How to cite this thesis / dissertation (APA referencing method):

Surname, Initial(s). (Date). *Title of doctoral thesis* (Doctoral thesis). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Surname, Initial(s). (Date). *Title of master's dissertation* (Master's dissertation). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Genotyping and whole genome classification of group A rotaviruses originating from an urban and rural site in Mozambique

By

Lithabiso Motanyane

Dissertation submitted in fulfilment of the requirements for the degree

Magister Scientiae

in the

Department of Microbial, Biochemical and Food Biotechnology

Faculty of Natural and Agricultural Sciences

University of the Free State

Bloemfontein

South Africa

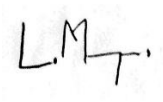
Supervisor: Dr. Hester G. O'Neill

Co-supervisor: Prof. A. Christiaan Potgieter

December 2017

Declaration

"I, Lithabiso Motanyane, declare that the dissertation hereby submitted by me, for the Magister Scientiae degree at the University of the Free State, is my own independent work and has not previously been submitted at another University/Faculty. I furthermore cede copyright of the dissertation in favour of the University of the Free State".



Signature: _____

Lithabiso Motanyane

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

South Africa

Acknowledgements

I would like to thank the following people and organizations who made it possible for me to reach this stage. Truly no man is an island and without their varied contributions, today would have told a different story about me.

My supervisor**, Dr. Hester G. O'Neill,** for providing me with the opportunity to study towards my MSc within her research group. Thank you for all the guidance and patience you portrayed throughout. May the Lord truly bless you, you are an amazing supervisor.

My co-supervisor, **Prof. Christiaan A. Potgieter**, for his assistance and especially for showing me how to analyse NGS data.

Dr. Nilsa de Deus and **Eva Dora Joao** our Mozambican collaborators.

Silke Arndt, for the next generation sequencing services provided at Inqaba Biotec.

Dr. Hamilton Ganesan, for providing me with a CLCBio activation key so I can use it for my data analysis.

My **colleagues** in the Clinical Biochemistry and Molecular Biology laboratories.

My **family,** for their unwavering support and prayers, a special thank you to my mum and dad, I will always love you. My friends, especially **Sello Kekana**, for always believing that I will finally reach the end of my MSc journey even when I did not feel like it will come to an end.

Lesotho Government, **Poliomyelitis Research Foundation** (PRF) and the **Department of Microbial, Biochemical and Food Biotechnology** for financial support.

Table of contents

LIST OF FIGURES

Chapter 3

LIST OF TABLES

Chapter 1 **Page no.**

Chapter 2

Chapter 3

Conference contributions

Motanyane, L, De Deus, N, Potgieter A.C., O'Neill, H.G. Molecular characterization of Mozambican group A rotavirus field strains. *Virology Africa 2015,* Cape Town, 1-3 December 2015. (poster)

O'Neill, H.G., Motanyane, L, João, E.D., De Deus N. Phylogenetic analysis of VP4 and VP7 coding sequences of Mozambican rotavirus strains. *6th European Rotavirus Biology Meeting*, Dijon, France, 17-20 May 2015. (poster).

Motanyane, L., Strydom, A., Potgieter, C., João, E.D., De Deus, N. & O'Neill, H.G. Determination of the whole genome constellations for Mozambican type A rotavirus strains using next generation sequencing. 12th International Rotavirus Symposium, Melbourne, Australia. 7-9 September 2016. (poster)

Strydom, A., Motanyane, L., Nyaga, M.M., João, E.D., De Deus, N. & O'Neill, H.G. Whole genome constellations of Rotavirus A detected in southern Mozambique prior to the introduction of vaccination. *11th African Rotavirus Symposium 2017*, Lilongwe, Malawi. 28-30 May 2017. (poster)

Publication

João, E.D., Strydom, A., O'Neill, H.G., Cuamba, A., Cassocera, M., Acácio, S., Mandomando, I., Motanyane, L., Page, N., De Deus, N. Rotavirus A strains obtained from children with acute gastroenteritis in Mozambique, 2012-2013: G and P genotypes and phylogenetic analysis of VP7 and partial VP4 genes. ***Archives of Virology*** (Scientifically accepted)

Chapter 1

Rotavirus classification and molecular epidemiology with specific emphasis on Sub-Saharan Africa

1.1 Introduction

Rotaviruses affect a wide variety of animal species including humans, non-human primates (such as macaques), livestock (such as bovine and porcine), household pets (such as dogs and cats), game (such as antelope and giraffes), rodents, as well as avian species (Schumann et al., 2009).

They belong to the family of double-stranded RNA (dsRNA) viruses. Double-stranded RNA viruses are divided into seven different families, namely- *Hypoviridae, Totiviridae, Birnaviridae, Partitiviridae, Cystoviridae, Chrysoviridae, and Reoviridae*. Double-stranded RNA viruses are made up of segmented genomes (Van Regenmortel and Fauquet, 2000). Rotaviruses belong to the genus *Rotavirus* within the *Reoviridae* family, which is the largest family of all the dsRNA virus families (Mertens et al., 2005). The *Reoviridae* family includes 11 other genera namely- *Orthoreovirus, Orbivirus, Coltivirus, Aquareovirus, Cypovirus, Fijivirus, Phytoreovirus, Oryzavirus, Seadornavirus, Idnoreovirus, and Mycoreovirus* (Mertens et al., 2005). The genus, rotavirus, is divided to eight groups, namely A to H (Matthijnssens et al., 2011). However, this review and the rest of this dissertation will only focus on group A rotaviruses.

Viruses in the *Reoviridae* family have a wide host range and despite being in the same family, differ in their pathogenicity. The name rotavirus is derived from the Latin word “rota” which means wheel, this pertains to the structure of these viruses which resemble a wheel as shown in Figure 1.1 (Flewett et al., 1974). Members of the *Reoviridae* family are non-enveloped, have icosahedral capsids, and their genomes are made up of segmented dsRNA molecules. Initially, negative-stain electron microscopy studies indicated that rotaviruses have an average diameter of 70 nm while later studies using cryo-electron microscopy indicated that they have a much larger diameter of 100 nm (Estes and Greenberg, 2013). Before the name rotavirus was accepted, several names, such as reovirus-like, orbivirus-like, duovirus-like, infantile

gastroenteritis virus and “new” virus, were used to refer to these viruses. The name rotavirus was only accepted and used after their discovery (Bishop, 2009).

Human rotaviruses were first linked to diarrhoea by Ruth Bishop and colleagues in 1973 after electron micrographs of sections of the duodenal mucosa, taken from children with diarrhoea, were investigated. These particles were later observed in stool specimens of children with severe diarrhoea (Bishop et al., 1974). Animal rotaviruses were discovered before human rotaviruses but were only identified as rotaviruses after the discovery of human rotaviruses due to their similar characteristics. For example, the Simian Agent 11 (SA11) was isolated in 1958 (Bishop, 2009).

Rotavirus transmission occurs through the faecal-oral route. It causes gastroenteritis which is characterised by vomiting and diarrhoea, amongst other symptoms in humans. It is mostly children under the age of 5 years world-wide that are adversely affected by group A rotaviruses (Kotloff et al., 2013). Rotaviruses negatively affect human health, therefore research on rotavirus infection mechanisms and treatment has been widely conducted. It has been reported that each year about 215 000 deaths occur world-wide due to rotavirus infections (Tate et al., 2016). Most of these deaths occur in developing countries in sub-Saharan Africa and Asia (Parashar et al., 2009). In developing countries, deaths due to rotavirus infections are more common because of limited access to medical care. In addition, the use of vaccines may be limited due to financial constraints. Poor nutrition in some of these countries also leads to worsening of rotavirus symptoms (Elliott, 2007).

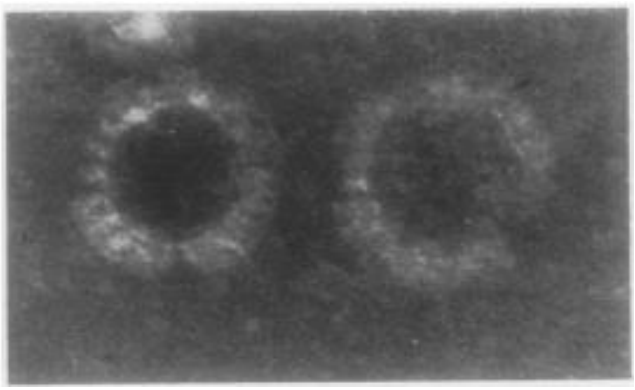


Figure 1.1 Negative stain electron micrograph of rotavirus particles showing that their morphology represents a wheel (Adapted from Flewett et al., 1974)**.**

1.2 Genome Organization

Rotavirus genomes consist of 11 dsRNA segments and the sizes of these segments differ. For the prototype virus, Simian agent 11 (SA11), they range in size between 667 bp (segment 11) to 3302 bp (segment 1). On average, rotavirus genomes are 18 522 bp in size (Estes and Greenberg, 2013). Figure 1.2 is a representation of the genome organisation of all 11 rotavirus genome segments. Each genome segment starts with a 5' guanidine residue followed by a 5' non-coding region that is made up of a set of conserved sequences. The 5' non-coding region is in turn followed by an open-reading frame (ORF). Most genome segments contain only one ORF. However, genome segment 7, 9 and 10 have an additional in-phase ORF, while genome segment 11 has an additional out-of-phase ORF (Estes and Greenberg, 2013). The end of the ORF is designated by a stop codon, marking the beginning of the 3' termini. The 3' end begins with a non-coding region that also contains a set of conserved sequences. Both the 5' and 3' non-coding regions have varying lengths for different genome segments. The 3' end of rotavirus genome segments end with a conserved sequence, which in most cases is UGUGACC. This sequence, together with the conserved sequence at the 5' end, has signals that aid the processes of transcription, translation, as well as genome replication and the packaging of the genome segments into the viral capsid. Unlike other eukaryotes and some viruses, the rotavirus 3' end lacks a poly-A tail (Estes and Greenberg, 2013).

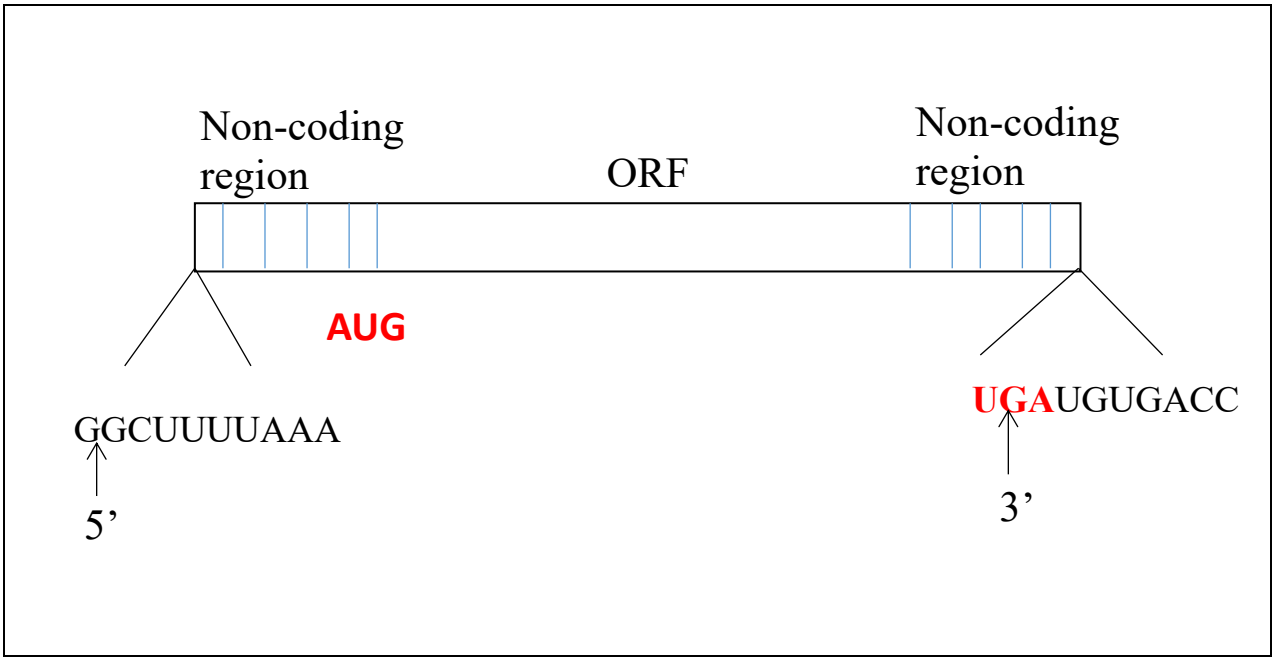


Figure 1.2: A schematic representation of the general genome organisation of rotavirus genome segments. Each genome segment begins with a guanidine residue at the 5' end non-coding region followed by an ORF that is located between the non-coding regions at the 5'end and 3'end. The 3'end of the genome segment ends with a conserved sequence UGUGACC (Adapted from Estes and Greenberg, 2013).

1.3 Rotavirus proteins

The 11 genome segments of rotavirus code for 12 proteins. Six of these proteins are structural and six are non-structural. The structural proteins are named viral protein (VP) 1, 2, 3, 4, 6 and 7, while the non-structural proteins (NSP) are named NSP1, NSP2, NSP3, NSP4, NSP5 and NSP6. Each of the 11 genome segments code for one protein with the exception of genome segment 11 which codes for two non-structural proteins namely NSP5 and NSP6. However, NSP6 is not encoded by all rotavirus strains (Mattion et al, 1991). The different sizes and functions of the rotavirus proteins are summarised in Table 1.1. The structural proteins are arranged into three layers that make up the icosahedral capsid of the virus. The non- structural proteins are only produced during viral infections, and play a role in virus replication. The inner capsid layer encases the viral genome and is made up of 120 copies of VP2 with 11 or 12 copies of both VP1 and VP3 complexes located at the five-fold vertices. The middle layer is made up

of 780 copies of VP6, and the outer layer consists of VP7 and VP4 (Estes and Greenberg, 2013; Yoder and Dormitzer, 2006). VP4 is attached to VP6 and passes through VP7.

Rotavirus particles can exist as either singled-layered particles (SLP), double-layered particles (DLP), or triple-layered particles (TLP). Complete rotavirus particles are referred to as TLPs and they have all three protein layers. DLPs lack the outer protein layer made up of VP4 and VP7 and these particles are not capable of infecting host cells (Teimoori et al., 2014). Single layered particles lack the two outer protein layers (VP6, VP7 and VP4).

Table 1.1: Proteins encoded by the 11 genome segments and their functions, using the prototype strain, SA11, as reference

Genome segment #	Translation product and (no. of copies in capsid structure)	Protein size (kDa)	Function	Reference
1	VP1 (11-12)	125	Functions as RNA-dependant RNA polymerase for transcribing and replicating the viral genome.	(Valenzuela et al., 1991)
2	VP2 (120)	102	Binds to the RNA through N-terminal and is essential for maintaining optimum space in between RNA segments.	(Jayaram et al, 2004) (Bican et al., 1982)
3	VP3 (11-12)	88	Functions as a 5' end of guanylyl methyl transferase and plays a role in capping the mRNA transcripts during transcription.	(Liu et al., 1992)
4	VP4 (120)	88	Plays a role in the attachment of the virus to the host cell by binding to host cell receptors, plays a role in cell entry processes.	(Prasad et al., 1990)
5	NSP1	58	Function not yet characterised, studies suggest that it counteracts the host's innate immune response by preventing apoptosis	(Arnold and Patton, 2011)
6	VP6 (780)	45	Required for double layered particle to become transcriptionally active, used for classification of rotaviruses into different groups and subgroups.	(Prasad et al., 1988)
7	NSP3	36	Binds to the viral mRNA at the 3' end consensus sequence, allowing delivery to host ribosomes and mRNA translation, and halts host cell mRNA translation.	(Jayaram et al ., 2004)
8	NSP2	37	Works in conjunction with NSP5 and VP1 during viral genome replication and packaging, plays a role in viroplasm formation, has NTPase, ssRNA binding, and helix destabilising activities.	(Vende et al., 2003) (Taraporewala et al., 1999)

Table 1.1: Proteins encoded by the 11 genome segments and their functions, using the prototype strain, SA11, as reference (continue)

Genome segment #	Translation product and (no. of copies in capsid structure)	Protein size (kDa)	Function	Reference
9	VP7	38	Plays a role in virus attachment to the host cell and entry by regulating the function of VP4 during these processes.	(Hasenack et al., 2002)
10	NSP4	20	Involved in the budding of DLPs from viroplasms into the ER after viral genome replication, involved in encapsidation. Implicated as an enterotoxin, resulting in diarrhoea by hindering the ability of the intestines to absorb fluid.	(Van Doorn et al., 2009) (Ericson et al., 1983) (Hyser et al., 2010)
11	NSP5	22	Works in conjunction with NSP2 and VP1 to replicate and package the viral genome, regulates the ability of NSP2 to bind to RNA, and plays a role in the formation of viroplasms.	(Afrikanova et al., 1998)
11	NSP6	12	It is a major component of the viroplasm, and works in conjunction with NSP5 during replication.	(González et al., 1998)

1.4 Rotavirus life-cycle

Rotavirus replication occurs in the cytoplasm of mature epithelium cells that line the small intestines. A schematic representation of the replication cycle is shown in Figure 1.3. For a rotavirus to successfully infect the host cell, it first has to attach to the cell receptors in order to penetrate the membrane and enter the cytoplasm of the host cell (Figure 1.3A). Animal rotavirus strains only attach to receptors that contain sialic acid (SA) while human rotavirus strains do not require SA containing receptors for attachment. Upon entrance of the virus into the host cell cytoplasm, the outer protein layer of the virus capsid is lost in the endocytic vesicles (Figure 1.3B), VP7 requires calcium to maintain capsid stability, and the low calcium concentration in the cytoplasm results in double-layered particle formation (Desselberger, 2014).

After entry and uncoating, transcription of the viral mRNA occurs (Figure 1.3D). The virus provides all the necessary enzymes. During transcription, the dsRNA molecules are processed into mRNA and used for the production of viral proteins by the host machinery as well as a template for the generation of negative sense RNA molecules. Viral mRNA transcription is carried out by virally encoded RNA-dependant RNA polymerase in the DLP. The transcripts are capped at the 5' end by the VP3 protein before exiting the DLP (Desselberger, 2014).

The positive sense 3'capped mRNA accumulates in the cytoplasm where it is translated into viral structural and non-structural proteins by the host cell ribosomes (Figure 1.3E). NSP3 facilitates translation by binding to the viral mRNA 3' consensus sequence through the N-terminal. The NSP3 C-terminal, in turn, binds to the host cell eIF4G preventing translation of the host cell mRNA (Vende et al., 2000). Synthesis of minus sense RNA and viral genome replication occurs in the cytoplasm of the host in viroplasms made up of NSP2 and NSP5 (Patton and Spencer, 2000). Synthesis of minus sense RNA and genome replication is followed by assembly of the DLP (Figure 1.3F). The DLP buds through the membrane of the endoplasmic reticulum (ER) with the help of NSP4 (Figure 1.3G). In the ER, VP4 and VP7 proteins are recruited to the DLP and the virus exits the ER as a mature TLP (Figure 1.3H). The virus exit the host cell by lysis (Estes and Greenberg, 2013).

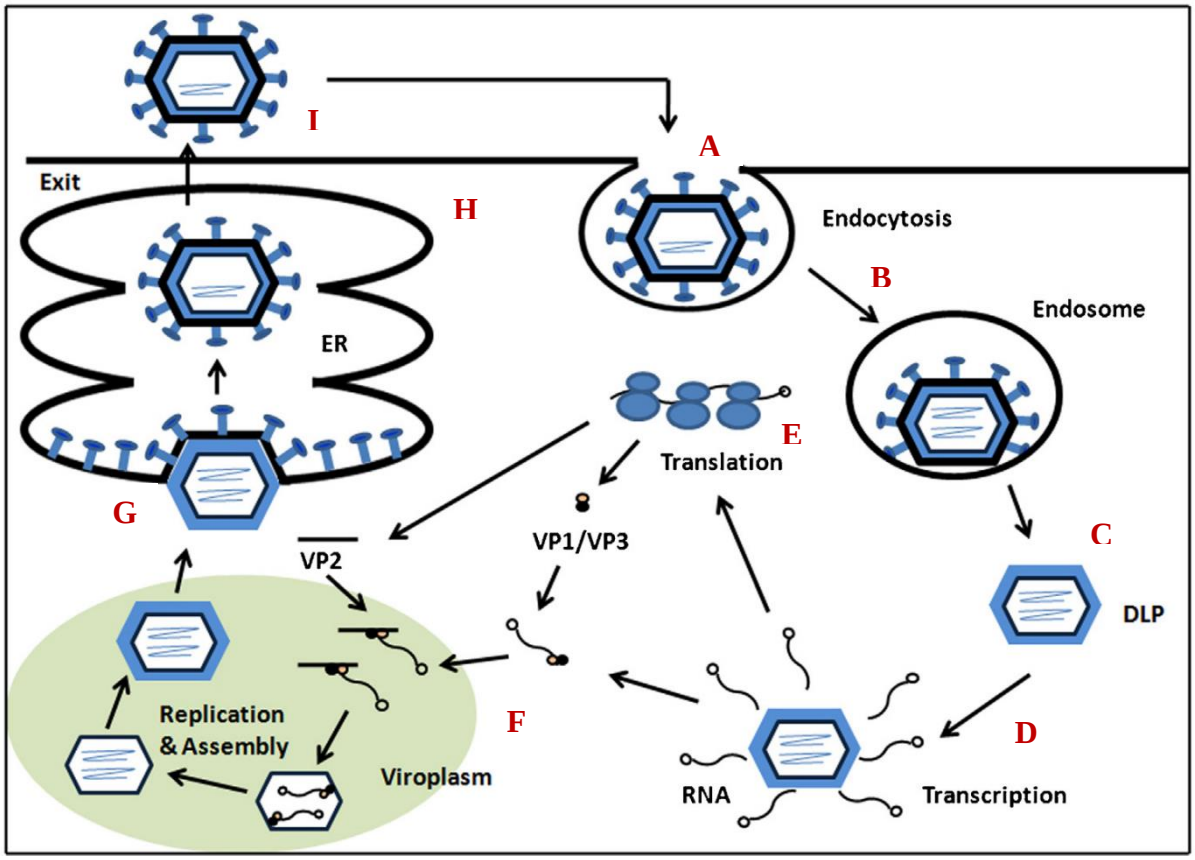


Figure 1.3: A schematic representation of the rotavirus replication cycle. The virus attaches to the host cell membrane and penetrates the membrane with the help of VP4 and VP7 capsid proteins (**A**). The virus is transported to the endosomes **(B)** where lower calcium concentrations lead to the loss of the outer layer, leaving the virus as DLP **(C)** that are released into the cytoplasm where transcription of the viral mRNA **(D)** begins, followed by translation **(E)**. Genome replication and assembly of the DLP occur in the viroplasm **(F).** The DLP are transported to the ER **(G)** where it obtains the third outer layer **(H)** and exits the cell through lysis **(I)** (Adapted from Jain et al., 2014).

1.5 Rotavirus classification

Four different rotavirus classification systems have been used, namely: serological classification into groups, electropherotype classification, binary classification, and whole genome classification. Each system has its own advantages and disadvantages. The choice of system depends mainly on the nature of the question being addressed and on the availability of resources. Rotaviruses are currently classified into eight groups based on the VP6 protein: group A to H (Matthijnssens et al., 2011). Groups A to C have been found to affect both humans

and animals, with group A viruses being the most common cause of diarrhoea in humans (Parashar et al., 2006; Yen et al., 2011). Group A rotaviruses negatively impact human health and lead to great financial losses through hospitalisation of infected individuals. In addition, studies have reported that group A rotaviruses account for a high percentage of all diarrhoeal cases in children below the age of 5 years (Ahmed et al., 2009; Kotloff et al., 2013; Nguyen et al., 2005).

Group B rotaviruses were first described in China in 1983, and were found to be responsible for diarrhoea and even death due to chronic diarrhoea in adults (Hung et al., 1983, Fang et al., 1989). Group C rotaviruses affect children between the ages 4 to 7 years and have been associated with food-born contamination (Trojnar et al., 2010). Rotaviruses belonging to groups D, E, F and G have only been associated with illness in animals, primarily avian species. Group H rotaviruses have only recently been added as the eighth group, based on the sequence analysis of the VP6 encoding genome segment (Matthijnssens et al., 2012). Group H viruses include the adult rotavirus strain, ADRV-N, isolated in 1997 from an outbreak of rotavirus in adults in China (Yang et al., 2004). Various studies have also reported the isolation of Group H strains from porcine (Marthaler et al., 2014; Nyaga et al., 2016).

1.5.1 Electropherotyping and subgrouping

Electropherotyping is a classification system in which rotaviruses are characterised according to their RNA migration patterns using polyacrylamide gel electrophoresis (PAGE). Using this system, rotaviruses can be grouped into two main groups: the “long” electropherotype and the “short” electropherotype (Luz et al., 2005). The difference between the two groups is based on the migration pattern of genome segments 10 and 11. In the “long” electropherotype group, gene segment 10 and 11 migrate according to size, whereas the “short” electropherotype, gene segment 11 migrates slower because it is larger than segment 10 resulting in a shorter pattern (Coulson et al., 1987).

1.5.2 Serological classification into serogroups

Serological classification is based on the reactivity of antibodies against epitopes on the VP6 protein. VP6 is the most immunogenic rotavirus protein and therefore readily detected (Esteban et al., 2013). In addition to electropherotyping, rotavirus strains are classified into two subgroups i.e. subgroup 1 (SGI) and subgroup 2 (SGII). Subgrouping is carried out using VP6 specific antibodies. Generally, SGI VP6 specific antibodies react with “short” electropherotype

strains, while SGII VP6 specific antibodies react with "long" electropherotype rotavirus strains. However, there are other rotavirus strains that react with both SGI and SGII VP6 specific antibodies or react with neither of the subgroup antibodies (Munford et al., 2007). The problem with subgrouping is the lack of available monoclonal antibodies for different subgroups. The low efficacy level was another concern as 20-30% of samples failed in most typing studies (Taniguchi et al., 1990). For these reasons, there was a need to develop new and more efficient typing techniques for rotaviruses (Taniguchi et al., 1992).

1.5.3 Binary classification

In order to overcome the problems associated with serotyping, a genotyping approach was developed for group A rotaviruses. Genotyping is based on sequence analysis of VP4 and VP7 encoding genome segments. The VP7 genotype is called the G type because VP7 is a **g**lycoprotein and the VP4 genotype is called the P type because VP4 is a **p**rotease sensitive protein (Rahman et al., 2003). Limitations of genotyping are that it is only applicable to group A rotaviruses. In addition, primers used for genotyping are not 100% accurate all the time, sometimes there are primer failures or cross reactivity of primers resulting in incorrect typing false results. Another limitation is that, since only two outer capsid proteins of the virus are used for classification, information on other genome segments are lost.

Based on this classification system, at least 27 different G genotypes and 37 different P genotypes have been determined to date (Matthjnssens et al., 2012). G serotypes assigned to rotavirus strains are the same as G genotypes assigned, and, therefore, rotaviruses are commonly described using G serotype. Conversely, P serotypes do not always correspond to P genotypes and, therefore, rotavirus strains are normally described with their P genotype denoted in brackets (Santos and Hoshino, 2005). One of the limitations of this method of classification is that since only two outer capsid proteins of the virus are used to classify viruses information on other genome segments such as viral virulence or evolutionary mechanisms cannot be inferred from this system.

1.5.4 Whole genome classification

Whole genome classification of rotaviruses are a recent development which involves nucleotide sequencing of all 11 viral genome segments and it was first described by Matthijnssens and co-workers (2008), all the 11 genome segments are assigned a genotype. In this system, the notation Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx is used to represent the

genotypes of the 11 genome segments that code for the viral proteins in the following order VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6 (Matthijnssens et al., 2008). Further details about the classification are shown in Table 1.2 below.

Table 1.2: Nucleotide percentage identity cut-off values used for assigning genotypes to the 11 rotavirus genome segments. The letters in bold are those used in the classification notation. (Adapted from Matthijnssens et al., 2008).

Gene product	Percentage identity Cut-off values	Genotype names
VP7	80	**G**lycosylated
VP4	80	**P**rotease-sensitive
VP6	85	**I**nner capsid
VP1	83	**R**NA-dependent RNA polymerase
VP2	84	**C**ore protein
VP3	81	**M**ethyltransferase
NSP1	79	**I**nterferon Antagonist
NSP2	85	**N**TPase
NSP3	85	**T**ranslation enhancer
NSP4	85	**E**nterotoxin
NSP5	91	p**H**osphoprotein

1.6 Rotavirus evolutionary mechanisms

Rotaviruses evolve using mechanism-like point mutations, genome reassortment, genome rearrangement, and recombination (Desselberger et al., 2001). These evolutionary mechanisms result in genetic diversity, either individually or in combination, and this is evidenced by the high numbers of G and P types as previously described in section 1.5.3. The commonly circulating G and P combinations globally include; G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] (Esona et al., 2010). While strain diversity differs from country to country, most diversity is reported from low income countries (Jain et al., 2001).

1.6.1 Point mutations

A point mutation occurs when there is a change in a single nucleotide at one location in the RNA sequence. Point mutations involve either base substitutions or frame-shift mutations if the mutation affects the open reading frame. Base substitutions occur when a single base is replaced or substituted by a different base. Frame-shift mutations occur in two different ways. Firstly, through an insertion of a base into the RNA sequence or secondly, through a deletion of a single nucleotide from the RNA sequence. Rotavirus genomes are likely to experience point mutations because their genome replication process is subjected to mutations since their RNA dependent RNA polymerase lacks proof-reading-activity (Estes and Greenberg, 2013). In rotaviruses, there is approximately one point mutation per genome replication (Blackhall et al., 1996). The accumulation of point mutations may lead to changes in the antigenic regions which result in genetic diversity (Kirkwood, 2010). In addition, point mutations can lead to the introduction of premature stop codons and in genome rearrangement (Desselberger et al., 2001).

1.6.2 Genome reassortment

Genome reassortment is the exchange of viral RNA segments between two or more viruses. It occurs when a single host cell is co-infected with two or more rotavirus strains. Therefore, genome segment exchange is likely to occur between the viruses resulting in a reassortment virus that carries genome segments from the different parent viruses. Rotavirus strains within the same rotavirus group have the ability to exchange genome segments that encode for antigenic determinants (Müller and Johne, 2007). Genome reassortment is a continuous evolutionary mechanism of which the assembling of the diversity of these viruses are limitless. Each genome segment can exchange with any other genome segment.

Genome reassortment results in unusual G/P type combinations. Unusual genotype combinations referring to those that do not commonly occur globally such as G3P[19]. Strains with unusual genotype combinations have been shown to cause a high burden of disease mostly in low-income countries, and this could be due to the fact that there are more cases of co-infections (20%) in low-income countries in comparison to the high-income countries (>5%) (Armah et al., 2010; Iturriza-Gómara et al., 2011). In addition, strains with unusual genotype combinations differ from one region to another.

Interspecies transmission is another form of genome reassortment. It occurs when a single host cell is infected by 2 or more rotavirus strains originating from different species. For example, between humans and animal strains, or between different species of animals, such as bovine and porcine strains, hence the name "interspecies transmission". Interspecies transmission is common in areas where humans live in close proximity to animals, such as cattle and pigs. It results in human reassortments that have an animal origin. It is believed that interspecies transmission is responsible for the wide diversity of rotaviruses in low-income countries where people live in close proximity to animals (Heylen et al., 2014). The development of whole genome characterisation has made it possible to study reassortment events in more detail and many reassortments have been described (Jere et al., 2012; Ramani et al., 2009).

1.6.3 Genome rearrangement

Rotavirus genome rearrangement is a change in more than one nucleotide in a sequence of a genome segment. Genome rearrangement is a result of a partial duplication of a genome segment (Desselberger, 1996). The consequences of genome rearrangement is the production of new viral proteins with altered functions, while the normal proteins are not produced. Partial duplication of a genome segment is the most common form of rearrangement, whereby the ORF is still intact but has an extended 3' untranslated region. The majority of genome rearrangements occur in non-structural genome segments, more especially in segments 5, 8, 10 and 11. However, genome rearrangement of a VP6 encoding segment has also been reported (Shen et al., 1994). Unlike genome reassortment and other evolutionary mechanisms, strain diversity due to genome rearrangement is not very common (Desselberger and Gray, 2013).

1.6.4 Recombination

Genome recombination may result in the formation of novel rotavirus strains, and requires that a single host cell be infected by 2 or more rotavirus strains (Estes and Greenberg, 2013). In

genome recombination, the nucleotide sequence of a genome segment consists of regions that originate from more than one virus strain. This is evidenced by the alignment of 2 or more regions from different strains that have contributed to the formation of the recombinant genome segment. Recombination can be intergenic, between viruses with the same genotype (Parra et al., 2004) or intragenic, between viruses with different genotypes (Jere et al., 2011a). In addition, recombination can occur in more than one genome segment of a single strain. A study by Jere and co-workers reported on recombination events in more than one genome segment in various strains characterised in a mixed infection (Jere et al., 2011b).

1.7 Rotavirus detection and characterisation

1.7.1 Rotavirus detection

Rotavirus infections in young children are characterised by diarrhoea, vomiting and fever. However, many bacteria and viruses elicit similar symptoms and rotaviruses cannot be diagnosed on symptoms alone. Laboratory tests need to be performed in order to confirm the diagnosis. There are several laboratory tests used for rotavirus detection. All these tests are based on either the detection of the actual rotavirus particles, or the viral antigens found in stool samples. There are advantages and disadvantages associated with these tests.

Rotavirus particles can be directly observed in stool samples obtained from people with rotavirus infections using an electron microscope (EM). With an EM, samples containing the virus particles are stained before viewing. Negative stains, such as uranyl acetate or phosphotungstic acid, are used. The advantage of using an EM for detection of rotavirus particles is that the method is accurate and reliable. However, the method of staining affects the results, and an EM requires a highly skilled operator (Estes and Cohen, 1989). In addition, EM uses expensive equipment that may not be available in laboratories.

The most commonly used diagnostic test is the enzyme-linked immunosorbent assay (ELISA), because it is easily available and relatively inexpensive. In this assay, rotavirus antigens are detected using antibodies specific for VP6 group antigens. The antibody used is normally fused to an enzyme that, upon binding of the antigen to the antibody, the enzyme fluoresces or changes colour when the substrate specific for the enzyme is added to the reaction. There are different types of ELISAs that can be used and these are direct, indirect, competitive, or sandwich ELISA. In direct ELISA, only one antibody is used and therefore is a quick method.

Another advantage is that there is no cross reactivity that may be caused by the secondary antibody. A disadvantage associated with this method is the labelling of the primary antibody with an enzyme might negatively affect the antibody's immunoreactivity. There is no flexibility in terms of choice when it comes to the labelling of primary antibodies from one experiment to another.

In indirect ELISA, a secondary labelled antibody is used and it is added to the primary antibody. An advantage is that there is a wide variety of secondary antibodies that may be used and are available commercially. A disadvantage is that the use of a secondary antibody might result in cross reactivity. Sandwich ELISA is similar to indirect ELISA in that a primary and secondary antibody is used. The only difference is that sandwich ELISA adds the antigen in between the two antibodies. The advantages of sandwich ELISA are that it has a high specificity and sensitivity, since two antibodies are used. Competitive ELISAs also make use of a primary and a secondary antibody, but an additional substance is added to compete with the binding of the antigen to the antibody so as to ensure that the antigen is bound to the antibody that it has a high affinity for. Advantages are that competitive ELISAs are also highly sensitive and there is no need for prior purification of antigens implying that even an impure antigen sample may be used (Estes and Greenberg, 2013).

The other method that has been used for rotavirus detection is dsRNA profiles on agarose gels. In this method, dsRNA is extracted from stool samples, and separated by gel electrophoresis. Rotavirus genomes produce a characteristic dsRNA profile that can be used for detection (WHO, 2009). The advantage associated with this method is that the results generated are not ambiguous. In contrast, the disadvantage associated with this method is that it is time consuming, not as sensitive, and also requires a skilled operator.

Rotaviruses can also be detected using a reverse transcriptase polymerase chain reaction (RT-PCR). In RT-PCR, the viral RNA is converted into cDNA. Primers that bind specific sections of the viral genome are added to the reaction, and multiple copies of these sections of the viral genome are amplified through multiple cycles of the PCR step. The results are then evaluated with agarose gel electrophoresis. The advantage of using the RT-PCR method is that it is also possible to separate genotype viruses into P and G genotypes using genotype specific primers (Estes and Greenberg, 2013).

1.7.2 Rotavirus Genotyping

Group A rotaviruses are classified into P and G genotypes according to the binary classification described in section 1.5.3. Genotyping PCR is used in the dual typing classification and it is based on different primer sets designed to target and amplify specific regions on rotavirus cDNA. For G genotyping, there is a universal primer set that is used to target and amplify the VP7, encoding full length genome segments for all group A rotavirus strains. The forward primer is Beg9, and it is 28 nucleotides long. The reverse primer is End9, and it is 27 nucleotides long (Gouvea et al., 1990). In PCR genotyping, the universal primer set is used in the first round of amplification, and the resulting amplicon is used in the second round of amplification as a template. In the second round of amplification, Beg9 and a cocktail of genotype specific primers are used, the sizes of the amplified amplicons which are analysed by agarose gel electrophoresis correspond to particular genotypes.

In P genotyping, a universal primer set also exists that target and amplify a partial sequence on the VP4 encoding genome segment of group A rotavirus strains. Similar to G typing, in P typing the universal primer set is used in the first round of amplification. In the second round of amplification, a cocktail of genotype specific primers is also used for the detection of present genotypes (Gentsh et al., 1992). PCR genotyping is not 100% accurate due to certain limitations, for instance, sometimes there may be false positive results due to cross-reactivity of primers. At times, primers used may fail to type if there are mutations at the primer binding sites, and, therefore, may require alternative primers to be designed (Iturizza-Gómara et al., 2000).

1.8 Rotavirus Epidemiology

Rotavirus infections are a huge problem world-wide. The infections range from mild watery diarrhoea to severe dehydration which may lead to death (Gray et al., 2008). By the age of 5 years, over 90% of children world-wide would have been exposed to rotavirus (Parashar et al., 2003). Tate and co-workers reported that, annually, rotavirus infections cause approximately 215 000 deaths world-wide and more than half of these occur in Sub-Saharan African countries and other low income countries (Tate et al., 2016). Studies such as the one reported by Sanchez-Padilla et al., 2009 are some of the few rotavirus surveillance studies reported in African countries. In high-income countries, there are a lot of studies that have been published on

rotavirus surveillance due to more active surveillance programmes in these countries. Lack of resources in African countries to conduct such studies has been the main reason for the poor surveillance in these countries.

Data from rotavirus surveillance in Africa suggests that the burden of rotavirus disease is high. A study conducted between 2006 to 2008 in 9 African countries namely - Cameroon, Ethiopia, Ghana, Kenya, Tanzania, Togo, Uganda, Zambia, and Zimbabwe showed a 41% average detection rate of rotaviruses in hospitalised children under the age of 5 years (Mwenda et al., 2010). Other similar studies conducted in African countries have reported a median detection rate of rotaviruses in hospitalised children under the age of 5 years to be 20 - 32.8% (Mwenda et al., 2010, Alexander and Hay, 1986).

In high-income countries, rotavirus infections are also a problem, the difference being the number of deaths attributable to rotavirus. Despite fewer death rates in high-income countries, many children below the age of 5 years still contract rotavirus leading to hospitalisation (Tucker et al., 1998).

Children in low-income countries contract rotavirus at a much earlier age in comparison to children in high-income countries (Parashar et al, 1998). Bresee and co-workers reported that the average age at which children develop symptomatic rotavirus infections in low-income countries is 3 months, while in high-income countries is between 9 to 15 months (Bresee et al., 1999). A study on rotavirus in the European Union member countries concluded that each year about 3.6 million rotavirus episodes occur. Most of the infections in children from these countries appear to be mild enough to be managed at home (Soriano-Gabarro´ et al., 2006). Out of the 3.6 million episodes, only 231 cases lead to death. This is significantly lower compared to deaths in low-income countries (Soriano-Gabarro´ et al., 2006).

Rotavirus detection is high all year round in countries with tropical climates, whereas infections in temperate climates peak during the fall and winter seasons (Cook et al., 1990). The reason for this occurrence is not yet well understood. In South Africa, rotavirus have the same trend as areas with temperate climates (Steele and Alexander, 1987). In a study by Mwenda and co-workers conducted in 2006 and 2008 in various African countries, the seasonality of rotavirus differed by country due to climatic conditions. In Ghana, the highest rotavirus infections occurred during the cool and dry months in January and February. In Zimbabwe, the highest infections occurred during the winter season in May and August, while in Ethiopia, infections peaked between October and December (Mwenda et al., 2010).

