; Neutral charge to neutral charge	The change in polarity may alter the biochemical properties
Four 2011 strains	D157E	C-terminal binding domain	Hydrophilic to hydrophilic; Positive charge to negative charge	The change in charge may alter the biochemical properties

The table indicates the aa differences that were observed during the post-vaccine period in different South African NSP3 RVA strains when compared to cognate regions in pre-vaccine strains.

5.3.3.6 Analyses of selection pressure

Selection pressure analysis by MEME, FUBAR and FEL showed that the most of the aa sites in the 11 gene segments were undergoing purifying selection except for the amino acid sites 7 and 245 in VP3 which were undergoing positive selection (Table 5.14; Appendix 9).

Table 5.15: Positively selected sites as identified by FEL, FUBAR and MEME analysis

Method	Amino acid sites in the gene segments under positive selection										
	VP1	**VP2**	**VP3**	**VP4**	**VP6**	**VP7**	**NSP1**	**NSP2**	**NSP3**	**NSP4**	**NSP5**
MEME	2,3,5,89, 243, 357, 733, 909, 1082	583	**7,** 118, 137**,** **245**,310, 728	23,28, 44,90, 169, 201, 576,578, 668, 773, 774, 775	85,238	-	14, 32, 90, 132, 154, 175, 181, 225, 253, 261,267, 293,392, 396,398, 404, 405, 416, 425	55,258,315	204	25,168	3
FUBAR	67, 357	12, 36	**7, 245**	-	-	-	-	75	308	-	-
FEL	3	-	**7, 245**	-	-	-	-	-	-	-	-

Bold and underlined sites are those identified by all the three methods (FEL, FUBAR and MEME) and these were considered for positive selection. For MEME and FEL, statistical significance was assessed at $p \leq 0.1$ while for FUBAR it was assessed at posterior probability ≥ 0.9. The dash (-) sign indicates no site was identified. Additional analyses reports are provided in Appendix 9.

5.4 Discussion

The present study reports whole-genome analysis of G1P[8] RVA strains collected seven years before and seven years after the introduction of the Rotarix® vaccine in South Africa. Whole-genome analysis of South African G1P[8] strains showed that they possessed a Wa-like genotype constellation. These findings are comparable to what has been observed in several countries (Le *et al*., 2010; Rahaman *et al*., 2010; Banyai *et al*., 2012; Shintani *et al*., 2012; Santos *et al*., 2019). The prevalent association of G1P[8] genotypes with the Wa-like genetic background has been hypothesized to be due to epidemiological fitness (Ghosh and Kobayashi, 2011).

Of the seven established G1 lineages (Le *et al*., 2006), South African pre- and post-vaccine G1 strains in this study segregated into G1-lineage I and G1-lineage II. Similar observations have been reported in several other studies (Le *et al*., 2010; da Silva *et al*., 2015; Santos *et al*., 2019) highlighting the likelihood of epidemiological fitness of G1 strains in these two lineages which has led to their prevalent circulation over time. Of the four P[8] lineages that have been established previously (Le *et al*., 2010), majority (84.2%) of the South African P[8] strains in this study clustered in P[8] lineage III, an observation comparable to other studies (da Silva *et al*., 2015; Dulgheroff *et al*., 2016; Almeida *et al*., 2017; Santos *et al*., 2019). Additionally, the South African P[8] strains were observed in P[8] lineage I and P[8] lineage IV (also designated as OP354 like lineage) (Cunliffe *et al*., 2001), an observation reported in a study of Indian P[8] strains (Tatte and Chitambar, 2011). Both pre-and post-vaccine strains were found present in this P[8] lineage III underscoring its global predominance in circulation (Gouvea *et al*., 1999; Cunliffe *et al*., 2001; Arista *et al*., 2006; Ansaldi *et al*., 2007; Espinola *et al*., 2008; Tatte and Chitambar, 2011).

The aa differences in the neutralization epitopes of VP7 compared to the cognate region in Rotarix® identified at positions N94S, S123N, K291R, and M217T appeared in both pre and post-vaccine strains and have been previously reported in Belgium and Brazil (Zeller *et al*., 2012; Santos *et al*., 2019). Since these amino acid substitutions were present even before the introduction of the vaccine, they can be attributed to the natural RVA evolutionary processes. The N147D amino acid substitution occurred in eight post-vaccine G1 strains and resulted in a change from neutral to negatively charged amino acid residue while T24A occurred in one post-

vaccine G1 strain which resulted in change from hydrophilic to hydrophobic residue (Betts and Russell, 2003). The change in charge may affect the protein's chemical properties, while the shift in polarity suggests possible inaccessibility of the epitope as it becomes more hydrophobic (Betts and Russell, 2003). Differences in VP4 amino acids between the study strains and Rotarix® were observed in epitopes 8-1 and 8-3, similar to those already reported in Belgium, Brazil and Tunisia (Zeller *et al*., 2012; Fredj *et al*., 2013; Santos *et al*., 2019). The polarity changes (hydrophilic to hydrophobic) observed at position T88I only in a post-vaccine strain may affect antibody binding at this region of the 8-4 epitope as the resulting epitope becomes relatively inaccessible due to hydrophobic effect (Betts and Russell, 2003). The loss of glycosylation site at position N89S only observed in post-vaccine period in one strain may alter the functioning of the protein at the 8-4 epitope (Caust *et al*., 1987).

To gain further insight into the impact of aa substitution occurring in the neutralization epitopes of VP7 and VP4 on the stability of the protein structures, we performed folding free energy change analysis. A protein's folding free energy change is an essential aspect of the protein's stability with a direct association with the protein's function (Zhang *et al*., 2012). The N147D substitution in VP7 and N89S substitution in VP8* resulted in a predicted destabilizing effect on the protein structure. Due to this destabilization effect on the protein structure as indicated by energy change analysis, we hypothesize that these amino acid substitutions observed during post-vaccine era may not be favored in the longterm to enhance viral fitness. We did not observe consistently occurring amino acid substitutions in the 11 gene segments both in antigenic and non-antigenic regions that occurred in strains throughout the post-vaccine period when compared to pre-vaccine period. However, there were yearly observations such as L167M (in VP1), R44K (in VP2), I234V (in VP3), V112I (in VP4), T120I (in NSP1) and V73I (in NSP4) that occurred in all the eight 2010 strains. Amino acid substitutions: K164R (in VP2), A411V (in VP3), A46T (in VP7), I341V (in NSP1), V254P (in NSP2) and D157E (in NSP5) of the four out of five 2011 strains and K96R (in VP1), V459I (in VP3), V600L (in VP4), and A259T (in NSP1) of all the eight 2014 and 2015 G1P[8] strains.

Selection pressure analysis demonstrated that most of the codon sites in the 11 genome segments of South African G1P[8] strains were undergoing purifying selection probably to purge deleterious polymorphisms that arise due to the inherent error-prone nature of the RNA polymerase enzyme (Estes and Greenberg, 2013). Amino acid site 7 and 245 in VP3 was under positive selection and site 245 fell within the guanine-N7-methyltransferase (N7-MTpase) domain suggested to catalyze methylation during capping of nascent rotavirus transcripts (Ogden *et al.*, 2014). Methylation capping by VP3 of RVA serves as a critical strategy to evade host antiviral innate immune response (Ogden *et al*., 2014).

5.5 Conclusions

The study shows that there were no consistently conserved amino acid substitutions occurring throughout during post-RVA vaccine period in the antigenic and non-antigenic regions of South African G1P[8] strains as most of the amino acid substitutions found in post-vaccine period were already present during pre-vaccine period. Therefore, Rotarix® did not appear to have an impact on the genetic changes of South African G1P[8] post-vaccine strains. However, continued long-term whole-genome surveillance to monitor any consistently occurring genetic changes during post-vaccine period, which may hint at vaccine-induced selective pressure, remains essential. This study was limited mainly by the significant differences in the sample size between pre and post-vaccine period, although the significant decline in G1P[8] strains in South Africa during post-vaccination period has been reported before (Page *et al*., 2018).

CHAPTER SIX

INVESTIGATION OF WHOLE-GENOME AND PHYLODYNAMICS OF SOUTH AFRICAN G2P[4] ROTAVIRUS STRAINS PRE- AND POST-VACCINATION INTRODUCTION

Chapter six is presented in a manuscript format entitled "Genetic variations between pre- and post-South African G2P[4] rotavirus A strains over 14 years (2002-2017)"

Mwangi, P. N., Mogotsi, M. T., Seheri, M.L., Mphahlele, M .J., Peenze, I., Esona, M. D., Kumwenda, B., Steele, A. D., Kirkwood, C. D., Ndze, V. N., Dennis, F. E. and Nyaga M. M. Whole genome analyses of South African G2P[4] strains pre-and post-rotavirus vaccine introduction.

6.1 Introduction

RVA is the primary etiological agent for AGE, accountable for an annual global mortality rate of approximately 125,000 in children under five years (Troeger *et al*., 2018). Rotavirus belongs to *Reoviridae* family and its genome comprises 11 gene segments that encode six structural proteins and five or depending on the species, six non-structural proteins (Estes and Greenberg, 2013).

Conventionally, RVA is classified into G and P-type as per the properties of the VP7 (denoted G for its glycoprotein property) and VP4 (denoted P for its property as a protease-sensitive protein) (Estes and Kapikian, 2007). The binary (G/P) classification system was extended to include the other nine gene segments to decipher the RVA diversity (Matthijnssens *et al*., 2011). Whole-genome classification provides the basis of the three established RVA genogroups: Wa-like, DS-1-like and AU-1-like. To date, the number of RVA genotypes are 36G (VP7), 51P (VP4), 26I (VP6), 22R (VP1), 20C (VP2), 20M (VP3), 31A (NSP1), 22N (NSP2), 22T (NSP3), 27E (NSP4) and 22H (NSP5) (https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification).

The most medically important RVA strains are G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8] (Doro *et al*., 2014). The G2P[4] is the second most prevalent RVA strain globally after G1P[8] (Doro *et al*., 2014). The epidemiological predominance of G2P[4] in certain geographical regions is well documented (Doan *et al*., 2011; Vizzi *et al*., 2017; Khandoker *et al*., 2018; Thanh *et al*., 2018). To combat RVA disease burden, four RVA vaccines: Rotarix® (GlaxoSmithKline, Rixenstart, Belgium), Rotasiil® (Serum Institute of India, India), RotaTeq® (Merck & Co, USA) and Rotavac® (Bharat Biotech, India) have been prequalified by WHO for global use (WHO, 2020). The

introduction of RVA vaccination has substantially led to a reduction of RVA mortality globally (Shah et *al.*, 2017).

South Africa introduced Rotarix® in 2009 into its EPI (WHO, 2009). Although RVA vaccination has reduced the burden of RVA disease, there are concerns of potential vaccine-induced selective pressure on the distribution of RVA strains (Leshem *et al*., 2014). For instance, there has been an upsurge of G2P[4] strains in several countries that have introduced Rotarix® (Gurgel *et al*., 2007; Donato, 2012, Vizzi *et al*., 2017; Zeller *et al*., 2017, Khandoker *et al*., 2018; Mokomane *et al*., 2019). It is unresolved whether the detection of non-G1P[8] strains such as G2P[4] is primarily vaccine-induced or due to the natural RVA evolutionary mechanisms. The G2P[4] evolution dynamics are further confounded by the fact that it is associated with cyclic peaks (Linhares and Justino, 2014). There have been outbreaks of G2P[4] infections every three to four years and significant epidemics every ten years in South Africa(Page and Steele, 2004). The epidemiological importance of this genotypical cycling is hypothesized to serve as a strategy for RVA to evade group immunity acquired from previous infections enabling the RVA to prevail in circulation (Parra, 2009). There is a dearth of whole-genome based analysis of G2P[4] RVA strains in South Africa. This study analyzed South African G2P[4] samples collected seven years pre- and seven years post-vaccine introduction to elucidate its genetic diversity and phylodynamics.

6.2 Materials and methods

6.2.1 Ethics statement

Diarrhoeal stool samples conventionally genotyped as part of WHO-coordinated RVA surveillance were retrieved from the archival storage of the Diarrhoeal Pathogens Research Unit (DPRU), Sefako Makgatho Health Sciences University (SMU) and Center for Enteric Diseases, National Institute for Communicable Diseases (NICD), South Africa. Further processing of the stool samples for whole-genome sequencing and data analysis was performed at the University of the Free State (UFS), Bloemfontein South Africa facilitated by material transfer agreements (SMU-UFS.1 and NICD-UFS.2). The study was approved by the HSREC of the UFS under ethics number UFS-HSD2018/0510/3107 (Appendix 5).

6.2.2 Strain description

A total of 98 rotavirus positive stool samples (59 retrieved from Center for Enteric Diseases, NICD, Johannesburg and 39 retrieved from DPRU, Pretoria) previously characterized as G2P[4] were distributed as follows: 2003 (n=6), 2006(n=1), 2007(n=9), 2008(n=13), 2009(n=13), 2010 (n=10), 2012 (n=11), 2013(n=15), 2014 (n=12), 2015 (n=4), 2016 (n=3) and 2017 (n=1). An additional five whole-genome sequences for G2P[4] strains collected from South Africa were mined from the GenBank database.

6.2.3 Rotavirus RNA extraction and purification

Extraction of rotaviral dsRNA was performed as previously described by Mwangi *et al*. (2020). The extracted RNA was subsequently purified using the MinElute PCR purification kit (Qiagen, Hilden, Germany).

6.2.4 Synthesis of complementary DNA (cDNA)

cDNA synthesis was performed using the Maxima H Minus double-stranded synthesis Kit (Thermo Fischer Scientific, Waltham, MA). Briefly, we introduced a 5 minute denaturing step at 95^{o}C that was performed in the thermocyler. A 1µl volume of Random Hexamer primer, 100µM was then added. The reaction mixture was incubated in a thermocyler at 65^{o}C for 5 minutes. Afterward, 5µl volume of the 4X First-Strand Reaction Mix and 1µl of First Strand Enzyme Mix was added and the reaction mixture incubated in a thermocyler pre-programmed as follows (10 minutes at 25^{o}C, 120 minutes at 50^{o}C, 5 minutes at 85^{o}C). For second-strand cDNA synthesis step, a 55µl volume of Nuclease-Free water, 20µl of 5X Second Strand Reaction Mix and 5µl of Second Strand Enzyme Mix was added and incubated in the thermocyler at 16^{o}C for 60 minutes. A 6µl volume of EDTA, pH 8.9 was added followed by 10µl of RNase I. The synthesized cDNA was purified using MSB® Spin PCRapace Kit by following manufacturer's protocol (Stratec Molecular, Berlin, Germany).

6.2.5 DNA library preparation and whole-genome sequencing

The Nextera® XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) was used to prepare the DNA libraries. Briefly, a 5µl volume of purified cDNA within the 0.2ng/µl concentration range was pipetted to each well of UltraFlux® 96-well PCR plate (Scientific Specialties Inc, Lodi, CA, USA).

Then 10µl of Tagment DNA buffer was added to each well of the PCR plate followed by 5µl of Ampilicon Tagment Mix. The reaction mixture was centrifuged at 280 *x g* at room temperature for 1 minute followed by incubation in a thermocyler programed at 55°C for 5 min with a 10°C hold temperature. A 5µl volume of Neutralize Tagment buffer was added and the reaction mixture was centrifuged (Hermle Labortechnik, Wehingen, Germany) at 280 *x g* at room temperature for 1 minute followed by 5 minutes incubation at room temperature. A 5µl volume of index 1 and index 2 primers (Nextera® XT index Kit) was added followed by 15µl volume of Nextera® PCR Mastermix and the reaction mixture was centrifuged at 280 x g at 20°C for 1 minute. Afterward, incubation was performed on a thermocyler programmed with pre-heated lid option, 72°C for 3 minutes and 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds then 72°C for 5 minutes and 10°C hold temperature. The post-tagmented and amplified DNA was cleaned up using AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA) and freshly prepared 80% ethanol (Sigma-Aldrich®, St Louis, MO, USA).

6.2.6 Genome assembly

The FASTQ raw data from the Illumina MiSeq® was subjected to both *de novo* and reference-based rotavirus genome assembly by using Geneious Prime® 2020.1.1 (Kearse *et al*., 2012). The reference mapping and *de novo* tools integrated into the Geneious Prime were utilized for subsequent assembly process. An in-house designed genome assembly pieline and CLC Genomics Workbench (https://www.qiagenbioinfomatics.com/), were utilized as complementary genome assembly resources.

6.2.7 Generation of whole-genome constellations

The determination of whole-genome constellations was performed using the Virus Pathogen Database and Analysis Resource (ViPR) (http://www.vipbrc.org)).

6.2.8 Phylogenetic analysis

Phylogenetic analysis was performed as described previously (Mwangi *et al*., 2020). Briefly, sequence alignments were performed using the MAFFT tool in Geneious Prime 2020 (Kearse *et al*., 2012). The best evolutionary model was estimated using the Model Test in MEGA6 (Tamura

et al., 2013). Maximum-likelihood phylogenetic trees with a 1000 bootstrap support were constructed using MEGA 6 (Tamura *et al*., 2013

6.2.9 Analysis of selection pressure

The suite of tools in the DataMonkey webserver were utilized to analyse for selective pressure (Weaver *et al*., 2018). The tools utilized were: Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR) and Mixed Effects Model of Evolution (MEME) (Kosakovsky *et al.*, 2005; Murrell *et al*., 2012; Murrell *et al*., 2013).

6.3 Results

6.3.1 Whole-genome constellations

The 103 G2P[4] strains analysed in this study had a DS-1-like genotype constellation (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) (Appendix 10). The full-genome lengths and their respective open reading frames (ORFs) were determined (Appendix 10).

6.3.2 Phylogenetic analyses

6.3.2.1 Phylogenetic analysis of VP7

The VP7 phylogenetic tree was constructed using the four lineage designations previously described by Doan *et al*. (2011). South African G2 strains clustered in lineage IV and within this lineage, the G2 strains segregated further into sub-lineages (Figure 6.1). The nt (aa) identities of South African G2 pre-vaccine strains amongst each other ranged from 95.5-100% (96.3-100%) while the similarities amongst post-vaccine strains ranged from 94.0-100% (97.2-100%). The nt (aa) comparison of pre and post-vaccine G2 strains was 95.0-100% (96.0-100%).

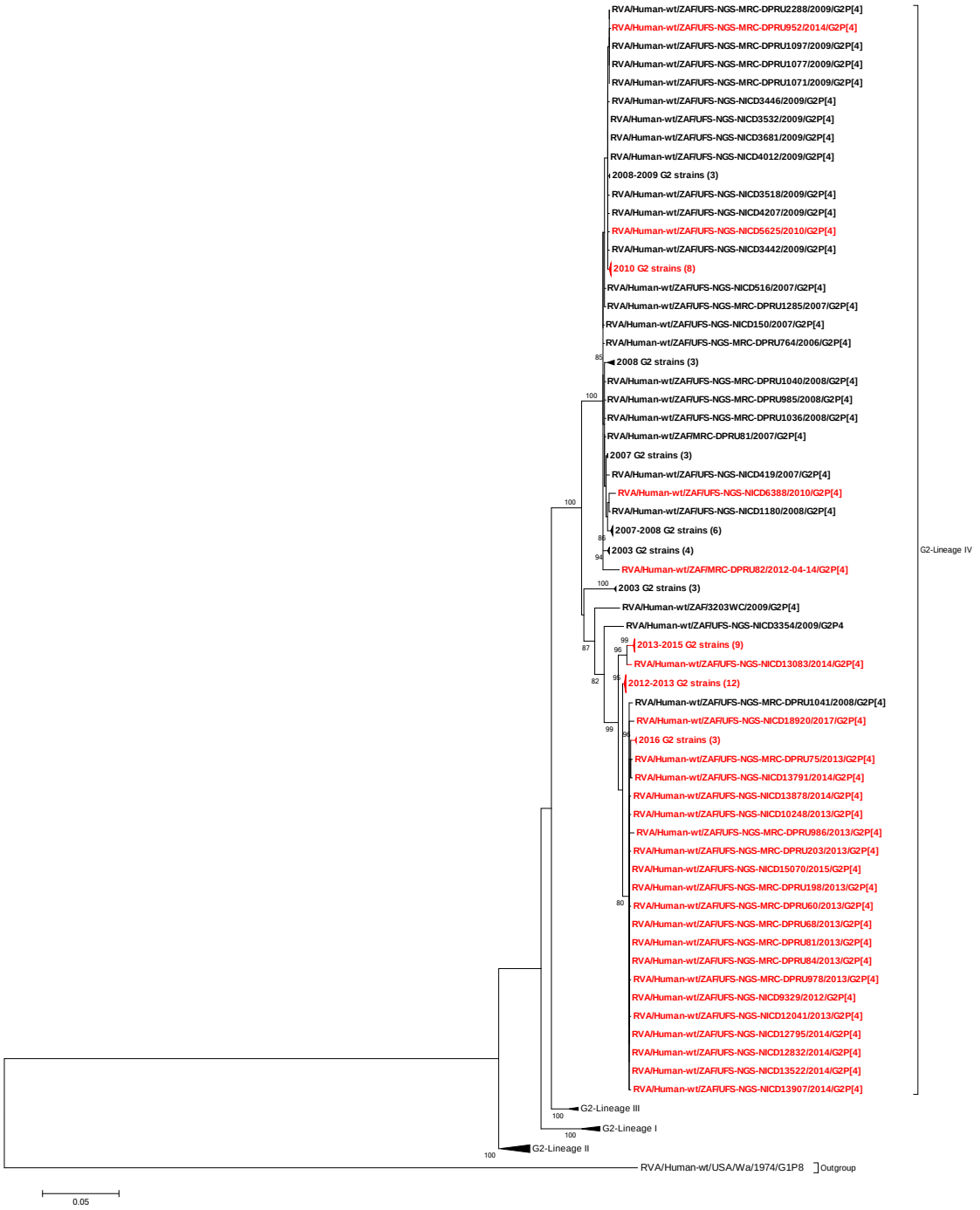


Figure 6.1: VP7 phylogenetic tree of South African G2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP7 gene segment.South African pre-vaccine G2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

6.3.2.2 Phylogenetic analysis of VP4

Of the four P[4] lineages described by Doan *et al.* (2011), South African P[4] strains clustered into lineage II and lineage IV (Figure 6.2). Lineage II comprised three strains that circulated in 2003 while the rest (100) of P[4] strains clustered in lineage IV (Figure 6.2). Within this lineage IV, South African P[4] strains segregated further into sub-lineages (Figure 6.2). The identities of P[4] strains amongst each other ranged from 92.6%-100% (95.6%-100%) while amongst post-vaccine strains it ranged from 92.5%-100% (95.5%-100%). The comparison of pre and post-vaccination P[4] strains was in the range of 92.6%-100% (95.5%-100%).

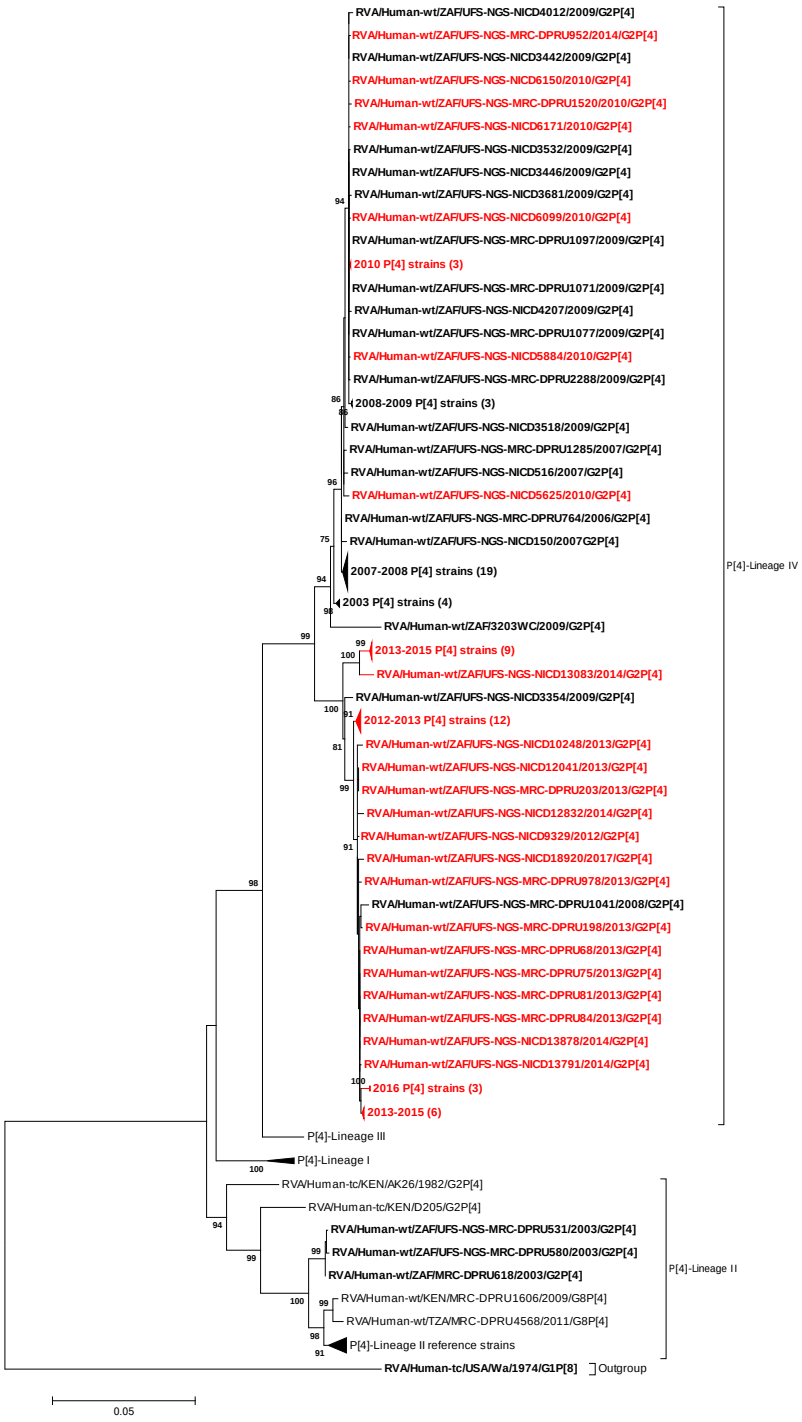


Figure 6.2: VP4 phylogenetic tree of South African P[4] strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP4 gene segment. South African pre-vaccine P[4] strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

6.3.2.3 Comparative analysis of neutralizing antigenic VP7 epitopes of South African post-vaccine G2 strains with pre-vaccine G2 strains

We compared VP7 neutralization epitopes previously described by Aoki *et al*. (2009) from post-vaccine VP7 genes with pre-vaccine VP7 genes. The aa profile between the pre- and post-vaccine G2 strains was maintained across the 26 of the 29 aa residues that span the VP7 neutralization epitopes with exception of three aa differences: T87S (identified in one post-vaccine strain), N146D (in one strain) and V212M (in one strain) (Appendix 11).

6.3.2.4 Comparative analysis of non-antigenic sites of South African post-vaccine G2 strains with pre-vaccine G2 strains

The following aa differences were identified when aa composition of post-vaccine G2 strains outside non-antigenic regions was performed against cognate regions in pre-vaccine G2 strains: L18V (in three strains), L29I (in one strain), A42V/T (in three strains), F47L (in two strains), T50A (in eight strains), V/A66I (in nine strains), S72G/N (in two strains), I246M (in one strain) and P254S (in one strain). At aa sites 15, 37, 139 and 242, serine, isoleucine, isoleucine and asparagine, respectively, occurred at a significantly higher proportion (60-80%) in post-vaccine strains compared to a negligible proportion (of 1-2%) in pre-vaccine strains (Table 6.1).

Table 6.1 Amino acid sites with significant amino acid differences between pre-and post-vaccine G2 strains

Representative G2 strains	Amino acid sites			
	15	37	139	242
Pre-vaccine South African G2 strains				
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	F	F	V	S
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4](2)	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	.	.	N
Post-vaccine South African G2 strains				
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	.	.	N
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	S	I	I	N
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	S	I	I	N
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	S	I	I	N
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	S	I	I	N
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	S	I	I	N
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	S	I	I	N

A strain per year was randomly selected to represent G2 strains that circulated in that particular year. South African pre-vaccine G2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre- and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.5 Comparative analysis of neutralizing antigenic VP4 epitopes of South African post-vaccine P[4] strains with pre-vaccine P[4] strains

We compared aa profile of South African post-vaccine VP4 strains in previously described epitope regions (Dormitzer *et al*., 2002) with cognate regions in South African pre-vaccine P[4] strains. The amino acid differences identified were D192Y (in three strains), T115I (in one strain) and S133N (in one strain) (Appendix 12).

6.3.2.6 Comparative analysis of non-antigenic sites of South African post-vaccine P[4] strains with pre-vaccine P[4] strains

The following aa differences were observed when South African post-vaccine P[4] strains were compared with P[4] pre-vaccine strains: R581K (in three strains), S596N (in 10 strains), D606N (in 10 strains), V597I (in seven strains) and S631N (in four strains).

At aa sites 130, 147, 162 and 607, isoleucine, asparagine, arginine and isoleucine, respectively occurred at a significantly higher frequency (60-80%) in post-vaccine strains compared to a negligible proportion (of 1-2%) in pre-vaccine strains (Table 6.2).

