UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The molecular epidemiology and diversity of gastroenteritis viruses in HIV-infected, -exposed and -unexposed children under the age of five years in Pretoria, South Africa

By

Esmari Rossouw

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

MSc (Medical Virology)

Department of Medical Virology

Faculty of Health Sciences

University of Pretoria

March 2020

Supervisor: Dr J Mans

Co-Supervisor: Dr M Brauer

Declaration

I, Esmari Rossouw, hereby declare that this work was not copied or repeated from any other studies either national or international publications. Procedures were carried out in accordance with the ethical rules as prescribed by the Faculty of Health Science Research Ethics Committee, University of Pretoria.

Signature: _____

Date: _____

Acknowledgements

Firstly, and most importantly, I would like to acknowledge my supervisor, Dr Janet Mans, for her continued support throughout the entire study. I would like to thank her for the countless times that she went through my dissertation, always guiding and helping when possible. Without you, Dr Mans, this could never be possible, and I thank you for that. I would also like to thank all my co-supervisors, Dr Marieke Brauer, for your support and guidance, and for funding to do this project. Prof Theunis Avenant and Dr du Plessis for their assistance at Kalafong Provincial Tertiary Hospital. And I would like to thank Prof Pieter Meyer and Ms Siobhan Johnstone for their help with all the statistical analysis.

I would like to acknowledge my Rossouw family and my Botha family, who's unending support has gotten me through the highs and lows of this degree. Ruan Botha for the continuous supply of coffee and popcorn for the late nights. My friends as well, Cara Peens, Lizaan van Niekerk, Adelè Steinmann, Kasha Ströh and Chanel Kuca, for the smiles and laughs and motivation to keep going.

May I give the same encouragement to you in the future.

Lastly I would like to thank my colleagues and students in arms, Victor Mabasa, Alicia Vermeulen, Jaime MacDonald, Colyn Grobelaar, Phil van Jaarsveldt and especially Adam Kuca, for inspiring me to look further than just today, for giving me the courage to give my best every day and teaching me what science is really about.

May you all find your dreams and conquer them. I wish you all the best in your future endeavours and look forward to seeing more of the great things that you are sure to accomplish.

"If I have seen further, it is by standing on the shoulders of Giants." Isaac Newton

Executive summary

Viruses are common causes of both endemic and epidemic gastroenteritis, infecting millions of people per year, with norovirus, rotavirus and adenovirus-F as the main causative agents, and sapovirus and astrovirus as contributing viruses. These viruses are highly infectious and most severe in the very young, old, or individuals who are immunocompromised. The viral infection usually causes self-limited gastroenteritis, although chronic infection has been observed in highly immunocompromised patients. African and South-East Asian regions are disproportionally affected by diarrhoeal disease. These regions (especially South Africa) are also more severely affected by human immunodeficiency virus (HIV) infections. It has been suggested that immunocompromised individuals may form part of a reservoir for novel virus variants and recombinants.

It should be taken into account that not every person is equally susceptible to infection after pathogen exposure and that not all infected persons develop clinical symptoms (Ramani and Giri, 2019). One host genetic factor that can influence susceptibility to enteric infection is the expression of histo-blood group antigens (HBGAs). Histo-blood group antigens are a major group of complex carbohydrates and are determinants of both human and animal ABO blood groups and the Lewis blood group systems, which are distributed in abundance on the mucosal epithelia of the gastrointestinal tract. Histo-blood group antigens have been proven to influence susceptibility to rotavirus and norovirus infections.

Saliva, blood and stool specimens (n=205) have previously been collected from children (≤ 5 years of age) hospitalised with gastroenteritis at Kalafong Provincial Tertiary Hospital from June 2016 to December 2017. Follow up stool specimens were then collected six weeks after enrolment when possible. A descriptive questionnaire was completed by each child's guardian, giving information on age, residential area, HIV status etc. of the participating child. The stool specimens were screened for six gastroenteritis causing viruses (norovirus GI and –GII, rotavirus, sapovirus, astrovirus and adenovirus) by multiplex PCR. Forty-seven percent (96/205) of specimens tested positive for at least one gastroenteritis causing virus. Rotavirus predominated (46/205), followed by norovirus (32/205), adenovirus (15/205), sapovirus (9/205) and astrovirus (3/205). A total of 27/32 norovirus (GI.3, GII.2, GII.3, GII.4, GII.7, GII.12 and GII.21), 44/46 rotavirus (G1P[8], G2P[4], G2P[6], G3P[4], G3P[8], G8P[4], G8P[6], G9P[6] and G9P[8]) and 8/9 sapovirus (GI.1, GI.2, GII.1, GII.4 and GII.8) strains have been genotyped, of which norovirus GII.4 and rotavirus G3P[4] predominated. A total of 46/205 children submitted a follow up stool specimen to be tested.

Of the 46 children, 9 tested positive for norovirus infection with initial stool specimen testing. Follow up screening resulted in 13/46 (28%) specimens testing positive for either norovirus GI or GII, with all patients presenting as asymptomatic. After genotyping it was observed that only one of the follow up specimens were identical to the original sequence genotyped, indicating prolonged shedding. FUT2 genotyping of 205/205 children showed a 71%:29% ratio between secretors and non-secretors. Eighty percent (77/96) of the virus-infected children were secretors whereas only 20% (19/96) were non-secretors. Rotavirus ($p<0.01$) and norovirus GII.4 ($p<0.05$) specifically were found to be more prevalent in secretors. In this study, no statistical significance was observed in terms of severity of and susceptibility to gastroenteritis viruses between HIV-infected, HIV-exposed uninfected or HIV-uninfected individuals. Histo-blood group phenotyping has resulted in various combinations, with Le(b) being the most prevalent antigen found.

Next generation sequencing was unsuccessful. In future, fresh specimens should be considered for testing, with more funding and time for optimisation of this process and to give adequate results.

In summary, gastroenteritis is still a leading cause of childhood morbidity and mortality, with all advancements in understanding the disease helping to decrease the impact of it. This study again reinforced the importance of these viruses, as they are circulating in such high abundance. It also reinforced the concept that susceptibility to noro- and rotavirus infection is affected by the secretor status of a person. This could in future help with better understanding the viral infection mechanisms and in turn help with vaccine development and treatment.

Key words: Gastroenteritis, HBGAs, Secretor status, HIV exposure, Norovirus, Rotavirus, Sapovirus, South Africa

National Presentations:

- Rossouw E, Brauer M, Avenant T, du Plessis NM, Mans J.

  THE MOLECULAR EPIDEMIOLOGY, DIVERSITY AND HOST INTERACTIONS OF GASTROENTERITIS VIRUSES IN CHILDREN UNDER THE AGE OF FIVE YEARS IN PRETORIA, SOUTH AFRICA. (Oral Presentation). Faculty Day, Faculty of Health Sciences, University of Pretoria, Prinshof Campus, Pretoria. 21-22 August 2018.

- Rossouw E, Brauer M, Avenant T, du Plessis NM, Mans J.
  ROTAVIRUS INFECTIONS IN CHILDREN UNDER THE AGE OF FIVE YEARS IN PRETORIA, SOUTH AFRICA BETWEEN JULY 2016 – DECEMBER 2017. (Poster). 12th African Rotavirus Symposium, Emperors Palace, Johannesburg. 30 July – 1 August 2019.

- Rossouw E, Brauer M, Avenant T, du Plessis NM, Mans J.
  THE MOLECULAR EPIDEMIOLOGY, DIVERSITY AND HOST INTERACTIONS OF GASTROENTERITIS VIRUSES IN CHILDREN UNDER THE AGE OF FIVE YEARS IN PRETORIA, SOUTH AFRICA. (Oral Presentation). Faculty Day, Faculty of Health Sciences, University of Pretoria, Prinshof Campus, Pretoria. 20-21 August 2019.

- Rossouw E, Brauer M, Meyer P, du Plessis NM, Avenant T, Mans J.
  THE MOLECULAR EPIDEMIOLOGY AND DIVERSITY OF GASTROENTERITIS VIRUSES IN HIV-INFECTED, -EXPOSED AND - UNEXPOSED CHILDREN UNDER THE AGE OF FIVE YEARS IN PRETORIA, SOUTH AFRICA. (Oral presentation). Virology Africa 2020 Conference, Radisson Blu Hotel, Cape Town. 10-14 February 2020.

Table of Contents

List of figures

List of Tables

List of abbreviations

BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine serum albumin
Cap	Capsid
cDNA	Complementary deoxyribonucleic acid
cPCR	Colony PCR
Ct	Cycle threshold
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide triphosphate
E.coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
EIA	Enzyme immunoassay
EM	Electron microscopy
EtBr	Ethidium bromide
FUT	Fucosyltransferase
G	Genogroup
g	Gram
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GEMS	Global Enteric Multicentre Study

HAART	Highly active antiretroviral therapy
HBGAs	Histo-blood group antigens
HEU	HIV-exposed uninfected
HI	HIV infected
HIV	Human immunodeficiency virus
HRP	Horseradish peroxidase
HU	HIV unexposed
H_2SO_4	Sulfuric Acid
IEM	Immune electron microscopy
IQR	Inter-quartile Range
kb	Kilo bases
kDa	Kilodaltons
KPTH	Kalafong Provincial Tertiary Hospital
Le	Lewis
MAb	Monoclonal Antibody
MAL-ED	Malnutrition and Enteric Disease Study
MCS	Microscopy, culture and sensitivity testing
MGB	Minor groove-binding
min	Minutes
mM	Millimolar
n	Number
NA	Non-assigned
NEB	New England Biolabs

NGS	Next generation sequencing
NHLS	National Health Laboratory Service
NICD	National Institute for Communicable Disease
nm	Nano meters
nM	Nano Molar
NS	Non-structural
nt	Nucleotide
NTC	No template control
NTPase	Nucleoside triphosphatase
NV	Norwalk virus
ORF	Open reading frame
ORS	Oral rehydration solution
P	Polymerase
p	Probability
PBS	Phosphate buffered saline
PBS-T	PBS containing 0.05% Tween 20
PCR	Polymerase chain reaction
Pd	Protruding domain
pH	Power of hydrogen
pmol	Pico Molar
Poly(A)	Polyadenylated
PPM	Parts per million
qPCR	Quantitative real-time PCR

RCWG	Rotavirus Classification Working Group
RdRp	RNA-dependent RNA polymerase
R_n	Normalised reporter
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR
S	Shell domain
SA	South Africa
SCS	SalivaBio Children's Swab
SD	Standard deviation
Se	Secretor
sec	Seconds
SeSe	Homozygous secretor
Sese	Heterozygous secretor
sese	Homozygous non-secretor
Se^{428}	Non-secretor allele
SNP	Single-nucleotide polymorphism
SNV	Single-nucleotide variation
Spp.	Species pluralis
ST-ETEC	Stable Enterotoxigenic *Escherichia coli*
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TMB	Tetramethylbenzidine
UEA-I	Lyophilised *Ulex europaeus* agglutinin
Unt	Untypable

VLP	Virus-like particle
VP	Viral protein
VPg	Viral genome-linked protein
WHO	World Health Organization
°C	Degrees Celsius
µL	Micro litre
µM	Micro molar

Chapter 1 : Literature review

1.1 Introduction

Diarrhoea is defined as the passage of three or more loose or liquid stools in 24 hours, according to the World Health Organization (WHO). Diarrhoeal disease, including viral gastroenteritis, is the second leading cause of infectious disease morbidity and one of the top ten leading causes of mortality worldwide (IHME, 2018), resulting in an estimated 90 million disability-adjusted life years (ranking sixth in global disability-adjusted life years burden (Hale *et al.*, 2015)). In 2017 diarrhoeal disease accounted for approximately 8% (or ≈ 499000) of global, and 25% of African and South-East Asian deaths among children under the age of five years (Troeger *et al.*, 2017, UNICEF, 2019). Although this is an improvement when compared to 2010 statistics, it still translates to over 1,300 young children dying per day, with the highest amount of deaths from diarrhoea occurring in children under the age of two years.

According to recent large scale studies, such as the MAL-ED (Platts-Mills *et al.*, 2015) and GEMS (Liu *et al.*, 2016a) studies, some of the most attributable pathogens of gastroenteritis in children, in descending order, are *Shigella* spp., rotavirus, adenovirus-F, *ST-ETEC*, *cryptosporidium spp*., and *campylobacter spp*. (as shown in Figure 1.1) (Liu *et al.*, 2016a). These studies showed that the pathogen-attributable burden was approximately 89.3%, with the six abovementioned pathogens accounting for approximately 77.8% of all attributable diarrhoea.

Viruses are amongst the most common causes of both endemic and epidemic gastroenteritis, infecting millions of people per year, with norovirus, rotavirus and adenovirus as the main causative agents, and sapovirus and astrovirus as contributing viruses (Parashar *et al.*, 2009, Patel *et al.*, 2009, Liu *et al.*, 2016a). There are >50 serotypes of adenovirus known, with types 40 and 41 mostly associated with diarrhoea (Uhnoo *et al.*, 1984). Astroviruses that cause diarrhoeal disease in humans belong to the Mamastrovirus genus (types 1–8), in the family *Astroviridae* (Bosch *et al.*, 2010, Kotloff, 2017). All of these viruses are highly infectious and most severe in people at the extremities of age, or who are immunocompromised (Patel *et al.*, 2009). The viral infection usually causes self-limiting gastroenteritis, although chronic infection has been observed in highly immunocompromised patients (Atmar, 2010, Bok and

Green, 2012). Even though general prevention methods, such as better sanitation infrastructure and an increase in the awareness of these viruses, have been set in place and found to be broadly effective against these enteric infections, these control programmes still rely on surveillance to understand the relative burden of the individual pathogens.

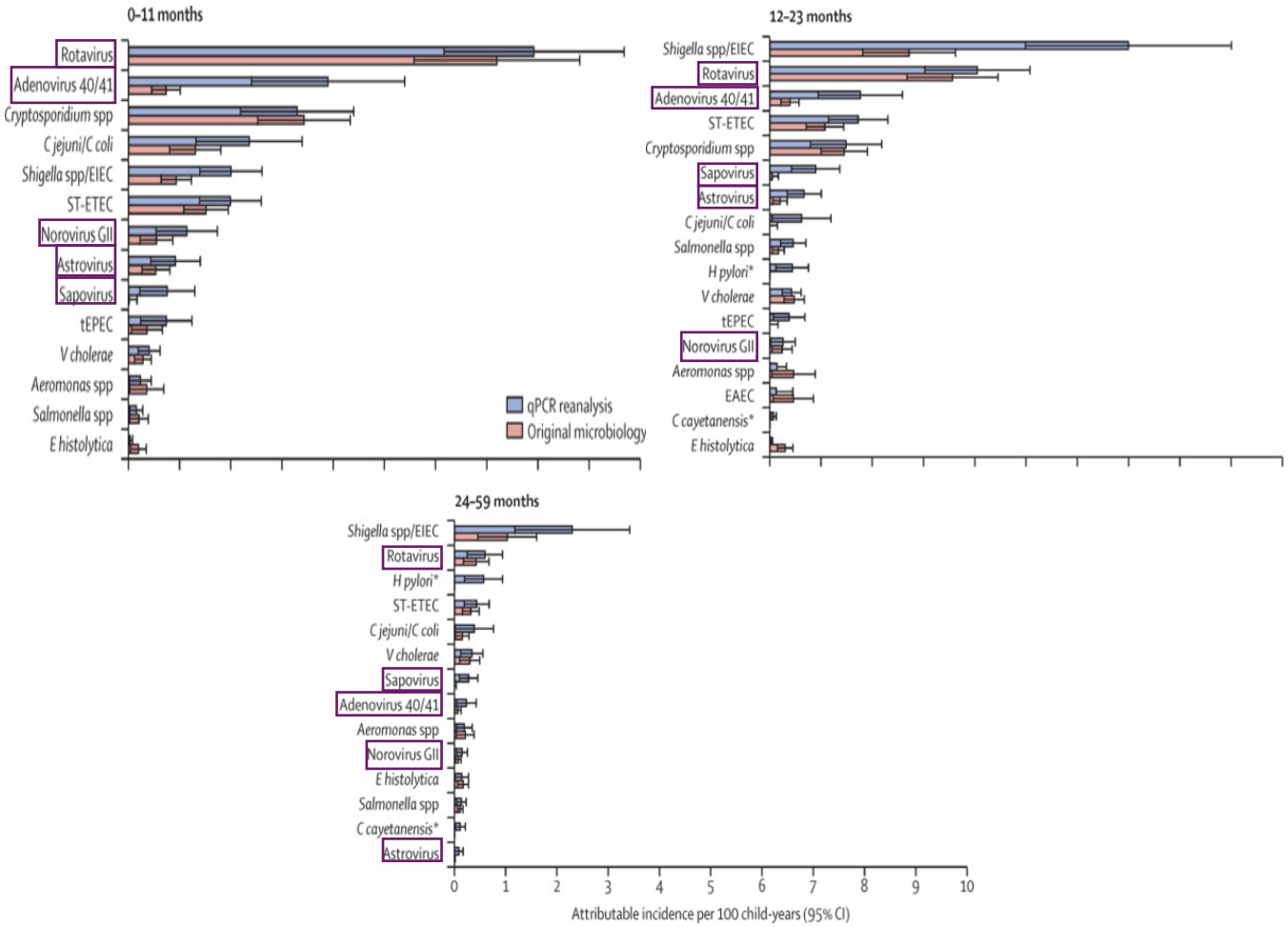


Figure 1.1: Attributable incidence of pathogen-specific moderate to severe diarrhoea per 100 child-years, by age stratum, across study sites. Adapted from the original GEMS study, along with the qPCR reanalysis, with blocks highlighting the specific viruses most commonly associated with gastroenteritis, Liu *et al.*, 2016.

The ease and prevalence of world travel have greatly facilitated the encounter between viruses and new hosts (Enquist *et al.*, 2015). As mentioned above, African and South-East Asian

regions are disproportionally affected by diarrhoeal diseases (Figure 1.2A) (Lozano *et al.*, 2012, Walker *et al.*, 2012, UNICEF, 2019). These regions are also more severely affected by human immunodeficiency virus (HIV) infection (Figure 1.2B) (WHO, 2017, IHME, 2018). Despite the heavy burden of diarrhoeal disease, the situation has improved and from 2000 to 2017 the number of deaths caused by diarrhoea among children under 5 years has reduced substantially (≈ 60%) (UNICEF, 2019). This is a great achievement, but as there are still children dying from this disease, further improvement is needed.