Rotavirus strain diversity is high, with the highest diversity reported in low-income countries. Prevalent strains in high-income countries include G1P[8], G4P[8], G3P[8], and G2P[4], while in developing countries include G2P[6], G3P[6], G12P[8], and G9P[6]. Strains that are prevalent global circulating strains are G1 to G4, and G8 in combination with P[4], P[6], and P[8] (Santos and Hoshino, 2005).

Seheri and co-workers reported on rotavirus strains circulating in Africa for a period of 5 years between 2007 and 2011. Countries included in the study were in East Africa, West Africa, Southern Africa and Central Africa. Genotype G1 was the most predominant strain detected in all the regions except in Central Africa, where it was less frequently detected. The second most common genotype was G9 followed by G2. In countries such as Uganda, Tanzania and Cameroon, G2 was not common while in countries such as South Africa, Gambia, DRC and Guinea Bissau it was frequently detected. Other genotypes detected included G8, G12 and G3.

The most prevalent P genotype was P[8] followed by P[6], and P[4]. P[8] was found in combination with G1, G3, G4, G9, G8 and G12, P[6] in combination with G1, G2, G3, G9, G8, and G12 and P[4] in association with G2, G9, and G8. A high percentage of untypable strains, as well as mixed genotypes, were also described. Data from this study shows that there is a high rotavirus strain diversity in African countries (Seheri et al., 2014).

1.9 Rotavirus vaccines and impact

Treatment of gastroenteritis caused by rotavirus aims at replacing fluids and electrolytes that have been lost through diarrhoea and vomiting (Santosham et al., 1997). In addition, rotavirus vaccines have been introduced to prevent severe gastroenteritis caused by rotavirus. Improved hygiene levels, as well as cleaner water supplies, have not been proven to decrease disease due to rotavirus in both industrialised and low-income countries (Parashar et al., 2009).

Rotavirus vaccine development projects aim at targeting diseases due to group A viruses, since these viruses are responsible for most deaths in humans. RotaShield® (Wyeth) was the first rotavirus vaccine to be licenced for use in 1998. It was administered to infants in 3 doses at 2, 4, and 6 months of age. The use of this vaccine was, however, discontinued due to the suspected association with intussusception (Shadman, 2000).

Currently, there are 2 rotavirus vaccines licensed for use globally, RotaTeq® (Merck) and Rotarix® (GlaxoSmith-Kline). The former was first licensed for use in 2004. It is a three-dose human-bovine reassortment pentavalent vaccine administered at two, four and six months of age (Heaton et al., 2005). It contains human G1, G2, G3, and G4 rotavirus reassortments and a bovine P[8] reassortment. Rotarix® is a G1P[8] two-dose, monovalent, live-attenuated vaccine administered at two and four months of age. Rotarix® was first licensed for use in 2006 (Dennehy, 2008). Both vaccines are used in many countries, and have proven to be effective in decreasing the severity of rotavirus infections as well as the number of deaths (Richardson et al., 2010). However, there are still numerous countries in Africa where these vaccines are not used in routine immunisation programmes. This is mainly due to a lack of funds to procure the vaccines.

The efficacy of rotavirus vaccines differs in high, middle, and low income countries. The highest efficacy is observed in the high income countries followed by middle then lower income countries. For example studies have reported the efficacy of Rotarix® in Latin America at 85%, in South Africa at 77%, and in Malawi at 49% (Madhi et al., 2010; Patel et al., 2009). A similar trend has also been reported for RotaTeq® where efficacy in Asia was 51%, and in Africa it was 64.2% (Patel et al., 2009).

Reasons for low rotavirus vaccine efficacy in low income countries are not yet clearly understood, but can be explained in part by a few factors. Firstly, the co-administration of rotavirus vaccines and oral poliovirus vaccines (OPV) which are administered more frequently to children in low income countries and less frequently to children from high income countries (Patel et al., 2012). Secondly, rotavirus maternal antibodies that have been transferred from the mother to the child through the placenta or from breast milk, may lower rotavirus vaccine efficacy by neutralizing rotavirus strains in the vaccine (Moon et al., 2010; Moon et al., 2015). Studies have reported that blood serum samples of children from low income countries have more rotavirus antibodies compared to those of children from high income countries, and a similar observation applies for rotavirus antibodies in breast milk (Moon et al., 2010). Thirdly, the age at which children first contract rotavirus infection may explain the low rotavirus vaccine efficacy observed in low income countries. Rotavirus vaccines are more effective when administered before natural immunity is acquired through natural infections, therefore, administering the vaccine after rotavirus infections, renders the vaccine less effective. In low income countries, it has been reported that children are infected with rotavirus at a much earlier age than children from high/middle income countries (Sanderson et al., 2011). In low income

countries, there is much wider variety of rotavirus genotypes and the vaccine may offer partial protection due to the presence of genotypes that are not included in the vaccine. Other factors such as malnutrition, which is common in low income countries, can also contribute to low rotavirus vaccine efficacy (Patel et al., 2009).

1.10 Problem Statement and Rational Behind Study

Initially, rotavirus surveillance was not conducted in Africa due to a lack of resources. However, researchers and the World Health Organisation (WHO) established the African Rotavirus Network to overcome this gap in information. Mwenda and co-workers reported an average rotavirus prevalence of 41% (16 to 57%) in children below the age of 5 years in a study conducted in 11 selected African countries on the burden and epidemiology of rotavirus before the introduction of vaccines (Mwenda et al., 2010). In the study, the disease burden was found to be high, especially in children in the age group 3 to 12 months. In addition, besides the detection of the globally common genotypes, a larger proportion of unusual and mixed genotypes were detected.

In South Africa, rotaviruses were first isolated in 1976 (Steele and Glass, 2011), and rotavirus surveillance was established in the 1990s. An average rotavirus prevalence of 32.8% was reported in South Africa with a peak of infections occurring in autumn, late winter, or early spring (Steele and Glass, 2011). Rotavirus strains circulating in South Africa, as with other African countries, indicate a wide range of strain diversity. Various studies have also reported the detection of mixed infections and animal strains (Esona et al., 2011).

In contrast to South Africa, Mozambique has limited information on rotavirus strain circulation. To date, only one study (Langa et., al 2016) has been published on rotavirus genotypes in Mozambique. A rotavirus surveillance project was initiated by Dr Nilsa de Deus (National Institute of Health, Maputo) at the Manhiça Health Research Centre (CISM) in 2012. The aim of the project was to study the prevalence and the genotypes of strains in two regions: Mavalane (urban site) and Manhiça, (rural site). The rotavirus prevalence in these regions was 40%, similar to the prevalence reported by Mwenda and co-workers in other African countries. A collaboration between CISM, Mozambique, and the University of the Free State on the characterisation of Mozambican rotavirus strains was established. The work reported in this study forms part of the collaboration.

The main aims of the study were to characterise Mozambican rotavirus P and G genotypes by means of RT-PCR and Sanger sequencing. Furthermore, whole genome analysis of selected Mozambican strains using next generation sequencing technology was performed to provide additional information on the remaining nine genes.

The specific objectives were:

(i) To confirm the genotypes of rotavirus strains circulating in Mozambique that have been determined with genotyping PCR, including mixed samples.

(ii) To determine the genotypes of samples which could not be typed using genotyping PCR.

(iii) To determine the whole-genome constellations for selected Mozambican strains.

(iv) To compare rotavirus genes of strains circulating in Mavalane and of those circulating in Manhica.

(v) To assess whether some genotyping primers that failed might be due to mismatches at the primer binding sites.

(vi) To do phylogenetic analysis of Mozambican rotavirus strains in context with strains circulating in Sub-Saharan Africa and globally circulating strains.

Chapter 2[1]

Dual Genotyping of Mozambican rotavirus strains using reverse transcriptase polymerase chain reaction and Sanger sequencing

2.1 Introduction

The introducting of molecular techniques including PCR and RT-PCR for rotavirus characterisation has been a major advancement, allowing improved classification of rotaviruses. The discovery of six conserved amino acid sequences on the VP7 protein provided type specific regions for PCR genotyping. The amino acid regions differ between serotypes, but are highly conserved within a given serotype (Gouvea et al., 1990).

The genotyping multiplex PCR is a cocktail of genotyping-specific primers in a PCR reaction. The PCR products are analysed on an agarose gel, and genotypes are differentiated based on the sizes of the amplicons (Gouvea et al., 1990, Gentsch et al., 1992). Sanger sequencing, developed by Frederick Sanger and co-workers in 1977, determines the nucleotide sequence of a gene (Sanger et al., 1977). It is used in combination with BLASTn (Basic Local Alignment Search Tool nucleotide database) analysis to identify the genotype of rotaviruses. In addition, phylogenetic inference can be used to study the relationships between rotaviruses (Brinkman and Leipe, 2001).

There is limited information on rotavirus surveillance in Mozambique. In addition, there is no published information on rotavirus genotypes that are circulating in the country. In 2015, Nhampossa and co-workers reported a decrease in rotavirus disease burden over time, but despite this observation, disease and death, due to diarrhoea, remained an important health concern that required attention. The study was conducted in Manhiça, a rural area in the southern part of the country, about 80 kilometres away from the capital Maputo. The focus of the study was to establish the burden, risk factors, and the aetiology of diarrhoeal diseases among children 0-59 months old. In this study, the authors concluded that rotaviruses were the most prevalent cause of diarrhoeal diseases (Nhampossa et al., 2015).

[1] Data generated in this chapter was included in a manuscript entitled "Rotavirus A strains obtained from children with acute gastroenteritis in Mozambique, 2012-2013: G and P genotypes and phylogenetic analysis of VP7 and partial VP4 genes" that has been scientifically accepted for publication in *Archives of Virology*.

In Mozambique, rotavirus surveillance that was initiated in 2012, focused on two study sites, Manhiça, a rural site, and Mavalane, an urban site. This surveillance study indicated a prevalence of 40% for rotavirus infections in Mozambique (de Deus et al., 2017). This chapter describes the confirmation of previously PCR genotyped Mozambican rotavirus strains using RT-PCR and Sanger sequencing of the partial VP4 and full length VP7 encoding genome segments. The results were compared to the genotyping RT-PCR results previously generated by our Mozambican collaborators. Finally, phylogenetic analysis of the various P and G genotypes of the Mozambican strains was performed.

2.2 Materials and Methods

2.2.1 Rotavirus strains and ethical clearances

Mozambican rotavirus strains were obtained from Dr Nilsa de Deus (National Institute of Health, Mozambique) under a material transfer agreement (MTA) between CISM and the University of the Free State (UFS). The project was approved by the Mozambique ministry of health (IRB00002657/2010) and the UFS ethics committee (UFS ECUFS NR 201/2013). Ethics approval certificates are included in Appendix A. The Rotavirus strains collected at two sites, Mavalane and Manhiça, shown in figure 2.1, are located in the southern part of Mozambique. Informed consent was obtained from the parents of the patients. The 58 specimens received from Mozambique are listed in Appendix B, in Table 1 (Mavalane) and Table 2 (Manhiça).

Figure 2.1 Map of Maputo province in southern Mozambique indicating the study sites Manhiça and Mavalane which are suburbs of the capital Maputo.

2.2.2 RNA extraction

Total RNA was extracted from 58 stool samples using TRI-reagent® (Sigma-Aldrich, st Louise, WI), a phenol-chloroform extraction method. These samples were previously determined to be positive for the rotavirus antigen using an enzyme-linked immunosorbent assay (ELISA). Solid stool samples were diluted to 10% suspensions in phosphate buffered saline (PBS), and 500 µl of the suspension was used for extraction. For liquid stool samples, 300-500 µl of the sample was used for RNA extraction, without diluting the samples. One millilitre of the TRI-reagent® was added and mixed with the stool samples, followed by addition of 300 µl chloroform (Sigma). The mixture was then centrifuged at 4 ˚C for 15 minutes at 15 800 revolutions per minute (rpm; Eppendorf Model 5804R). After centrifugation, the clear supernatant was removed and added to 650 µl isopropanol (Merck), centrifuged at room temperature for 15 minutes at 15.8 rcf (Eppendorf Model 5804R), and the supernatant discarded. The pellet was then re-suspended in 95 µl elution buffer MinElute® Gel Extraction kit (Qiagen). Extracted RNA was stored at -20ºC. The integrity of the RNA was analysed with agarose gel electrophoresis (Section 2.2.4).

2.2.3 cDNA synthesis and PCR amplification

The extracted RNA was reverse transcribed into cDNA using Avian Myeloblastosis Virus (AMV) reverse-transcriptase (Thermo-Scientific). The Beg9 (5'- GGC TTT AAA AGA GAG AAT TTC CGT CTG G-3') and End9 (5'- GGT CAC ATC ATA CAA TTC TAA TCT AAG-3') primer set was used for cDNA synthesis and PCR amplification of the VP7 encoding genome segment (Gouvea et al.,1990). Primers used for cDNA synthesis and PCR amplification of the VP4 encoding genome segment were Con2 (5'-ATT TCG GAC CAT TTA TAA CC-3') and Con3 (5'- TGG CTT CGC TCA TTT AT AGA CA-3') (Gentsch et al.,1992). All cDNA reactions had a final volume of 10 µl. The reactions consisted of the following components: primers at a final concentration of 2 pmol.µl^{-1}, AMV buffer (ThermoScientific) at a 2X final concentration, RNAse inhibitor (Promega) at a final concentration of 2 U/µl, dNTPs (Kapa Biosystems) at a final concentration of 2 mM, AMV reverse transcriptase (ThermoScientific) at a final concentration of 1 U/µl, and molecular grade water. cDNA synthesis reactions were incubated at 40 ^{o}C for one hour.

The synthesised cDNA was used to amplify partial VP4 and full length VP7 encoding genome segments. Five microliters from the cDNA reaction was added to a mixture containing KAPA HiFi polymerase (Kapa Biosystems) at a final concentration of 1 U/µl, KAPA dNTP mix (KAPA Biosystems) at a final concentration of 0.3 mM, forward, reverse primers at a final concentration of 0.3 pmol.µl^{-1} , and 5X KAPA HiFi buffer (Kapa Biosystems) at a 1X final concentration. All PCR reactions had a final volume of 50 µl. Initial denaturation step took place at 95 °C for 3 min, followed by annealing temperature of 42 °C for 30 seconds and a final extension temperature of 72 ^{o}C (1 min/kb). These cycling conditions were repeated 30 times. The PCR product was analysed on a 1% agarose gel in TAE buffer (Tris Acetic-EDTA). RT (reverse transcriptase) and PCR (polymerase chain reaction) reactions were run in a G-Storm Thermo Cycle (G-Storm™).

2.2.4 Gel electrophoresis

For agarose gel electrophoresis, 1% agarose gels were used for analysis of the extracted RNA and PCR products of genome segments 4 and 7. The gels were prepared with 1 g (gram) agarose powder (Lonza) and 100 ml of 1X Tris-Acetic acid-EDTA (TAE) buffer at pH 8.5 (40 m Tris, 2 mM EDTA, 20 mM glacial acetic acid). For analysis of extracted RNA, 0.5X Tris-Borate-EDTA buffer at pH 8.0 (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA) was used to dissolve

the agarose powder. The gel was electrophoresed at 90 Volts for 45 minutes. Ethidium bromide (Thermo-Scientific) at a final concentration of 0.6 mg/ml was used to stain the gel which was visualised with ChemiDoc XRS (Bio-Rad Laboratories) UV light. GeneRuler DNA ladder mix (Thermo-Scientific) was used to estimate the sizes of the bands of interest. For both extracted RNA and PCR products, 10 µl were loaded for analysis and for the marker 5µl was loaded.

2.2.5 DNA gel extraction

The fragments, that had the expected sizes after PCR product analysis on agarose gel electrophoresis, were cut out with a razor blade under a DarkReader^{TM} transilluminator (Clare Chemicals). The excised gel pieces were transferred to a microfuge tube that had been previously weighed. The weight of the gel slice was subsequently determined. Extraction of the DNA from the gel was performed using the Macherey-Nagel NucleoSpin® Gel and PCR clean-up kit. Buffer NT1 (700 µl) was added to the gel slice and incubated at 50^{o}C for about 10 minutes or until the gel slice was completely dissolved in the buffer. The contents were transferred to a spin column and centrifuged at 11 000 rcf (Eppendorf Model 5418) for 30 seconds and the flow through discarded. DNA on the spin column was washed with 700 µl of buffer NT3 and centrifuged at the same speed and time as the previous centrifugation step. This was followed by elution of the DNA from the column using 30 µl of the elution buffer. The success of the DNA extraction was determined by analysing the extracted DNA using agarose gel electrophoresis (section 2.2.4).

2.2.6 Nucleotide sequence determination using Sanger sequencing

The gel extracted PCR product was used for Sanger sequencing using the BigDye terminator v.3.1 cycle sequencing kit (Applied Biosystems) and analysed on an automated DNA sequencer (ABI PRISM 3100). The primers used for sequencing reactions were Beg9 and End9 for VP7 encoding genome segment and Con2 and Con3 for VP4 encoding genome segment determination, all at a final concentration of 0.32 pmol.µl^{-1}. Sequencing buffer was at a final concentration of 1X and all PCR reactions had a final volume of 10 µl. Denaturation of the synthesised cDNA (section 2.2.3) was at 96 °C for one minute, and 25 cycles were performed at 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. The concentrations of the cDNA templates were in the range of 5-20 ng.

Following sequencing PCR, the PCR product was purified using the EDTA/Ethanol precipitation protocol according to the manufacturer's instructions. The PCR product was mixed with molecular water in a 1:1 ratio and then added to a mixture of 125 mM EDTA and 60 µl absolute ethanol and centrifuged at 20 000 x g for 15 minutes at 4 ^{o}C. The supernatant was completely aspirated before the addition of 60 µl 70% ethanol and centrifuged at the same speed and temperature as the previous centrifugation step but for 5 minutes. After centrifugation, the supernatant was completely aspirated and the DNA dried in a speed-Vac (Thermo-Scientific) for 5 minutes and submitted for electrophoresis on an automated DNA sequencer (ABI PRISM 3100; Department of Microbial, Biochemical and Food Biotechnology UFS).

2.2.7 Data Analysis

Sanger sequencing results were edited using Genious 6.06 (Biomatters) and consensus sequences were generated using forward and reverse sequences of each amplicon. The consensus sequences were submitted to NCBI for BLASTn analysis, to determine the Mozambican rotavirus strains genotype and to identify sequences to construct phylogenetic trees. Phylogenetic analysis was used to infer phylogenetic relatedness of the Mozambican rotavirus strains with other international rotavirus group A strains. The GeneBank accession numbers for these rotavirus strains are given in Appendix C. Molecular Evolutionary Genetics Analysis software version 5.2 (MEGA 5.2) (Tamura et al., 2011) was used to construct phylogenetic trees and sequences were aligned with the Muscle alignment tool in MEGA 5.2. Phylogenetic trees were constructed using the Neighbour Joining method (Saitou and Nei, 1987) with 1000 bootstrap replicates. Kimura 2 parameter (Kimura, 1980) was used for the determination of genetic distances and the method was determined by MEGA according to the best fit model.

2.2.8 Evaluation of genotyping primer binding regions

Sanger sequencing results we used to evaluate genotyping primer binding sites for the primers used by the Mozambican group.

2.2.9 Determination of differences in the VP4 and VP7 antigenic epitopes of Mozambican strains and vaccine rotavirus strains.

Nucleotide sequences of rotavirus strains that were included in the analysis were firstly translated into protein sequences. Differences in the VP4 and VP7 antigenic epitopes between Mozambican strains and vaccine strains were evaluated.

2.3 Results and Discussion

2.3.1 RNA extraction and RT-PCR amplification of the VP4 and VP7 encoding genome

Fifty eight stool samples collected from Mozambican children presenting with diarrhoea in various hospitals in Manhiça and Mavalane, were received from CISM. Of the 58 stool samples received, RNA was successfully extracted from 40 samples (Appendix B, Table 1 and 2). A representative gel of the results obtained after TRI-reagent® RNA extraction is shown in Figure 2.2.

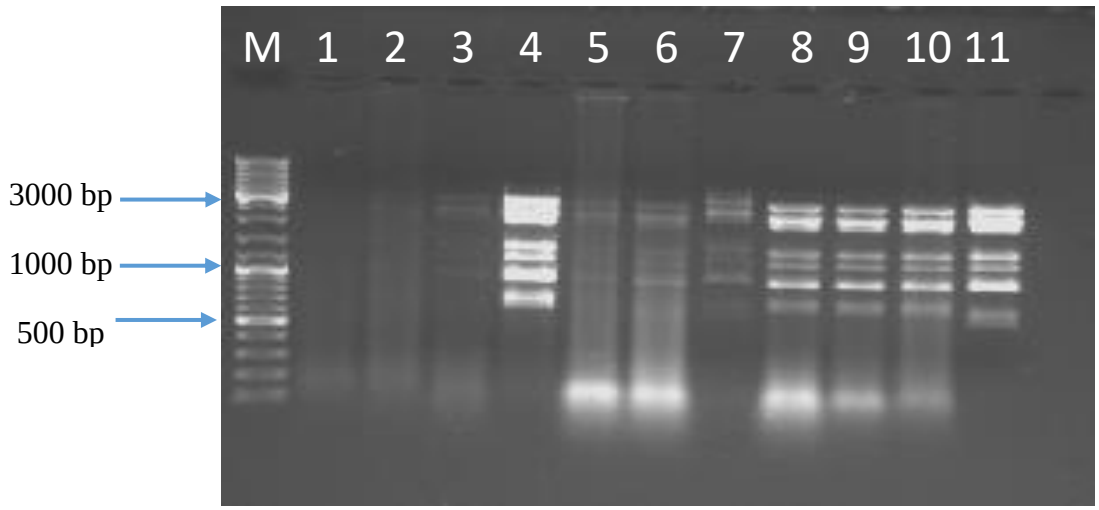


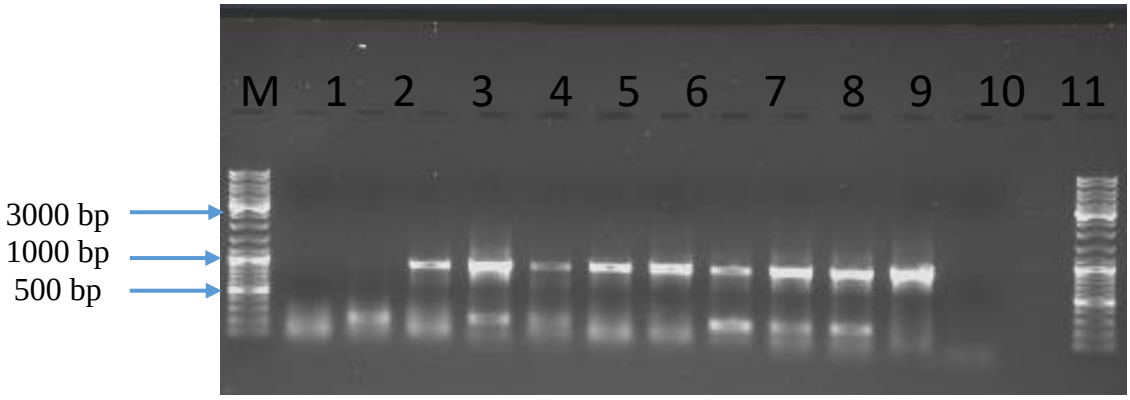
Figure 2.2: Representative gel showing analysis of the extracted dsRNA on 1% agarose gel. Lanes 1-10 are samples 0002, 0277, 0278, 0285, 0286, 0287, 0288, 0293, 0304, 0308 and 0310. Lane M is the GeneRuler DNA ladder Mix molecular weight marker (Thermo-Scientific) that was used to confirm the presence of the rotavirus segmented genome.

Of the 18 samples that RNA could not be extracted from, 11 samples were from Manhiça and 7 samples from Mavalane. Negative samples for RNA extraction (lanes 1-2 in Figure 2.2)

displayed an absence of distinct bands and appeared as smears on the gel. These samples were ELISA positive for the rotavirus antigen but no RNA was extracted using the phenol-chloroform method. This could be due to the fact that RNA in the stool samples was degraded, or that the dsRNA was at a very low concentration such that it could not be detected.

For those samples which RNA was extracted, partial PCR amplification of the VP4 and VP7 encoding genome segments was successful for the majority of samples as analysis on 1 % agarose gel indicated in Figures 2.3 and 2.4. Genome segment 4 which encodes for VP4 is 2362 bp in size, based on the genome sizes described for SA11. Genome segment 9 which encodes for VP7 is 1062 bp long based on SA11. Beg9 and End9 used for the amplification of full length genome segment 9 amplify a region that is 1062 bp long. In contrast, Con2 and Con3 primers used for the amplification of genome segment 4, do not amplify the whole segment but instead amplify a region that is 876 bp long.

Figure 2.3 shows that PCR amplification of the partial VP4 encoding genome segment using KAPA HiFi DNA polymerase was successful as the bands obtained were at the expected sizes as confirmed by the molecular weight marker. Similarly, Figure 2.4 shows that PCR amplification of VP7 encoding genome segment was successful as bands obtained were approximately 1000 bp. Six samples namely; 0040, 0042, 0102, 0113, 0223 and 0439 were negative for RNA extraction but PCR amplification of VP4 and VP7 encoding genome segments was successful. This confirmed results obtained at CISM that these samples were positive for the rotavirus antigen. The reason why RNA extraction was negative for these samples could be because RNA form this samples was at a very low concentration such that the bands were not visible on agarose gel electrophoresis. For most samples that RNA extraction was successful, both VP4 and VP7 enconding segments could be amplified with the exception of three samples including sample 0038 shown in Figures 2.3 and 2.4 where only one of the two genome segments were amplified.



2.3: Representative gel showing the analysis of the partially amplified VP4 encoding genome segment on a 1% agarose gel. Lane M is the GeneRuler DNA ladder mix molecular weight marker (Thermo-Scientific) and lanes 1-10 are samples 0024, 0038, 0051, 0146, 0153, 0208, 0289, 0297, 0428, 0440 and 0441.

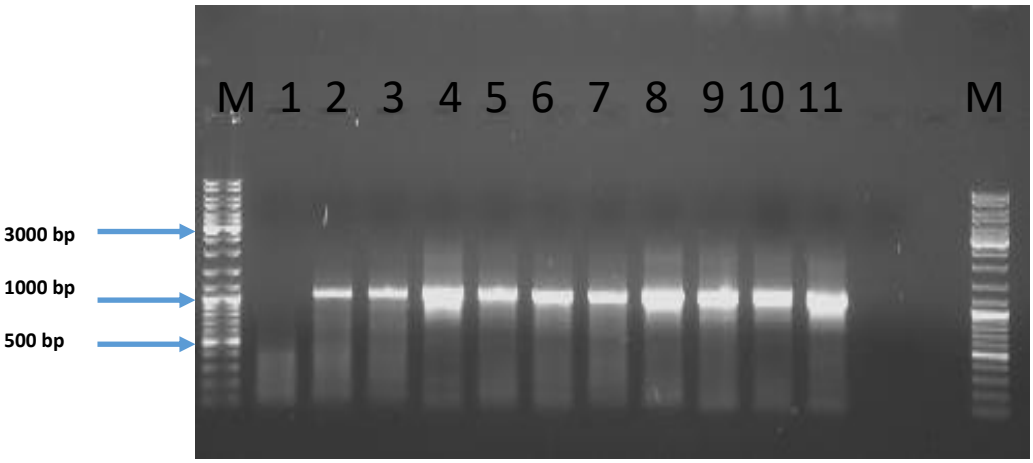


Figure 2.4: Representative gel showing the analysis of the amplified VP7 encoding genome segment on a 1% agarose gel. Lane M is the GeneRuler DNA ladder mix molecular weight marker. Lanes 1-11 are amplified VP7 encoding genome segments for Mozambican samples 0024, 0038, 0051, 0146, 0153, 0208, 0289, 0297, 0428, 0440 and 0441.

The PCR amplified VP4 and VP7 encoding genome segments were gel-purified using the Macherey-Nigel gel clean-up kit to remove primer dimers and undesirable smears that may be present. The results obtained confirmed that the purification process was successful (Figure 2.5). There were no undesirable bands and primer dimers observed after analysis of the purified DNA.

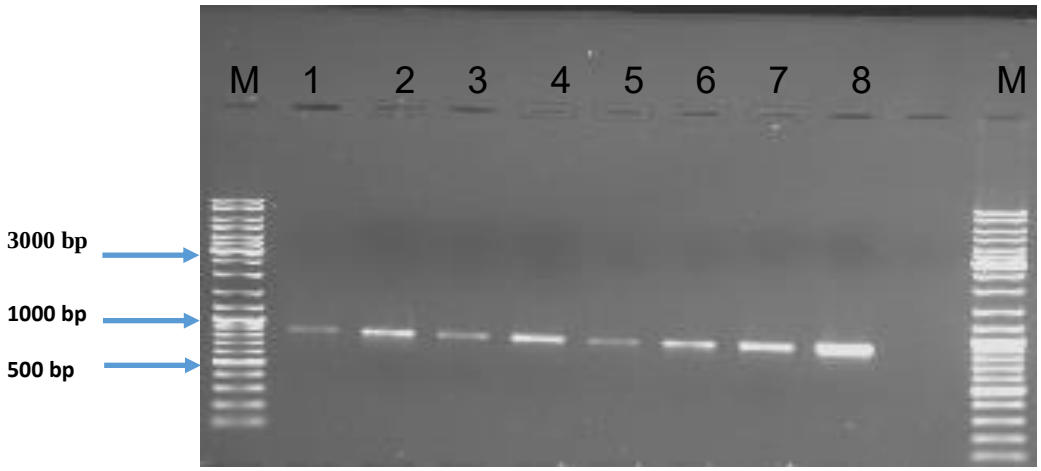


Figure 2.5: Representative gel showing the analysis of the purified PCR product for the VP4 encoding genome segment on a 1% agarose gel. Lane M is the GeneRuler DNA ladder mix molecular weight marker (Thermo-Scientific). Lanes 1-7 are purified PCR products for Mozambican samples 0208, 0441, 0146, 0051, 0428, and 0297. Lane 8 is the PCR purified product of the positive control G12P[6] human rotavirus sample.

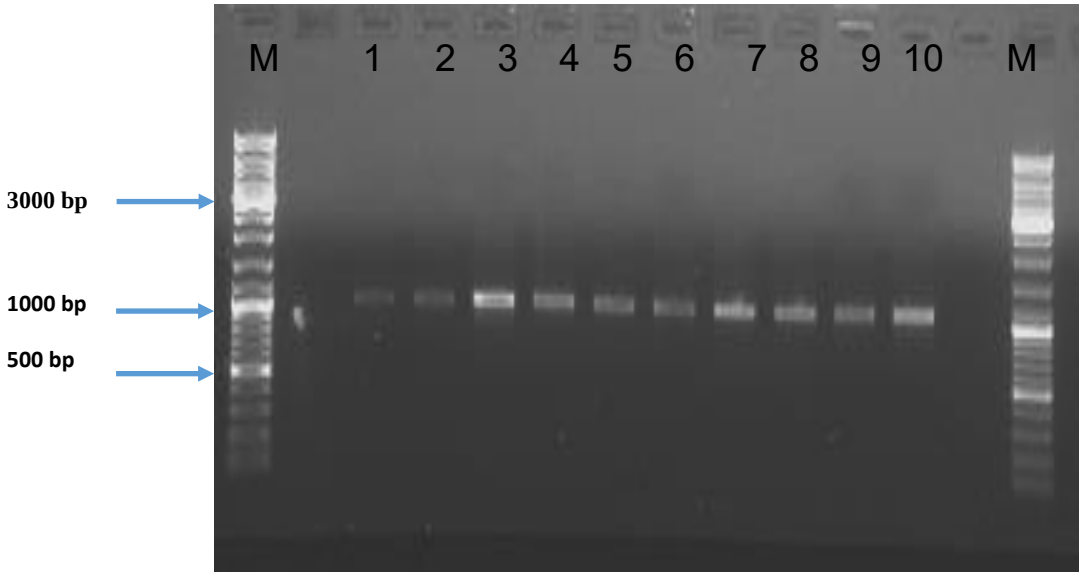


Figure 2.6: Representative gel showing the analysis of the purified PCR product for the VP7 encoding genome segment on a 1% agarose gel. Lane M is the GeneRuler DNA ladder mix molecular weight marker. Lanes 1-9 are purified PCR products for Mozambican samples 0208, 0153, 0289, 0038, 0428, 0297, 0441, 0051 and 0164. Lane 10 is the purified PCR product for the positive control G12P[6].

2.3.2 Sanger sequencing and BLASTn analysis

Results obtained from Sanger sequencing were submitted for BLASTn analysis to determine the genotypes of the Mozambican strains. Tables 2.1 and 2.2 indicate genotypes of strains from Mavalane and Manhiça obtained from Sanger sequencing and BLASTn analysis.

Table 2.1. Confirmation of G and P genotypes for Mavalane samples using Sanger sequencing and BLASTn analysis.

Sample ID/Study no	PCR genotyping previously performed	Year of collection	Genotypes obtained with BLASTn analysis
0208	G12P[6]	2012	Confirmed
0211	G12P[6]	2012	Confirmed
0223	GNTP[8]P[6]	2012	-ND
0251	NT	2012	-ND
0253	GNTP[4]P[6]	2012	-ND
0257	G12G8P[6]	2012	-ND
0277	G12P[6]	2012	Confirmed
0278	G12P[6]	2012	Confirmed
0285	G8P[4]	2012	Confirmed
0286	G12P[6]	2012	Confirmed
0287	G8P[4]	2012	Confirmed
0288	G12P[8]	2012	Confirmed
0289	G12P[6]	2012	Confirmed
0293	G8P[4]	2012	Confirmed
0297	G8P[4]	2012	Confirmed
0304	G12P[6]	2012	Confirmed
0308	G2P[4]	2012	Confirmed
0310	G8P[4]	2012	Confirmed
0311	G8P[8]	2012	Unable to extract RNA
0314	G2G9P[8]	2012	-ND
0349	G12P[8]	2012	Unable to extract RNA
0364	NT	2012	-ND
0412	NT	2013	G2P[4]
0428	G2P[4]	2013	Confirmed
0439	G2P[4]	2013	Confirmed
0440	G2P[4]	2013	Confirmed
0441	G2PNT	2013	-ND
0448	G2P[4]	2013	Confirmed

ND: not determined by means of Sanger sequencing and BLASTn analysis. Confirmed: Sanger sequencing and BLASTn analysis confirmed genotyping results obtained at CISM. NT: genotype of the particular sample could not be determined using genotyping PCR.

Table 2.2. Confirmation of G and P genotypes for Manhiça samples using Sanger sequencing and BLASTn analysis.

Sample ID	PCR genotyping previously performed	Year of collection	Genotypes obtained with BLASTn analysis
0002	NT	2012	Unable to extract RNA
0024	G9PNT	2012	Unable to extract RNA
0026	G9PNT	2012	Unable to extract RNA
0038	G12PNT	2012	-ND
0040	G12P[6]	2012	Confirmed
0042	GNTP[8]	2012	G12P[6]
0043	G2G8G9PNT	2012	-ND
0044	G8G12PNT	2012	-ND
0045	G8G12PNT	2012	-ND
0050	G12PNT	2012	G12P[6]
0051	G12P[6]	2012	Confirmed
0052	G8G12PNT	2012	ND
0060	G12PNT	2012	G12P[8]
0067	NT	2013	Unable to extract RNA
0101	NT	2013	Unable to extract RNA
0102	NT	2013	-ND
0106	NT	2013	Unable to extract RNA
0146	G2P[4]	2013	Confirmed
0113	GNTP[4]	2013	G2P[4]
0117	GNTP[4]	2013	G2P[4]
0125	GNTP[4]	2013	-ND
0126	GNTP[4]	2013	-ND
0131	G2P[4]	2013	Confirmed
0143	GNTP[4]	2013	-ND
0144	NT	2013	-ND
0151	GNTP[4]	2013	G2P[4]
0152	GNTP[4]	2013	-ND
0153	GNTP[4]	2013	G2P[4]
0154	GNTP[4]	2013	Unable to extract RNA
0156	NT	2013	Unable to extract RNA

ND: not determined by means of Sanger sequencing and BLASTn analysis. Confirmed: Sanger sequencing and BLASTn analysis confirmed the genotyping results obtained at CISM. NT: genotype of the particular sample could not be determined using genotyping PCR.

Sanger sequencing of the VP4 and VP7 encoding genome segments, as well as BLASTn analysis of the sequences, showed that the most common genotypes were G2 and G12 in combination with P[4] and P[6], respectively. A few other samples were typed as G8P[4] and G12P[8]. Of the samples that were successfully typed by means of Sanger sequencing, G2P[4] genotypes accounted for 44.4 %. Fifty percent of these were from Mavalane and the other 50% from Manhiça. G12P[6] accounted for 37% of the strains, 25.9% from Mavalane and 11.1% from Manhiça. G8P[4] accounted for only 11% of the samples and these strains were all collected from Mavalane. Finally, G12P[8] accounted for 7.4% of the total samples and half of the samples were from Mavalane and the other half from Manhiça.

In general, more samples from Mavalane were typed by BLASTn analysis as compared to samples from Manhiça as shown in Table 2.1 and 2.2. It was not possible to extract RNA from a large number of samples from Manhiça in comparison to samples from Mavalane. A possible reason could be due to low levels of RNA, such that it could not be detected, or even lack of RNA. It is also possible that degradation of RNA in some of the samples could have occurred either during collection or storage. Most samples from Manhiça were either untypeable or typed as mixed genotype by genotyping PCR and only a few of them were processed by RT-PCR and Sanger sequencing.

Sanger sequencing and BLASTn analysis of Mavalane samples collected during the first year of the surveillance, i.e 2012, indicated the presence of mostly genotype G12P[6] and G8P[4]. In the same area, in the second year of surveillance, only genotype G2P[4] was observed to be circulating. According to genotyping PCR results obtained by the Mozambican group, analysis of samples collected in Manhiça during the first year of surveillance were either typed as mixed, untypeable or partially typed meaning that only the P or G genotype could be identified. Sanger sequencing and BLASTn analysis of samples collected in Manhiça during the first year of the surveillance, indicated the presence of genotype G12P[6] and G12P[8] during the first year of surveillance. In the second year of surveillance in Manhiça, genotype G2P[4] was the only circulating genotype according to Sanger sequencing and BLASTn analysis. The fact that similar genotypes were circulating in these two areas during the same time is not surprising as people move around between the two places. Figure 2.7, indicates that more samples were collected from Mavalane as compared to Manhiça. Of the G genotype, G2 was a frequently detected genotype and equal number of strains were collected from the two regions. This was followed by G12 then G8. Of the P genotype, P[4] was detected most in the study followed by P[6] and P[8] with equal number of samples collected from the two regions.

G2P[4] genotype is among the common genotypes that circulate globally while G12P[6] and G8P[4] are genotypes that are mainly common in low income countries. The G12P[6] genotype is reported to be a recently emerging genotype with an unknown origin (Ghosh et al., 2006). The first G12 strain was isolated from stool sample of a child from India in 1987, in combination with P[6] genotype (Taniguchi et al., 1992). In Africa, the first G12 genotype was described in 2009 in South Africa (Page et al., 2009). After the description of the first G12 strain in 1987, it only re-emerged 10 years later. Since then it has been detected in higher numbers, especially in low income countries. Since the rotavirus surveillance project in Mozambique was only initiated in 2012, it is also not surprising that the majority of the genotypes observed include G12P[6].