Table 6.2 Amino acid sites with significant amino acid differences between pre-and post-vaccine P[4] strains

Representative P[4] strains	Amino acid sites			
	130	147	162	607
Pre-vaccine South African P[4] strains				
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	**V**	**S**	**G**	**V**
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	.	.	.
Post-vaccine South African P[4] strains				
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	**I**	**N**	**R**	**I**
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	**I**	**N**	**R**	**I**
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	**I**	**N**	**R**	**I**
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	**I**	**N**	**R**	**I**
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	**I**	**N**	**R**	**I**
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	**I**	**N**	**R**	**I**

A strain per year was randomly selected to represent P[4] strains that circulated in that particular year. South African pre-vaccine P[4] strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre- and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.7 Phylogenetic analysis of VP1-VP3 and VP6

Phylogenetic trees of South African VP1-VP3 and VP6 genes exhibited a consistent clustering pattern whereby most of post-vaccine strains (2012-2017), with exception of 2010 strains, clustered distinctly away from pre-vaccine strains that circulated from 2003-2009 (Figures 6.3-6.6). The comparison of pre and post-vaccine R2, C2, M2 and I2 genes ranged from nt (aa) 91.0-

100.0% (98.2-100%), 96.2-99.8% (98.4-100%), 87.1-100% (92.3-100%) and 96.3-100% (98.5-100%), respectively.

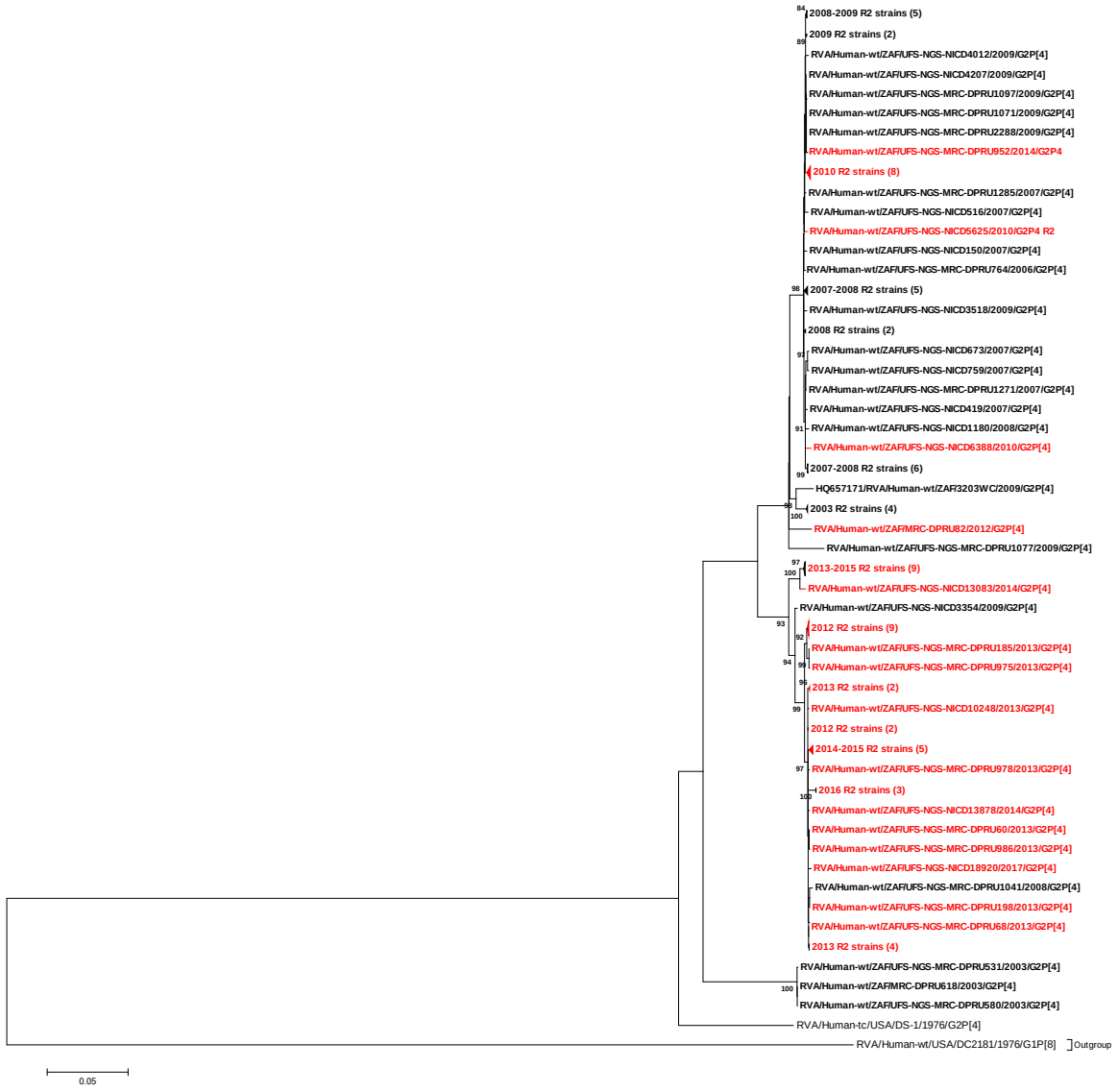


Figure 6.3: VP1 phylogenetic tree of South African R2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP1 gene segment. South African pre-vaccine R2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

Figure 6.4: VP2 phylogenetic tree of South African C2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP2 gene segment. South African pre-vaccine C2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

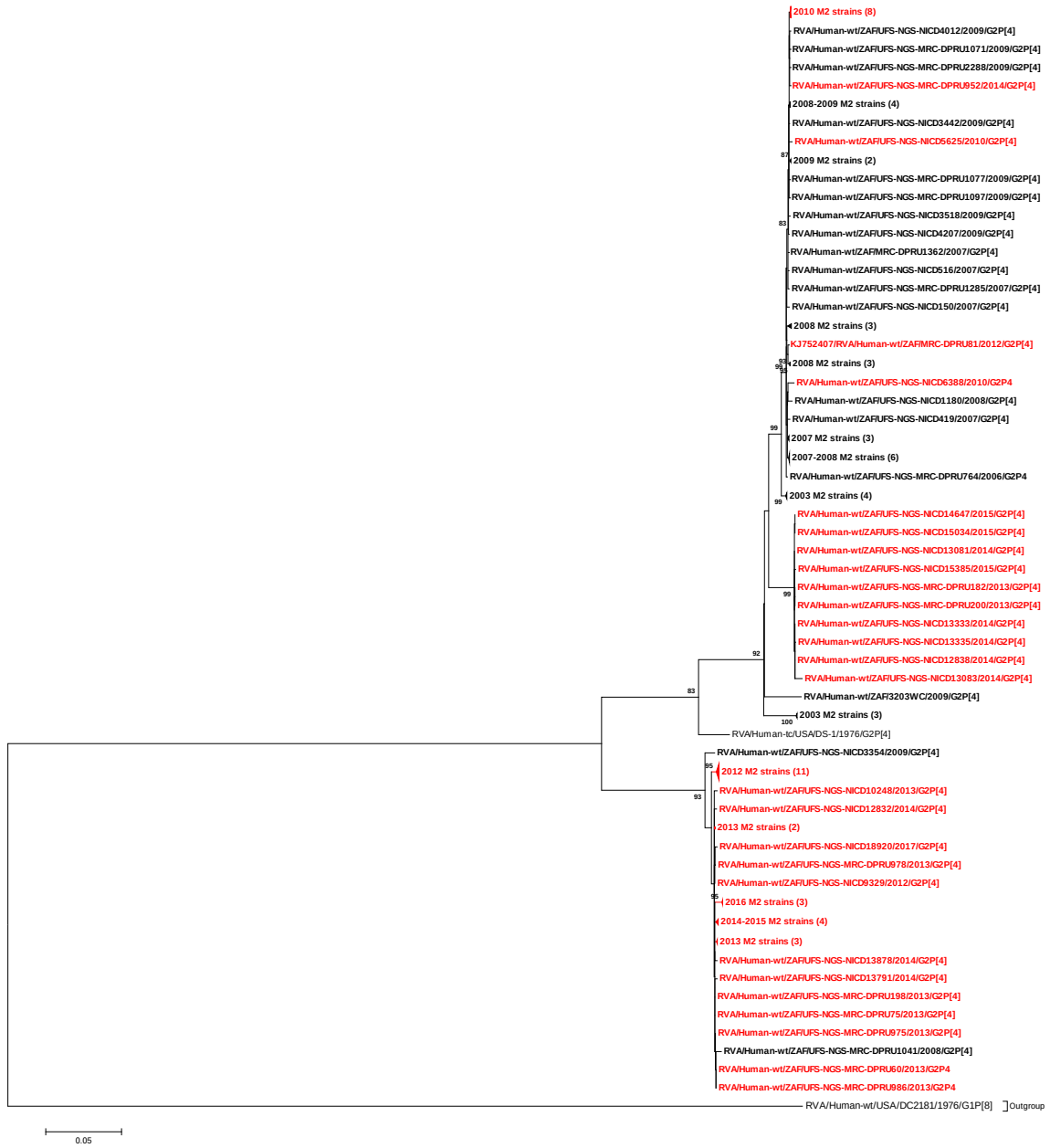


Figure 6.5: VP3 phylogenetic tree of South African M2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP3 gene segment. South African pre-vaccine M2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

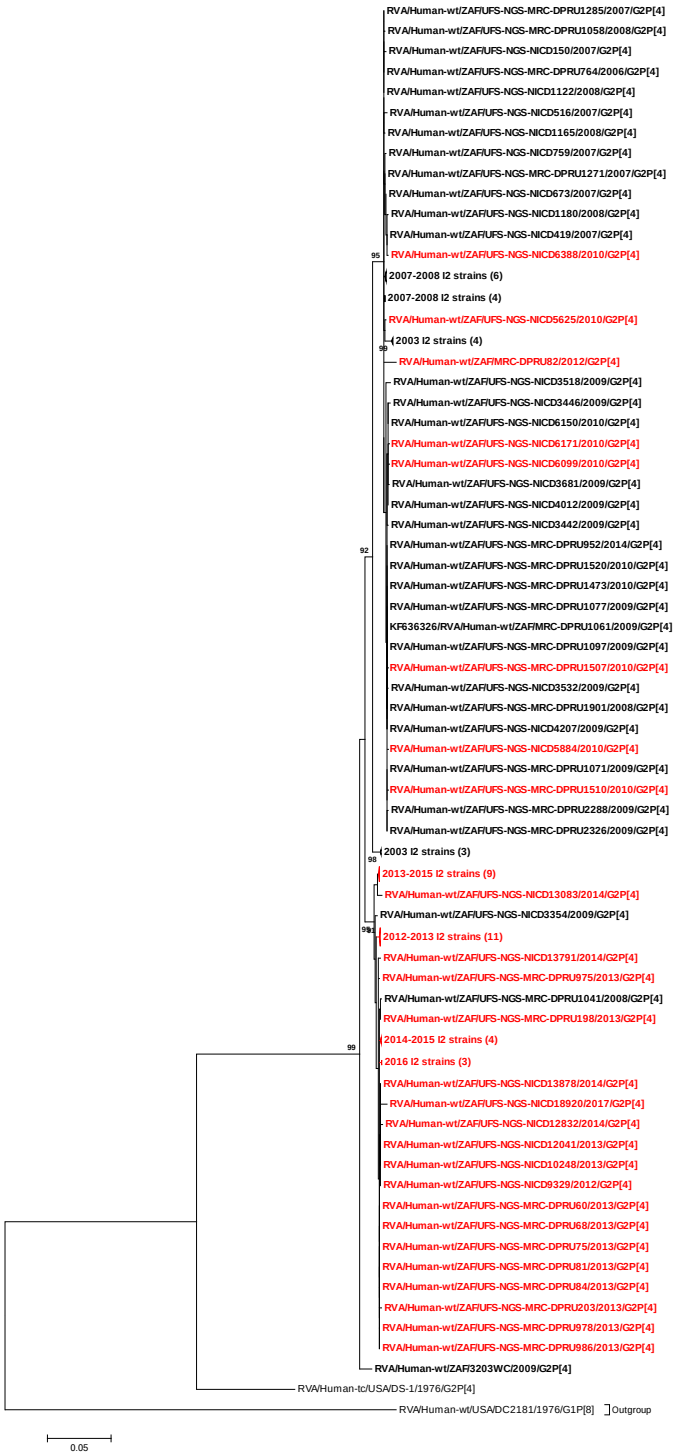


Figure 6.6: VP6 phylogenetic tree of South African I2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the VP6 gene segment. South African pre-vaccine I2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

6.3.2.8 Analysis of amino acid differences in VP1 sequences

The following aa differences were observed when South African post-vaccine R2 strains were compared with R2 pre-vaccine strains: A187T (in five strains), V332I (in nine strains), I482T (in three strains), A546S (in four strains) and I911V (in 11 strains). At aa sites 56, 159, 293, 294, 432 and 888, arginine, arginine, aspartate, asparagine, lysine and asparagine, respectively occurred at a significantly higher proportion (60-80%) in post-vaccine strains compared to an insignificant percentage (of 1-2%) in pre-vaccine strains (Table 6.3).

Table 6.3 Amino acid sites with significant amino acid differences between pre-and post-vaccine R2 strains

Representative R2 strains	Amino acid sites					
	56	159	293	294	432	888
Pre-vaccine South African R2 strains						
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	K	K	N	S	R	S
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU985/2008/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1036/2008/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1040/2008/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	.	.	.	.	.
Post-vaccine South African R2 strains						
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	R	R	D	N	K	N
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	R	R	D	N	K	N
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	R	R	D	N	K	N
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	R	R	D	N	K	N
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	R	R	D	N	K	N
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	R	R	D	N	K	N

A strain per year was randomly selected to represent R2 strains that circulated in that particular year. South African pre-vaccine R2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.9 Analysis of amino acid differences in VP2 sequences

The aa difference Q481H in three post-vaccine strains was observed when post-vaccine C2 strains were compared with C2 pre-vaccine strains.

6.3.2.10 Analysis of amino acid differences in VP3 sequences

When we compared post-vaccine VP3 sequences with pre-vaccine ones we identified aa differences at 14 aa sites. These sites were V19M (in 11 strains), K/Q68N (in 10 strains), T125A (in three strains), V206I (in 10 strains), Q227H (in 11 strains), I32M (in 10 strains, L331F (in 11 strains), T/N347P (in 10 strains), D414N (in 10 strains), I444L (in 10 strains), T517A (in 10 strains), I601M (in 10 strains), I/T750V (in 10 strains) and D/G828N(in 10 strains). A significantly higher occurrence (~80%) of aa residues (tabulated below) during post-vaccine period as compared to ~ 4% occurrence during pre-vaccine period was observed.

Table 6.4 Amino acid sites with significant amino acid differences between pre-and post-vaccine M2 strains

Representative M2 strains	Amino acid sites																								
	28	49	50	69	88	107	109	132	143	161	167	204	209	222	226	274	275	277	281	282	316	326	327	347	348
Pre-vaccine South African M2 strains																									
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	E	N	V	I	V	H	I	S	R	D	S	D	I	Y	R	N	I	H	I	V	N	N	D	T	V
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Post-vaccine South African M2 strains																									
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	M	R	V	I	D	D	N	N	M
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	M	R	V	I	D	D	N	N	M
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	M	R	V	I	D	D	N	N	M
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	M	R	V	I	D	D	N	N	M
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	.	R	.	I	D	D	N	N	M
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	D	S	I	M	I	Y	V	N	K	N	N	S	V	H	K	D	M	R	V	I	D	D	N	N	M

Pre-vaccine South African M2 strains	355	356	359	363	379	412	425	437	458	459	468	518	605	687	688	703	730	734	736	740	749	767	798	816
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	K	M	R	D	V	V	I	T	I	I	L	N	V	K	V	V	T	V	T	R	E	I	V	Y
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.
Post-vaccine South African M2 strains																								
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	M	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	R	V	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	K	D	V	I	H
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	R	V	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	K	D	V	I	H
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	R	V	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	K	D	V	I	H
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	R	V	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	K	D	V	I	H
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	R	I	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	.	D	V	I	H
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	R	V	E	S	I	I	V	I	V	T	F	S	I	R	A	I	I	A	A	K	D	V	I	H

A strain per year was randomly selected to represent M2 strains that circulated in that particular year. South African pre-vaccine R2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre- and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.11 Amino acid differences in VP6 sequences

When compared against the VP6 gene sequence of pre-vaccine strains, aa difference M342T was identified in nine post-vaccine strains. Isoleucine at position 281 occurred only in only one out of 46 pre-vaccine strains while it occurred in 35 out of 57 post-vaccine strains.

6.3.2.12 Phylogenetic analysis of NSP1-NSP5

Phylogenetic trees of South African NSP1-NSP5 genes exhibited a consistent clustering pattern whereby most of post-vaccine strains (2012-2017), with exception of 2010 strains, clustered distinctly away from pre-vaccine strains that circulated from 2003-2009 (Figures 6.7-6.11). The comparison of pre and post-vaccine A2, N2, T2,E2 and H2 genes ranged from nt (aa) 94.9-100%(95.9-100%), 95.6-100%(97.2-100%), 95.8-100%(96.1-100%), 84.4-100% (88.0-100%) and 94.3-100% (95.5-100%), respectively.

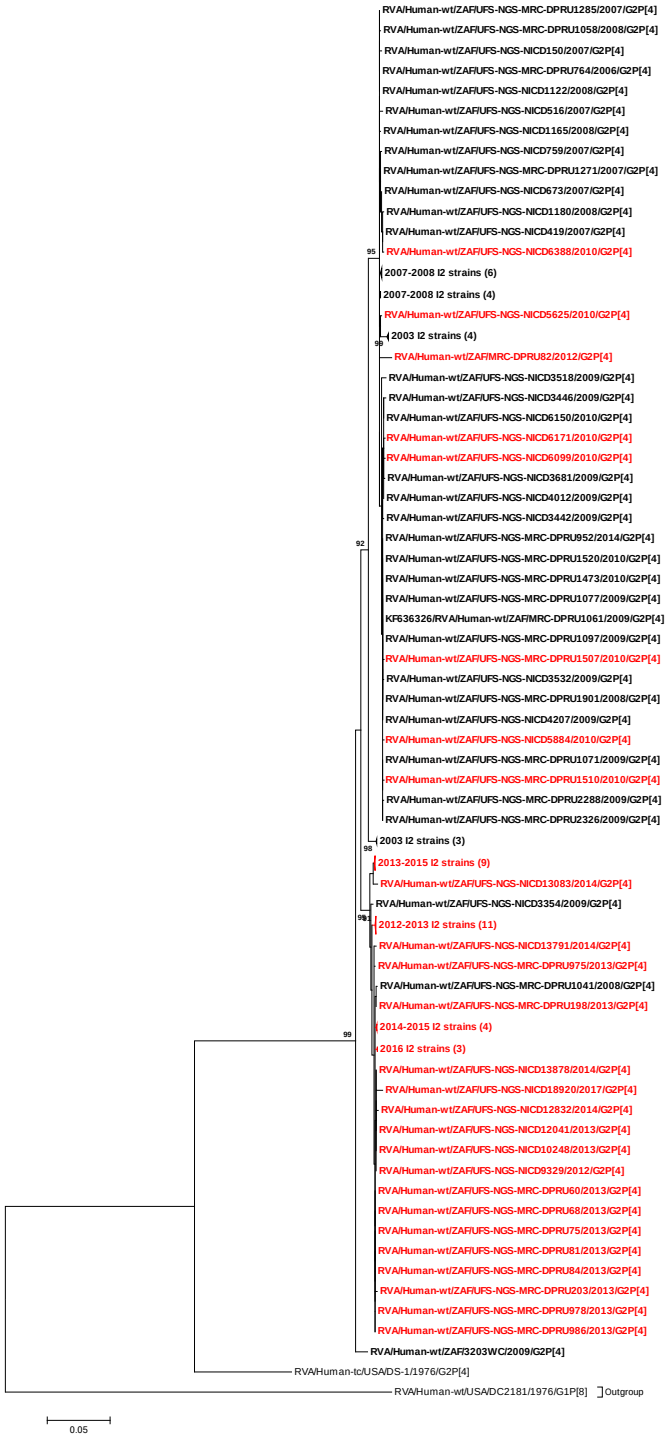


Figure 6.7: NSP1 phylogenetic tree of South African A2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP1 gene segment. South African pre-vaccine A2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

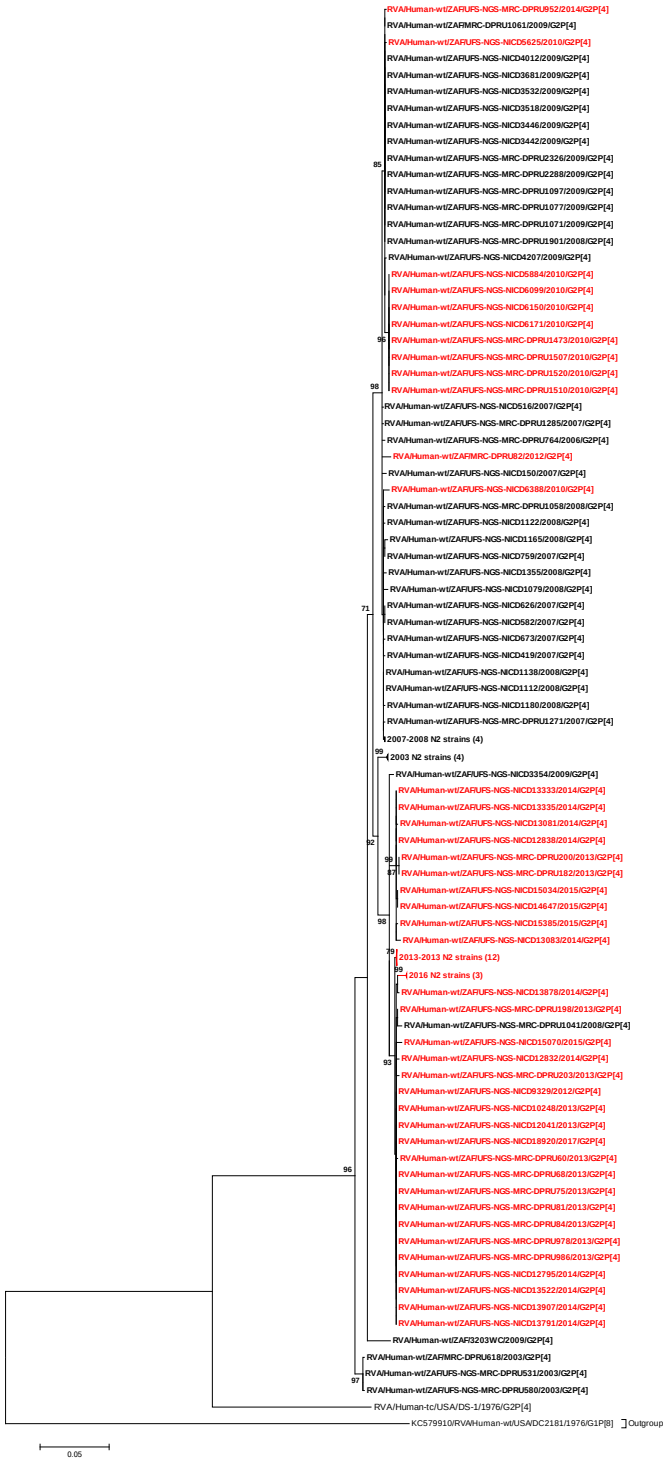


Figure 6.8: NSP2 phylogenetic tree of South African N2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP2 gene segment. South African pre-vaccine N2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

Figure 6.9: NSP3 phylogenetic tree of South African T2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP3 gene segment. South African pre-vaccine T2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

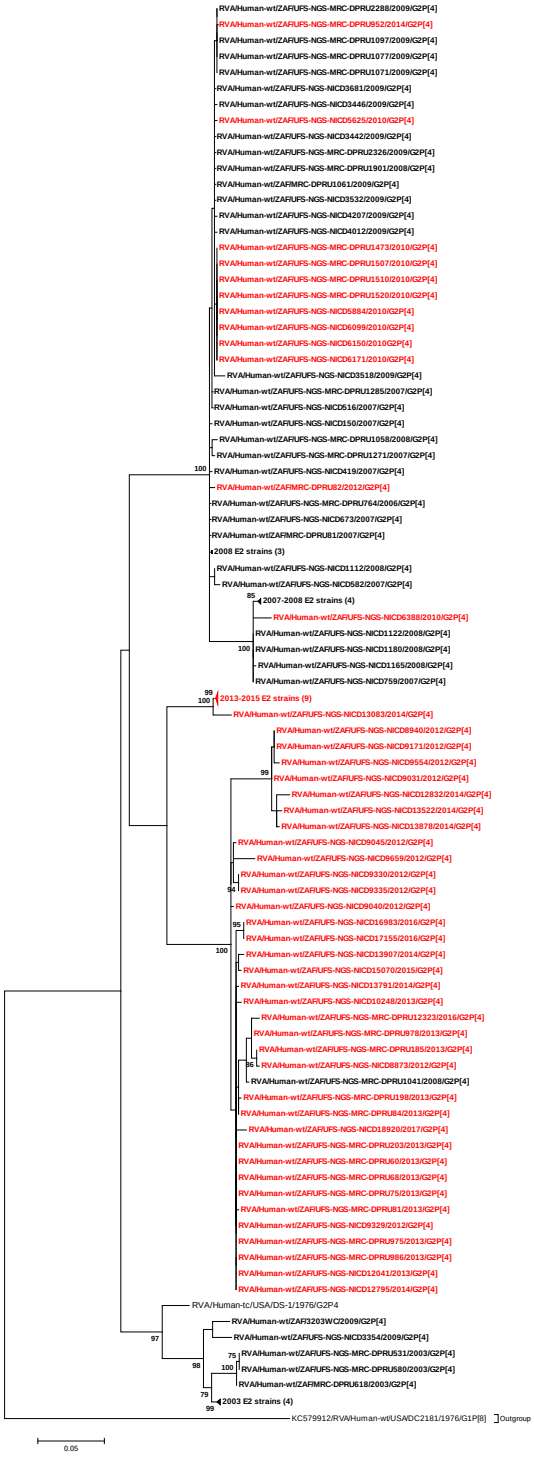


Figure 6.10: NSP4 phylogenetic tree of South African E2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP4 gene segment. South African pre-vaccine E2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

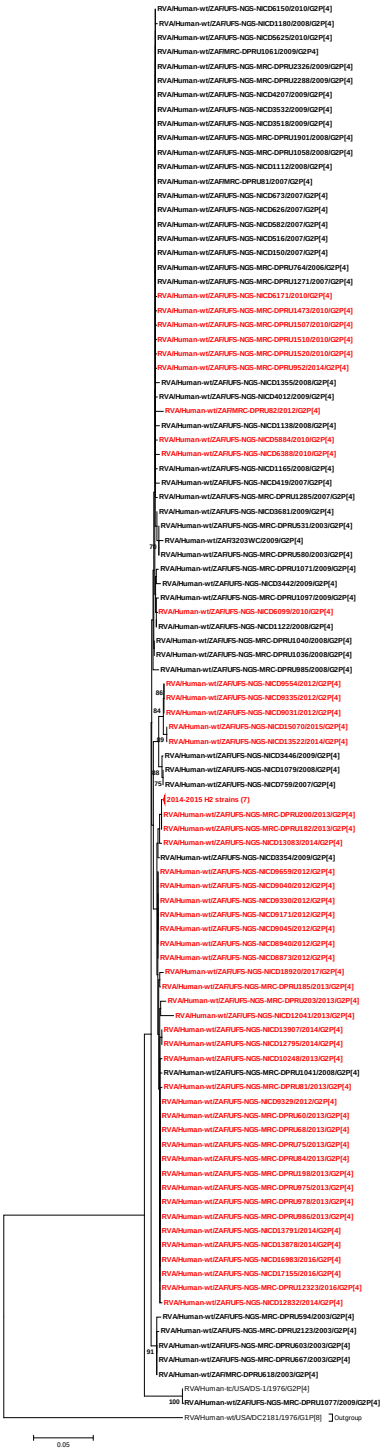


Figure 6.11: NSP5 phylogenetic tree of South African H2 strains

Maximum likelihood phylogenetic tree as per the complete nucleotide sequences of the NSP5 gene segment. South African pre-vaccine H2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The number in brackets denotes the number of compressed strains. Lineages are indicated in roman numerals. Percent bootstrap values (70% or higher) are indicated at each branch node. The genetic distance is indicated by the scale bar.

6.3.2.13 Amino acid differences in NSP1 sequences

When South African post-vaccine A2 strains were compared with cognate pre-vaccine A2 strains, aa differences Q10P (in six post-vaccine strains) and N76D (in nine post-vaccine strains) were identified. At aa sites 103, 234, 246, 273, 319, 333, 334, 429 and 464, aspartate, isoleucine, leucine, phenylalanine, arginine, leucine, asapargine, arginine and threonine, respectively occurred at a significantly higher proportion (60-80%) in post-vaccine strains compared to an insignificant percentage (of 1-2%) in pre-vaccine strains (Table 6.5).

Table 6.5 Amino acid sites with significant amino acid differences between pre-and post-vaccine A2 strains

Representative A2 strains	Amino acid sites								
	103	234	246	273	319	333	334	429	464
Pre-vaccine South African A2 strains									
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	N	N	I	L	H	W	T	K	N
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	G	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	G	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	G	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	G	.	.	.	.	.	.	.	.
Post-vaccine South African A2 strains									
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	D	.	.	.	.	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	D	I	L	F	R	L	N	R	T
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	D	I	L	F	R	L	N	R	T
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	D	I	L	F	R	L	N	R	T
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	D	I	L	F	R	L	N	R	T
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	D	I	L	F	R	L	N	R	T

A strain per year was randomly selected to represent A2 strains that circulated in that particular year. South African pre-vaccine A2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.14 Amino acid differences in NSP2 sequences

When South African post-vaccine N2 strains were compared with cognate pre-vaccine N2 strains, aa differences S16Y (in three post-vaccine strains) and K58R (in ten post-vaccine strains) were observed. The aa residues asparagine, alanine, methionine and leucine at positions 74, 75, 82 and 254, respectively, occurred at a higher prevalence during post-vaccine period as compared to pre-vaccine period (Table 6.6).