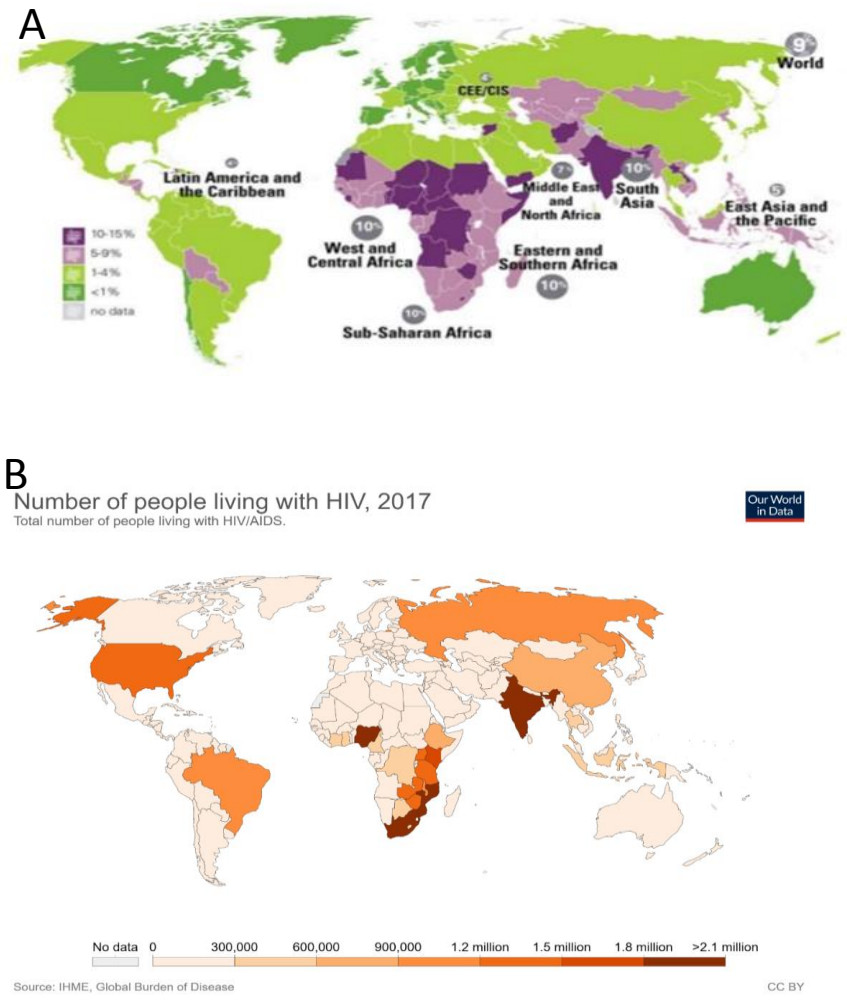


Figure 1.2: A) World map representing the percentage of deaths among children under the age of five, which can be attributed to diarrhoea in 2015 reproduced from UNICEF, 2016. B) World map representing the estimated prevalence of HIV in 2017 reproduced from the global burden of disease, IHME, 2018.

1.2 *Caliciviridae*

The *Caliciviridae* family consists of five genera, including Norovirus, Sapovirus, Lagovirus, Nebovirus, and Vesivirus, and six new genera that have been proposed, namely Bavovirus, Nacovirus, Recovirus, Salovirus, Minovirus and Valovirus (Oka *et al.*, 2015, Vinjé *et al.*, 2019). All of these are small (27-40 nano meters [nm]), single-stranded, positive sense, non-enveloped, non-segmented viruses with ribonucleic acid (RNA) genomes (Blacklow and Greenberg, 1991), but only norovirus and sapovirus have been known to cause gastroenteritis in humans and animals (Karst *et al.*, 2015).

1.2.1 History

Norovirus was likely described for the first time as "winter vomiting disease" in 1929 by Dr. John Zahorsky (Zahorsky, 1929). It was characterised by the abrupt onset of limited vomiting and diarrhoea, typically most prevalent in winter months. Identification of the specific viral particles only occurred in 1972 after an outbreak in a primary school in Norwalk, Ohio in 1968. Kapikian and his team used immune electron microscopy (IEM) to identify these particles from stool (Kapikian *et al.*, 1972, Kapikian, 2000). The researchers identified the Norwalk virus (Figure 1.3A) specifically, which is the prototype agent of the genus Norovirus (Patel *et al.*, 2009). This discovery presented electron microscopists with a new way to identify viral pathogens in faecal specimens, by using IEM to characterise the immune response, leading to other enteric viruses such as rotavirus (Figure 1.3B) to be discovered, which were then classified by their morphological appearance by electron microscopy (EM) (Lopman *et al.*, 2014b).

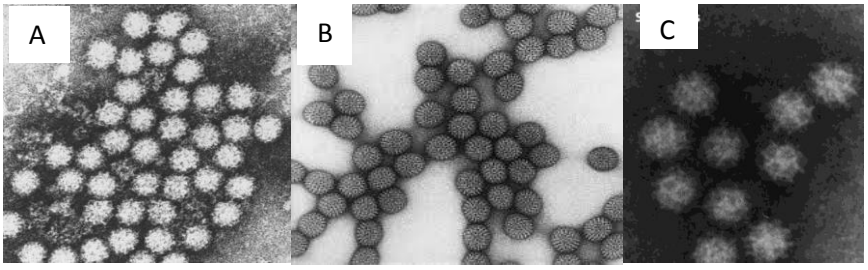


Figure 1.3: Electron microscopy images of A) norovirus particles reproduced from Kapikian *et al.*, 1972, B) rotavirus particles and C) sapovirus particles reproduced from Oka *et al.*, 2015.

In 1976, sapovirus (originally known as Sapporo-like virus), was discovered during an EM study of diarrhoeal stool samples in the United Kingdom (Madeley and Cosgrove, 1976). The prototype strain of the sapovirus genus was characterised from an outbreak in 1982, Sapporo, Japan, (strain Hu/SaV/Sapporo/1982/JPN) (Figure 1.3C) (Chiba *et al.*, 2000).

1.2.2 Clinical and epidemiological features

Clinically, noro- and sapovirus are characterised by self-limiting symptoms such as an abrupt onset of nausea, abdominal pain, non-bloody diarrhoea, and vomiting, with an incubation time of 24-48 hours after infection for norovirus and a duration of approximately two to three days for norovirus and two to six days for sapovirus (Patel *et al*., 2009, Lee *et al.*, 2012, Oka *et al*., 2015). The average duration of the illness can be prolonged in patients infected during hospital outbreaks, children under the age of 11 (Patel *et al.*, 2009), or in immunocompromised patients (Atmar, 2010, Vinjé, 2015). Incubation of sapovirus can range from a day to four days (Yamashita *et al.*, 2010), and both noro- and sapovirus shedding can continue up to four weeks after symptoms have subsided (Oka *et al*., 2015).

In recent years, norovirus has become the leading cause of non-bacterial epidemic gastroenteritis in children in middle- to high-income countries where the rotavirus vaccine has been introduced successfully (Tate *et al.*, 2012, Griffin, 2013), with genogroup GII noroviruses reported as the most common strain (Patel *et al.,* 2009). Estimates from several studies state that norovirus causes approximately 685 million cases and 200,000 deaths globally per year (Havelaar *et al.*, 2015, Pires *et al.*, 2015). Globally, norovirus alone results in a total of approximately $4.2 billion in direct health system costs, with a further $60.3 billion in societal costs per year (Bartsch *et al.*, 2016). Norovirus is associated with 18% of all cases of acute gastroenteritis (Ahmed *et al.*, 2014), although it is found to be more commonly associated with mild acute gastroenteritis and has been reported to be detected in approximately 30% of healthy individuals (Lopman, 2015, Miura *et al.*, 2018). It should be noted that even though these individuals are asymptomatic, they still shed the virus in detectable amounts (Atmar *et al.*, 2008, Miura *et al.*, 2018). Asymptomatic results should be viewed with caution, as the possible reason for asymptomatic infection may include long-term shedding from a previous symptomatic episode (less than 6 weeks from previous episode), and not be a truly asymptomatic infection (which is due to the lack of susceptible factors for symptomatic infection) (Qi *et al.*, 2018). Cohort studies have determined that approximately 66%-90% of children experience at least

one norovirus infection (whether symptomatic or asymptomatic) in early childhood, with the highest incidence found in children under 11 months of age (Platts-Mills *et al.*, 2018, Cannon *et al.*, 2019). The severity of sapovirus infection is in general not as great as that of rota- or norovirus infections (Sakai *et al.*, 2001), but as with most diseases, the symptoms, severity and duration are all affected by individual host factors (Oka *et al.*, 2015). Norovirus outbreak peaks vary depending on the hemisphere. In the northern hemisphere, a clear peak can be seen in the winter season (Patel *et al.*, 2009). In the southern hemisphere, more specifically African regions, norovirus seasonality is not as clear. Peaks have been reported in summer months, others throughout the year with no clear seasonal pattern, and others more prevalent in the cool, dry season (Mans *et al.*, 2016). Sapovirus outbreaks occur year round with no discrimination for age or setting of the host (Yan *et al.*, 2003).

1.2.3 Structure

1.2.3.1 Norovirus Structure

The human norovirus particle is non-enveloped and ~27 nm wide (Kapikian *et al.*, 1972). The genome is approximately 7.5 kilobases (kb) nucleotides long, with a 3' polyadenylated (poly(A)) tail and a covalently linked viral protein (VP) at the 5' end. (Hansman *et al.*, 2010). The genome consists of three open reading frames (ORFs); ORF1, ORF2, and ORF3. Open reading frame 1 (~5 kb), consists of two-thirds of the genome. This ORF encodes a ~200 kilo Daltons (kDa) polyprotein which is essential for viral replication (Donaldson *et al.*, 2010). Open reading frame 1 encodes multiple proteins with specific functions: p48, an amino-terminal protein (~48 kDa); nucleoside triphosphatase (NTPase), a 2C-like protein; p22, a 22 kDa 3A-like protein; viral genome-linked protein (VPg), and RNA-dependent RNA polymerase (RdRp), a 3D-like protein. The p48 and p22 proteins are known to block the host secretory pathways (Ettayebi and Hardy, 2003, Fernandez-Vega *et al.*, 2004) , which has been suggested to interfere with intercellular protein trafficking (Roth and Karst, 2016). Open reading frame 2 is ~1.6 kb in length and encodes the 57 kDa major structural capsid protein, viral protein 1 (VP1). Viral protein 1 is divided into two domains, the shell (S) domain (yellow) and the protruding domain (Pd), divided into two parts known as P1 (shown in blue) and P2 (shown in red) (Figures 1.4 and 1.5). Deletion experiments confirmed that the P domain is needed for interaction with receptors (Tan *et al.*, 2004).

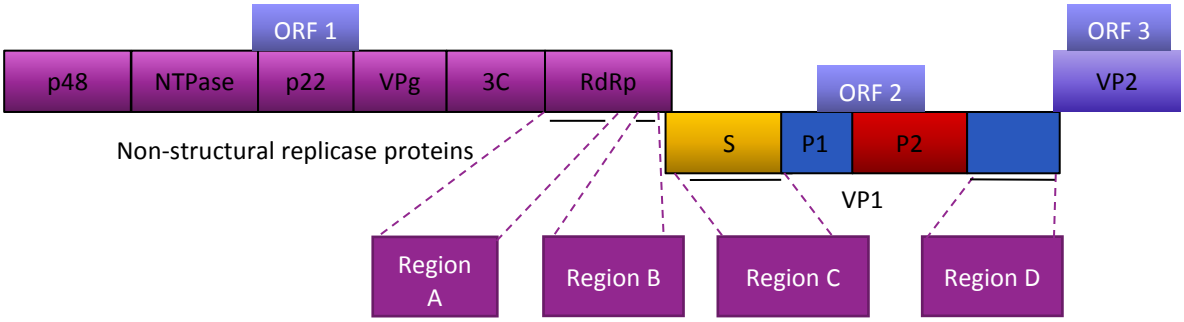


Figure 1.4: Schematic diagram representing the genomic organisation and reading frames of norovirus. Adapted from de Graaf *et al.*, 2016 with a schematic presentation of norovirus genome and positions of regions (A–D) that are commonly used for detection and genotyping, adapted from Vinjé *et al.*, 2004.

Open reading frame 3 is between 208–268 amino acids in length and encodes a 22 kDa minor basic structural protein, VP2, found at the interior surface of the capsid (at the S domain), which is thought to stabilise the capsid and has been proposed to be involved in capsid assembly and genome encapsulation (Seah *et al.*, 1999, Someya, 2002, Tan *et al.*, 2004, Donaldson *et al.*, 2010, Vongpunsawad *et al.*, 2013, Robilotti *et al.*, 2015, de Graaf *et al.*, 2016). The capsid of each viral particle consists of 90 dimers of VP1. The S domains form the shell of the capsid and the P domains extend upward from the shell surface, stabilised by many intermolecular dimer interactions (Tan *et al.*, 2004).

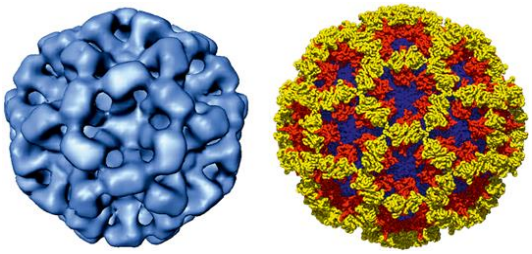


Figure 1.5: The structure of the calicivirus capsid demonstrated by cryo-image reconstruction of the recombinant Norwalk virus-like particles (Left). X-ray structure of the Norwalk virus capsid (Right) with the Shell, Protruding 1, and Protruding 2 domains coloured in blue, red and yellow, respectively. Reproduced from Clarke *et al* 2012.

1.2.3.2 Sapovirus structure

Sapovirus particles range from 30 to 38 nm in diameter, they are icosahedral, and have cup-shaped depressions on their surface (the typical morphology for a calicivirus), as well as the characteristic 'Star-of-David' appearance (Madeley, 1979, Hansman *et al.*, 2007). The genome organisation of the sapovirus differs greatly from that of the noroviruses (Yan *et al.*, 2003). The sapovirus genome is between 7.1 and 7.7 kb in size, with a 3' poly (A) tail (Chang *et al.*, 2005). Like norovirus it has a VPg linked to the 5' end of the viral RNA which is critical for the genome replication, transcription and translation (Oka *et al.*, 2015). It has two conformations, with genogroups GII and GIII consisting of two and the other genogroups consisting of three ORFs respectively (Oka *et al.*, 2012). The genomic organisation can be seen in Figure 1.6, showing both ORFs, as well as the extra cleavage site, indicated with an arrow, to show where the VP1 is cleaved from the rest of the ORF1 polyprotein.

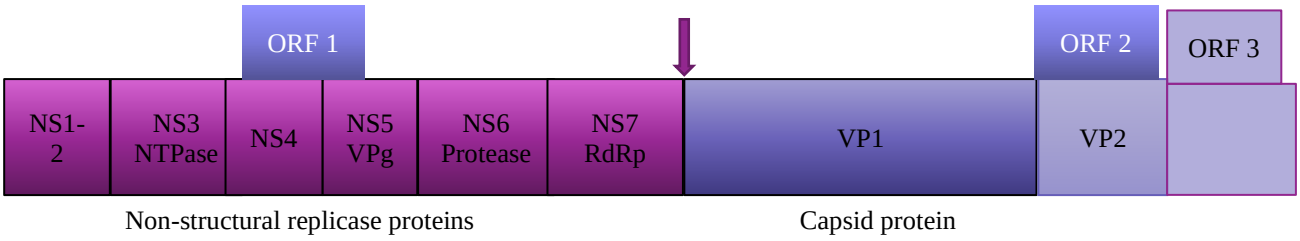


Figure 1.6: Genomic organisation and reading frame usage of sapovirus. Adapted from Oka *et al.*, 2015.

Open reading frame 1 encodes for a large polyprotein which contains all the non-structural proteins, as well as the major capsid protein VP1 (Yan *et al.*, 2003). The capsid protein is believed to contain all of the elements for viral attachment and antigenicity (Clarke and Lambden, 2000). The capsid is formed from 180 molecules of VP1 (Oka *et al.*, 2015). The minor structural protein VP2, similar to the VP2 of norovirus is encoded by ORF2 (Green, 2007, Hansman, 2007, Oka *et al.*, 2015). Several human and bat sapovirus strains are predicted to have a third ORF (Oka *et al.*, 2015). The ORF3 encodes proteins of unknown function, and is only present in genotype GI, GIV and GV (Hansman, 2007). All *Caliciviridae* viruses encode for at least seven protein functions indicated in Figure 1.6 as non-structural (NS) proteins NS1 through NS7, with processing differing between different viruses. An extra cleavage site is present in the ORF1 of caliciviruses such as sapovirus in which the capsid protein sequence is

in frame with the non-structural polyprotein encoded in ORF1. Cleavage at this site (indicated with an arrow in Figure 1.6) is thought to release the VP1 from the polyprotein so that the RdRp (NS7) can assume an active conformation in the beginning of the replicative cycle (Green, 2007, Oka *et al.*, 2015).

1.2.4 Classification and molecular detection

There are at least ten genogroups of norovirus known to date, of which genotypes GI, GII, GIV and GIX infect humans (Figure 1.7 and Figure 1.8) (Chhabra *et al.*, 2019).

These genogroups can further be divided into multiple genotypes. Norovirus genotyping studies have classified at least 36 human norovirus genotypes, of which GII.4 is most prevalent, (Chhabra *et al.*, 2018), followed by GII.3, GII.6, GII.2 and GII.7 respectively (Kumthip *et al.*, 2019). The GII.4 genotype has been identified in more than 80% of cases (Kambhampati *et al.*, 2015) and over 50% of norovirus epidemics worldwide (Dai *et al.*, 2015), although multiple genotypes are circulating at any given time. Globally, GII has a prevalence of 96%, GI a prevalence of 3.6% and GIV with an average of 4% (Tran *et al.*, 2013). The detection rate of GII.2, GII.3 and GII.17 have a higher prevalence in low resource countries, whereas GII.6 and GII.7 are higher in developed countries (Kumthip *et al.*, 2019).

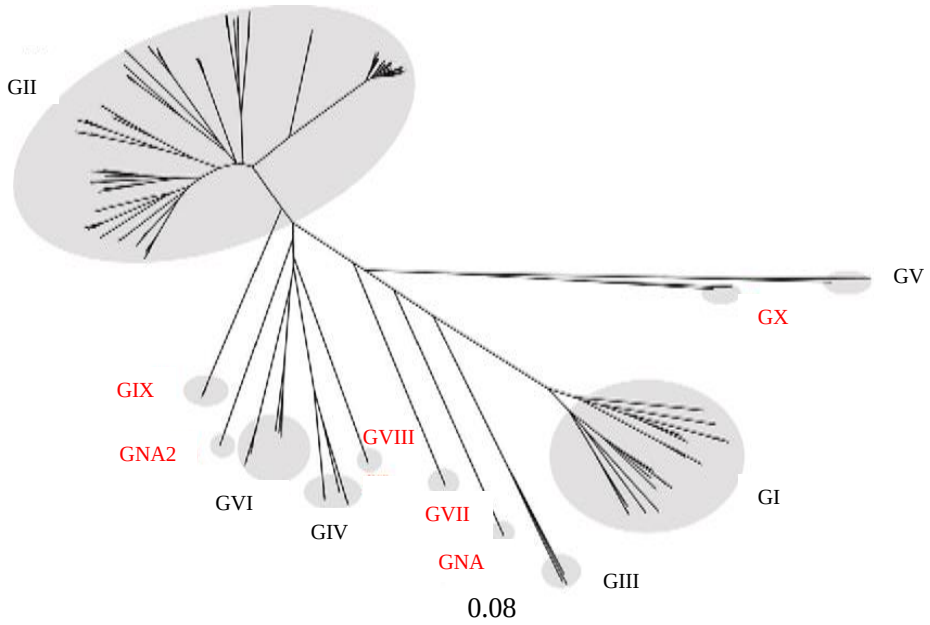


Figure 1.7: Phylogenetic classification of noroviruses based on VP1 amino acid sequences into ten norovirus genogroups and one non-assigned (NA) genogroup. Phylogenetic analysis was performed using maximum likelihood (PhyML). Resulting trees were plotted and edited in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Figure reproduced from Chhabra *et al*. 2019.