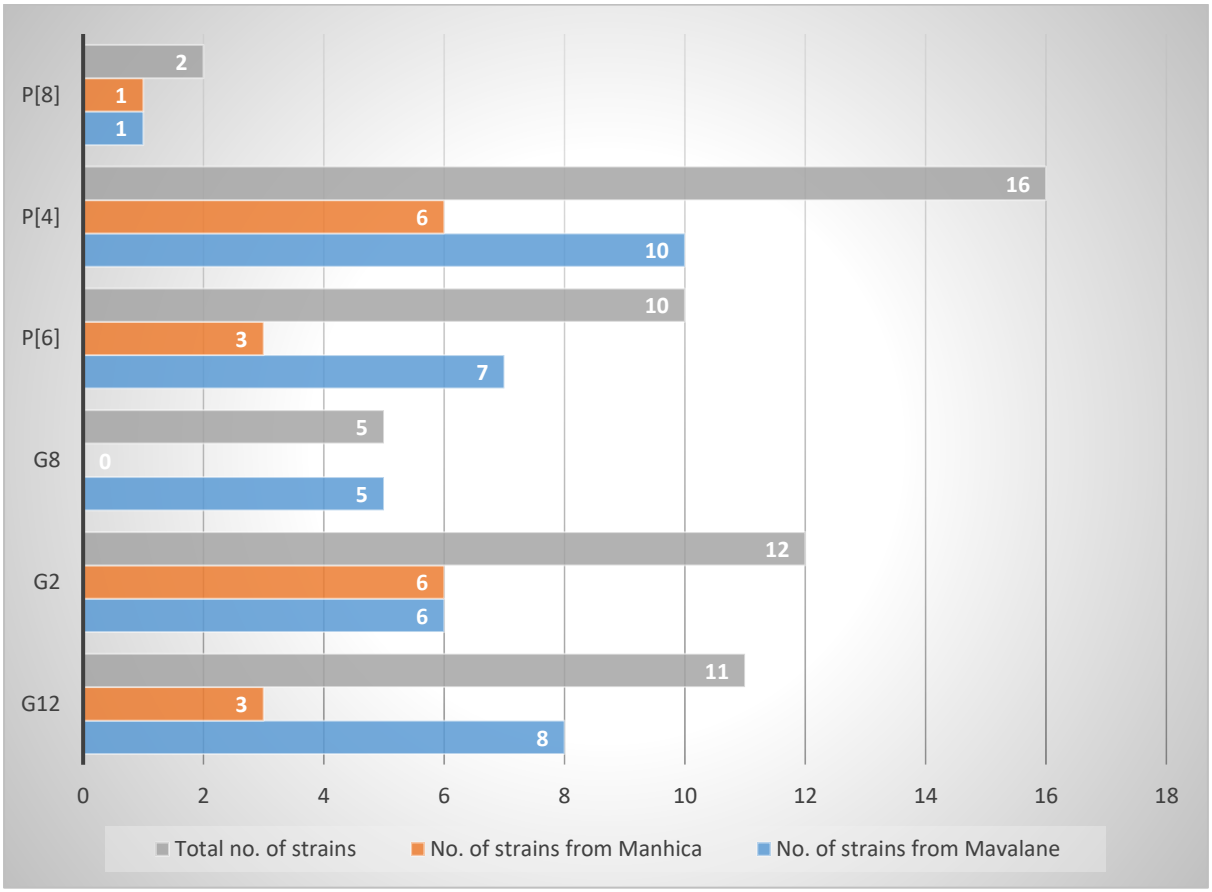


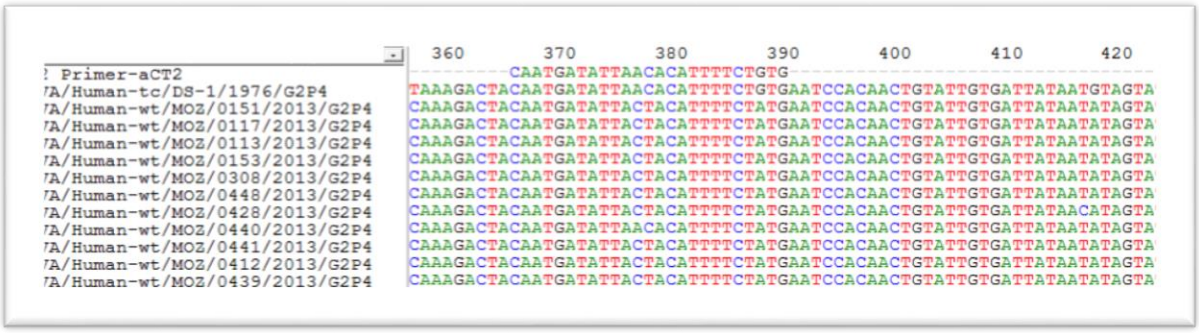
Figure 2.7: Summary of Mozambican strains genotype results generated by means of RT-PCR and Sanger sequencing. The number of strains in each genotype from Manhiça and Mavalane are indicated in orange and blue colours respectively.

2.3.3 Evaluation of genotyping primer binding sites

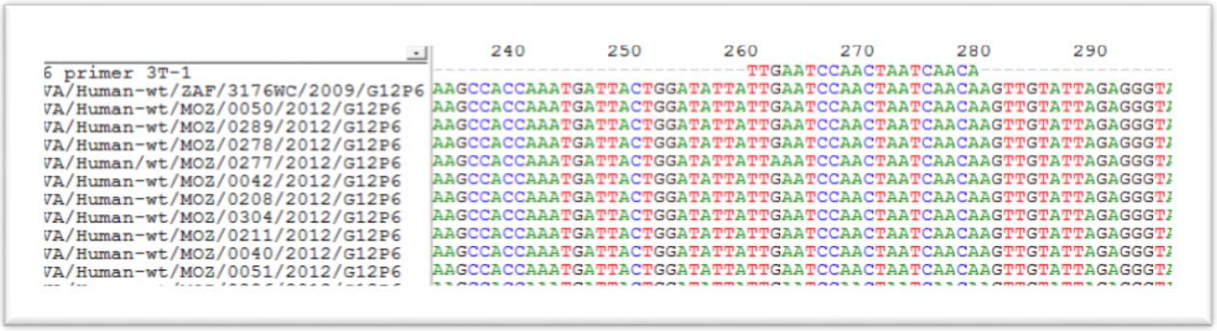
Genotyping primer binding sites were evaluated in an attempt to account for the possibility of why some of the primers failed to type the strains or only partially typed them while with RT-PCR and Sanger sequencing it was possible to determine their genotypes. To achieve this, multiple sequence alignments of different primer binding regions were evaluated. When determining the primer binding regions of strains which were partially typed such as RVA/Human-wt/MOZ/0050/2012/G12P[6] shown in Figure 2.8 (B) and RVA/Human-wt/MOZ/0113/2013/G2TP[4] shown in Figure 2.8 (A), the results obtained do not suggest any problems with primer binding sites.

Using genotyping primers, both the G and P genotypes of strain RVA/Human-wt/MOZ/0412/2013/GNTPNT Figure 2.8 (A) and (D) could not be typed but with Sanger sequencing and RT-PCR the strain was successfully typed as a G2P[4]. G2 genotyping primer Figure 2.8 (A) and P[4] genotyping primer Figure 2.8 (D) bind to the same region on strain RVA/Human-wt/MOZ/0412/2013/G2P[4] as the other G2P[4] strains that were successfully typed with genotyping primers. Strain RVA/Human-wt/MOZ/0042/2012/GNTP[6] (Figure 2.8 (C1)) was typed as a P[8] using genotyping primers but with Sanger sequencing and RT-PCR it was typed as a P[6]. When aligning P[8] primer 1T-1D with the sequence of strain RVA/Human-wt/MOZ/0042/2012/GNTP[6], the primer binds at nucleotide position 339 – 358 which is the region that would be expected and amplified for P[8] genotype which explains why the P[6] was typed as P[8]. The G[8] specific primer is a degenerate primer so maybe a more specific primer could be designed and used. The G genotype of the Mozambican strain RVA/Human-wt/MOZ/0412/2013/GNTP[4] could not be determined possibly due to poor quality of dsRNA or even due to insufficient amounts of RNA therefore it was not possible to evaluate why the P[6] specific genotype could not bind to and amplify the correct region. One of the reasons why the genotyping primers failed could be because of the poor quality of dsRNA used which were a result of the extraction process.

(A)


360 370 380 390 400 410 420
Primer-aCT2
/A/Human-tc/DS-1/1976/G2P4
/A/Human-wt/MOZ/0151/2013/G2P4
/A/Human-wt/MOZ/0117/2013/G2P4
/A/Human-wt/MOZ/0113/2013/G2P4
/A/Human-wt/MOZ/0153/2013/G2P4
/A/Human-wt/MOZ/0308/2013/G2P4
/A/Human-wt/MOZ/0448/2013/G2P4
/A/Human-wt/MOZ/0428/2013/G2P4
/A/Human-wt/MOZ/0440/2013/G2P4
/A/Human-wt/MOZ/0441/2013/G2P4
/A/Human-wt/MOZ/0412/2013/G2P4
/A/Human-wt/MOZ/0439/2013/G2P4

(B)


240 250 260 270 280 290
6 primer 3T-1
/A/Human-wt/ZAF/3176WC/2009/G12P6
/A/Human-wt/MOZ/0050/2012/G12P6
/A/Human-wt/MOZ/0289/2012/G12P6
/A/Human-wt/MOZ/0278/2012/G12P6
/A/Human/wt/MOZ/0277/2012/G12P6
/A/Human-wt/MOZ/0042/2012/G12P6
/A/Human-wt/MOZ/0208/2012/G12P6
/A/Human-wt/MOZ/0304/2012/G12P6
/A/Human-wt/MOZ/0211/2012/G12P6
/A/Human-wt/MOZ/0040/2012/G12P6
/A/Human-wt/MOZ/0051/2012/G12P6

(C1)


320 330 340 350 360 370
8 Primer 1T-1D
/A/Human-tc/USA/Wa/1974/G1P8
/A/Human-wt/MOZ/0042/2012/G12P6

(C2)


310 320 330 340 350 360 370
8 Primer 1T-1D
/A/Human-tc/USA/Wa/1976/G1P8
/A/Human-wt/MOZ/0208/2012/G12P6
/A/Human-wt/MOZ/0211/2012/G12P6
/A/Human-wt/MOZ/0278/2012/G12P6
/A/Human-wt/MOZ/0286/2012/G12P6
/A/Human-wt/MOZ/0289/2012/GXP6
/A/Human-wt/MOZ/0304/2012/G12P6

(D)

Figure 2.8 (A) Sequence alignment of G2 primer binding region and Mozambican G2 strains (0113, 0117, 0151 and 0153) that could not be typed by genotyping primers but were successfully typed as G2 by Sanger sequencing and BLASTn analysis. G2 strains that were correctly typed with both genotyping primers and Sanger sequencing were also included in the analysis. Strain RVA/Human-tc/USA/DS-1/1976/G2P[4] was used as a reference. **(B)** Sequence alignment of primer binding regions of P[6] primer and Mozambican P[6] sequences. Mozambican strain 0050 was typed as G12PNT with genotyping primers but was typed as G12P[6] with Sanger sequencing and BLASTn analysis while strain 0042 was typed as GNTP[8] with genotyping primers but was typed as G12P[6] with Sanger sequencing and BLASTn analysis. The other Mozambican P[6] strains included in the analysis were typed as P[6] according to genotyping primers and Sanger sequencing, Strain RVA/Human-wt/ZAF/3176WC/2009/G12P[6] was used as a reference. **(C1)** Sequence alignment of primer binding region of P[8] primer and a Mozambican P[6] strain that was incorrectly typed as P[8]. Strain RVA/Human-tc/USA/Wa/1976/G1P[8] was used as a reference. **(C2)** P[8] primer and Mozambican P[6] strains that were typed as P[6] with Sanger sequencing and genotyping primers. Strain RVA/Human-tc/USA/Wa/1976/G1P[8] was used as a reference. **(D)** Sequence alignment of primer binding region of P[4] primer and a Mozambican P[4] strain 0412 that could not be typed with genotyping primers but was typed as G2P[4] with Sanger sequencing and BLASTn analysis. The other Mozambican P[4] strains included in the analysis were typed as P[4] with both Sanger sequencing and genotyping primers, strain RVA/Human-wt/USA/DS-1/1976 was included in the analysis as a reference.

2.3.4 Determination of amino-acid differences in the VP4 and VP7 antigenic epitopes between Mozambican rotavirus strains and strains present in Rotarix and RotaTeq vaccines

Rotavirus vaccines provide protection against infections by inducing neutralising-antibody pressure. It is beneficial to have fewer amino acid differences in the VP4 and VP7 neutralising epitopes of human rotavirus strains and strains that make up the vaccines as this contributes to vaccine effectiveness. Determination of amino acid differences in the VP4 and VP7 antigenic epitopes between circulating human rotaviruses and vaccine viruses help to infer the affectivity of the vaccine.

Rotavirus VP7 protein consists of two antigenic epitopes, namely 7-1 and 7-2, with epitope 7-1 subdivided into 7-1a and 7-1b. The VP7 antigenic epitope is made up of 29 amino acids. Rotavirus VP4 has four antigenic epitopes labelled 8-1 to 8-4 in its VP8 component and five antigenic epitopes named 5-1 to 5-5 in its VP5 component. The VP4 antigenic epitope is made up of 37 amino acids. Results obtained for the determination of amino acid differences in the VP7 antigenic epitopes of Mozambican rotavirus strains and vaccine strains are shown in Table 2.3. Analysis of Mozambican G12 strains showed that their VP7 antigenic epitopes had up to fifteen amino acid differences with G1 component of RotaTeq for samples 0278 and 0211 and seventeen amino acid differences with G1 component of Rotarix for all three analysed G12 strains. Mozambican G12 strains were also compared to G2, G3, G4 and G6 components of RotaTeq and there were a few common amino acids between the other vaccine components and G12 strains. There were up to nine amino acids that were not present in any of the vaccine components and only present in G12 strains as indicated by yellow shading.

Analysis of Mozambican G8 sequences showed up to 17 amino acid differences with both RotaTeq and Rotarix. When G8 strains were compared to other components of RotaTeq, there were similar amino acids between Mozambican strains and vaccine virus strains indicated by orange shading for similarity with G6 vaccine component, blue shading for G3 components and green for the G2 component. In cases where Mozambican strains have amino acids that are similar to more than one vaccine component, the Mozambican strain was indicated in red shading in Table 2.3.

In the analysis of Mozambican G2 strains, the VP7 antigenic epitopes had at most five amino acid differences with the G2 component of RotaTeq, with the majority of the differences being in 7-1a and 7-1b epitopes. Mozambican G2 strains had 22 antigenic epitopes that were

completely conserved when compared to epitopes of the G2 component of RotaTeq. Analysis of G2 antigenic epitopes suggest that RotaTeq might be highly effective against Mozambican G2 rotavirus infections. Amino acids at positions; 98, 104, 291, 201 and 264 were completely conserved between the two vaccine rotaviruses and all Mozambican strains analysed for VP7 antigenic epitopes. For VP4 antigenic epitope analysis, amino acids at positions 100, 188, 190, 193, 180, 183 and 132, Table 2.4. There are also many amino acids for Mozambican strains that differ from those of the P[8] strains of vaccine as indicated by yellow shading.

Table 2.3: Alignment of VP7 antigenic residues between rotavirus vaccine strains and Mozambican strains. Amino acids that are different between RotaTeq and Rotarix are bolded. The yellow colour is for amino acids that are unique to the Mozambican strains and not present in any of the vaccine strains. The green colour is for Mozambican strains that have the same amino acids as G2 (SC2-9) component of RotaTeq. The blue colour is for strains with the same amino acids as G3 component of RotaTeq, while the purple and orange colours are used for Mozambican strains with similar amino acids to G4 and G6 RotaTeq strains respectively. The red colour is for strains whose amino acids are similar to amino acids found in more than one component of RotaTeq and lastly the grey colour is for strains that have similar amino acids to the G1 component of RotaTeq.

	7 – 1a														7 – 1b						7 - 2								
	87	91	94	96	97	98	99	100	104	123	125	129	130	291	201	211	212	213	238	242	143	145	146	147	148	190	217	221	264
RotaTeq G1	T	T	N	G	**D**	W	K	D	Q	S	V	V	D	K	Q	N	V	D	N	T	K	D	Q	**S**	L	S	M	N	G
Rotarix G1	T	T	N	G	**E**	W	K	D	Q	S	V	V	D	K	Q	N	V	D	N	T	K	D	Q	**N**	L	S	M	N	G
50 G12P[6]	S	T	T	P	D	W	T	N	Q	D	S	V	D	K	Q	D	V	T	N	N	Q	Q	N	S	L	S	E	A	G
278 G12P[6]	S	T	T	P	D	W	T	N	Q	D	S	V	D	K	Q	D	V	T	N	N	Q	Q	N	S	L	S	E	A	G
211 G12P[6]	S	T	T	P	D	W	T	N	Q	D	S	V	D	K	Q	D	V	T	N	N	Q	Q	N	S	L	S	E	A	G
297 G8P[4]	A	T	A	S	S	W	K	D	Q	D	A	I	N	K	Q	D	T	T	N	T	K	N	A	D	S	S	E	T	G
310 G8P[4]	A	T	A	S	S	W	K	D	Q	D	A	I	N	K	Q	D	T	T	N	T	K	N	A	D	S	S	E	T	G
G2P[5] (SC2-9) Of RotaTeq™	**A**	**N**	**S**	**D**	**E**	**W**	**E**	**N**	**Q**	**D**	**T**	**M**	**N**	**K**	**Q**	**D**	**V**	**S**	**N**	**S**	**R**	**D**	**N**	**T**	**S**	**D**	**I**	**S**	**G**
0113 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	I	N	K	Q	D	V	T	N	N	R	D	N	T	S	D	I	S	G
0131 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	M	N	K	Q	D	V	D	N	N	R	D	N	T	S	D	I	S	G
0153 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	M	N	K	Q	D	V	D	N	N	R	D	N	T	S	D	I	S	G
0412 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	M	N	K	Q	D	V	D	N	N	R	D	N	T	S	D	I	S	G
0428 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	M	N	K	Q	D	V	D	N	N	R	D	D	T	S	D	I	S	G
0440 G2P[4]	T	N	S	N	E	W	E	N	Q	D	N	M	N	K	Q	D	V	D	N	N	R	D	N	T	S	D	I	S	G
0448 G2P[4]	T	N	S	N	E	W	E	N	Q	D	T	M	N	K	Q	D	V	D	N	N	R	D	N	T	S	D	I	S	G
G3P[5]	T	T	N	N	S	W	K	D	Q	D	A	V	D	K	Q	D	A	N	K	D	K	D	A	T	L	S	E	A	G
G4P[5]	T	T	S	T	E	W	K	D	Q	N	L	I	D	K	Q	D	T	A	D	T	R	A	S	G	E	S	T	S	G
G6P[8]	V	N	A	T	E	W	K	D	Q	D	A	V	E	K	Q	N	P	D	N	A	K	D	S	T	Q	S	T	T	G

Table 2.4: Alignment of antigenic residues in VP4 between Mozambican strains and rotavirus vaccine strains. Strains with amino acids different from both vaccine strains are indicated in yellow. Mozambican strains with amino acids similar to Rotarix but not RotaTeq are shown in green. Strains with amino acids similar to RotaTeq but not Rotarix are shown grey colouring.

13	8-1											8-2		8-3								8-4			
	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
Rotateq	D	S	Q	E	S	T	N	L	N	**D**	I	T	A	N	P	V	D	**N**	**R**	N	D	**D**	N	T	N
Rotarix	D	S	Q	E	S	T	N	L	N	**N**	I	T	A	N	P	V	D	**S**	**S**	N	D	**N**	N	T	N
42	D	N	S	E	S	T	N	L	S	E	V	T	A	T	N	Q	S	T	E	N	N	D	T	N	Q
60	D	S	Q	D	S	T	D	L	N	G	I	T	A	N	P	V	D	N	R	N	D	D	N	T	N
113	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
117	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
131	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
146	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
151	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
153	D	S	Q	D	S	T	N	L	N	N	I	T	A	S	P	T	N	N	E	N	S	N	N	T	D
208	D	N	S	E	S	T	N	L	S	E	V	T	A	T	N	Q	S	T	E	N	N	N	T	N	Q
211	D	N	S	E	S	T	D	L	S	E	V	T	A	T	N	Q	S	T	E	N	N	D	T	N	Q
285	D	S	Q	E	S	T	N	L	N	N	I	T	A	S	Q	I	N	N	E	N	N	N	N	T	N
286	D	N	S	E	S	T	D	L	S	E	V	T	A	T	N	Q	S	T	E	N	N	D	T	N	Q
288	D	S	Q	D	S	T	D	L	N	G	I	T	A	N	P	V	D	N	R	N	D	D	N	T	N
297	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	I	N	N	E	N	N	D	N	T	N
308	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
310	D	S	Q	E	S	T	D	L	N	N	I	T	A	S	Q	I	N	N	E	N	N	D	N	T	N
440	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
441	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
448	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

2.3.3 Phylogenetic analysis of genome segments encoding VP4 and VP7 of Mozambican rotavirus strains

Following the confirmation and assignment of genotypes using BLASTn, the nucleotide sequences determined with Sanger sequencing were used to infer phylogenetic relationships to provide insight into the molecular epidemiology of Mozambican rotavirus. Strains are labelled according to the recommendation of the rotavirus classification working group (RCWG) (Matthijnssens et al., 2011).

2.3.3.1 P[4] genotype

In the P[4] phylogenetic tree, Mozambican strains belonged to two separate lineages, lineage I (Nakagomi et al., 2013) and lineage V (Chaimongkol et al., 2012, Gómez et al., 2014, Figure 2.9. These groups are supported by 100% and 99% bootstrap values, respectively, which is an indication of a high confidence of their grouping. Mozambican strains in lineage I were strains collected from Mavalane, between July and August in 2012. Most of the Mozambican P[4] strains in lineage I were in combination with the G8 genotype with only one sample in combination with G2 genotype. Lineage I exclusively comprises strains collected from countries in Sub-Saharan Africa such as South Africa, Zambia, Malawi, Zimbabwe, Uganda, and Kenya. Mavalane strains in lineage I grouped closely with South African strains RVA/Human-wt/ZAF/MRC-DPRU82/2012/G2P[4] and RVA/Human-wt/ZAF/MRC-DPRU1533/2007/G2P[4].

In contrast to strains in lineage I, strains in lineage V were collected from countries within Sub-Saharan Africa and other countries such as, India, the Philippines, Australia, and Russia (Figure 2.8). Six of ten Mozambican strains in lineage V were collected from Manhiça while the remaining four samples were from Mavalane. All Mozambican P[4] strains in this lineage are in association with G2 genotype. In addition, these Mozambican strains were collected between 2012 and 2013. Seven Mozambican strains in lineage V, grouped closely with each other and with one Zimbabwean strain RVA/human-wt/ZWE/MRC-DPRU1132/XXXX/G2P[4]. These strains also clustered with six Australian strains namely- RVA/Human-wt/Aus/MON016/2010/G2P[4], RVA/Human-wt/Aus/WAPC681/2010/G2P[4], RVA/Human-wt/Aus/WAPC686/2010/G2P[4], RVA/Human-wt/Aus/MON016/2010/G2P[4], RVA/Human-wt/Aus/WAPC703/2010/G2P[4]. All these Australian strains were collected in 2010 after the introduction of rotavirus vaccines and the emergence of G2P[4] genotype in

Australian states. Before the introduction of vaccines in Australia, G2P[4] genotype were detected at a lower percentage (Donato et al., 2014).

In addition in lineage V, Mozambican rotavirus strains RVA/Human-wt//0151/G2P[4] and RVA/Humna-wt/MOZ/0428/2013/G2P[4], grouped closely with a strain from Canada RVA/Human-wt/CAN/25/2011/G2P[4] and a strain from India, RVA/Human-wt/IND/NIVI12445/2012G2P[4]. The Indian strain was detected in a study investigating group A rotavirus strains causing gastroenteritis in adolescents and adults between 2008 and 2012. In an earlier study in 2004-2007, the authors reported an increased detection of group A rotaviruses in adolescents and adults compared to the detection of rotaviruses in 1993-1996. In the 2008-2012 study, rotavirus detection in adolescents and adults was reported to have increased as compared to the 2004 to 2007 study and therefore the authors attempted to understand the temporal variation of rotavirus infections observed (Tatte et al., 2014).

Mozambican rotavirus strain RVA/Human-wt/MOZ/0153/2013/G2P[4] from Mavalane in lineage V, grouped closely with two Indian stains namely RVA/Human-wt/IND/MSC65/2011/G8P[4] and RVA/Human-wt/IND/KOL10109/2009/G2P[4] and three Brazilian stains namely; RVA/Human-wt/BRA/MG19041/2010/G2P[4], RVA/Human-wt/BRA/RJI8346/2010/G2P[4] and RVA/Human-wt/BRA/19112/2010/G2P[4]. Similarities between strains circulating in Brazil and Mozambique would be expected as both are Portuguese speaking countries, and frequent movement of people between the two countries is expected. The three Brazilian strains were detected in a study where the detection of G2P[4] genotype increased after the introduction of Rotarix in 2006 in Brazil (Gomez et al., 2014). The Indian strain RVA/Human-wt/IND/Msc65/2010/G8P[4] was among the first G8P[4] strains reported in India. Additionally, the same Indian stain has been reported to be the result of animal-human zoonotic transmission. These results are not surprising as in some parts of India, people live in close proximity to animals.

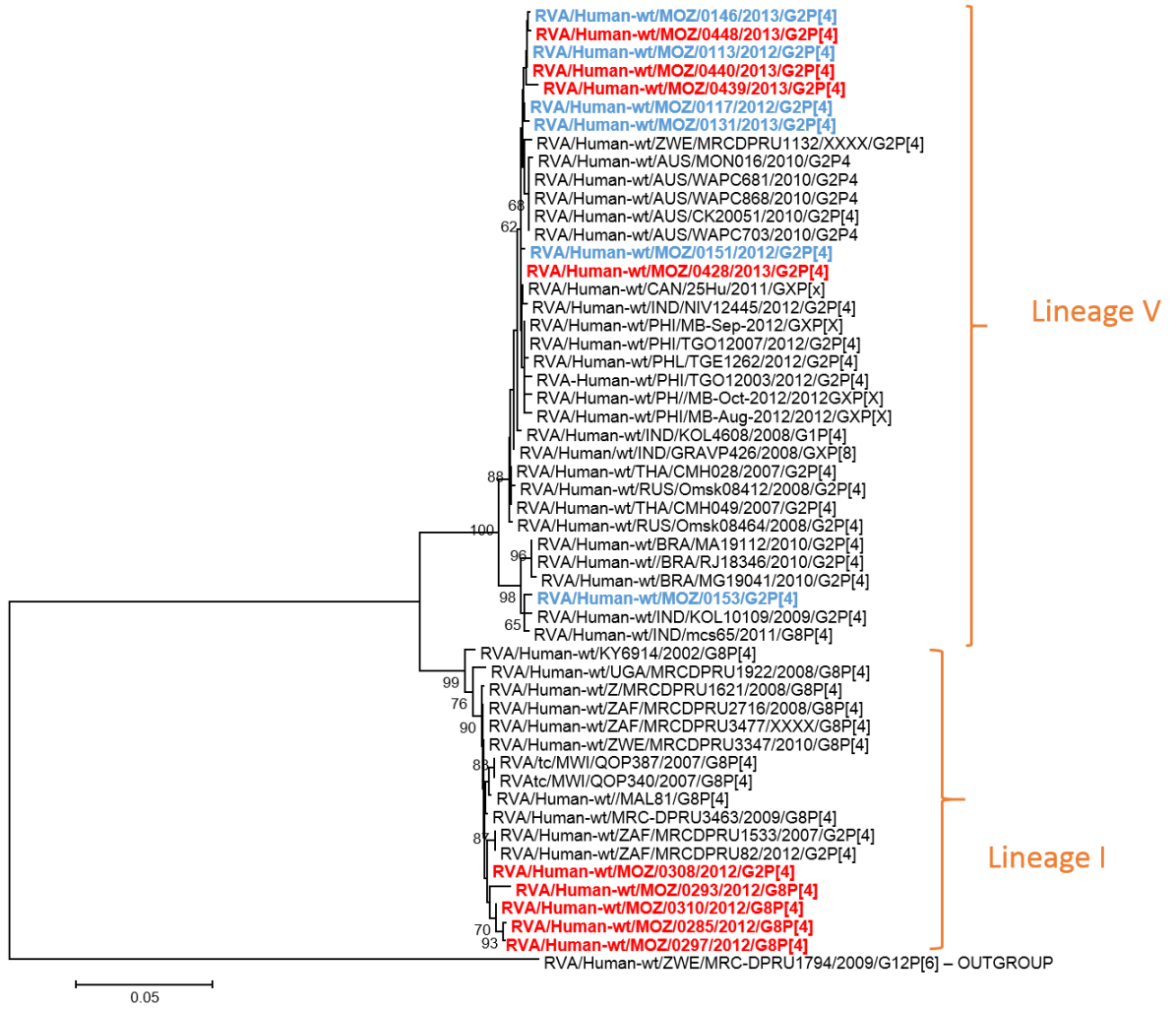


Figure 2.9: Phylogenetic analysis of the P[4] genotype of Mozambican strains. The tree was constructed using MEGA5.2 and the evolutionary history inferred using the neighbour-joining method. The evolutionary distances were computed using the Kimura-2 parameter method and are in the units of the number of base substitutions per site. Only bootstrap values higher than 60% are shown on the appropriate branches. The scale bar represents the genetic distances as formulated by MEGA 5.2. Strain RVA/Human-wt/ZWE/MRC-DPRU1794/20009/G12P[6] was used as an out-group. Strains originating from Mavalane are highlighted in red while Manhiça strains are highlighted in blue.

2.3.3.2 P[6] genotype

Mozambican strains in the P[6] phylogenetic tree belonged to lineage I as shown in Figure 2.10 (Ide et al., 2015). All strains in the P[6] phylogenetic tree were collected from low income countries in Asia with a few strains from South Africa. In Asia, the majority of P[6] strains were collected in India and Bangladesh. All Mozambican P[6] strains were in combination with G12 genotype, with 7 samples from Mavalane and 4 from Manhiça. Mozambican strain RVA/Human-wt/MOZ/0289/G12P[6] grouped closely with one South African strain RVA/Human-wt/ZAF/MRCD-PRU4090/2011/G12P[6] demonstrating 100% nucleotide identity between the two stains. However, there is no published information regarding this South African strain.

Mozambican strains also clustered with Indian strains RVA/Human-wt/IND/UKHLD/2011/G12P[6] and RVA/Human-wt/IND/KOL2509/2009/G12P[6]. The latter strain was detected in a community based study of rotaviruses circulating in India before the introduction of vaccines in the national immunisation programme (Mullick et al., 2014,)**.** Indian strains from the same study also grouped with Mozambican G12 strains in the G12 phylogenetic tree (Figure 2.11). Mozambican strain RVA/Human-wt/MOZ/0304/2012/G12P[6] did not group with other Mozambican strains but instead grouped with an Indian strain RVA/Human-tc/IND/MMRA23/2011/G12P[6] that was detected during whole genome analysis of human G12P[6] and G12P[8] strains emerging in Myanmar, India (Ide et al., 2015). All Mozambican strains in the P[6] phylogenetic tree were collected in 2012 and strains that clustered closely with Mozambican strains were collected around the same time as the Mozambican strains. This implies that similar strains were circulating globally during the same time period.

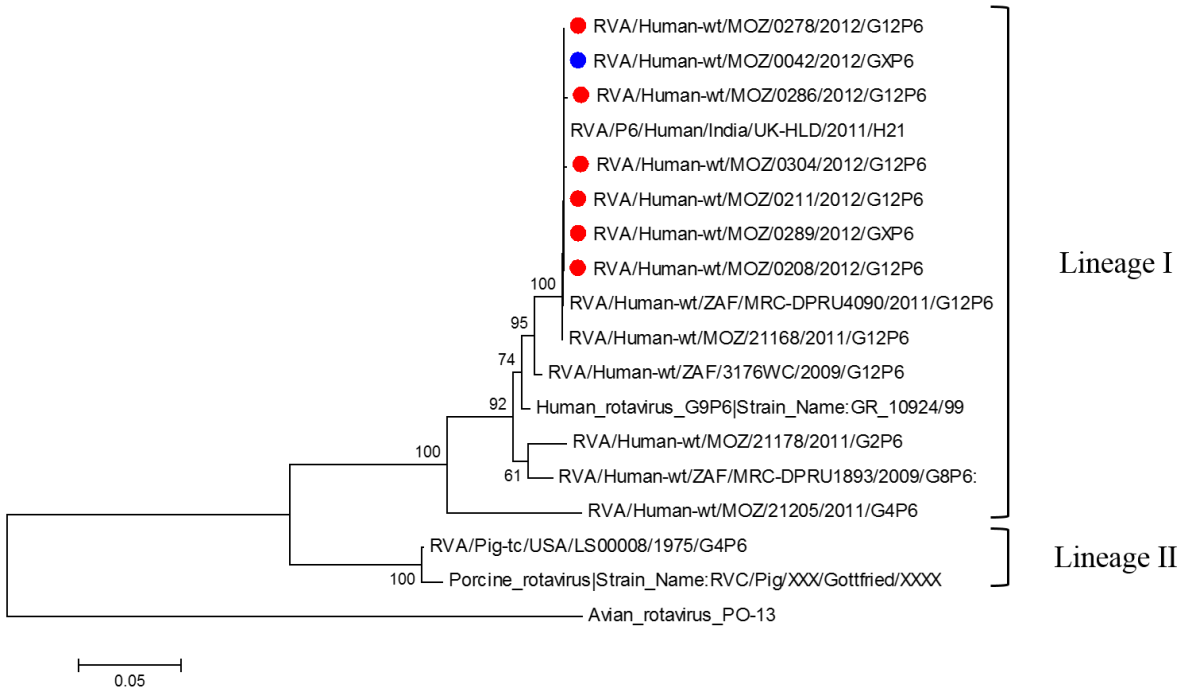


Figure 2.10: Phylogenetic analysis of the P[6] genotype of Mozambican strains. The tree was constructed using MEGA5.2 and the evolutionary history inferred using the neighbour-joining method. The evolutionary distances were computed using the Kimura-2 parameter method and are in the units of the number of base substitutions per site. Only bootstrap values higher than 60% are shown on the appropriate branches. The scale bar represents the genetic distances as formulated by MEGA 5.2. Strain Aviant/PO-13 was used as an out-group to root the tree. Strains originating from Mavalane are highlighted in red while Manhiça strains are highlighted in blue.

2.3.3.3 P[8] genotype

In this study, only two Mozambican strains with P[8] genotype were detected, RVA/Human-wt/MOZ/0060/2012/G12P[8] from Manhiça and RVA/Human-wt/MOZ/0288/2012/G12P[8] from Mavalane. The two strains belonged to lineage III, Figure 2.11. Mozambican strains grouped closely and with two other Mozambican strains described by Langa et al., 2016, RVA/Human-wt/MOZ/21135/2011/G1P[8] and RVA/Human-wt/MOZ/21195/2011/G12P[8] and also with a strain from Uganda RVA/Human-wt/UGA/MRC-DPRU4620/2011/G12P[8]. In the same lineage Mozambican strains also clustered with strains from countries such as South Africa, Brazil, and Argentina which were in combination with G1, G9 and G12 genotypes. The majority of the strains that clustered with Mozambican strains were collected

around the same time period as the Mozambican strains, which implies that similar genotypes were circulating around the same time period. It is expected that Mozambican strains described in the current study and those described by Langa and co-workers in 2016 which were collected from Chokwe district, Gaza province and Southern Mozambique would group together as people living in these areas are expected to also travel to Mavalane and Manhiça. More Mozambican P[8] strains are needed in other to make an accurate conclusion regarding P[8] strains circulating in Mozambique.

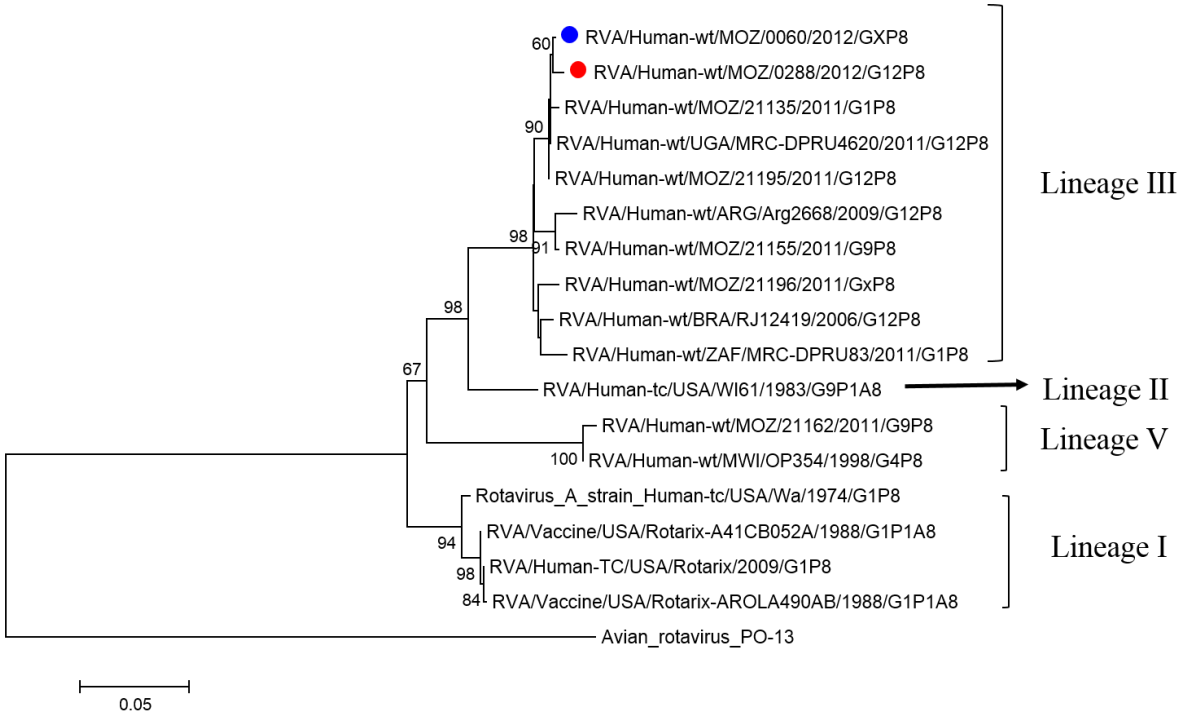


Figure 2.11: Phylogenetic analysis of the P[8] genotype of Mozambican strains. The tree was constructed using MEGA5.2 and the evolutionary history inferred using the neighbour-joining method. The evolutionary distances were computed using the Kimura-2 parameter method and are in the units of the number of base substitutions per site. Only bootstrap values higher than 60% are shown on the appropriate branches. The scale bar represents the genetic distances as formulated by MEGA 5.2. Strain Avian/rotavirus/PO-13 was used as an out-group to root the tree.

2.3.3.4 G12 genotype

In Figure 2.12, all Mozambican strains belonged to lineage G12-III (Ide et al., 2015, Mullick et al., 2014, Zeller et al., 2012). Mozambican G12 strains grouped into two main clusters, the first cluster has the majority of Mozambican G12 strains in combination with P[6] genotype and the second cluster has two Mozambican strains namely RVA/Human-wt/MOZ/0288/2012/G12P[8] and RVA/Human-wt/MOZ/0060/2012/G12P[8]. Mozambican G12 strains in combination with P[6] genotype grouped closely with other G12 strains from low income countries such as India, South Africa, and Mauritius. Mozambican strain RVA/Human-wt/MOZ/0277/2012/G12P[6] grouped closely with an Indian strain RVA/Human-wt/IND/KOL2409/2009/G12P[6]. The Indian strain RVA/Human-wt/KOL2409/2009/G12P[6] was detected in a community based rotavirus surveillance study in an urban slum in Kolkata, India. This study was conducted in 2008 to 2010 before the introduction of rotavirus vaccines in the national immunisation programmes in India (Mullick et al., 2014,). Mozambican strain RVA/Human-wt/MOZ/0278/2012/G12P[6] also grouped closely with an Indian strain RVA/Human-wt/IND/KOL-2-10/2010/G12P[8].

Mozambican G12 strains which are in combination with P[8] genotype, grouped closely with three strains collected from Mozambique which were described by Langa and co-workers (2016) and with one strain collected from Uganda. In the G12 phylogenetic tree, strains from high income countries such as Germany and Brazil did not group very closely with Mozambican strains which might suggest differences between G12 strains circulating in Mozambique and those circulating in high income countries. Mozambican G12 strains circulating in Mavalane and in Manhica do not appear to be very diverse from each other as they grouped closely with each other in the G12P[6] cluster and also in the G12P[8] cluster.