Table 6.6 Amino acid sites with significant amino acid differences between pre-and post-vaccine N2 strains

Representative N2 strains	Amino acid sites			
	74	75	82	254
Pre-vaccine South African N2 strains				
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	T	T	V	P
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	V	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	V	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	.	V	.	S
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	.	V	.	.
Post-vaccine South African N2 strains				
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]	.	V	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	N	A	M	L
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	N	A	M	L
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	N	A	M	L
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	N	A	M	L
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	N	A	M	L
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	N	A	M	L

A strain per year was randomly selected to represent N2 strains that circulated in that particular year. South African pre-vaccine A2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.15 Amino acid differences in NSP4 sequences

When we compared post-vaccine NSP4 sequences with pre-vaccine NSP4 sequences we identified amino acid differences: Y131H (in ten strains), T/I139A (in ten strains) and C/D/N140G (in ten strains). The aa residues tabulated below were identified at a significantly higher proportion (60-80%) during post-vaccine period than pre-vaccine period (1-2%).

Table 6.7 Amino acid sites with significant amino acid differences between pre-and post-vaccine E2 strains

Representative E2 strains	Amino acid sites					
	74	135	137	139	140	163
Pre-vaccine South African E2 strains						
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	T	I	R	T	N	K
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	.	V	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD1138/2008/G2P[4]	.	V	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007/G2P[4]	.	V	.	.	.	.
Post-vaccine South African E2 strains						
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010G2P[4]	.	V	.	.	.	.
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	A	M	Q	I	C	R
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	A	M	Q	I	C	R
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	A	M	Q	I	C	R
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	V	M	Q	I	C	R

A strain per year was randomly selected to represent E2 strains that circulated in that particular year. South African pre-vaccine A2 strains are highlighted in bold-face while post-vaccine strains are highlighted in bold-red. The aa residues highlighted in sky-blue differed significantly between pre- and post-vaccine strains. The aa residues highlighted in black dot (.) are similar to the aa residue featured in the consensus study strain RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4].

6.3.2.16 Amino acid difference in NSP5 sequences

When we compared South African post-vaccine H2 strains with pre-vaccine H2 strains we identified amino acid difference N175D in seven post-vaccine strains.

6.3.3 Analysis of selection pressure

Selection pressure analysis by MEME, FUBAR and FEL showed that most of the codon sites in the 11 gene segments were undergoing purifying selection. However, aa sites 14 (in VP2), 7 (in VP3),

7(in VP4), 13 (in VP6), 254 (in NSP2), 6 (in NSP3) and 5 (in NSP4) were undergoing positive selection (Table 6.8).

Table 6.8: Positively selected sites as identified by FEL, FUBAR and MEME analysis

Method	Amino acid sites in the gene segments under positive selection										
	VP1	**VP2**	**VP3**	**VP4**	**VP6**	**VP7**	**NSP1**	**NSP2**	**NSP3**	**NSP4**	**NSP5**
MEME	2,62,90, 97,128, 157,158, 187,332, 364,371, 406,546, 654,1033, 1041	5,13,**14**, 723,869, 873,874, 877,878	**7**,25,133,153,322, 346,473, 704,788, 828,835	**7**,192, 625,627	**13**	3,5	10,212, 271,313	249,**254**	3,**6**, 16,27	2,3,**5**, 9,10, 16,17, 51,79, 170	3,7
FUBAR	60,364	**14**, 878	**7**	**7**	**13**	66	10,11, 103,380, 402	75,**254**	**6**	**5**	-
FEL	-	**14**	**7**,828	**7**	**13**	-	-	**254**	**6**	**5**,9	-

Bold and underlined sites were identified using all the three methods (FEL, FUBAR and MEME) and these were considered for positive selection. For MEME and FEL, statistical significance was assessed at $p \leq 0.1$ while for FUBAR it was assessed at posterior probability ≥ 0.9. The dash (-) sign indicates that no site was identified. Additional analyses reports are provided in Appendix 13.

6.4 Discussion

The study reports whole-genome analysis of South African G2P[4] RVA strains collected seven years before and seven years after the introduction of the Rotarix® vaccine. The whole-genome analysis showed that all the 103 G2P[4] strains analysed possessed the typical DS-1-like (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) constellation. These findings are consistent with whole-genome studies of G2P[4] strains in several countries (Dennis *et al*., 2014; Donato *et al*., 2014; Doan *et al*.., 2015; Agbemabiese *et al*., 2016).

Phylogenetic analysis of South African G2 strains showed that, of the four previously described G2 lineages (Doan *et al*., 2011), South African G2 strains clustered in G2-lineage IV. All the South African G2 strains had asparagine at position 96 (N96), which has been the hallmark of sub-lineage IVa linked with re-emergence of G2 strains in the last decade (2001-2009) (Doan *et al*., 2011). We noted that most of the pre-vaccine G2 strains that circulated from 2003-2009 and those of the first year after vaccine introduction kam(2010, when coverage was still low) clustered distinctly from most of the post-vaccine strains that circulated from 2012-2017. This distinct segregation appeared to be corroborated by the sharp increase in relative frequency at

positions 15, 37, 139 and 242 of aa residues serine, isoleucine, isoleucine and asparagine, respectively, during post-vaccine period compared to pre-vaccine period. According to previous studies, the asparagine residue at position 242 is associated with the re-emergence of the sub-lineage IVa-3 to where most modern G2 sequences globally belong (Doan *et al*., 2011; Agbemabiese *et al*., 2016). The P[4]-lineage IV to which majority (97%) of South African P[4] strains clustered, emerged during 2000s (Gomez *et al*., 2011) and is suggested to have prevalently circulated worldwide, possibly due to epidemiological fitness (Giammanco *et al*., 2014).

The aa profile of VP7 antigenic domains was nearly the same between pre- and post-vaccine G2 strains, with the exception of three aa differences (T87S and V212M identified in one strain and N146D identified in one strain). The amino acid differences T87S and V212M are conservative in nature while the N146D substitution resulted in a change from uncharged to a negatively charged residue which may contribute to antigenic change (Betts and Russell, 2003). Similarly, the aa composition in VP4 antigenic domains was maintained during pre- and post-vaccine period with the exception of D192Y in three strains, T115I in one strain and S133N in one strain. While S133N aa substitution was conservative, D192Y resulted in a change from a negatively charged to uncharged residue while T115I resulted in change from polar to nonpolar residue. The change in polarity and charge may affect the biochemical properties of the protein at these regions and potentially alter antigen-antibody binding which can promote vaccine-escape strains (Betts and Russell, 2003).

We analysed aa differences outside known antigenic sites to gain further insight into the genetic make-up between pre- and post-vaccine strains. We observed aa differences that were significantly prevalent, in about 60-80% corresponding to 35-45 out of 57 post-vaccine strains, compared with negligible proportions of 1-2% corresponding to 1 -2 out of 46 pre-vaccine strains. These aa differences potentially explain the distinct clustering of most post-vaccine strains away from pre-vaccine strains (with the exception of 2010 post-vaccine strains which clustered alongside the pre-vaccine strains). We hypothesize that these aa differences that occurred significantly during the post-vaccine period could be involved in conditioning the strains for more efficient dissemination into a susceptible population. It appears the aa differences observed in

post-vaccine strains may be involved with the dynamic natural fluctuation of G2P[4] strains and not necessarily influenced by the vaccine. There are two reasons why we think this could be the case. One, these aa differences were only apparent from 2012 and did not immediately coincide with vaccine introduction in 2009 as reported in Rotarix® jurisdictions in Australia, Belgium and Brazil (Gurgel *et al*., 2007; Donato, 2012; Zeller *et al*., 2017) where researchers hinted at the possibility of vaccine-induced selective pressure. The other reason is although significant aa differences were observed during post-vaccine period, the aa residues still occurred, albeit in negligible proportions during pre-vaccine period.

Most aa sites in the 11 gene segments were undergoing purifying selection. Rotavirus RNA dependent polymerase is error-prone and promotes high mutation rates (Estes and Greenberg, 2013) and therefore purifying selection may be a strategy to purge any resultant deleterious effects. However sites 14 in VP2, 7 in VP3, 7 in VP4, 13 in VP6, 254 in NSP2 and 5 in NSP4 were undergoing positive selection possibly to enhance viral fitness although the present study did not assess the impact of the positive selection at these sites

It is noteworthy to mention that in our dataset, we had nearly an even distribution of G2P[4] strains during pre and post-vaccine periods (46 pre-vaccine and 57 post-vaccine). Based on the distribution of the collected archived samples, we did not observe the three to four-year cyclical epidemiological occurrence pattern reported for South African G2P[4] (Page and Steele, 2004).

6.5 Conclusion

The study shows that the aa composition in the VP7 and VP4 antigenic regions of South African G2P[4] strains were nearly the same during pre-and post-vaccination period. However, aa differences were observed in the gene segments of post-vaccine strains relative to pre-vaccine strains in the non-antigenic regions. We hypothesize the aa changes consistently prevalent during the post-vaccination period (in 60-80% of the post-vaccine strains) to be due to natural RVA evolutionary processes as they were present in negligible percentages (1-2%) during pre-vaccine period. Continued whole-genome RVA characterization and analysis of both antigenic and non-antigenic regions of RVA gene segments is certainly warranted to fully assess the impact of vaccine-induced selective pressure.

CHAPTER SEVEN: DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

7.1 Discussion

The present study aimed to elucidate evolutionary changes of South African G1P[8] and G2P[4] RVA strains in a 14-year period and assess the impact of the Rotarix® vaccine introduction on the evolution of these RVA strains. The G1P[8] and G2P[4] RVA strains are the two most predominant RVA strains, globally (Doro *et al*., 2014). In South Africa, G1P[8] was a predominant strain pre-vaccine introduction (Steele *et al*., 2003), while non-G1P[8] RVA strains, such as G2P[4], which are not incorporated in the Rotarix® vaccine, have predominated during the post-vaccine introduction period (Page *et al*., 2018). The overall predominance of G1P[8] strain in South Africa before the RVA vaccine introduction and the emergence/ re-emergence of non-G1P[8] strains during the post-vaccination period formed the basis of the selection criteria of G1P[8] and G2P[4] strains in this study.

The consideration to perform the study in a 14 year period aimed to circumnavigate the limitations of earlier studies that had tried to assess vaccine-induced selective pressure (Steyer *et al*., 2014; Zeller *et al*., 2017; Santos *et al*., 2019). To comprehensively conclude on any apparent vaccine-induced genetic changes, the opinion among prominent RVA researchers was that an evaluation of genetic changes in at least seven years pre-and seven years post-vaccine introduction could be most appropriate (Patton, 2012). For G1P[8], the total number of sequenced samples was 103, with an additional 68 South African G1P[8] sequences acquired from GenBank. Although the study did not principally elucidate the epidemiological distribution of G1P[8] strains, we noted an overall significant difference between the pre- (*n*=150) and post-G1P[8] (*n*=23) RVA strains. This observation corroborated the report regarding the substantial decline of G1P[8] RVA strains during the post-vaccination period in South Africa (Page *et al*., 2018). For G2P[4], the total number of sequenced samples was 98, with an additional five South African G2P[4] sequences extracted from GenBank. Based on archived records, there were no available G2P[4] strains in 2002, 2004, 2005 and 2011. Notably, there was nearly an equal distribution of pre (*n*=46) and post-vaccine (*n*=57) G2P[4] strains from our dataset and this

observation is converse to previously reported three to four-year cyclical appearance of G2P[4] strains in South Africa (Page and Steele, 2004).

The rationale for whole-genome based sequencing was to comprehensively assess genetic changes that might be occurring across the genome (Bwogi *et al*., 2017; Esona *et al.*, 2017; Jere *et al*., 2018; Maringa *et al*., 2020; Rasebotsa *et al*., 2020) considering that the rotavirus genome is segmented with an inherent ability for reassortment events (Estes and Greenberg, 2013). The comprehensive elucidation of RVA diversity is facilitated by the utilization of the whole-genome RVA classification scheme that was introduced in 2008 (Matthijnssens *et al.*, 2008). Furthermore, the adoption of NGS technology has greatly capacitated whole-genome surveillance of RVA (Ghosh and Kobayashi, 2011; Nyaga *et al*., 2020) and this study utilized the Illumina MiSeq® sequencing platform.

The four objectives of the study were tackled in an overlapping manner through manuscript format. The first objective investigated the genotype variation and relative emergence of potential vaccine escape strains. This objective was mainly addressed by the findings of atypical G1P[8] strains that were published (Mwangi *et al*., 2020). The two atypical G1P[8] strains possessed a DS-1-like genotype constellation (G2-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) rather than the typical Wa-like genotype constellation (G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1) (Matthijnssens and Van Rast, 2012). These atypical strains circulated in 2008, which was before vaccine introduction in South Africa. These findings were unique as most studies that have reported on atypical G1P[8] strains were during the post-vaccination period (Fujii *et al*., 2014; Kuzuya *et al*., 2014; Yamamoto *et al*., 2014; Komoto *et al*., 2015; Komoto *et al*., 2016; Jere *et al*., 2018; Luchs *et al*., 2019) which led the researchers to suggest the possibility of vaccine-induced strain changes. However, we hypothesize that atypical G1P[8] strains have been in circulation even before RVA vaccine introduction. Therefore, the increased reports of atypical G1P[8] strains during the post-vaccination period may have to do with the adoption of whole-genome based surveillance techniques. Phylogenetic analysis showed that the two atypical G1P[8] strains were generated through reassortment mechanisms involving contemporary local South African G1P[8] strains and the DS-1-like backbone of G2P[6] strains, as they clustered distinctly away from other globally reported atypical G1P[8] strains. These findings highlight the genetic reassortment

mechanism as a driver of RVA evolution. The 1.2% frequency of occurrence of the DS-1like G1P[8] strains from our dataset of G1P[8] strains implies that these strains lack a degree of genetic fitness to be able to infect and spread rapidly in the human population (Gentsch *et al*., 2005; Patton *et al*., 2006).

To investigate the evolutionary dynamics of South African pre-and post-G1P[8] and G2P[4] strains, which was the second objective of the study, we utilized phylogenetic and pairwise sequence analysis. For G1P[8] strains, the majority of the genes of outer capsid proteins, clustered in lineages G1-lineage I and II (for VP7) and P[8]-lineage III (for VP4). G1-lineage I and II and P[8]-lineage III are the predominant lineages that have persistently circulated over time globally, potentially due to their epidemiological fitness (Le *et al*., 2010; da Silva *et al*., 2015). In all the 11 gene segments of G1P[8] strains, there was a tendency of clustering by yearly status, with post-vaccine strains (2010, 2014 and 2015) forming their sub-lineages. One strain, RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2250/2013/G1P[8], clustered with Rotarix® and had ≥ 99.9% nucleotide identity with Rotarix® in all its structural and non-structural genes, indicative of vaccine shedding. Rotarix® vaccine shedding is well documented in the literature and plays a role in herd immunity (Hsieh *et al*., 2014).

South African G2 strains clustered in globally predominant G2-lineage IV, while the majority (99.7%) of P[4] strains, clustered in the prevalent P[4]-lineage IV. All the G2 strains had the asparagine residue at position 96 which is a hallmark of sub-lineage IVa that emerged and spread globally from the year 2000 (Doan *et al*., 2011). The P[4]-lineage IV similarly emerged during 2000s and spread globally (Gomez *et al*., 2011). It appears the association of G2-lineage IV and P[8]-lineage IV confers these strains with some competitive advantage to infect susceptible populations. Overall, most of the pre-vaccine G2P[4] strains in all the gene segments clustered in a pattern whereby most of the pre-vaccine strains (2003-2009) clustered distinctly away from post-vaccine strains (2012-2017). The 2010 post-vaccine strains were a significant exception as they segregated alongside the pre-vaccine strains.

The third objective investigated genetic changes between pre and post-vaccine strains and was performed primarily through pairwise sequence comparisons and also the assessment of

evolution in the aa sites through selection pressure analysis. When the different gene segments of G1P[8] post-vaccine strains were compared with cognate pre-vaccine strains, we identified aa differences that potentially explained the distinct yearly clustering of post-vaccine strains into sub-lineages. However, we did not observe aa differences that were consistently conserved throughout the post-vaccine period for the G1P[8] strains. It appears that these aa differences observed for yearly strains during post-vaccine period are due to the natural RVA evolutionary processes. When post-vaccine G2P[4] strains were compared with pre-vaccine strains, we observed aa differences that were not present during pre-vaccine period, albeit being not conserved throughout the post-vaccine period. Interestingly, we identified aa residues that were consistently prevalent and appeared at a significantly higher frequency in 60-80% of post-vaccine strains (which corresponds to 35-45 out of 57 post-vaccine strains) at various sites in the gene segments compared to their negligible proportions of 1-2% of pre-vaccine strains (which corresponds to 1-2 out of 46 pre-vaccine strains). By virtue of the presence of these aa differences, notwithstanding negligible proportions during pre-vaccine period, we hypothesize that they were not necessarily vaccine-driven. Rather, we suggest the replacement of aa residues at several aa sites which was found to be prevalent during post-vaccination period to be due to the natural RVA evolutionary process. Such aa substitutions such as D96N and S242N for G2 strains have been reported even before RVA vaccine introduction (Doan *et al*., 2011). The G2P[4] strains evolve dynamically with aa substitutions at particular sites, explaining the dynamic cyclical appearance reported in several countries (Linhares and Justino, 2014). Overall, selection pressure analysis in the gene segments of G1P[8] and G2P[4] strains indicated that most of the sites were undergoing purifying selection potentially to phase out the deleterious effects of the error-prone RNA polymerase. However, some few sites (such as site 7 and 245 in M1 genes, 13 in I2 genes, 254 in N2 genes, 14 in C2 genes) appeared to undergo positive selection, possibly to enhance viral fitness.

The study's fourth objective investigated the aa profiles within the mapped antigenic regions of the neutralization proteins, VP7 and VP4. Our approach for G1P[8] strains was to perform a pairwise comparison of the pre-and post-vaccine strains with the cognate regions in the monovalent G1P[8] vaccine strain and correlate any genetic changes through computational

protein modelling. The aa composition of post-vaccine strains was nearly the same with that of pre-vaccine strains at the antigenic domains of VP7 and VP4, with a major exception of N147D substitution that occurred in eight 2014 and 2015 strains. The N147D substitution resulted in a change in charge from uncharged residue to negatively charged residue. Free energy change analysis showed that this substitution resulted in a loss of energy that significantly destabilized the resultant protein structure. Therefore, we hypothesized that since the mutation resulted in destabilizing the protein structure, such aa changes may not be favored in the longterm to enhance viral fitness. Thus, the aa substitution appears to be a spontaneous RVA evolutionary effect and not necessarily vaccine-induced. The overall aa profile in VP7 and VP4 antigenic domains of G2P[4] strains remained mostly unchanged between pre and post-vaccine strains.

7.2 Study limitations

The study was limited by less distribution of G1P[8] strains during post-vaccination period compared to the pre-vaccine G1P[8] strains. However, it is essential to note that the substantial decline of G1P[8] strains in South Africa is well documented (Page *et al*., 2018). For G2P[4] strains, we did not obtain archived samples in 2002, 2004 and 2005. G2P[4] strains have been associated with the cyclical appearance in some geographical regions attributable to the natural RVA evolutionary processes.

7.3 Conclusions

In conclusion, the present study demonstrated the identification of atypical G1P[8] strains to be driven by genetic reassortment events of local, contemporary South African G1P[8] and DS-1-like backbone of G2P[6] strains. The identification of these atypical G1P[8] strains during the pre-vaccination period as opposed to their detection during the post-vaccination period, shows that they are not vaccine-induced. The study showed that Rotarix® did not appear to impact the aa profile in the antigenic domains of the neutralization proteins, VP7 and VP4 of G1P[8] strains. The aa composition in the antigenic profiles for both G1P[8] and G2P[4] strains remained mostly unchanged during pre- and post-vaccination period, with just a few aa differences that were not consistently conserved throughout the post-vaccination period. Similarly, for non-antigenic sites in G1P[8] strains, there were no consistently conserved aa differences in the post-vaccine period

except in yearly strains which explained the distinct yearly clustering of the post-vaccine strains. However, for G2P[4] strains, we identified aa residues at several sites in the gene segments that occurred in a significantly higher proportion (60-80%) in post-vaccine strains compared to their negligible amounts (1-2%) in pre-vaccine strains. While these aa changes could be due to natural RVA evolutionary processes of G2P[4] strains, it is not evident whether RVA vaccination with the monovalent G1P[8] Rotarix® vaccine creates a condition in which G2P[4] strains acquire selective advantage through such changes.

7.4 Recommendations and suggestions for future studies

The study has demonstrated the benefits of NGS technology for whole-genome surveillance of prevailing RVA strains and recommends rotavirus surveillance sites to extend the routine G/P surveillance towards a whole-genome characterization. Whole-genome surveillance allows identification of complete genotype constellation enabling to decipher RVA diversity. We recommend long-term assessment of RVA genome in both antigenic and non-antigenic sites to identify any consistently conserved aa differences which may provide hints of any vaccine-induced selective pressure. Apart from *in silico* analysis, future studies should further validate potential vaccine-escape mutants by undertaking cross-neutralization studies. We further suggest consideration of mutagenesis studies to elucidate how selective pressure occurring at various aa sites in the RVA gene segments impacts RVA evolution.

REFERENCES

Abdel-Haq, N.M., Thomas, R.A., Asmar, B.I., Zacharova, V. and Lyman, W.D., 2003. Increased prevalence of G1P [4] genotype among children with rotavirus-associated gastroenteritis in metropolitan Detroit. *Journal of Clinical Microbiology*, *41*(6), pp.2680-2682.

Abid, N., Pietrucci, D., Salemi, M. and Chillemi, G., 2020. New Insights Into the Effect of Residue Mutations on the Rotavirus VP1 Function Using Molecular Dynamic Simulations. *Journal of Chemical Information and Modeling*.

Agbemabiese, C.A., Nakagomi, T., Doan, Y.H., Do, L.P., Damanka, S., Armah, G.E. and Nakagomi, O., 2016. Genomic constellation and evolution of Ghanaian G2P [4] rotavirus strains from a global perspective. *Infection, Genetics and Evolution*, *45*, pp.122-131.

Agócs, M.M., Serhan, F., Yen, C., Mwenda, J.M., de Oliveira, L.H., Teleb, N., Wasley, A., Wijesinghe, P.R., Fox, K., Tate, J.E. and Gentsch, J.R., 2014. WHO global rotavirus surveillance network: a strategic review of the first 5 years, 2008–2012. *MMWR. Morbidity and Mortality Weekly Report*, *63*(29), p.634.

Ahmed, K., Nakagomi, T. and Nakagomi, O., 2005. Isolation and molecular characterization of a naturally occurring non-structural protein 5 (NSP5) gene reassortant of group A rotavirus of serotype G2P [4] with a long RNA pattern. *Journal of Medical Virology*, *77*(2), pp.323-330.

Ahmed, T., Svennerholm, A.M., Al Tarique, A., Sultana, G.N. and Qadri, F., 2009. Enhanced immunogenicity of an oral inactivated cholera vaccine in infants in Bangladesh obtained by zinc supplementation and by temporary withholding breast-feeding. *Vaccine*, *27*(9), pp.1433-1439.

Ali, A., Kazi, A.M., Cortese, M.M., Fleming, J.A., Moon, S., Parashar, U.D., Jiang, B., McNeal, M.M., Steele, D., Bhutta, Z. and Zaidi, A.K., 2015. Impact of withholding breastfeeding at the time of vaccination on the immunogenicity of oral rotavirus vaccine—a randomized trial. *PLoS One*, *10*(6).

Almeida, T.N.V., de Sousa, T.T., da Silva, R.A., Fiaccadori, F.S., Souza, M., Badr, K.R. and de Paula Cardoso, D.D.D., 2017. Phylogenetic analysis of G1P [8] and G12P [8] rotavirus A samples obtained in the pre-and post-vaccine periods, and molecular modeling of VP4 and VP7 proteins. *Acta Tropica*, *173*, pp.153-159.

Alvarez-Cubero, M.J., Saiz, M., Martinez-Garcia, B., Sayalero, S.M., Entrala, C., Lorente, J.A. and Martinez-Gonzalez, L.J., 2017. Next generation sequencing: an application in forensic sciences?. *Annals of Human Biology*, *44*(7), pp.581-592.

Anh, D.D., Van Trang, N., Thiem, V.D., Anh, N.T.H., Mao, N.D., Wang, Y., Jiang, B., Hien, N.D. and Rotavin-M1 Vaccine Trial Group, 2012. A dose-escalation safety and immunogenicity study

of a new live attenuated human rotavirus vaccine (Rotavin-M1) in Vietnamese children. *Vaccine*, *30*, pp.A114-A121.

Ansaldi, F., Pastorino, B., Valle, L., Durando, P., Sticchi, L., Tucci, P., Biasci, P., Lai, P., Gasparini, R. and Icardi, G., 2007. Molecular characterization of a new variant of rotavirus P [8] G9 predominant in a sentinel-based survey in central Italy. *Journal of Clinical Microbiology*, *45*(3), pp.1011-1015.

Aoki, S.T., Settembre, E.C., Trask, S.D., Greenberg, H.B., Harrison, S.C. and Dormitzer, P.R., 2009. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science*, *324*(5933), pp.1444-1447.

Appaiahgari, M.B., Glass, R., Singh, S., Taneja, S., Rongsen-Chandola, T., Bhandari, N., Mishra, S. and Vrati, S., 2014. Transplacental rotavirus IgG interferes with immune response to live oral rotavirus vaccine ORV-116E in Indian infants. *Vaccine*, *32*(6), pp.651-656.

Arana, A., Montes, M., Jere, K.C., Alkorta, M., Iturriza-Gómara, M. and Cilla, G., 2016. Emergence and spread of G3P [8] rotaviruses possessing an equine-like VP7 and a DS-1-like genetic backbone in the Basque Country (North of Spain), 2015. *Infection, Genetics and Evolution*, *44*, pp.137-144.

Arias, C.F., Silva-Ayala, D. and López, S., 2015. Rotavirus entry: a deep journey into the cell with several exits. *Journal of Virology*, *89*(2), pp.890-893.

Arista, S., Giammanco, G.M., De Grazia, S., Ramirez, S., Biundo, C.L., Colomba, C., Cascio, A. and Martella, V., 2006. Heterogeneity and temporal dynamics of evolution of G1 human rotaviruses in a settled population. *Journal of Virology*, *80*(21), pp.10724-10733.

Armah, G., Lewis, K.D., Cortese, M.M., Parashar, U.D., Ansah, A., Gazley, L., Victor, J.C., McNeal, M.M., Binka, F. and Steele, A.D., 2016. A randomized, controlled trial of the impact of alternative dosing schedules on the immune response to human rotavirus vaccine in rural Ghanaian infants. *The Journal of Infectious Diseases*, *213*(11), pp.1678-1685.

Armah, G.E., Sow, S.O., Breiman, R.F., Dallas, M.J., Tapia, M.D., Feikin, D.R., Binka, F.N., Steele, A.D., Laserson, K.F., Ansah, N.A. and Levine, M.M., 2010. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. *The Lancet*, *376*(9741), pp.606-614.

Arnold, M.M., Barro, M. and Patton, J.T., 2013. Rotavirus NSP1 mediates degradation of interferon regulatory factors through targeting of the dimerization domain. *Journal of Virology*, *87*(17), pp.9813-9821.

Arnoldi, F., Campagna, M., Eichwald, C., Desselberger, U. and Burrone, O.R., 2007. Interaction of rotavirus polymerase VP1 with nonstructural protein NSP5 is stronger than that with NSP2. *Journal of Virology*, *81*(5), pp.2128-2137.

Arora, R. and Chitambar, S.D., 2011. Full genomic analysis of Indian G1P [8] rotavirus strains. *Infection, Genetics and Evolution*, *11*(2), pp.504-511.

Asmah, R.H., Green, J., Armah, G.E., Gallimore, C.I., Gray, J.J., Iturriza-Gómara, M., Anto, F., Oduro, A., Binka, F.N., Brown, D.W. and Cutts, F., 2001. Rotavirus G and P genotypes in rural Ghana. *Journal of Clinical Microbiology*, *39*(5), pp.1981-1984.

Atherly, D.E., Lewis, K.D., Tate, J., Parashar, U.D. and Rheingans, R.D., 2012. Projected health and economic impact of rotavirus vaccination in GAVI-eligible countries: 2011–2030. *Vaccine*, *30*, pp.A7-A14.

Babji, S. and Kang, G., 2012. Rotavirus vaccination in developing countries. *Current Opinion in Virology*, *2*(4), pp.443-448.

Ball, J.M., Mitchell, D.M., Gibbons, T.F. and Parr, R.D., 2005. Rotavirus NSP4: a multifunctional viral enterotoxin. *Viral Immunology*, *18*(1), pp.27-40.

Ball, J.M., Tian, P., Zeng, C.Q.Y., Morris, A.P. and Estes, M.K., 1996. Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. *Science*, *272*(5258), pp.101-104.

Bányai, K., László, B., Duque, J., Steele, A.D., Nelson, E.A.S., Gentsch, J.R. and 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*, pp.A122-A130.