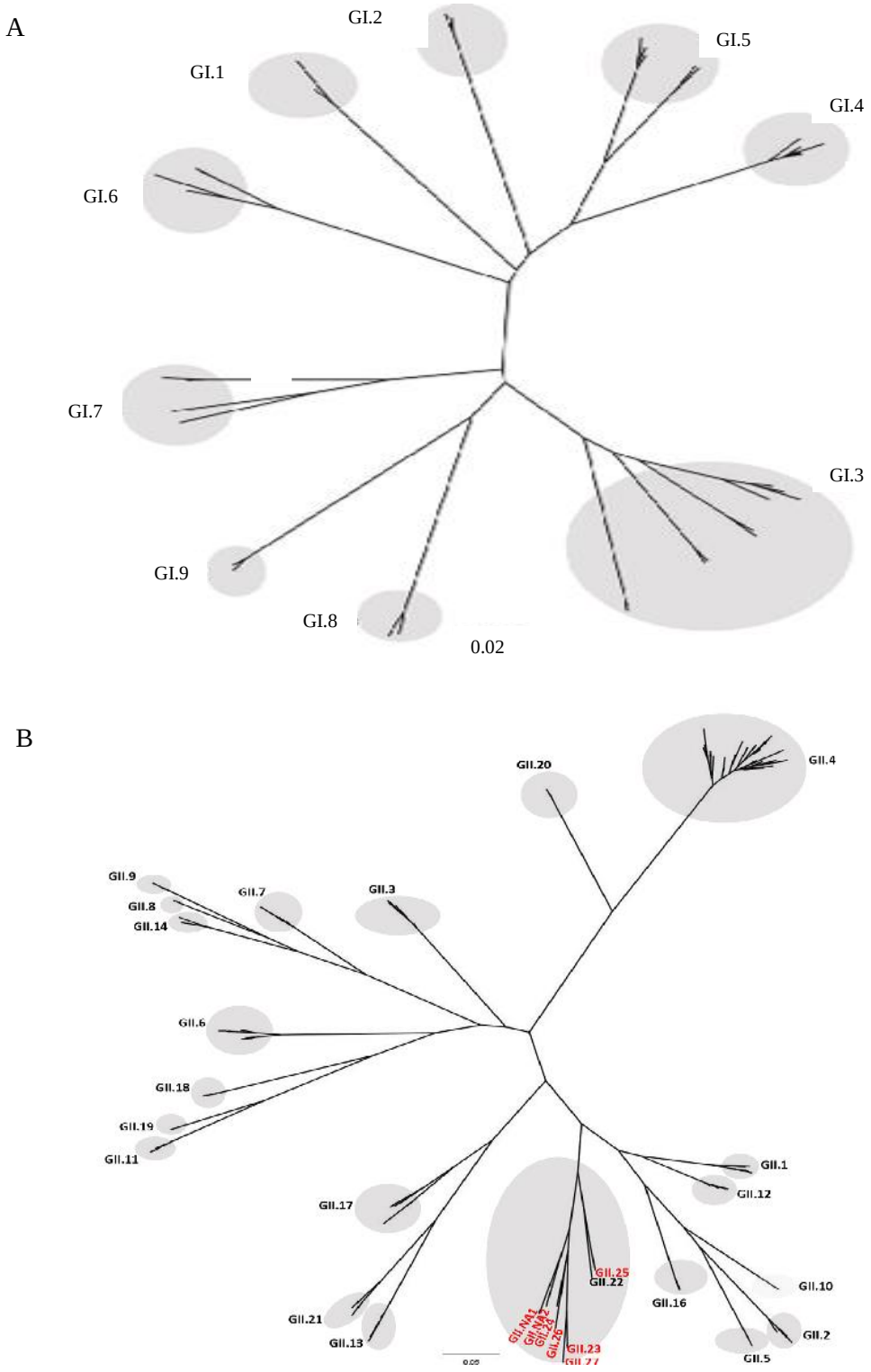


Figure 1.8: Phylogenetic classification of noroviruses based on VP1 amino acid sequences into A) GI genotypes and B) GII genotypes. Phylogenetic analysis was performed using maximum likelihood (PhyML). Resulting trees were plotted and edited in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Newly identified genogroups and genotypes are labelled in red. Reproduced from Chhabra *et al.*, 2019.

Noroviruses are classified based on the analysis of the complete amino acid sequence of the capsid protein, VP1 found in ORF2 (Vinjé, 2015), as well as part of the viral polymerase, found in ORF1 (Kroneman *et al.*, 2013). Originally a 15% sequence difference was used as a cut-off threshold for new genotypes (Zheng *et al.*, 2006). In recent years the classification of new genotypes has been dictated by the internationally accepted 2 × SD criterion of VP1 divergence, and the agreement among the Norovirus Working Group that new norovirus genotypes need to be supported by VP1 sequences from at least two different countries (Kroneman *et al.*, 2013). Genotypes GI and GII are divided into at least 9 and 22 different genotypes respectively (Kroneman *et al*., 2013).

As the clinical symptoms for viral gastroenteritis are so similar, laboratory diagnosis is needed to identify the causative pathogen accurately. The current detection and genotyping methods of genogroup I and II noroviruses consist of a two-step approach including detection of viral RNA by TaqMan real-time polymerase chain reaction (qPCR) (gold standard) (Kageyama *et al.*, 2003), followed by conventional RT-PCR and sequencing of partial regions B and C of ORF1 and ORF2 (Figure 1.4). Recent updating of the classification of norovirus genogroups and genotypes has resulted in the grouping of nucleotide sequences from the partial RdRp region into polymerase [P]-groups and P-types, which are independent from the classification of the corresponding capsid genogroups and genotypes (in other words dual-typing, to include the diversity at the level of partial RdRp sequences in strain designations (Kroneman *et al*., 2013)). This was done to provide a uniform basis for norovirus classification, especially when taking into consideration recombination observed in these strains (example of new proposed dual-typing designations of norovirus strains shown in Table 1.1) (Chhabra *et al*., 2019). Further classification of prototype strains in current and new genotype assignments can be found in the article.

Table 1.1: Examples of new proposed dual-typing designations of norovirus strains adapted from (Chhabra *et al*., 2019).

Previous designation	New designation
GI.P6-GI.6	GI.6[P6]
GI.Pd-GI.3	GI.3[P13]
GII.P1-GII.1	GII.1[P1]
GII.Pe-GII.4 Sydney	GII.4 Sydney [P31]
GII.P16-GII.4 Sydney	GII.4 Sydney[P16]
GII.P15-GII.15	GIX.1[P15]

The genetic diversity of human norovirus is very apparent when taking into account that VP1 amino acid sequences of GII.4 strains differ by 5% to 7% in just the GII genotype (Bok *et al.*, 2009). Recombination has been identified as a potentially important way for these viruses to generate diversity (Table 1.2) (Bull *et al.*, 2007, Kroneman *et al*., 2013). It is known that GII.4 noroviruses undergo antigenic drift, which can influence the HBGA repertoire available to allow attachment. This antigenic drift is likely selected by host herd immunity (Dai *et al*., 2015). The Sydney 2012 GII.4 is the last known strain to have caused a pandemic, although novel recombinants of the pandemic strain have been described. A recent recombinant norovirus GII.P16/GII.4 has emerged in the United States and Europe and spread worldwide (Cheung *et al.*, 2019). It has been documented that the GII.P16 polymerase has recombined with ≥8 capsid genotypes (Van Beek *et al.*, 2018), such events could lead to changes in norovirus epidemiology (Cheung *et al*., 2019).

Table 1.2: Proposed epidemic norovirus GII.4 variants (Kroneman *et al*., 2013).

Proposed epidemic variant name	GenBank no.[1]
US95_96	AJ004864[2]
Farmington_Hills_2002	AY485642[3]
Asia_2003	AB220921[3]
Hunter_2004	AY883096[2]
Yerseke_2006a	EF126963[2]
Den Haag_2006b	EF126965[2]
NewOrleans_2009	GU445325[3]
Sydney_2012	JX459908[3]

[1]GenBank accession number of the first submitted capsid sequence of this variant
[2]Capsid sequence
[3]Complete genome

In Asia, during the 2014 and 2015 gastroenteritis season, norovirus GII.4 was replaced by GII.17 as the most predominant genotype in circulation (de Graaf *et al.*, 2015). This strain has been detected in South African environmental surveillance from wastewaters since as early as 2015 (Mabasa *et al.*, 2018), and has been identified in clinical specimens from 2017 (Vermeulen, 2019).

New norovirus GII genotypes and strains have been proposed recently, which were classified into one existing genotype (GII.22), three novel (GII.23, GII.24, and GII.25) and three tentative novel (GII.NA1, GII.NA2, and GII.NA3) genotypes (Chhabra *et al*., 2018). These newly identified strains again show the importance of global norovirus surveillance to keep track of the changing genotype distributions, as well as identify emerging novel genotypes. If the sequences were only supported by one country, a genotype has preliminarily been labelled as NA (non-assigned).

In South Africa specifically a study on wastewaters between 2015 and 2016, determined that norovirus was circulating abundantly in raw sewage and effluent water samples collected from multiple wastewater treatment plants (Mabasa *et al*., 2018). This study also found that norovirus GII.2 and GII.17 predominated in the wastewaters and not GII.4 which is known to predominate in clinical surveillance in children < 5 years of age (Mabasa *et al*., 2018).

Sapoviruses are classified into 19 genogroups, of which GI, GII, GIV and GV infect humans, these groups are further divided into 18 genotypes (Farkas *et al.*, 2004, Hansman *et al*., 2007, Oka *et al.*, 2018, Vinjé *et al*., 2019)). Viruses in the other genogroups have been detected in swine (GIII and GV-GXI), sea lions (GV), mink (GXII), dogs (GXIII), bats (GXIV, GXVI-GXIX) and rats (GXV) (Yinda *et al.*, 2017). Sapovirus GI and GII can each be subdivided into seven and eight genotypes respectively, GIII and GIV each contain a single genotype, whereas GV has two known genotypes (Oka *et al*., 2012, Oka *et al*., 2018). Detection and classification of sapovirus genogroups are done by performing a TaqMan RT-qPCR (gold standard) and complete nucleotide sequencing of the capsid region (VP1, which is approximately 1700 nt long), and then comparing the amino acid sequence, to other complete capsid region sequences, as was proposed by the International Calicivirus Scientific Committee at the Fourth International Conference on Caliciviruses in Chile (in 2010) (Oka *et al*., 2012). The partial RdRp or partial VP1 region or both of these regions can be used to partially characterise detected sapoviruses (Oka *et al*., 2015). A pairwise distance cut-off value of ≤0.169 is used to distinguish different sapovirus genotypes and ≤0.488 to distinguish different genogroups. Previous studies in South Africa identified GIV as the most prevalent strain in children hospitalised with gastroenteritis, followed by GI.2 (Murray *et al.*, 2016).

Although human sapovirus has not been cultivatable to date, recombinant VP1 can be expressed in baculovirus- or mammalian expression systems, which promotes the self-assembly of virus-

like particles (VLPs) which are morphologically similar to the native virus particles (Jiang *et al.*, 1999, Oka *et al*., 2012).

1.2.5 Transmission

Individuals in any closed environment in which a lot of people from different locations congregate for an extended period are at risk of contracting norovirus. These environments include prisons, long flights, dormitories, day care facilities, elderly care facilities and as recently observed, sports congregations (Olympics 2017) (Mellou *et al.*, 2012, Enquist *et al*., 2015). Noroviruses are extremely contagious and spread through the faecal-oral route (Sarvestani *et al.*, 2016). Transmission by infectious vomit, both through indirect spread from the environment (in other words through hand or mouth contact) and through aerosols created by explosive vomiting (Griffin, 2013, Alsved *et al.*, 2019) might explain the rapid spread of outbreaks in closed settings (Robilotti *et al*., 2015). Often the first case in an outbreak is caused by contact with contaminated food or water (Sarvestani *et al*., 2016). This contamination is then spread further through person-to-person contact (Patel *et al*., 2009). Some characteristics of gastroenteritis viruses such as noro-, rota- and sapovirus allow for the quick spread of these viruses. These characteristics include but are not limited to a low infectious dose (with only as little as 10 virus particles needed for infection), shedding at a high viral load, being relatively stable in the environment, with multiple modes of transmission available. These viruses can survive 10 ppm chlorine and heating to 60˚C, that is why noroviruses are known to be maintained in steamed oysters (Carroll *et al.*, 2015). This results in the need to decontaminate all shared surfaces with chlorine-containing solutions after an outbreak (Enquist *et al*., 2015).

Sapovirus outbreaks are not as common as norovirus outbreaks, but they do occur throughout the year (Yan *et al*., 2003). The virus is transmitted through the faecal-oral route, through contact with contaminated faeces, vomit, or via consumption of contaminated food and drinking water (Oka *et al*., 2015). Food-borne outbreaks have also been suggested as a route of transmission (Kobayashi *et al.*, 2012). Outbreaks usually occur sporadically in similar settings as those of norovirus, such as schools, day-care centres and hospitals (Hansman, 2007).

1.2.6 Susceptibility

Host genetics and heterogeneous host-virus interactions influence the pathogenicity of noroviruses; as a result, individuals are not equally susceptible to norovirus infection

(Lindesmith *et al.*, 2003). Pathogenicity of norovirus is affected by several key factors. These include the viral attachment of norovirus to histo-blood group antigens (HBGAs), the tissue and cellular tropism of a strain in the human host, the host's immune response to infection and the bacterial microbiota of the host (de Graaf *et al*., 2016). A number of significant steps in the understanding of norovirus gene expression and replication have also been made using murine noroviruses. Murine noroviruses (MNoVs) are known to bind to terminal sialic acids and internalised using cholesterol and dynamin. The specific proteinaceous cellular receptor was determined for MNoV, after knockout testing was performed on the *Cd300lf* gene, showing that Cd300lf and Cd300ld function as receptors for MNoV and are essential for viral entry *in vitro* (Haga *et al.*, 2016, Orchard *et al.*, 2016). As the mechanism for human norovirus internalisation is still unknown, it has been suggested that it may also be dependent on cholesterol and dynamin in some manner and that cell tissue tropism may be determined by proteinaceous receptors that are interacting with permissiveness co-factors present at different sites (Wobus *et al.*, 2004, Thorne and Goodfellow, 2014, de Graaf et al., 2016, Orchard et al., 2016). There have been numerous studies and observations of asymptomatic norovirus infection (García *et al.*, 2006, Phillips *et al.*, 2010, Liu *et al*., 2016a), although the mechanism of how norovirus infection can result in asymptomatic or symptomatic infection is still unknown. Possible explanations for the detection of norovirus in asymptomatic individuals include long-term shedding from a previous symptomatic episode as well as true asymptomatic infection due to lack of susceptible factors for symptomatic infection (Qi *et al*., 2018). In South Africa, there is almost no data on asymptomatic norovirus infection, although one study did find a prevalence of 36% (Kabue *et al.*, 2016), which is considerably higher than that reported elsewhere (Qi *et al*., 2018). Improved understanding of asymptomatic norovirus infection is needed to enable a true picture of the overall contribution of norovirus to diarrhoeal disease.

No form of susceptibility or resistance to human sapovirus disease has been identified yet. Histo-blood group antigens and blood types do not seem to play a role in susceptibility (Matussek *et al.*, 2015, Oka *et al*., 2015).

1.2.7 Treatment, prevention, and control

Rehydration is the mainstay treatment of gastrointestinal infection. In most cases, dehydration can be effectively treated with oral rehydration solution (ORS). Adults with severe dehydration should receive intravenous fluids (Enquist *et al*., 2015). No anti-viral agents are available for

the treatment of norovirus specifically as of yet, but recently a drug, called Nitazoxanide, an oral anti-parasitic agent, has been found to have broad antiviral activity and to be active against aetiologies of viral gastroenteritis (Tan *et al.*, 2017). Virus-like particle-based vaccine development for norovirus is still ongoing (Treanor *et al.*, 2014, Leroux-Roels *et al.*, 2017), thus the only current method of control is prevention of spread. The prevention of norovirus and sapovirus is done by interrupting their mode of transmission. The upkeep of strict hygiene by food handlers, surveillance of water to ensure no possible contamination, and maintenance of personal hygiene by sick individuals, avoiding contact with environmental surfaces and other persons, are all suggested as methods to try and prevent the spread of noro- and sapovirus (Patel *et al*., 2009).

Norovirus was known as an uncultivatable pathogen, until recently when replication could be demonstrated in human intestinal enteroids by Ettayebi and team, and the cultivation could be effectively replicated in other laboratories (Ettayebi *et al.*, 2016, Costantini *et al.*, 2018). This enables human host-pathogen research on previously non-cultivatable pathogens, and could give rise to new methods of prevention and treatment of human norovirus infections (Ettayebi *et al*., 2016).The control of norovirus outbreaks is considered a large challenge, necessitating further studies to improve our understanding of the virus and possible vaccine development (Glass *et al.*, 2009). Due to the high disease burden of noroviruses, a functional, effective vaccine is in high demand for high-risk populations, especially for the young, elderly and immunocompromised (Dai *et al*., 2015).

1.3 Rotavirus

1.3.1 History

In 1973 Ruth Bishop and her team identified an abundance of particles of an unknown virus in the cytoplasm of mature epithelial cells lining duodenal villi and in faeces from children admitted to the Royal Children's Hospital, Melbourne with severe gastroenteritis. This led to the discovery of rotavirus (Bishop *et al.*, 1973), which had a wheel-like appearance from which the genus derives its name (Latin rota, "wheel") (Figure 1.9).

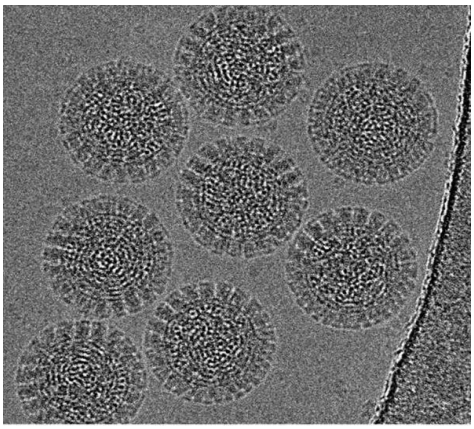


Figure 1.9: Rotavirus particles as seen through EM, showing the characteristic wheel-like appearance. Reproduced from Grant and Grigorieff, 2015.

The human rotavirus forms part of the genus Rotavirus in the family *Reoviridae* (Estes and Kapikian, 2007). Historically, rotavirus has been the most common cause of acute gastroenteritis among infants and children in the developed world. The first rotavirus vaccine, RotaShield (Wyeth Laboratories, Marietta, Pennsylvania) was licensed in 1998, and was advised for distribution in 1999 with high hopes of decreasing rotavirus severity and mortality rates (Control and Prevention, 1999). After the introduction of the RotaShield vaccine, a correlation was observed between vaccinated children and symptoms of intussusception. It was determined that the vaccine was linked to an increased risk of intussusception (Murphy *et al.*, 2001). The vaccine was then taken off the market and this was a large drawback for rotavirus vaccine development. Only in 2006, a new paediatric rotavirus vaccination was broadly recommended in the United States, which has led to a sharp decline in paediatric diarrheal hospitalisations and medical expenditures (Enquist *et al*., 2015).