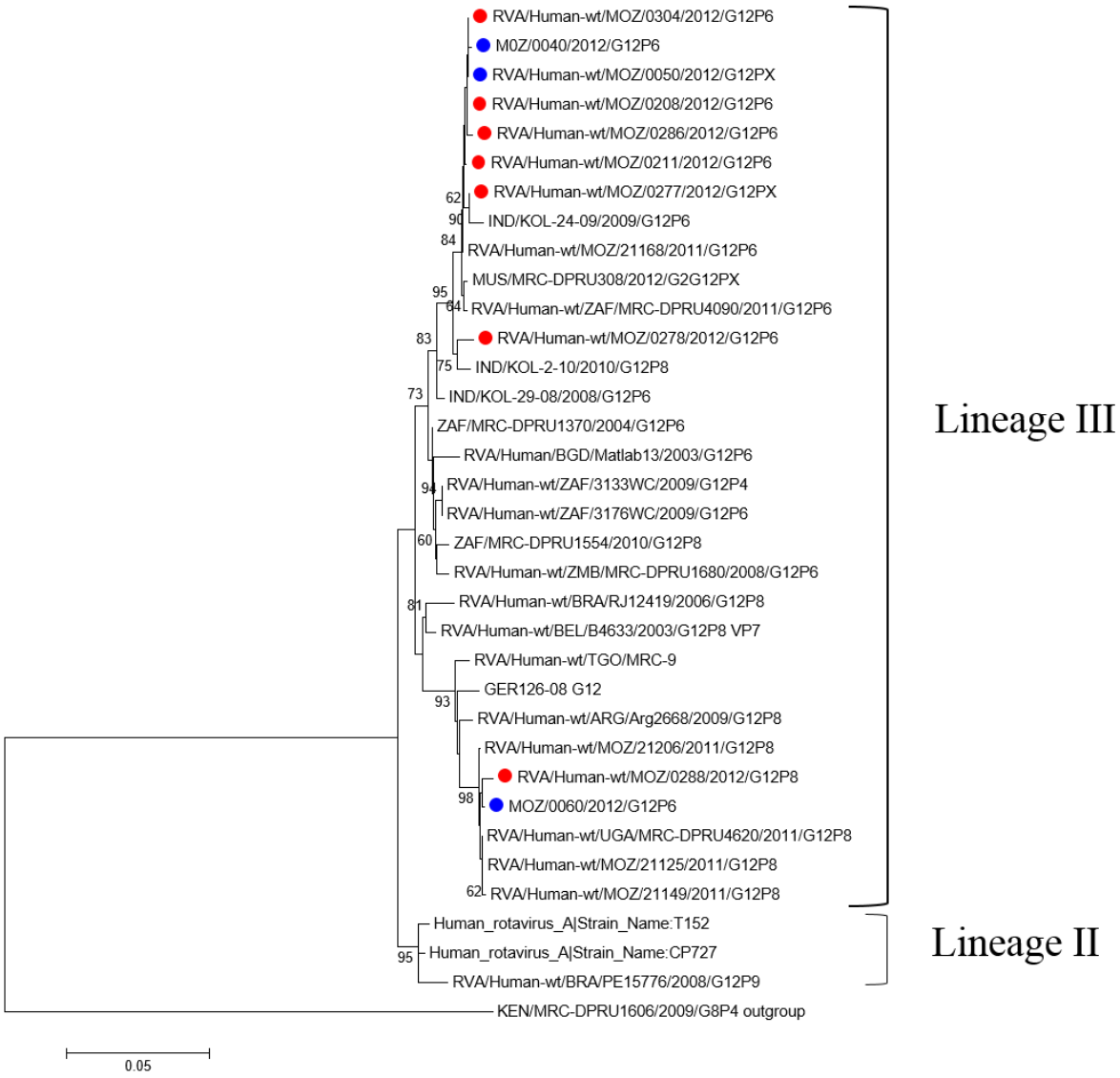


Figure 2.12: Phylogenetic analysis of the G12 genotype of Mozambican rotavirus strains
The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch length = 0.46301151 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura 3-parameter method [3] and are in the units of the number of base substitutions per site. The analysis involved 35 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair. There were a total of 1074 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [4]. Strain RVA/KEN/MRC-DPRU1606/2009//G8P[4] was uses as an outgroup.

2.3.3.5 G2 genotype

All Mozambican G2 strains characterised in this study belong to lineage II (Mullick, et al., 2014a). Most strains in the G2 phylogenetic tree were collected from countries outside Africa, with one strain from an African country. G2 phylogenetic analysis showed that Mozambican strains clustered with one Zimbabwean strain RVA/Human-wt/ZWE/MRCDPRU1132/XXXX/G2P[4] that was isolated from an eight month old female infant (Figure 2.13). However, there is no further published information available regarding this strain. Most Mozambican strains in the G2 tree clustered together with the exception of strains RVA/Human-wt/MOZ/O448/2013/G2P[4] and RVA/Human-wt/MOZ/0308/2012/G2P[4], which grouped together.

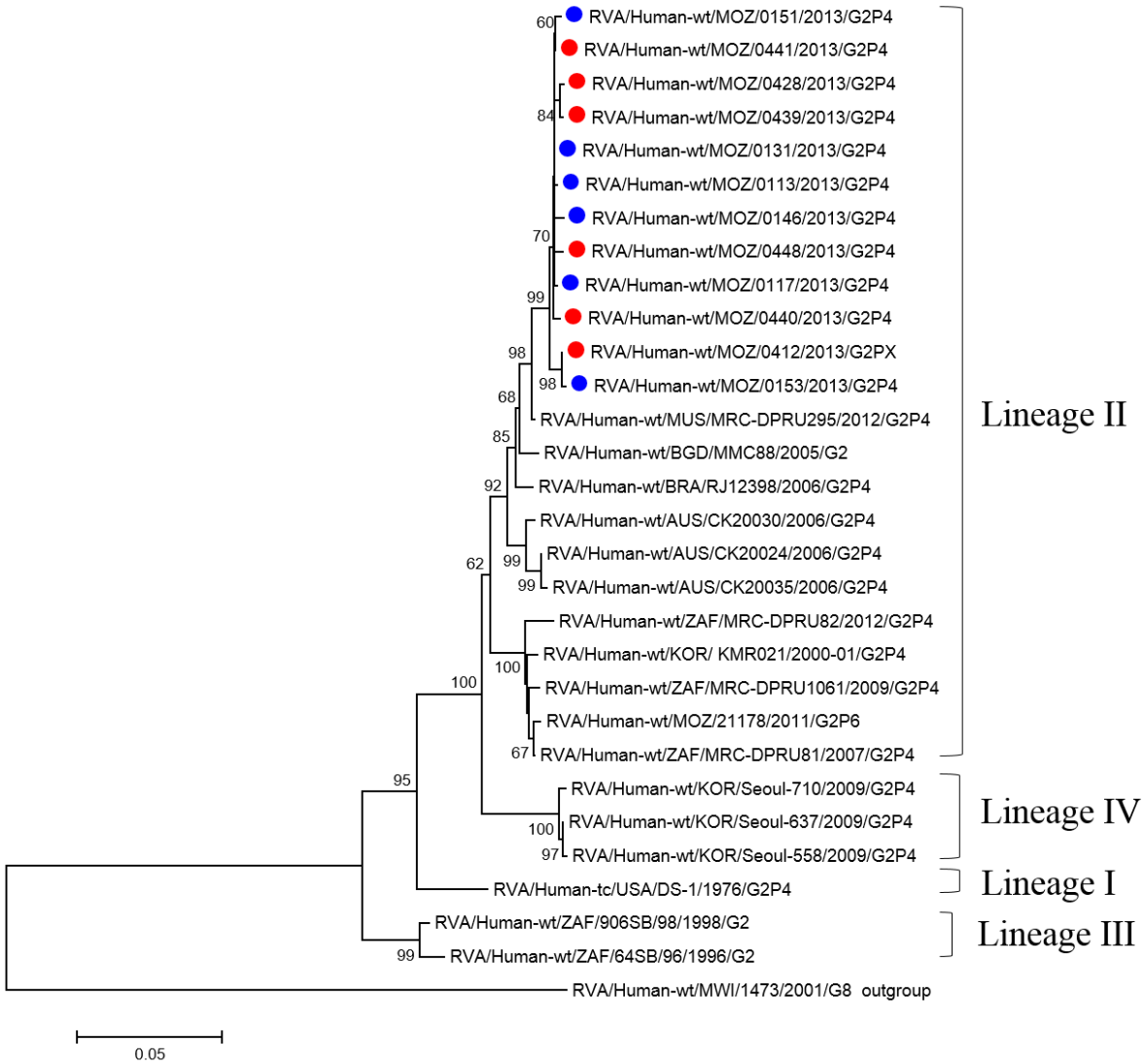


Figure 2.13: Phylogenetic analysis of the G2 genotype of Mozambican strains. The tree was constructed using MEGA5.2 and the evolutionary history inferred using the neighbour-

joining method. The evolutionary distances were computed using the Kimura-2 parameter method and are in the units of the number of base substitutions per site. Only bootstrap values higher than 60% are shown on the appropriate branches. The scale bar represents the genetic distances as formulated by MEGA 5.2. Strain RVA/Human-wt/MWI/MAL81/G8P[4] was used as an out-group.

2.3.3.6 G8 genotype

Mozambican G8 strains detected in this study belonged to lineage II (Figure 2.14). Only three G8 strains were detected in the current study. The three strains were all collected from Mavalane and they were in association with P[4] genotype. Mozambican G8 strains grouped closely with each other and with strains collected from low-income countries in Sub-Saharan Africa such as Zambia, Kenya, Uganda and Tanzania. Mozambican G8 rotavirus strains grouped distantly with a Nigerian bovine G8 rotavirus strain and a Nigerian human derived rotavirus strain RVA/Human-tc/NGA/HMG035/1999/G8P[1]. This observation might imply that Mozambican G8 strains are not a result of direct interspecies transmission between a human and a bovine strain but a result of human to human transmission. Rotavirus G8 genotypes are commonly associated with bovine strains but they also occur in humans with most strains described in low-income countries in Africa. In most cases where G8 genotype has been described, it is normally detected at lower rates and rarely the predominant genotype (Adah & Wade, 2001, Armah et al., 2001). In the current study, G8 genotype was also detected at a low rate and the same observation was the case for Mozambican rotavirus strains described by Langa and co-workers (2016). Genotype G8 phylogenetic analysis of Mozambican strains in relation to global G8 strains showed that G8 strains circulating in Mozambique are similar to other G8 strains circulating in low-income countries in sub-Saharan Africa.

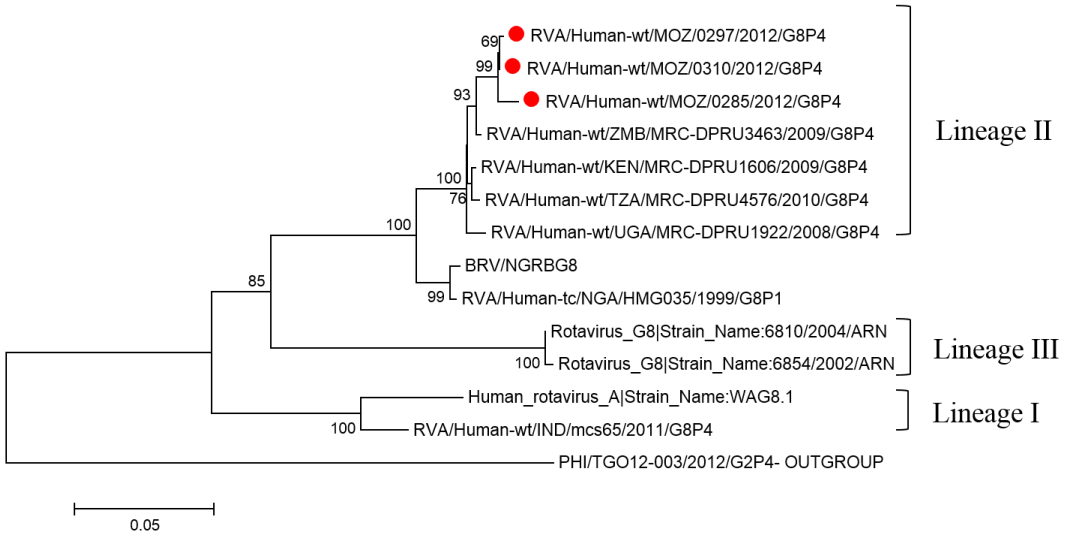


Figure 2.14: Phylogenetic analysis of the G8 genotype of Mozambican rotavirus strains. The tree was constructed using MEGA5.2 and the evolutionary history inferred using the neighbour-joining method. The evolutionary distances were computed using the Kimura-2 parameter method and are in the units of the number of base substitutions per site. Only bootstrap values higher than 60% are shown on the appropriate branches. The scale bar represents the genetic distances as formulated by MEGA 5.2. Strain RVA/Human-wt/PHI/TGO12-003/2012/G2P[4] was used as an out-group.

2.4 Conclusion

In conclusion, the aims and objectives of the work done in this chapter were achieved. Analyses of the G and P types for rotavirus strains originating from a rural (Manhiça) and urban (Mavalane) suggested that similar strains were circulating in the two regions. It is important to note that fewer samples could be analysed for the rural site than the urban site (Tables 2.1 & 2.2). The various P and G phylogenetic analysis of Mozambican strains in relation to other globally circulating strains indicated that Mozambican strains are mostly similar to strains circulating in developing countries in sub-Saharan Africa and in Asia. A few of the Mozambican strains also grouped with strains circulating in countries outside of Africa such as Brazil, the USA and Australia. In this study, G12P[6] and G2P[4] genotypes were detected at higher levels as compared to G8P[4] and G12P[8]. Genotype G2P[4] is among the commonly circulating genotypes worldwide therefore its detection in Mozambique can also be expected (Estes and Greenberg, 2013, Santos and Hoshino, 2005). Genotypes G12P[6] and G12P[8] are among the commonly detected genotypes in low income countries and similarly their detection in Mozambique is also expected (Seheri et al., 2014). In most phylogenetic trees for example

G12, P[4] and P[6] most Mozambican strains grouped closely with each other, indicating high similarly between the strains.

It is expected that strains circulating in Mozambique would be similar to strains circulating in neighbouring countries such as South Africa, Zimbabwe, Malawi, Zambia etc. as indicated by the results obtained. This observation is expected as there is movement of people between these countries. People from Mozambique travel to South Africa to seek employment or for holiday purposes and vice versa. The rotavirus vaccine Rotarix®, which was introduced in South Africa in 2009 has proven to be effective in lowering rotavirus cases by about 50% despite it being only about 61% efficacious (Madhi et al., 2010). This implies that if similar strains are circulating in both Mozambique and South Africa, then the same rotavirus vaccine used in South Africa would be expected to be as effective when introduced in Mozambique. In addition, the effectiveness of rotavirus vaccines in lowering the disease burden in other African countries such as Kenya, Malawi, Mali and Ghana have also been reported, this further confirms that if similar strains are circulating in some of these countries and in Mozambique, then rotavirus vaccines can be expected to be effective when introduced in Mozambique (Armah et al., 2010).

Chapter 3

Whole Genome Characterisation of Mozambican rotavirus strains using next generation sequencing

3.1 Introduction

The late 1970s marked the beginning of an era for research, not only in virology, but in various other fields as well. This was made possible by the development of DNA sequencing technologies by Sanger and colleagues (Sanger et al., 1977) and by Maxim and Gilbert (Maxim and Gilbert, 1977). Maxim and Gilbert developed a DNA sequencing technology in which the nucleotide sequence of a piece of DNA are determined by chemically fragmenting DNA that has been labelled with a phosphate at either the 3' end or the 5'end (Maxam and Gilbert, 1977). Sanger and co-workers' used a chain termination method today known as Sanger sequencing. The Sanger sequencing method is preferred over the Maxim and Gilbert's method because the latter involves the use of toxic chemicals and isotopes (Van Dijk et al., 2014).

Recently, improved methods for sequencing DNA/RNA have been introduced and are referred to as second generation sequencing technologies or next generation sequencing (NGS) technologies. Next generation sequencing technologies were developed after the completion of the human genome project using Sanger sequencing (Lander et al., 2001). The process took a long time and was expensive. Therefore, alternative sequencing technologies that were faster, cheaper and yielded high throughput data, were developed (Van Dijk et al., 2014).

There are various NGS platforms available for use. The choice of platform depends on various factors including, the kind of question the study aims to answer, the size of the genome to be sequenced, the level of accuracy required and the depth of coverage required (Radford et al., 2012). Examples of NGS platforms include 454 pyrosequencing (Roche Diagnostics), Illumina sequencing (Illumina), Ion torrent (Life Technologies), Solid Sequencing (Life Technologies) and PacBio sequencing (Pacific Biosciences) to mention a few. In this chapter, only a few NGS platforms will be discussed, specifically those that have been used to study rotaviruses.

For all NGS methods, the DNA must be suitable for sequencing. The first step is library preparation, and is achieved by fragmenting DNA molecules of interest into smaller fragments that are suitable for sequencing on the NGS platform. The smaller DNA fragments are blunt ended and specific adapters are ligated to the blunt ends. Adapter ligation can also be achieved

by A/T overhangs. For most technologies after adapter ligation and clean-up, the library is ready for sequencing. Some NGS methods may require additional steps prior to sequencing reactions.

Life Sciences introduced the first NGS technology (454 pyrosequencing) in 2005. The methods were developed by Jonathan Rothberg and the technology was later bought by Roche Diagnostics (Van Dijk et al., 2014). It is based on a sequencing by synthesis principle. In this technology, after library preparation, the fragmented DNA molecules are individually attached to small beads. Emulsion PCR amplification is then performed where primers, PCR reagents and emulsion oil are added to the DNA-attached-beads and multiple copies of the same DNA sequence are generated on each bead.

PCR amplification is followed by sequencing where PCR amplicons are firstly denatured into ssDNA molecules and bound to sequencing primers. The ssDNA molecules are loaded into wells of a picotiter plate and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase, apyrase and the substrates adenosine 5' phosphosulfate (APS) and luciferin. The fragments are sequenced by adding a single nucleotide at a time and if it is complementary it is bound to the DNA strand by DNA polymerase releasing pyrophosphate and hydrogen. If the added nucleotide is not complementary, it is degraded by apyrase. The technology exploits the released pyrophosphate which is converted into light by ATP. This light is detected by a camera which captures light in the wells after each nucleotide mixture has been run over the plate (Radford et al., 2012, Van Dijk et al., 2014). As the process of adding nucleotides is repeated, a dsDNA molecule is synthesised while at the same time the nucleotide sequence is determined which is why the technology is based on a sequencing by synthesis principle. One of the main advantages of 454 pyrosequencing is the availability of sequencing result within a short period of time and also the generation of high through-put data (>200 000). There is also flexibility in terms of choosing primers to be used. Limitations associated with pyrosequencing include: high error rates, expensive reagents used for short read lengths and cumbersome data analysis (Fakruddin & Chowdhury, 2012).

Ion Torrent, by Ion Torrent Systems Inc. was commercially available in 2010. It was developed by the same person who developed 454 pyrosequencing (Van Dijk et al., 2014). As with 454 pyrosequencing, amplification of DNA forms after ligation of adapters with known sequences to beads in emulsion PCR reactions. After amplification, the beads are coated with numerous

copies of the amplified DNA. The DNA is denatured and bound to sequencing primers. The DNA is then loaded into wells attached to an ion-sensitive semiconductor. During the extension step, when the polymerase incorporates a nucleotide base, both hydrogen ion and pyrophosphate are released. The Ion Torrent exploits the released hydrogen ion, by detecting a small change in pH (Radford et al., 2012). Advantages of Ion torrent include the speed at which results are generated and lower costs involved per sample.

Illumina sequencing technology was introduced in 2006 by Illumina. In the library preparation step, DNA or cDNA molecules are fragmented into smaller molecules than in 454 or Ion Torrent sequencing. With Illumina technology it is also possible to combine DNA fragmentation and adapter ligation into one step, making library preparation more efficient. The adapter ligated DNA fragments are then amplified by PCR followed by gel purification. The purified library is amplified using bridge amplification. In bridge amplification, the purified library is loaded onto a flow cell coated with oligonucleotides that are complementary to the ligated adapters ligated to the library. One end of the library DNA binds to the oligonucleotides on the flow cell and the other end hybridises to a complementary oligonucleotide on the flow cell, forming a looped structure. The library is then amplified into numerous clusters. Following bridge amplification, DNA fragments are ready for sequencing after being denatured. Illumina sequencing is based on sequencing by synthesis which makes use of four fluorescently labelled reversible terminator-bound dNTPs. During the extension step, single bases are detected as they are incorporated into the growing DNA fragment and the fluorescent dye is detected by an imaging system (Radford et al., 2012, Van Dijk et al., 2014).

Next generation sequencing technologies can also be used to sequence RNA. In RNA sequencing, the first step is total RNA library preparation followed by removal of ribosomal RNA (rRNA) which accounts for a large percentage of RNA present in all cells. Removal of rRNA can be achieved by use of commercial kits that either remove rRNA or enrich for mRNA. After removal of rRNA, total RNA is then converted into cDNA. NGS library preparation and sequencing steps follow depending on which platform is used (Wilhelm and Landry, 2009). Next generation sequencing platforms have also been used for virome sequencing with the aim of detecting both DNA and RNA viruses present in biological samples. Some studies have reported detection of novel viruses using NGS for virome studies (Loh et al., 2009; Victoria et al., 2009).

Prior to the application of NGS technologies for studying segmented dsRNA viruses, sequencing genome segments of these viruses was achieved by, either sequencing individually cloned genome segments (Attoui et al., 2000), or sequencing amplicons of the individual genome segments (Maan et al., 2007). These methods were time consuming. In 2009, Potgieter and co-workers reported the first application of NGS using 454 pyrosequencing coupled with sequence-independent genome amplification for sequencing of dsRNA viruses from four different virus families including rotaviruses from the *Reoviridae* family (Potgieter et al., 2009). For rotaviruses, the application of 454 pyrosequencing for whole genome analysis was also reported by Jere and co-workers (Jere et al., 2011a, Jere et al., 2011b, Jere et al., 2012). Various NGS platforms such as the Illumina Miseq have been used for rotavirus whole-genome characterisation (Nyaga et al., 2014; Tacharoenmuang et al., 2015). Of late, various studies on virome sequencing from biological samples such as faecal specimens have also been reported with the detection of novel viruses (Conceição-Neto et al., 2015).

The aim of this chapter was to use NGS to resolve the genotypes of selected Mozambican rotavirus strains that were previously typed as mixed genotypes by genotyping PCR, or that could not be typed with genotyping PCR. Two additional samples that were successfully typed with genotyping PCR were also included in whole genome analysis to compare the results obtained. The genotypes of the study strains are described according to the whole genome classification notation that was described by Matthijnssens and co-workers where -Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx is used to represent the genotypes of the 11 genome segments that code for the viral proteins in the following order -VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6 (Matthijnssens et al, 2008).

3.2 Materials and Methods

3.2.1 Rotavirus strains

Nine stool specimens, previously obtained under MTA from CISM (see section 2.2.1) were selected for whole genome analysis (Table 3.1). Genotypes assigned to the samples were determined by the Mozambican group using a standard genotyping PCR, and either typed as mixed or untypeable.

Table 3.1: Mozambican rotavirus strains targeted for next generation sequencing and genotyping results obtained by the Mozambican group using according to genotyping PCR. The site and year in which the samples were collected are also shown.

Strains	Genotype	Site of collection	Year of collection
0043	G2G8G9P[NT]	Manhiça	2012
0044	G8G12P[NT]	Manhiça	2012
0045	G8G12P[NT]	Manhiça	2012
0052	G8G12P[NT]	Manhiça	2012
0060	G12P[NT]	Manhiça	2012
0144	NT	Manhiça	2013
0257	G8G12P[6]	Mavalane	2012
0278	G12P[6]	Mavalane	2012
0314	G2G9P[8]	Mavalane	2012

NT: Genotype of the particular sample could not be determined. P[NT]: P genotype could not be determined for the particular sample.

3.2.2 RNA extraction

Total RNA was extracted from the nine specimens (Table 3.1) using TRI-reagent® (Sigma) and the same procedure described in section 2.2.2. The concentration of the extracted RNA was determined using Nanodrop 2000 (Nanodrop) and the extracted RNA was analysed on 1% TBE agarose gels as described in section 2.2.4.

3.2.3 LiCl clean-up of the extracted double-stranded RNA

Lithium chloride (LiCl) (Sigma-Aldrich) at a final concentration of 2 M was used to purify the extracted RNA by removing ssRNA molecules. Lithium chloride was prepared in distilled water and 30 µl of this was added to 90 µl of the extracted total RNA and incubated at 4 ^{o}C overnight. Following incubation, the LiCl- RNA mixture was centrifuged at 16 000 x g for 30 minutes at 4 $^{\circ}$C to precipitate the LiCl. The supernatant containing dsRNA was purified using the MinElute kit (Qiagen). The supernatant was added to 330 µl of buffer QG. The mixture was transferred to a spin column and centrifuged at 17 900 x g for a minute, the flow through

was discarded, and the RNA was washed with 700 µl of buffer PE. Following the wash step, the dsRNA was eluted with 32 µl of elution buffer and 10 µl was run on an agarose gel.

3.2.4 Illumina Miseq Sequencing

The samples were sent for sequencing after LiCl treatment of dsRNA-enriched samples. Samples were diluted to concentrations of 40 and 80 ng/µl with the same elution buffer used during RNA extraction. Whole genome characterisation of nine selected Mozambican samples was done using Illumina Miseq platform at Inqaba Biotech (Pretoria, South Africa). For library preparation, Illumina's Nextera sample preparation kit was used. Briefly, RNA was converted to genomic DNA (gDNA) and a library prepared by fragmenting the DNA into smaller molecules and ligating bar-coded adapters to the DNA in one step. PCR amplification of the tagged DNA followed and the libraries were sequenced. The kit used for sequencing was the Miseq v3, 600 cycle.

3.2.5 Data analysis

3.2.5.1 Analysis of Illumina data and genotype assignment

CLC Bio Genomics Workbench version 8.5.1 (CLC Bio, Denmark) was used to assemble Illumina sequencing reads into a number of contigs with each contig corresponding to a specific rotavirus genome segment. Illumina reads were first assembled into *de novo* contigs. The resulting contigs were submitted for BLASTn (http://blast.ncbi.nlm.nih.gov/) analysis to determine if they were rotavirus sequences and which genome segments they coded for. Once it was established that the contigs were rotavirus encoding genome segment sequences, the contigs were submitted to RotaC $^{2.0}$ (http://rotac.regatools.be/). RotaC $^{2.0}$ is an automated online rotavirus genotyping tool for group A rotaviruses to assign genotypes to the genome segments. RotaC $^{2.0}$ was developed by Maes and co-workers (Maes et al., 2009).

3.2.5.2 Phylogenetic analysis

Sequences for the genome segments of nine Mozambican strains characterised in this study were submitted to MOLE-BLAST (https://blast.ncbi.nlm.nih.gov/moleblast/moleblast.cgi) to obtain sequences with a high similarity to the Mozambican strains. Accession numbers of the nucleotide sequences used in the analysis are given in Appendix D. Together with sequences

of Mozambican strains, these sequences were used to construct phylogenetic trees for the 11 genome encoding segments. Sequences were aligned using the Muscle alignment tool in MEGA 5.2. Phylogenetic trees were drawn using the Neighbour Joining method (Saitou and Nei, 1987) with 1000 bootstrap replicates, using the best fit model as determined by MEGA.

3.3 Results and Discussion

To resolve the genotypes of rotavirus samples that could either not be typed by genotyping PCR or were previously identified as mixed infections, nine specimens were subjected to next generation sequencing for whole-genome characterisation. Following RNA extraction from stools, these specimens were treated with LiCl to remove single-stranded-RNA molecules. Before being sent for sequencing, dsRNA concentrations were diluted as indicated in section 3.2.4. NGS sequencing results were analysed and contigs determined. This was followed by assignment of genotypes and phylogenetic analysis.

3.3.1 Preparation of dsRNA for NGS

Using TRI-reagent®, RNA was successfully extracted from the nine specimens selected for NGS. RNA concentrations determined using Nanodrop ranged from 81 ng/µl for sample 0043 to 570 ng/µl for sample 0314. These specimens were diluted to RNA concentrations between 40 ng/µl and 81 ng/µl. Lithium chloride treatment was successful and genome segments 1 to 9 can be seen after agarose gel electrophoresis while segments 10 and 11 are not visible due to the dye font caused by high concentrations of ethidium bromide (Figure 3.1).

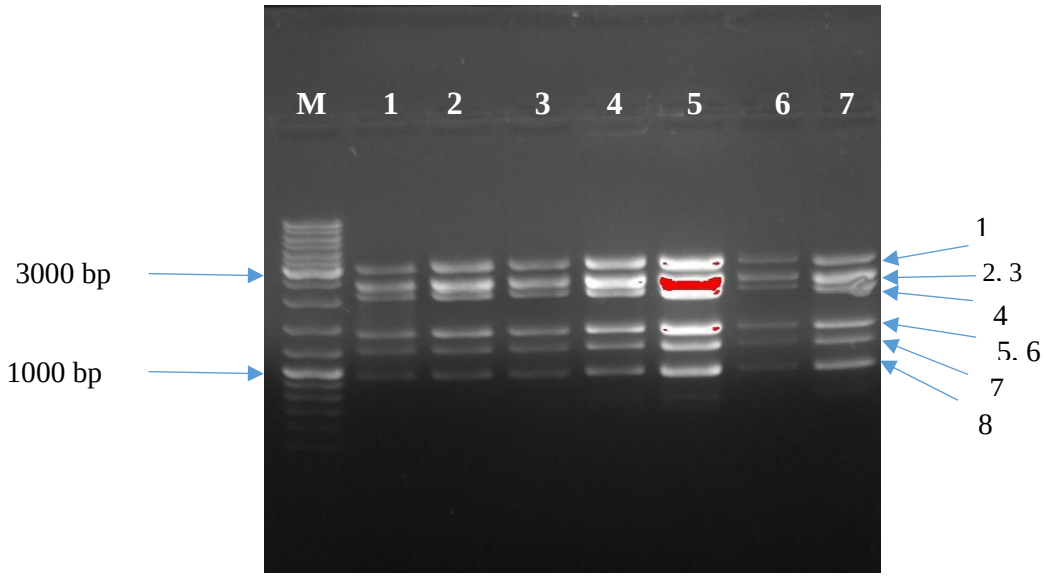


Figure 3.1: A representative gel showing analysis of the extracted dsRNA post LiCl treatment on a 1 % TBE agarose gel. Lanes 1 – 7 are samples 0043, 0044, 045, 0052, 0144, 0257 and 0314. Lane M is the GeneRuler DNA ladder Mix molecular weight marker (Thermo-Scientific) that was used to confirm the different sizes of the RNA segments. For the molecular weight marker, 5 µl was loaded on the gel and for each sample 10 µl were loaded on the gel. Genome segments 1 to 8 can be seen on the gel.

3.3.2 Next generation sequencing data results and data assembly

Paired-end Illumina Miseq reads, were initially assembled *de novo* to obtain contigs. The resulting contigs were submitted to BLASTn analysis and the contigs which were confirmed to be rotavirus genome segments were further submitted to RotaC for genotype assignment. RotaC results gave an indication of the identity of the virus genes. Virus genes were further assembled against appropriate reference strains. Samples with Wa-like genotypes were assembled against reference strain RVA/Human-tc/USA/Wa/1974/G1P1A[8] (Wentzel et al., 2013) while samples with DS-1-like genotype were assembled against reference strain RVA/Human-tc/USA/DS-1/1976/G2P1B[4] (Mlera et al., 2011).

Table 3.2. shows the lengths of the various genome segments of Mozambican rotavirus strains obtained after NGS data assembly and the percentage of the length obtained in comparison to the full-lengths of the various genome segments for the prototype rotavirus strain SA11 (Mlera et al 2013). The minimum length obtained for genome segment 1 in comparison to that of SA11 is 44.31% for sample 0060 while for the rest of the samples, the proportion of genome segment 1in comparison to the reference strain ranged from 82.02% to 100%. The proportion of the

lengths obtained for genome segment 2 in comparison to the referenced strain ranged from 95.50% to 100% for all Mozambican samples. For genome segment 3, the lengths obtained ranged from 82.17% to 100%. Mozambican strains characterised in this study had the same lengths for the VP4 encoding genome segment which is 99.87% of the full-length of the reference strain with the exception of the second VP4 encoding genome segment of sample 0060. For genome segment 5, the lengths obtained ranged from 79.86% to 97.09%. Full-length sequences were obtained for five samples of segment 6 and the remaining four had lengths of genome segment 6 ranging from 43.65% to 99.93%. The minimum length obtained for genome segment 7 was 69.41% and maximum 97.10% while full-lengths were obtained for genome segment 8 for six Mozambican strains with the remaining lengths between 98.77% and 99.53%. Sizes for segment 9 ranged from 40.45% to 100%. Full-lengths were obtained for all samples for genome segment 10 and the minimum length obtained for segment 11 in comparison to the reference strain is 90.08%.

The lengths of the open reading frames (ORFs) for all nine Mozambican rotavirus strains were successfully obtained after assembly of the data as shown in Table 3.2. Full-length ORFs for all nine samples for genome segments encoding for NSP2 and NSP4 proteins were obtained. The lengths obtained for the ORFs of segment 4 for all nine Mozambican samples were the same (99.87%) with the exception of the second segment 4 of sample 0060. For VP3 ORFs, only three samples did not have the entire OFRs. For segment 6, seven samples had full length ORFs while six Mozambican samples had full-length ORFs for segment 2. For segment 1, the proportion of the ORFs obtained in comparison to the reference strain ranged from 43.77% to 100%. More than 90% of the full-length ORFs were obtained for segment 7 and 11, while more than 75% of the full-length ORFs were obtained for segment 5 for all Mozambican strains. VP7 ORFs ranged from 43.82% to 100% as shown in Table 3.2 and for the first ORF of sample 0060 with genotype G12P[8] respectively. NGS data assembly was therefore successful as the individual genome segments and ORFs for all samples were determined. Average coverage ranged from 95.66 to 4233.76. Accession numbers for the SA11 strain genome segments that the Mozambican strain ORFs (open reading frames were calculated based on are in Appendix D).

Table 3.2: Description of assembled genome segments

Genome segment (GS)	Strains	43	44	45	52	144	257	278	314	60	
1 (VP1)	Length of GS Length of ORF	3250 3240	3301 3267	3120 3120	3301 3267	3181 3020	2707 2680	3302 3267	2721 2690	3302 3267	1463 1430
	% Length of GS % Length of ORF	98.4% 99.17	99.97 100	94.49 95.50	99.97 100	96.34 92.44	81.98 82.03	100 100	64.23 82.34	100 100	44.31 43.77
	Average Coverage	936.55	461.38	471.19	133.49	392.36	4233.76	3062.72	2572.45	3229.11	2791.44
2 (VP2)	Length of GS Length of ORF	2683 2665	2596 2573	2684 2665	2569 2541	2572 2530	2684 2665	2693 2665	2684 2665	2685 2665	
	% length of GS % Length of ORF	99.7% 100	96.51 96.56	99.78 100	95.50 95.35	95.61 94.93	99.78 100	100 100	99.78 100	99.78 100	
	Average Coverage	1174.93	544.64	381.03	225.41	751.73	3076.44	2110.98	2177.94	1946.99	
3 (VP3)	Length of GS Length of ORF	2590 2508	2177 2128	2590 2508	2194 2100	2129 1903	2591 2508	2587 2508	2591 2508	2590 2508	
	% length of GS % Length of ORF	99.96 100	84.02 84.85	99.96 100	84.68 83.73	82.17 75.88	100 100	99.85 100	100 100	99.96 100	
	Average Coverage	931.35	307.58	430.20	2930.63	218.44	3117.52	1884.67	2102.28	1897.25	
4 (VP4)	Length of GS Length of ORF	2359 2328	2359 2328	2359 2328	2359 2328	2359 2328	2359 2328	2359 2328	2359 2328	2359 2328	1232 1190
	% length of GS % length of ORF	99.87 99.87	99.87 99.87	99.87 99.87	99.87 99.87	99.87 99.87	99.87 99.87	84.72 99.87	99.87 99.87	99.87 99.87	52.42 51.05
	Average Coverage	1431.07	714.37	385.26	1261.00	774.11	3216.37	1144.16	2221.67	3427.04	110.53
5 (NSP1)	Length of GS Length of ORF	1564 1428	1289 1208	1567 1482	1256 1166	1321 1291	1564 1482	1561 1460	1566 1482	1565 1460	
	% length of GS % Length of ORF	96.90 95.77	79.86 80.62	97.09 99.40	77.81 78.20	81.85 86.57	97.08 99.40	96.90 97.92	97.02 99.40	96.96 97.92	
	Average coverage	599.27	159.32	165.26	725.60	358.31	688.84	3042.46	551.40	2415.44	
6 (VP6)	Length of GS Length of ORF	1356 1194	1295 1124	1356 1194	1116 854	1356 1194	1356 1194	1355 1194	1356 1194	1354 1194	592 580
	% length of GS % Length of ORF	100 100	95.51 94.14	100 100	82.30 71.52	100 100	100 100	99.93 100	100 100	99.85 100	43.65 48.58
	Average Coverage	953.28	144.65	324.43	897.15	273.84	3165.17	2545.60	1874.13	2128.49	95.66

Table 3.2: Description of assembled genome segments (continued)

Genome segment (GS)	Strains	43	44	45	52	144	257	278	314	60	
7 (NSP3)	Length of GS Length of ORF	1064 943	1069 948	1066 943	1064 943	1064 943	1066 943	1073 933	1067 943	1073 933	
	% length of GS % Length of ORF	96.29 99.47	96.74 100	96.47 99.47	69.41 99.47	96.29 99.47	96.47 99.47	97.10 98.42	96.56 99.47	97.10 98.42	
	Average coverage	819.47	359.08	301.16	626.42	358.33	2540.47	1497.53	2115.67	1048.23	
8 (NSP2)	Length of GS Length of ORF	1059 954	1058 954	1059 954	1059 954	1054 954	1059 954	1059 954	1059 954	1046 954	
	% Length of GS % Length of ORF	100 100	99.91 100	100 100	100 100	99.53 100	100 100	100 100	100 100	98.77 100	
	Average Coverage	422.97	153.86	188.31	832.25	199.47	1840.17	1426.96	1023.52	1389.07	
9 (VP7)	Length of GS Length of ORF	1063 981	947 832	779 700	832 783	785 725	879 830	939 890	934 885	953 908	430 430
	% length of GS % Length of ORF	100 100	89.09 84.81	73.28 71.3	78.34 79.82	73.85 73.90	82.69 84.61	88.33 90.72	87.86 90.21	89.65 92.56	40.45 43.82
	Average Coverage	597.81	256.27	209.54	144.24	166.55	1923.87	1235.81	993.27	437.45	120.75
10 (NSP4)	Length of GS Length of ORF	746 528	746 528	751 528	743 528	749 528	751 528	750 528	751 528	750 528	
	% length of GS % length of ORF	99.33 100	99.33 100	100 100	98.93 100	99.73 100	100 100	99.87 100	100 100	99.86 100	
	Average Coverage	592.22	248.18	183.91	117.89	221.37	956.12	1622.65	565.89	1472.96	
11 (NSP 5/6)	Length of GS Length of ORF	600 570	612 584	653 597	616 587	623 549	650 583	662 597	660 597	663 597	
	% length of GS % Length of ORF	90.08 95.48	91.89 97.82	98.05 100	92.49 98.32	93.54 91.96	98.48 97.65	99.40 100	99/01 100	99.54 100	
	Average Coverage	243.18	110.54	630.86	122.57	189.92	602.80	697.60	384.81	343.43	

GS: genome segment, ORF: Open reading frame, NSP; non-structural protein, VP; viral protein

3.3.3. Evaluation of genotyping primer binding sites in relation to strains that were previously typed as mixed or untypeable using genotyping PCR

In this section, primer binding regions of strains, which were typed as mixed using genotyping PCR, were evaluated to determine if the results obtained were due to genotyping primers failure. Mozambican strains RVA/Human-wt/MOZ/0043/2012/G2P[4] and RVA/Human-wt/MOZ/0314/G8P[4] were typed as mixed genotypes containing G9 genotype while with next generation sequencing, the two samples were typed as G2P[4] and G8P[4], respectively. Evaluation of G9 primer binding region in Figure 3.2 (A) indicated that the two Mozambican strains do bind to the primer although not a position where a G9 primer is expected to bind and with about six mismatches. The fact that the G9 primer binds explains why there was a G9 genotype annotation as a corresponding region was amplified during the PCR reaction. For Mozambican samples, new G9 primers could be designed or alternative primers used to determine if there will not be an improvement in the results obtained.

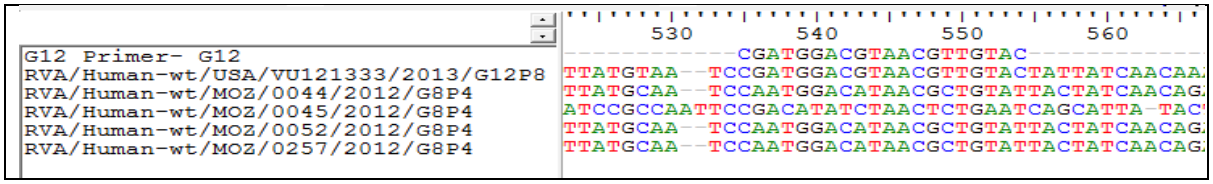
Four Mozambican G8 strains, Figure 3.2 (B), were typed as mixed and containing G12 genotype. The reason why G12 genotype was detected when using genotyping PCR is because the four Mozambican strains bind to the G12 primer although there are mismatches but they do not seem to hinder the binding of G12 primers to G8 sequences. In addition, G12 primers bind to Mozambican G8 strains at nucleotide positions that intersect with the correct binding regions for G12 primers. To avoid this, alternative G12 sequencing primers for Mozambican G12 strains should be used or designed.