Banyai, K., Mijatovic-Rustempasic, S., Hull, J.J., Esona, M.D., Freeman, M.M., Frace, A.M., Bowen, M.D. and Gentsch, J.R., 2011. Sequencing and phylogenetic analysis of the coding region of six common rotavirus strains: Evidence for intragenogroup reassortment among co-circulating G1P [8] and G2P [4] strains from the United States. *Journal of Medical Virology*, *83*(3), pp.532-539.

Barro, M. and Patton, J.T., 2005. Rotavirus nonstructural protein 1 subverts innate immune response by inducing degradation of IFN regulatory factor 3. *Proceedings of the National Academy of Sciences*, *102*(11), pp.4114-4119.

Barro, M. and Patton, J.T., 2007. Rotavirus NSP1 inhibits expression of type I interferon by antagonizing the function of interferon regulatory factors IRF3, IRF5, and IRF7. *Journal of Virology*, *81*(9), pp.4473-4481.

Becker-Dreps, S., Vilchez, S., Velasquez, D., Moon, S.S., Hudgens, M.G., Zambrana, L.E. and Jiang, B., 2015. Rotavirus-specific IgG antibodies from mothers' serum may inhibit infant immune responses to the pentavalent rotavirus vaccine. *The Pediatric Infectious Disease Journal*, *34*(1), p.115.

Berglund, E.C., Kiialainen, A. and Syvänen, A.C., 2011. Next-generation sequencing technologies and applications for human genetic history and forensics. *Investigative Genetics*, *2*(1), p.23.

Bergmann, C.C., Maass, D., Poruchynsky, M.S., Atkinson, P.H. and Bellamy, A.R., 1989. Topology of the non-structural rotavirus receptor glycoprotein NS28 in the rough endoplasmic reticulum. *The EMBO Journal*, *8*(6), pp.1695-1703.

Bernstein, D.I., Sack, D.A., Rothstein, E., Reisinger, K., Smith, V.E., O'Sullivan, D., Spriggs, D.R. and Ward, R.L., 1999. Efficacy of live, attenuated, human rotavirus vaccine 89–12 in infants: a randomised placebo-controlled trial. *The Lancet*, *354*(9175), pp.287-290.

Berois, M., Sapin, C., Erk, I., Poncet, D. and Cohen, J., 2003. Rotavirus nonstructural protein NSP5 interacts with major core protein VP2. *Journal of Virology*, *77*(3), pp.1757-1763.

Betts, M.J. and Russell, R.B., 2003. Amino acid properties and consequences of substitutions. *Bioinformatics for Geneticists*, *317*, p.289.

Bhandari, N., Rongsen-Chandola, T., Bavdekar, A., John, J., Antony, K., Taneja, S., Goyal, N., Kawade, A., Kang, G., Rathore, S.S. and Juvekar, S., 2014a. Efficacy of a monovalent human-bovine (116E) rotavirus vaccine in Indian children in the second year of life. *Vaccine*, *32*, pp.A110-A116.

Bhandari, N., Rongsen-Chandola, T., Bavdekar, A., John, J., Antony, K., Taneja, S., Goyal, N., Kawade, A., Kang, G., Rathore, S.S. and Juvekar, S., 2014b. Efficacy of a monovalent human-bovine (116E) rotavirus vaccine in Indian infants: a randomised, double-blind, placebo-controlled trial. *The Lancet*, *383*(9935), pp.2136-2143.

Bishop, R., Davidson, G.P., Holmes, I.H. and Ruck, B.J., 1973. Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *The Lancet*, *302*(7841), pp.1281-1283.

Blackhall, J., Fuentes, A. and Magnusson, G., 1996. Genetic stability of a porcine rotavirus RNA segment during repeated plaque isolation. *Virology*, *225*(1), pp.181-190.

Borgan, M.A., Mori, Y., Ito, N., Sugiyama, M. and Minamoto, N., 2003. Antigenic analysis of nonstructural protein (NSP) 4 of group A avian rotavirus strain PO-13. *Microbiology and Immunology*, *47*(9), pp.661-668.

Borodavka, A., Dykeman, E.C., Schrimpf, W. and Lamb, D.C., 2017. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *Elife*, *6*, p.e27453.

Bronner, I.F. and Quail, M.A., 2019. Best Practices for Illumina Library Preparation. *Current Protocols in Human Genetics*, *102*(1), p.e86.

Broquet, A.H., Hirata, Y., McAllister, C.S. and Kagnoff, M.F., 2011. RIG-I/MDA5/MAVS are required to signal a protective IFN response in rotavirus-infected intestinal epithelium. *The Journal of Immunology*, *186*(3), pp.1618-1626.

Burke, R.M., Tate, J.E., Kirkwood, C.D., Steele, A.D. and Parashar, U.D., 2019. Current and new rotavirus vaccines. *Current Opinion in Infectious Diseases*, *32*(5), p.435.

Buttafuoco, A., Michaelsen, K., Tobler, K., Ackermann, M., Fraefel, C. and Eichwald, C., 2020. Conserved rotavirus NSP5 and VP2 domains interact and affect viroplasm. *Journal of Virology*, *94*(7).

Bwogi, J., Jere, K.C., Karamagi, C., Byarugaba, D.K., Namuwulya, P., Baliraine, F.N., Desselberger, U. and Iturriza-Gomara, M., 2017. Whole genome analysis of selected human and animal rotaviruses identified in Uganda from 2012 to 2014 reveals complex genome reassortment events between human, bovine, caprine and porcine strains. *PloS one*, *12*(6), p.e0178855.

Bzik, V.A., Medani, M., Baird, A.W., Winter, D.C. and Brayden, D.J., 2012. Mechanisms of action of zinc on rat intestinal epithelial electrogenic ion secretion: insights into its antidiarrhoeal actions. *Journal of Pharmacy and Pharmacology*, *64*(5), pp.644-653.

Caddy, S.L., Vaysburd, M., Wing, M., Foss, S., Andersen, J.T., O 'Connell, K., Mayes, K., Higginson, K., Iturriza-Gómara, M., Desselberger, U. and James, L.C., 2020. Intracellular neutralisation of rotavirus by VP6-specific IgG. *PLoS Pathogens*, *16*(8), p.e1008732.

Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M. and Peersen, O.B., 2015. Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *Journal of Virology*, *89*(1), pp.275-286.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. and Gormley, N., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), pp.1621-1624.

Carvalho, M.F. and Gill, D., 2019. Rotavirus vaccine efficacy: current status and areas for improvement. *Human Vaccines & Immunotherapeutics*, *15*(6), pp.1237-1250.

Castanier, C. and Arnoult, D., 2011. Mitochondrial localization of viral proteins as a means to subvert host defense. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1813*(4), pp.575-583.

Caust, J., Dyall-Smith, M.L., Lazdins, I. and Holmes, I.H., 1987. Glycosylation, an important modifier of rotavirus antigenicity. *Archives of Virology*, *96*(3-4), pp.123-134.

Centers for Disease Control and Prevention, 2004. Managing acute gastroenteritis among children: oral rehydration, maintenance, and nutritional therapy. *Pediatrics*, *114*(2), pp.507-507.

Chan, J., Nirwati, H., Triasih, R., Bogdanovic-Sakran, N., Soenarto, Y., Hakimi, M., Duke, T., Buttery, J.P., Bines, J.E., Bishop, R.F. and Kirkwood, C.D., 2011. Maternal antibodies to rotavirus:

could they interfere with live rotavirus vaccines in developing countries?. *Vaccine*, *29*(6), pp.1242-1247.

Chandola, T.R., Taneja, S., Goyal, N., Antony, K., Bhatia, K., More, D., Bhandari, N., Cho, I., Mohan, K., Prasad, S. and Harshavardhan, G.V.J.A., 2017. ROTAVAC® does not interfere with the immune response to childhood vaccines in Indian infants: A randomized placebo controlled trial. *Heliyon*, *3*(5), p.e00302.

Chang-Graham, A.L., Perry, J.L., Strtak, A.C., Ramachandran, N.K., Criglar, J.M., Philip, A.A., Patton, J.T., Estes, M.K. and Hyser, J.M., 2019. Rotavirus calcium dysregulation manifests as dynamic calcium signaling in the cytoplasm and endoplasmic reticulum. *Scientific Reports*, *9*(1), pp.1-20.

Chen, J.Z., Settembre, E.C., Aoki, S.T., Zhang, X., Bellamy, A.R., Dormitzer, P.R., Harrison, S.C. and Grigorieff, N., 2009. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proceedings of the National Academy of Sciences*, *106*(26), pp.10644-10648.

Choi, A.H., Basu, M., McNeal, M.M., Flint, J., VanCott, J.L., Clements, J.D. and Ward, R.L., 2000. Functional mapping of protective domains and epitopes in the rotavirus VP6 protein. *Journal of Virology*, *74*(24), pp.11574-11580.

Choi, Y. and Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), pp.2745-2747.

Chomczynski, P. and Sacchi, N., 2006. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nature Protocols*, *1*(2), pp.581-585.

Churgay, C.A. and Aftab, Z., 2012. Gastroenteritis in children: Part II. Prevention and management. *American Family Physician*, *85*(11), pp.1066-1070.

Ciarlet, M. and Schödel, F., 2009. Development of a rotavirus vaccine: clinical safety, immunogenicity, and efficacy of the pentavalent rotavirus vaccine, RotaTeq®. *Vaccine*, *27*, pp.G72-G81.

Ciarlet, M., and Estes, M., 2002. Rotaviruses: basic biology, epidemiology and methodologies. In Britton G(ed) Encyclopedia of environmental microbiology. Wiley, New York, pp 2573-2773.

Ciarlet, M., Hyser, J.M. and Estes, M.K., 2002. Sequence Analysis of the VP4, VP6, VP7, and NSP4 Gene Products ofthe Bovine Rotavirus WC3. *Virus Genes*, *24*(2), pp.107-118.

Coldiron, M.E., Guindo, O., Makarimi, R., Soumana, I., Seck, A.M., Garba, S., Macher, E., Isanaka, S. and Grais, R.F., 2018. Safety of a heat-stable rotavirus vaccine among children in Niger: Data from a phase 3, randomized, double-blind, placebo-controlled trial. *Vaccine*, *36*(25), pp.3674-3680.

Condemine, W., Eguether, T., Couroussé, N., Etchebest, C., Gardet, A., Trugnan, G. and Chwetzoff, S., 2019. The C terminus of rotavirus VP4 protein contains an actin binding domain which requires cooperation with the coiled-coil domain for actin remodeling. *Journal of Virology*, *93*(1), pp.e01598-18.

Contreras-Treviño, H.I., Reyna-Rosas, E., León-Rodríguez, R., Ruiz-Ordaz, B.H., Dinkova, T.D., Cevallos, A.M. and Padilla-Noriega, L., 2017. Species A rotavirus NSP3 acquires its translation inhibitory function prior to stable dimer formation. *PloS One*, *12*(7).

Correia, J.B., Patel, M.M., Nakagomi, O., Montenegro, F.M., Germano, E.M., Correia, N.B., Cuevas, L.E., Parasha, U.D., Cunliffe, N.A. and Nakagomi, T., 2010. Effectiveness of monovalent rotavirus vaccine (Rotarix) against severe diarrhea caused by serotypically unrelated G2P [4] strains in Brazil. *The Journal of Infectious Diseases*, *201*(3), pp.363-369.

Corthésy, B., Benureau, Y., Perrier, C., Fourgeux, C., Parez, N., Greenberg, H. and Schwartz-Cornil, I., 2006. Rotavirus anti-VP6 secretory immunoglobulin A contributes to protection via intracellular neutralization but not via immune exclusion. *Journal of Virology*, *80*(21), pp.10692-10699.

Coulson, B.S. and Kirkwood, C., 1991. Relation of VP7 amino acid sequence to monoclonal antibody neutralization of rotavirus and rotavirus monotype. *Journal of Virology*, *65*(11), pp.5968-5974.

Cowley, D., Donato, C.M., Roczo-Farkas, S. and Kirkwood, C.D., 2016. Emergence of a novel equine-like G3P [8] inter-genogroup reassortant rotavirus strain associated with gastroenteritis in Australian children. *Journal of General Virology*, *97*(2), pp.403-410.

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. and Desselberger, U., 2017. Rotavirus infection. *Nature Reviews Disease Primers*, *3*(1), pp.1-16.

Criglar, J.M., Anish, R., Hu, L., Crawford, S.E., Sankaran, B., Prasad, B.V. and Estes, M.K., 2018. Phosphorylation cascade regulates the formation and maturation of rotaviral replication factories. *Proceedings of the National Academy of Sciences*, *115*(51), pp.E12015-E12023.

Criglar, J.M., Hu, L., Crawford, S.E., Hyser, J.M., Broughman, J.R., Prasad, B.V. and Estes, M.K., 2014. A novel form of rotavirus NSP2 and phosphorylation-dependent NSP2-NSP5 interactions are associated with viroplasm assembly. *Journal of Virology*, *88*(2), pp.786-798.

Cunliffe, N.A., Gondwe, J.S., Graham, S.M., Thindwa, B.D.M., Dove, W., Broadhead, R.L., Molyneux, M.E. and Hart, C.A., 2001. Rotavirus strain diversity in Blantyre, Malawi, from 1997 to 1999. *Journal of Clinical Microbiology*, *39*(3), pp.836-843.

Cunliffe, N.A., Witte, D., Ngwira, B.M., Todd, S., Bostock, N.J., Turner, A.M., Chimpeni, P., Victor, J.C., Steele, A.D., Bouckenooghe, A. and Neuzil, K.M., 2012. Efficacy of human rotavirus

vaccine against severe gastroenteritis in Malawian children in the first two years of life: a randomized, double-blind, placebo controlled trial. *Vaccine*, *30*, pp.A36-A43.

da Silva, M.F.M., Rose, T.L., Gómez, M.M., Carvalho-Costa, F.A., Fialho, A.M., de Assis, R.M.S., de Andrade, J.D.S.R., de Mello Volotão, E. and Leite, J.P.G., 2015. G1P [8] species A rotavirus over 27 years–pre-and post-vaccination eras–in Brazil: full genomic constellation analysis and no evidence for selection pressure by Rotarix® vaccine. *Infection, Genetics and Evolution*, *30*, pp.206-218.

Das, B.K., Gentsch, J.R., Cicirello, H.G., Woods, P.A., Gupta, A., Ramachandran, M., Kumar, R., Bhan, M.K. and Glass, R.I., 1994. Characterization of rotavirus strains from newborns in New Delhi, India. *Journal of Clinical Microbiology*, *32*(7), pp.1820-1822.

De Vos, B., Vesikari, T., Linhares, A.C., Salinas, B., Pérez-Schael, I., Ruiz-Palacios, G.M., de Lourdes Guerrero, M., Phua, K.B., Delem, A. and Hardt, K., 2004. A rotavirus vaccine for prophylaxis of infants against rotavirus gastroenteritis. *The Pediatric Infectious Disease Journal*, *23*(10), pp.S179-S182.

DeLano, W.L., 2009. The PyMOL Molecular Graphics System http://www. pymol. org, 2002. *DeLano Scientific*.

Dennis, A.F., McDonald, S.M., Payne, D.C., Mijatovic-Rustempasic, S., Esona, M.D., Edwards, K.M., Chappell, J.D. and Patton, J.T., 2014. Molecular epidemiology of contemporary G2P [4] human rotaviruses cocirculating in a single US community: footprints of a globally transitioning genotype. *Journal of Virology*, *88*(7), pp.3789-3801.

Deo, R.C., Groft, C.M., Rajashankar, K.R. and Burley, S.K., 2002. Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer. *Cell*, *108*(1), pp.71-81.

Desai, S., Rathi, N., Kawade, A., Venkatramanan, P., Kundu, R., Lalwani, S.K., Dubey, A.P., Rao, J.V., Narayanappa, D., Ghildiyal, R. and Gogtay, N.J., 2018. Non-interference of Bovine-Human reassortant pentavalent rotavirus vaccine ROTASIIL® with the immunogenicity of infant vaccines in comparison with a licensed rotavirus vaccine. V*accine*, *36*(37), pp.5519-5523.

Desselberger, U., 1996. Genome rearrangements of rotaviruses. In *Advances in virus research* (Vol. 46, pp. 69-95). Academic Press.

Desselberger, U., 2014. Rotaviruses. *Virus Research*, *190*, pp.75-96.

Díaz-Salinas, M.A., Casorla, L.A., López, T., López, S. and Arias, C.F., 2018. Most rotavirus strains require the cation-independent mannose-6-phosphate receptor, sortilin-1, and cathepsins to enter cells. *Virus Research*, *245*, pp.44-51.

Ding, S., Zhu, S., Ren, L., Feng, N., Song, Y., Ge, X., Li, B., Flavell, R.A. and Greenberg, H.B., 2018. Rotavirus VP3 targets MAVS for degradation to inhibit type III interferon expression in intestinal epithelial cells. *Elife*, *7*, p.e39494.

Do, L.P., Nakagomi, T., Otaki, H., Agbemabiese, C.A., Nakagomi, O. and Tsunemitsu, H., 2016. Phylogenetic inference of the porcine Rotavirus A origin of the human G1 VP7 gene. *Infection, Genetics and Evolution*, *40*, pp.205-213.

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

Doan, Y.H., Nakagomi, T., Cunliffe, N.A., Pandey, B.D., Sherchand, J.B. and Nakagomi, O., 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), p.1969.

Donato, C., 2012. Molecular epidemiology of rotavirus in the era of vaccination. *Microbiology Australia*, *33*(2), pp.64-66.

Donato, C.M., Zhang, Z.A., Donker, N.C. and 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*, *28*, pp.398-412.

Donker, N.C., Boniface, K. and Kirkwood, C.D., 2011b. Phylogenetic analysis of rotavirus A NSP2 gene sequences and evidence of intragenic recombination. *Infection, Genetics and Evolution*, *11*(7), pp.1602-1607.

Donker, N.C., Foley, M., Tamvakis, D.C., Bishop, R. and Kirkwood, C.D., 2011a. Identification of an antibody-binding epitope on the rotavirus A non-structural protein NSP2 using phage display analysis. *Journal of General Virology*, *92*(10), pp.2374-2382.

Dormitzer, P.R., Greenberg, H.B. and Harrison, S.C., 2000. Purified recombinant rotavirus VP7 forms soluble, calcium-dependent trimers. *Virology*, *277*(2), pp.420-428.

Dormitzer, P.R., Nason, E.B., Prasad, B.V. and Harrison, S.C., 2004. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. *Nature*, *430*(7003), p.1053.

Dormitzer, P.R., Sun, Z.Y.J., Wagner, G. and Harrison, S.C., 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *The EMBO Journal*, *21*(5), pp.885-897.

Dóró, R., László, B., Martella, V., Leshem, E., Gentsch, J., Parashar, U. and 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*, *28*, pp.446-461.

Dulgheroff, A.C.B., Silva, G.A.V.D., Naveca, F.G., Oliveira, A.G.D. and Domingues, A.L.D.S., 2016. Diversity of group A rotavirus genes detected in the Triângulo Mineiro region, Minas Gerais, Brazil. B*razilian Journal of Microbiology*, *47*(3), pp.731-740.

Edwards, M.J. and Sier, A.M., 1960. Bovine epizootic diarrhoea in Western Australia. *Australian Veterinary Journal*, *36*, pp.402-404.

Egli, A., Santer, D.M., O'Shea, D., Tyrrell, D.L. and Houghton, M., 2014. The impact of the interferon-lambda family on the innate and adaptive immune response to viral infections. *Emerging Microbes & Infections*, *3*(1), pp.1-12.

Eichwald, C., Arnoldi, F., Laimbacher, A.S., Schraner, E.M., Fraefel, C., Wild, P., Burrone, O.R. and Ackermann, M., 2012. Rotavirus viroplasm fusion and perinuclear localization are dynamic processes requiring stabilized microtubules. *PloS One*, *7*(10).

Eichwald, C., Jacob, G., Muszynski, B., Allende, J.E. and Burrone, O.R., 2004. Uncoupling substrate and activation functions of rotavirus NSP5: phosphorylation of Ser-67 by casein kinase 1 is essential for hyperphosphorylation. *Proceedings of the National Academy of Sciences*, *101*(46), pp.16304-16309.

Eisenberg, D., Lüthy, R. and Bowie, J.U., 1997. [20] VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in Enzymology,* 277, pp. 396-404).

Ella, R., Bobba, R., Muralidhar, S., Babji, S., Vadrevu, K.M. and Bhan, M.K., 2018. A Phase 4, multicentre, randomized, single-blind clinical trial to evaluate the immunogenicity of the live, attenuated, oral rotavirus vaccine (116E), ROTAVAC®, administered simultaneously with or without the buffering agent in healthy infants in India. *Human Vaccines & Immunotherapeutics*, *14*(7), pp.1791-1799.

Emperador, D.M., Velasquez, D.E., Estivariz, C.F., Lopman, B., Jiang, B., Parashar, U., Anand, A. and Zaman, K., 2016. Interference of monovalent, bivalent, and trivalent oral poliovirus vaccines on monovalent rotavirus vaccine immunogenicity in rural Bangladesh. *Clinical Infectious Diseases*, *62*(2), pp.150-156.

Engevik, M.A., Banks, L.D., Engevik, K.A., Chang-Graham, A.L., Perry, J.L., Hutchinson, D.S., Ajami, N.J., Petrosino, J.F. and Hyser, J.M., 2020. Rotavirus infection induces glycan availability to promote ileum-specific changes in the microbiome aiding rotavirus virulence. *Gut Microbes*, pp.1-24.

Esona, M.D., Roy, S., Rungsrisuriyachai, K., Gautam, R., Hermelijn, S., Rey-Benito, G. and Bowen, M.D., 2018. Molecular characterization of a human G20P [28] rotavirus a strain with multiple genes related to bat rotaviruses. *Infection, Genetics and Evolution*, *57*, pp.166-170.

Esona, M.D., Roy, S., Rungsrisuriyachai, K., Sanchez, J., Vasquez, L., Gomez, V., Rios, L.A., Bowen, M.D. and Vazquez, M., 2017. Characterization of a triple-recombinant, reassortant rotavirus strain from the Dominican Republic. *The Journal of General Virology*, *98*(2), p.134.

Espinola, E.E., Amarilla, A., Arbiza, J. and Parra, G.I., 2008. Sequence and phylogenetic analysis of the VP4 gene of human rotaviruses isolated in Paraguay. *Archives of Virology*, *153*(6), pp.1067-1073.

Estes, M. K., and Greenberg, H. B., 2013. Rotaviruses In: Knipe DM, Howley PM, editors. Fields Virology.6th edition Wolters Kluwer Heath/Lippincott Williams& Wilkins, Philadelphia,PA, pp.1347-1401

Estes, M.K. and Kapikian, A.Z., 2007. *Rotaviruses*. In: Knipe D, Griffin D, Lamb R, Martin M, Roizman B, Straus S, editors. Fields Virology. Wolters Kluwer Health; Lippincott, Williams and Wilkins; Philadelphia, PA, USA, pp. 1917–1975.

Fiore, L., Greenberg, H.B. and Mackow, E.R., 1991. The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin. *Virology*, *181*(2), pp.553-563.

Flewett, T.H., Davies, H., Bryden, A.S. and Robertson, M.J., 1974. Diagnostic electron microscopy of faeces: II Acute gastroenteritis associated with reovirus-like particles. *Journal of Clinical Pathology*, *27*(8), pp.608-614.

Franco, M.A. and Greenberg, H.B., 1995. Role of B cells and cytotoxic T lymphocytes in clearance of and immunity to rotavirus infection in mice. *Journal of Virology*, *69*(12), pp.7800-7806.

Franco, M.A., Angel, J. and Greenberg, H.B., 2006. Immunity and correlates of protection for rotavirus vaccines. *Vaccine*, *24*(15), pp.2718-2731.

Franco, M.A., Lefevre, P., Willems, P., Tosser, G., Lintermanns, P. and Cohen, J., 1994. Identification of cytotoxic T cell epitopes on the VP3 and VP6 rotavirus proteins. *Journal of General Virology*, *75*(3), pp.589-596.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. and Dubchak, I., 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research*, *32*(suppl_2), pp.W273-W279.

Fredj, M.B.H., BenHamida-Rebaï, M., Heylen, E., Zeller, M., Moussa, A., Kacem, S., Van Ranst, M., Matthijnssens, J. and Trabelsi, A., 2013. Sequence and phylogenetic analyses of human rotavirus strains: Comparison of VP7 and VP8∗ antigenic epitopes between Tunisian and vaccine strains before national rotavirus vaccine introduction. *Infection, Genetics and Evolution*, *18*, pp.132-144.

Fu, C., 2010. Effectiveness of Lanzhou lamb rotavirus vaccine against hospitalized gastroenteritis: further analysis and update. *Human Vaccines*, *6*(11), pp.953-953.

Fu, C., He, Q., Xu, J., Xie, H., Ding, P., Hu, W., Dong, Z., Liu, X. and Wang, M., 2012. Effectiveness of the Lanzhou lamb rotavirus vaccine against gastroenteritis among children. *Vaccine*, *31*(1), pp.154-158.

Fujii, Y., Nakagomi, T., Nishimura, N., Noguchi, A., Miura, S., Ito, H., Doan, Y.H., Takahashi, T., Ozaki, T., Katayama, K. and Nakagomi, O., 2014. Spread and predominance in Japan of novel G1P [8] double-reassortant rotavirus strains possessing a DS-1-like genotype constellation typical of G2P [4] strains. *Infection, Genetics and Evolution*, *28*, pp.426-433.

Fujii, Y., Oda, M., Somura, Y. and Shinkai, T., 2019. Molecular characteristics of novel mono-reassortant G9P [8] rotavirus A strains possessing the NSP4 gene of the E2 genotype detected in Tokyo, Japan. *Japanese Journal of Infectious Diseases*, pp.JJID-2019.

Gaffey, M.F., Wazny, K., Bassani, D.G. and Bhutta, Z.A., 2013. Dietary management of childhood diarrhea in low-and middle-income countries: a systematic review. *BMC Public Health*, *13*(S3), p.S17.

Gardet, A., Breton, M., Fontanges, P., Trugnan, G. and Chwetzoff, S., 2006. Rotavirus spike protein VP4 binds to and remodels actin bundles of the epithelial brush border into actin bodies. *Journal of Virology*, *80*(8), pp.3947-3956.

Ge, Y., Mansell, A., Ussher, J.E., Brooks, A.E., Manning, K., Wang, C.J. and Taylor, J.A., 2013. Rotavirus NSP4 triggers secretion of proinflammatory cytokines from macrophages via Toll-like receptor 2. *Journal of Virology*, *87*(20), pp.11160-11167.

Gentsch, J.R., Laird, A.R., Bielfelt, B., Griffin, D.D., Bányai, K., Ramachandran, M., Jain, V., Cunliffe, N.A., Nakagomi, O., Kirkwood, C.D. and Fischer, T.K., 2005. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *Journal of Infectious Diseases*, *192*(Supplement_1), pp.S146-S159.

Kampf, G., 2018. Efficacy of ethanol against viruses in hand disinfection. *Journal of Hospital Infection*, *98*(4), pp.331-338.

Ghosh, S. and Kobayashi, N., 2011. Whole-genomic analysis of rotavirus strains: current status and future prospects. *Future Microbiology*, *6*(9), pp.1049-1065.

Giammanco, G.M., Bonura, F., Zeller, M., Heylen, E., Van Ranst, M., Martella, V., Banyai, K., Matthijnssens, J. and De Grazia, S., 2014. Evolution of DS-1-like human G2P [4] rotaviruses assessed by complete genome analyses. *Journal of General Virology*, *95*(1), pp.91-109.

Glass, R.I., Jiang, B. and Parashar, U., 2018. The future control of rotavirus disease: Can live oral vaccines alone solve the rotavirus problem?. *Vaccine*, *36*(17), pp.2233-2236.

Gómez, M.M., de Mendonça, M.C.L., Volotão, E.D.M., Tort, L.F.L., da Silva, M.F.M., Cristina, J. and Leite, J.P.G., 2011. Rotavirus A genotype P [4] G2: genetic diversity and reassortment events

among strains circulating in Brazil between 2005 and 2009. *Journal of Medical Virology*, *83*(6), pp.1093-1106.

Gomez-Rial, J., Sánchez-Batán, S., Rivero-Calle, I., Pardo-Seco, J., Martinón-Martínez, J.M., Salas, A. and Martinón-Torres, F., 2019. Rotavirus infection beyond the gut. *Infection and Drug Resistance*, *12*, p.55.

González, S.A. and Burrone, O.R., 1991. Rotavirus NS26 is modified by addition of single O-linked residues of N-acetylglucosamine. *Virology*, *182*(1), pp.8-16.

Gouvea, V., Lima, R.C., Linhares, R.E., Clark, H.F., Nosawa, C.M. and Santos, N., 1999. Identification of two lineages (WA-like and F45-like) within the major rotavirus genotype P [8]. *Virus Research*, *59*(2), pp.141-147.

Graham, K.L., Halasz, P., Tan, Y., Hewish, M.J., Takada, Y., Mackow, E.R., Robinson, M.K. and Coulson, B.S., 2003. Integrin-using rotaviruses bind α2β1 integrin α2 I domain via VP4 DGE sequence and recognize αXβ2 and αVβ3 by using VP7 during cell entry. *Journal of Virology*, *77*(18), pp.9969-9978.

Groome, M.J., Moon, S.S., Velasquez, D., Jones, S., Koen, A., Niekerk, N.V., Jiang, B., Parashar, U.D. and Madhi, S.A., 2014. Effect of breastfeeding on immunogenicity of oral live-attenuated human rotavirus vaccine: a randomized trial in HIV-uninfected infants in Soweto, South Africa. *Bulletin of the World Health Organization*, *92*, pp.238-245.