1.3.2 Clinical and epidemiological features

The symptoms for rotavirus are the same as for norovirus, but typically more severe, leading to more hospitalisations (Griffin, 2013). Rotavirus infects villus epithelial cells of the small intestine without damaging or affecting the gastric mucosa and colon. Virus replication occurs in the cytoplasm of these cells and thus impairs transport of nutrients (Monavari *et al.*, 2017). Rotaviruses are known to cause severe dehydrating gastroenteritis in children under five years old, and has been shown to account for approximately two million childhood hospitalisations and roughly 128 500 child deaths in 2016 (Figure 1.10) (Troeger *et al.*, 2018). This represents

a significant reduction when these numbers are compared to the 528 000 child deaths due to this infection in 2000 (Parashar *et al*., 2009, Tate *et al.*, 2016). It should be noted that this data could underrepresent the actual rotavirus prevalence, as it is only based on reported cases, and up to 50% of cases are asymptomatic and a further 25% are not as severe and do not require hospitalisation. In addition, not all hospitalised patients presenting with gastroenteritis are necessarily screened for rotavirus (Griffin, 2013).

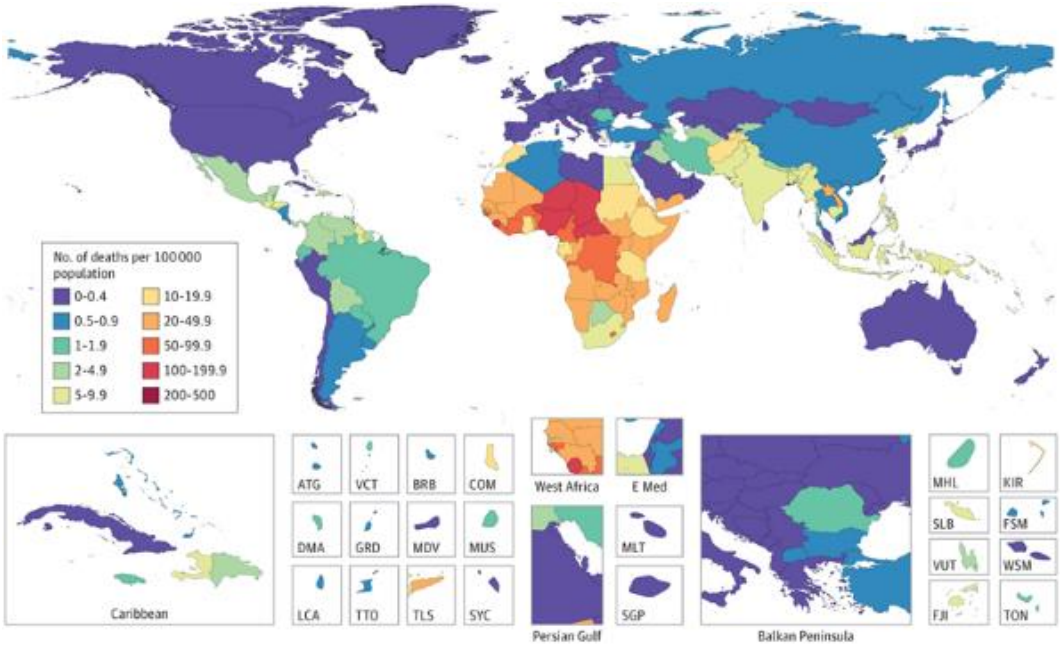


Figure 1.10: Geographic distribution of rotavirus-associated mortality rates among children younger than 5 years in 2016. Reproduced from Troeger *et al*., 2018.

Rotavirus is very contagious, very stable and is shed in great amounts in faecal matter which enhances its transmission. The incubation is usually between 24 to 48 hours, and the duration of the illness lasts between three to five days. Shedding can range from four to 29 days, with an average of about seven days (Richardson *et al.*, 1998). Rotavirus occurs most frequently in children under the age of two years in both high and low resource countries, adults are also frequently infected but tend to be asymptomatic. The virus has been found to be more prevalent in winter months, but no true seasonality can be seen in tropical climates.

1.3.3 Structure

Mature rotavirus particles are non-enveloped, have an icosahedral symmetry and are 100 nm in diameter (when including the spikes) with a T=13 icosahedral protein coat. They also have a capsid with three layers (triple layer particle), each consisting of a different VP (Carter and Saunders, 2007), the inner layer is formed by VP2, the intermediate shell by VP6, and the outer layer is formed by glycosylated VP7 (Xu *et al.*, 2019). Rotavirus particles have a total of 60 projecting spikes and 11 segments of linear double-stranded RNA as is shown in Figure 1.11.

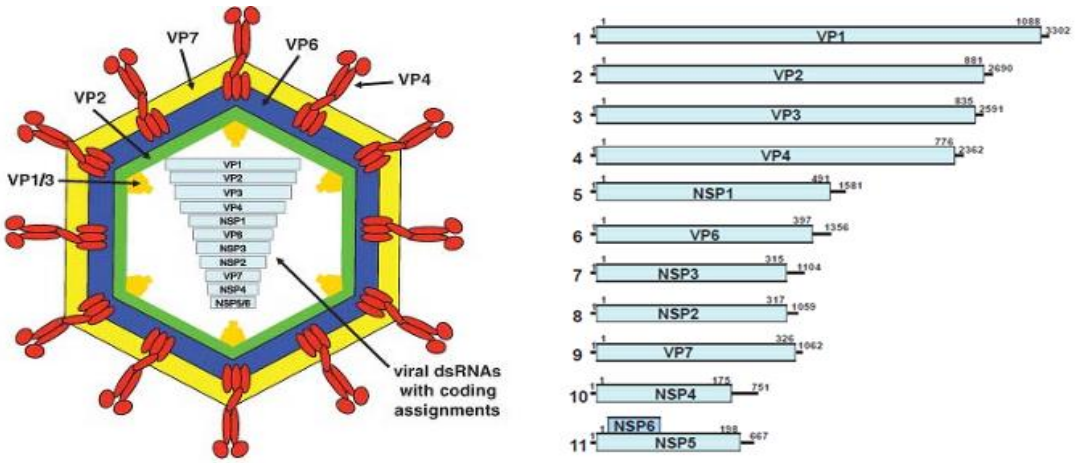


Figure 1.11: Schematic diagram of a rotavirus particle with three protein layers, projecting spikes and 11 segments of double-stranded RNA. Reproduced from Griffin, 2013.

These segments vary in length, from 667 bp (base pairs) to 3302 bp, with a total genome size of 18555 bp and can be separated by size through electrophoresis. Conserved noncoding regions can be found on each RNA segment at the 5' and 3' ends. These regions differ between different rotavirus groups and are important for transcription, replication and genetic reassortment. One protein is encoded per segment, except for NSP5 and NSP6, which are both encoded by segment 11 (Carter and Saunders, 2007). Thus 12 proteins are encoded, which include six structural proteins, and six non-structural proteins (Carter and Saunders, 2007). Table 1.3 gives a short description of each RNA segment, the protein it transcribes and its function, allowing a better understanding of the virus.

Table 1.3: Rotavirus RNA segments, encoded proteins and their functions, adapted from Griffin, 2013.

RNA segment	Protein/Enzyme	Function
1) VP1	RNA-dependent RNA polymerase	ss-RNA binding, forms complex with VP3
2) VP2	Core Protein	Required for replicase activity of VP1
3) VP3	Methyltransferase	Guanylyl transferase
4) VP4	Protease-sensitive structural protein	Neutralisation agent and viral attachment protein
5) NSP1		Interferon antagonist
6) VP6	Structural protein	Inner capsid: Required for transcription
7) NSP3		Translation enhancer: Binds 3' end of viral mRNAs
8) NSP2	NTPase	NTPase: Forms viroplasms with NSP5
9) VP7		Glycosylated: Neutralisation agent
10) NSP4	Enterotoxin	Viroporin, modulates intracellular calcium
11) NSP5	Phosphoprotein	Protein kinase, forms viroplasm with NSP2
NSP6	Phosphoprotein	Protein kinase, forms viroplasm with NSP2

VP4 is cleaved into VP5 and VP8 fragments through proteolysis. The VP8 fragment, which is the globular head of the VP4 spike, has the most variable sequence and interacts with the host's receptors, allowing the virions to attach to the host cells, while VP5 (the stalk of the spike) is thought to be responsible for the penetration of the virus into the cells (Liu *et al.*, 2012b). The two outer capsid proteins, VP4 and VP7 are the determinants of rotavirus P and G types, respectively, which are used to classify rotavirus types on the basis of a dual-nomenclature system (Trojnar *et al.*, 2013). It has been suggested that the different P-type rotaviruses show distinct binding specificities, which may be responsible for determining the host ranges as well as zoonotic transmission (Xu *et al*., 2019).

1.3.4 Classification

There are nine different genogroups of rotavirus, which are further divided into multiple genotypes (Lefkowitz *et al.*, 2017). Groups A to C infect humans as well as animals, and groups D to I only infect animals (Parashar *et al*., 2009, Linhares *et al.*, 2011). Reassortment can occur within a group, but not between the different groups, indicating that each rotavirus group is considered a unique species (Griffin, 2013). Different groups are classified using a 53% amino acid similarity cut-off when comparing the VP6 gene (Matthijnssens *et al.*, 2012). Some rotavirus types may be contained in a specific area, while others have a global prevalence. Emergence, disappearance, and re-emergence are seen to be a common trait with some types (O'Ryan, 2009). Group A rotaviruses are the most prevalent cause of epidemic gastroenteritis, and are especially linked to infections in infants (Griffin, 2013). To date there are 36 G and 51 P accepted genotypes of rotavirus group A (Rotavirus Classification Working Group) (RCWG,

2017) with P[8] and P[4] being first and second predominant types, respectively, followed by P[6] when it comes to causing rotavirus diarrhoea (Zhang *et al.*, 2016b). Predominant G genotypes have been identified as G1 (72.9%), G9 (10.81%) and G2 (8.10%) (Mohanty *et al.*, 2017). These genotypes are responsible for over 95% of human rotavirus infections (Liu *et al.*, 2012b). In South Africa, the combination of G9P[8] and G3P[8] are the most common combinations found according to recent studies (Page *et al.*, 2016b).

1.3.5 Transmission

Rotavirus is transmitted via the faecal-oral route, but it is also suggested that transmission may occur by means of aerosolised particles (Griffin, 2013). Confined settings facilitate transmission, favouring person-to-person spread, with water-borne virus transmission now also considered as a significant factor in rotavirus transmission (Kiulia *et al.*, 2015) and food-borne viral transmission a minor, but important factor in rotavirus transmission as well (Gastañaduy *et al.*, 2013, Lüthi, 2018) .

1.3.6 Susceptibility

The mechanisms responsible for inducing protective immunity against rotavirus infection is not yet fully understood, but increasing levels of immunity occur with repeated exposures or vaccination (Griffin, 2013). As is seen with norovirus, the susceptibility to rotavirus infection is affected by HBGA phenotypes, as rotavirus also uses HBGAs as attachment factors. Malnutrition, as well as the age of the patient, and whether the patient is immunocompromised, can also influence the disease severity.

1.3.7 Prevention, treatment and control

Since 2006, two operational rotavirus vaccines have been licensed. RotaTeq (Merck and Co, PA, USA), is a pentavalent, live bovine-human vaccine containing five reassortant strains (G1, G2, G3, G4 and P8 type). The other vaccine is Rotarix (GSK Biologicals, Rixensart, Belgium) which contains a live attenuated human G1P[8] rotavirus strain (Kirkwood *et al.*, 2011). Vaccination is the best preventative method, but due to limited resources is not always readily available in developing countries (Monavari *et al.*, 2017). Several new, more cost-effective vaccines have recently been pre-approved by the WHO, which will increase options for vaccination in future (Pecenka *et al.*, 2018).

The vaccine protects against severe disease but does not prevent infection (Groome *et al.*, 2014). The development of these vaccines has led to a decrease in the occurrence of severe rotavirus gastroenteritis, especially in high-income countries (Burnett *et al.*, 2017), with a >30% decline from previous estimates (Figure 1.12) (Tate *et al.*, 2016, Troeger *et al.*, 2018). It has been observed that current vaccines are less effective in low-income, high-mortality countries (Groome *et al.*, 2014, Jonesteller *et al.*, 2017). Vaccination has been available in South Africa since 2009 and has led to a marked decrease in mortality due to rotavirus. Vaccine efficacy in South Africa is estimated to be between 54 to 76.9%, which is much lower than the 85% efficacy reached in European and North American countries (Ruiz-Palacios *et al.*, 2006, Groome *et al.*, 2014). This discrepancy is thought to be caused by multiple factors such as malnutrition, poor water and sanitation quality, and the variability of the intestinal microbiota (Patel *et al.*, 2009, Harris *et al.*, 2016).

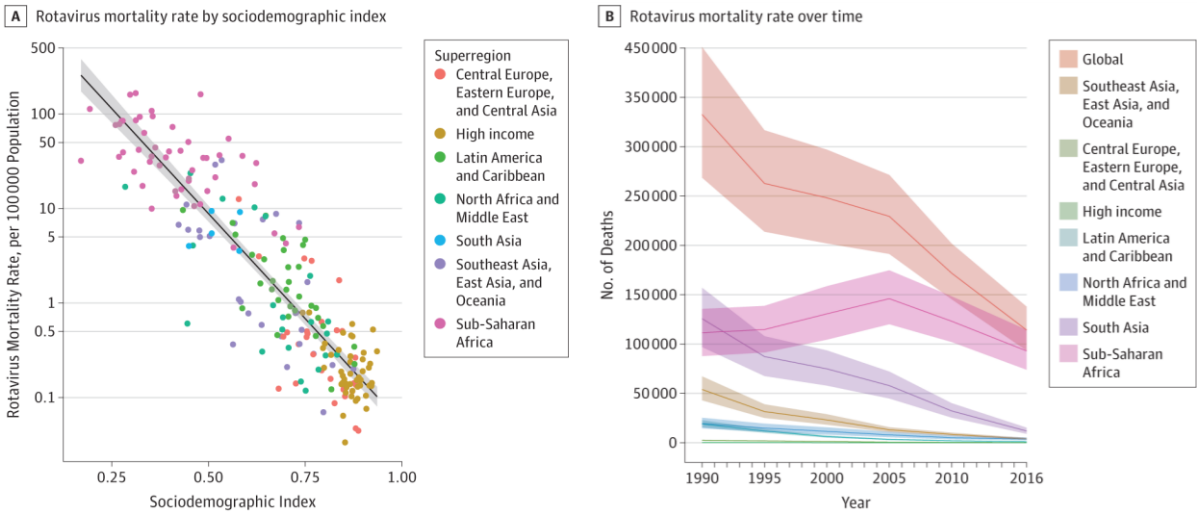


Figure 1.12: A) Rotavirus-associated mortality rate by sociodemographic index in 195 countries, estimated for 2016. B) Rotavirus mortality rate over time, globally and by super region. Reproduced from Troeger *et al.*, 2018.

1.4 Histo-blood group antigens and fucosyltransferase 2

1.4.1 Chemical composition and functions

Histo-blood group antigens are a major group of complex carbohydrates and are determinants of both human and animal ABO blood groups and the Lewis blood group systems (Hansman

et al., 2010, de Graaf *et al*., 2016), which are abundant on the mucosal epithelia of the gastrointestinal tract (Zhang *et al*., 2016b). The secretion and distribution of HBGAs are dependent on secretor (*Se)* enzyme activity [encoded by *fucosyltransferase* 2 (*FUT2*)] (Hansman *et al*., 2010). Fucosyltransferase 2 is an enzyme that regulates the expression of the H antigen in secretory glands and the intestinal mucosa (McGovern *et al.*, 2010). Gene expression of active *FUT2* leads to the synthesis of H type 1 and H type 2 HBGAs which are detectable on mucosal surfaces, in saliva and other bodily fluids, such as blood, and milk (Hansman *et al*., 2010, Frenck *et al.*, 2012). Individuals with non-functional *FUT2* genes are given a non-secretor status, and lack certain HBGAs on their cells and in secretions (de Graaf *et al*., 2016). Histo-blood group antigens are generated through the transfer of N-Acetylgalactosamine (GalNAc) and Galactose (Gal) to an H precursor structure, regardless of the carbohydrate core structure (Shirato, 2011). The HBGA precursors can be divided into 4 major types: Type 1 (Galβ1-3GlcNAcβ), Type 2 (Galβ1-4GlcNAcβ), Type 3 (Galβ1-3GalNAcα) and Type 4 (Galβ1-3GalNAcβ) (Hansman *et al*., 2010). Figure 1.13 illustrates the synthesis of HBGAs (de Graaf *et al*., 2016). One should take into account that although A, B and H HBGAs are the same as the A, B and O blood groups, the ABO blood group system is not dependent on the secretor status (Shirato *et al.*, 2008). This is because these HBGAs, expressed on the surfaces of erythrocytes, are synthesised by *FUT1* rather than *FUT2*.

It must be noted that HBGA phenotyping is not as straightforward as would originally appear. A recent study has highlighted the effects of child as well as maternal HBGAs on symptomatic and asymptomatic infections, specifically in early childhood (Colston *et al.*, 2019). The study found that FUT2 positive infants and mothers have a statistically significant risk of diarrhoea, with a reduced time to the first diarrhoeal episode. Infants with FUT2 positive mothers showed an increase in cumulative incidence (85.2%) when compared to infants of non-secretor mothers (52%), with a 2.24 times higher risk of infection up to 5 months of age (Colston *et al*., 2019). This is due to the fact that FUT2 and FUT3 phenotype combinations in breastfeeding mothers can influence the distribution and concentration of human milk oligosaccharides, which are expressed in breastmilk. These changes in breastmilk composition are then seen to alter the child's microbiome, which in turn may alter resistance to enteric infections (Zivkovic *et al.*, 2011, Lewis *et al.*, 2015, Colston *et al*., 2019).

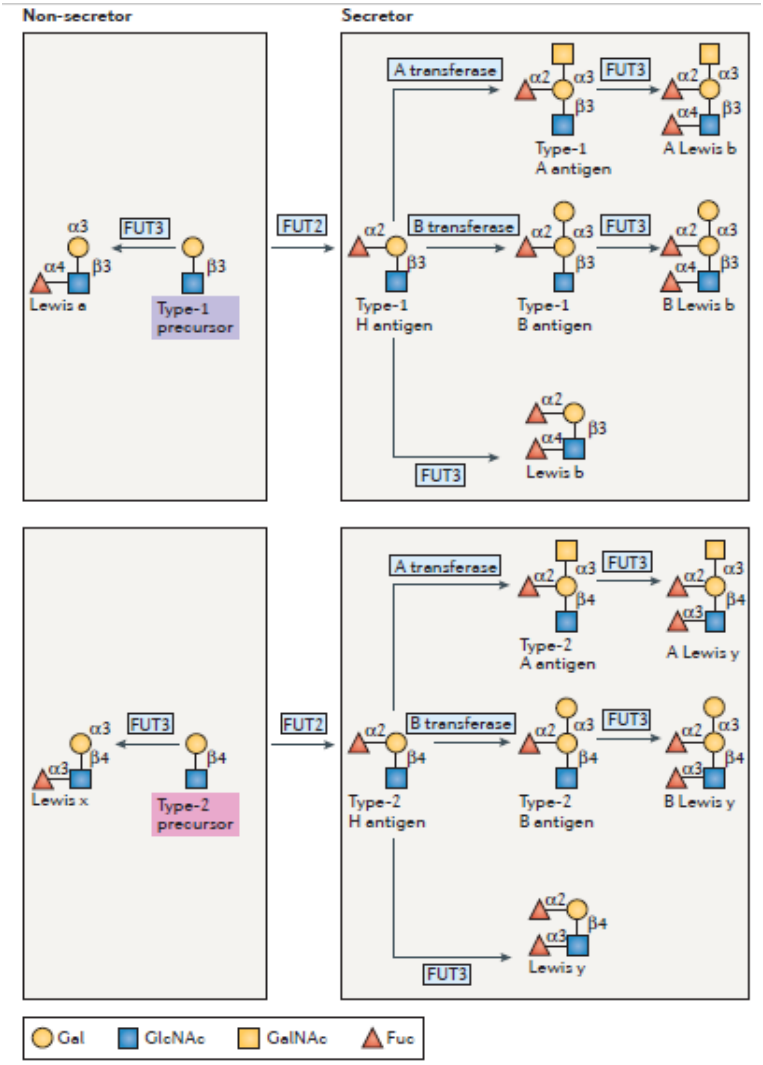


Figure 1.13: HBGA Synthesis Pathways. Type 1 and Type 2 HBGA precursors are modified by α (1, 2)-FUT2 to produce H HBGAs. These are further modified by A and B transferases to produce A and B HBGAs, reproduced from de Graaf *et al*, 2016.