Mozambican strain RVA/Human-wt/2013/G2P[4] was untypeable with genotyping PCR but was typed as G2P[4] with next generation sequencing. Evaluation of G2 primer binding site, Figure 2.9 (C) does not suggest any primer failure as the primer matches perfectly with the Mozambican strain. This result is consistent with the results obtained when evaluating G2 primer binding sites in relation to Mozambican strains sequenced with Sanger sequencing and RT-PCR. Similarly, evaluation of P[4] primer binding region, Figure 3.2 (E) does not suggest any primer failure and this is the same result obtained from the evaluation of P[4] primers in relation to Sanger sequencing and RT-PCR sequencing.

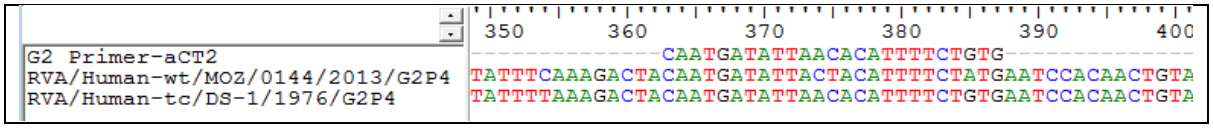
Mozambican strain RVA/Human-wt/MOZ/0314/G8P[4] was typed as mixed with genotype P[8] with genotyping PCR but was typed as P[4] with NGS. Evaluation of P8 primer binding site, Figure 3.2 (D) indicated that the P[8] primer binds to the Mozambican strain sequence and results in a fragment that correspond to the P[8] primer fragment. The P[8] primers is a degenerate primer and a solution could be using alternative P[8] primers.

A.

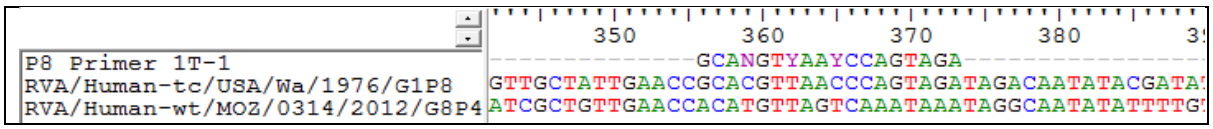
B.



C.



D.



E.

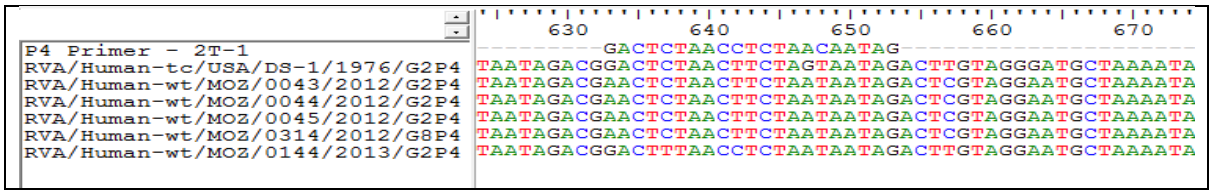


Figure 3.2 (A) Sequence alignment of G9 primer binding region and Mozambican strains 0043 and 0314 that were typed as mixed, containing G9 genotype according to genotyping PCR. With next generation sequencing technology using the Illumina Myseq platform, the two strains do not contain G9 genotype. Strain RVA/Human-wt/ZAF/5001DB/1997/G9P[6] was used as a reference. (**B)** Genotype G12 primer binding region sequence alignment and Mozambican strains that were typed as mixed and containing G12 genotype according to genotyping PCR. Strain RVA/Human-wt/USA/VU121333/2013/G12P[6] was used as a reference. **(C)** Sequence alignment for G2 primer binding region and Mozambican strain 0144 which could not be typed using genotyping PCR. Strain RVA/Human-wt/USA/DS-1/1976/G2P[4] was used a reference. **(D)** Sequence alignment for P[8] primer binding region and Mozambican strain 0314 that was typed as mixed and containing P[8] genotype according to genotyping PCR. Strain RVA/Human-tc/USA/Wa/1976/G1P[8] was used as a reference. **(E)** Genotype P[4] primer binding region sequence alignment and Mozambican strains which the

G genotype could not be determined using genotyping PCR. Strain RVA/Human-wt/USA/DS-1/1976/G2P[4] was used as a reference for the correct primer binding region.

3.3.4 Resolution of genotypes that were previously typed as mixed or were untypeable

Using NGS, genotypes of samples which could not be determined with genotyping PCR were resolved (Table 3.3). Samples which were previously typed as mixed appeared as single genotypes according to NGS analysis. Genotype results obtained by the Mozambican group are, therefore, an indication of genotyping primer cross reaction or mis-priming failure which shows a limitation of genotyping PCR. Whole-genome analysis of nine Mozambican strains revealed G8P[4] with a DS-1-like backbone as the genotype detected for most samples. In addition, there were also two samples with a DS-1-like genotype that were typed as G2P[4]. Sample 0060 was typed as G12PNT with genotyping PCR and as G12P[8] with Sanger sequencing and BLASTn analysis. However, whole-genome analysis of this sample showed it to be a mixed genotype i.e G12G6P[8]P[14] of strains with a DS-1 backbone and a Wa-like backbone.

Rotavirus dual typing of strains into P and G genotypes is a quick and cheaper method for surveillance studies. However, it limits the amount of information generated on rotavirus strains to genome segments 4 and 9. Additional PCR bias may also occur in samples with high cDNA concentration where only the dominant strain is amplified. An example showing PCR bias is the result obtained in this study for sample 0060. Following RT-PCR, Sanger sequencing and BLASTn analysis, this sample was assigned G12P[8]. Whole-genome characterisation revealed the presence of a G6 and a P[14] genotype in the same sample at lower levels indicated by low coverage for genome segments 4, 6 and 9. Similar observations have been reported before where single genotype strains are revealed to be a mixture (Jere et al., 2011b). In addition, whole-genome characterisation of strain 0060 revealed two different R (R1 and R2) and I (I1 and I2) genotypes for the VP1 and VP6 encoding segments, respectively. It was also revealed that sample 0060 contained a mixed infection with a genotype of an animal origin G6P[14] (Tacharoenmuang et al., 2015). Sample 0060 was collected from Manhiça which is a rural area and therefore contact of people with animals is possible.

Whole-genome characterisation of nine Mozambican strains indicated that the majority of strains have the DS-l-like backbone and no reassortment events were suggested by the data

obtained. None of the strains previously suggested by genotyping PCR to be mixed infections were mixed samples.

Table 3.3: Nine Mozambican rotavirus strains that were selected for next generation sequencing and the genotypes obtained for all the 11 genome segments with RotaC[2.0]

Strains	Genotype										
	VP7	**VP4**	**VP6**	**VP1**	**VP2**	**VP3**	**NSP1**	**NSP2**	**NSP3**	**NSP4**	**NSP5**
0043	G2	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0044	G8	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0045	G8	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0052	G8	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0060	G12	P8	I1	R1	C1	M1	A1	N1	T1	E1	H1
	G6	P14	I2	R2							
0144	G2	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0257	G8	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
0278	G12	P6	I1	R1	C1	M1	A1	N1	T1	E1	H2
0314	G8	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2
Wa-like	G1	P8	I1	R1	C1	M1	A1	N1	T1	E1	H2
DS-1	G2	P4	I2	R2	C2	M2	A2	N2	T2	E2	H2

Red: DS-1-like genotype backbone
Green: Wa-like genotype backbone
Purple: Genotype with bovine origin

3.3.5 Phylogenetic analysis of the 11 genome segments of Mozambican rotavirus strains.

Phylogenetic trees were constructed and used to infer the evolutionary relatedness of Mozambican strains with other globally circulating strains with more emphasis on strains circulating in southern Africa. Sequences with close similarity to the Mozambican strains study were obtained from NCBI (National Centre for Biotechnology Information) and used to construct phylogenetic trees for all 11 genome segments. In all the phylogenetic trees, Mozambican strains originating from Mavalane are highlighted in red and strains from Manhiça in blue.

3.3.5.1 Phylogenetic analysis of genome segment 1 which encodes for VP1

Rotavirus genome segment 1 sequences in Figure 3.3 grouped into lineage I (Nakagomi et al., 2013) and VII (Giammanco et al., 2014) and further clustered into two separate clusters, labelled A and B (Figure 3.2). In lineage I, Mozambican strains RVA/Human-wt/MOZ/0044/2012/G8P[8], RVA/Human-wt/0257/2012/G8P[4], RVA/Human-wt/0043/2012/G2P[4] and RVA/Human-wt/0314/2012/G8P[4] grouped closely together indicating high similarity between the nucleotide sequences of these strains. Additionally, in lineage I, Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped closely with two Kenyan strains RVA/Human-wt/KEN/MRC-DPRU1606/2009/G8P[4] and RVA/Human-wt/ KEN/MRC-DPRU4155/XXXX/G8P[X] and three Tanzanian strains namely RVA/Human-wt/TZA/MRC-DPRU4570/2011/G8P[4], RVA/Human-wt/TZA/MRC-DPRU4568/2011/G8P[4] and RVA/Human-wt/TZA/MRC-DPRU4576/2010/G8P[4].

In lineage VII, Mozambican strains RVA/Human-wt/MOZ/0144/2013/G2P[4] and RVA/Human-wt/MOZ/0045/2012/G2P[4] grouped with one Australian strain, one Zimbabwean strain and three strains from the Philippines. In cluster A, Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] with R2 genotype grouped closely to an antelope rotavirus RVA/Antelope-wt/ZAF/RC-18-08/G6P[14] and this grouping had a 99% bootstrap support. This cluster also included three South African rotavirus strains isolated from bovine. There are two sequences for the VP1 encoding genome segment of Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14], one with a DS-1-like origin and the other with a Wa-like origin suggesting co-infection with two different strains. In cluster C, two Mozambican strains RVA/Human-wt/MOZ/0278/2012/G12P[6] and RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] with R1 genotyped grouped with three strains collected from South Africa between 2011 to 2012.

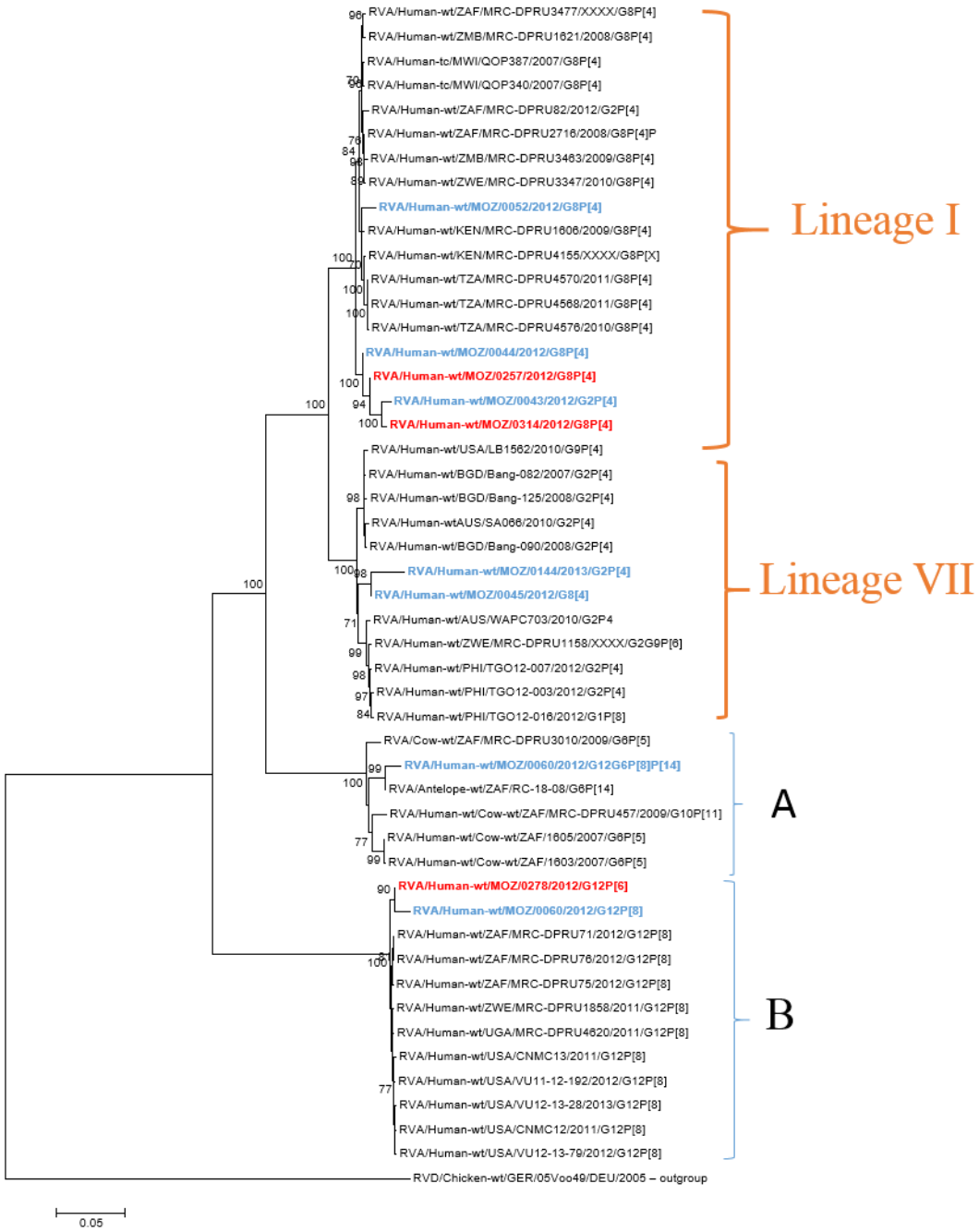


Figure 3.3: Phylogenetic analysis of genome segment 1 of Mozambican strains in relation to other international stains. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Lineages which could be identified are indicated in brackets and the clusters which the sequences form are labelled A and B. The outgroup used is the VP1 encoding genome segment of strain RVD/Chicken-wt/GER/05V0049/DEU/2005.

3.3.5.2 Phylogenetic analysis of genome segment 2 which encodes for VP2

In the phylogenetic analysis of VP2 encoding genome segment, lineages II (Nakagomi et al., 2013) and VIII (Giammanco et al., 2014) could be identified as indicated in Figure 3.4. There are also Mozambican strains which grouped into four separate clusters, A to D, whose lineages could not be identified. In lineage II, Mozambican strains RVA/Human-wt/MOZ/0044/2012/G8P[4], RVA/Human-wt/MOZ/0314/2012/G8P[4] and RVA/Human-wt/MOZ/0257/2012/G8P[4] grouped with Malawian tissue culture adapted rotavirus strains RVA/Human-tc/MWI/QOP383/2007/G8P[4] and RVA/Human-tc/MWI/QOP340/2007/G8P[4].

Still in lineage II, Mozambican strain RVA/Human-wt/0052/2012/G8P[4] grouped with strains collected from Kenya and Tanzania similar to its grouping for genome segment 1 in Figure 3.2. In lineage VIII, Mozambican strain RVA/Human-wt/MOZ/0045/2012/G8P[4] had a high nucleotide identity >99% with a human reassortment strain from the United States of America RVA/Human-wt/USA/LB1562/2010/G9P[4] (Lewis et al., 2014). In addition, the same Mozambican strain also grouped with strains from Thailand, Uganda, Ethiopia, and the United States of America.

In cluster A, Mozambican strain RVA/Human-wt/MOZ/0144/2013/G2P[4] grouped with strains from the Philippines. In cluster B which consists of strains with Wa-like genotypes, Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] grouped with two strains from South Africa isolated from bovine. The two strains RVA/Cow-wt/ZAF/1605/2007/G6P[5] and RVA/Cow-wt/ZAF/1603/2007/G6P[5] were obtained from stool samples of calves and whole genome analysis of the samples done using 454 pyrosequencing. The results obtained from the study suggested that these bovine rotaviruses might have emerged as a result of reassortment events between human and bovine rotaviruses (Jere et al., 2012). In addition, the Mozambican strain grouped with one strain from South Africa isolated from an antelope, one from Hungary and one from Egypt. The strain isolated from Egypt was reported as the first strain to be isolated in the country with a G6P[14] genotype (El Sherif et al., 2011). The strain from Hungary was described in a study where whole genome characterisation of rotavirus strains revealed a common origin between human Wa-like rotaviruses and porcine rotaviruses (Matthijnssens et al., 2008). In cluster C, Mozambican strain RVA/Human-wt/MOZ/0278/2012/G12P[6] grouped with strains from Zimbabwe, South Africa, Kenya and Uganda. In cluster D, Mozambican strain RVA/Human-

wt/MOZ/0043/2012/G2P[4] grouped closely with a strain collected in South Africa and with two other strains originating from Zambia and Australia.

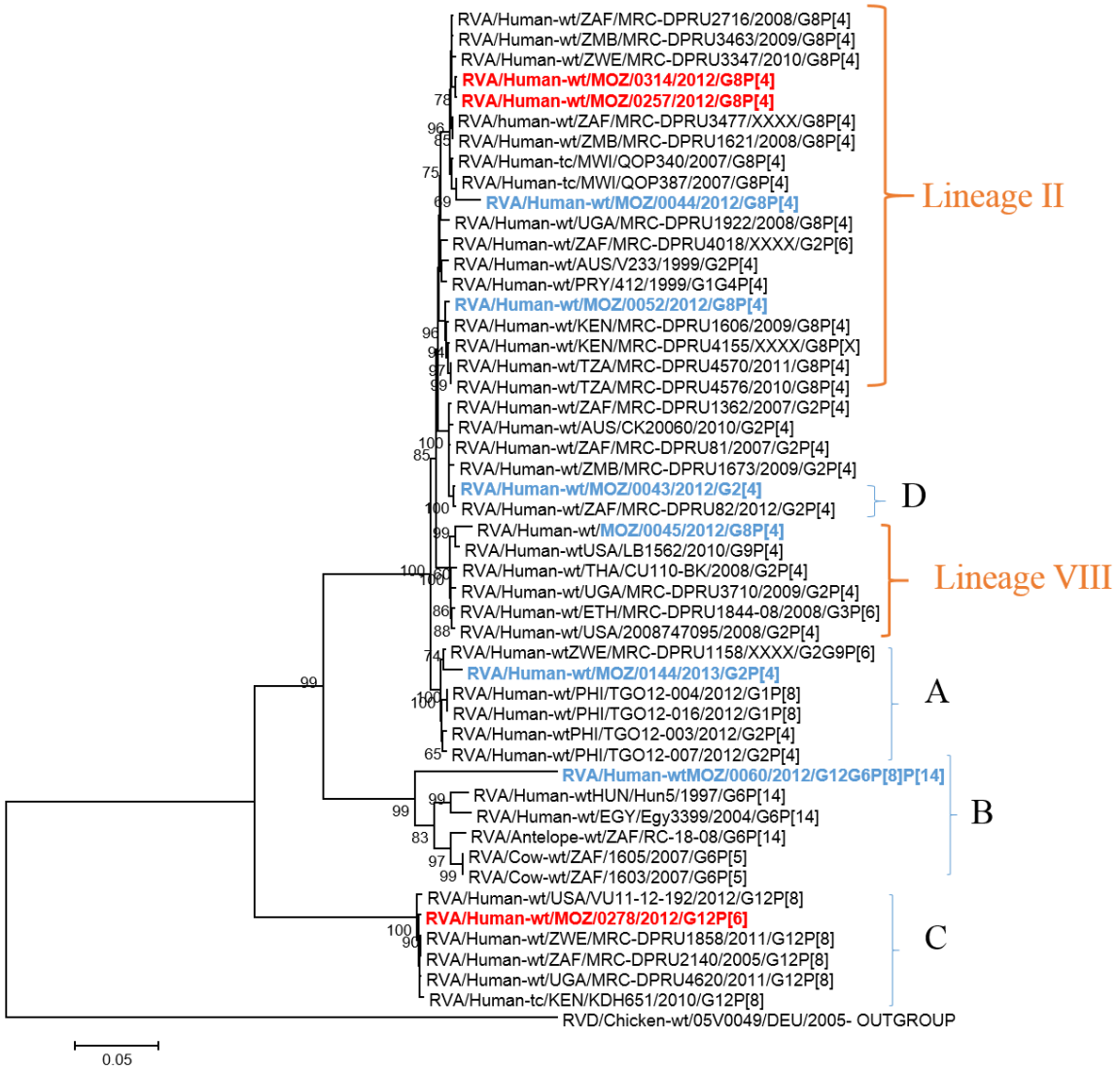


Figure 3.4: Phylogenetic analysis of genome segment 2 of Mozambican strains in relation to other international stains. MEGA 5.2 was used to construct the tree and the evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are labelled cluster A to D. Segment 2 encoding segment of strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup.

3.3.5.3 Phylogenetic analysis of genome segment 3 which encodes for VP3

In the phylogenetic analysis of genome segment 3, lineages I and II (Nakagomi et al., 2013) were identified (Figure 3.5). Strains whose lineages could not be identified are described in

clusters A to C. In lineage I, four Mozambican strains RVA/Human-wt/MOZ/0043/2012/G2P[4], RVA/Human-wt/MOZ/0044/2012/G8P[4], RVA/Human-wt/MOZ/0257/2012/G8P[4] and RVA/Human-wt/MOZ/0314/2012/G8P[4] grouped together and with one South African strain RVA/Human-wt/ZAF/MRC-DPRU-82/2012/G2P[4]. In lineage II, Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] grouped with tissue culture adapted strains from Malawi.

In cluster A, Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with two strains from Kenya and three strains from Tanzania which is similar to grouping observed for genome segments 1 and 2. In cluster B, Mozambican strains RVA/Human-wt/MOZ/0144/2013/G2P[4] and RVA/Human-wt/MOZ/0045/2012/G8P[4] grouped with a South African strain and a strain from the USA, respectively. Phylogenetic analysis of genome segment 3 shows that Mozambican strains are similar to strains circulating in low-income countries in southern Africa. The USA is the only high-income country that Mozambican strains group with. Strains in cluster C, have a Wa-like genotype. Mozambican strains RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] and RVA/Human-wt/MOZ/0278/2012/G12P[6] in cluster C, grouped with strains from Uganda, Zimbabwe and South Africa.

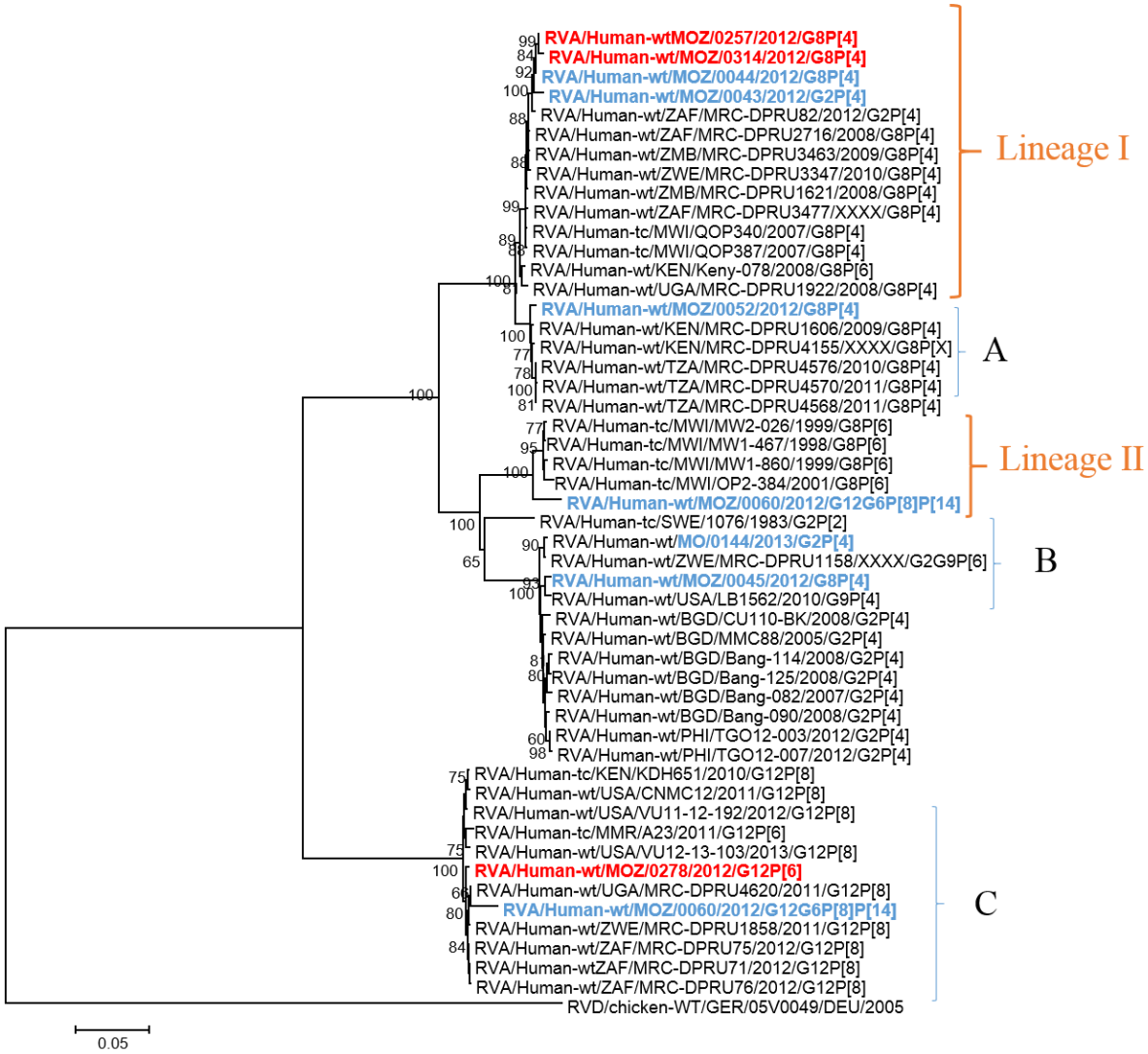


Figure 3.5: Phylogenetic analysis of genome segment 3 of Mozambican strains in relation to other international stains. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. VP3 encoding genome segment of strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in terms of clusters labelled A to C.

3.3.5.4 Phylogenetic analysis of genome segment 4 which encodes for VP4

For the VP4 encoding genome segment, lineages I (Nakagomi et al., 2013), II (Ide et al., 2015) and III (Komoto et al., 2014, Souvik Ghosh et al., 2011) could be identified (Figure 3.6). In lineage I, Mozambican strain RVA/Human-wt/MOZ/0043/G2P[4] grouped with a South African strain isolated from a pig. However, this grouping had a bootstrap support lower than

60%. Mozambican strains RVA/Human-wt/MOZ/0044/2012/G2P[4], RVA/Human-wt/MOZ/0314/2012/G8P[4], RVA/Human-wt/MOZ/0257/2012/G8P[4] and RVA/Human-wt/MOZ/0045/2012/G8P[4] in lineage I, grouped close to each other and to Malawian tissue culture adapted strains. Strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with one Kenyan strain at >90% nucleotide identity and three Tanzanian strains. This is a similar grouping to that observed with genome segment 1, 2 and 3 of the same strain.

In lineage III, Mozambican strain RVA/Human-wt/MOZ/0144/2012/G2P[4] grouped with two strains isolated from the Philippines, namely - RVA/Human-wt/PHI/GO-12-007/2012/G2P[4] and RVA/Human-wt/PHI/GO-12-003/2012/G2P[4] and three Australian strains. Also in lineage III, Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] with P[8] genotype grouped with one strain from Uganda and one strain from Zimbabwe. In addition, the Mozambican strain also grouped with three strains from the USA.

In lineage II, there is only one Mozambican strain, RVA/Human-wt/MOZ/0278/2012/G12P[6], and it had high nucleotide identity 99% with a South African strain RVA/Human-wt/ZAF/MRC-DPRU/4090 /2011/G12P[6]. The Mozambican strain also grouped with a strain from Germany RVA/Human-wt/GER/Ger172-08/2008/G12P[6] and three strains from Bangladesh. In cluster A, Mozambican strain RVA/Human-wt/MOZ/0060/G12G6P[8]P[14] with P[14] genotype had a high nucleotide identity with a strain isolated from Thailand RVA/Human-wt/THA/SKT-27/2012/G6P[14] which is a strain that is a product of interspecies transmission and reassortment events (Tacharoenmuang et al., 2015). Mozambican strain in cluster A also grouped with a rotavirus strain isolated from a sheep in Spain and a strain isolated from a rabbit in China. The grouping of this Mozambican strain in this cluster, suggests that the Mozambican strain is also a product of interspecies transmission. Strain RVA/Human-wt/MOZ/0060/G12G6P[8][14] was obtained from a patient in Manhiça, a rural area in Mozambique where people and animals live in close proximity.

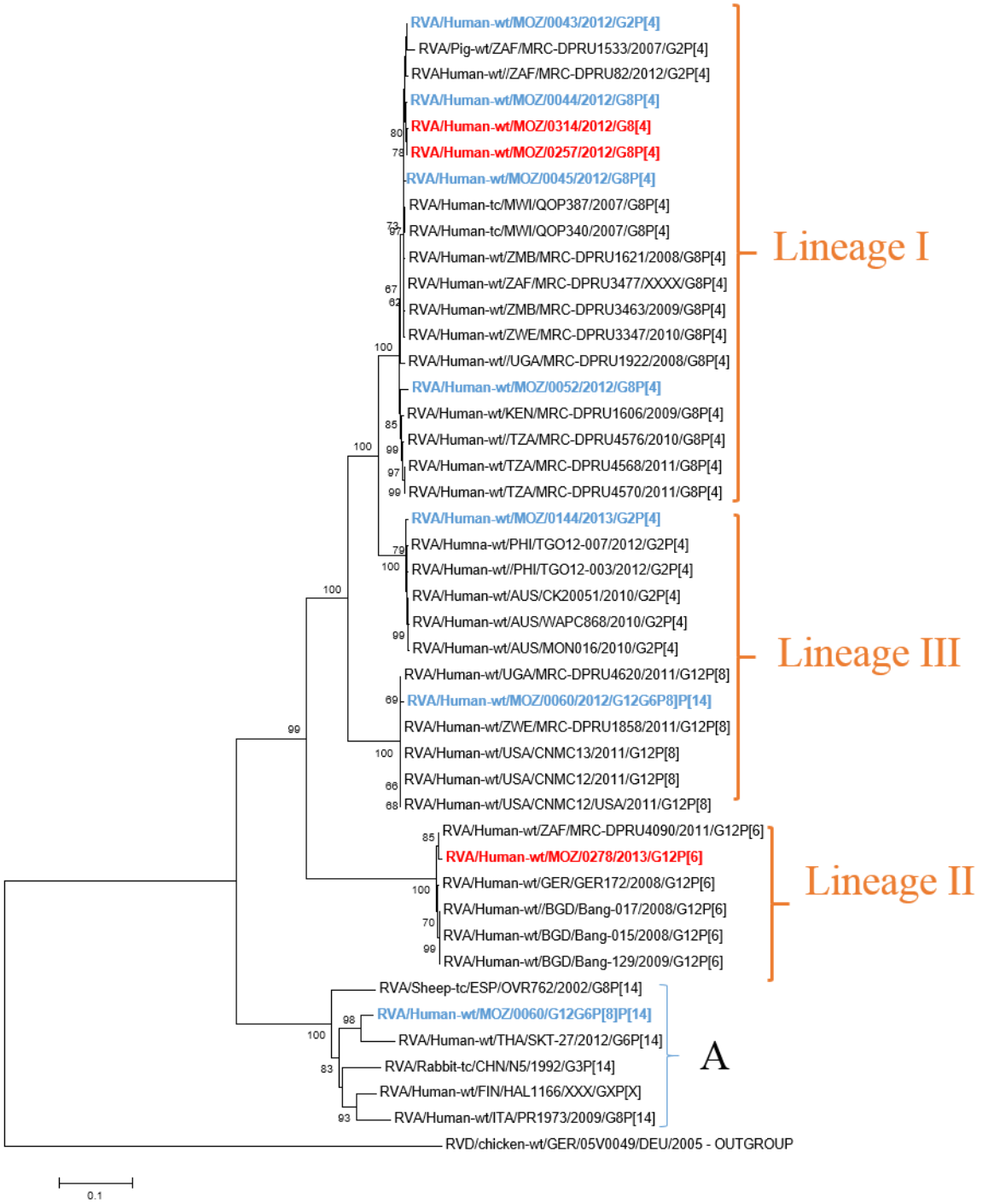


Figure 3.6: Phylogenetic analysis of genome segment 4 of Mozambican strains in relation to other international stains. MEGA 5.2 was used to construct the tree and the evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. VP4 encoding genome segment of strain RVD/Chicken-wt/GER/05v0049/DEU/2005 was used as an outgroup. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are in cluster A.

3.3.5.5 Phylogenetic analysis of genome segment 5 which encodes for NSP1

Lineage I (Nakagomi et al., 2013) strains could be identified for genome segment 5 (figure 3.7). In lineage I, Mozambican strain RVA/Human-wt/MOZ/0043/2012/G8P[4] grouped with a South African strain RVA/Human-wt/ZAF/MRC-DPRU82/2012/G2P[4] and in turn this cluster also grouped with another Mozambican strain RVA/Human-wt/MOZ/0044/2012/G8P[4]. Four Mozambican strains RVA/Human-wt/MOZ/0045/2012/G8P[4], RVA/Human-wt/MOZ/0052/2012/G8P[4], RVA/Human-wt/MOZ/0257/2012/G8P[4] and RVA/Human-wt/MOZ/0314/2012/G8P[4] grouped closer together.

In cluster A, Mozambican strain RVA/Human-wt/MOZ/0144/2013/G12P[6] had high nucleotide identity of 99% with a Zimbabwean strain Human-wt/ZWE/MRC-DPRU1158/XXX/G2G9P[6]. Both strains also grouped with strains isolated in the Philippines. In cluster B, Mozambican strains RVA/Human-wt/MOZ/0060/2012/G12GP[8]P[14] and RVA/Human-wt/MOZ/0278/2012/G12P[6] grouped closely together and with other strains from South Africa and Zimbabwe.

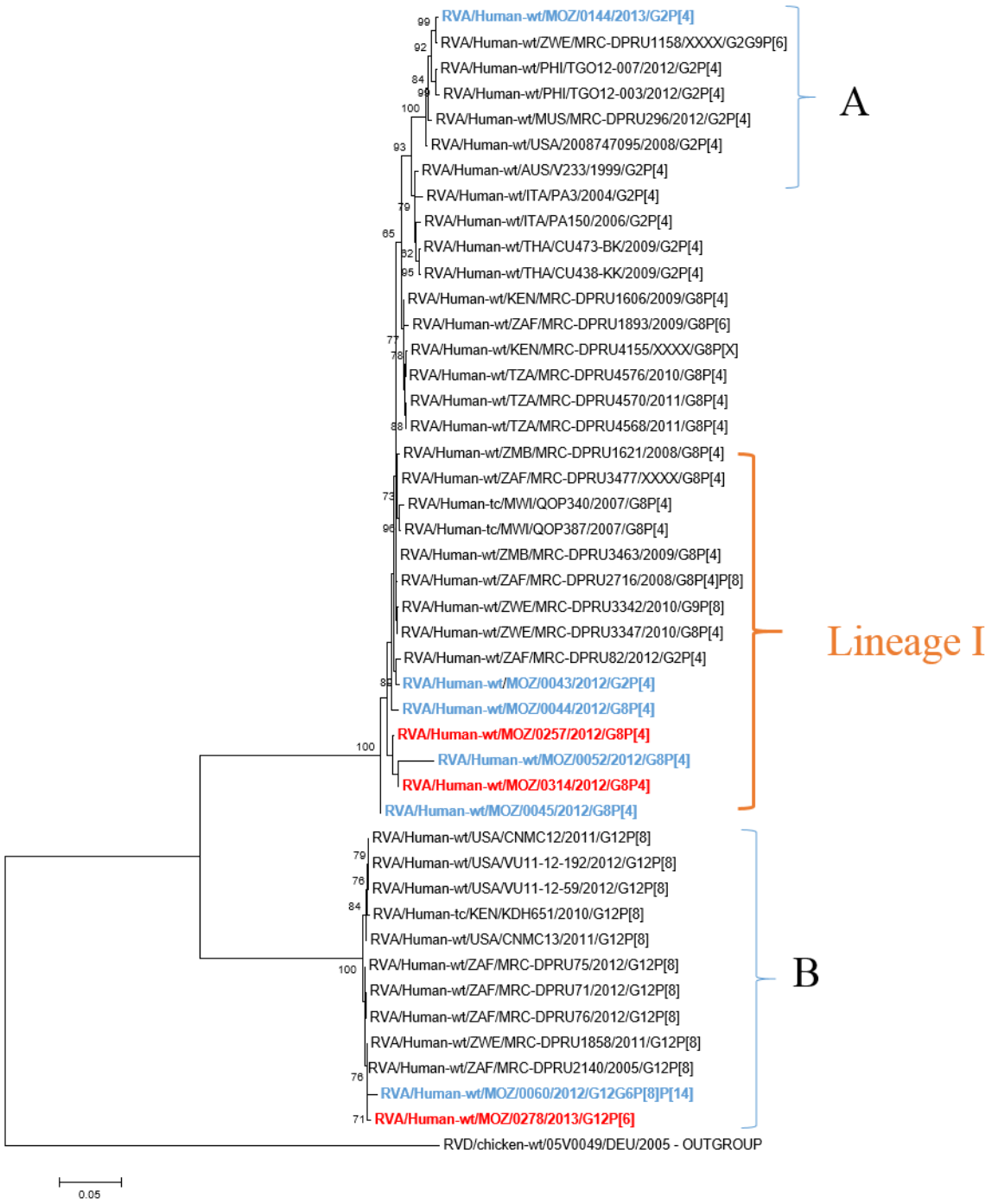


Figure 3.7: Phylogenetic analysis of genome segment 5 of Mozambican strains in relation to other international stains. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in terms of clusters labelled A and B. NSP1 encoding genome segment of strain RVD/Chicken-wt/05V0049/DEU/2005 was used as an outgroup.

3.3.5.6 Phylogenetic analysis of genome segment 6 which encodes for VP6

Strains belonging to lineages I and VII could be identified for genome segment 6. Mozambican strains grouped into five clusters, labelled A to E whose lineages could not be identified in Figure 3.8 In cluster A, Mozambican strains RVA/Human-wt/MOZ/0044/2012/G8P[4], RVA/Human-wt/MOZ/0257/2012/G8P[4] and RVA/Human-wt/MOZ/0314/2012/G8P[4] grouped closely together and also grouped with a single strain from Zambia and another strain from Zimbabwe. In lineage VII (Giammanco et al., 2014), Mozambican strain RVA/Human-wt/MOZ/0045/2012/G8P[4] grouped with one strain from the USA and had 99% nucleotide identity with it. In cluster B, Mozambican strain RVA/Human-wt/0144/2013/G12P[6] grouped closely to a strain from Zimbabwe and to two strains from the Philippines In cluster C, Mozambican strain RVA/Human-wt/MOZ/0043/2012/G2P[4] grouped with one human South African strain and one porcine South African strain and various studies have reported reassortment events between South African human and animal strains (Komoto et al., 2016, Gentsch et al., 2005, Jere et al., 2011b). In addition, the same strain showed a similar grouping for its genome segment 4, which could suggest the possibility of reassortment in the Mozambican strain. In cluster D, the second VP6 encoding genome segment of Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] with I2 genotype did not group with any strain.

In lineage I (Nakagomi et al., 2013), Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with strains from Kenya and Tanzania. However, there is no published information regarding those strains from Kenya and Tanzania. In cluster E, Mozambican strains RVA/Human-wt/MOZ/0060/2012/G12GP[8]P[14] and RVA/Human-wt/MOZ/0278/2012/G12P[6] grouped together with a single strain from Uganda and Kenya and strains from the USA.