Guarino, A., Ashkenazi, S., Gendrel, D., Vecchio, A.L., Shamir, R. and Szajewska, H., 2014. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition/European Society for Pediatric Infectious Diseases evidence-based guidelines for the management of acute gastroenteritis in children in Europe: update 2014**.** *Journal of Pediatric Gastroenterology and Nutrition*, *59*(1), pp.132-152.

Guerra, S.F., Fecury, P.C., Bezerra, D.A., Lobo, P.S., Júnior, E.T.P., Júnior, E.C.S., Mascarenhas, J.D.A.P., Soares, L.S., Justino, M.C.A. and Linhares, A.C., 2019. Emergence of G12P [6] rotavirus strains among hospitalised children with acute gastroenteritis in Belém, Northern Brazil, following introduction of a rotavirus vaccine. *Archives of Virology*, pp.1-11.

Guerra, S.F.S., Soares, L.S., Lobo, P.S., Júnior, E.T.P., Júnior, E.C.S., Bezerra, D.A.M., Vaz, L.R., Linhares, A.C. and Mascarenhas, J.D.A.P., 2016. Detection of a novel equine-like G3 rotavirus associated with acute gastroenteritis in Brazil. *Journal of General Virology*, *97*(12), pp.3131-3138.

Guntapong, R., Tacharoenmuang, R., Singchai, P., Upachai, S., Sutthiwarakom, K., Komoto, S., Tsuji, T., Tharmaphornpilas, P., Yoshikawa, T., Sangkitporn, S. and Taniguchi, K., 2017. Predominant prevalence of human rotaviruses with the G1P [8] and G8P [8] genotypes with a short RNA profile in 2013 and 2014 in Sukhothai and Phetchaboon provinces, Thailand. *Journal of Medical Virology*, *89*(4), pp.615-620.

Guo, Y., Wentworth, D.E., Stucker, K.M., Halpin, R.A., Lam, H.C., Marthaler, D., Saif, L.J. and Vlasova, A.N., 2020. Amino Acid Substitutions in Positions 385 and 393 of the Hydrophobic Region of VP4 May Be Associated with Rotavirus Attenuation and Cell Culture Adaptation. *Viruses*, *12*(4), p.408.

Gurgel, R.Q., Cuevas, L.E., Vieira, S.C., Barros, V.C., Fontes, P.B., Salustino, E.F., Nakagomi, O., Nakagomi, T., Dove, W., Cunliffe, N. and Hart, C.A., 2007. Predominance of rotavirus P [4] G2 in a vaccinated population, Brazil. *Emerging Infectious Diseases*, *13*(10), p.1571.

Gutiérrez, M., Isa, P., Sánchez-San Martin, C., Pérez-Vargas, J., Espinosa, R., Arias, C.F. and López, S., 2010. Different rotavirus strains enter MA104 cells through different endocytic pathways: the role of clathrin-mediated endocytosis. *Journal of Virology*, *84*(18), pp.9161-9169.

Hagbom, M., Sharma, S., Lundgren, O. and Svensson, L., 2012. Towards a human rotavirus disease model. *Current Opinion in Virology*, *2*(4), pp.408-418.

Halasz, P., Holloway, G. and Coulson, B.S., 2010. Death mechanisms in epithelial cells following rotavirus infection, exposure to inactivated rotavirus or genome transfection. *Journal of General Virology*, *91*(8), pp.2007-2018.

Harris, V.C., Armah, G., Fuentes, S., Korpela, K.E., Parashar, U., Victor, J.C., Tate, J., de Weerth, C., Giaquinto, C., Wiersinga, W.J. and Lewis, K.D., 2017. Significant correlation between the infant gut microbiome and rotavirus vaccine response in rural Ghana. *The Journal of Infectious Diseases*, *215*(1), pp.34-41.

Hoa-Tran, T.N., Nakagomi, T., Vu, H.M., Do, L.P., Gauchan, P., Agbemabiese, C.A., Nguyen, T.T.T., Nakagomi, O. and Thanh, N.T.H., 2016. Abrupt emergence and predominance in Vietnam of rotavirus A strains possessing a bovine-like G8 on a DS-1-like background. *Archives of Virology*, *161*(2), pp.479-482.

Hoa-Tran, T.N., Nakagomi, T., Vu, H.M., Nguyen, T.T.T., Takemura, T., Hasebe, F., Dao, A.T.H., Anh, P.H.Q., Nguyen, A.T., Dang, A.D. and Nakagomi, O., 2020. Detection of three independently-generated DS-1-like G9P [8] reassortant rotavirus A strains during the G9P [8] dominance in Vietnam, 2016–2018. *Infection, Genetics and Evolution*, p.104194.

Holloway, G., Dang, V.T., Jans, D.A. and Coulson, B.S., 2014. Rotavirus inhibits IFN-induced STAT nuclear translocation by a mechanism that acts after STAT binding to importin-α. *Journal of General Virology*, *95*(8), pp.1723-1733.

Hoque, S.A., Khandoker, N., Thongprachum, A., Khamrin, P., Takanashi, S., Okitsu, S., Nishimura, S., Kikuta, H., Yamamoto, A., Sugita, K. and Baba, T., 2020. Distribution of rotavirus genotypes in Japan from 2015 to 2018: Diversity in genotypes before and after introduction of rotavirus vaccines. *Vaccine*.

Hoshino, Y. and Kapikian, A.Z., 1994. Rotavirus vaccine development for the prevention of severe diarrhea in infants and young children. *Trends in Microbiology*, *2*(7), pp.242-249.

Hoxie, I. and Dennehy, J.J., 2020. Intragenic recombination influences rotavirus diversity and evolution. *Virus Evolution*, *6*(1), p.59.

Hsieh, Y.C., Wu, F.T., Hsiung, C.A., Wu, H.S., Chang, K.Y. and Huang, Y.C., 2014. Comparison of virus shedding after lived attenuated and pentavalent reassortant rotavirus vaccine. *Vaccine*, *32*(10), pp.1199-1204.

Hua, J., Chen, X. and Patton, J.T., 1994. Deletion mapping of the rotavirus metalloprotein NS53 (NSP1): the conserved cysteine-rich region is essential for virus-specific RNA binding. *Journal of Virology*, *68*(6), pp.3990-4000.

Huang, P., Xia, M., Tan, M., Zhong, W., Wei, C., Wang, L., Morrow, A. and Jiang, X., 2012. Spike protein VP8* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. *Journal of Virology*, *86*(9), pp.4833-4843.

Hull, J.J., Teel, E.N., Kerin, T.K., Freeman, M.M., Esona, M.D., Gentsch, J.R., Cortese, M.M., Parashar, U.D., Glass, R.I. and Bowen, M.D., 2011. United States rotavirus strain surveillance from 2005 to 2008: genotype prevalence before and after vaccine introduction. *The Pediatric Infectious Disease Journal*, *30*(1), pp.S42-S47.

Huson, D.H. and Scornavacca, C., 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, *61*(6), pp.1061-1067.

Hyser, J.M., Zeng, C.Q.Y., Beharry, Z., Palzkill, T. and Estes, M.K., 2008. Epitope mapping and use of epitope-specific antisera to characterize the VP5∗ binding site in rotavirus SA11 NSP4. *Virology*, *373*(1), pp.211-228.

Isa, P., Arias, C.F. and López, S., 2006. Role of sialic acids in rotavirus infection. *Glycoconjugate Journal*, *23*(1-2), pp.27-37.

Iturriza-Gómara, M. and Gray, J., 2011. Euro Rota Net, 4th Year Report.

Iturriza-Gómara, M., Green, J., Brown, D.W., Ramsay, M., Desselberger, U. and Gray, J.J., 2000. Molecular epidemiology of human group A rotavirus infections in the United Kingdom between 1995 and 1998. *Journal of Clinical Microbiology*, *38*(12), pp.4394-4401.

Iturriza-Gómara, M., Isherwood, B., Desselberger, U. and Gray, J.I.M., 2001. Reassortment in vivo: driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. *Journal of virology*, *75*(8), pp.3696-3705.

Ivashechkin, A.A., Yuzhakov, A.G., Grebennikova, T.V., Yuzhakova, K.A., Kulikova, N.Y., Kisteneva, L.B., Smetanina, S.V., Bazarova, M.V. and Almazova, M.G., 2020. Genetic diversity of group A rotaviruses in Moscow in 2018-2019. *Archives of Virology*, *165*(3), pp.691-702.

Jaimes, M.C., Rojas, O.L., González, A.M., Cajiao, I., Charpilienne, A., Pothier, P., Kohli, E., Greenberg, H.B., Franco, M.A. and Angel, J., 2002. Frequencies of virus-specific CD4+ and CD8+ T lymphocytes secreting gamma interferon after acute natural rotavirus infection in children and adults. *Journal of Virology*, *76*(10), pp.4741-4749.

Jayaram, H., Estes, M.K. and Prasad, B.V., 2004. Emerging themes in rotavirus cell entry, genome organization, transcription and replication. *Virus Research*, *101*(1), pp.67-81

Jayaram, H., Taraporewala, Z., Patton, J.T. and Prasad, B.V., 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature*, *417*(6886), pp.311-315.

Jere, K.C., Chaguza, C., Bar-Zeev, N., Lowe, J., Peno, C., Kumwenda, B., Nakagomi, O., Tate, J.E., Parashar, U.D., Heyderman, R.S. and French, N., 2018. Emergence of double-and triple-gene reassortant G1P [8] rotaviruses possessing a DS-1-like backbone after rotavirus vaccine introduction in Malawi. *Journal of Virology*, *92*(3), pp.e01246-17.

Jere, K.C., Mlera, L., Page, N.A., van Dijk, A.A. and O'Neill, H.G., 2011. Whole genome analysis of multiple rotavirus strains from a single stool specimen using sequence-independent amplification and 454® pyrosequencing reveals evidence of intergenotype genome segment recombination. *Infection, Genetics and Evolution*, *11*(8), pp.2072-2082.

Jiang, X., Jayaram, H., Kumar, M., Ludtke, S.J., Estes, M.K. and Prasad, B.V., 2006. Cryoelectron microscopy structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: implications for genome replication. *Journal of Virology*, *80*(21), pp.10829-10835.

Jing, Z., Zhang, X., Shi, H., Chen, J., Shi, D., Dong, H. and Feng, L., 2018. A G3P [13] porcine group A rotavirus emerging in China is a reassortant and a natural recombinant in the VP 4 gene. *Transboundary and Emerging Diseases*, *65*(2), pp.e317-e328.

Kaljot, K.T., Shaw, R.D., Rubin, D.H. and Greenberg, H.B., 1988. Infectious rotavirus enters cells by direct cell membrane penetration, not by endocytosis. *Journal of Virology*, *62*(4), pp.1136-1144.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. and Thierer, T., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), pp.1647-1649.

Khandoker, N., Thongprachum, A., Takanashi, S., Okitsu, S., Nishimura, S., Kikuta, H., Yamamoto, A., Sugita, K., Baba, T., Kobayashi, M. and Hayakawa, S., 2018. Molecular epidemiology of rotavirus gastroenteritis in Japan during 2014-2015: Characterization of re-emerging G2P [4] after rotavirus vaccine introduction. *Journal of Medical Virology*, *90*(6), pp.1040-1046.

Kim, B., Feng, N., Narváez, C.F., He, X.S., Eo, S.K., Lim, C.W. and Greenberg, H.B., 2008. The influence of CD4+ CD25+ Foxp3+ regulatory T cells on the immune response to rotavirus infection. *Vaccine*, *26*(44), pp.5601-5611.

Kirkwood, C.D. and Steele, A.D., 2018. Rotavirus vaccines in China: improvement still required. *JAMA Network Open*, *1*(4), pp.e181579-e181579.

Kirkwood, C.D., 2010. Genetic and antigenic diversity of human rotaviruses: potential impact on vaccination programs. *Journal of Infectious Diseases*, *202*(Supplement_1), pp.S43-S48.

Kirkwood, C.D., Boniface, K., Barnes, G.L. and Bishop, R.F., 2011. Distribution of rotavirus genotypes after introduction of rotavirus vaccines, Rotarix® and RotaTeq®, into the National Immunization Program of Australia. *The Pediatric Infectious Disease Journal*, *30*(1), pp.S48-S53.

Kirkwood, C.D., Ma, L.F., Carey, M.E. and Steele, A.D., 2019. The rotavirus vaccine development pipeline. *Vaccine*, *37*(50), pp.7328-7335.

Kitamoto, N., Mattion, N.M. and Estes, M.K., 1993. Alterations in the sequence of the gene 4 from a human rotavirus after multiple passages in HepG2 liver cells. *Archives of Virology*, *130*(1-2), pp.179-185.

Kobayashi, N., Alam, M., Kojima, K., Mise, K., Ishino, M. and Sumi, A., 2003. Genomic diversity and evolution of rotaviruses: an overview. *Genomic Diversity and Molecular Epidemiology of Rotaviruses*, pp.75-90.

Komoto, S., Ide, T., Negoro, M., Tanaka, T., Asada, K., Umemoto, M., Kuroki, H., Ito, H., Tanaka, S., Ito, M. and Fukuda, S., 2018. Characterization of unusual DS-1-like G3P [8] rotavirus strains in children with diarrhea in Japan. *Journal of Medical Virology*, *90*(5), pp.890-898.

Komoto, S., Tacharoenmuang, R., Guntapong, R., Ide, T., Haga, K., Katayama, K., Kato, T., Ouchi, Y., Kurahashi, H., Tsuji, T. and Sangkitporn, S., 2015. Emergence and characterization of unusual DS-1-like G1P [8] rotavirus strains in children with diarrhea in Thailand. *PLoS One*, *10*(11), p.e0141739.

Komoto, S., Tacharoenmuang, R., Guntapong, R., Ide, T., Tsuji, T., Yoshikawa, T., Tharmaphornpilas, P., Sangkitporn, S. and Taniguchi, K., 2016. Reassortment of human and animal rotavirus gene segments in emerging DS-1-like G1P [8] rotavirus strains. *PLoS One*, *11*(2), p.e0148416.

Kondo, K., Tsugawa, T., Ono, M., Ohara, T., Fujibayashi, S., Tahara, Y., Kubo, N., Nakata, S., Higashidate, Y., Fujii, Y. and Katayama, K., 2017. Clinical and molecular characteristics of human rotavirus G8P [8] outbreak strain, Japan, 2014. *Emerging Infectious Diseases*, *23*(6), p.968.

Korpe, P.S. and Petri Jr, W.A., 2012. Environmental enteropathy: critical implications of a poorly understood condition. *Trends in Molecular Medicine*, *18*(6), pp.328-336.

Kosakovsky Pond, S.L. and Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), pp.1208-1222.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. and Frost, S.D., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, *23*(10), pp.1891-1901.

Krieger, E. and Vriend, G., 2014. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, *30*(20), pp.2981-2982.

Krieger, E., Koraimann, G. and Vriend, G., 2002. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Structure, Function, and Bioinformatics*, *47*(3), pp.393-402.

Kühn, U. and Wahle, E., 2004. Structure and function of poly (A) binding proteins. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1678*(2-3), pp.67-84.

Kulkarni, P.S., Desai, S., Tewari, T., Kawade, A., Goyal, N., Garg, B.S., Kumar, D., Kanungo, S., Kamat, V., Kang, G. and Bavdekar, A., 2017. A randomized Phase III clinical trial to assess the efficacy of a bovine-human reassortant pentavalent rotavirus vaccine in Indian infants. *Vaccine*, *35*(45), pp.6228-6237.

Kulski, J.K., 2016. Next-generation sequencing—an overview of the history, tools, and "Omic" applications. *Next Generation Sequencing–Advances, Applications and Challenges*, pp.3-60.

Kumar, D., Yu, X., Crawford, S.E., Moreno, R., Jakana, J., Sankaran, B., Anish, R., Kaundal, S., Hu, L., Estes, M.K. and Wang, Z., 2020. 2.7 Å cryo-EM structure of rotavirus core protein VP3, a unique capping machine with a helicase activity. *Science Advances*, *6*(16), p.eaay6410.

Kuzuya, M., Fujii, R., Hamano, M., Kida, K., Mizoguchi, Y., Kanadani, T., Nishimura, K. and Kishimoto, T., 2014. Prevalence and molecular characterization of G1P [8] human rotaviruses possessing DS-1-like VP6, NSP4, and NSP5/6 in Japan. *Journal of Medical Virology*, *86*(6), pp.1056-1064.

Lartey, B.L., Damanka, S., Dennis, F.E., Enweronu-Laryea, C.C., Addo-Yobo, E., Ansong, D., Kwarteng-Owusu, S., Sagoe, K.W., Mwenda, J.M., Diamenu, S.K. and Narh, C., 2018. Rotavirus strain distribution in Ghana pre-and post-rotavirus vaccine introduction. *Vaccine*, *36*(47), pp.7238-7242.

Le, V.P., Chung, Y.C., Kim, K., Chung, S.I., Lim, I. and Kim, W., 2010. Genetic variation of prevalent G1P [8] human rotaviruses in South Korea. *Journal of Medical Virology*, *82*(5), pp.886-896.

Le, L.T., Nguyen, T.V., Nguyen, P.M., Huong, N.T., Huong, N.T., Huong, N.T., Hanh, T.B., Ha, D.N., Anh, D.D., Gentsch, J.R. and Wang, Y., 2009. Development and characterization of candidate

rotavirus vaccine strains derived from children with diarrhoea in Vietnam. *Vaccine*, *27*, pp.F130-F138.

Lee, B., Dickson, D.M., deCamp, A.C., Ross Colgate, E., Diehl, S.A., Uddin, M.I., Sharmin, S., Islam, S., Bhuiyan, T.R., Alam, M. and Nayak, U., 2018. Histo–Blood Group Antigen Phenotype Determines Susceptibility to Genotype-Specific Rotavirus Infections and Impacts Measures of Rotavirus Vaccine Efficacy. *The Journal of Infectious Diseases*, *217*(9), pp.1399-1407.

Leshem, E., Lopman, B., Glass, R., Gentsch, J., Bányai, K., Parashar, U. and 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), pp.847-856.

Li, R.C., Huang, T., Li, Y., Wang, L.H., Tao, J., Fu, B., Si, G., Nong, Y., Mo, Z., Liao, X. and Luan, I., 2016. Immunogenicity and reactogenicity of the human rotavirus vaccine, RIX4414 oral suspension, when co-administered with routine childhood vaccines in Chinese infants. *Human Vaccines & Immunotherapeutics*, *12*(3), pp.785-793.

Lin, J.D., Feng, N., Sen, A., Balan, M., Tseng, H.C., McElrath, C., Smirnov, S.V., Peng, J., Yasukawa, L.L., Durbin, R.K. and Durbin, J.E., 2016. Distinct roles of type I and type III interferons in intestinal immunity to homologous and heterologous rotavirus infections. *PLoS Pathogens*, *12*(4), p.e1005600.

Ling-Qiao, Z., 2001. A rotavirus vaccine licensed in China. *Health News*, *31*(1).

Linhares, A.C. and Justino, M.C.A., 2014. Rotavirus vaccination in Brazil: effectiveness and health impact seven years post-introduction. *Expert Review of Vaccines*, *13*(1), pp.43-57.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M., 2012. Comparison of next-generation sequencing systems. *BioMed Research International*, *2012*.

López, S. and Arias, C.F., 2012. Rotavirus–host cell interactions: an arms race. *Current Opinion in Virology*, *2*(4), pp.389-398.

López, S. and Arias, C.F., 2017. Rotavirus Biology. In *Human Virology in Latin America* (pp. 19-42). Springer, Cham.

Luchs, A., da Costa, A.C., Cilli, A., Komninakis, S.C.V., Carmona, R.D.C.C., Morillo, S.G., Sabino, E.C. and Timenetsky, M.D.C.S.T., 2019. First Detection of DS-1-like G1P [8] Double-gene Reassortant Rotavirus Strains on The American Continent, Brazil, 2013. *Scientific Reports*, *9*(1), p.2210.

Ludert, J.E., Feng, N., Yu, J.H., Broome, R.L., Hoshino, Y. and Greenberg, H.B., 1996. Genetic mapping indicates that VP4 is the rotavirus cell attachment protein in vitro and in vivo. *Journal of Virology*, *70*(1), pp.487-493.

Ludert, J.E., Michelangeli, F., Gil, F., Liprandi, F. and Esparza, J., 1987. Penetration and uncoating of rotaviruses in cultured cells. *Intervirology*, *27*(2), pp.95-101.

Madhi, S. A., Cunliffe, N. A., Steele, D., Witte, D., Kirsten, M., Louw, C., Ngwira, B., Victor, J. C., Gillard, P. H., Cheuvart, B. B., Han, H. H., and Neuzil, K. M. 2010. Effect of human rotavirus vaccine on severe diarrhea in African infants. *The New England Journal of Medicine*, *362*(4), 289–298.

Madhi, S.A., Kirsten, M., Louw, C., Bos, P., Aspinall, S., Bouckenooghe, A., Neuzil, K.M. and Steele, A.D., 2012. Efficacy and immunogenicity of two or three dose rotavirus-vaccine regimen in South African children over two consecutive rotavirus-seasons: a randomized, double-blind, placebo-controlled trial. *Vaccine*, *30*, pp.A44-A51.

Madhi, S. A. *et al.* 2016. 'Effect of human rotavirus vaccine on severe diarrhea in African infants', *Malawi Medical Journal*, 28(3), pp. 108–114. doi: 10.1056/NEJMoa0904797.

Maes, P., Matthijnssens, J., Rahman, M. and Van Ranst, M., 2009. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. *BMC Microbiology*, *9*(1), p.238.

Magagula, N.B., Esona, M.D., Nyaga, M.M., Stucker, K.M., Halpin, R.A., Stockwell, T.B., Seheri, M.L., Steele, A.D., Wentworth, D.E. and Mphahlele, M.J., 2015. Whole genome analyses of G1P [8] rotavirus strains from vaccinated and non-vaccinated South African children presenting with diarrhea. *Journal of Medical Virology*, *87*(1), pp.79-101.

Maiorov, V.N. and Crippen, G.M., 1994. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins.

Mansell, E.A. and Patton, J.T., 1990. Rotavirus RNA replication: VP2, but not VP6, is necessary for viral replicase activity. *Journal of Virology*, *64*(10), pp.4988-4996.

Mapaseka, S.L., Dewar, J.B., Van Der Merwe, L., Geyer, A., Tumbo, J., Zweygarth, M., Bos, P., Esona, M.D., Duncan Steele, A. and Sommerfelt, H., 2010. Prospective hospital-based surveillance to estimate rotavirus disease burden in the Gauteng and North West Province of South Africa during 2003–2005. *Journal of Infectious Diseases*, *202*(Supplement_1), pp.S131-S138.

Mardis, E.R., 2013. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, *6*, pp.287-303.

Maringa, W.M., Mwangi, P.N., Simwaka, J., Mpabalwani, E.M., Mwenda, J.M., Peenze, I., Esona, M.D., Mphahlele, M.J., Seheri, M.L. and Nyaga, M.M.,2020. Molecular Characterisation of a Rare Reassortant Porcine-Like G5P[6] Rotavirus Strain Detected in an Unvaccinated Child in Kasama, Zambia. *Pathogens*, *9*.pp 663.

Markkula, J., Hemming-Harlo, M., Salminen, M.T., Savolainen-Kopra, C., Pirhonen, J., Al-Hello, H. and Vesikari, T., 2017. Rotavirus epidemiology 5–6 years after universal rotavirus vaccination: persistent rotavirus activity in older children and elderly. *Infectious Diseases*, *49*(5), pp.388-395.

Marthaler, D., Rossow, K., Culhane, M., Goyal, S., Collins, J., Matthijnssens, J., Nelson, M. and Ciarlet, M., 2014. Widespread rotavirus H in commercially raised pigs, United States. *Emerging Infectious Diseases*, *20*(7), p.1203.

Martin, D. and Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, *16*(6), pp.562-563

Martin, D., Ouldali, M., Ménétrey, J. and Poncet, D., 2011. Structural organisation of the rotavirus nonstructural protein NSP5. *Journal of Molecular Biology*, *413*(1), pp.209-221.

Martinez-Laso, J., Roman, A., Rodriguez, M., Cervera, I., Head, J., Rodriguez-Avial, I. and Picazo, J.J., 2009. Diversity of the G3 genes of human rotaviruses in isolates from Spain from 2004 to 2006: cross-species transmission and inter-genotype recombination generates alleles. *Journal of General Virology*, *90*(4), pp.935-943.

Mathieu, M., Petitpas, I., Navaza, J., Lepault, J., Kohli, E., Pothier, P., Prasad, B.V., Cohen, J. and Rey, F.A., 2001. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. *The EMBO Journal*, *20*(7), pp.1485-1497.

Matthijnssens, J. and Van Ranst, M., 2012. Genotype constellation and evolution of group A rotaviruses infecting humans. *Current Opinion in Virology*, *2*(4), pp.426-433.

Matthijnssens, J., Bilcke, J., Ciarlet, M., Martella, V., Bányai, K., Rahman, M., Zeller, M., Beutels, P., Van Damme, P. and Van Ranst, M., 2009. Rotavirus disease and vaccination: impact on genotype diversity. *Future Microbiology*, *4*(10), pp.1303-1316.

Matthijnssens, J., Ciarlet, M., McDonald, S.M., Attoui, H., Bányai, K., Brister, J.R., Buesa, J., Esona, M.D., Estes, M.K., Gentsch, J.R. and Iturriza-Gómara, M., 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Archives of Virology*, *156*(8), pp.1397-1413.

Matthijnssens, J., Ciarlet, M., Rahman, M., Attoui, H., Bányai, K., Estes, M.K., Gentsch, J.R., Iturriza-Gómara, M., Kirkwood, C.D., Martella, V. and Mertens, P.P., 2008. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. *Archives of Virology*, *153*(8), pp.1621-1629.

Matthijnssens, J., Zeller, M., Heylen, E., De Coster, S., Vercauteren, J., Braeckman, T., Van Herck, K., Meyer, N., Pirçon, J.Y., Soriano-Gabarro, M. and Azou, M., 2014. Higher proportion of G2P [4] rotaviruses in vaccinated hospitalized cases compared with unvaccinated hospitalized cases, despite high vaccine effectiveness against heterotypic G2P [4] rotaviruses. *Clinical Microbiology and Infection*, *20*(10), pp.O702-O710.

Mattion, N.M., Cohen, J., Aponte, C. and Estes, M.K., 1992. Characterization of an oligomerization domain and RNA-binding properties on rotavirus nonstructural protein NS34. *Virology*, *190*(1), pp.68-83.

McClain, B., Settembre, E., Temple, B.R., Bellamy, A.R. and Harrison, S.C., 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 Å resolution. *Journal of Molecular Biology*, *397*(2), pp.587-599.

McDonald, S.M., Nelson, M.I., Turner, P.E. and Patton, J.T., 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature Reviews Microbiology*, *14*(7), p.448.

McDonald, S.M., Tao, Y.J. and Patton, J.T., 2009. The ins and outs of four-tunneled Reoviridae RNA-dependent RNA polymerases. *Current Opinion in Structural Biology*, *19*(6), pp.775-782.

Mesa, M.C., Gutiérrez, L., Duarte-Rey, C., Angel, J. and Franco, M.A., 2010. A TGF-β mediated regulatory mechanism modulates the T cell immune response to rotavirus in adults but not in children. *Virology*, *399*(1), pp.77-86.

Mihalov-Kovács, E., Gellért, Á., Marton, S., Farkas, S.L., Fehér, E., Oldal, M., Jakab, F., Martella, V. and Bányai, K., 2015. Candidate new rotavirus species in sheltered dogs, Hungary. *Emerging Infectious Diseases*, *21*(4), p.660.

Mishra, N., Reslan, L., El-Husseini, M., Raoof, H., Finianos, M., Guo, C., Thakkar, R., Inati, A., Dbaibo, G., Lipkin, W.I. and Zaraket, H., 2020. Full genome characterization of human G3P [6] and G3P [9] rotavirus strains in Lebanon. *Infection, Genetics and Evolution*, *78*, p.104133.

Mokomane, M., Esona, M.D., Bowen, M.D., Tate, J.E., Steenhoff, A.P., Lechiile, K., Gaseitsiwe, S., Seheri, L.M., Magagula, N.B., Weldegebriel, G. and Pernica, J.M., 2019. Diversity of Rotavirus Strains Circulating in Botswana before and after introduction of the Monovalent Rotavirus Vaccine. *Vaccine*, *37*(43), pp.6324-6328.

Monnier, N., Higo-Moriguchi, K., Sun, Z.Y.J., Prasad, B.V., Taniguchi, K. and 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), pp.1513-1523.

Moon, S.S., Groome, M.J., Velasquez, D.E., Parashar, U.D., Jones, S., Koen, A., van Niekerk, N., Jiang, B. and Madhi, S.A., 2016. Prevaccination rotavirus serum IgG and IgA are associated with lower immunogenicity of live, oral human rotavirus vaccine in South African infants. *Clinical Infectious Diseases*, *62*(2), pp.157-165.

Morelli, M., Ogden, K.M. and Patton, J.T., 2015. Silencing the alarms: innate immune antagonism by rotavirus NSP1 and VP3. *Virology*, *479*, pp.75-84.

Morozova, O.V., Sashina, T.A., Fomina, S.G. and 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), pp.1693-1703.

Moyo, S.J., Blomberg, B., Hanevik, K., Kommedal, O., Vainio, K., Maselle, S.Y. and Langeland, N., 2014. Genetic diversity of circulating rotavirus strains in Tanzania prior to the introduction of vaccination. *Public Library of Science One*, *9*(5).