No association has been made between secretor status of mother or child and risk of infection with sapovirus (Matussek *et al*., 2015, Oka *et al*., 2015, Colston *et al*., 2019).

1.5 HIV-exposed, uninfected infants

In the late 1990's diarrhoeal disease has been described as a frequent complication as well as a common cause of hospitalisation and death of HIV infected (HI) children (Lew *et al.*, 1997). Human immunodeficiency virus was first isolated in Paris in 1983 (Gallo RC, 1987). Over the years progress has been made to combat HIV, with highly active antiretroviral therapy (HAART) playing a central role in both treatment and secondary prevention of spread of the disease, as recommended in national guidelines.

Highly active antiretroviral therapy has shown a marked improvement in the health and the long-term prognosis of HIV infected patients. This therapy has also markedly reduced the risk of mother-to-child transmission of the virus (Claire Thorne, 2005), which, in turn, has led to an increase in the proportion of HIV-exposed uninfected (HEU) children (Claire Thorne, 2005). Recent studies have shown that a total of approximately 7.7 million people of all ages are infected with HIV in South Africa. These studies also indicate that the transmission rate from mother to child in South Africa is approximately 9% (shown in Figure 1.14) (UNAIDS, 2019).

Figure 1.14: Cascade of services for preventing vertical transmission, numbers of new HIV infections and transmission rate, eastern and southern Africa, 2018. Reproduced from UNAIDS, 2019 data.

The first observation that children who were HEU might have an increased susceptibility to infections was in Kenya 1992, where HEU infants were found to have a high incidence of measles (Embree *et al.*, 1992). Various studies have been performed on HEU children, to determine the effects of HIV exposure. These effects included children having increased mortality rates (up to three times higher), increased infectious morbidity, impaired growth as well as a higher chance of HIV infection and other potential intracellular infections, when compared to HIV unexposed (HU) infants (Feiterna-Sperling *et al.*, 2007, Marinda *et al.*, 2007,

Epalza *et al.*, 2010, Slogrove *et al.*, 2012, Bunders *et al.*, 2014, Evans *et al.*, 2016). One must consider that this data may be distorted and highly affected by feeding practices, which are ever-changing as breastfeeding recommendations for HI mother's change over time (Rollins *et al.*, 2013, Evans *et al.*, 2016). Another consideration for HEU infants is that they are also at higher risk of mortality when compared to HU children due to factors such as poor care given and lower socio-economic status because of the sick mother (and the father as well). These infants also stand a chance of increased exposure to infections from their immunocompromised parents (Marinda *et al.*, 2007).

1.6 Motivation

Understanding the prevalence of and relationship between causative viruses of gastroenteritis can assist with appropriate infection control and vaccine development (Brown *et al.*, 2016). Further studies are required to gain a better understanding of the factors, which could potentially influence the susceptibility of subjects to norovirus and rotavirus, such as HIV exposure. More detailed characterisation of the pathogen- and strain-specific effects of HBGAs on enteric infections can inform the development of precision public health and improve the success of regionalised and targeted interventions (Colston *et al.*, 2019). Efforts are underway to develop a norovirus vaccine and information regarding the diversity of noroviruses in the aforementioned groups, would be critical to informing developers which norovirus strains to include in the candidate vaccines. Saliva, blood and stool samples which have previously been collected, were analysed and compared to see if there is a correlation between *FUT2* genotype, HBGA phenotype, the occurrence and type of virus found in children with gastroenteritis and if HIV exposure has any effect on this.

1.7 Aim

The overall aim of this study was to investigate the relationship between gastroenteritis virus infections and host secretor status, as well as to determine whether HIV exposure affects gastroenteritis virus infections in children in terms of pathogen diversity and severity.

1.8 Objectives

1. To screen stool specimens from children under the age of five years, hospitalised with diarrhoea for gastroenteritis viruses using multiplex RT-PCR.

2. To genotype the norovirus, sapovirus and rotavirus strains detected in these children by RT-PCR and nucleotide sequencing.
3. To screen follow up samples at 6 weeks post gastroenteritis (where available) of children for norovirus GI and GII with singleplex RT-PCR assays.
4. To determine the fucosyltransferase-2 secretor status of the children by real-time PCR.
5. To determine the saliva carbohydrate phenotype of all the children.
6. To confirm *FUT2* genotype by nucleotide sequencing if discrepant *FUT2* genotype and saliva carbohydrate phenotype was observed.
7. To combine the data on virus detection, genotyping, secretor status and saliva phenotype generated in objectives 1-4 on specimens collected from March 2017 to December 2017 with data previously generated for specimens collected from July 2016 to February 2017.
8. To determine the co-infection rate of the virus detection results with previously generated data on bacteria and parasite detection in the same specimens.
9. To compare the prevalence and severity of norovirus and rotavirus in HIV-exposed uninfected and HIV unexposed children.
10. To perform next-generation sequencing on a norovirus-positive specimen from one HI, one HEU and one HU child respectively.

Chapter 2 : Population demographics and virus epidemiology

2.1 Introduction

Viruses are among the most common causes of both endemic and epidemic gastroenteritis, infecting millions of people per year, leading to high morbidity and mortality (Hale *et al.*, 2015, IHME, 2018). African and South East Asian regions are disproportionately affected by viral gastroenteritis (UNICEF, 2016, IHME, 2018). The most attributable viruses, in descending order, include rotavirus, adenovirus-F, norovirus, astrovirus and sapovirus (Liu *et al.*, 2016a).

Severity of rotavirus infection has decreased dramatically with the introduction of rotavirus vaccines, especially in high-income countries (Burnett *et al.*, 2017). However, the prevalence of this virus in hospitalisations in South Africa has been estimated to still be at least 15% (Makgatho *et al.*, 2019), with the combination of G9P[8] and G3P[8] the most common strains found (Page *et al.*, 2015). The overall prevalence of noroviruses in children with gastroenteritis in low-income countries in Africa has been estimated to be 13.5% (Mans *et al.*, 2016). One particular study on wastewaters in South Africa between 2015 and 2016, determined that norovirus was detected in large quantities of water specimens collected from multiple wastewater treatment plants (Mabasa *et al.*, 2018). This study also found that norovirus GII.2 and GII.17 predominated in the wastewaters, as opposed to GII.4 which is known to predominate in clinical surveillance (Mabasa *et al.*, 2018). Understanding the prevalence of asymptomatic norovirus infections is also important as asymptomatic individuals may facilitate the transmission of this virus. This information could be useful in successfully presenting and applying public health control policies. Previous studies in children in SA also identified sapovirus infection at a prevalence of approximately 8% (Page *et al.*, 2016a) with GIV as the most prevalent strain found in the country, followed by GI.2 (Murray *et al.*, 2016). Astrovirus prevalence was observed to be 7% in this population (Nadan *et al.*, 2019), and is more often observed in co-infections with other gastroenteritis viruses in hospitalised patients (Nadan *et al.*, 2019).

A thorough understanding of the global burden of gastroenteritis virus infection and illness as well as genotype diversity, dominant strains, and strain replacement patterns is essential for infection control and effective vaccine development.

2.2 Materials and methods

2.2.1 Ethical approval

This project formed part of a larger study, which obtained approval by the Research Ethics Committee, Faculty of Health Sciences of the University of Pretoria – Ref: 362/2015. A pilot study with a sample size of 85 children was completed in 2017 and all findings have been combined with the 120 children from this study. The pilot study obtained ethics approval, Ref: 90/2017. Further ethics approval was obtained in April 2018 by the Research Ethics Committee, Faculty of Health Sciences of the University of Pretoria – Ref: 182/2018 (renewed on 10/04/2019). The relevant documents are available in Appendix B. A unique study number was assigned to each participant to ensure patient confidentiality.

2.2.2 Study design

The overall strategy was to combine analysis of gastroenteritis viruses with that of FUT2 genotype and saliva carbohydrate phenotype in order to study the relationship between the host HBGA profile and gastroenteritis virus infection. Furthermore, the virus type and diversity, as well as severity of infection, was compared between HI-, HEU- and HU children to evaluate the effect of HIV exposure on viral gastroenteritis in children. Stool specimens collected from children with gastroenteritis were screened for six known gastroenteritis viruses (norovirus GI and GII, sapovirus, astrovirus, adenovirus-F and rotavirus) and the norovirus, rotavirus and sapovirus strains detected in these children were genotyped. In order to study the link between norovirus and rotavirus and host susceptibility, secretor status was determined through the FUT2 genotyping of blood (or stool samples if blood was not available) as well as the HBGA phenotyping from saliva. All the data from the previous pilot study (n=85) and the current research project (n=120) were combined (n=205) and analysed to determine the relationship between gastroenteritis viruses and secretor status (FUT2 genotype and saliva carbohydrate phenotype) of patients who are either HI, HEU or HU.

2.2.3 Study population

From June 2016 to December 2017, 221 children below the age of five years hospitalised with diarrhoea, at Kalafong Provincial Tertiary Hospital (KPTH), were enrolled in this study. Informed consent was obtained from each parent/caregiver and demographic and clinical information as well as HIV status information, was collected with a questionnaire (University of Pretoria, Faculty of Health Sciences Human Ethics Committee Protocol 362/2015, Appendix B1). Stool specimens, whole blood and saliva specimens were collected from these participants. Parents/caregivers were requested to submit a follow up stool specimen six weeks after enrolment. Participants were divided into three categories, based on their HIV status, namely HI, HEU and HU. Severity of disease was categorised using the Vesikari clinical severity scoring system (Table 2.1), into either mild, moderate or severe categories (Ruuska and Vesikari, 1990, Freedman *et al.*, 2010, Schnadower *et al.*, 2013). Specific criteria had to be met to be considered for this study. These criteria included all children under the age of five years to be hospitalised with gastroenteritis, with adequate stool and saliva specimens obtained from each participant. A total of 205 children met this criteria and were accepted into the study.

Table 2.1: A) Vesikari clinical severity scoring system, reproduced from Ruuska and Vesikari, 1990, B) Vesikari Clinical Severity Scoring System Severity Rating Scale.

A)	**Score**		
Parameter	1	2	3
Diarrhoea			
Maximum number stools per day	1-4	5	≥6
Diarrhoea duration (Days)	1-4	5	≥6
Vomiting			
Maximum number vomiting episodes per day	1	2-4	≥5
Vomiting Duration (Days)	1	2	≥3
Temperature	37.1-38.4	38.5-38.9	≥39.0
Dehydration	N/A	1-5%	≥6%
Treatment	Rehydration	Hospitalization	N/A

B) Severity Category			
Mild	Moderate	Severe	Maximum score
<7	7-10	≥11	20

2.2.4 Specimen preparation

The 205 stool specimens were stored at -20°C and 10% suspensions were prepared using ultrapure sterile water (Promega, Madison, WI,). The 10% suspensions were centrifuged at 14000 x g for five minutes and the supernatant used for total nucleic acid extraction.

2.2.5 Total nucleic acid extraction

Automated total nucleic acid extraction was performed on 200 µL of 10% stool suspension supernatant using the automated NucliSENS® EasyMAG® Instrument (BioMérieux, Marcy-l'Étoile, France). The nucleic acids were then eluted in 50 µL and stored at -80°C. An internal control included in the screening kit was added to each specimen during each extraction experiment according to the manufacturer's instructions.

2.2.6 Enteric pathogen detection

A multiplex real-time reverse transcription PCR (Allplex^{TM} Gastrointestinal Virus Panel, Seegene, Seoul, South Korea), was used to screen each specimen for six gastroenteritis viruses (Norovirus GI & GII, sapovirus, rotavirus, adenovirus and astrovirus), on the Bio-Rad CFX platform (Bio-Rad Laboratories, Hercules, CA). The one-step RT-PCR master mix (20 micro-litre [µL]) was prepared according to the kit's instructions. Nucleic acids were incubated at 95˚C for two minutes (min) and then placed on ice for two minutes to ensure the separation of rotavirus's double-stranded RNA and then 5 µL was added to the master mix for each specimen. The reaction mix was then capped and centrifuged to ensure that all the liquid was at the bottom of the PCR tube and to eliminate bubbles. A positive and a negative control, provided with the kit, were also included in each test run to assess the validity of the test results, as well as an internal control for each sample (added before nucleic acid extraction). The cycling conditions were as follows: an initial step of 50˚C for 20 min to generate complementary deoxyribonucleic acid (cDNA), followed by initial denaturation at 95˚C for 15 min, thereafter, 45 cycles of denaturation (95˚C, 10 seconds (sec)), annealing (60˚C, 1 min) and extension (72˚C, 30 sec) were performed and the fluorescence was measured at the annealing and extension steps. After each run, Seegene Viewer analysis software (Seegene Inc.) was used to analyse the results. A Ct value over 39 was considered negative for all viruses.

Stool specimens were screened for bacteria and parasites associated with gastroenteritis by standard microscopy and culture microbiological detection methods at the National Health Laboratory Service (NHLS) Tshwane Academic Division diagnostic microbiology laboratory.

All patient caregivers were asked to return with a follow-up stool specimen from the participant six weeks after initial enrolment into the project. Nucleic acid was extracted from all follow-up specimens as described above and these were screened for norovirus GI and GII with a real-time RT-PCR using the Qiagen QuantiFast Pathogen RT-PCR and IC kit (Qiagen Incorporated, Hilden, Germany), according to manufacturer's instructions, with specific primers and probes for norovirus GI and GII as shown in Table 2.2.

Table 2.2: Primers and probes for detection of norovirus GI and GII

Virus	Primer/Probe	Sequence
Norovirus GI	Forward primer: QNIF4	(5'-CGC TGG ATG CGN TTC CAT-3')
	Reverse primer: NV1LCR	(5'-CCT TAG ACG CCA TCA TTT AC-3')
	Probe: Norovirus GI	FAM - TGG ACA GGA GAY CGC RAT G – TAMRA
Norovirus GII	Forward primer: QNIF2	(5'-ATG TTC AGR TGG ATG AGR TTC TCW GA-3')
	Reverse primer: COG2R	(5'-TCG ACG CCA TCT TCA TTC ACA-3')
	Probe: QN1FS	FAM - AGC ACG TGG GAG GGC GAT CG - TAMRA

2.2.7 cDNA synthesis, amplification and sequencing

For viral genotyping, cDNA was generated via reverse transcriptase PCR. The same cDNA protocol with Protoscript II (New England Biolabs., (NEB), Ipswich, MA) and random primers (Roche Diagnostics Corp., Mannheim, Germany) was used for norovirus, sapovirus and rotavirus genotyping (unless stated otherwise). The reaction mix consisted out of 60 picomol (pmol) random hexamer, 1 millimolar (mM) dNTPs, 1X Protoscript II RT Reaction Buffer, 0.02 M DTT, 40 U Protector RNase inhibitor (Roche), 200 U Protoscript II reverse transcriptase and was made up to the final volume (10 µL) with nuclease-free water. Ten microliters of each RNA sample were then added to the reaction mixtures (final volume: 20 µL) and cDNA was synthesized. Since rotavirus is double-stranded, an extra step, in which

the RNA sample was boiled for two minutes and then placed on ice for two minutes was performed before adding the RNA to the reaction mixture. The exact PCR cycling conditions for each virus's cDNA synthesis and amplification is shown in Table 2.3.

Table 2.3: Cycling parameters for the amplification of the genotyping regions of norovirus, rotavirus and sapovirus.

Specific Virus Amplification	Norovirus One-step kit	Norovirus (Round 1)	Norovirus (Round 2)	Rotavirus (Round 1-2)	Sapovirus (Round 1)	Sapovirus (Round 2)
Kit Used	One-Step RT-PCR kit (Qiagen Inc)	EmeraldAmp® MAX HS PCR (Thermo Scientific, Waltham, MA)		GoTaq Hotstart Polymerase kit (Promega)	The EmeraldAmp® MAX HS PCR	
cDNA Preparation	42˚C; 30 min	25°C; 10 min 42°C; 60 min 85°C; 5 min	-	25°C; 10 min 42°C; 60 min 85°C; 5 min	25°C; 10 min 42°C; 60 min 85°C; 5 min	-
Initial denaturation	95˚C; 5 min	95˚C; 10 min	95˚C; 10 min	95˚C; 1 min	95˚C; 10 min	95˚C; 10 min
Denaturation	95˚C; 1 min	94˚C; 30 sec	94˚C; 30 sec	95˚C; 1 min	94˚C; 30 sec	94˚C; 30 sec
Annealing	45˚C; 1 min	50˚C; 30 sec	55˚C; 30 sec	42˚C; 1 min	50˚C; 30 sec	52˚C; 30 sec
Extension	72˚C; 80 sec	72˚C; 2 min	72˚C; 2 min	72˚C; 1 min	72˚C; 2 min	72˚C; 1 min
Cycles	40	40	40	35	40	45
Final extension	72˚C; 10 min	72˚C; 10 min	72˚C; 10 min	72˚C; 7 min	72˚C; 10 min	72˚C; 10 min

Reagents used for the PCR reactions were obtained from New England Biolabs, Ipswich, MA, unless stated otherwise. Nuclease-free water was obtained from Promega and primers from Inqaba Biotechnical Industries (Pretoria, South Africa).

2.2.8 Caliciviruses:

2.2.8.1.1 Norovirus amplification

The norovirus-positive total nucleic acid was subjected to RT-PCR for genotyping using the One-Step RT-PCR kit (Qiagen Inc). The reverse transcription and PCR were carried out sequentially in the same tube using this kit, so no extra cDNA preparation was needed. The reaction mixture (25 µL) comprised of 1X One-Step RT-PCR Buffer, dNTP mix (0.6 mM), 1.5 µL One Step RT-PCR Enzyme Mix, RNase-free water, primers (1.2 µM) (Inqaba Biotec; Table 2.4) and 5 µL extracted nucleic

acid with PCR cycling conditions shown in Table 2.3.

Table 2.4: Primers used for the amplification of norovirus ORF1/ORF2 (A-C or RdRp/Cap) junction for genotyping.

Genogroup	Primer Name (polarity)	Sequence (5' – 3')	Product Size (bp)
GI	JV12Y	ATACCACTATGATGCAGAYTA	~1.1 kb
	G1SKR (-)	CCAACCCARCCATTRTACA	
GII	JV12Y	ATACCACTATGATGCAGAYTA	~1.1 kb
	G2SKR (-)	CCRCCNGCATRHCCRTTRTACAT	

The amplified region spanned across the ORF1/ORF2 (A-C region or RdRp/Cap) junction and the size of the PCR-products were approximately 1.1 kb for GI and GII.

If amplification with the one-step was unsuccessful, a two-step PCR was performed to amplify the B-C region. The EmeraldAmp® MAX HS PCR Master Mix (Thermo Scientific) was used for the genotyping of viruses in these specimens. The PCR mixture (50 µL) consisted out of 25 µL EmeraldAmp MAX HS PCR Master Mix, <500 ng template, 0.2 µM of the forward and reverse primers (shown in Table 2.5), 5 µL cDNA (section 2.2.7) and nuclease free water. This mixture was prepared for each genotyping PCR.