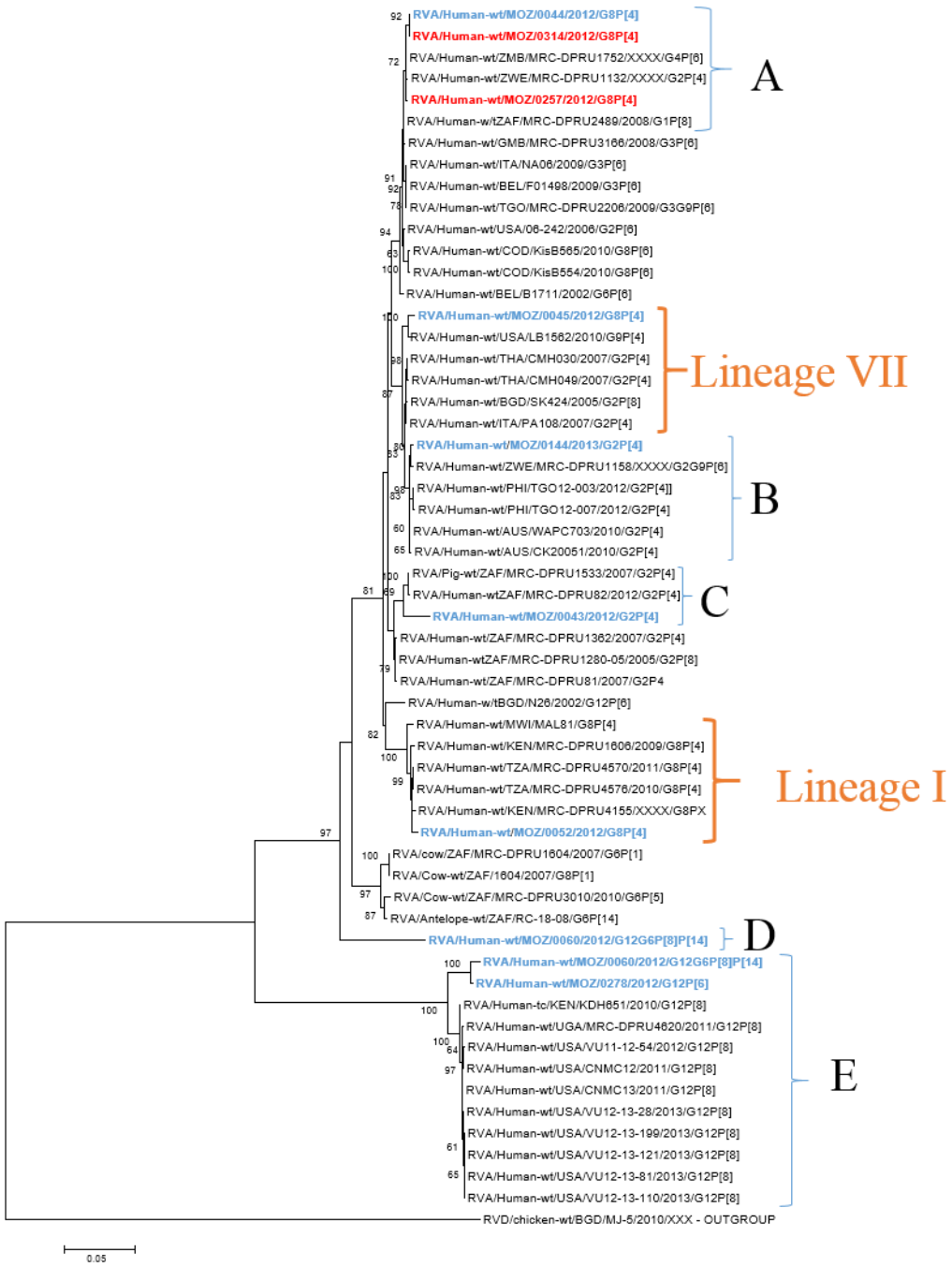


Figure 3.8: Phylogenetic analysis of genome segment 6 of Mozambican strains in relation to other international stains. MEGA 5.2 was used to construct the tree and the evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in terms of clusters labelled A to E. Strain RVD/chicken-wt/BGD/MJ-5/2010/XXX was used as an outgroup.

3.3.5.7 Phylogenetic analysis of genome segment 7 which encodes for NSP3

In genome segment 7 phylogenetic analysis, lineage I like strains were identified in six Mozambican strains (Nakagomi et al., 2013). The remaining three Mozambican strains grouped into two clusters as indicated in Figure 3.9. In lineage I, Mozambican strains RVA/Human-wt/MOZ/0044/2012/G8P[4], RVA/Human-wt/MOZ/0045/2012/G8P[4], RVA/Human-wt/MOZ/0043/2012/G2P[4], RVA/Human-wt/MOZ/0257/2012/G8P[4] and RVA/Human-wt/MOZ/0314/2012/G8P[4] grouped with each other indicating a high similarity between their sequences. In addition, in lineage I, Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with two strains from Kenya and two strains from Tanzania.

In cluster A, Mozambican strain RVA/Human-wt/MOZ/0144/2013/G2P[4] grouped with two strains from Zimbabwe, two from the Philippines and one from India. Strains in cluster B have a Wa-like genotype. In this cluster, two Mozambican strains RVA/Human-wt/MOZ/0060/2012/G12GP[8]P[14] and RVA/Human-wt/MOZ/0278/2012/G2P[4] grouped closest to each other with a 100% bootstrap support. The two Mozambican strains grouped with a Zimbabwean and two South African strains, as well as two porcine strains from South Africa. Phylogenetic analysis of genome segment 7, therefore, shows that rotavirus strains circulating in Mozambique are similar to strains circulating in southern Africa especially Zimbabwe and South Africa.

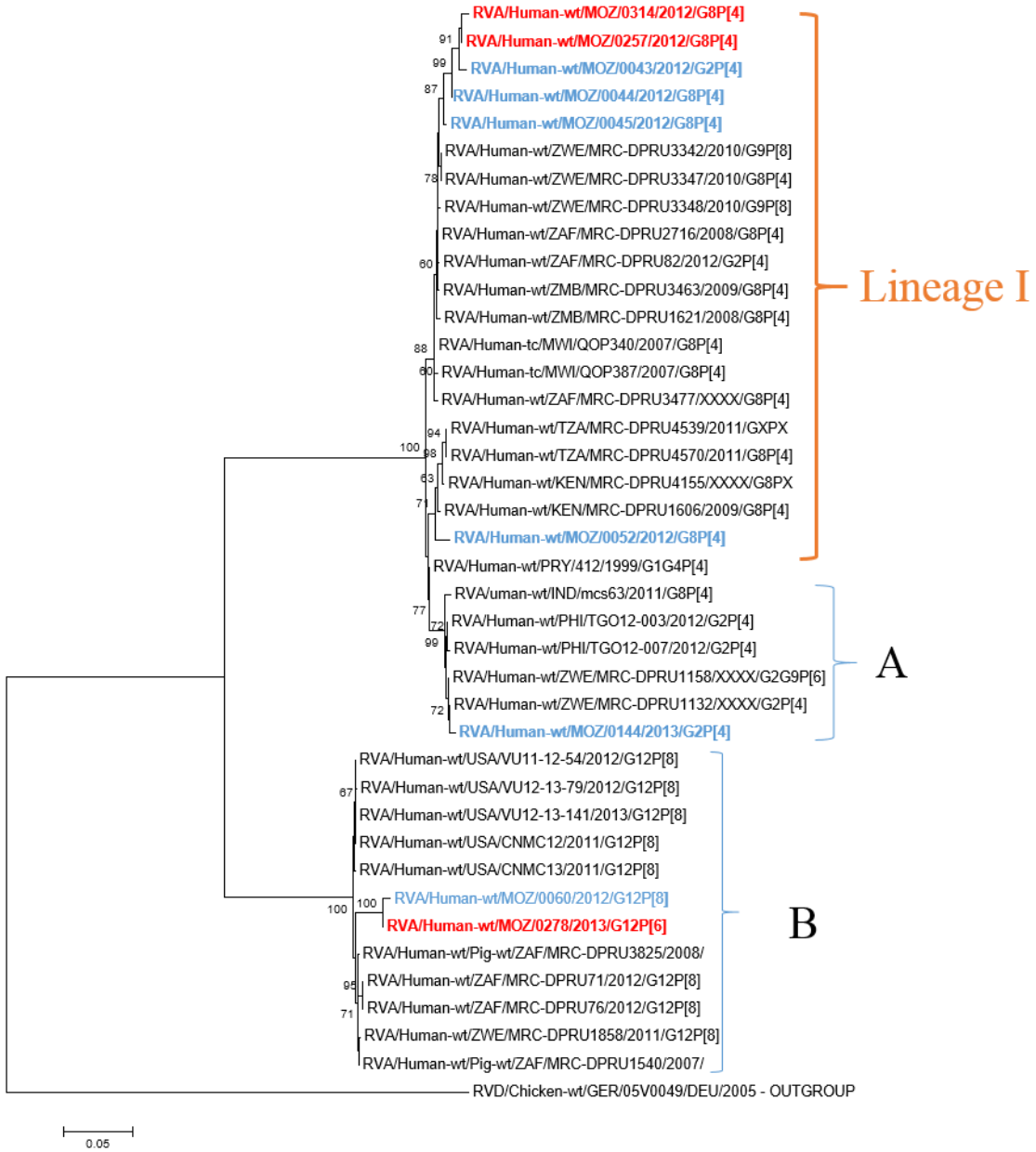


Figure 3.9: Phylogenetic analysis of genome segment 7 of Mozambican strains in relation to other international stains. MEGA 5.2 was used to construct the tree and the evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. The evolutionary distances were computed using the Jukes-Cantor method and are in the units of the number of base substitutions per site. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in terms of clusters labelled A to B. Strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup.

3.3.5.8 Phylogenetic analysis of genome segment 8 which encodes for NSP2

From Figure 3.10, Mozambican strains grouped in lineage I (Nakagomi et al., 2013) and also in three separate clusters labelled A to C. In cluster A, Mozambican strain RVA/Human-wt/MOZ/0144/2012/G2P[4] grouped with four strains from the Philippines and had 99% nucleotide identity to these strains. In addition, the same Mozambican strain also grouped with a strain from Australia RVA/Human-wt/AUS/WAP703/2010/G2P[4]. In cluster B, Mozambican strain RVA/Human-wt/MOZ/0043/2012/G2P[4] grouped with strains from South Africa, Australia, Zambia, Swaziland and Ethiopia.

In cluster C, Mozambican strain RVA/Human-wt/MOZ/0052/2012/G2P[4] grouped with one Kenyan strain RVA/Humna-wt/KEN/MRC-DPRU1606/2009/G8P[4]. In cluster D, Mozambican strains RVA/Human-wt/MOZ/0278/2013/G12P[6] and RVA/Human-wt/MOZ/0060/G12P[8] grouped with strains from South Africa, the USA, Uganda, Kenya, India and Australia. In lineage I, Mozambican strains; RVA/Human-wt/MOZ/0045/2012/G2P[4], RVA/Human-wt/MOZ/0044/2012/G2P[4], RVA/Human-wt/MOZ/0314/2012/G2P[4], RVA/Human-wt/MOZ/0278/2012/G2P[4] and RVA/Human-wt/MOZ/0257/2012/G2P[4] clustered with strains from Zambia, Zimbabwe, Malawi and South Africa.

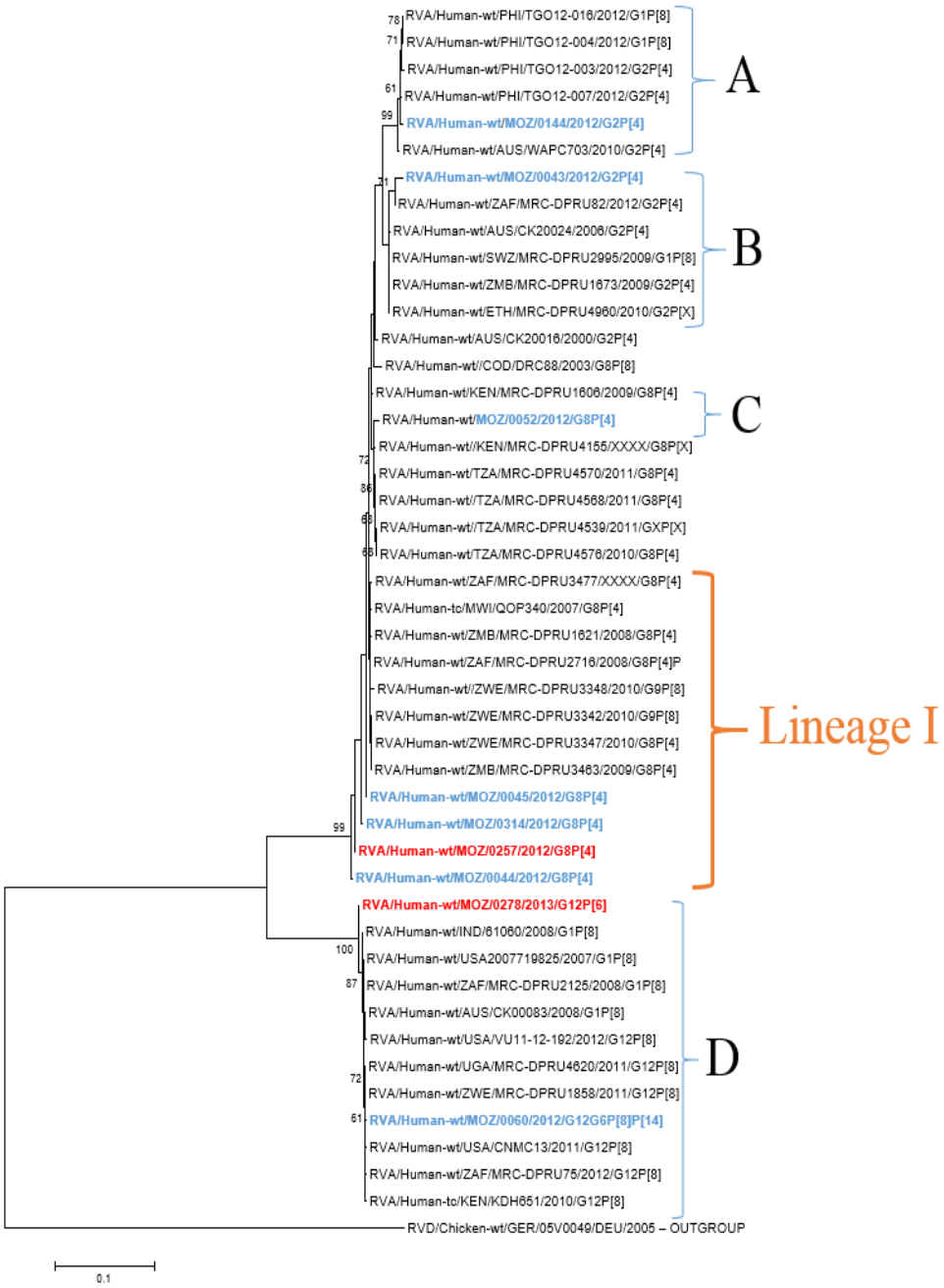


Figure 3.10: Phylogenetic analysis of genome segment 8 of Mozambican strains in relation to other international stains. The evolutionary history was inferred using the Neighbor-Joining method. Evolutionary distances were computed using the Jukes-Cantor method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. Lineages which could be identified are indicated in brackets and sequences which their 5lineages could not be identified are described in terms of clusters labelled A to D. Strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup.

3.3.5.9 Phylogenetic analysis of genome segment 9 which encodes for VP7

In Figure 3.11 genome segment 9 sequences grouped into 4 lineages, I-a, II-d, III (Mullick et al., 2014b, Komoto et al., 2014) and IV and VI (Mullick et al., 2014a),. In lineage VI, Mozambican strains RVA/Human-wt/MOZ/0314/2012/G8P[4], RVA/human-wt/MOZ/0257/G8P[4] and RVA/human-wt/0044/2012/G8P[4] grouped closer with each other indicating high similarity between them. Another Mozambican strain RVA/Human-wt/0045/2012/G8P[4] grouped with a strain from Zambia. Strain RVA/Human-wt/0052/2012/G8P[4] grouped with two South African strains, one Zambian strain and one Malawian strain in lineage I-a. In lineage II-d, Mozambican strain RVA/Human-wt/MOZ/0060/G12G6P[8]P[14] grouped with a South African strain isolated from an antelope. In lineage III, Mozambican strain RVA/Human-wt/MOZ/0278/2012/G12P[6] grouped with an Indian strain and the two strains share a 99% nucleotide identity. In addition, Mozambican sample RVA/Human-wt/MOZ/0060/G12G6P[8]P[14] grouped with one strain from South Africa, one from Zimbabwe and three Australian strains. In lineage IV, Mozambican strain RVA/Human-wt/MOZ/0144/2013/G2P[4] grouped with a strain from Zimbabwe and a strain from Mauritius while sample RVA/Human-wt/MOZ/0043/2012/G2P[4] grouped with two strains from Hungary and one strain from Zambia.

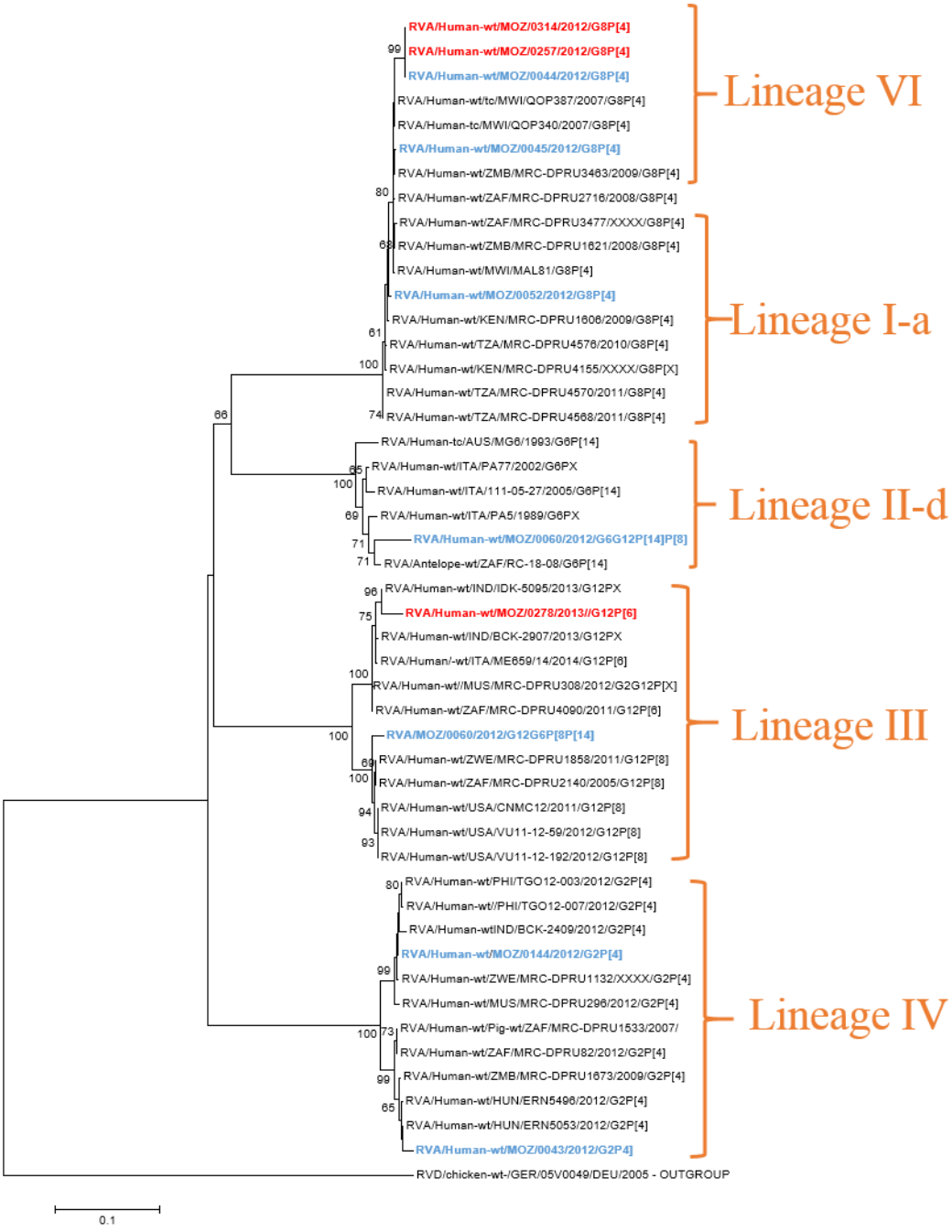


Figure 3.11: Phylogenetic analysis of genome segment 9 of Mozambican strains in relation to other international stains. MEGA 5.2 was used to construct the tree and the evolutionary history inferred using the Neighbor-Joining method and the Juke-Cantor method was used to compute evolutionary distances. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in cluster A. Strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup.

3.3.5.10 Phylogenetic analysis of genome segment 10 which encodes for NSP4

In the phylogenetic analysis of genome segment 10, Mozambican strains grouped into three separate clusters labelled A to C and two strains grouped in lineage IV (Nakagomi et al., 2013) (Figure 3.12). Another cluster comprised of strains with Wa-like genotype. In cluster A, four Mozambican strains RVA/Human-wt/MOZ/0044/2012/G8P[4], RVA/Human-wt/MOZ/0043/2012/G2P[4], RVA/Human-wt/MOZ/0314/2012/G8P[4] and RVA/Human-wt/MOZ/0257/2012/G8P[4] grouped with each other and with a Russian strain RVA/Human-wt/RUS/Omsk08-475/2008/G2P[4]. In cluster B, Mozambican strains RVA/Human-wt/0060/2012/G12G6P[8]P[14] and RVA/Human-wt/0278/2012/G12P[6] grouped with a strain from Uganda and Zimbabwe and two additional strains from South Africa. In cluster C, Mozambican strain RVA/Human-wt/0144/G2P[4] grouped with an Indian strain and formed a distinct grouping compared to its other genome segments, suggesting that segment 10 of this strain might be following a different evolutionary pattern as compared to its other 10 genome segments. In lineage IV, Mozambican strain RVA/Human-wt/0045/2012/G8P[4] grouped with two strains from Zambia, two from Zimbabwe and one from South Africa. In addition, Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with two strains from Kenya and three other strains from Tanzania, which is the same grouping as the other genome segments.

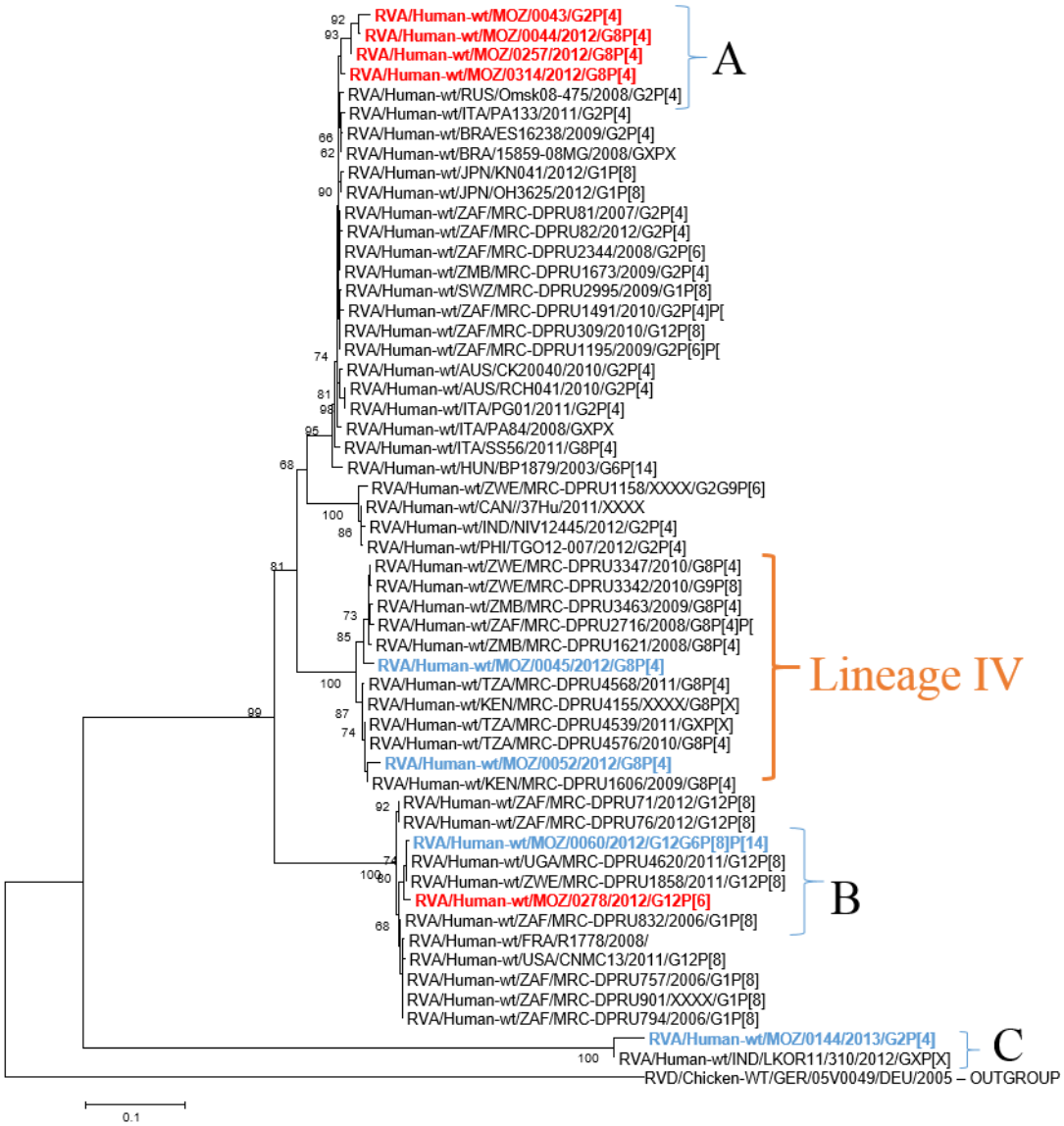


Figure 3.12: Phylogenetic analysis of genome segment 10 of Mozambican strains

MEGA 5.2 was used to construct the tree and the evolutionary history inferred using the Neighbor-Joining method and the evolutionary distances computed using the Jukes-Cantor method. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in cluster A to C. Strain RVD/Chicken-wt/GER/05V0049/DEU/2005 was used as an outgroup.

3.3.5.11 Phylogenetic analysis of genome segment 11 which encodes for NSP5/6

In the phylogenetic analysis of genome segment 11 (Figure 3.13), Mozambican strains clustered in three separate clusters, A to C, and in lineage III (Giammanco et al., 2014). In cluster A, Mozambican strain RVA/Human-wt/MOZ/0043/2012/G2P[4] grouped with a strain from South Africa and had a 99% nucleotide identity with the South African strain. Additionally, in cluster A, Mozambican strains RVA/Human-wt/MOZ/0257/2012/G8P[4], RVA/Human-wt/MOZ/0314/2012/G8P[4] and RVA/Human-wt/MOZ/0044/2012/G8P[4] grouped closely together and with a strain from Italy RVA/Human-wt/ITA/S69/2011/G8P[4] and another strain from Australia RVA/Human-wt/AUS/RCH041/2010/G2P[4].

In lineage III, Mozambican strain RVA/Human-wt/MOZ/0144/2013/G2P4], grouped with a strain from Italy RVA/Human-wt/ITA/PA133/2011/G2P[4], one from Australia RVA/Human-wt/AUS/WAPCO703/2010/G2P[4], and two stains from the Philippines. The Mozambican strain in lineage III grouped with the two strains from the Philippines for segments, 1, 2, 4, 5, 6, 7 and 8 indicating a close similarity between these strains. In cluster B, Mozambican strains RVA/Human-wt/MOZ/0052/2012/G8P[4] and RVA/Human-wt/MOZ/0045/2012/G8P[4] grouped with two strains from Kenya and one strain from Tanzania. Mozambican strain RVA/Human-wt/MOZ/0052/2012/G8P[4] grouped with the same strains from Tanzania and Kenya for the majority of its other genome segments, which confirms the close similarity between these strains. In cluster C, Mozambican strain RVA/Human-wt/MOZ/0278/2012/G12P[6] grouped with one strain from Bangladesh and one the USA, while Mozambican strain RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] grouped with two strains from China, one from Kenya and one from Thailand.

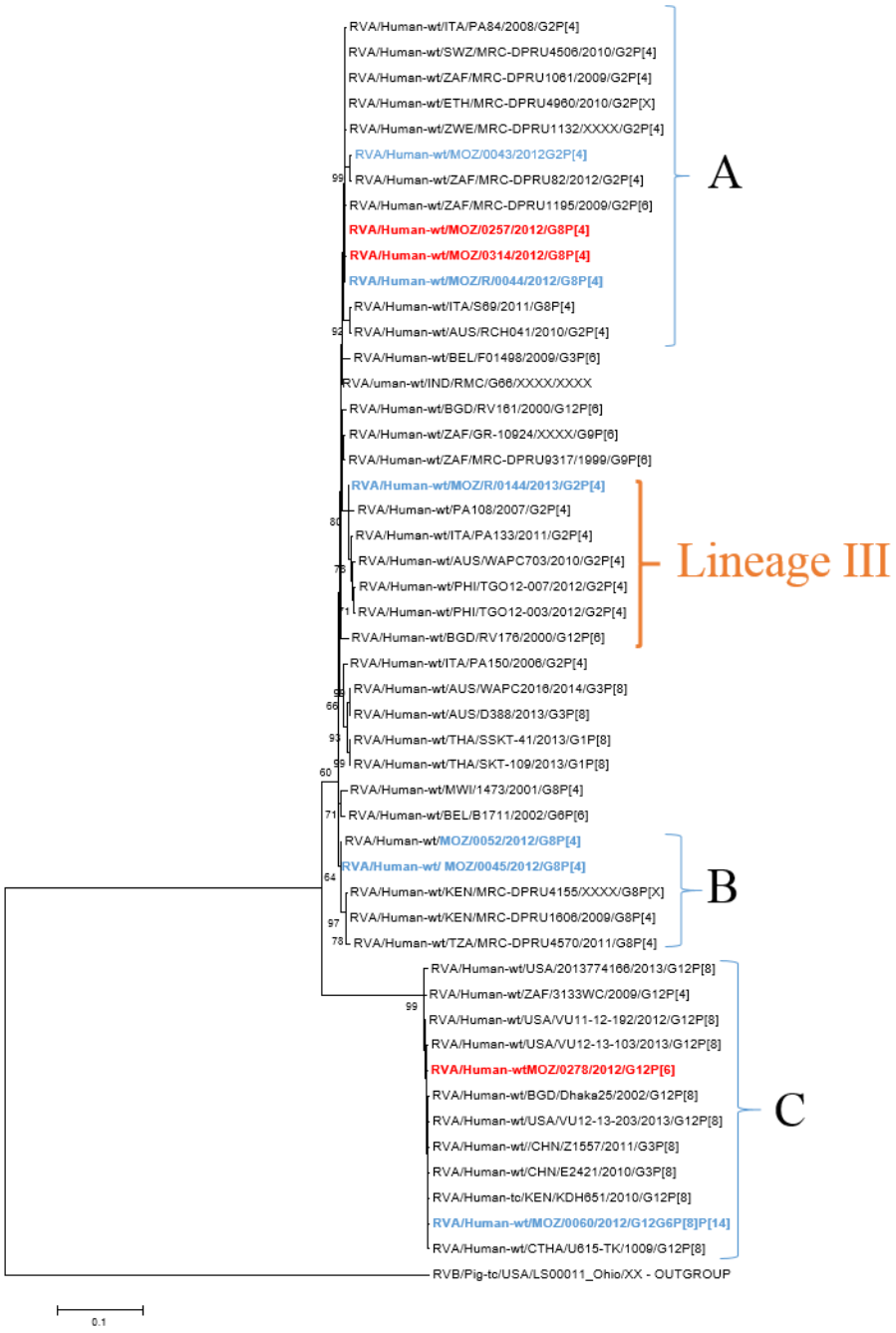


Figure 3.13: Phylogenetic analysis of genome segment 11 of Mozambican rotavirus strains.

MEGA 5.2 was used to construct the tree and the evolutionary history inferred using the Neighbor-Joining method. The evolutionary distances computed using the Kimura-2-parameter. The tree is drawn to scale. Bootstrap values of less than 60 % are not shown. Strains from Manhiça are indicated in blue while strains from Mavalane are indicated in red. Lineages which could be identified are indicated in brackets and sequences which their lineages could not be identified are described in cluster A to C. Strain RVB/Pig-tc/USA/LS00011/XXX as an outgroup.

3.4 Conclusion

In this chapter, the aims of the study were achieved using NGS technology to characterise rotavirus strains. Results from this study also confirmed that NGS is a powerful tool that can be used as an alternative to Sanger sequencing and RT-PCR. Next generation sequencing technologies, specifically Illumina Myseq platform used in the current study is highly automated thus minimising human errors that come with less automated sequencing technologies. Results, obtained from NGS sequencing methods, are especially useful for phylogenetic studies because of the high depth of sequencing.

As observed in chapter 2 with G and P type analyses alone, no differences in the strains circulating in the rural (Manhiça) and urban (Mavalane) sites were observed. From phylogenetic analysis of the 11 genome segments of Mozambican rotavirus strains it was evident that Mozambican strains grouped with strains mostly from neighbouring countries in sub-Saharan Africa. These countries include; South Africa, Malawi, Kenya, Tanzania, and Uganda. Mozambican strains also grouped with strains from Asian countries more specifically Bangladesh and India. Mozambican strains also grouped with a few strains from the United States of America, Australia and Germany. Genotyping PCR as performed by the Mozambican group indicated that some Mozambican strains were mixed infections but no evidence of this was detected with NGS. These results were further confirmed by colleagues in the same group at the University of the Free State (UFS), who did RT-PCR cloning on some of the Mozambican samples previously typed as mixed and analysis of the various clones did not indicate mixed infections.

Next generation sequencing revealed that some Mozambican strains grouped with animal strains in the phylogenetic analysis suggesting possible interspecies transmission between human and animal rotavirus strains. In addition, Mozambican strain, RVA/Human-wt/MOZ/0052/2012/G8P[4] consistently grouped with the same strains from Tanzania and Kenya for all genome segments indicating similarity between these strains. This observation suggests that this strain is likely an introduction into Mozambique from East Africa or continuously circulating in East Africa. Mozambican strain RVA/Human-wt/0144/2013/G2P[4] consistently grouped with similar strains from the Philippines for most of its genome segments also suggesting the possibility of an introduction of this strain from Asia into Mozambique. For genome segment 10, it showed a distinct grouping pattern and grouped with an Indian strain which suggests a different evolutionary route for segment 10 of this strain.

Mozambican strains, RVA/Human-wt/0043/2012/G2P[4], RVA/Human-wt/0044/2012/G8P[4], RVA/Human-wt/0257/2012/G8P[4] and RVA/Human-wt/0314/2013/G8P[4] grouped together for most of their genome segments which suggests a similar evolutionary pathway between these viruses. Mozambican rotavirus strains indicate some level of diversity as they group in different lineages of the same genotype and do not exclusively group amongst themselves.

Next generation sequencing results can also be used to evaluate the efficiency of genotyping primers used in genotyping PCR. Useful information regarding which primers to continue using and which need modification can be obtained. From the evaluation done on genotyping primer binding regions used by the Mozambican group it is clear that G12 and P[8] primers need to be modified so that more accurate results can be obtained. When conducting rotavirus surveillance studies over a large area, genotyping PCR is one of the cheapest methods that can be used therefore it is important that more accurate primers are used.

In conclusion, Mozambican group A rotaviruses are similar to strains circulating in low-income countries in Africa and Asia as a result of the introduction of these strains into Mozambique possibly through migration of people between these countries. Next generation sequencing also serves as a more accurate method that can be used for the confirmation of genotypes previously determined by other methods as well as for in depth analysis or characterisation of strains.

Chapter 4

Concluding Remarks and Way Forward

Rotavirus characterisation studies have evolved from characterisation of strains using e5lectropherotyping to serological characterisation with monoclonal antibodies against VP6, VP4 and VP7 (Titone et al., 1986, Gentsch et al., 2005). Over time serological characterisation became uncommon as not all antibodies were available for all established rotavirus groups. In addition, cross-reactivity of antibodies between different virus strains became a problem (Kang et al., 1993), and serotyping was replaced by characterisation into P and G genotypes using Sanger sequencing and PCR. With the advent of NGS technologies in the early 2000s (Barzon et al., 2011), rotavirus whole genome characterisation studies have become more common. In this study, Mozambican rotavirus strains that were previously typed using genotyping PCR were typed into G and P genotypes using RT-PCR and Sanger sequencing. In addition, nine selected rotavirus strains were analysed with NGS technology using the Illumina Miseq platform.

Extraction of RNA from stool samples using TRI-reagent (a phenol-chloroform method) was successful for most samples and other studies have also reported successful results using the same method (Jere et al., 2012). In addition, Mozambican samples negative for RNA extraction on agarose gels were successfully amplified. This observation suggested that the extracted RNA concentration was too low to be detected by ethidium bromide-stained agarose gel electrophoresis. This is an indication of success obtained using phenol-chloroform method as even lower RNA concentrations can be extracted. For samples which were negative for RNA extraction, there could have been inhibitors present in the samples which led to extraction failure. Alternative extraction methods could be used for such samples to see if the same results are obtained

Amplification of synthesised cDNA from extracted RNA using KAPA Hi-Fi DNA polymerase and primer sets for VP4 (Beg 9, End9) and VP7 (Con 2, Con 3) encoding genome segments was also successful. Success of PCR amplification for the two genome segments was determined by the presence of clear distinct bands at the expected sizes of 876 and 1062 bp for VP4 and VP7, respectively, after analysis of PCR products on agarose gel.

Genotyping of Mozambican strains with RT-PCR and Sanger sequencing was successful for most samples. For samples which sequencing failed it could be due to impure DNA template. The DNA might have contaminants such as RNA, chromosomal DNA, traces of organic chemicals such as chloroform and ethanol. Contaminants could also come from excess PCR reagents used such dNTPs, Taq polymerase or even primers therefore it is important to make sure that for each PCR reaction sufficient concentrations of reagents are used. Impure DNA template should be cleaned-up using spin columns before sequencing. Sequencing may also fail due to lower template concentrations and in such cases template concentrations should be 5increased. Genotypes were determined by submitting Sanger sequencing results to BLASTn analysis.

With RT-PCR and Sanger sequencing, genotypes of 28 of 30 samples typed by genotyping PCR were confirmed. Of the two remaining samples, one was typed as GNTP[8] with genotyping PCR while with Sanger sequencing it was typed as G12P[6]. The other sample could not be typed with genotyping PCR but was typed as G2P[4] with Sanger sequencing. Following BLASTn analysis, G2P[4] and G12P[6] appeared to be the most frequently detected genotypes and accounted for 44% and 37%, respectively. Phylogenetic analysis of Mozambican strains showed that Mozambican rotavirus strains are similar to globally circulating strains. Results obtained in this study proved that genotyping by RT-PCR and Sanger sequencing to be more efficient as compared to genotyping PCR.

Whole-genome characterisation of nine Mozambican samples with NGS was successful. Using genotyping PCR, seven of nine samples were typed as mixed but with whole-genome characterisation they were shown to contain single genotypes. The two remaining samples were typed as non-mixed by genotyping PCR and Sanger sequencing. Whole genome characterisation of the two samples confirmed the genotype obtained for one sample and interestingly, the other sample previously typed as G12P[8] was typed as G12G6P[8][14] with NGS. With NGS, seven of nine samples had DS-1-like backbone with no sign of intergenogroup reassortment and the other two samples had a Wa-like genotype. Whole-genome characterisation of sample RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14], w5hich typed as G12P[8] by genotyping PCR and Sanger sequencing revealed a mixture of a porcine rotavirus strain with a Wa-like genotype strain. In addition, sample RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] was found to have a second VP1 encoding genome segment with a DS-1-like R2 genotype. This could mean that sample RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] was a mixture of more than two rotavirus strains or it

was a mixture with a human-bovine reassortment strain. Interestingly, the study did not indicate a difference in strains circulating between the two sites, Manhiça and Mavalane.

In this study, whole genome characterisation of Mozambican rotavirus strains as expected proved to be a good tool for characterising rotaviruses. Both *de novo* assembly and assembly against reference rotavirus strains of raw data worked equally well as all 11 genome segment could be identified. For most samples, nearly full length ORF for the 11 genome segments were obtained. The advantages of whole-genome characterisation over characterisation with binary typing into P and G genotypes using Sanger sequencing is that information about all 11 genome segments is known. With binary typing, only information about genome segment 4 and 9 is known. Information about reassortment events or recombination is missed. In this study, RT-PCR and Sanger sequencing of sample RVA/Human-wt/MOZ/0060/2012/G12G6P[8]P[14] is an example of RT-PCR and Sanger sequencing limitation. The aims and objectives of the current study were partially achieved and phylogenetic analysis of Mozambican strains indicated that these strains are similar to strains circulating in southern Africa. This would be expected because of similar living conditions of between people in low-income countries, and the frequent migration of people to and from these countries.

For future work, alternative RNA extraction methods could be used to determine if different results will be obtained. New genotyping primers for G12, G9 and P[8] strains can be designed or alternative primer sets can be evaluated for genotyping PCR. This study was conducted before the introduction of rotavirus vaccine RotaTeq® in Mozambique which was introduced in November 2015. The results obtained in this study will, therefore, be important for comparison of strains circulating in Mozambique before introduction of the vaccine and results obtained after introduction of the vaccine. Whole-genome characterisation of rotavirus strains using NGS is a quick and easy method which the Mozambican group could adopt for their future surveillance studies.

References

Adah, M. I., & Wade, A. (2001). Molecular Epidemiology of Rotaviruses in Nigeria : Detection of Unusual Strains with G2P[6] and G8P[1] Specificities. *Journal of Clinical Microbiology*, *39*(11), 3969–3975.