Msimang, V.M., Page, N., Groome, M.J., Moyes, J., Cortese, M.M., Seheri, M., Kahn, K., Chagan, M., Madhi, S.A. and Cohen, C., 2013. Impact of rotavirus vaccine on childhood diarrheal hospitalization after introduction into the South African public immunization program. *The Pediatric Infectious Disease Journal*, *32*(12), pp.1359-1364.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. and Scheffler, K., 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), pp.1196-1205.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. and Pond, S.L.K., 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), p.e1002764.

Mwangi, P.N., Mogotsi, M.T., Rasebotsa, S.P., Seheri, M.L., Mphahlele, M.J., Ndze, V.N., Dennis, F.E., Jere, K.C. and Nyaga, M.M., 2020. Uncovering the first atypical DS-1-like G1P [8] rotavirus strains that circulated during pre-rotavirus vaccine Introduction era in South Africa. *Pathogens*, *9*(5), p.391.

Mwenda, J.M., Mandomando, I., Jere, K.C., Cunliffe, N.A. and Steele, A.D., 2019. Evidence of reduction of rotavirus diarrheal disease after rotavirus vaccine introduction in national immunization programs in the African countries: Report of the 11th African rotavirus.

Mwenda, J.M., Ntoto, K.M., Abebe, A., Enweronu-Laryea, C., Amina, I., Mchomvu, J., Kisakye, A., Mpabalwani, E.M., Pazvakavambwa, I., Armah, G.E. and Seheri, L.M., 2010. Burden and epidemiology of rotavirus diarrhea in selected African countries: preliminary results from the African Rotavirus Surveillance Network. *Journal of Infectious Diseases*, *202*(Supplement_1), pp.S5-S11.

Mwenda, J.M., Tate, J.E., Parashar, U.D., Mihigo, R., Agócs, M., Serhan, F. and Nshimirimana, D., 2014. African rotavirus surveillance network: a brief overview. *The Pediatric Infectious Disease Journal*, *33*, pp.S6-S8.

Naik, S.P., Zade, J.K., Sabale, R.N., Pisal, S.S., Menon, R., Bankar, S.G., Gairola, S. and Dhere, R.M., 2017. Stability of heat stable, live attenuated Rotavirus vaccine (ROTASIIL®). *Vaccine*, *35*(22), pp.2962-2969.

Naylor, C., Lu, M., Haque, R., Mondal, D., Buonomo, E., Nayak, U., Mychaleckyj, J.C., Kirkpatrick, B., Colgate, R., Carmolli, M. and Dickson, D., 2015. Environmental enteropathy, oral vaccine failure and growth faltering in infants in Bangladesh. *EBioMedicine*, *2*(11), pp.1759-1766.

Nordgren, J., Nitiema, L.W., Ouermi, D., Simpore, J. and Svensson, L., 2013. Host genetic factors affect susceptibility to norovirus infections in Burkina Faso. *PloS one*, *8*(7).

Nordgren, J., Sharma, S., Bucardo, F., Nasir, W., Günaydın, G., Ouermi, D., Nitiema, L.W., Becker-Dreps, S., Simpore, J., Hammarström, L. and Larson, G., 2014. Both Lewis and secretor status mediate susceptibility to rotavirus infections in a rotavirus genotype–dependent manner. *Clinical Infectious Diseases*, *59*(11), pp.1567-1573.

Nyaga, M.M., Esona, M.D., Jere, K.C., Peenze, I., Seheri, M.L. and Mphahlele, M.J., 2014a. Genetic diversity of rotavirus genome segment 6 (encoding VP6) in Pretoria, South Africa. *SpringerPlus*, *3*(1), pp.1-5.

Nyaga, M.M., Jere, K.C., Esona, M.D., Seheri, M.L., Stucker, K.M., Halpin, R.A., Akopov, A., Stockwell, T.B., Peenze, I., Diop, A. and Ndiaye, K., 2015. Whole genome detection of rotavirus mixed infections in human, porcine and bovine samples co-infected with various rotavirus strains collected from sub-Saharan Africa. *Infection, Genetics and Evolution*, *31*, pp.321-334.

Nyaga, M.M., Peenze, I., Potgieter, C.A., Seheri, L.M., Page, N.A., Yinda, C.K., Steele, A.D., Matthijnssens, J. and Mphahlele, M.J., 2016. Complete genome analyses of the first porcine rotavirus group H identified from a South African pig does not provide evidence for recent interspecies transmission events. *Infection, Genetics and Evolution*, *38*, pp.1-7.

Nyaga, M.M., Sabiu, S., Ndze, V.N., Dennis, F.E. and Jere, K.C., 2020. Report of the 1st African Enteric Viruses Genome Initiative data and Bioiniformatics workshop on whole genome analysis of some African rotavirus strains held in Bloemfontein, South Africa. *Vaccine*, *38,* pp.5402-5407.

Nyaga, M.M., Stucker, K.M., Esona, M.D., Jere, K.C., Mwinyi, B., Shonhai, A., Tsolenyanu, E., Mulindwa, A., Chibumbya, J.N., Adolfine, H. and Halpin, R.A., 2014b. Whole-genome analyses of DS-1-like human G2P [4] and G8P [4] rotavirus strains from Eastern, Western and Southern Africa. *Virus genes*, *49*(2), pp.196-207.

Nyaga, M.M., Tan, Y., Seheri, M.L., Halpin, R.A., Akopov, A., Stucker, K.M., Fedorova, N.B., Shrivastava, S., Steele, A.D., Mwenda, J.M. and Pickett, B.E., 2018. Whole-genome sequencing and analyses identify high genetic heterogeneity, diversity and endemicity of rotavirus genotype P [6] strains circulating in Africa. *Infection, Genetics and Evolution*, *63*, pp.79-88.

O'Ryan, M., Lucero, Y., O'Ryan-Soriano, M.A. and Ashkenazi, S., 2010. An update on management of severe acute infectious gastroenteritis in children. *Expert Review of Anti-infective Therapy*, *8*(6), pp.671-682.

Ogden, K.M., Snyder, M.J., Dennis, A.F. and Patton, J.T., 2014. Predicted structure and domain organization of rotavirus capping enzyme and innate immune antagonist VP3. *Journal of Virology*, *88*(16), pp.9072-9085.

Page, N.A., Seheri, L.M., Groome, M.J., Moyes, J., Walaza, S., Mphahlele, J., Kahn, K., Kapongo, C.N., Zar, H.J., Tempia, S. and Cohen, C., 2018. Temporal association of rotavirus vaccination and genotype circulation in South Africa: Observations from 2002 to 2014. *Vaccine*, *36*(47), pp.7231-7237.

Page, N., Kruger, T., Seheri, M., Peenze, I., Quan, V., Groome, M. and Madhi, S., 2016. Rotavirus surveillance report, South Africa, 2014-2015: a comparison with previous rotavirus seasons. *Foreword contents. Communicable Diseases Surveillance Bulletin*, *14*(4), p.126.

Page, N.A. and Steele, A.D., 2004. Antigenic and genetic characterization of serotype G2 human rotavirus strains from South Africa from 1984 to 1998. *Journal of Medical Virology*, *72*(2), pp.320-327.

Parashar, U.D., Gibson, C.J., Bresee, J.S. and Glass, R.I., 2006. Rotavirus and severe childhood diarrhea. *Emerging Infectious Diseases*, *12*(2), p.304.

Parra, G.I., 2009. Seasonal shifts of group A rotavirus strains as a possible mechanism of persistence in the human population. *Journal of Medical Virology*, *81*(3), pp.568-571.

Parra, M., Herrera, D., Jácome, M.F., Mesa, M.C., Rodríguez, L.S., Guzmán, C., Angel, J. and Franco, M.A., 2014. Circulating rotavirus-specific T cells have a poor functional profile. *Virology*, *468*, pp.340-350.

Patton, J.T., 2012. Rotavirus diversity and evolution in the post-vaccine world. *Discovery Medicine*, *13*(68), p.85.

Patton, J.T., Vasquez-Del Carpio, R., Tortorici, M.A. and Taraporewala, Z.F., 2006. Coupling of rotavirus genome replication and capsid assembly. *Advances in Virus Research*, *69*, pp.167-201.

Paul, A., Babji, S., Sarkar, R., Lazarus, R.P. and Kang, G., 2016. Rotavirus-specific salivary and fecal IgA in Indian children and adults. *Indian pediatrics*, *53*(7), pp.601-606.

Perchetti, G.A., Huang, M.L., Peddu, V., Jerome, K.R. and Greninger, A.L., 2020. Stability of SARS-CoV-2 in phosphate-buffered saline for molecular detection. *Journal of Clinical Microbiology*, *58*(8).

Pesavento, J.B., Crawford, S.E., Estes, M.K. and Prasad, B.V., 2006. Rotavirus proteins: structure and assembly. In *Reoviruses: Entry, Assembly and Morphogenesis* (pp. 189-219). Springer, Berlin, Heidelberg.

Piron, M., Delaunay, T., Grosclaude, J. and Poncet, D., 1999. Identification of the RNA-binding, dimerization, and eIF4GI-binding domains of rotavirus nonstructural protein NSP3. *Journal of Virology*, *73*(7), pp.5411-5421.

Piron, M., Vende, P., Cohen, J. and Poncet, D., 1998. Rotavirus RNA-binding protein NSP3 interacts with eIF4GI and evicts the poly (A) binding protein from eIF4F. *The EMBO Journal*, *17*(19), pp.5811-5821.

Poelaert, D., Pereira, P., Gardner, R., Standaert, B. and Benninghoff, B., 2018. A review of recommendations for rotavirus vaccination in Europe: arguments for change. *Vaccine*, *36*(17), pp.2243-2253

Potgieter, A.C., Page, N.A., Liebenberg, J., Wright, I.M., Landt, O. and Van Dijk, A.A., 2009. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *Journal of General Virology*, *90*(6), pp.1423-1432.

Potgieter, N., De Beer, M.C., Taylor, M.B. and Duncan Steele, A., 2010. Prevalence and diversity of rotavirus strains in children with acute diarrhea from rural communities in the Limpopo Province, South Africa, from 1998 to 2000. *Journal of Infectious Diseases*, *202*(Supplement_1), pp.S148-S155.

Prasad, B.V., Wang, G.J., Clerx, J.P. and Chiu, W., 1988. Three-dimensional structure of rotavirus. *Journal of Molecular Biology*, *199*(2), pp.269-275.

Qin, L., Ren, L., Zhou, Z., Lei, X., Chen, L., Xue, Q., Liu, X., Wang, J. and Hung, T., 2011. Rotavirus nonstructural protein 1 antagonizes innate immune response by interacting with retinoic acid inducible gene I. *Virology Journal*, *8*(1), pp.1-9.

Rahman, M., Matthijnssens, J., Saiada, F., Hassan, Z., Heylen, E., Azim, T. and Van Ranst, M., 2010. Complete genomic analysis of a Bangladeshi G1P [8] rotavirus strain detected in 2003 reveals a close evolutionary relationship with contemporary human Wa-like strains. *Infection, Genetics and Evolution*, *10*(6), pp.746-754.

Rainsford, E.W. and McCrae, M.A., 2007. Characterization of the NSP6 protein product of rotavirus gene 11. *Virus Research*, *130*(1-2), pp.193-201.

Ramani, S., Mamani, N., Villena, R., Bandyopadhyay, A.S., Gast, C., Sato, A., Laucirica, D., Clemens, R., Estes, M.K. and O'Ryan, M.L., 2016. Rotavirus serum IgA immune response in children receiving rotarix coadministered with bOPV or IPV. *The Pediatric Infectious Disease Journal*, *35*(10), pp.1137-1139.

Rasebotsa, S., Mwangi, P.N., Mogotsi, M.T., Sabiu, S., Magagula, N.B., Rakau, K., Uwimana, J., Mutesa, L., Muganga, N., Murenzi, D. and Tuyisenge, L., 2020. Whole genome and in-silico analyses of G1P [8] rotavirus strains from pre-and post-vaccination periods in Rwanda. *Scientific Reports*, *10*(1), pp.1-22.

Rivera, R., Forney, K., Castro, M.R., Rebolledo, P.A., Mamani, N., Patzi, M., Halkyer, P., Leon, J.S. and Iñiguez, V., 2013. Rotavirus genotype distribution during the pre-vaccine period in Bolivia: 2007–2008. *International Journal of Infectious Diseases*, *17*(9), pp.e762-e767.

Rodríguez, J.M. and Luque, D., 2019. Structural insights into rotavirus entry. In *Physical Virology* (pp. 45-68). Springer, Cham.

Rosa, M.E.S., Pires, I.D.C. and Gouvea, V., 2002. 1998-1999 rotavirus seasons in Juiz de Fora, Minas Gerais, Brazil: detection of an unusual G3P [4] epidemic strain. Journal of Clinical Microbiology, 40(8), pp.2837-2842.

Rose, T.L., da Silva, M.F.M., Goméz, M.M., Resque, H.R., Ichihara, M.Y.T., de Mello Volotão, E. and Leite, J.P.G., 2013. Evidence of vaccine-related reassortment of rotavirus, Brazil, 2008–2010. *Emerging Infectious Diseases*, *19*(11), p.1843.

Ruiz-Palacios, G.M., Pérez-Schael, I., Velázquez, F.R., Abate, H., Breuer, T., Clemens, S.C., Cheuvart, B., Espinoza, F., Gillard, P., Innis, B.L. and Cervantes, Y., 2006. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *New England Journal of Medicine*, *354*(1), pp.11-22.

Santos, F.S., Junior, E.S., Guerra, S.F.S., Lobo, P.S., Junior, E.P., Lima, A.B.F., Vinente, C.B.G., Chagas, E.H.N., Justino, M.C.A., Linhares, A.C. and Matthijnssens, J., 2019. G1P [8] Rotavirus in children with severe diarrhea in the post-vaccine introduction era in Brazil: Evidence of reassortments and structural modifications of the antigenic VP7 and VP4 regions. *Infection, Genetics and Evolution*, *69*, pp.255-266.

Sastri, N.P., Crawford, S.E. and Estes, M.K., 2016. Pleiotropic Properties of Rotavirus Nonstructural Protein 4 (NSP4) and Their Effects on Viral Replication and Pathogenesis. In *Viral Gastroenteritis* (pp. 145-174). Academic Press.

Schellack, N., Naested, C., Schellack, G., Meyer, H., Motubatse, J., Mametja, K. and Makola, F., 2017. The management of Rotavirus disease in children. *SA Pharmaceutical Journal*, *84*(6), pp.46-51.

Schuck, P., Taraporewala, Z., McPhie, P. and Patton, J.T., 2001. Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes. *Journal of Biological Chemistry*, *276*(13), pp.9679-9687.

Seheri, L,M.; Ngomane, G.; Page, N.A.; Mokomane, M.; Maphalala, G.P.; Weldegebriel, G.; Peenze, I.; Magagula, N.B.; Nyaga, M.M.; Lisoga, J.; Dlamini, X.; Lukhele, N.; Mine, M.O.; Kuswani, K.; Mwenda, J.M. Mphahlele., 2019. Outbreak investigation of diarrheal disease in Botswana and Eswatini in 2018. 12th African Rotavirus Symposium, Johannesburg, South Africa, 30 July – 01 August 2019.

Seheri, L.M., Magagula, N.B., Peenze, I., Rakau, K., Ndadza, A., Mwenda, J.M., Weldegebriel, G., Steele, A.D. and Mphahlele, M.J., 2018. Rotavirus strain diversity in Eastern and Southern African countries before and after vaccine introduction. *Vaccine*, *36*(47), pp.7222-7230.

Settembre, E.C., Chen, J.Z., Dormitzer, P.R., Grigorieff, N. and Harrison, S.C., 2011. Atomic model of an infectious rotavirus particle. *The EMBO Journal*, *30*(2), pp.408-416.

Shah, M.P., Tate, J.E., Mwenda, J.M., Steele, A.D. and Parashar, U.D., 2017. Estimated reductions in hospitalizations and deaths from childhood diarrhea following implementation of rotavirus vaccination in Africa. *Expert Review of Vaccines*, *16*(10), pp.987-995.

Shillcutt, S.D., LeFevre, A.E., Fischer Walker, C.L., Taneja, S., Black, R.E. and Mazumder, S., 2016. Economic costs to caregivers of diarrhoea treatment among children below 5 in rural Gujarat India: findings from an external evaluation of the DAZT programme. *Health Policy and Planning*, *31*(10), pp.1411-1422.

Shintani, T., Ghosh, S., Wang, Y.H., Zhou, X., Zhou, D.J. and Kobayashi, N., 2012. Whole genomic analysis of human G1P [8] rotavirus strains from different age groups in China. *Viruses*, *4*(8), pp.1289-1304.

Siepel, A.C., Halpern, A.L., Macken, C. and Korber, B.T., 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Research and Human Retroviruses*, *11*(11), pp.1413-1416.

Silverman, R.H. and Weiss, S.R., 2014. Viral phosphodiesterases that antagonize double-stranded RNA signaling to RNase L by degrading 2-5A. *Journal of Interferon & Cytokine Research*, *34*(6), pp.455-463.

Silvestri, L.S., Taraporewala, Z.F. and Patton, J.T., 2004. Rotavirus replication: plus-sense templates for double-stranded RNA synthesis are made in viroplasms. *Journal of Virology*, *78*(14), pp.7763-7774.

Silvestri, L.S., Tortorici, M.A., Vasquez-Del Carpio, R. and Patton, J.T., 2005. Rotavirus glycoprotein NSP4 is a modulator of viral transcription in the infected cell. *Journal of Virology*, *79*(24), pp.15165-15174.

Simwaka, J.C., Mpabalwani, E.M., Seheri, M., Peenze, I., Monze, M., Matapo, B., Parashar, U.D., Mufunda, J., Mphahlele, J.M., Tate, J.E. and Mwenda, J.M., 2018. Diversity of rotavirus strains circulating in children under five years of age who presented with acute gastroenteritis before and after rotavirus vaccine introduction, University Teaching Hospital, Lusaka, Zambia, 2008–2015. *Vaccine*, *36*(47), pp.7243-7247.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, *34*(2), pp.126-129.

Steele, A.D. and Glass, R., 2011. Rotavirus in South Africa: From discovery to vaccine introduction. *Southern African Journal of Epidemiology and Infection*, *26*(4), pp.184-190.

Steele, A.D., Cunliffe, N., Tumbo, J., Madhi, S.A., De Vos, B. and Bouckenooghe, A., 2009. A review of rotavirus infection in and vaccination of human immunodeficiency virus–infected children. *The Journal of infectious diseases*, *200*(Supplement_1), pp.S57-S62.

Steele, A.D., De Vos, B., Tumbo, J., Reynders, J., Scholtz, F., Bos, P., de Beer, M.C., Van der Merwe, C.F. and Delem, A., 2010. Co-administration study in South African infants of a live-attenuated oral human rotavirus vaccine (RIX4414) and poliovirus vaccines. *Vaccine*, *28*(39), pp.6542-6548.

Steele, A.D., Neuzil, K.M., Cunliffe, N.A., Madhi, S.A., Bos, P., Ngwira, B., Witte, D., Todd, S., Louw, C., Kirsten, M. and Aspinall, S., 2012. Human rotavirus vaccine Rotarix™ provides protection against diverse circulating rotavirus strains in African infants: a randomized controlled trial. *BMC Infectious Diseases*, *12*(1), p.213.

Steele, A.D., Peenze, I., De Beer, M.C., Pager, C.T., Yeats, J., Potgieter, N., Ramsaroop, U., Page, N.A., Mitchell, J.O., Geyer, A. and Bos, P., 2003. Anticipating rotavirus vaccines: epidemiology and surveillance of rotavirus in South Africa. *Vaccine*, *21*(5-6), pp.354-360.

Steele, A.D., Van Niekerk, M.C. and Mphahlele, M.J., 1995. Geographic distribution of human rotavirus VP4 genotypes and VP7 serotypes in five South African regions. *Journal of Clinical Microbiology*, *33*(6), pp.1516-1519.

Steele, A.D., Victor, J.C., Carey, M.E., Tate, J.E., Atherly, D.E., Pecenka, C., Diaz, Z., Parashar, U.D. and Kirkwood, C.D., 2019. Experiences with rotavirus vaccines: can we improve rotavirus vaccine impact in developing countries?. *Human Vaccines & Immunotherapeutics*, *15*(6), pp.1215-1227.

Steger, C.L., Brown, M.L., Sullivan, O.M., Boudreaux, C.E., Cohen, C.A., LaConte, L.E. and McDonald, S.M., 2019. In Vitro Double-Stranded RNA Synthesis by Rotavirus Polymerase Mutants with Lesions at Core Shell Contact Sites. *Journal of Virology*, *93*(20), pp.e01049-19.

Steyer, A., Sagadin, M., Kolenc, M. and Poljšak-Prijatelj, M., 2014. Molecular characterization of rotavirus strains from pre-and post-vaccination periods in a country with low vaccination coverage: the case of Slovenia. *Infection, Genetics and Evolution*, *28*, pp.413-425.

Strokach, A., Corbi-Verge, C. and Kim, P.M., 2019. Predicting changes in protein stability caused by mutation using sequence-and structure-based methods in a CAGI5 blind challenge. *Human Mutation*, *40*(9), pp.1414-1423.

Strydom, A., João, E.D., Motanyane, L., Nyaga, M.M., Potgieter, A.C., Cuamba, A., Mandomando, I., Cassocera, M., de Deus, N. and O'Neill, H.G., 2019. Whole genome analyses of DS-1-like Rotavirus A strains detected in children with acute diarrhoea in southern Mozambique suggest several reassortment events. *Infection, Genetics and Evolution*, *69*, pp.68-75.

Sun, X., Li, D., Qi, J., Chai, W., Wang, L., Wang, L., Peng, R., Wang, H., Zhang, Q., Pang, L. and Kong, X., 2018. Glycan binding specificity and mechanism of human and porcine P [6]/P [19] rotavirus VP8* s. *Journal of Virology*, *92*(14).

Syedbasha, M. and Egli, A., 2017. Interferon lambda: modulating immunity in infectious diseases. *Frontiers in Immunology*, *8*, p.119.

Tacharoenmuang, R., Komoto, S., Guntapong, R., Ide, T., Sinchai, P., Upachai, S., Yoshikawa, T., Tharmaphornpilas, P., Sangkitporn, S. and Taniguchi, K., 2016. Full genome characterization of novel DS-1-like G8P [8] rotavirus strains that have emerged in Thailand: reassortment of bovine and human rotavirus gene segments in emerging DS-1-like intergenogroup reassortant strains. *PloS one*, *11*(11).

Tait, S.W. and Green, D.R., 2013. Mitochondrial regulation of cell death. *Cold Spring Harbor Perspectives in Biology*, *5*(9), p.a008706.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), pp.2725-2729.

Tanaka, Y., Yokokawa, R., Rong, H.S., Kishino, H., Stek, J.E., Nelson, M. and Lawrence, J., 2017. Concomitant administration of diphtheria, tetanus, acellular pertussis and inactivated poliovirus vaccine derived from Sabin strains (DTaP-sIPV) with pentavalent rotavirus vaccine in Japanese infants. *Human Vaccines & Immunotherapeutics*, *13*(6), pp.1352-1358.

Taniguchi, K., Hoshino, Y., Nishikawa, K., Green, K.Y., Maloy, W.L., Morita, Y., Urasawa, S., Kapikian, A.Z., Chanock, R.M. and Gorziglia, M., 1988. Cross-reactive and serotype-specific neutralization epitopes on VP7 of human rotavirus: nucleotide sequence analysis of antigenic mutants selected with monoclonal antibodies. *Journal of Virology*, *62*(6), pp.1870-1874.

Tao, Y., Farsetta, D.L., Nibert, M.L. and Harrison, S.C., 2002. RNA synthesis in a cage—structural studies of reovirus polymerase λ3. *Cell*, *111*(5), pp.733-745.

Tate, J.E., Burton, A.H., Boschi-Pinto, C., Parashar, U.D., World Health Organization–Coordinated Global Rotavirus Surveillance Network, Agocs, M., Serhan, F., de Oliveira, L., Mwenda, J.M., Mihigo, R. and Ranjan Wijesinghe, P., 2016. Global, regional, and national estimates of rotavirus mortality in children< 5 years of age, 2000–2013. *Clinical Infectious Diseases*, *62*(suppl_2), pp.S96-S105.

Tatte, V.S. and Chitambar, D.S.D., 2011. Intragenotypic diversity in the VP4 encoding genes of rotavirus strains circulating in adolescent and adult cases of acute gastroenteritis in Pune, Western India: 1993 to 1996 and 2004 to 2007. *Journal of General and Molecular Virology*, *3*(4), pp.53-62.

Taylor, J.A., O'Brien, J.A. and Yeager, M., 1996. The cytoplasmic tail of NSP4, the endoplasmic reticulum-localized non-structural glycoprotein of rotavirus, contains distinct virus binding and coiled coil domains. *The EMBO Journal*, *15*(17), pp.4469-4476.

Thanh, H.D., Lim, I. and 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*, *8*(1), pp.1-10.

Tian, P., Estes, M.K., Hu, Y., Ball, J.M., Zeng, C.Q. and Schilling, W.P., 1995. The rotavirus nonstructural glycoprotein NSP4 mobilizes Ca2+ from the endoplasmic reticulum. *Journal of Virology*, *69*(9), pp.5763-5772.

Todd, S., Page, N.A., Steele, A.D., Peenze, I. and Cunliffe, N.A., 2010. Rotavirus strain types circulating in Africa: review of studies published during 1997–2006. *Journal of Infectious Diseases*, *202*(Supplement_1), pp.S34-S42.

Torres-Vega, M.A., González, R.A., Duarte, M., Poncet, D., López, S. and Arias, C.F., 2000. The C-terminal domain of rotavirus NSP5 is essential for its multimerization, hyperphosphorylation and interaction with NSP6. *Journal of General Virology*, *81*(3), pp.821-830.

Tosser, G., Labbé, M., Bremont, M. and Cohen, J., 1992. Expression of the major capsid protein VP6 of group C rotavirus and synthesis of chimeric single-shelled particles by using recombinant baculoviruses. *Journal of Virology*, *66*(10), pp.5825-5831.

Tregnaghi, M.W., Abate, H.J., Valencia, A., Lopez, P., Da Silveira, T.R., Rivera, L., Medina, D.M.R., Saez-Llorens, X., Ayala, S.E.G., De León, T. and Van Doorn, L.J., 2011. Human rotavirus vaccine is highly efficacious when coadministered with routine expanded program of immunization vaccines including oral poliovirus vaccine in Latin America. *The Pediatric Infectious Disease Journal*, *30*(6), pp.e103-e108.

Troeger, C., Khalil, I.A., Rao, P.C., Cao, S., Blacker, B.F., Ahmed, T., Armah, G., Bines, J.E., Brewer, T.G., Colombara, D.V. and Kang, G., 2018. Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years. *JAMA Pediatrics*, *172*(10), pp.958-965.

Tsugawa, T., Rainwater-Lovett, K. and Tsutsumi, H., 2015. Human G3P [9] rotavirus strains possessing an identical genotype constellation to AU-1 isolated at high prevalence in Brazil, 1997–1999. *The Journal of General Virology*, *96*(3), p.590.

Turnpenny, P.D. and Ellard, S., 2016. *Emery's Elements of Medical Genetics E-Book*. Elsevier Health Sciences.

Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J. and Rousseau, F., 2011. A graphical interface for the FoldX forcefield. *Bioinformatics*, *27*(12), pp.1711-1712.

Velasquez, D.E. and Jiang, B., 2019. Evolution of P [8], P [4], and P [6] VP8* genes of human rotaviruses globally reported during 1974 and 2017: possible implications for rotavirus vaccines in development. *Human Vaccines & Immunotherapeutics*, *15*(12), pp.3003-3008.

Velasquez, D.E., Parashar, U. and Jiang, B., 2018. Decreased performance of live attenuated, oral rotavirus vaccines in low-income settings: causes and contributing factors. *Expert Review of Vaccines*, *17*(2), pp.145-161

Vende, P., Tortorici, M.A., Taraporewala, Z.F. and Patton, J.T., 2003. Rotavirus NSP2 interferes with the core lattice protein VP2 in initiation of minus-strand synthesis. *Virology*, *313*(1), pp.261-273.

Vesikari, T., Karvonen, A.V., Majuri, J., Zeng, S.Q., Pang, X.L., Kohberger, R., Forrest, B.D., Hoshino, Y., Chanock, R.M. and Kapikian, A.Z., 2006. Safety, efficacy, and immunogenicity of 2 doses of bovine-human (UK) and Rhesus–Rhesus-Human rotavirus reassortant tetravalent vaccines in Finnish children. *The Journal of Infectious Diseases*, *194*(3), pp.370-376.

Vesikari, T., Uhari, M., Renko, M., Hemming, M., Salminen, M., Torcel-Pagnon, L., Bricout, H. and Simondon, F., 2013. Impact and effectiveness of RotaTeq® vaccine based on 3 years of surveillance following introduction of a rotavirus immunization program in Finland. *The Pediatric Infectious Disease Journal*, *32*(12), pp.1365-1373.

Viskovska, M., Anish, R., Hu, L., Chow, D.C., Hurwitz, A.M., Brown, N.G., Palzkill, T., Estes, M.K. and Prasad, B.V., 2014. Probing the sites of interactions of rotaviral proteins involved in replication. *Journal of Virology*, *88*(21), pp.12866-12881.