The norovirus strains were characterised by amplifying and sequencing the partial RdRp (region B) and capsid (region C) gene region with the primers as described previously (Kojima *et al.*, 2002, Richards *et al.*, 2004b) (Table 2.5). The following cycling parameters were used (according to EmeraldAmp specifications): enzyme activation and initial denaturation at 95°C for 2 min, followed by 30 cycles of 98˚C for 10 sec, 60˚C for 30 sec and 72˚C for 1 min/kb.

Table 2.5: Sequences of the primers used for region BC genotyping of norovirus GI and GII.

Norovirus	Primer	Sequence (5'-3')	Polarity	Location
GI	MON432[a]	TGGACICGYGGICCYAAYCA	+	5291-5308
	G1SKF[b]	CTGCCCGAATTYGTAAATGA	+	5342-5361
	G1SKR[b]	CCAACCCARCCATTRTACA	-	5653-5671
GII	MON431[a]	TGGACIAGRGGICCYAAYCA	+	5012-5037
	G2SKF[b]	CNTGGGAGGGCGATCGCAA	+	5046-5064
	G2SKR[b]	CCRCCNGCATRHCCRTTRTACAT	-	5367-5389

[a](Richards *et al.*, 2004a), [b] (Kojima *et al.*, 2002).

2.2.8.2 Sapovirus amplification

Sapoviruses were characterised by amplifying and sequencing approximately 350-400 bp of a partial 5'-region of the capsid gene. The partial capsid region of sapovirus was amplified by nested PCR, using the primers described (Sano *et al.*, 2011) (Table 2.6). The EmeraldAmp® MAX HS PCR Assay (Thermo Scientific) was used for the genotyping of sapoviruses. The PCR mixture (50 µL) consisted of 25 µL EmeraldAmp MAX HS PCR Master Mix, <500 ng template, 0.2 µM of the forward and reverse primers (shown in Table 2.6), and nuclease-free water. This mixture was prepared for each genotyping PCR with cycling conditions as shown in Table 2.3.

Table 2.6: Primers for sapovirus PCR reproduced from Sano *et al.*, 2011.

Nested PCR	Primer/probe	Sequence (5'-3')	Polarity	Location (nt)
First round	SV-F13	GAYYWGGCYCTCGCYACCTAC	+	5074-5094
	SV-F14	GAACAAGCTGTGGCATGCTAC	+	5074-5094
	SVR-DS3	GGTGAVAVMCCATTYTCCAT	-	5857-5876
	SVR-DS4	GGHGAHATNCCRTTBYSCAT	-	5857-5876
Second round	SaV1245Rfwd	TAGTGTTTGARATGGAGGG	+	5159-5177
	SVR-DS5	CCCCACCCKGCCCACAT	-	5482-5498
	SVR-DS6	CCCCAMCCMGCMMACAT	-	5482-5498

In the first round of PCR, 2 µl cDNA was added to the PCR mixture. The nested PCR then contained 5 µl of the first-round product.

2.2.9 Rotavirus

Rotavirus genotyping was performed with a nested PCR using the EmeraldAmp® MAX HS PCR Assay (Thermo Scientific) and could be divided into two rounds, summarised in Figure 2.1.

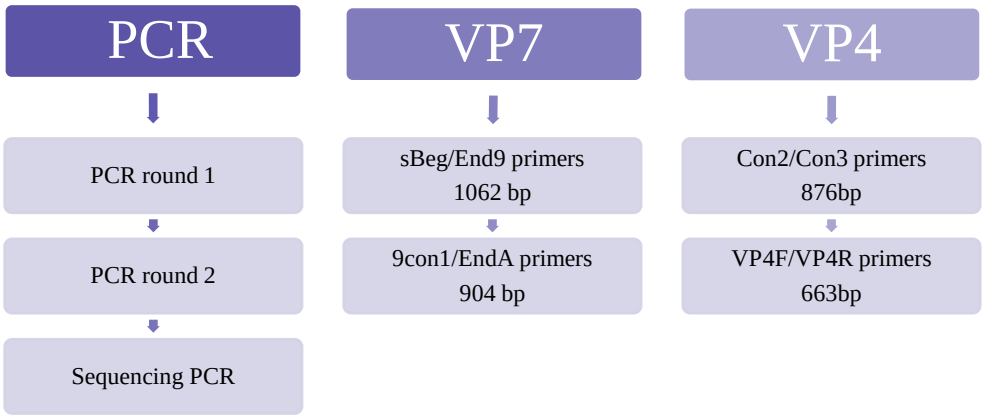


Figure 2.1: Brief overview of rotavirus nested PCR.

For the first round of VP7 and VP4 PCRs the following master mixes were prepared: 0.25 µM primer (sBeg/End9 for VP7 and Con2/Con3 for VP4) with 25 µL Emerald master mix (Thermo Scientific) made up to a total volume of 40 µL with ddH_2O (Takara Bio Incorporated., Shiga, Japan). After the master mix was aliquoted, 10 µL cDNA of each sample was added and amplified.

One microliter of the first round PCR product was used as a template for the nested PCR, with an identical master mix (50 µL), and primers 9con1/EndA for VP7 and VP4F/VP4R for VP4 (Table 2.7). The cycling conditions were identical to those of the first round (Table 2.3).

Table 2.7: VP7- and VP4 specific primers utilised during nested PCRs.

VP type	Round	Primer	Sequence 5'-3'	Position
VP7	First	sBeg9[a] (F)	GGCTTTAAAAGAGAGAATTTC	1-21
		End 9[a]	GGTCACATCATACAATTCTAATCTAAG	1062-1036
	Second	9con1[b](F)	TAGCTCCTTTTAATGTATGG	37-56
		EndA[c] (R)	ATAGTATAAAATACTTGCCACCA	944-922
VP4	First	Con3[d] (F)	TGG CTT CGC TCA TTT ATA GAC A	11–32
		Con2[d] (R)	ATT TCG GAC CAT TTA TAA CC	868–887
	Second	VP4F[e]	TATGCTCCAGTNAATTGG	775-795
		VP4R[e]	ATTGCATTTCTTTCCATAATG	795-775

[a](Gouvea *et al.*, 1990), [b](Das *et al.*, 1994), [c](Gault *et al.*, 1999), [d](Gentsch *et al.*, 1992), [e](Simmonds *et al.*, 2008)

2.2.10 PCR product analysis

RT-PCR products were separated on 1.5% LE agarose gels. The gels were prepared using 1.5 – 2.25 grams (g) SeaKem® LE Agarose (FMC Corporation, Philadelphia, PA) and 100 – 150 mL 1 x Tris-acetate-ethylenediaminetetraacetic acid (TAE) buffer (volumes differed depending on gel size). The solution was then boiled and cooled before adding 5-7.5 µL of 10 mg/mL ethidium bromide (EtBr) (Thermo Scientific). A 1 kb plus DNA ladder (Thermo Scientific) was used as reference to determine the size of the products. The voltage was set between 90 and 120 volts depending on the size of the gel. The Gel DocTM XR+ System (Bio-Rad Laboratories) was used to visualise the PCR products after electrophoresis.

2.2.10.1 DNA cloning and colony PCR

In the event of mixed norovirus, sapovirus or rotavirus infections, the CloneJET PCR Cloning Kit (Fermentas Inc., Glen Burnie, MD) was used to clone amplified segments. The RT-PCR products were purified using the DNA Clean and Concentrator-25 kit (Zymo Research, Irvine, CA) and eluted in 30 µL. The products were cloned into linearised pJet 1.2 vectors which were transformed into chemically competent E.coli® cells (Lucigen ® Corporation, Middleton, WI) according to the manufacturer's instructions. The cells were subsequently plated onto agar plates with 100 µg/mL ampicillin and incubated overnight at 37˚C. Five clones were randomly selected and subjected to colony PCR (cPCR). The cPCR reaction mixture (20 µL) comprised of OneTaq® Quick-Load® 1X Master Mix with Standard Buffer (NEB Inc.), 0.2 µM forward and reverse primers (NEB Inc.) and nuclease-free water (Promega Corp.). The amplification protocol was as follows: initial denaturation

(95°C; 30 sec), followed by 30 cycles denaturation (94°C; 30 sec), annealing (60°C; 30 sec), extension (68°C; 1 min) and a final extension (68°C; 5 min).

2.2.10.2 Sanger sequencing

All amplicons were directly sequenced in both directions using the ABI PRISM BigDye® Terminator v. 3.1 Cycle sequencing kit (Applied Biosystems, Foster city, CA). For every sample, a forward and reverse sequencing reaction mixture was prepared. The reaction consisted of 3 µL of 5X sequencing buffer, 1 µL Terminator mix, 3.2 pmol forward or reverse primer and 13 µL nuclease free water in a total volume of 18 µL. Two microliters of the PCR products were then added to the sequencing PCR mix bringing the total volume to 20 µL. The sequencing PCR was performed using the same primers as described in Tables 2.5, 2.6 and 2.7 above for each virus, with the cycling parameters as follows: 94°C for 3 min; 94°C for 30 sec, 50°C for 10 sec, 60˚C for 4 min (25 cycles). If very low PCR product yields or mixed sapovirus or norovirus sequences were detected in a sample, the amplicons were cloned using the CloneJET™ PCR cloning kit as described in section 2.2.10.1, and randomly selected clones were sequenced using pJET1.2/blunt specific primers (Thermo Scientific).

After a sequencing PCR was performed the samples were sent to Inqaba Biotech for purification and analysis on an ABI 3130 automated analyser to determine nucleotide sequences.

2.2.11 Phylogenetic analysis

Sequences received were analysed using the Sequencher DNA Sequence Analysis Software (Gene Codes Corporation, MI, United States), BioEdit Sequence Alignment Editor (Hall, 1999) and BLAST-n (Altschul *et al.*, 1997). The polymerase and capsid genotypes for norovirus were determined using the Norovirus Genotyping Tool (**https://www.rivm.nl/mpf/typingtool/norovirus/**) (Kroneman *et al.*, 2011). Sequences for norovirus, rotavirus and sapovirus were aligned with reference strains using MAFFT version 6 (**http://mafft.cbrc.jp/alignment/server/index.html**) and phylogenetic analysis was performed in MEGA version 6.0. The evolutionary distances were determined using the Kimura 2-parameter method and phylogenetic analysis for each virus was performed using the neighbour-joining method, validated by 1000 bootstrap replicates. Genotypes

were assigned based on clustering with reference strains in the phylogenetic tree with >70% bootstrap support (Kimura, 1980, Felsenstein, 1985, Saitou and Nei, 1987).

2.3. Results

2.3.1 Population information

2.3.1.1 Enrolment information

Between July 2016 and December 2017, a total of 221 children hospitalised with gastroenteritis (<5 years of age) were recruited for the study at KPTH. The required stool, blood and saliva specimens were collected from 205 children. Enrolment varied from one month to the next, with an average of between seven and 17 children enrolled per month. Maximum enrolment was observed in January 2017 (33 children).

2.3.1.2 Age demographics

The median age of the study population was 10 months (13 days minimum, 64 months maximum). Eighty-two percent of specimens (169/205) were collected from children ≤2 years of age, with the largest contribution from children between 7 and 12 months.

2.3.1.3 Residential area demographics

Residential areas were also recorded, with a wide distribution throughout Gauteng (Figure 2.2). Most enrolments were from Saulsville (n=48) and Olievenhoutbosch (n=42) (Figure 2.3).

Figure 2.2: Distribution of residential areas of enrolled patients. Fifteen residential areas in the Pretoria/Johannesburg region are indicated by , and KPTH, indicated by .

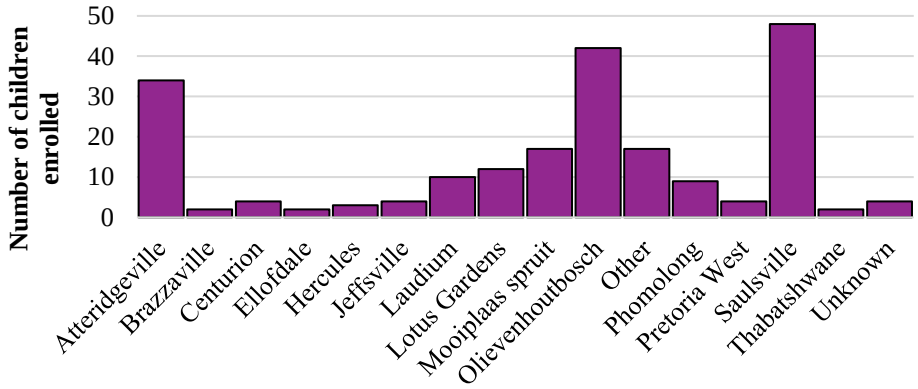


Figure 2.3: Different residential areas of the 205 children (< 5 years old), who were hospitalised with gastroenteritis at KPTH between July 2016 and December 2017.

2.3.1.4 Patient demographics and severity scoring

The patient demographics along with clinical characteristics of the entire cohort (205) and children infected with norovirus or rotavirus as well as children that tested negative for viruses are summarised in Table 2.8.

Table 2.8: Demographic and clinical characteristics of children with acute gastroenteritis (n = 205).

CHARACTERISTIC	COHORT (N=205)	NOROVIRUS (+) COHORT (N=32)	ROTAVIRUS (+) COHORT (N=46)	VIRUS (-) COHORT (N=109)
DEMOGRAPHIC CHARACTERISTICS				
AGE (IN MONTHS) AT ENTRY, MEDIAN (MIN-MAX)	10 (0.5-64)	10 (0.5-46)	8.5 (0.5-48)	14 (1-64)
0 - <12, N (%)	115 (56)	21 (66)	32 (70)	48 (44)
12 – < 24, N (%)	51 (25)	9 (28)	10 (22)	29 (27)
24 – < 48, N (%)	28 (14)	2 (6)	3 (7)	22 (20)
48 – 65, N (%)	12 (16)	0	1 (1)	10 (9)
GENDER, N (%)				
MALE	122 (60)	17 (53)	26 (57)	64 (59)
FEMALE	83 (40)	15 (47)	20 (43)	45 (41)
RACE, N (%)				
WHITE	1 (0.5)	1 (3)*	1 (2)*	0
BLACK	201 (98)	31 (97)	45 (98)	106 (97)
COLOURED	2 (1)	0	0	2 (2)
ASIAN	1 (0.5)	0	0	1 (1)
ENVIRONMENTAL FEATURES				
WATER SOURCE				
INDOOR TAP, N (%)	79 (39)	15 (47)	24 (52)	36 (33)
OTHER, N (%)	**126 (61)**	17 (53)	22 (48)	71 (65)
SANITATION TYPE				
FLUSH TOILET, N (%)	126 (61)	16 (50)	34 (74)	65 (60)
OTHER, N (%)	**79 (39)**	16 (50)	12 (26)	44 (40)
CLINICAL CHARACTERISTICS				
NUMBER OF DAYS WITH DIARRHEA, MEDIAN (IQR)	3 (3)	3 (2)	3 (3)	3 (3)
1-4 DAYS, N (%)	145 (71)	25 (78)	35 (75)	36 (33)
5 DAYS, N (%)	21 (10)	2 (6)	3 (7)	25 (23)
≥ 6 DAYS, N (%)	39 (19)	5 (16)	8 (17)	48 (44)
MAXIMUM # OF DIARRHEAL EPISODES IN 24H PERIOD, MEDIAN (IQR)	5 (3)	5 (3.75)	5 (3)	5 (2.5)
1-4 EPISODES, N (%)	70 (35)	14 (43)	14 (30)	36 (33)
5 EPISODES, N (%)	46 (22)	5 (16)	12 (26)	25 (23)
≥ 6 EPISODES, N (%)	89 (43)	13 (41)	20 (44)	48 (44)
VOMITING, N (%)	118 (58)	22 (69)	28 (61)	57 (28)
NUMBER OF DAYS WITH VOMITING, MEDIAN (IQR)				
1 DAY, N (%)	64 (54)	7 (32)	16 (57)	36 (63)
2 DAYS, N (%)	20 (17)	4 (18)	8 (29)	6 (11)
≥3 DAYS, N (%)	34 (29)	11 (50)	4 (14)	15 (26)
MAXIMUM # VOMITING EPISODES IN 24H PERIOD, MEDIAN (IQR)	1 (2)	2 (2)	1 (1)	1 (2)

1 EPISODE, N (%)	61 (52)	7 (32)	16 (57)	30 (53)
2-4 EPISODES, N (%)	45 (38)	12 (54)	9 (32)	21 (37)
≥5 EPISODES, N (%)	12 (10)	3 (14)	3 (11)	6 (10)
FEVER IN PREVIOUS 48H, N (%)‡	76 (37%)	5 (16)	3 (7)	17 (16)
≤37.0°C	171 (83)	26 (81)	43 (93)	88 (81)
37.1 – 38.4°C	16 (8)	3 (9.5)	2 (4)	8 (7)
38.5 - 38.9°C	14 (7)	3 (9.5)	1 (2)	9 (8)
≥39.0°C	4 (2)	0	0	4 (4)
DEHYDRATION SCORE				
NO DEHYDRATION (0), N (%)	20 (10)	3 (9)	5 (11)	10 (9)
MILD (1 – 5), N (%)	15 (7)	1 (3)	3 (7)	9 (8)
MODERATE – SEVERE (≥6), N (%)	170 (83)	28 (88)	38 (83)	90 (83)
DIARRHOEA TYPE				
WATERY, N (%)	175 (85)	26 (81)	43 (93)	89 (82)
DYSENTERY, N (%)	30 (15)	6 (19)	3 (7)	20 (18)
CHILD'S HIV STATUS				
UNINFECTED, UNEXPOSED, N (%)	134 (65)	22 (69)	33 (72)	70 (64)
UNINFECTED, EXPOSED, N (%)	61 (30)	8 (25)	12 (26)	33 (30)
INFECTED, N (%)	10 (5)	2 (6)	1 (2)	6 (6)
BASELINE VESIKARI SCORE, MEAN (SD)	10.39 (2.67)	11.063 (2.5)	10.09 (2.2)	10.33 (2.9)
MILD (<7), N (%)	16 (8)	1 (3)	2 (4)	12 (11)
MODERATE (7 – 10), N (%)	88 (43)	11 (34)	24 (52)	45 (41)
SEVERE (11 – 20), N (%)	101 (49)	20 (63)	20 (44)	52 (48)

‡Temperature above 38° (axillary measurement)

When the children were divided into different age groups, the majority were between newborn and 18 months old. More severe disease was also observed in children under the age of 18 months (Figure 2.4)

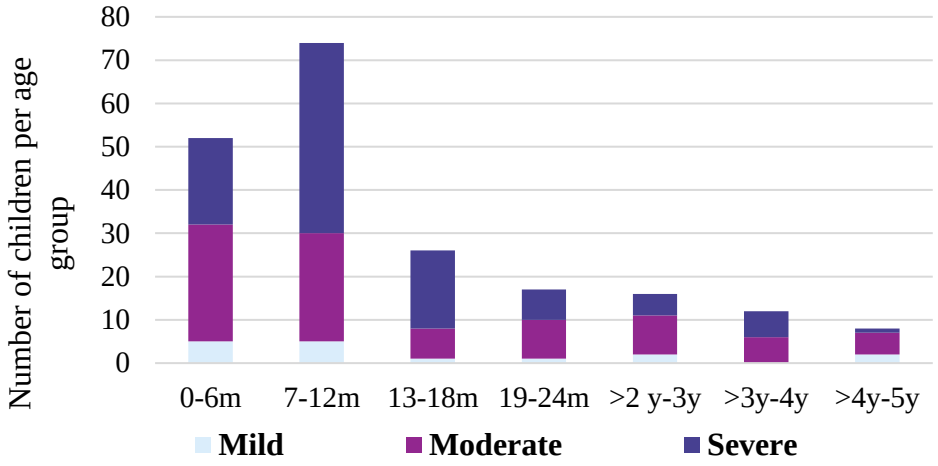


Figure 2.4: Severity of disease for different age groups of children ≤5 years who were hospitalised with acute gastroenteritis.