Afrikanova, I., Fabbretti, E., Miozzo, M. C., & Burrone, O. R. (1998). Rotavirus NSP5 phosphorylation is up-regulated by interaction with NSP2. *Journal of General Virology*, *79*(11), 2679–2686.

Ahmed, S., Kabir, A R. M. L., Rahman, A., Hussain, M., Phil, M., & Khatoon, S. (2009). Severity of Rotavirus Diarrhea in Children : One Year Experience in a Children Hospital of Bangladesh. *Iranian Journal of Pediatrics*, *19*(2), 108–116.

Alexander, J. J., & Hay, I. A. N. T. (1986). Rotavirus-Associated Gastroenteritis, *Journal of Clinical Microbiology*, *23*(5), 992–994.

Armah, G. E., Pager, C. T., Asmah, R. H., Anto, F. R., Oduro, A. R., Binka, F., & Steele, D. (2001). Prevalence of Unusual Human Rotavirus Strains in Ghanaian Children. *Journal of Medical Virology*, (71), 67–71.

Armah, G. E., Sow, S. O., Breiman, R. F., Dallas, M. J., Tapia, M. D., Feikin, D. R., Neuzil, K. 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), 606–614.

Arnold, M. M., & Patton, J. T. (2011). Diversity of interferon antagonist activities mediated b5y NSP1 proteins of different rotavirus strains. *Journal of Virology*, *85*(5), 1970–9.

Attoui, H., Billoir, F., Cantaloube, J. F., Biagini, P., De Micco, P., & De Lamballerie, X. (2000). Strategies for the sequence determination of viral dsRNA genomes. *Journal of Virological Methods*, *89*(1-2), 147–158.

Barzon, L., Lavezzo, E., Militello, V., & Toppo, S., (2011). Applications of next-generation sequencing technologies to diagnostic virology. *International Journal of Molecular Sciences*, *12*(11),

Bican, P., Cohen, J., Charpilienne, A., & Scherrer, R. (1982). Purification and characterization of bovine rotavirus cores. *Journal of Virology*, *43*(3), 1113–1117.

Bishop, R. (2009). Discovery of rotavirus: Implications for child health. *Journal of Gastroenterology and Hepatology*, *24 Suppl 3*, S81–S85.

Bishop, R., Davidson, G. P., Holmes, I. H., & Ruck, B. J. (1974). detection of a new virus by electron microscopy of fæcal extracts from children with acute gastroenteritis. *The Lancet*, *303*(7849), 149–151.

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

Bresee, J. S., Glass, R. I., Ivanoff, B., & Gentsch, J. R. (1999). Current status and future priorities for rotavirus vaccine development, evaluation and implementation in developing countries. *Vaccine*, *17*(18), 2207–2222.

Brinkman, F. S. L., & Leipe, D. D. (2001). *Phylogenetic Analysis*. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins* (Vol. 43).

Chaimongkol, N., Khamrin, P., Malasao, R., Thongprachum, A., Ushijima, H., & Maneekarn, N. (2012). Genotypic linkages of gene segments of rotaviruses circulating in pediatric patients with acute gastroenteritis in Thailand. *Infection, Genetics and Evolution*, *12*(7), 1381–1391.

Conceição-Neto, N., Zeller, M., Heylen, E., Lefrère, H., Mesquita, J. R., & Matthijnssens, J. (2015). Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. *Virology Journal*, *12*, 79. doi:10.1186/s12985-015-0305-5

Cook, S. M., Glass, R. I., LeBaron, C. W., & Ho, M. S. (1990). Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*, *68*(2), 171–177.

Coulson, B. S., Unicomb, L. E., Pitson, G. A, & Bishop, R. F. (1987). Simple and specific enzyme immunoassay using monoclonal antibodies for serotyping human rotaviruses. *Journal of Clinical Microbiology*, *25*(3), 509–15.

de Deus, N., João, E., Cuamba, A., Cassocera, M., Luís, L., Acácio, S., Mandomando, I., Augusto, O., & Page, N. (2017). Epidemiology of Rotavirus Infection in Children from a Rural and Urban Area, in Maputo, Southern Mozambique, before Vaccine Introduction. *Journal of Tropical Pediatrics* doi: 10.1093/tropej/fmx032 [Epub ahead of print]

Dennehy, P. H. (2008). Rotavirus vaccines: An overview. *Clinical Microbiology Reviews*, *21*(1), 198–208.

Desselberger, U. (1996). Genome rearrangements of rotaviruses. *Advances in Virus Research*, *46*, 69–95.

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

Desselberger, U., & Gray, J. (2013). Viral gastroenteritis. *Medicine*, *41*(12), 700–704.

Desselberger, U., Iturriza-Gómara, M., & Gray, J. J. (2001). Rotavirus epidemiology and surveillance. *Novartis Foundation Symposium*, *238*, 125–47.

Donato, C. M., Zhang, Z. A., Donker, N. C., & Kirkwood, C. D. (2014). Characterization of G2P[4] rotavirus strains associated with increased detection in Australian states using the RotaTeq vaccine during the 2010-2011 surveillance period. *Infection, Genetics and Evolution*, *28*, 398–412.

El Sherif, M., Esona, M. D., Wang, Y., Gentsch, J. R., Jiang, B., Glass, R. I., Abou Baker, S., Klena, J. D. (2011). Detection of the first G6P[14] human rotavirus strain from a child with diarrhea in Egypt. *Infection, Genetics and Evolution*, *11*(6), 1436–1442.

Elliott, E. J. (2007). Acute gastroenteritis in children. *British Medical Journal*, *334*(7583), 35–40.

Ericson, B. L., Graham, D. Y., Mason, B. B., Hanssen, H. H., & Estes, M. K. (1983). Two types of glycoprotein precursors are produced by the simian rotavirus SA11. *Virology*, *127*(2), 320–332.

Esona, M. D., Armah, G. E., & Steele, A. D. (2010). Rotavirus VP4 and VP7 genotypes circulating in Cameroon: Identification of unusual types. *The Journal of Infectious Diseases*, *202 Suppl*(Suppl 1), S205–11

Esona, M. D., Banyai, K., Foytich, K., Freeman, M., Mijatovic-Rustempasic, S., Hull, J., Gentsch, J. R. (2011). Genomic characterization of human rotavirus G10 strains from the

African Rotavirus Network: Relationship to animal rotaviruses. *Infection, Genetics and Evolution*, *11*(1), 237–241.

Esteban, L. E., Temprana, C. F., M. H., Glikmann, G., & Castello, A. A. (2013). Antigenicity and immunogenicity of rotavirus VP6 protein expressed on the surface of lactococcus lactis. *BioMed Research International*, *2013*.

Estes MK, Greenberg HB. Rotaviruses. In: Knipe DM, Howley PM, editors. Fields Virology. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013. p. 1347–95.

Estes, M. K., & Cohen, J. (1989). Rotavirus gene structure and function. *Microbiological Reviews*, *53*(4), 410–49.

Fakruddin & Chowdhury, A. (2012). Pyrosequencing-An Alternative to Traditional Sanger Sequencing, *American journal of Biochemistry and Biotechnology*, *8*(1), 14–20.

Fang, Z. Y., Ye, Q., Ho, M. S., Dong, H., Qing, S., Penaranda, M. E., Glass, R. I. (1989). Investigation of an outbreak of adult diarrhea rotavirus in China. *The Journal of Infectious Diseases*, *160*(6), 948–953.

Flewett, T. H., Davies, H., Bryden, A. S., & Robertson, M. J. (1974). Acute gastroenteritis associated with reovirus-like particles. *Journal of Clinical Pathology*, *27*, 608–614.

Gómez, M. M., Carvalho-Costa, F. A., de Mello Volotão, E., Rose, T. L., da Silva, M. F. M., Fialho, A. M., Leite, J. P. G. (2014). Prevalence and genomic characterization of G2P[4] group A rotavirus strains during monovalent vaccine introduction in Brazil. *Infection, Genetics and Evolution*, *28*, 486–494.

Gentsch, J. R., Glass, R. I., Woods, P., Gouvea, V., Gorziglia, M., Flores, J., Bhan, M. K. (1992). Identification of group A rotavirus gene 4 types by polymerase chain reaction. *Journal of Clinical Microbiology*, *30*(6), 1365–1373.

Gentsch, J. R., Laird, A. R., Bielfelt, B., Griffin, D. D., Banyai, K., Ramachandran, M., Glass, R. I. (2005). Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *The Journal of Infectious Diseases*, *192 Suppl* (Suppl 1), S146–S159.

Ghosh, S., Paul, S. K., Hossain, M. A., Alam, M. M., Ahmed, M. U., & Kobayashi, N. (2011). Full genomic analyses of two human G2P[4] rotavirus strains detected in 2005: Identification of a caprine-like VP3 gene. *Journal of General Virology*, *92*(5), 1222–1227.

Ghosh, S., Varghese, V., Samajdar, S., Bhattacharya, S. K., Kobayashi, N., & Naik, T. N. (2006). Molecular characterization of a porcine Group A rotavirus strain with G12 genotype specificity. *Archives of Virology*, *151*(7), 1329–1344.

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

González, R. A., Torres-Vega, M. A., López, S., & Arias, C. F. (1998). In vivo interactions among rotavirus nonstructural proteins. *Archives of Virology*, *143*(5), 981–996.

Gouvea, V., Glass, R. I., Woods, P., Taniguchi, K., Clark, H. F., Forrester, B., & Fang, Z. (1990). Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool Polymerase Chain Reaction Amplification and Typing of Rotavirus Nucleic Acid

from Stool Specimens. *Journal of Clinical Microbiology*, *28*(2), 276–282.

Gray, J., Vesikari, T., Van Damme, P., Giaquinto, C., Mrukowics, J., Guarino, A., Dagan., R., Szajewska, H. & Usonis,V. (2008). Rotavirus. *Journal of Pediatric Gastroentorology and Nutrition*, 46:S24-S31.

Hasenack, B. S., Botelho, M. V. J., Lauretti, F., De Melo, F. L., Orlandi, J. M., Linhares, R. E. C., Ueda, M., Nozawa, C. M. (2002). The effect of concanavalin A on the replication of rotavirus (SA-11) in cell culture. *Brazilian Archives of Biology and Technology*, *45*(2), 125–135.

Heaton, P. M., Goveia, M. G., Miller, J. M., Offit, P., & Clark, H. F. (2005). Development of a Pentavalent Rotavirus Vaccine against Prevalent Serotypes of Rotavirus Gastroenteritis. *Journal of Infectious Disease*, *192*(Supplement 1), S17–21.

Heylen, E., Likele, B. B., Zeller, M., Stevens, S., Coster, S. De, Conceica˜oNeto, N., Matthijnssens, J. (2014). Rotavirus Surveillance in Kisangani, the Democratic Republic of the Congo, Reveals a High Number of Unusual Genotypes and Gene Segments of Animal Origin in Non-Vaccinated Symptomatic Children. *PLoS ONE*, *9*(6), e100953.

Hung, T., G. Chen, C. Wang, Z. Chou, T. Chao, W. Ye, H. Yao, and K. Meng. (1983). Rotavirus-like agent in adult non-bacterial diarrhoea in China. The Lancet ii:1078–1079.

Hyser, J. M., Collinson-Pautz, M. R., Utama, B., & Estes, M. K. (2010). Rotavirus disrupts calcium homeostasis by NSP4 viroporin activity. *mBio*, *1*(5), 1–12.

Ide, T., Komoto, S., Higo-Moriguchi, K., Htun, K. W., Myint, Y. Y., Myat, T. W., Taniguchi, K. (2015). Whole genomic analysis of human G12P[6] and G12P[8] rotavirus strains that have emerged in Myanmar. *PLoS ONE*, *10*(5).

Iturriza-Gòmara, M., Green, J. O. N., Brown, D. W. G., Desselberger, U., & Gray, J. J. (2000). Diversity within the VP4 Gene of Rotavirus P [8] Strains : Implications for Reverse Transcription-PCR Genotyping. *Journal of Clinical Microbiology*, *38*(2), 898–901.

Iturriza-Gómara, M., Dallman, T., Bányai, K., Böttiger, B., Buesa, J., Diedrich, S., Gray, J. (2011). Rotavirus genotypes co-circulating in Europe between 2006 and 2009 as determined by EuroRotaNet, a pan-European collaborative strain surveillance network. *Epidemiology and Infection*, *139*(6), 895–909.

Jain, S., Vashistt, J., & Changotra, H. (2014). Rotaviruses: Is their surveillance needed? *Vaccine*, *32*(27), 3367–3378.

Jain, V., Das, B. K., Bhan, M. K., Glass, R. I., Gentsch, J. R., Bhambal, S. S., Thakur, S. (2001). Great diversity of group A rotavirus strains and high prevalence of mixed rotavirus infections in India. *Journal of Clinical Microbiology*, *39*(10), 3524–3529.

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

Jere, K.C., Mlera, L., O'Neill, H. G., Potgieter, A. C., Page, N. A., Seheri, M. L., & van Dijk, A. A. (2011a). Whole genome analyses of African G2, G8, G9, and G12 rotavirus strains using sequence-independent amplification and 454® pyrosequencing. *Journal of Medical Virology*, 83(11):2018-42.

Jere, K. C., Mlera, L., Page, N. A., van Dijk, A. A., & O'Neill, H.G. (2011b). 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):2072-82.

Jere, K. C., Mlera, L., O'Neill, H. G., Peenze, I., & Van Dijk, A. A. (2012). Whole genome sequence analyses of three African bovine rotaviruses reveal that they emerged through multiple reassortment events between rotaviruses from different mammalian species. *Veterinary Microbiology*, *159*(1-2), 245–250.

Kang, S. Y., Benfield, D. A., Gorziglia, M., & Saif, L. J. (1993). Characterization of the neutralizing epitopes of VP7 of the Gottfried strain of porcine rotavirus. *Journal of Clinical Microbiology*, *31*(9), 2291–2297.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120.

Kirkwood, C. D. (2010). Genetic and antigenic diversity of human rotaviruses: potential impact on vaccination programs. *Journal of Infectious Diseases*, *202 Suppl*, S43–8.

Komoto, S., Tacharoenmuang, R., Guntapong, R., Ide, T., Tsuji, T., Yoshikawa, T., Taniguchi, K. (2016). Reassortment of Human and Animal Rotavirus Gene Segments in Emerging DS-1-Like G1P[8] Rotavirus Strains. *Plos One*, *11*(2), e0148416.

Komoto, S., Wandera Apondi, E., Shah, M., Odoyo, E., Nyangao, J., Tomita, M., Taniguchi, K. (2014). Whole genomic analysis of human G12P[6] and G12P[8] rotavirus strains that have emerged in Kenya: Identification of porcine-like NSP4 genes. *Infection, Genetics and Evolution*, *27*, 277–293.

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., Wu, Y., Sow, S.O., Sur, D., Breiman, R.F., Faruque, A.S., Zaidi, A.K., Saha, D., Alonso, P.L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J.B., Omore, R., Oundo, J.O., Hossain, A., Das, S.K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R.A., Antonio, M., Hossain, M.J., Akinsola, A., Mandomando, I., Nhampossa, T., Acácio, S., Biswas, K., O'Reilly, C.E., Mintz, E.D., Berkeley, L.Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R.M., Levine, M.M. (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. *Lancet*, *382*, 209–222.

Lander, E. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921.

Langa, S., Thompson, R., Arnaldo, P., Resque, H. R., Rose, T., Enosse, S. M.,Figueira, M. (2016). Epidemiology of Rotavirus A Diarrhea in Chokwe Southern Mozambique, From February to September 2011. *Journal of Medical Virology*, *88:* 1751–1758.

Lewis, J., Roy, S., Esona, M. D., Mijatovic-Rustempasic, S., Hardy, C., Wang, Y., Bowen, M. D. (2014). Full Genome Sequence of a Reassortment Human G9P[4] Rotavirus Strain. *Genome Announcements*, *2*(6), 9–10.

Liu, M., Mattion, N. M., & Estes, M. K. (1992). Rotavirus VP3 expressed in insect cells possesses guanylyltransferase activity. *Virology*, *188*(1), 77–84.

Loh, J., Zhao, G., Presti, R. M., Holtz, L. R., Finkbeiner, S. R., Droit, L., Virgin, H. W. (2009). Detection of novel sequences related to african Swine Fever virus in human serum and

sewage. *Journal of Virology*, *83*(24), 13019–13025.

Luz, C. R. N. E. da, Mascarenhas, J. D. P., Gabbay, Y. B., Motta, A. R. B., Lima, T. V. R., Soares, L. da S., & Linhares, A. C. (2005). Rotavirus serotypes and electropherotypes identified among hospitalised children in São Luís, Maranhão, Brazil. *Revista Do Instituto de Medicina Tropical de São Paulo*, *47*(5), 287–93.

Maan, S., Rao, S., Maan, N. S., Anthony, S. J., Attoui, H., Samuel, A. R., & Mertens, P. P. C. (2007). Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *Journal of Virological Methods*, *143*(2), 132–139.

Madhi, S. A., Cunliffe, N. A., Steele, D., Witte, D., Kirsten, M., Louw, C., 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.

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

Marthaler, D., Rossow, K., Culhane, M., Goyal, S., Collins, J., Ciarlet, M. (2014). Rotavirus H in United States. *Journal of Infectious Diseases 20*(7), 1195–1198.

Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S. M., Van Ranst, M. (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *Journal of Virology*, *82*(7), 3204–3219.

Matthijnssens, J., Ciarlet, M., McDonald, S. M., Attoui, H., Banyai, K., Brister, J. R., van Ranst, M. (2011). Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Archives of Virology*, *156*(8), 1397–1413.

Matthijnssens, J., Otto, P. H., Ciarlet, M., Desselberger, U., van Ranst, M., & Johne, R. (2012). VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation *Archives of Virology*, *157*(6), 1177–1182.

Mattion, N. M., Mitchell, D. B., Both, G. W., & Estes, M. K. (1991). Expression of rotavirus proteins encoded by alternative open reading frames of genome segment 11. *Virology*, *181*(1), 295–304.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(2), 560–564.

Mertens, P. P. C., Attoui, H., Duncan, R. & Dermody, T. S. (2005). Reoviridae. In Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses, pp. 447–454. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press

Mlera L, O'Neill HG, Jere KC, van Dijk AA (2013). Whole-genome consensus sequence analysis of a South African rotavirus SA11 sample reveals a mixed infection with two close derivatives of the SA11- H96 strain. *Archives of Virology*. 158:1021–1030

Mlera, L., Jere, K. C., Van Dijk, A. A. & O'Neill, H. G. (2011). Determination of the whole-genome consensus sequence of the prototype DS-1 rotavirus using sequence-independent genome amplification and 454® pyrosequencing. *Journal of Virological Methods*, 175, 266-71.

Moon, S., Wang, Y., Shane, A.L., Nguyen, T., Ray, P. & Dennehy, P. (2010). Inhibitory effect

of breast milk on infectivity of live oral rotavirus vaccine candidates. *Journal of Pediatric Infectious Diseases*, *29*, 919-923.

Moon, S.S., Groome, M.J., Velasquez, D.E., Parashar, U.D., Jones, S., Koen, A., van Niekerk, N., Jiang, B., Madhi, S.A. (2015). Pre-vaccination rotavirus serum IgG and IgA are associated with lower immunogenecity of live-oral human rotavirus vaccine in South African infants. *Clinical Infectious Diseases*, *62* (2), 157-165.

Müller, H., & Johne, R. (2007). Rotaviruses: diversity and zoonotic potential--a brief review. *Berliner Und Munchener Tierarztliche Wochenschrift*, *120*(3-4), 108–112.

Mullick, S., Mandal, P., Nayak, M. K., Ghosh, S., De, P., Rajendran, K., Chawla-Sarkar, M. (2014a). Hospital based surveillance and genetic characterization of rotavirus strains in children (<5 years) with acute gastroenteritis in Kolkata, India, revealed resurgence of G9 and G2 genotypes during 2011-2013. *Vaccine*, *32*(S1), A20–A28.

Mullick, S., Mukherjee, A., Ghosh, S., Pazhani, G. P., Sur, D., Manna, B., Chawla-Sarkar, M. (2014b). Community based case-control study of rotavirus gastroenteritis among young children during 2008-2010 reveals vast genetic diversity and increased prevalence of G9 Strains in Kolkata. *PLoS ONE*, *9*(11).

Munford, V., Souza, E. C., Alessandra, T., Caruzo, R., Martinez, M. B., Rácz, M. L., Farmaceuticas, F. D. C. (2007). Serological and Molecular Diversity of Human Rotavirus in São Paulo, *Brazilian Journal of Microbiology*, 459–466.

Mwenda, J. M., Ntoto, K. M., Abebe, A., Enweronu-Laryea, C., Amina, I., Mchomvu, J., Steele, A. D. (2010). Burden and epidemiology of rotavirus diarrhea in selected African countries: preliminary results from the African Rotavirus Surveillance Network. *The Journal of Infectious Diseases*, *202 Suppl*(Suppl 1), S5–S11.

Nakagomi, T., Doan, Y. H., Dove, W., Ngwira, B., Iturriza-Gumara, M., Nakagomi, O., & Cunliffe, N. A. (2013). G8 rotaviruses with conserved genotype constellations detected in Malawi over 10 years (1997-2007) display frequent gene reassortment among strains co-circulating in humans. *Journal of General Virology*, *94*(PART 6), 1273–1295.

Nguyen, T. V., Van, P. Le, Huy, C. Le, & Weintraub, A. (2005). Diarrhea caused by enterotoxigenic Bacteroides fragilis in children less than 5 years of age in Hanoi, Vietnam. *Anaerobe*, *11*(1-2), 109–114.

Nhampossa, T., Mandomando, I., Acacio, S., Quintó, L., Vubil, D., Ruiz, J., Alonso, P. (2015). Diarrheal Disease in Rural Mozambique: Burden, Risk Factors and Etiology of Diarrheal Disease among Children Aged 0–59 Months Seeking Care at Health Facilities. *Plos One*, *10*(5), e0119824.

Nyaga, M. M., Esona, M. D., Jere, K. C., Peenze, I., Seheri, M. L., & Mphahlele, M. J. (2014). Genetic diversity of rotavirus genome segment 6 (encoding VP6) in Pretoria, South Africa. *SpringerPlus*, *3*(179), 1–5.

Page, N.A., de Beer, M.C., Seheri, L.M., Dewar, J.B. & Steele, A.D. (2009). The detection and molecular characterization of human G12 genotypes in South Africa. *Journal of Medical Virology, 1*(1), 106-113.

Parashar, U. D., Bresee, J. S., Gentsch, J. R., & Glass, R. I. (1998). Rotavirus. *Emerging Infectious Diseases*, *4*(4), 561–570.

Parashar, U. D., Burton, A., Lanata, C., Boschi-Pinto, C., Shibuya, K., Steele, D., Glass, R. I.

(2009). Global mortality associated with rotavirus disease among children in 2004. *The Journal of Infectious Diseases*, *200 Suppl* (5), S9–S15.

Parashar, U. D., Gibson, C. J., Bresee, J. S., & Glass, R. I. (2006). Rotavirus and severe childhood diarrhea. *Emerging Infectious Diseases*, *12*(2), 304–306.

Parashar, U. D., Hummelman, E. G., Bresee, J. S., Miller, M. A., & Glass, R. I. (2003). Global illness and deaths caused by rotavirus disease in children. *Emerging Infectious Diseases*, *9*(5), 565–572.

Parra, G. I., Bok, K., Martìnez, M., & Gòmez, J. A. (2004). Evidence of rotavirus intragenic recombination between two sublineages of the same genotype. *Journal of General Virology*, *85*(6), 1713–1716.

Patel, M., Shane, A. L., Parashar, U. D., Jiang, B., Gentsch, J. R., & Glass, R. I. (2009). Oral rotavirus vaccines: how well will they work where they are needed most? *The Journal of Infectious Diseases*, *200*(Suppl 1), S39–S48.

Patel, M. M., Pitzer, V., Alonso, W. J., (2012). Global Seasonality of Rotavirus Diseases. *Journal of Pediatric Infectious Diseases*, *32:134-147*.

Patton, J. T., & Spencer, E. (2000). Genome replication and packaging of segmented double-stranded RNA viruses. *Virology*, *277*(2), 217–225.

Potgieter, A. C., Page, N. A., Liebenberg, J., Wright, I. M., Landt, O., & 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), 1423–1432.

Prasad, B. V. V., Burns, J. W., Marietta, E., Estes, M. K., & Chiu, W. (1990). Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy. *Nature*, *343*(February), 476–479.

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

Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., & Hall, N. (2012). Application of next-generation sequencing technologies in virology. *Journal of General Virology*, *93*(PART 9), 1853–1868.

Rahman, M., Leener, K. De, Goegebuer, T., Wollants, E., Donck, I. Van Der, Ranst, M. Van. (2003). Genetic Characterization of a Novel, Naturally Occurring Recombinant Human G6P[6] Rotavirus. *Journal of Clinical Microbiology*, *41*(5), 2088–2095.

Ramani, S., Iturriza-Gomara, M., Jana, A. K., Kuruvilla, K. A., Gray, J. J., Brown, D. W., & Kang, G. (2009). Whole genome characterization of reassortment G10P[11] strain (N155) from a neonate with symptomatic rotavirus infection: Identification of genes of human and animal rotavirus origin. *Journal of Clinical Virology*, *45*(3), 237–244.

Richardson, V., Hernandez-Pichardo, J., Quintanar-Solares, M., Esparza-Aguilar, M., Johnson, B., Gomez-Altamirano, C. M., Patel, M. (2010). Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *The New England Journal of Medicine*, *362*(4), 299–305.

Ryan, M. O., Giaquinto, C., Benninghoff, B., Ryan, M. O., Giaquinto, C., Human, B. B., Giaquinto, C. (2015). Human Rotavirus Vaccine (Rotarix): Focus on Effectiveness and Impact 6 years After First Introduction in Africa. *Online Journal*, *0584*.

Saitou, N., & Nei, M. (1987). The neighgbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Sanderson, C., Clark, A., Taylor, D., & Bolanos, B. (2011). Global Review of Rotavirus Morbidity and Mortality Data by Age and Region, 1–42.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7.

Santos, N., & Hoshino, Y. (2005). Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Reviews in Medical Virology*, *15*(1), 29–56.

Santosham, M., Keenan, E. M., Tulloch, J., Broun, D., & Glass, R. (1997). Oral rehydration therapy for diarrhea: an example of reverse transfer of technology. *Pediatrics*, *100*, E10.

Schumann, T., Hotzel, H., Otto, P., & Johne, R. (2009). Evidence of interspecies transmission and reassortment among avian group A rotaviruses. *Virology*, *386*(2), 334–343.

Seheri, M., Nemarude, L., Peenze, I., Netshifhefhe, L., Nyaga, M. M., Ngobeni, H. G., Mphahlele, J. M. (2014). Update of rotavirus strains circulating in Africa from 2007 through 2011. *The Pediatric Infectious Disease Journal*, *33 Suppl 1*(1), S76–84.

Shadman, R. (2000). The withdrawal of the rotashield rotavirus vaccination due to an association with intussusception: fact or fiction? *Vaccine Revolution-Stanford University*, 1–5.

Shen, S., Burke, B., & Desselberger, U. (1994). Rearrangement of the VP6 gene of a group A rotavirus in combination with a point mutation affecting trimer stability. *Journal of Virology*, *68*(3), 1682–8.

Soriano-Gabarro, M. (2006). Burden of Rotavirus Disease in European Union Countries. *Pediatric Infectious Disease Journal 25*(1), 7–11.

Steele, A. D., & Alexander, J. J. (1987). Molecular epidemiology of rotavirus in black infants in South Africa. *Journal of Clinical Microbiology*, *25*(12), 2384–2387.

Steele, A., & Glass, R. (2011). Rotavirus in South Africa: From discovery to vaccine introduction. *184Part I) South African Journal of Epidemiology and Infections*, *2626*(44), 184–190.

Tacharoenmuang, R., Komoto, S., Guntapong, R., Ide, T., Haga, K., Katayama, K., … Taniguchi, K. (2015). Whole genomic analysis of an unusual human G6P[14] rotavirus strain isolated from a child with diarrhea in Thailand: Evidence for bovine-to-human interspecies transmission and reassortment events. *PLoS ONE*, *10*(9).

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739.

Taniguchi, K., Urasawa, T., Kobayashi, N., Gorziglia, M., & Urasawa, S. (1990). Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup I specificity and long RNA pattern: implication for new G serotype specificity. *Journal of Virology*, *64*(11), 5640–4.

Taniguchi, K., Wakasugi, F., Pongsuwanna, Y., Urasawa, T., Ukae, S., Chiba, S., & Urasawa, S. (1992). Identification of human and bovine rotavirus serotypes by polymerase chain reaction. *Epidemiology and Infection*, *109*, 303–312.

Taraporewala, Z., Chen, D., & Patton, J. T. (1999). Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. *Journal of Virology*, *73*(12), 9934–43.

Tate, J. E., Burton, A. H., Boschi-Pinto, C., & Parashar, U. D. (2016). Global, Regional and National Estimates of Rotavirus Mortality in children <5 years of age, 2000-2013. *Clinical Infectious Diseases*, *62*(S2), S96-105.

Tatte, V. S., Chothe, N. S., & Chitambar, S. D. (2014). Characterisation of rotavirus strains identified in adolescents and adults with acute gastroenteritis highlights circulation of non-typeable strains: 2008-2012. *Vaccine*, *32*(S1), A68–A74.

Teimoori, A., Soleimanjahi, H., & Makvandi, M. (2014). Characterization and transferring of human rotavirus double-layered particles in MA104 cells. *Jundishapur Journal of Microbiology*, *7*(6), 1–5.

Titone, L., Giovannelli, L., Passarani, N., & Gerna, G. (1986). Electrophrerotyping of Human Rotaviruses: An Epidermiological survey of Rotavirus Infections in Sicily. *European Journal of Epidermiology*, 2 (2), 104–107.

Trojnar, E., Otto, P., Roth, B., Reetz, J., & Johne, R. (2010). The genome segments of a group D rotavirus possess group A-like conserved termini but encode group-specific proteins. *Journal of Virology*, *84*(19), 10254–10265.

Tucker, A. W., Haddix, A. C., Bresee, J. S., Holman, R. C., Parashar, U. D., & Glass, R. I. (1998). Cost-effectiveness analysis of a rotavirus immunization program for the United States. The *Journal of the American Medical Association*, *279*(17), 1371–1376.

Valenzuela, S., Pizarro, J., Sandino, a M., Vásquez, M., Fernández, J., Hernández, O., … Spencer, E. (1991). Photoaffinity labeling of rotavirus VP1 with 8-azido-ATP: identification of the viral RNA polymerase. *Journal of Virology*, *65*(7), 3964–7.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418-426.

Van Doorn, L. J., Kleter, B., Hoefnagel, E., Stainier, I., Poliszczak, A., Colau, B., & Quint, W. (2009). Detection and genotyping of human rotavirus VP4 and VP7 genes by reverse transcriptase PCR and reverse hybridization. *Journal of Clinical Microbiology*, *47*(9), 2704–2712.

Van Regenmortel, M. H., & Fauquet, C. M. (2000). Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. *Acad. Press*, *103*, 166–87

Vende, P., Piron, M., Castagné, N., & Poncet, D. (2000). Efficient translation of rotavirus mRNA requires simultaneous interaction of NSP3 with the eukaryotic translation initiation factor eIF4G and the mRNA 3' end. *Journal of Virology*, *74*(15), 7064–71.

Vende, P., Tortorici, M. A., Taraporewala, Z. F., & Patton, J. T. (2003). Rotavirus NSP2 interferes with the core lattice protein VP2 in initiation of minus-strand synthesis. *Virology*, *313*(1), 261–273.

Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Delwart, E. (2009). Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *Journal of Virology*, *83*(9), 4642–51.

Wentzel, J. F., Yuan, L., Rao, S., Dijk, A. A. Van, & Neill, H. G. O. (2013). Infection , Genetics and Evolution Consensus Sequence Determination and Elucidation of the Evolutionary History of a Rotavirus Wa Variant Reveal a Close Relationship to Various Wa Variants Derived from the original Wa strain. *Infection, Genetics and Evolution*, *20*, 276–283.

Wilhelm, B. T., & Landry, J. R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, *48*(3), 249–257.

WHO (2009). Manual of Rotavirus Detection and Characterization Methods. *Immunization, Vaccines and Biologicals*. (n.d.).

Yang, H., Makeyev, E. V., Kang, Z., Ji, S., Bamford, D. H., & Van Dijk, A. A. (2004). Cloning and sequence analysis of dsRNA segments 5, 6 and 7 of a novel non-group A, B, C adult rotavirus that caused an outbreak of gastroenteritis in China. *Virus Research*, *106*(1), 15–26.

Yen, C., Tate, J. E., Patel, M. M., Cortese, M. M., Lopman, B., Fleming, J., … Parashar, U. D. (2011). Rotavirus vaccines: Update on global impact and future priorities. *Human Vaccines*, *7*(12), 1282–1290.

Yoder, J. D., & Dormitzer, P. R. (2006). Alternative intermolecular contacts underlie the rotavirus VP5* two- to three-fold rearrangement. *The European Molecular Biology Organisation*, *25*(7), 1559–68.

Zeller, M., Patton, J. T., Heylen, E., Coster, S. De, Ciarlet, M., Ranst, M. Van, & 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*, 966–976.

Chapter 5

Summary

In 2015, a study by Nhampossa and co-workers reported that rotaviruses are among the leading causes of severe gastroenteritis in children in Mozambique. However, there is not much published information on rotavirus surveillance in this country. In addition, currently there has not been any published information on the genotypes of strains circulating in Mozambique. In 2012, Dr Nilsa de Deus (National Institute of Health, INS) initiated a rotavirus surveillance project at Manhiça Health Research Centre (CISM) where the aim was to study the prevalence as well as the genotypes of strains circulating in two regions of the country i.e Manhiça and Mavalane, rural and urban site, respectively.

The work done in the current study is part of a collaborative project on Mozambican rotavirus strain characterisation, molecular epidemiology and rotavirus strain evolution between the research groups of Dr de Deus at INS/CISM and the University of the Free State (UFS). The aims were to use reverse-transcriptase polymerase chain reaction (RT-PCR) and Sanger sequencing to confirm the genotypes obtained by the Mozambican researchers with genotyping PCR and to use next generation sequencing (NGS) to determine genotypes of rotavirus previously typed as mixed genotypes or untypeable with genotyping PCR.

The Mozambican reseachers collected stool specimens of children with severe diarrhoea seeking medical attention in hospitals in Mavalane and Manhiça. Positive specimens for the rotavirus antigen were typed by genotyping PCR and were sent to UFS. RNA was successfully extracted from 40/58 specimens received. From the 40 RNA positive specimens, characterisation of strains into P and G genotypes using RT-PCR and Sanger sequencing was performed for 30 specimens. Phylogenetic analysis of strains whose genotypes could be determined by RT-PCR and Sanger sequencing showed that strains circulating in Mozambique were similar to strains circulating in the low income countries in southern Africa and India. In Mozambique, the genotypes detected included G2P[4], G12P[6], G8P[4] and G12P[8].

Whole genome sequencing of nine selected Mozambican samples using NGS was successful and samples which were previously typed as mixed by genotyping PCR appeared as single infections. Interestingly, one specimen that was previously typed with genotyping PCR as G12P[NT], and subsequently as G12P[6] with RT-PCR in combination with Sanger sequencing was typed as a mixed genotype with NGS, revealing a relatively rare P[14] genotype. Phylogenetic analysis of specimens characterised with NGS indicated that

Mozambican rotavirus strains are similar to strains circulating in southern Africa, India and Bangladesh. Rotavirus infections are a problem in low-income countries such as Mozambique as their rate of detection is high, with various genotypes circulating and making continual surveillance a necessity before and after the introduction of vaccines.

Keywords: Mozambique, Rotavirus, Surveillance, Genotyping, Genotyping PCR, RT-PCR, Sanger sequencing, Phylogenetic analysis, Next generation sequencing, Illumina Miseq

Chapter 6

Opsomming

'n Studie deur Nhampossa en mede-werkers het in 2015 gerapporteer dat rotavirusse een van die grootste oorsake van ernstige gastro-enteritis by kinders in Mosambiek is. Daar is egter nie veel gepubliseerde inligting oor rotavirus waarneming in hierdie land nie. Daar is ook geen gepubliseerde inligting oor die genotipes van rotavirus stamme wat in Mosambiek voorkom nie. Dr. Nilsa de Deus (Nasionale Instituut van Gesondheid, INS) het ’n projek in 2012 by die Manhiça Gesondheid Navorsingsentrum (CISM) begin wat die verspreiding en voorkoms van genotipes van rotavirus stamme in die Manhica en Mavalane streke in Mozambiek bestudeer. Hierdie streke verteenwoordig beide landelike en stedelike gebiede van die land.

Die werk wat in hierdie studie beskryf word, is deel van 'n samewerkingsprojek op Mosambiek rotavirus stam karakterisering, molekulêre epidemiologie en rotavirus stam evolusie tussen die navorsingsgroepe van Dr. de Deus by INS/CISM en die Universiteit van die Vrystaat (UV) (sien bylaag 1 vir meer inligting). Die eerste doel van hierdie studie was om genotipes van rotavirus stamme wat voorheen deur die Mosambiekse groep met genotipering polimerase kettingreaksie (G-PKR) bepaal is, te bevestig met tru-transkriptase polimerase kettingreaksie (TT-PKR), sowel as Sanger basisopeenvolgorde bepaling. Tweedens is volgende generasie basisopeenvolgorde bepaling (VGB) gebruik om genotipes van rotavirus stamme wat voorheen as gemengde genotipes geïdentifiseer was, te onderskei en verder is monsters waarvan die genotipes nie met G-PKR vasgestel kon word nie, ook deur VGB bepaal.

Die Mosambiekse groep het stoelmonsters van kinders met ernstige diarree in hospitale in Mavalane en Manhiça ingesamel. Monsters wat deur G-PKR getipeer is, is na die UV gestuur. RNS is suksesvol vanuit 40 (n=58) monsters geëkstraeer. Dertig van hierdie RNS monsters is met dubbele genotipering van die P en G genotipes deur middel van TT-PKR en Sanger basisopeenvolgorde bepaling suksesvol ontleed. Filogenetiese ontleding gebaseer op die TT-PKP en Sanger basis volgorde bepaling data het getoon dat rotavirus stamme in Mosambiek soortgelyk is aan stamme wat hoofsaaklik in die ontwikkelende lande van Suider-Afrika en in Indië voorkom. In Mosambiek het 44% van die monsters ’n G2P[4] genotipe kombinasie gehad, gevolg deur 37% met G12P[6], G8P[4] met 11.7% en G12P[8] met 7.4%.

Heel-genoom basis opeenvolgordes van nege monsters was suksesvol bepaal met VGB. Dit was ook moontlik om genotipes van monsters vas te stel wat voorheen nie met G-PKR geskei

kon word nie. Een spesifieke monster wat met G-PKR as G12P [NT] geïdentifiseer was, is met TT-PKR gevolg deur Sanger basis opeenvolgorde bepaling as G12P [8] bepaal. Interessant genoeg het VGB ook 'n relatief seldsame P[14] genotipe in hierdie monster geïdentifiseer. Filogenetiese analise gebaseer op VGB data het ook aangedui dat die Mosambiek stamme soortgelyk is aan rotavirus stamme wat in Suider-Afrika en in ontwikkelende lande in Indië en Bangladesj voorkom. Die resultate van hierdie studie het getoon dat rotavirus infeksies 'n probleem in ontwikkelende lande soos Mosambiek is. Aangesien daar ’n verskeidenheid rotavirus genotipes in Mosambiek voorkom, is voortdurende waarneming van hierdie stamme noodsaaklik voor en na die bekendstelling van entstowwe.

Sleutelwoorde: Mosambiek, Rotavirus, waarneming, Genotipering, G-PKR, TT-PKR, Sanger basisopeenvolgorde bepaling, Filogenetiese analise, Volgende generasie basisopeenvolgorde bepaling, Illumina Miseq

Appendix A

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS UV

Research Division
Internal Post Box G40
☎(051) 4052812
Fax (051) 4444359
Ms H Strauss/hv

E-mail address: StraussHS@ufs.ac.za

2013-11-28

REC Reference nr 230408-011
IRB nr 00006240

DR HG O'NEILL
DEPT OF MICROBIAL, BIOCHEMICAL AND
FOOD BIOTECHNOLOGY
FACULTY OF NATURAL AND AGRICULTURAL SCIENCES
UFS

Dear Dr O'Neill

ECUFS NR 201/2013
DR HG O'NEILL DEPT OF MICROBIAL, BIOCHEMICAL AND FOOD BIOTECHNOLOGY
PROJECT TITLE: WHOLE GENOME CONSENSUS SEQUENCE DETERMINATION OF MOZAMBICAN ROTAVIRUS STRAINS.