Vizzi, E., Piñeros, O.A., Oropeza, M.D., Naranjo, L., Suárez, J.A., Fernández, R., Zambrano, J.L., Celis, A. and Liprandi, F., 2017. Human rotavirus strains circulating in Venezuela after vaccine introduction: predominance of G2P [4] and reemergence of G1P [8]. *Virology Journal*, *14*(1), p.58.

Vlasova, A.N., Kandasamy, S., Chattha, K.S., Rajashekara, G. and Saif, L.J., 2016. Comparison of probiotic lactobacilli and bifidobacteria effects, immune responses and rotavirus vaccines and infection in different host species. *Veterinary Immunology and Immunopathology*, *172*, pp.72-84.

Ward, M.L., Mijatovic-Rustempasic, S., Roy, S., Rungsrisuriyachai, K., Boom, J.A., Sahni, L.C., Baker, C.J., Rench, M.A., Wikswo, M.E., Payne, D.C. and Parashar, U.D., 2016. Molecular characterization of the first G24P [14] rotavirus strain detected in humans. *Infection, Genetics and Evolution*, *43*, pp.338-342.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L. and Lepore, R., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), pp.W296-W303.

Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V. and 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), pp.773-777.

Webb, B. and Sali, A., 2016. Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, *54*(1), pp.5-6.

World Health Organization, 2013. Rotavirus vaccines: WHO position paper—January 2013. *Weekly Epidemiological Record*, *88*(05), pp.49-64.

World Health Organization, 2020. *Diarrhoeal disease*. Retrieved 16 August 2020, from http://www.who.int/mediacentre/factsheets/fs330/en/

World Health Organization. 2009 Meeting of the immunization Strategic Advisory Group of Experts. *The Weekly Epidemiological Record* 2009, *84*(23), pp.220-236.

Yamamoto, S.P., Kaida, A., Kubo, H. and Iritani, N., 2014. Gastroenteritis outbreaks caused by a DS-1–like G1P [8] rotavirus strain, Japan, 2012–2013. *Emerging Infectious Diseases*, *20*(6), p.1030.

Yoder, J.D., Trask, S.D., Vo, P.T., Binka, M., Feng, N., Harrison, S.C., Greenberg, H.B. and Dormitzer, P.R., 2009. VP5* rearranges when rotavirus uncoats. *Journal of Virology*, *83*(21), pp.11372-11377.

Yodmeeklin, A., Khamrin, P., Kumthip, K., Malasao, R., Ukarapol, N., Ushijima, H. and Maneekarn, N., 2018. Increasing predominance of G8P [8] species A rotaviruses in children admitted to hospital with acute gastroenteritis in Thailand, 2010-2013. *Archives of Virology*, *163*(8), pp.2165-2178.

Yu, G., Lam, T.T.Y., Zhu, H. and Guan, Y., 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Molecular Biology and Evolution*, *35*(12), pp.3041-3043.

Zachos, N.C., Baetz, N.W., Gupta, A., Kapoor, A., Cole, R.N., Verkman, A.S., Turner, J.R., Estes, M.K. and Donowtiz, M., 2019. Human Rotavirus Diarrhea Is Associated with Altered Trafficking and Expression of Apical Membrane Transport Proteins. B*ioRxiv*, p.784264.

Zade, J.K., Kulkarni, P.S., Desai, S.A., Sabale, R.N., Naik, S.P. and Dhere, R.M., 2014. Bovine rotavirus pentavalent vaccine development in India. *Vaccine*, *32*, pp.A124-A128.

Zaman, K., Anh, D.D., Victor, J.C., Shin, S., Yunus, M., Dallas, M.J., Podder, G., Thiem, V.D., Luby, S.P., Coia, M.L. and Lewis, K., 2010. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. *The Lancet*, *376*(9741), pp.615-623.

Zaman, K., Fleming, J.A., Victor, J.C., Yunus, M., Bari, T.I.A., Azim, T., Rahman, M., Mowla, S.M.N., Bellini, W.J., McNeal, M. and Icenogle, J.P., 2016. Noninterference of rotavirus vaccine with measles-rubella vaccine at 9 months of age and improvements in antirotavirus immunity: a randomized trial. *The Journal of Infectious Diseases*, *213*(11), pp.1686-1693.

Zárate, S., Romero, P., Espinosa, R., Arias, C.F. and López, S., 2004. VP7 mediates the interaction of rotaviruses with integrin αvβ3 through a novel integrin-binding site. *Journal of Virology*, *78*(20), pp.10839-10847.

Zeller, M., Heylen, E., Tamim, S., McAllen, J.K., Kirkness, E.F., Akopov, A., De Coster, S., Van Ranst, M. and Matthijnssens, J., 2017. Comparative analysis of the Rotarix™ vaccine strain and G1P [8] rotaviruses detected before and after vaccine introduction in Belgium. *PeerJ*, *5*, p.e2733.

Zeller, M., Nuyts, V., Heylen, E., De Coster, S., Conceição-Neto, N., Van Ranst, M. and Matthijnssens, J., 2016. Emergence of human G2P [4] rotaviruses containing animal derived gene segments in the post-vaccine era. *Scientific Reports*, *6*, p.36841.

Zeller, M., Patton, J.T., Heylen, E., De Coster, S., Ciarlet, M., Van Ranst, M. and Matthijnssens, J., 2012. Genetic analyses reveal differences in the VP7 and VP4 antigenic epitopes between human rotaviruses circulating in Belgium and rotaviruses in Rotarix and RotaTeq. *Journal of Clinical Microbiology*, *50*(3), pp.966-976.

Zeller, M., Rahman, M., Heylen, E., De Coster, S., De Vos, S., Arijs, I., Novo, L., Verstappen, N., Van Ranst, M. and Matthijnssens, J., 2010. Rotavirus incidence and genotype distribution before and after national rotavirus vaccine introduction in Belgium. *Vaccine*, *28*(47), pp.7507-7513.

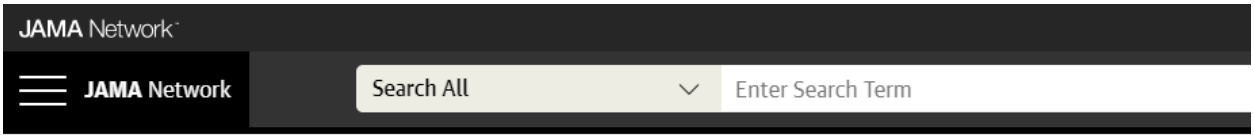
Zhang, B., Chassaing, B., Shi, Z., Uchiyama, R., Zhang, Z., Denning, T.L., Crawford, S.E., Pruijssers, A.J., Iskarpatyoti, J.A., Estes, M.K. and Dermody, T.S., 2014. Prevention and cure of rotavirus infection via TLR5/NLRC4–mediated production of IL-22 and IL-18. *Science*, *346*(6211), pp.861-865.

Zhang, Z., Wang, L., Gao, Y., Zhang, J., Zhenirovskyy, M. and Alexov, E., 2012. Predicting folding free energy changes upon single point mutations. *Bioinformatics*, *28*(5), pp.664-671.

Zhu, S., Ding, S., Wang, P., Wei, Z., Pan, W., Palm, N.W., Yang, Y., Yu, H., Li, H.B., Wang, G. and Lei, X., 2017. Nlrp9b inflammasome restricts rotavirus infection in intestinal epithelial cells. *Nature*, *546*(7660), pp.667-670.

APPENDICES

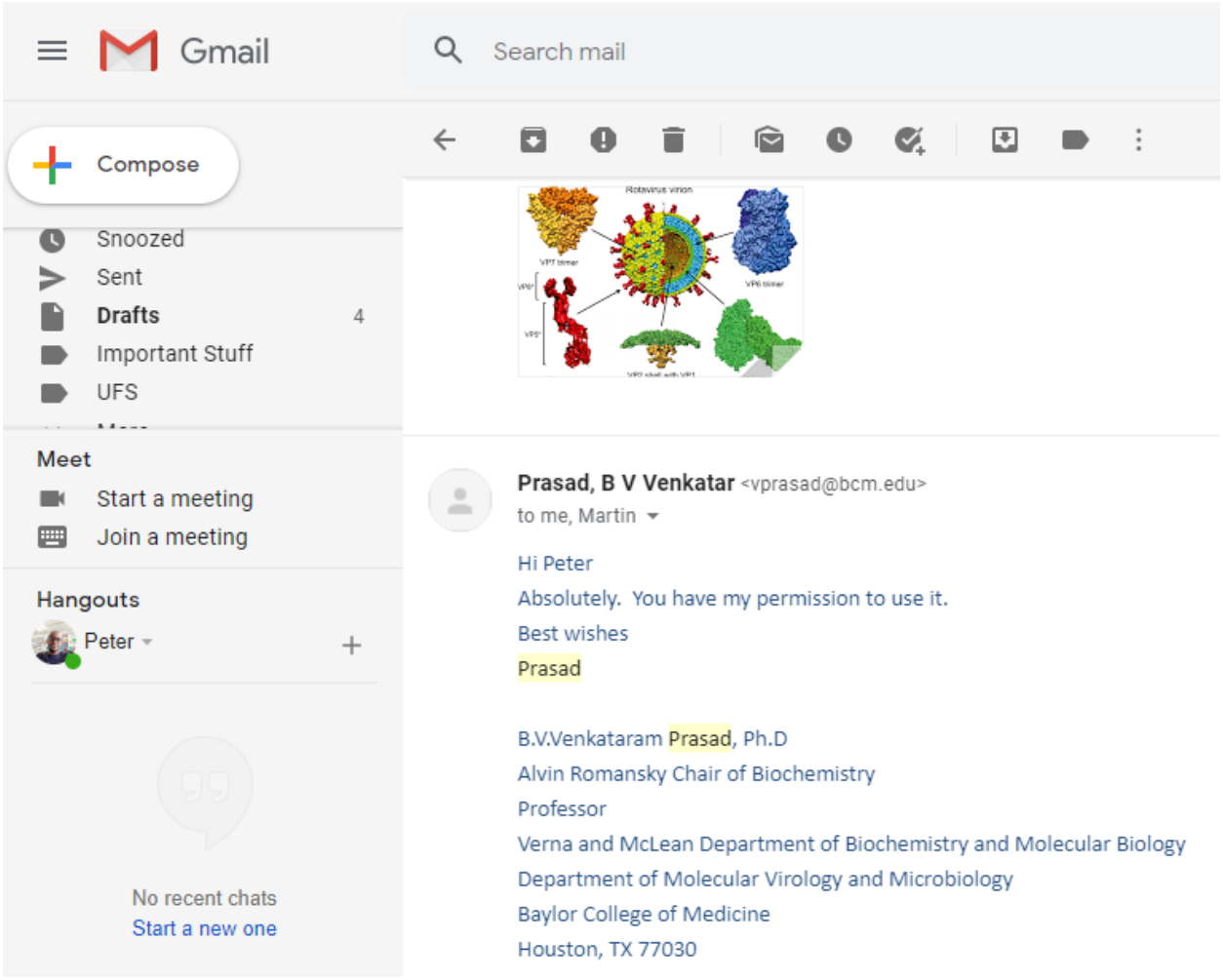
Appendix 1: Permission to use Figure 2.1



CC-BY License Permissions

This is an open access article distributed under the terms of the CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium. You are not required to obtain permission to reuse this article content, provided that you credit the author and journal.

Appendix 2: Permission to use Figure 2.2



Appendix 3: Permission to use Figure 2.3

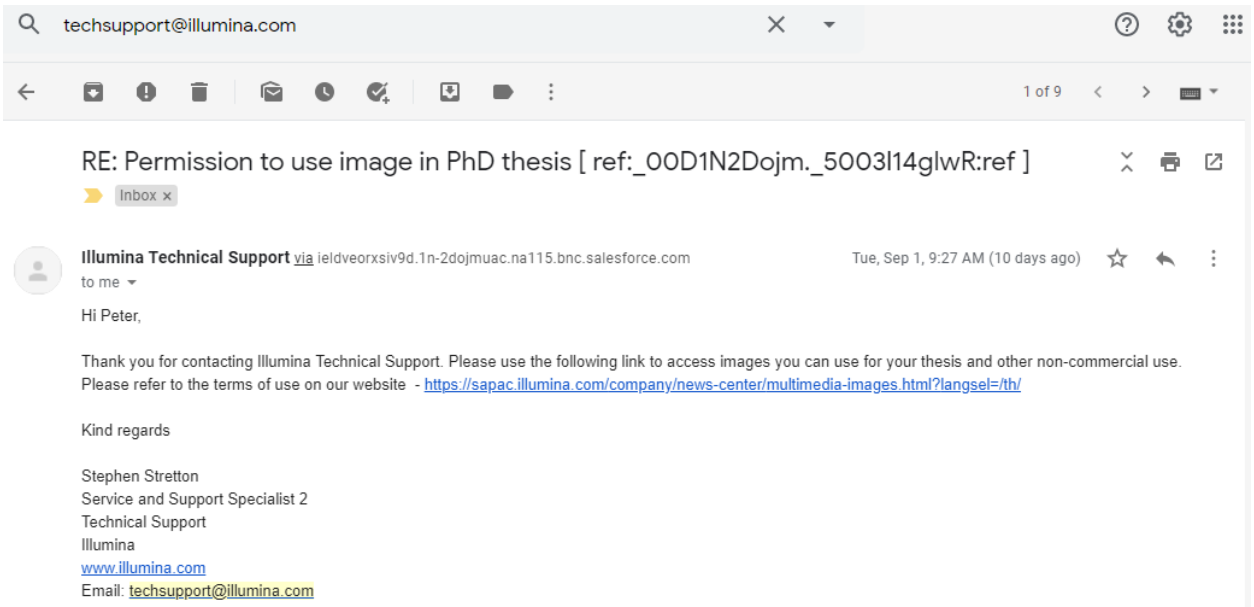
SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 11, 2020

This Agreement between University of the Free State -- Peter Mwangi ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4898230008796
License date	Aug 29, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Rotavirus Biology
Licensed Content Author	Susana López, Carlos F. Arias

Appendix 4: Permission to use figure 2.4



techsupport@illumina.com

1 of 9

RE: Permission to use image in PhD thesis [ref:_00D1N2Dojm._5003l14glwR:ref]

Inbox ×

Illumina Technical Support via ieldveorxsiv9d.1n-2dojmuac.na115.bnc.salesforce.com Tue, Sep 1, 9:27 AM (10 days ago)

to me

Hi Peter,

Thank you for contacting Illumina Technical Support. Please use the following link to access images you can use for your thesis and other non-commercial use. Please refer to the terms of use on our website - https://sapac.illumina.com/company/news-center/multimedia-images.html?langsel=/th/

Kind regards

Stephen Stretton
Service and Support Specialist 2
Technical Support
Illumina
www.illumina.com
Email: techsupport@illumina.com

Appendix 5: Ethics approval letter

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

Health Sciences Research Ethics Committee

16-Aug-2018

Dear **Mr Peter Mwangi**

Ethics Number: UFS-HSD2018/0510/3107

Ethics Clearance: **Molecular characterization of rotavirus strains from pre- and post-vaccine introduction in South Africa**

Principal Investigator: **Mr Peter Mwangi**

Department: **Medical Microbiology (Bloemfontein Campus)**

SUBSEQUENT SUBMISSION APPROVED

With reference to your recent submission for ethical clearance from the Health Sciences Research Ethics Committee. I am pleased to inform you on behalf of the HSREC that you have been granted ethical clearance for your request as stipulated below:

Title Change to Molecular characterization of rotavirus strains from pre- and post-vaccine introduction in South Africa

The HSREC functions in compliance with, but not limited to, the following documents and guidelines: The SA National Health Act. No. 61 of 2003; Ethics in Health Research: Principles, Structures and Processes (2015); SA GCP(2006); Declaration of Helsinki; The Belmont Report; The US Office of Human Research Protections 45 CFR 461 (for non-exempt research with human participants conducted or supported by the US Department of Health and Human Services- (HHS), 21 CFR 50, 21 CFR 56; CIOMS; ICH-GCP-E6 Sections 1-4; The International Conference on Harmonization and Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Tripartite), Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines, Constitution of the HSREC of the Faculty of Health Sciences.

For any questions or concerns, please feel free to contact HSREC Administration: 051-4017794/5 or email EthicsFHS@ufs.ac.za.

Thank you for submitting this request for ethical clearance and we wish you continued success with your research.

Yours Sincerely

Dr. SM Le Grange
Chair : Health Sciences Research Ethics Committee

Health Sciences Research Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 401 7795/7794 | E: ethicsfhs@ufs.ac.za
IRB 00006240; REC 230408-011; IORG0005187; FWA00012784
Block D, Dean's Division, Room D104 | P.O. Box/Posbus 339 (Internal Post Box G40) | Bloemfontein 9300 | South Africa
www.ufs.ac.za

Appendix 6: Whole-genome constellations of 171 South African G1P[8] rotavirus strains

	Gene segment	VP7	VP7 Lineage	VP4	VP4 Lineage	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
	Base pair size for full length sequences	1062		2359		1356	3302	2717	2591	1567	1059	1074	750	664
	Base pair size for open reading frame (ORF) sequences	978		2325		1191	3264	2682	2505	1458	951	930	525	591
	Genotype constellations													
	Sequenced for this study													
	Pre-vaccine strains													
1	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3883/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
2	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3892/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
3	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3951/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
4	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3956/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
5	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4018/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
6	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4019/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
7	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4023/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
8	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4024/2002/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
9	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4026/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
10	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4049/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
11	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4055/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
12	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4102/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
13	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4269/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
14	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4271/2002/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
15	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU456/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
16	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU493/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
17	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU609/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
18	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU632/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
19	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2117/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
20	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2145/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

21	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2156/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
22	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1104/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
23	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1116/2004/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
24	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1131/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
25	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1135/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
26	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1139/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
27	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1145/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
28	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1148/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
29	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1184/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
30	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1192/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
31	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1205/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
32	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1230/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
33	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1287/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
34	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3410/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
35	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3449/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
36	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU451/2005/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
37	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU658/2005/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
38	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU682/2005/G1P[8]**	G1	II	P[8]	IV	I1	R1	C1	M1	A1	N1	T1	E1	H1
39	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU687/2005/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
40	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU745/2005/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
41	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU747/2005/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
42	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU732/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
43	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU749/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
44	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU753/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
45	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU757/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
46	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU768/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
47	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU783/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
48	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU849/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
49	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU902/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
50	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU981/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
51	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU999/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

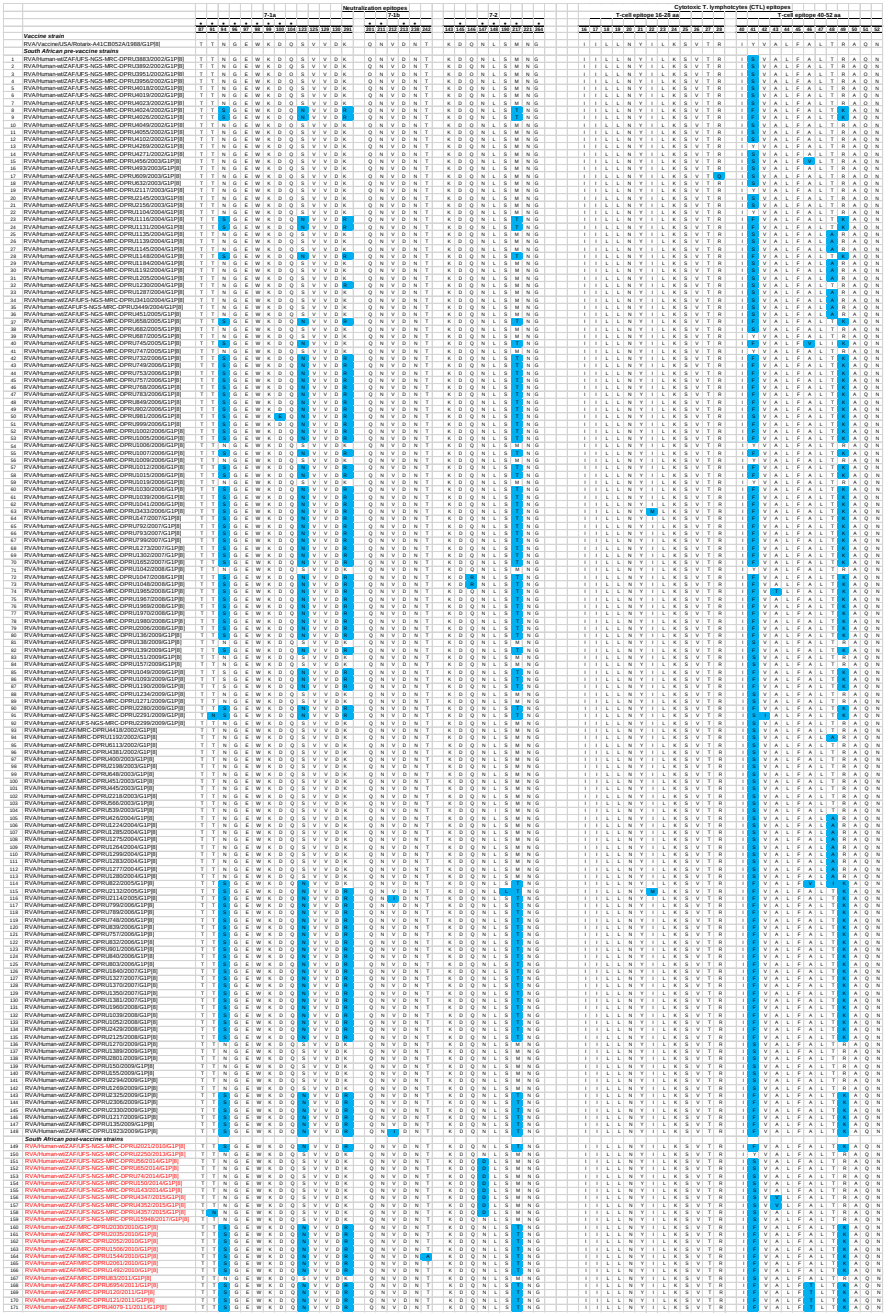
52	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1002/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
53	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1005/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
54	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1006/2006/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
55	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1007/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
56	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1009/2006/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
57	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1012/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
58	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1015/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
59	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1019/2006/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
60	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1030/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
61	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1039/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
62	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1041/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
63	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU3433/2006/G1P[8]**	G1	I	P[8]	IV	I1	R1	C1	M1	A1	N1	T1	E1	H1
64	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU147/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
65	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU792/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
66	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU793/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
67	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU799/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
68	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1273/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
69	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1302/2007/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
70	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1652/2007/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
71	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1042/2008/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
72	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1047/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
73	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1048/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
74	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1965/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
75	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1967/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
76	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1969/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
77	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1970/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
78	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1980/2008/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
79	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2006/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
80	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU136/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
81	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU138/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
82	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU139/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

83	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU151/2009/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
84	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU157/2009/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
85	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1049/2009/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
86	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1093/2009/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
87	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1190/2009/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
88	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1234/2009/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
89	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1271/2009/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
90	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2280/2009/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
91	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2291/2009/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
92	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2299/2009/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
	Post-vaccine strains													
93	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2021/2010/G1P[8]	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
94	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2250/2013/G1P[8]	G1	II	P[8]	I	I1	R1	C1	M1	A1	N1	T1	E1	H1
95	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU56/2014/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
96	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU65/2014/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
97	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU74/2014/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
98	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU143/2014/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
99	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU150/2014/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
100	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4347/2015/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
101	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4352/2015/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
102	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU4357/2015/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
103	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU15948/2017/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
	Acquired from GenBank													
	Pre-vaccine strains													
104	RVA/Human-wt/ZAF/MRC-DPRU4418/2002/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
105	RVA/Human-wt/ZAF/MRC-DPRU1192/2002/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
106	RVA/Human-wt/ZAF/MRC-DPRU6113/2002/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
107	RVA/Human-wt/ZAF/MRC-DPRU4381/2002/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
108	RVA/Human-wt/ZAF/MRC-DPRU400/2003/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
109	RVA/Human-wt/ZAF/MRC-DPRU2198/2003/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
110	RVA/Human-wt/ZAF/MRC-DPRU648/2003/G1P[8]	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

111	**RVA/Human-wt/ZAF/MRC-DPRU451/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
112	**RVA/Human-wt/ZAF/MRC-DPRU445/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
113	**RVA/Human-wt/ZAF/MRC-DPRU2218/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
114	**RVA/Human-wt/ZAF/MRC-DPRU566/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
115	**RVA/Human-wt/ZAF/MRC-DPRU539/2003/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
116	**RVA/Human-wt/ZAF/MRC-DPRU426/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
117	**RVA/Human-wt/ZAF/MRC-DPRU1224/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
118	**RVA/Human-wt/ZAF/MRC-DPRU1285/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
119	**RVA/Human-wt/ZAF/MRC-DPRU1275/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
120	**RVA/Human-wt/ZAF/MRC-DPRU1264/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
121	**RVA/Human-wt/ZAF/MRC-DPRU1299/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
122	**RVA/Human-wt/ZAF/MRC-DPRU1283/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
123	**RVA/Human-wt/ZAF/MRC-DPRU1277/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
124	**RVA/Human-wt/ZAF/MRC-DPRU1280/2004/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
125	**RVA/Human-wt/ZAF/MRC-DPRU822/2005/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
126	**RVA/Human-wt/ZAF/MRC-DPRU2132/2005/G1P[8]**	G1	I	P[8]	IV	I1	R1	C1	M1	A1	N1	T1	E1	H1
127	**RVA/Human-wt/ZAF/MRC-DPRU2114/2005/G1P[8]**	G1	I	P[8]	IV	I1	R1	C1	M1	A1	N1	T1	E1	H1
128	**RVA/Human-wt/ZAF/MRC-DPRU799/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
129	**RVA/Human-wt/ZAF/MRC-DPRU789/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
130	**RVA/Human-wt/ZAF/MRC-DPRU748/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
131	**RVA/Human-wt/ZAF/MRC-DPRU839/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
132	**RVA/Human-wt/ZAF/MRC-DPRU757/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
133	**RVA/Human-wt/ZAF/MRC-DPRU832/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
134	**RVA/Human-wt/ZAF/MRC-DPRU901/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
135	**RVA/Human-wt/ZAF/MRC-DPRU840/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
136	**RVA/Human-wt/ZAF/MRC-DPRU803/2006/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
137	**RVA/Human-wt/ZAF/MRC-DPRU1840/2007/G1P[8]**	G1	I	P[8]	IV	I1	R1	C1	M1	A1	N1	T1	E1	H1
138	**RVA/Human-wt/ZAF/MRC-DPRU1327/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
139	**RVA/Human-wt/ZAF/MRC-DPRU1370/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
140	**RVA/Human-wt/ZAF/MRC-DPRU1350/2007/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
141	**RVA/Human-wt/ZAF/MRC-DPRU1381/2007/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

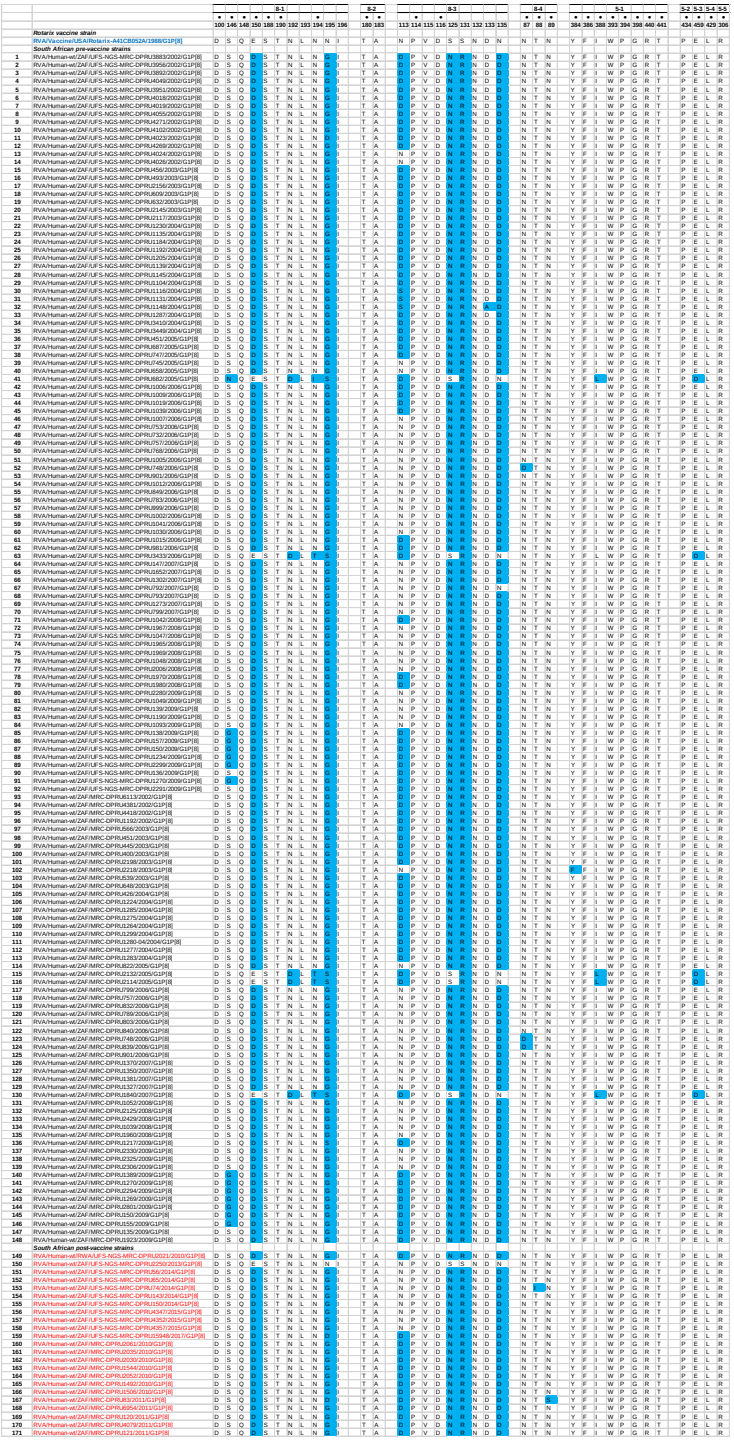
142	**RVA/Human-wt/ZAF/MRC-DPRU1960/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
143	**RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
144	**RVA/Human-wt/ZAF/MRC-DPRU1052/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
145	**RVA/Human-wt/ZAF/MRC-DPRU2429/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
146	**RVA/Human-wt/ZAF/MRC-DPRU2125/2008/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
147	**RVA/Human-wt/ZAF/MRC-DPRU1270/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
148	**RVA/Human-wt/ZAF/MRC-DPRU1389/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
149	**RVA/Human-wt/ZAF/MRC-DPRU2801/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
150	**RVA/Human-wt/ZAF/MRC-DPRU150/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
151	**RVA/Human-wt/ZAF/MRC-DPRU155/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
152	**RVA/Human-wt/ZAF/MRC-DPRU2294/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
153	**RVA/Human-wt/ZAF/MRC-DPRU1269/2009/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
154	**RVA/Human-wt/ZAF/MRC-DPRU2325/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
155	**RVA/Human-wt/ZAF/MRC-DPRU2306/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
156	**RVA/Human-wt/ZAF/MRC-DPRU2330/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
157	**RVA/Human-wt/ZAF/MRC-DPRU1217/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
158	**RVA/Human-wt/ZAF/MRC-DPRU135/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
159	**RVA/Human-wt/ZAF/MRC-DPRU1923/2009/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
	Post-vaccine strains													
160	**RVA/Human-wt/ZAF/MRC-DPRU2030/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
161	**RVA/Human-wt/ZAF/MRC-DPRU2035/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
162	**RVA/Human-wt/ZAF/MRC-DPRU2052/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
163	**RVA/Human-wt/ZAF/MRC-DPRU1506/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
164	**RVA/Human-wt/ZAF/MRC-DPRU1544/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
165	**RVA/Human-wt/ZAF/MRC-DPRU2061/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
166	**RVA/Human-wt/ZAF/MRC-DPRU1492/2010/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
167	**RVA/Human-wt/ZAF/MRC-DPRU83/2011/G1P[8]**	G1	II	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
168	**RVA/Human-wt/ZAF/MRC-DPRU6954/2011/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
169	**RVA/Human-wt/ZAF/MRC-DPRU120/2011/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
170	**RVA/Human-wt/ZAF/MRC-DPRU121/2011/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1
171	**RVA/Human-wt/ZAF/MRC-DPRU4079-11/2011/G1P[8]**	G1	I	P[8]	III	I1	R1	C1	M1	A1	N1	T1	E1	H1

Appendix 7: Neutralization epitopes in G1



Alignment of VP7 neutralizing and cytotoxic T-cell lympocytic epitopes residues of Rotarix® strain against wild type South African G1P[8] strains. Antigenic residues are divided in three epitopes (7-1a, 7-1b, and 7-2). Amino acids that differ with Rotarix® are colored in sky blue. Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. South Africa pre-vaccine G1 genotypes are indicated in boldface while post-vaccine strains are indicated in red.