2.3.1.5 HIV demographics

Most of the children were HIV unexposed (n=134; 65%), with HIV exposed children accounting for 30% (n=61) of the study population and HIV positive children for 5% (n=10).

HIV exposure did not affect the frequency of gastroenteritis virus infections in this cohort (p>0.5). A general trend in increasing ratios of co-infections was observed from HU to HEU to HI (Figure 2.5). Due to the small HI sample size, this could not be determined as statistically significant. The severity of disease was also evaluated against HIV status (Figure 2.6). The difference in severity of disease for HI, HEU and HU patients was not statistically significant (p=0.072).

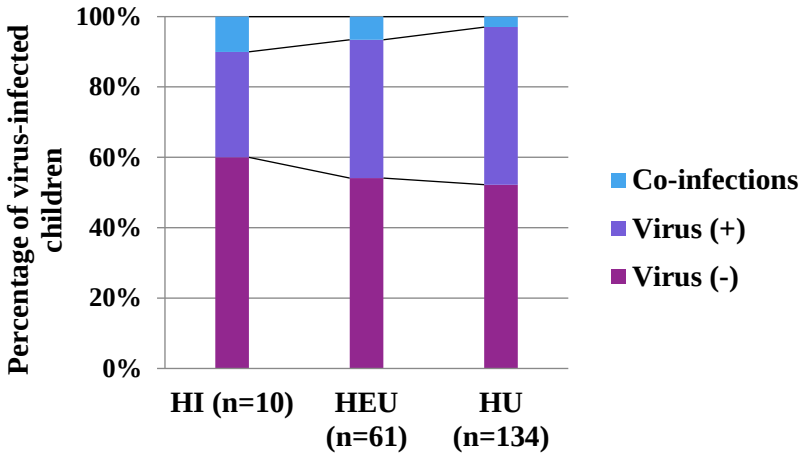


Figure 2.5: Percentage of HI, HEU and HU children infected with one virus or multiple gastroenteritis viruses. All data from the study population of 205 children (< 5 years), hospitalised with gastroenteritis at KPTH from July 2016 to December 2017.

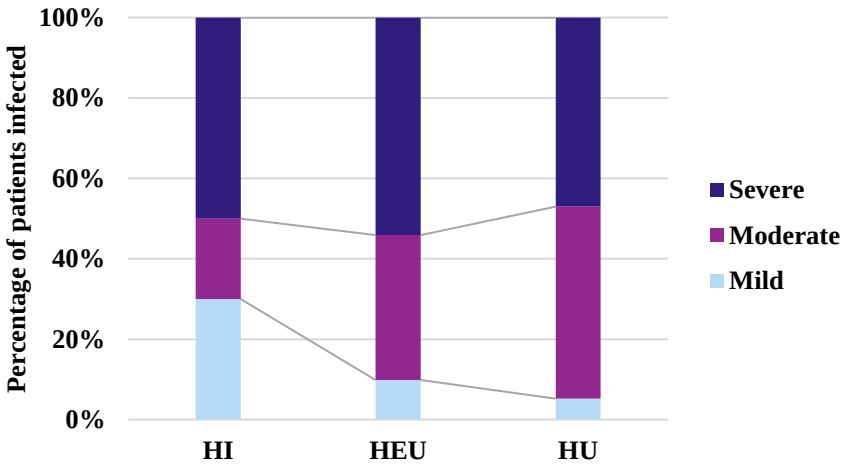


Figure 2.6: HIV status compared with severity of disease. Observed from 205 children ≤ 5 years, who were hospitalised with gastroenteritis. No statistically significant difference observed between different HIV status groups.

2.3.2 Enteric pathogen detection

Fifty-two percent (106/205) of the specimens tested positive for at least one gastroenteritis pathogen, with 47% (96/205) of specimens testing positive for at least one gastroenteritis virus. Nine percent (9/96) viral co-infections were identified and 5% (10/205) bacterial and parasite infections (Figure 2.7). Bacterial and parasite co-infection information was obtained from the NHLS, where microscopy, culture and sensitivity (MCS) testing was

performed on specimens from a total of 183/205 patients. Twenty-two patients' MCS data was unavailable.

A total of 5 parasite and 10 bacterial infections were observed, with 5 viral and microbial co-infections (Appendix C). Viral co-infections were also common, with 2/3 astrovirus infections being viral co-infections, while norovirus GI was only identified as single infections (Appendix C). Figure 2.9 provides a summary of all viral, bacterial and parasite infections.

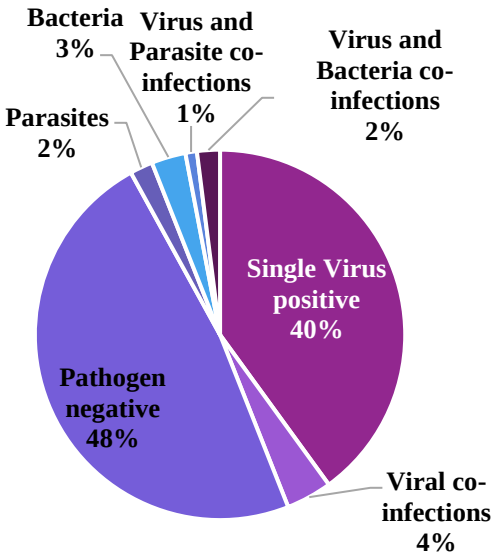


Figure 2.7: Pathogen distribution as observed through viral and MCS screening. All data from 205 children ≤5 years, who were hospitalised with gastroenteritis.

Rotavirus was the most prevalent virus detected (n=46; 22%), followed by norovirus GII (n=29; 14%) and then adenovirus (n=15; 7%) (Figure 2.8).

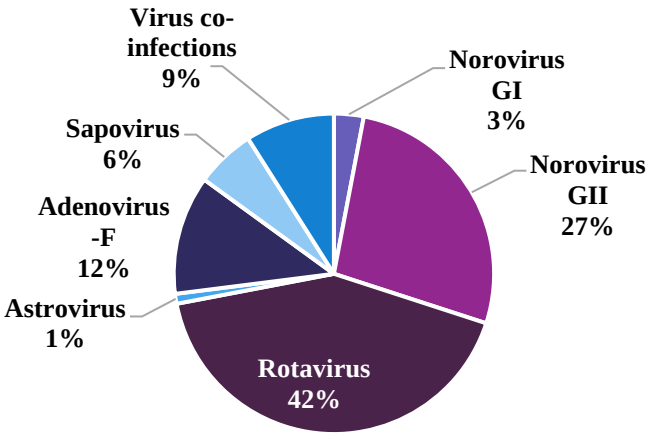


Figure 2.8: Distribution of gastroenteritis viruses (n=96) in children hospitalised with gastroenteritis as detected with the Allplex gastrointestinal multiplex panel.

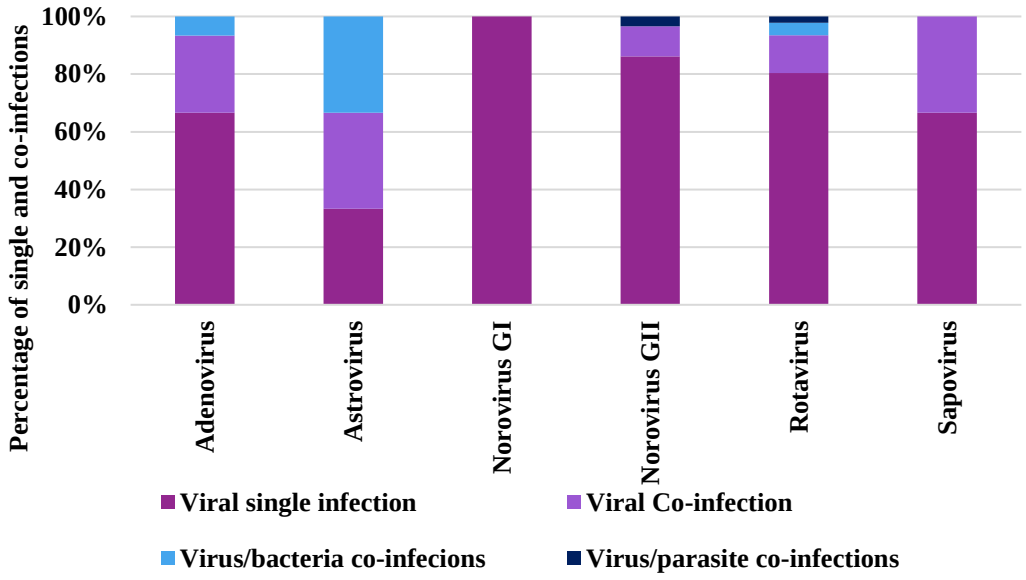


Figure 2.9: Distribution of gastroenteritis viruses in terms of single and co-infections. Observed in children (< 5 years), who were hospitalised with gastroenteritis at KPTH between July 2016 and December 2017.

The distribution of the different gastroenteritis viruses over the study period is shown in Figure 2.10. A clear peak could be observed for rotavirus from June to September 2017, and

for norovirus in summer months from November 2016 to February 2017. The peak observed in January 2017 is possibly due to bacterial infections (three known *Salmonella* positives in January) since only 4/33 tested positive for viruses.

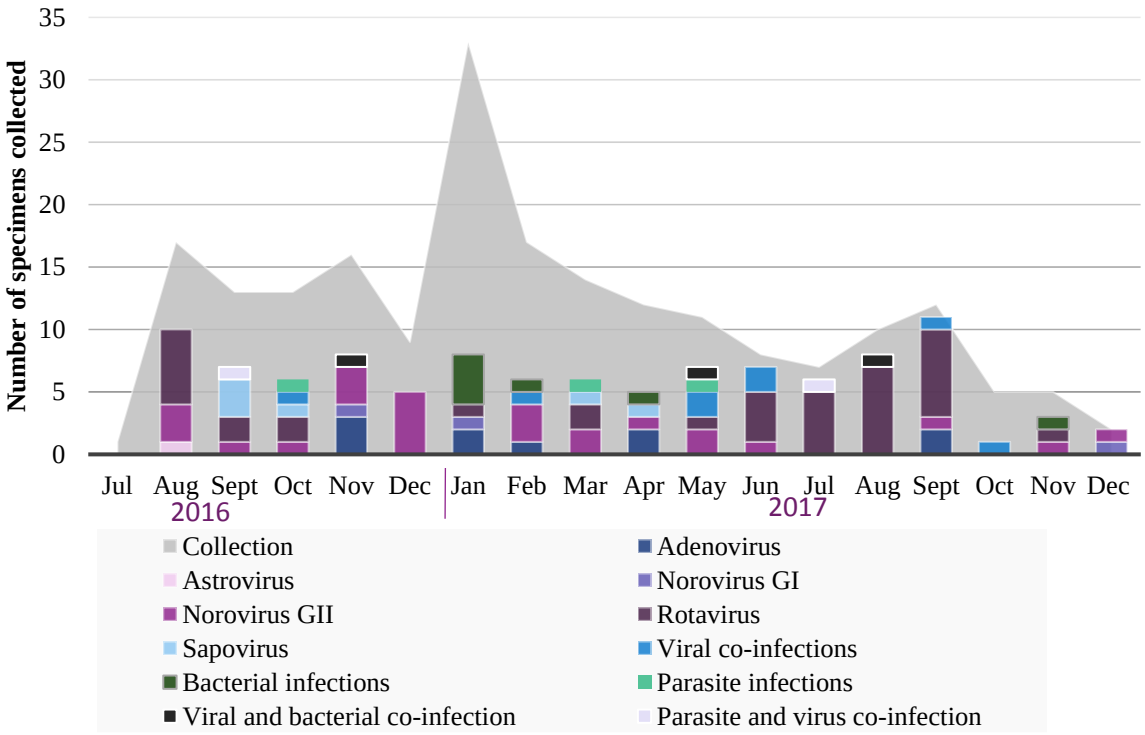


Figure 2.10: Gastroenteritis virus detection from July 2016 – December 2017. Detected from children (< 5 years), hospitalised with gastroenteritis at KPTH.

All viruses were detected in children under the age of 2 years, except 1 adenovirus, 2 norovirus, and 4 rotavirus infections that were detected in children aged between 24 months and 5 years (Figure 2.11).

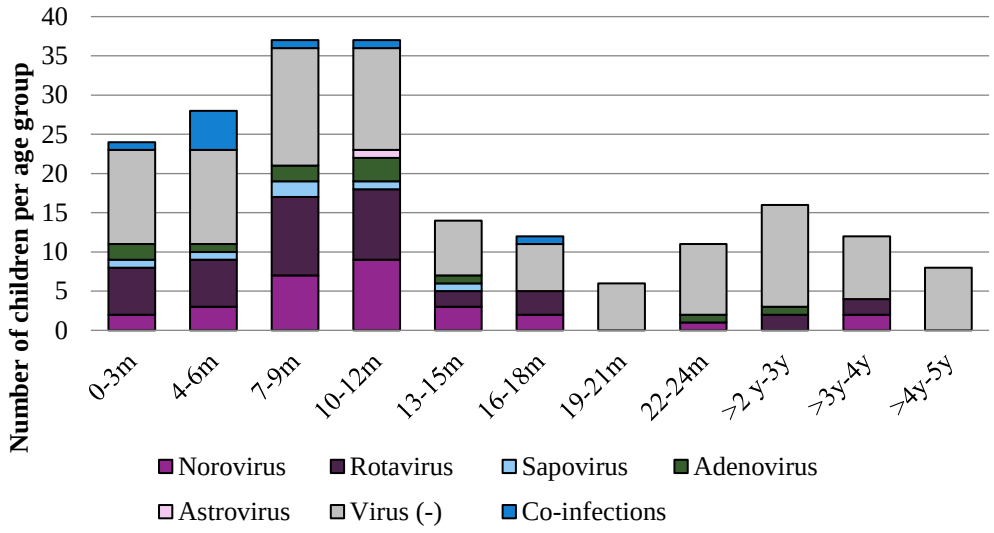


Figure 2.11: Prevalence of the five gastroenteritis viruses in different age groups of children (≤5 years) who were hospitalised with gastroenteritis between July 2016 and December 2017.

Follow up stool specimens were obtained for 46/205 patients. Thirty percent (14/46) of the specimens tested positive for norovirus GI (3) or GII (11), with all these patients being asymptomatic at the time the follow up specimen was collected (Figure 2.12).

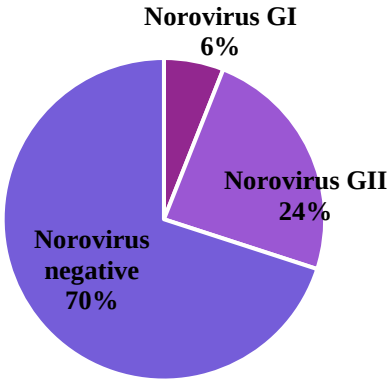


Figure 2.12: Distribution of asymptomatic norovirus infection obtained from 46 follow up specimens of children ≤5 who were initially hospitalised with viral gastroenteritis.

Disease severity scores and detected virus were compared for all the children, to determine if specific virus infections presented with more severe symptoms. Figure 2.13 indicates that norovirus infection most often coincided with severe illness. The child with a single astrovirus infection also presented with severe symptoms, although this should be interpreted with caution, due to the small sample size.

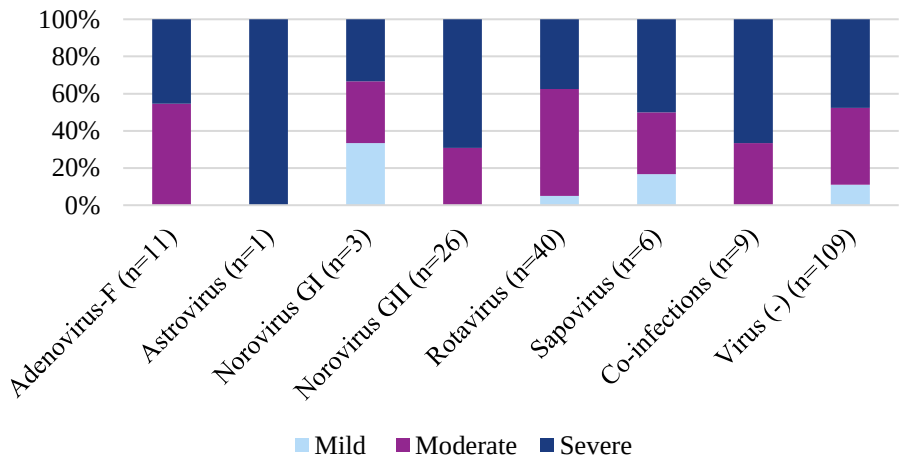


Figure 2.13: Disease severity observed during gastroenteritis virus infections (96) and virus unrelated gastroenteritis episodes (109) in children ≤ 5 years, hospitalised with gastroenteritis at KPTH from July 2016 to December 2017.

2.3.3 Virus Genotyping

2.3.3.1 Norovirus genotyping

Norovirus was detected in 15.6% (32/205) of specimens. Genogroups GI and GII represented 9.4% (3/32) and 90.6% (29/32) of norovirus cases, respectively, with no GI/GII co-infections detected. A dual infection with rotavirus, adenovirus or sapovirus was detected in three different specimens. Figure 2.14A is a representative gel electrophoresis image of the 1090 bp overlapping region (region A-C) that was amplified for both GI and GIIs, to determine partial RdRp and capsid gene sequences.

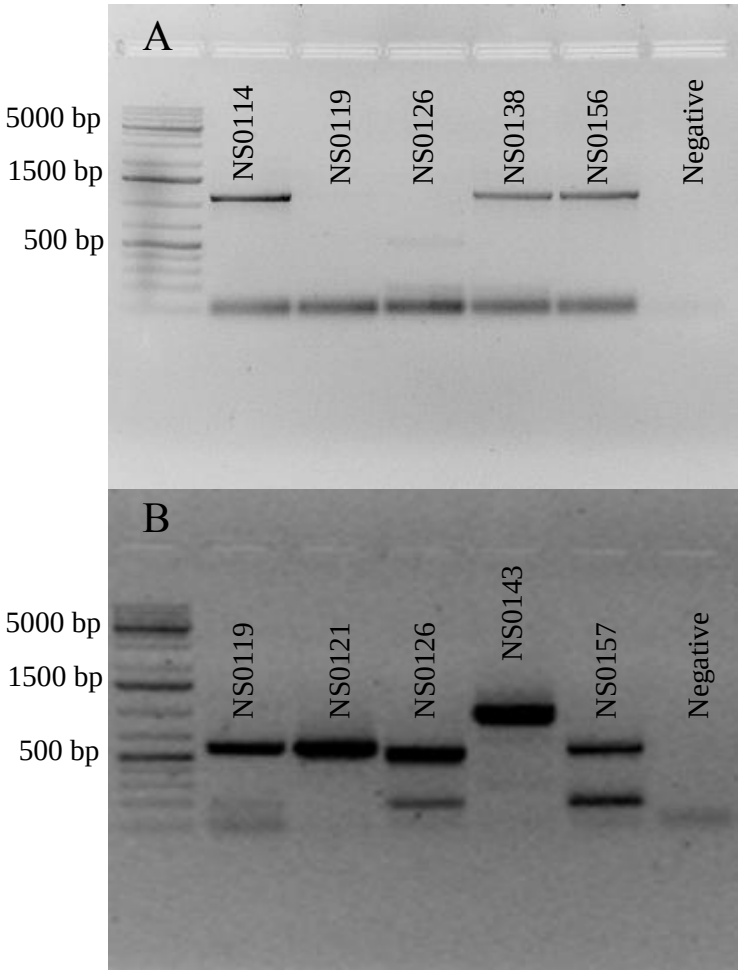


Figure 2.14: A) Gel electrophoresis analysis of region A-C one-step norovirus PCR products, indicating amplification of 1090 bp for NS0114, NS0138 and NS0156. B) Gel electrophoresis analysis of region B-C semi-nested norovirus PCR products, indicating amplification of ≈550 bp for NS0119, NS0126 and NS0157.