- You are hereby kindly informed that the Ethics Committee approved the above project at the meeting held on 26 November 2013.
- Committee guidance documents: Declaration of Helsinki, ICH, GCP and MRC Guidelines on Bio Medical Research. Clinical Trial Guidelines 2000 Department of Health RSA; Ethics in Health Research: Principles Structure and Processes Department of Health RSA 2004; Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa, Second Edition (2006); the Constitution of the Ethics Committee of the Faculty of Health Sciences and the Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines.
- Any amendment, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.
- The Committee must be informed of any serious adverse event and/or termination of the study.
- All relevant documents e.g. signed permission letters from the authorities, institutions, changes to the protocol, questionnaires etc. have to be submitted to the Ethics Committee before the study may be conducted (if applicable).
- A progress report should be submitted within one year of approval of long term studies and a final report at completion of both short term and long term studies.
- Kindly refer to the ETOVS/ECUFS reference number in correspondence to the Ethics Committee secretariat.

Yours faithfully

.....
FOR CHAIR: ETHICS COMMITTEE

REPÚBLICA DE MOÇAMBIQUE

MINISTÉRIO DA SAÚDE

COMITÉ NACIONAL DE BIOÉTICA PARA A SAÚDE

Exma Senhora
Drª Nilza de Deus

Ref.286/CNBS/10 Data 23 de Agosto de 2010

Assunto: *Aprovação do Protocolo "Comparative Study of Rotavirus infection between HIVpositive and negative children less than 5 years with acute and severe diarrhea ."*

O Comité Nacional de Bioética para a Saúde (CNBS) analisou as modificações efectuadas ao no protocolo intitulado: ***"Comparative Study of Rotavirus infection between HIVpositive and negative children less than 5 years with acute and severe diarrhe."*** sobre o mesmo chegou a seguinte conclusão:

O CNBS não vê nenhum inconveniente de ordem ética que impeça a realização do estudo pelo que, dá a sua devida aprovação.

Contudo, recomenda aos investigadores que o mantenham informado do decurso do estudo.

Faz notar que a aprovação ética não substitui a autorização administrativa.

Sem mais de momento as nossas cordiais saudações.

O Presidente do

Dr. João Manuel de Carvalho Sumane

ENDEREÇO:
MINISTÉRIO DA SAÚDE
C. POSTAL 264
Av. Eduardo Mondlane/Salvador Allende
MAPUTO – MOÇAMBIQUE

Telefones: 430814/427131(4)
Telex: 6-239 MISAU MO
FAX: 258 (1) 426547
258 (1) 33320

Appendix B

List of stains received from Mozambique

Table 1. Mozambican rotavirus strains collected from Mavalane. The sample identification number, genotype determined at CISM and the year in which the samples were collected are indicated.

Sample ID	Genotype	Year of collection
0208	G12P6	2012
0211	G12P6	2012
0223	GNTP8P6	2012
0251	NT	2012
0253	GNTP4P6	2012
0257	G12G8P6	2012
0277	G12P6	2012
0278	G12P6	2012
0285	G8P4	2012
0286	G12P6	2012
0287	G8P4	2012
0288	G12P8	2012
0289	G12P6	2012
0293	G8P4	2012
0297	G8P4	2012
0304	G12P6	2012
0308	G2P4	2012
0310	G8P4	2012
0311	G8P8	2012
0314	G2G9P8	2012
0349	G12P8	2012
0364	NT	2012
0412	NT	2013
0428	G2P4	2013
0439	G2P4	2013
0440	G2P4	2013
0441	G2PNT	2013
0448	G2P4	2013

NT in the table means the genotype of the particular sample could not be determined. GNT means the G genotype for the particular sample could not be determined while PNT means the P genotype could not be determined.

Table 2. Mozambican rotavirus strains collected from Manhica. The sample identification number, genotype determined at CISM and the year in which the samples were collected are indicated.

Sample ID	Genotype	Year of collection
0002	NT	2012
0024	G9PNT	2012
0026	G9PNT	2012
0038	G12PNT	2012
0040	G12P[6]	2012
0042	GNTP[8]	2012
0043	G2G8G9PNT	2012
0044	G8G12PNT	2012
0045	G8G12PNT	2012
0050	G12PNT	2012
0051	G12P[6]	2012
0052	G8G12PNT	2012
0060	G12PNT	2012
0067	NT	2012
0101	NT	2013
0102	NT	2013
0106	NT	2013
0146	G2P[4]	2013
0113	GNTP[4]	2013
0117	GNTP[4]	2013
0125	GNTP[4]	2013
0126	GNTP[4]	2013
0131	G2P[4]	2013
0143	GNTP[4]	2013
0144	NT	2013
0151	GNTP[4]	2013
0152	GNTP[4]	2013
0153	GNTP[4]	2013
0154	GNTP[4]	2013
0156	NT	2013

NT in the table means the genotype of the particular sample could not be determined. GNT means the G genotype for the particular sample could not be determined while PNT means the P genotype could not be determined.

APPENDIX C

Accession numbers of strains included in RT-PCR and Sanger sequencing phylogenetic trees

G2 Genotype

KP007148.1:PHI/TGO12-003/2012/G2P[4], KP007170.1:PHI/TGO12-007/2012/G2P[4], KJ752999.1:ZWE/MRC-DPRU1132/XXXX/G2P [4], JQ037744.1:Hum-wt/IRL/CIT-Bel9809/2011/G2P[4], JX965169.1:AUS/WAPC681/2010/G2P4, JX965171.1:AUS/WAPC818/2010/G2P4, KM008645.1:IND/KOL-101-09/2009/G2P [4], KP753240.1:MUS/MRC-DPRU296/2012/G2P[4], KP752805.1:MUS/MRC-DPRU308/2012/G2G12P[X] ,KJ753655.1:MUS/MRC-DPRU293/XXXX/G2P[4],KP752665.1:MUS/MRC-DPRU295/2012/G2P[4],KC443789.1:AUS/CK20051/2010/G2P[4],JX965172.1:AUS/WAPC868/2010/G2P4,JX965170.1:AUS/WAPC703/2010/G2P4

G12 Genotype

KM008689.1:IND/KOL-26-08/2008/G12P[6] ,KM008690.1:IND/KOL-30-08/2008/G1P[6] ,KP882231.1:BGD/Bang-129/2009/G12P[6] ,KP881934.1 :BGD/Bang-017/2008/G12P[6] ,KP881912.1BGD/Bang-015/2008/G12P[6] ,KP938517.1 :JU-SOL-5 , KM880064.1:JU-SOL-173 ,KP938519.1: JU-SOL-77 VP4 , KP938518.1:JU-SOL-58 , JX411973.1:India/UK-HLD/2011/H129 ,LC019055.1MMR/A23/2011/G12P[6], KM008684.1:IND/KOL-25-09/2009/G12P[6] ,JN638737.1:IVRI/India/2009/P[6] , KF598846.1 LKO/76 , KP017286.1:HRVA/HR-192/G12P[6]/CMBT/MDU/Rohtak ,JX411975.1: India/UK-HLD/2011/H21 ,KJ752817.1:ZAF/MRC-DPRU4090/2011/G12P[6]

G8 Genotype

KJ870784.1:COD/KisB101/2007/G8P[6],KJ752852.1:ZAF/MRCDPRU3477/XXXX/G8P[4] KJ753817.1:ZAF/MRC-DPRU2716/2008/G8P[4]P[8] ,KF636316.1:ZMB/MRC-DPRU3463/2009/G8P[4, KF636362.1,ZMB/MRC-DPRU1621/2008/G8P[4], JN591405.1:MAL81/G8P[4] ,AB749172.1:MWI/QOP340/2007/G8P[4], AB749173.1:MWI/QOP387/2007/G8P[4], KP752576.1:TZA/MRC-DPRU4570/2011/G8P[4] KF636250.1: /KEN/MRC-DPRU1606/2009/G8P[4], KJ753045.1:KEN/MRC-DPRU4155/XXXX/G8P[X] , KJ751607.1:TZA/MRC-DPRU4576/2010/G8P[4]

P[4] Genotype

FJ386448.1: KY6914/2002/G8P[4], KJ752665.1: UGA/MRC-DPRU1922/2008/G8P[4], KJ752375.1: ZAF/MRC-DPRU1533/2007/G2P[4], KJ752208.1: ZAF/MRC-DPRU82/2012/G2P[4], JN591400.1: MAL81/G8P[4], KF636360.1: ZMB/MRC-DPRU1621/2008/G8P[4], AB749198.1: /MWI/QOP387/2007/G8P[4], AB749197.1:

MWI/QOP340/2007/G8P[4], KJ753815.1: ZAF/MRC-DPRU2716/2008/G8P[4]P[8], KJ752850.1: ZAF/MRC-DPRU3477/XXXX/G8P[4], KJ752461.1: ZWE/MRC-DPRU3347/2010/G8P[4], KF636314.1: ZMB/MRC-DPRU3463/2009/G8P[4], JX965132.1: AUS/MON016/2010/G2P4 VP4, KC443787.1: /AUS/CK20051/2010/G2P[4], JX965112.1: AUS/WAPC681/2010/G2P4 VP8*,JX965126.1: AUS/WAPC868/2010/G2P4 VP4, JX965125.1: AUS/WAPC703/2010/G2P4 VP4, KJ752997.1: ZWE/MRC-DPRU1132/XXXX/G2P[4], HQ738607.1: Omsk08-412, HQ738608.1: Omsk08-464, KP007149.1: PHI/TGO12-003/2012/G2P[4], LC061050.1: MB Oct 2012 2, LC061047.1: LC061047.1: MB Aug 2012 2, AB550232.1: GRAVP426, JQ043280.1: CMH028/07, JQ043284.1: CMH049/07, LC061014.1: PHL/TGE12-62/2012/G2P[4], KF951367.1: NIV12445/IND/2012, KM008682.1: IND/KOL-46-08/2008/G1P[4] VP4, JX029841.1: 25Hu, LC061048.1: MB Sep 2012, KP007171.1: PHI/TGO12-007/2012/G2P[4], KJ940110.1: BRA/MG19041/2010/G2P[4] VP8 (VP8), KJ940109.1: BRA/RJ18346/2010/G2P[4] VP8 (VP8), KJ940113.1: BRA/MA19112/2010/G2P[4] VP8 (VP8), JX307596.1: IND/mcs65/2011/G8P[4], KM008679.1: IND/KOL-101-09/2009/G2P[4]

P[6] genotype

KJ751996.1:ZAF/MRC-DPRU2130-05/2005/G12P[6], KP752937.1:ZAF/MRC-DPRU1370/2004/G12P[6], KM008688.1 : IND/KOL-19-08/2008/G1P [6],KM008689.1:IND/KOL-26-08/2008/G12P[6], KM008690.1:IND/KOL-30-08/2008/G1P[6], KP882231.1:BGD/Bang-129/2009/G12P[6], KP881934.1:BGD/Bang-017/2008/G12P[6], KP881912.1: BGD/Bang-015/2008/G12P [6], KP938517.1: JU-SOL-5 VP4, KM880064.1: JU-SOL-, KP938519.1: JU-SOL-77, KP938518.1 : JU-SOL-58, JX411973.1 : India/UK-HLD/2011, LC019055.1:MMR/A23/2011/G12P[6], JN638737.1:S-25/IVRI/India/2009/P[6], KF598846.1: LKO/76,KP017286.1:HRVA/HR-192/G12P[6]/CMBT/MDU/Rohtak, KM008684.1 : IND/KOL-25-09/2009/G12P[6], MOZ_ROTA_0051/2012/G12P [6], JX411975.1:India/UK-HLD/2011/H21, KJ752817.1, ZAF/MRC-DPRU4090/2011/G12P[6]

P[8] genotype

AB905370.1:BTN/BTN-120/2010/G12P[8], KT920709.:USA/CNMC12/2011/G12P[8, KJ560517.1:CNMC12/USA/2011/G12P8,KT921094.1:USA/CNMC13/2011/G12P[8],KJ560526.1:CNMC13/USA/2011/G12P8, KP753249.1:ZAF/MRC-DPRU71/2012/G12P[8], KP753226.1:ZWE/MRC-DPRU1858/2011/G12P[8], KP753159.1:ZAF/MRC-DPRU76/2012/G12P[8], KJ753730.1:UGA/MRC-DPRU4620/2011/G12P[8], KJ752364.1:ZAF/MRC-DPRU75/2012/G12P[8]

APPENDIX D

Accession numbers for SA11 genome segments used to for the determination of open reading frames for Mozambican strains

Segment 1: JQ688673

Segment 2: JQ688674

Segment 3: JQ688675

Segment 4: JQ688676

Segment 5: JQ688677

Segment 6: JQ688678

Segment 7: JQ688679

Segment 8: JQ688680

Segment 9: JQ688681

Segment 10: JQ688682

Segment 11: JQ688683

Accession numbers of strains included in the phylogenetic analysis of the 11 genome segments of Mozambican strains characterised with NGS

Genome segment 1 (VP1)

KT921091.:CNMC13/2011/G12P[8] ,KT918732.1 :USA/VU11-12-192/2012/G12P[8], KT918424.1:USA/VU12-13-28/2013/G12P[8], KT920706.1 :USA/CNMC12/2011/G12P[8], KT918831.1:USA/VU12-13-79/2012/G12P[8], KJ752361.1:ZAF/MRC-DPRU75/2012/G12P[8], KP753246.1:ZAF/MRC-DPRU71/2012/G12P[8], KP753156.1:ZAF/MRC-DPRU76/2012/G12P[8], KJ753727.1:UGA/MRC-DPRU4620/2011/G12P[8], KP753223.1:ZWE/MRC-DPRU1858/2011/G12P[8], KP752910.1:ZAF/MRCDPRU457/2009/G10P[11], JN831231.1:ZAF/1605/2007/G6P[5], JN831209.1:ZAF/1603/2007/G6P[5], KJ752061.1:MRC-DPRU3010/2009/G6P[5], FJ495126.1:ZAF/RC-18-08/G6P[14], KF636245.1:KEN/MRC-DPRU1606/2009/G8P[4], KJ753041.1: KEN/MRC-DPRU4155/XXXX/G8P[X], KP752571.1: MRC-DPRU4570/2011/G8P[4] ,KJ751613.1:TZA/MRC-DPRU4568/2011/G8P[4, KJ751602.1:TZA/MRC-DPRU4576/2010/G8P[4], KJ752205.1:ZAF/MRC-DPRU82/2012/G2P[4], KF636311.1:ZMB/MRC-DPRU3463/2009/G8P[4] , KJ753811.1:ZAF/MRC-DPRU2716/2008/G8P[4]P[8] ,KJ752458.1:ZWE/MRC-DPRU3347/2010/G8P[4] ,AB749792.1: MWI/QOP387/2007/G8P[4], AB749791.1:MWI/QOP340/2007/G8P[4], KJ752847.1:ZAF/MRC-DPRU3477/XXXX/G8P[4], KF636357.1:ZMB/MRC-

DPRU1621/2008/G8P[4,KC782519.2:USA/LB1562/2010/G9P4 VP1 (VP1) ,KC571493.1:AUS/SA066/2010/G2P4, KP882079.1:BGD/Bang-082/2007/G2P[4] ,KP882090.1:BGD/Bang-090/2008/G2P[4], KP882222.1:BGD/Bang-125/2008/G2P[4] ,JX965133.1:AUS/WAPC703/2010/G2P4, KJ753827.1:ZWE/MRC-DPRU1158/XXXX/G2G9P[6], KP007151.1:PHI/TGO12-003/2012/G2P[4], KP007193.1: PHI/TGO12-016/2012/G1P [8], KP007173.1:TGO12-007/2012/G2P[4]

Genome segment 2 (VP2)

KP007163.1:PHI/TGO12-004/2012/G1P[8], KP007194.1 :PHI/TGO12-016/2012/G1P[8] : ZWE/MRC-DPRU1158/XXXX/G2G9P[6], KP007152.1:PHI/TGO12-003/2012/G2P[4] ,KP007174.1:PHI/TGO12-007/2012/G2P[4], KP753179.1:UGA/MRC-DPRU3710/2009/G2P[4], KJ752129.1:ETH/MRC-DPRU1844-08/2008/G3P[6], KC442893.1:USA/2008747095/2008/G2P[4], JN706481.1:KC782520.2:USA/LB1562/2010/G9P4VP2, KJ752206.1:ZAF/MRC-DPRU82/2012/G2P[4] ,KP752780.1:ZMB/MRC-DPRU1673/2009/G2P[4] ,KC443379.1:AUS/CK20060/2010/G2P[4],KJ752406.1:ZAF/MRC-DPRU81/2007/G2P[4] ,KJ753606.1:ZAF/MRC-DPRU1362/2007/G2P[4, KJ752459.1:ZWE/MRC-DPRU3347/2010/G8P[4] , AB751535.1:MWI/QOP340/2007/G8P[4], AB751536.1:MWI/QOP387/2007/G8P[4],KJ752848.1:ZAF/MRC-DPRU3477/XXXX/G8P[4] ,KF636358.1:ZMB/MRC-DPRU1621/2008/G8P[4] ,KF636312.1:ZMB/MRC-DPRU3463/2009/G8P[4] ,KJ753812.1:ZAF/MRC-DPRU2716/2008/G8P[4]P[8] , KF636246.1:KEN/MRC-DPRU1606/2009/G8P[4] , KJ753042.1:KEN/MRC-DPRU4155/XXXX/G8P[X] ,KP752572.1:TZA/MRC-DPRU4570/2011/G8P[4] ,KJ751603.1:TZA/MRC-DPRU4576/2010/G8P[4] ,KJ752663.1:UGA/MRC-DPRU1922/2008/G8P[4] , KP941130.1:KEN/Keny-061/2008/G9P[6] , KJ751648.1:ZAF/MRC-DPRU4018/XXXX/G2P[6],KJ627012.1:PRY/412/1999/G1G4P[4] ,KC834710.1:AUS/V233/1999/G2P4,

EF554105.1:HUN/Hun5/1997/G6P[14], HM113526.1:Egy3399, JN831232.1:ZAF/1605/2007/G6P[5],

JN831210.1:ZAF/1603/2007/G6P[5], FJ495127.1:ZAF/RC-18-08/G6P[14]

Genome segment 3 (VP3)

KP753225.1:ZWE/MRC-DPRU1858/2011/G12P[8], KJ753729.1: UGA/MRC-DPRU4620/2011/G12P[8], KP753158.1:/ZAF/MRC-DPRU76/2012/G12P[8] , KP753248.1:/ZAF/MRC-DPRU71/2012/G12P[8] , KJ752363.1:/ZAF/MRC-DPRU75/2012/G12P[8] , KP752573.1: /TZA/MRC-DPRU4570/2011/G8P[4], KJ751615.1:/TZA/MRC-DPRU4568/2011/G8P[4], KJ751604.1:/TZA/MRC-DPRU4576/2010/G8P[4], KF636247.1:/KEN/MRC-DPRU1606/2009/G8P[4], KJ753043.1:/KEN/MRC-DPRU4155/XXXX/G8P[X] , KP882719.1:/KEN/Keny-078/2008/G8P[6],KJ752664.1: /UGA/MRC-DPRU1922/2008/G8P[4],KJ753813.1:/ZAF/MRC-

DPRU2716/2008/G8P[4]P[8],KF636313.1:/ZMB/MRC-DPRU3463/2009/G8P[4], KJ752460.1: /ZWE/MRC-DPRU3347/2010/G8P[4], AB751561.1: /MWI/QOP340/2007/G8P[4], AB751562.1:/MWI/QOP387/2007/G8P[4],KF636359.1: /ZMB/MRC-DPRU1621/2008/G8P[4], KJ752849.1: /ZAF/MRC-DPRU3477/XXXX/G8P[4], KJ752207.1: /ZAF/MRC-DPRU82/2012/G2P[4] , AB751571.1:/MWI/OP2-384/2001/G8P[6], AB751566.1:/MWI/MW1-860/1999/G8P[6], AB751567.1:/MWI/MW2-026/1999/G8P[6], AB751565.1:/MWI/MW1-467/1998/G8P[6],JX416213.1: /SWE/1076/1983/G2P2A[6], JN706507.1: CU110-BK/08 , KP882191.1: /BGD/Bang-114/2008/G2P[4] ,KP882224.1:/BGD/Bang-125/2008/G2P[4] ,KP882081.1: /BGD/Bang-082/2007/G2P[4], HQ641366.1:MMC88, KP007153.1: PHI/TGO12-003/2012/G2P[4] , KP007175.1: /PHI/TGO12-007/2012/G2P[4], KP882092.1: /BGD/Bang-090/2008/G2P[4] , KJ753831.1:/ZWE/MRC-DPRU1158/XXXX/G2G9P[6] ,KC782521.2: /USA/LB1562/2010/G9P4

Genome segment 4 (VP4)

EF554151.1:/ESP/OVR762/2002/G8P[14], LC055550.1:/THA/SKT-27/2012/G6P[14], KP198648.1:/ITA/PR1973/2009/G8P[14], FJ747628.1: GER172-08, KP881934.1: /BGD/Bang-017/2008/G12P[6], KP881912.1: /BGD/Bang-015/2008/G12P[6], KP882231.1:/BGD/Bang-129/2009/G12P[6],KJ752817.1: /ZAF/MRC-DPRU4090/2011/G12P[6], KP753226.1:/ZWE/MRC-DPRU1858/2011/G12P[8],KJ753730.1:/UGA/MRC-DPRU4620/2011/G12P[8],KT921094.1: /USA/CNMC13/2011/G12P[8], KT920709.1:/USA/CNMC12/2011/G12P[8], KJ560517.1 /CNMC12/USA/2011/G12P8, JX965126.1: /AUS/WAPC868/2010/G2P4 ,KC443787.1: /AUS/CK20051/2010/G2P[4], JX965132.1: /AUS/MON016/2010/G2P4 , KP007171.1: /PHI/TGO12-007/2012/G2P[4],KP007149.1:/PHI/TGO12-003/2012/G2P[4], KJ752375.1:/ZAF/MRC-DPRU1533/2007/G2P[4], KF636248.1:/KEN/MRC-DPRU1606/2009/G8P[4], KJ751605.1:/TZA/MRC-DPRU4576/2010/G8P[4],KJ751616.1:/TZA/MRC-DPRU4568/2011/G8P[4], KP752574.1:/TZA/MRC-DPRU4570/2011/G8P[4], KJ752665.1:/UGA/MRC-DPRU1922/2008/G8P[4], KJ752208.1:/ZAF/MRC-DPRU82/2012/G2P[4],AB749198.1:/MWI/QOP387/2007/G8P[4],AB749197.1:/MWI/QOP340/2007/G8P[4], KF636360.1: /ZMB/MRC-DPRU1621/2008/G8P[4], KF636314.1:/ZMB/MRC-DPRU3463/2009/G8P[4],KJ752461.1:/ZWE/MRC-DPRU3347/2010/G8P[4], KJ752850.1:/ZAF/MRC-DPRU3477/XXXX/G8P[4]

Genome segment 5 (NSP1)

KT921086.1: USA/CNMC13/2011/G12P [8], AB861962.1: KEN/KDH651/2010/G12P [8], KT918727.1: USA/VU11-12-192/2012/G12P [8], KT918903.1: USA/VU11-12-59/2012/G12P [8], KT920701.1: USA/CNMC12/2011/G12P [8], KP753218.1: ZWE/MRC-DPRU1858/2011/G12P8, KJ753588.1: MRC-DPRU2140/2005/G12P [8], KP753241.1: MRC-DPRU71/2012/G12P[8], KP753151.1: ZAF/MRC-DPRU76/2012/G12P[8], KJ752356.1: ZAF/MRC-DPRU75/2012/G12P[8], KC442887.1: USA/2008747095/2008/G2P[4], KP753229.1: MUS/MRC-DPRU296/2012/G2P[4], KJ753819.1: ZWE/MRC-DPRU1158/XXXX/G2G9P[6], KP007176.1: PHI/TGO12-007/2012/G2P[4], KP007154.1: PHI/TGO12-003/2012/G2P[4], KJ751939.1: ZAF/MRC-DPRU1893/2009/G8P[6], KF636240.1: MRC-DPRU1606/2009/G8P[4], KJ753036.1: KEN/MRC-DPRU4155/XXXX/G8P[X], KP752566.1: TZA/MRC-DPRU4570/2011/G8P[4], KJ751608.1: TZA/MRC-DPRU4568/2011/G8P[4], KJ751597.1: TZA/MRC-DPRU4576/2010/G8P[4], KJ752200.1: ZAF/MRC-DPRU82/2012/G2P[4], AB753084.1: MWI/QOP340/2007/G8P[4], AB753085.1: MWI/QOP387/2007/G8P[4], KF636352.1: ZMB/MRC-DPRU1621/2008/G8P[4], KJ752842.1: MRC-DPRU3477/XXXX/G8P[4], KJ753806.1: ZAF/MRC-DPRU2716/2008/G8P[4]P[8], KF636306.1: ZMB/MRC-DPRU3463/2009/G8P[4], KP753100.1: ZWE/MRC-DPRU3342/2010/G9P[8], KJ752453.1: ZWE/MRC-DPRU3347/2010/G8P[4], KC178721.1: PA3/2004, KC834698.1: AUS/V233/1999/G2P4, KC178722.1: PA150/2006, JN706582.1: CU473-BK/09, JN706581.1: CU438-KK/09

Segment 6 VP6

KJ753731.1:/UGA/MRC-DPRU4620/2011/G12P[8] ,KT919319.1:/USA/VU11-12-54/2012/G12P[8], KT918428.1:/USA/VU12-13-28/2013/G12P[8],KT918527.1:USA/VU12-13-121/2013/G12P[8], KT918637.1:/USA/VU12-13-81/2013/G12P[8],KT918956.1:/USA/VU12-13-199/2013/G12P[8], KT918967.1:/USA/VU12-13-110/2013/G12P[8],KT920710.1:/USA/CNMC12/2011/G12P[8], KT921095.1:USA/CNMC13/2011/G12P[8],AB861960.1:/KEN/KDH651/2010/G12P[8],KF636260.1:ZAF/MRC-DPRU1604/2007/G6P[1],JN831224.1:/ZAF/1604/2007/G8P[1],KJ752065.1:/ZAF/MRC-DPRU3010/2009/G6P[5], FJ495131.1:/ZAF/RC-18-08/G6P[14],JN591410.1:/MAL81/G8P[4],KF636249.1:/KEN/MRC-DPRU1606/2009/G8P[4],KP752575.1:/TZA/MRC-DPRU4570/2011/G8P[4],KJ751606.1:/TZA/MRC-DPRU4576/2010/G8P[4], KJ753044.1:/KEN/MRC-DPRU4155/XXXX/G8P[X], DQ146686.1:/BGD/N26/2002/G12P[6], EU839963.1:SK424,KC178803.1:PA108/2007,JQ043294.1:CMH030/07,JQ043296.1:CMH049/07,KP007150.1:/PHI/TGO12-003/2012/G2P[4],KP007172.1:/PHI/TGO12-007/2012/G2P[4],JX965142.1:/AUS/WAPC703/2010/G2P4,KC443788.1:/AUS/CK20051/2010/G2P[4],KJ753833.1:/ZWE/MRC-DPRU1158/XXXX/G2G9P[6],KC782523.2:/USA/LB1562/2010/G9P4,EF554086.1:/BEL/B1711/2002/G6P[6],KJ870882.1:/COD/KisB565/2010/G8P[6],KJ870893.1:/COD/KisB554/2010/G8P[6,KP188812.1:/ITA/NA06/2009/G3P[6],JF460838.1:/BEL/F01498/2009/G3P[6],KP752497.1:/TGO/MRC-DPRU2206/2009/G3G9P[6] ,KP752509.1:GMB/MRC-DPRU3166/2008/G3P[6], JF460816.1:/USA/06-242/2006/G2P[6],KJ753122.1:ZAF/MRC-

DPRU2489/2008/G1P[8], KJ752998.1/ZWE/MRC-DPRU1132/XXXX/G2P[4],KJ753494.1:/ZMB/MRC-DPRU1752/XXXX/G4P[6],KJ752376.1:/ZAF/MRC-DPRU1533/2007/G2P[4],KJ752209.1:/ZAF/MRC-DPRU82/2012/G2P[4],KJ753609.1:/ZAF/MRC-DPRU1362/2007/G2P[4],KJ751805.1:/ZAF/MRC-DPRU1280-05/2005/G2P[8],KJ752409.1:/ZAF/MRC-DPRU81/2007/G2P[4].

Segment 7 (NSP 3)

KP753240.1: /MUS/MRC-DPRU296/2012/G2P[4] ,KP007148.1:/PHI/TGO12-003/2012/G2P[4], KP007170.1: /PHI/TGO12-007/2012/G2P[4], KF723269.1: BCK-2409/2012, KJ752999.1:/ZWE/MRC-DPRU1132/XXXX/G2P[4], KJ919895.1:/HUN/ERN5496/2012/G2P4, KP752784.1:/ZMB/MRC-DPRU1673/2009/G2P[4], KJ919836.1:/HUN/ERN5053/2012/G2P4, KJ752377.1:/ZAF/MRC-DPRU1533/2007/G2P[4], KJ752210.1:/ZAF/MRC-DPRU82/2012/G2P[4],KF723285.1:BCK-2907/2013, KU097007.1:/ITA/ME659/14/2014/G12P,KP752806.1:/MUS/MRC-DPRU308/2012/G2G12P[X], KJ752819.1:/ZAF/MRC-DPRU4090/2011/G12P[6],KF723287.1:IDK-5095/2013,KT918737.1:/USA/VU11-12-192/2012/G12P[8],KT920711.1:/USA/CNMC12/2011/G12P[8],KT918913.1:/USA/VU11-12-59/2012/G12P[8], KP753228.1:/ZWE/MRC-DPRU1858/2011/G12P[8],KJ753599.1:/ZAF/MRC-DPRU2140/2005/G12P[8], KJ752852.1:/ZAF/MRC-DPRU3477/XXXX/G8P[4], KF636316.1:/ZMB/MRC-DPRU3463/2009/G8P[4],KF636362.1:/ZMB/MRC-DPRU1621/2008/G8P[4],KJ753817.1:/ZAF/MRC-DPRU2716/2008/G8P[4]P[8],JN591405.1:/MAL81/G8P[4] ,AB749172.1:/MWI/QOP340/2007/G8P[4], AB749173.1:/MWI/QOP387/2007/G8P[4], KP752576.1:/TZA/MRC-DPRU4570/2011/G8P[4] ,KJ751618.1:/TZA/MRC-DPRU4568/2011/G8P[4],KF636250.1:/KEN/MRC-DPRU1606/2009/G8P[4] ,KJ751607.1: TZA/MRC-DPRU4576/2010/G8P[4], KJ753045.1:/KEN/MRC-DPRU4155/XXXX/G8P[X] , EF554098.1:/AUS/MG6/1993/G6P[14], FJ495133.1:/ZAF/RC-18-08/G6P[14],JF793945.1:PA5/89, JF793946.1:PA77/02,EF554142.1:/ITA/111-05-27/2005/G6P[14].

Segment 8 (NSP2)

DQ146667.1:BGD/Dhaka12/2003/G12P[6] , KJ870918.1: COD/KisB521/2008/G12P[6] ,HQ392301.1:BE00035/2008/G1P[8] ,KT918728.1: USA/VU11-12-192/2012/G12P[8] ,JX027814.1:AUS/CK00083/2008/G1P[8] ,HM773619.1:USA2009727047/2009/G9P[8], HQ609569.1:61060NSP2(NSP2),HM773751.1:USA2007719825/2007/G1P[8],AB861963.1: KEN/KDH651/2010/G12P[8],KP753219.1:ZWE/MRCDPRU1858/2011/G12P[8],KC020030.2:RUS/Nov09-D2/2009/G2G4P[4] ,AF506018.1:NR1 non-structural protein NSP2 gene, KC178732.1:PA83/2007,KC020021.1:RUS/Nov04-H390/2004/G1P[4],KC155686.1:Nov11-N2246/2011/G2P[8] ,EF554089.1:BEL/B1711/2002/G6P[6] , KC155671.1:RUS/Nov04-

H429/2004/P[6],DQ005107.1:COD/DRC88/2003/G8P[8],DQ005118.1:COD/DRC86/2003/G8P[6],KJ753037.1:KEN/MRC-DPRU4155/XXXX/G8P[X],KP752567.1:TZA/MRC-DPRU4570/2011/G8P[4],KJ751609.1:TZA/MRCDPRU4568/2011/G8P[4],KP753036.1:TZA/MRC-DPRU4539/2011/GXP[X] , KF636353.1:ZMB/MRC-DPRU1621/2008/G8P[4],KJ753807.1:MRCDPRU2716/2008/G8P[4]P[8],KP753102.1:MRCDPRU3342/2010/G9P[8],KJ752454.1:ZWE/MRCDPRU3347/2010/G8P[4],KF636307.1:ZMB/MRCDPRU3463/2009/G8P[4] ,

Segment 9 (VP7)

KP753240.1:MUS/MRC-DPRU296/2012/G2P[4], KP007148.1:PHI/TGO12-003/2012/G2P[4] KP007170.1:PHI/TGO12-007/2012/G2P[4], KF723269.1: BCK-2409, KJ752999.1:ZWE/MRC-DPRU1132/XXXX/G2P[4],KJ919895.1:HUN/ERN5496/2012/G2P4VP7KP752784.1:ZMB/MRC-DPRU1673/2009/G2P[4], KJ919836.1:HUN/ERN5053/2012/G2P4 ,KJ752377.1:ZAF/MRC-DPRU1533/2007/G2P[4], KJ752210.1:ZAF/MRC-DPRU82/2012/G2P[4], KF723285.1:BCK-2907/2013,KU097007.1:ITA/ME659/14/2014/G12P:KP752806.1:MUS/MRCDPRU308/2012/G2G12P[X] , KJ752819.1:ZAF/MRC-DPRU4090/2011/G12P[6], KF723287.1:IDK-5095/2013 , KT918737.1:USA/VU11-12-192/2012/G12P[8]: KT920711.1:USA/CNMC12/2011/G12P[8], KT918913.1:USA/VU111259/2012/G12P[8],KP753228.1::ZWE/MRCDPRU1858/2011/G12P[8],KJ753599.1:/ZAF/MRCDPRU2140/2005/G12P[8],KJ752852.1:ZAF/MRCDPRU3477/XXXX/G8P[4],KF636316.1:ZMB/MRCDPRU3463/2009/G8P[4],KF636362.1,ZMB/MRCDPRU1621/2008/G8P[4],KJ753817.1,ZAF/MRCDPRU2716/2008/G8P[4]P[8],JN591405.1MAL81/G8P[4],AB749172.1:MWI/QOP340/2007/G8P[4],AB749173.1:MWI/QOP387/2007/G8P[4],KP752576.1:MRCDPRU4570/2011/G8P[4],KJ751618.1:TZA/MRCDPRU4568/2011/G8P[4],KF636250.1:KEN/MRCDPRU1606/2009/G8P[4],KJ751607.1:TZA/MRCDPRU4576/2010/G8P[4],KJ753045.1,KEN/MRCDPRU4155/XXXX/G8P[X],EF554098.1:AUS/MG6/1993/G6P[14],FJ495133.1,ZAF/RC-18-08/G6P[14],JF793945.1:PA5/89,JF793946.1,EF554142.1:ITA/111-027/2005/G6P[14]

Genome segment 10 (NSP4)

KC851858.1:LKOR11/310,KP753221.1:ZWE/MRCDPRU1858/2011/G12P[8,KJ753725.1:UGA/MRCDPRU4620/2011/G12P[8],KP753244.1:ZAF/MRCDPRU71/2012/G12P[8],KP753154.1:ZAF/MRCDPRU76/2012/G12P[8],KT921089.1:USA/CNMC13/2011/G12P[8],HM035538.1:R1778/FRA/2008,KP752991.1:ZAF/MRCDPRU832/2006/G1P[8],KP752832.1::ZAF/MRCDPRU757/2006/G1P[8]KP752580.1:ZAF/MRC-DPRU901/XXXX/G1P[8],KP752435.1:ZAF/MRCDPRU794/2006/G1P[8],KJ753809.1:ZAF/MRCDPRU2716/2008/G8P[4]P[8],KF636355.1:ZMB/MRCDPRU1621/2008/G8P[4],KF636309.1:ZMB/MRCDPRU3463/2009/G8P[4],KJ752456.1:ZWE/MRCDPRU3347/2010/G8P[4,KP753106.1:ZWE/MRCDPRU3342/2010/G9P[8],KJ751611.1:TZA/MRCDPRU4568/2011/G8P[4],KJ753039.1:KEN/MRCDPRU4155/XXXX/G8P[X],KP753040.1:TZA/MRCDPR

U4539/2011/GXP[X],KJ751600.1:TZA/MRCDPRU4576/2010/G8P[4],KF636243.1:KEN/MRCDPRU1606/2009/G8P[4],KJ753823.1:ZWE/MRCDPRU1158/XXXX/G2G9P[6],KF951363.1:NIV12445/IND/2012 ,JX029835.1:37Hu ,KP007179.1:PHI/TGO12-007/2012/G2P[4] ,FN665686.1:HUN/BP1879/2003/G6P[14],KC870017.1:SS65 , KC870009.1:SS56 ,JX965155.1:AUS/RCH041/2010/G2P4 ,KC152930.1:ITA/PG01/2011/G2P4 ,GQ465023.1:Omsk08-475 ,KC443209.1:AUS/CK20040/2010/G2P[4] ,KC178752.1:PA84/2008 NSP4 , AB975935.1:KN041,AB796460.1 :JPN/OH3625/2012/G1P[8] ,KJ753003.1 :ZAF/MRC-DPRU1491/2010/G2P[4]P[8, KJ752203.1 :ZAF/MRC-DPRU82/2012/G2P[4] ,KP752777.1:ZMB/MRC-DPRU1673/2009/G2P[4] , KJ752403.1:ZAF/MRC-DPRU81/2007/G2P[4] ,KJ753714.1:ZAF/MRC-DPRU2344/2008/G2P[6] ,KJ753103.1:SWZ/MRC-DPRU2995/2009/G1P[8] ,KJ753546.1:ZAF/MRC-DPRU309/2010/G12P[8] ,KJ753785.1:ZAF/MRC-DPRU1195/2009/G2P[6]P[8] ,KC178753.1:PA133/2011 , KJ721750.1:BRA/ES16238/2009/G2P[4] , HM066187.1:15859_08MG

Genome segment 11 (NSP5/6)

AB861966.1: /KEN/KDH651/2010, KF372002.1: CHN/Z1557/2011/G3P [8], KF371828.1: CHN/E2421/2010/G3P [8], KT919336.1: USA/VU12-13-103/2013/G12P [8], JN706675.1: CU615-TK/09, KT918731.1: USA/VU11-12-192/2012/G12P [8], HQ657148.1: ZAF/3133WC/2009/G12P [4], KM116058.1: USA/2013774166/2013/G12P[8] NSP5 (NSP5), KT918544.1: USA/VU12-13-203/2013/G12P[8], DQ146659.1: BGD/Dhaka25/2002/G12P[8] NSP5, KP752570.1: TZA/MRC-DPRU4570/2011/G8P[4], KJ753040.1: /KEN/MRC-DPRU4155/XXXX/G8P[X], KF636244.1: /KEN/MRC-DPRU1606/2009/G8P[4], KC178759.1: strain PA108/2007 NSP5 mRNA, KP007180.1: PHI/TGO12-007/2012/G2P[4] , KP007158.1: PHI/TGO12-003/2012/G2P[4], KC178762.1: PA133/2011 NSP5 mRNA, JX965157.1: AUS/WAPC703/2010/G2P4 NSP5, KU059798.1: /AUS/WAPC2016/2014/G3P[8], KU059776.1: AUS/D388/2013/G3P[8], LC066671.1: THA/SSKT-41/2013, LC066660.1: THA/SKT-109/2013/G1P[8], JF460844.1: Hu/BEL/F01498/2009/G3P[6], KC178757.1: PA150/2006 NSP5, DQ490561.1BGD/RV176/2000/G12P[6] NSP5, KC870026.1: SS69, JX965158.1: AUS/RCH041/2010/G2P4 NSP5, KJ752204.1: ZAF/MRC-DPRU82/2012/G2P[4], HQ657137.1: MWI/1473/2001/G8P[4], EF554092.1: BEL/B1711/2002/G6P[6], DQ490544.1: BGD/RV161/2000/G12P[6], AY769694.1: RMC/G66, KC178761.1: PA84/2008, KJ753786.1: ZAF/MRC-DPRU1195/2009/G2P[6]P[8], KJ752991.1: ZWE/MRC-DPRU1132/XXXX/G2P[4], KP752731.1: ETH/MRC-DPRU4960/2010/G2P[X], KP752719.1: SWZ/MRC-DPRU4506/2010/G2P[4]P[8], KF636321.1: ZAF/MRC-DPRU1061/2009/G2P[4], FJ183362.1: GR 10924/99 NSP5, JN605447.1: MRC-DPRU9317