Appendix 8: Neutralization epitopes in P[8]



Alignment of antigenic residues in VP4 between the P[8] component of Rotarix® and wild type P[8] strains. Antigenic residues are divided into four antigenic epitopes in VP8* (8-1, 8-2, 8-3 and 8-4) and five antigenic epitopes in VP5* (5-1, 5-2, 5-3, 5-4 and 5-5). Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. Amino acids residues colored in sky blue differ from Rotarix®.

Appendix 9: Selection pressure analysis for G1P[8] strains-links to access detailed analysis reports

VP1 https://www.datamonkey.org/fel/5e84f06d89e95e52b24a708f https://www.datamonkey.org/fubar/5f3e4fca503f2f2348163677 http://datamonkey.org/meme/5e84f0e489e95e52b24a71ba	NSP1 https://www.datamonkey.org/fel/5e84f06d89e95e52b24a708f https://www.datamonkey.org/fubar/5e8604ba89e95e52b24ab539 https://www.datamonkey.org/meme/5e8604a389e95e52b24ab4db
VP2 https://www.datamonkey.org/fel/5f3e5860503f2f234816393a https://www.datamonkey.org/fubar/5f3e1581503f2f234816246f https://www.datamonkey.org/meme/5f3e159c503f2f23481624be	NSP2 https://www.datamonkey.org/fel/5e86f36c89e95e52b24add57 https://www.datamonkey.org/fubar/5e86f3a689e95e52b24adde9 https://www.datamonkey.org/meme/5e86f39589e95e52b24adda1
VP3 http://datamonkey.org/fel/5e85b95f89e95e52b24a9756 https://www.datamonkey.org/fubar/5e85b99d89e95e52b24a986d https://www.datamonkey.org/meme/5e85b98389e95e52b24a97d5	NSP3 https://www.datamonkey.org/fel/5e86ffcf89e95e52b24ae13a https://www.datamonkey.org/fubar/5e86fffc89e95e52b24ae1cc https://www.datamonkey.org/meme/5e86ffee89e95e52b24ae186
VP4 https://www.datamonkey.org/fel/5f3e851f503f2f2348164d1f https://www.datamonkey.org/fubar/5f3e84f6503f2f2348164c73 https://www.datamonkey.org/meme/5f3e850d503f2f2348164cc9	NSP4 https://www.datamonkey.org/fel/5e87257489e95e52b24ae584 https://www.datamonkey.org/fubar/5e87258989e95e52b24ae5be https://www.datamonkey.org/meme/5e8725a689e95e52b24ae5fb
VP6 https://www.datamonkey.org/fel/5f3e91df503f2f2348165103 https://www.datamonkey.org/fubar/5f3e91f6503f2f2348165189 https://www.datamonkey.org/meme/5f3e91ef503f2f2348165146	NSP5 https://www.datamonkey.org/fel/5e87747589e95e52b24af0ba https://www.datamonkey.org/fubar/5e87258989e95e52b24ae5be https://www.datamonkey.org/meme/5e8774a289e95e52b24af138
VP7 https://www.datamonkey.org/fel/5e80e11c2a9e58730e9056aa https://www.datamonkey.org/fubar/5e80e62e2a9e58730e905738 https://www.datamonkey.org/meme/5e80d8ce2a9e58730e905521	

Appendix 10: Whole-genome constellations of 103 South African G2P[4] rotavirus strains

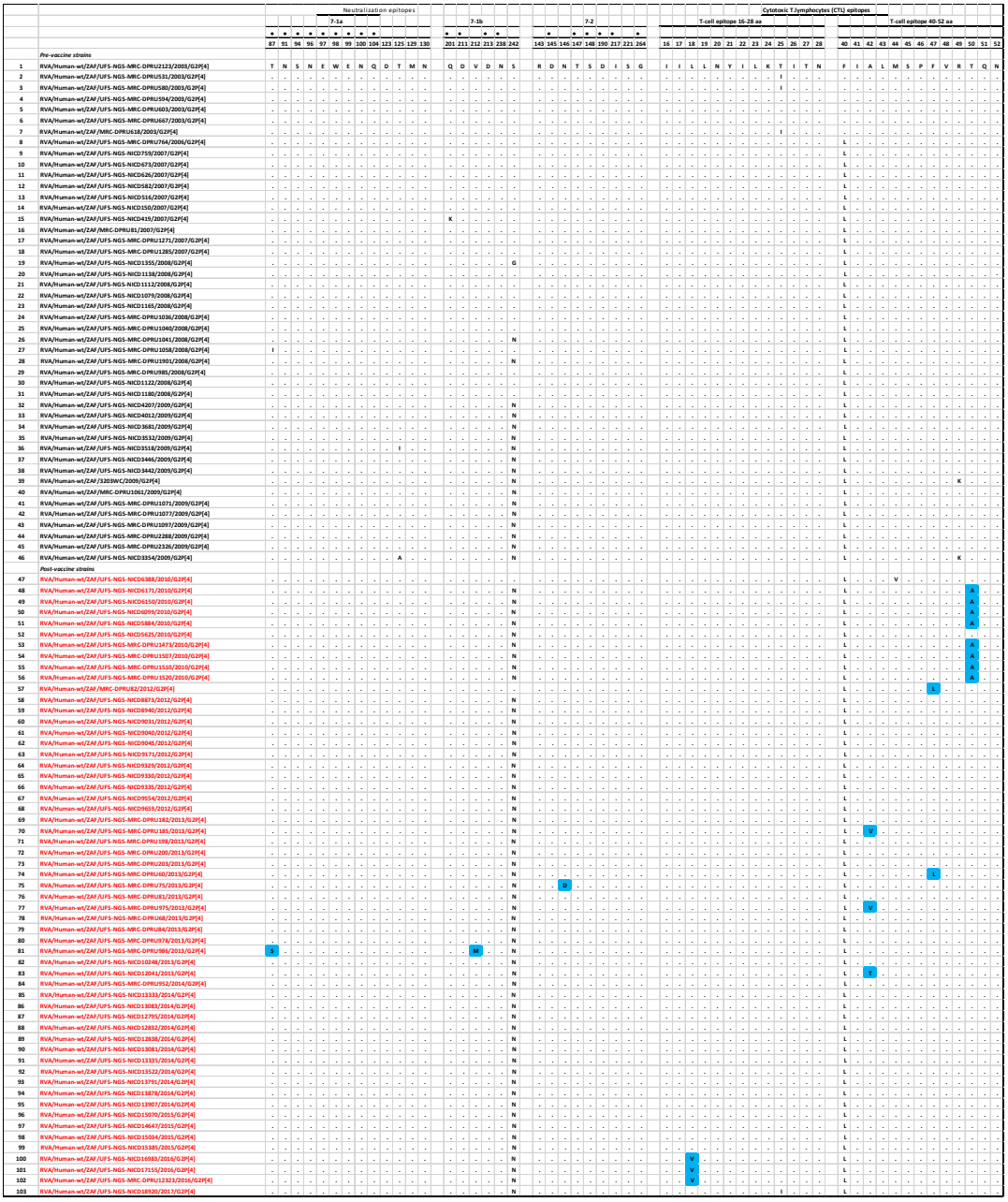
	Gene segment	VP7	VP[4]	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
	Base pair size for full length sequences	1062	2359	1356	3302	2684	2591	1563	1059	1064	751	821
	Base pair size for open reading frame (ORF) sequences	978	2325	1191	3264	2637	2505	1458	951	930	525	600
	Genotype constellations											
	Sequenced for this study											
	Pre-vaccine strains											
1	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
2	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU531/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
3	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU580/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
4	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU594/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
5	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU603/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
6	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU667/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
7	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
8	**RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
9	**RVA/Human-wt/ZAF/UFS-NGS-NICD419/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
10	**RVA/Human-wt/ZAF/UFS-NGS-NICD516/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
11	**RVA/Human-wt/ZAF/UFS-NGS-NICD582/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
12	**RVA/Human-wt/ZAF/UFS-NGS-NICD626/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
13	**RVA/Human-wt/ZAF/UFS-NGS-NICD673/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
14	**RVA/Human-wt/ZAF/UFS-NGS-NICD759/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
15	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1271/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
16	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1285/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
17	**RVA/Human-wt/ZAF/UFS-NGS-NICD1079/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
18	**RVA/Human-wt/ZAF/UFS-NGS-NICD1112/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
19	**RVA/Human-wt/ZAF/UFS-NGS-NICD1122/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
20	**RVA/Human-wt/ZAF/UFS-NGS-NICD1138/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
21	**RVA/Human-wt/ZAF/UFS-NGS-NICD1165/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2

22	**RVA/Human-wt/ZAF/UFS-NGS-NICD1180/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
23	**RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
24	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU985/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
25	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1036/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
26	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1040/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
27	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1041/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
28	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1058/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
29	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1901/2008/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
30	**RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
31	**RVA/Human-wt/ZAF/UFS-NGS-NICD3354/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
32	**RVA/Human-wt/ZAF/UFS-NGS-NICD4207/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
33	**RVA/Human-wt/ZAF/UFS-NGS-NICD4012/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
34	**RVA/Human-wt/ZAF/UFS-NGS-NICD3681/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
35	**RVA/Human-wt/ZAF/UFS-NGS-NICD3518/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
36	**RVA/Human-wt/ZAF/UFS-NGS-NICD3446/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
37	**RVA/Human-wt/ZAF/UFS-NGS-NICD3442/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
38	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1071/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
39	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1077/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
40	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1097/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
41	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2288/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
42	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2326/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
	Post-vaccine strains											
43	**RVA/Human-wt/ZAF/UFS-NGS-NICD6388/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
44	**RVA/Human-wt/ZAF/UFS-NGS-NICD6171/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
45	**RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
46	**RVA/Human-wt/ZAF/UFS-NGS-NICD6099/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
47	**RVA/Human-wt/ZAF/UFS-NGS-NICD5884/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
48	**RVA/Human-wt/ZAF/UFS-NGS-NICD5625/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
49	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1473/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
50	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1507/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
51	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1510/2010/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2

52	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1520/2010/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
53	RVA/Human-wt/ZAF/UFS-NGS-NICD8873/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
54	RVA/Human-wt/ZAF/UFS-NGS-NICD9659/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
55	RVA/Human-wt/ZAF/UFS-NGS-NICD9554/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
56	RVA/Human-wt/ZAF/UFS-NGS-NICD9335/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
57	RVA/Human-wt/ZAF/UFS-NGS-NICD9330/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
58	RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
59	RVA/Human-wt/ZAF/UFS-NGS-NICD9171/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
60	RVA/Human-wt/ZAF/UFS-NGS-NICD9045/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
61	RVA/Human-wt/ZAF/UFS-NGS-NICD9040/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
62	RVA/Human-wt/ZAF/UFS-NGS-NICD9031/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
63	RVA/Human-wt/ZAF/UFS-NGS-NICD8940/2012/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
64	RVA/Human-wt/ZAF/UFS-NGS-NICD12041/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
65	RVA/Human-wt/ZAF/UFS-NGS-NICD10248/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
66	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU60/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
67	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU68/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
68	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU75/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
69	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU81/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
70	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU84/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
71	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU182/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
72	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU185/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
73	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU198/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
74	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU200/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
75	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
76	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU975/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
77	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU978/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
78	RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU986/2013/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
79	RVA/Human-wt/ZAF/UFS-NGS-NICD13907/2014/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
80	RVA/Human-wt/ZAF/UFS-NGS-NICD13878/2014/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
81	RVA/Human-wt/ZAF/UFS-NGS-NICD13791/2014/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
82	RVA/Human-wt/ZAF/UFS-NGS-NICD13522/2014/G2P[4]	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2

83	**RVA/Human-wt/ZAF/UFS-NGS-NICD13335/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
84	**RVA/Human-wt/ZAF/UFS-NGS-NICD13333/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
85	**RVA/Human-wt/ZAF/UFS-NGS-NICD13083/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
86	**RVA/Human-wt/ZAF/UFS-NGS-NICD13081/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
87	**RVA/Human-wt/ZAF/UFS-NGS-NICD12838/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
88	**RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
89	**RVA/Human-wt/ZAF/UFS-NGS-NICD12795/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
90	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU952/2014/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
91	**RVA/Human-wt/ZAF/UFS-NGS-NICD15385/2015/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
92	**RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
93	**RVA/Human-wt/ZAF/UFS-NGS-NICD15034/2015/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
94	**RVA/Human-wt/ZAF/UFS-NGS-NICD14647/2015/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
95	**RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
96	**RVA/Human-wt/ZAF/UFS-NGS-NICD16983/2016/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
97	**RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU12323/2016/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
98	**RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
	Acquired from GenBank											
	Pre-vaccine strains											
99	**RVA/Human-wt/ZAF/MRC-DPRU618/2003/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
100	**RVA/Human-wt/ZAF/MRC-DPRU81/2007/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
101	**RVA/Human-wt/ZAF/MRC-DPRU1061/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
102	**RVA/Human-wt/ZAF/3203WC/2009/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
	Post-vaccine strains											
103	**RVA/Human-wt/ZAF/MRC-DPRU82/2012/G2P[4]**	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2

Appendix 11: Neutralization epitopes in G2



Alignment of VP7 neutralizing and cytotoxic T-cell lympocytic epitopes residues of South African pre-vaccine G2 strains against post-vaccine G2 strains. Antigenic residues are divided in three epitopes (7-1a, 7-1b, and 7-2). Amino acids that differ with pre-vaccine strains are colored in sky blue. Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. South Africa pre-vaccine G2 strains are indicated in boldface while post-vaccine strains are indicated in red.

Appendix 12: Neutralization epitopes in P[4]

	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	459	429	306
Pre-vaccine strains																																					
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2123/2003/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/MRC-DPRU618/2003/G2P[4]	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU531/2003/G2P[4]	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU580/2003/G2P[4]	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU594/2003/G2P[4]	D	S	Q	D	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	I	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU603/2003/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU667/2003/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU764/2006/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD150/2007G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD419/2007/G2P[4]	D	S	Q	D	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	N	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD516/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD582/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD626/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD673/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD759/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/MRC-DPRU81/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1271/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1285/2007/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1079/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1112/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1122/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1138/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1165/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1180/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD1355/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU985/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1036/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1040/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1041/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1058/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1901/2008/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3532/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3354/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD4207/2009/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD4012/2009/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3681/2009/G2P[4] VP[4]	D	S	Q	D	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	R	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3518/2009/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3446/2009/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD3442/2009/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/MRC-DPRU1061/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/3203WC/2009/G2P[4]	D	N	Q	D	S	T	D	L	N	N	I	T	A	N	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1071/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1077/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1097/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2288/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU2326/2009/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
Post-vaccine strains																																					
RVA/Human-wt/ZAF/UFS-NGS-NICD6388/2010/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD6171/2010/G2P[4] R2	D	S	Q	N	S	T	D	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD6150/2010/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD6099/2010/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD5884/2010/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD5625/2010/G2P[4] VP[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1473/2010/G2P[4]	D	S	Q	D	S	T	Y	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1507/2010/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1510/2010	D	S	Q	D	S	T	Y	L	N	N	I	T	A	S	Q	T	N	N	E	N	S	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1520/2010	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD8873/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9659/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9554/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9335/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9330/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9329/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9171/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9045/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9040/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD9031/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD8940/2012/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/MRC-DPRU82/2012/G2P[4]	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	I	N	N	E	N	N	D	N	T	D	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD12041/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD10248/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU60/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU68/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU75/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU81/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU84/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU182/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU185/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU198/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU200/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU203/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU975/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU979/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU986/2013/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13907/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13878/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13791/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13522/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13335/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13333/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13083/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD13081/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD12838/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD12832/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD12795/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU952/2014/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD15385/2015/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD15070/2015/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD15034/2015/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD14647/2015/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD17155/2016/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD16983/2016/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU12323/2016/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R
RVA/Human-wt/ZAF/UFS-NGS-NICD18920/2017/G2P[4]	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	Y	F	L	W	P	G	R	T	P	E	L	R

Alignment of antigenic residues in VP4 between the P[4] component of South African pre-vaccine and post-vaccine type P[4] strains. Antigenic residues are divided into four antigenic epitopes in VP8* (8-1, 8-2, 8-3 and 8-4) and five antigenic epitopes in VP5* (5-1, 5-2, 5-3, 5-4 and 5-5). Amino acid changes that have been shown to escape neutralization with monoclonal antibodies are indicated with a black dot. Amino acids residues colored in sky blue differ from pre-vaccine strains.

Appendix 13: Selection pressure analysis for G2P[4] strains-links to access detailed analysis reports

VP1 https://www.datamonkey.org/fel/5f5134e5611379489329119a https://www.datamonkey.org/fubar/5f513505611379489329119a https://www.datamonkey.org/meme/5f513513611379489329124e	NSP1 https://www.datamonkey.org/fel/5f5fbf490da7ba237c463351 https://www.datamonkey.org/fubar/5f5fbf6f0da7ba237c4633a8 https://www.datamonkey.org/meme/5f5fbf810da7ba237c4633ff
VP2 https://www.datamonkey.org/fel/5f52043b6113794893292d37 https://www.datamonkey.org/fubar/5f520a7c6113794893292ed8 https://www.datamonkey.org/meme/5f5204f46113794893292ded	NSP2 https://www.datamonkey.org/fel/5f6127f50da7ba237c466882 https://www.datamonkey.org/fubar/5f6127dc0da7ba237c466816 https://www.datamonkey.org/meme/5f6127ed0da7ba237c46684c
VP3 https://www.datamonkey.org/fel/5f55dadddc5d1378a560339d https://www.datamonkey.org/fubar/5f55db15dc5d1378a56033fd https://www.datamonkey.org/meme/5f55db24dc5d1378a560345b	NSP3 www.datamonkey.org/fel/5f612e640da7ba237c466910 https://www.datamonkey.org/fubar/5f612e7a0da7ba237c46694c https://www.datamonkey.org/meme/5f612e860da7ba237c466986
VP4 https://www.datamonkey.org/fel/5f4cc6c734c53e423ae77349 https://www.datamonkey.org/fubar/5f4cc6e034c53e423ae773a2 https://www.datamonkey.org/meme/5f4cc6ec34c53e423ae773f9	NSP4 https://www.datamonkey.org/fel/5f587685dc5d1378a560a09f https://www.datamonkey.org/fubar/5f5877aadc5d1378a560a0e2 https://www.datamonkey.org/meme/5f5877bcdc5d1378a560a127
VP6 https://www.datamonkey.org/fel/5f56351fdc5d1378a56042ad https://www.datamonkey.org/fubar/5f56354fdc5d1378a56042f1 https://www.datamonkey.org/meme/5f56355adc5d1378a5604332	NSP5 https://www.datamonkey.org/fel/5f6118d60da7ba237c46673a https://www.datamonkey.org/fubar/5f6118ea0da7ba237c46676d https://www.datamonkey.org/meme/5f6118f50da7ba237c46679e
VP7 https://www.datamonkey.org/fel/5f464b2434c53e423ae6a172 https://www.datamonkey.org/fubar/5f464b5834c53e423ae6a1b8 https://www.datamonkey.org/meme/5f464b6434c53e423ae6a1f7	

Appendix 14: Abstract of research article I

pathogens

MDPI

Article

Uncovering the First Atypical DS-1-like G1P[8] Rotavirus Strains That Circulated during Pre-Rotavirus Vaccine Introduction Era in South Africa

Peter N. Mwangi [1], Milton T. Mogotsi [1], Sebotsana P. Rasebotsa [1], Mapaseka L. Seheri [2], M. Jeffrey Mphahlele [3], Valantine N. Ndze [4], Francis E. Dennis [5], Khuzwayo C. Jere [6,7] and Martin M. Nyaga [1,*]

1 Next Generation Sequencing Unit, Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein 9300, South Africa; 2017219839@ufs4life.ac.za (P.N.M.); tmogotsi16@gmail.com (M.T.M.); RasebotsaS@ufs.ac.za (S.P.R.)
2 Diarrhoeal Pathogens Research Unit, Sefako Makgatho Health Sciences University, Medunsa 0204, Pretoria, South Africa; mapaseka.seheri@smu.ac.za
3 South African Medical Research Council, 1 Soutpansberg Road, Pretoria 0001, South Africa; Jeffrey.Mphahlele@mrc.ac.za
4 Faculty of Health Sciences, University of Buea, P.O. Box 63, Buea, Cameroon; valentinengum@yahoo.com
5 Noguchi Memorial Institute for Medical Research, University of Ghana, P.O. Box LG581, Legon, Ghana; FDennis@noguchi.ug.edu.gh
6 Centre for Global Vaccine Research, Institute of Infection and Global Health, University of Liverpool, Ronald Ross Building, 8 West Derby Street, Liverpool L69 7BE, UK; Khuzwayo.Jere@liverpool.ac.uk
7 Malawi-Liverpool-Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, Blantyre 312225, Malawi
* Correspondence: NyagaMM@ufs.ac.za; Tel.: +27-51-401-9158

Received: 13 April 2020; Accepted: 18 May 2020; Published: 20 May 2020

check for updates

Abstract: Emergence of DS-1-like G1P[8] group A rotavirus (RVA) strains during post-rotavirus vaccination period has recently been reported in several countries. This study demonstrates, for the first time, rare atypical DS-1-like G1P[8] RVA strains that circulated in 2008 during pre-vaccine era in South Africa. Rotavirus positive samples were subjected to whole-genome sequencing. Two G1P[8] strains (RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1971/2008/G1P[8] and RVA/Human-wt/ZAF/UFS-NGS-MRC-DPRU1973/2008/G1P[8]) possessed a DS-1-like genome constellation background (I2-R2-C2-M2-A2-N2-T2-E2-H2). The outer VP4 and VP7 capsid genes of the two South African G1P[8] strains had the highest nucleotide (amino acid) nt (aa) identities of 99.6–99.9% (99.1–100%) with the VP4 and the VP7 genes of a locally circulating South African strain, RVA/Human-wt/ZAF/MRC-DPRU1039/2008/G1P[8]. All the internal backbone genes (VP1–VP3, VP6, and NSP1-NSP5) had the highest nt (aa) identities with cognate internal genes of another locally circulating South African strain, RVA/Human-wt/ZAF/MRC-DPRU2344/2008/G2P[6]. The two study strains emerged through reassortment mechanism involving locally circulating South African strains, as they were distinctly unrelated to other reported atypical G1P[8] strains. The identification of these G1P[8] double-gene reassortants during the pre-vaccination period strongly supports natural RVA evolutionary mechanisms of the RVA genome. There is a need to maintain long-term whole-genome surveillance to monitor such atypical strains.

Keywords: atypical strains; genome constellation; reassortment; rotavirus; whole-genome characterization

Appendix 15: Abstract of research article II

vaccines

MDPI

Article

Whole Genome In-Silico Analysis of South African G1P[8] Rotavirus Strains before and after Vaccine Introduction over a Period of 14 Years

Peter N. Mwangi [1], Milton T. Mogotsi [1], Mapaseka L. Seheri [2], M. Jeffrey Mphahlele [2,3], Ina Peenze [2], Mathew D. Esona [2], Benjamin Kumwenda [4], A. Duncan Steele [5], Carl D. Kirkwood [5], Valantine N. Ndze [6], Francis E. Dennis [7], Khuzwayo C. Jere [8,9,10] and Martin M. Nyaga [1,*]

1 Next Generation Sequencing Unit and Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein 9300, South Africa; nthigapete@gmail.com (P.N.M.); tmogotsi16@gmail.com (M.T.M.)
2 Diarrheal Pathogens Research Unit, Sefako Makgatho Health Sciences University, Medunsa 0204, South Africa; mapaseka.seheri@smu.ac.za (M.L.S.); Jeffrey.Mphahlele@mrc.ac.za (M.J.M.); ina.peenze@smu.ac.za (I.P.); mathew.esona@gmail.com (M.D.E.)
3 South African Medical Research Council, Pretoria 0001, South Africa
4 College of Medicine, Department of Biomedical Sciences, Faculty of Biomedical Sciences and Health Professions, University of Malawi, Private Bag 360, Chichiri, Blantyre 3, Malawi; bkumwenda@medcol.mw
5 Enteric and Diarrheal Diseases, Global Health, Bill & Melinda Gates Foundation, P.O. Box 23350, Seattle, WA 98109, USA; Duncan.Steele@gatesfoundation.org (A.D.S.); Carl.Kirkwood@gatesfoundation.org (C.D.K.)
6 Faculty of Health Sciences, University of Buea, P.O. Box 63, Buea, Cameroon; valentinengum@yahoo.com
7 Noguchi Memorial Institute for Medical Research, University of Ghana, P.O. Box LG581, Legon, Ghana; fdennis@noguchi.ug.edu.gh
8 Center for Global Vaccine Research, Institute of Infection, Liverpool L697BE, UK; Khuzwayo.Jere@liverpool.ac.uk
9 Veterinary and Ecological Sciences, University of Liverpool, Liverpool L697BE, UK
10 Malawi-Liverpool-Wellcome Trust Clinical Research Program, Department of Medical Laboratory Sciences, College of Medicine, University of Malawi, Blantyre 312225, Malawi
* Correspondence: NyagaMM@ufs.ac.za; Tel.: +27-51-401-9158

Received: 7 September 2020; Accepted: 13 October 2020; Published: 14 October 2020

check for updates

Abstract: Rotavirus G1P[8] strains account for more than half of the group A rotavirus (RVA) infections in children under five years of age, globally. A total of 103 stool samples previously characterized as G1P[8] and collected seven years before and seven years after introducing the Rotarix® vaccine in South Africa were processed for whole-genome sequencing. All the strains analyzed had a Wa-like constellation (G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1). South African pre- and post-vaccine G1 strains were clustered in G1 lineage-I and II while the majority (84.2%) of the P[8] strains were grouped in P[8] lineage-III. Several amino acid sites across ten gene segments with the exception of VP7 were under positive selective pressure. Except for the N147D substitution in the antigenic site of eight post-vaccine G1 strains when compared to both Rotarix® and pre-vaccine strains, most of the amino acid substitutions in the antigenic regions of post-vaccine G1P[8] strains were already present during the pre-vaccine period. Therefore, Rotarix® did not appear to have an impact on the amino acid differences in the antigenic regions of South African post-vaccine G1P[8] strains. However, continued whole-genome surveillance of RVA strains to decipher genetic changes in the post-vaccine period remains imperative.

Keywords: evolution; rotavirus strains; Wa-like constellation; whole-genome; lineages