If no amplification was observed for region A-C (as with NS0119 and NS0126), a secondary semi-nested PCR was performed, which targeted a smaller region of approximately 550 bp (Region B-C) as shown in Figure 2.14B.

The amplicons from samples NS0010 and NS0038 were cloned to ensure sufficient PCR product for nucleotide sequencing. The cloning experiment had low efficacy, and only one clone of each of NS0010 and NS0038 was obtained (Figure 2.15).

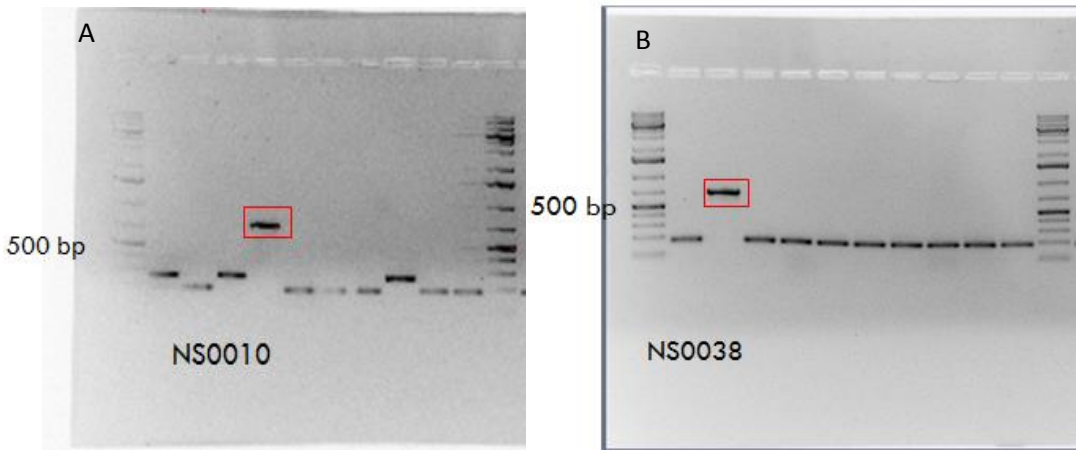


Figure 2.15: Gel electrophoresis analysis of colony PCR screening for positive clones, A) NS0010, B) NS0038. A red block indicates the correct sized PCR product.

In total, 27/32 norovirus strains could be genotyped, with four strains typed based on the A-C region and the remaining 23 strains on the B-C region. After several attempts at genotyping, no amplification was observed for five strains (2 GI and 3 GII). However, when re-tested on RT-PCR, the specimens remained positive for norovirus. These strains were considered untypable due to very low viral load [cycle threshold (Ct>38)]. The genotyped noroviruses represent seven genotypes, including emerging strains such as GII.2 and GII.21, with GII.4 being predominant (Figure 2.16). A summary of all norovirus strains detected and amplified in this study is indicated in Appendix D.

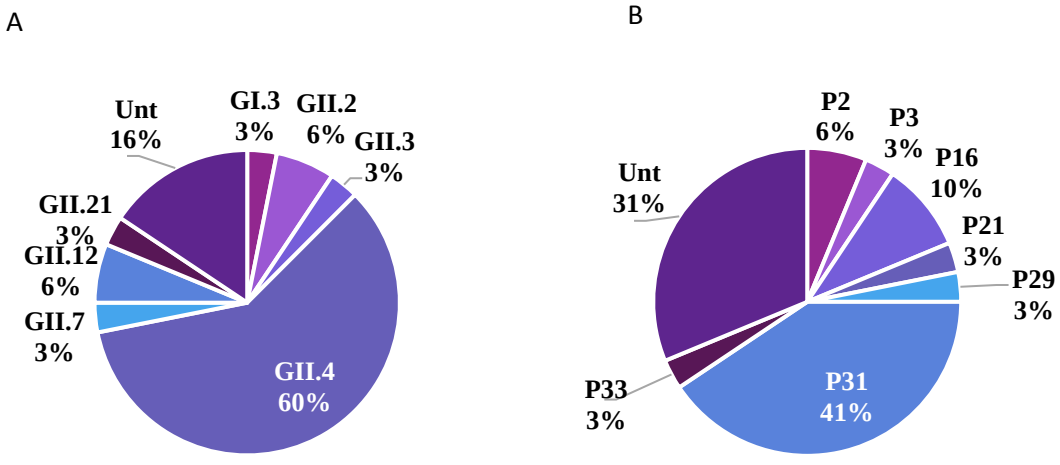


Figure 2.16: Norovirus capsid (A) and polymerase (B) genotype distribution in 32 children (≤ 5 years) who were hospitalised with gastroenteritis at KPTH from July 2016 to December 2017. Unt = untypable

Some sequences obtained were not of high enough quality and length to construct region AC or BC phylogenetic trees. Instead the norovirus sequences were divided into a capsid (n=28) and polymerase tree (n=21). Phylogenetic analysis of the partial capsid- and polymerase regions of the study strains determined with references from GenBank is shown in Figure 2.17 and 2.18. A total of seven capsid types and seven polymerase types were determined with norovirus sequencing analysis, with the GII.4[P31] observed most often.

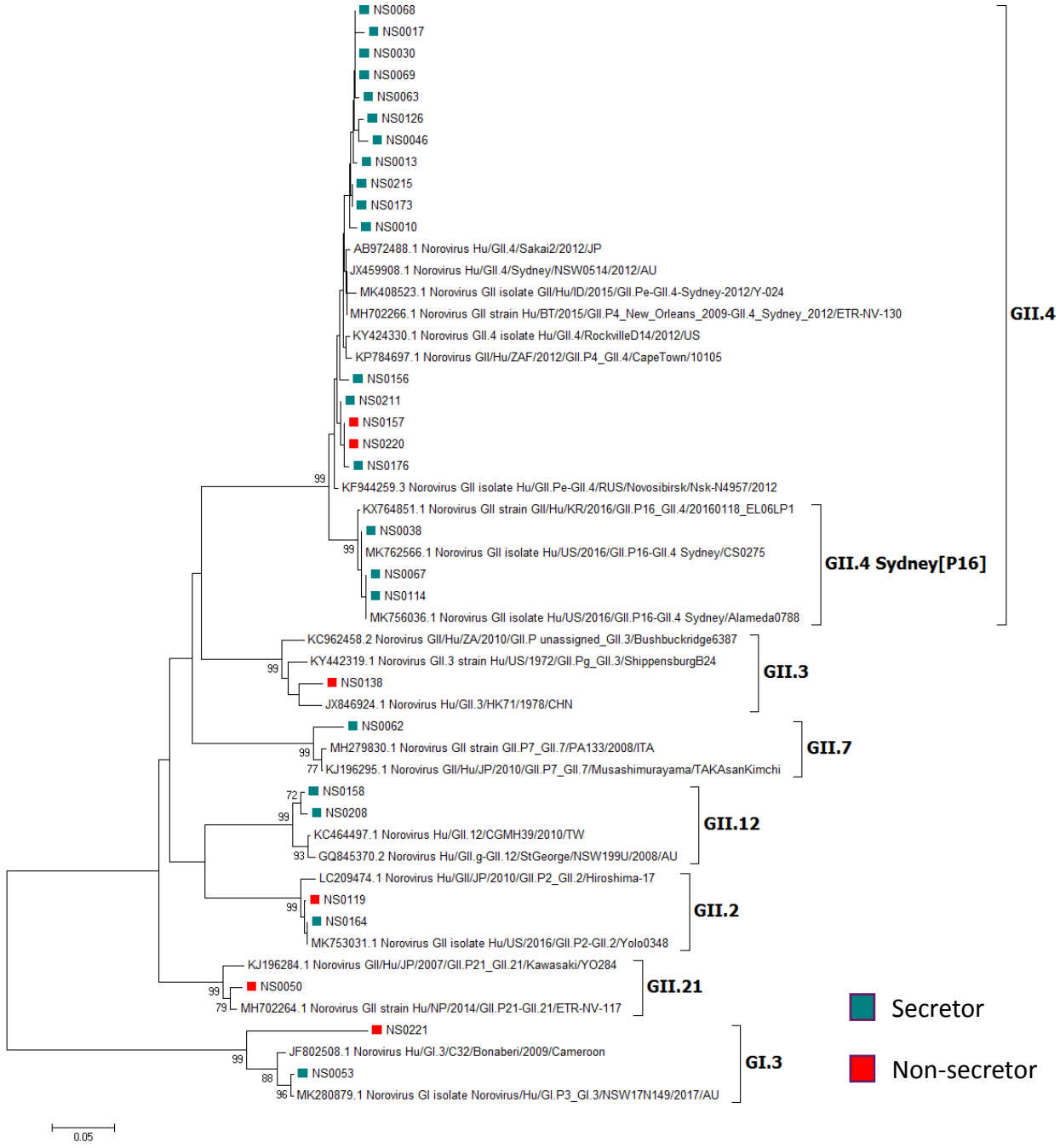


Figure 2.17: Neighbour joining phylogenetic analysis of the partial capsid (≈270 bp) of 29 norovirus strains detected in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region. Secretor status is indicated by coloured blocks.

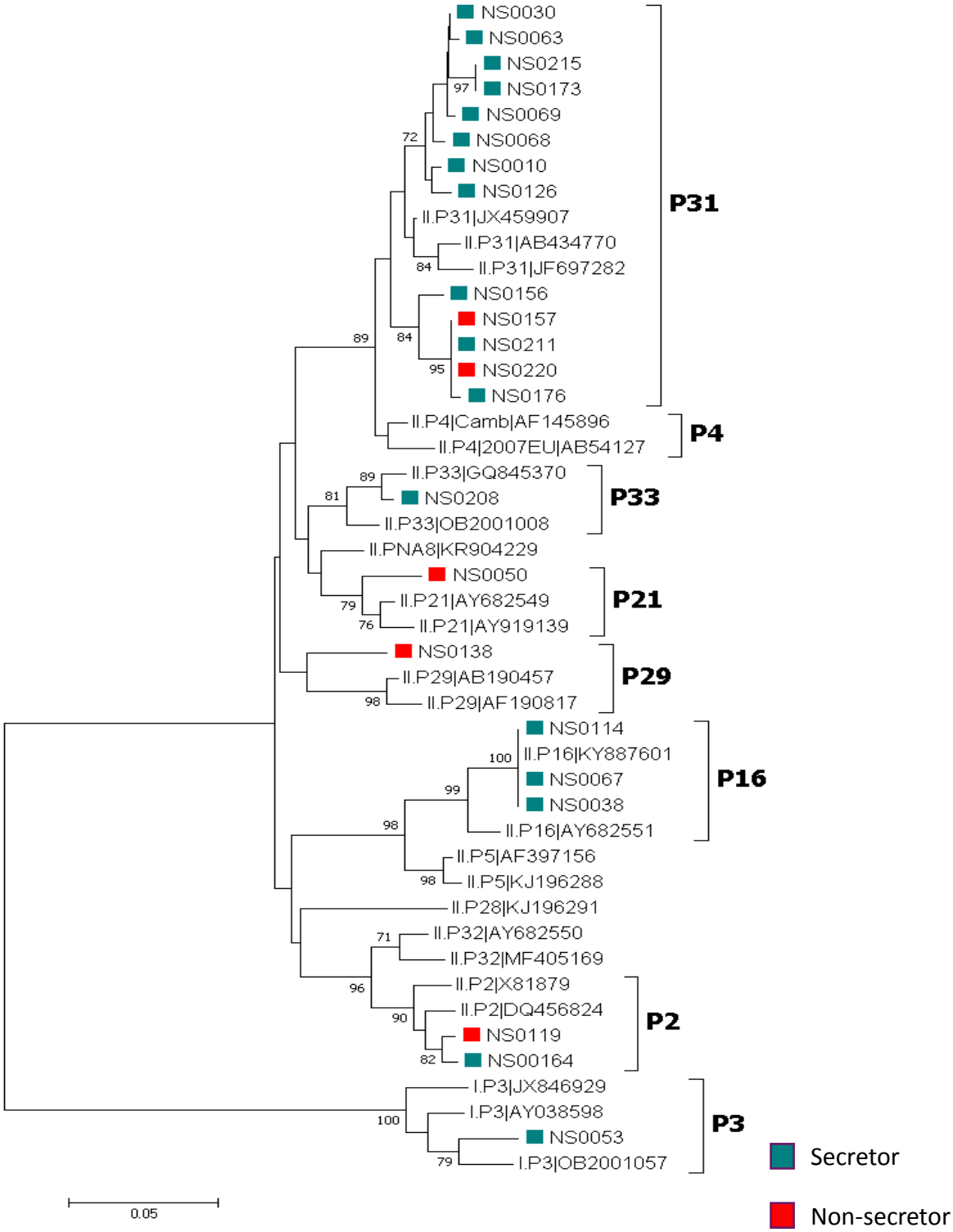


Figure 2.18: Neighbour joining phylogenetic analysis of the partial polymerase (≈260 bp) of 21 norovirus strains detected in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region. Secretor status is indicated by coloured blocks.

Of the 14 asymptomatically infected patients detected during follow up, five were infected with norovirus at enrolment. A total of 12/14 of the specimens could be genotyped, representing five genotypes, with GII.13 being the most prevalent, followed by GII.4 (Figure 2.19).

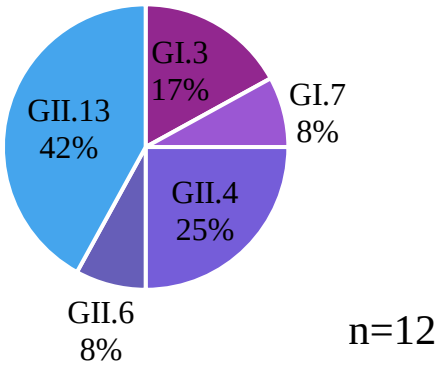


Figure 2.19: Norovirus genotype distribution in children (≤ 5 years) 6 weeks after their initial hospitalisation (n=12) with gastroenteritis at KPTH from July 2016 to December 2017. All the children were asymptomatic at the time of follow up stool specimen collection.

Of the five original norovirus positive specimens, only three could be genotyped. When comparing the initial and follow up strains, only NS0173 was found to be identical to the initial infecting strain (GII.4), indicating possible prolonged asymptomatic replication (Figure 2.20).

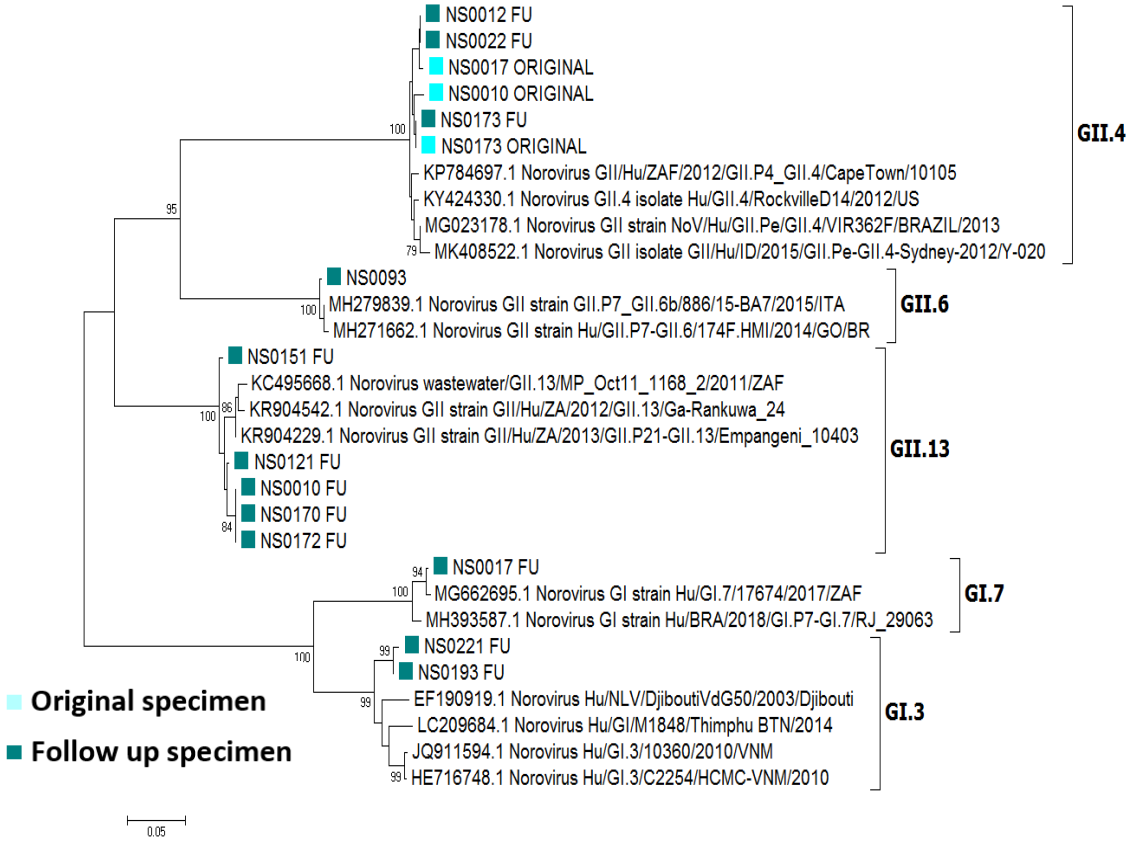


Figure 2.20: Neighbour joining phylogenetic analysis of the partial capsid (≈270 bp) of 12 norovirus strains detected in follow up specimens in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region.

2.3.3.2 Sapovirus

The majority of sapovirus strains (8/9) could be amplified and genotyped, including a GII.8, which represents the first report of this genotype in South Africa. One strain was untypable due to a high Ct value (35.5). Five genotypes were identified among the eight specimens with GI.2 being the most predominant (Figure 2.21). Figure 2.22 shows the phylogenetic tree created through neighbour-joining phylogenetic analysis.

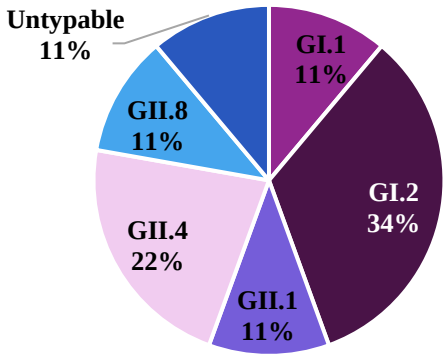


Figure 2.21: Sapovirus genotype distribution determined in children ≤ 5 years who were hospitalised with gastroenteritis at KPTH from July 2016 to December 2017.

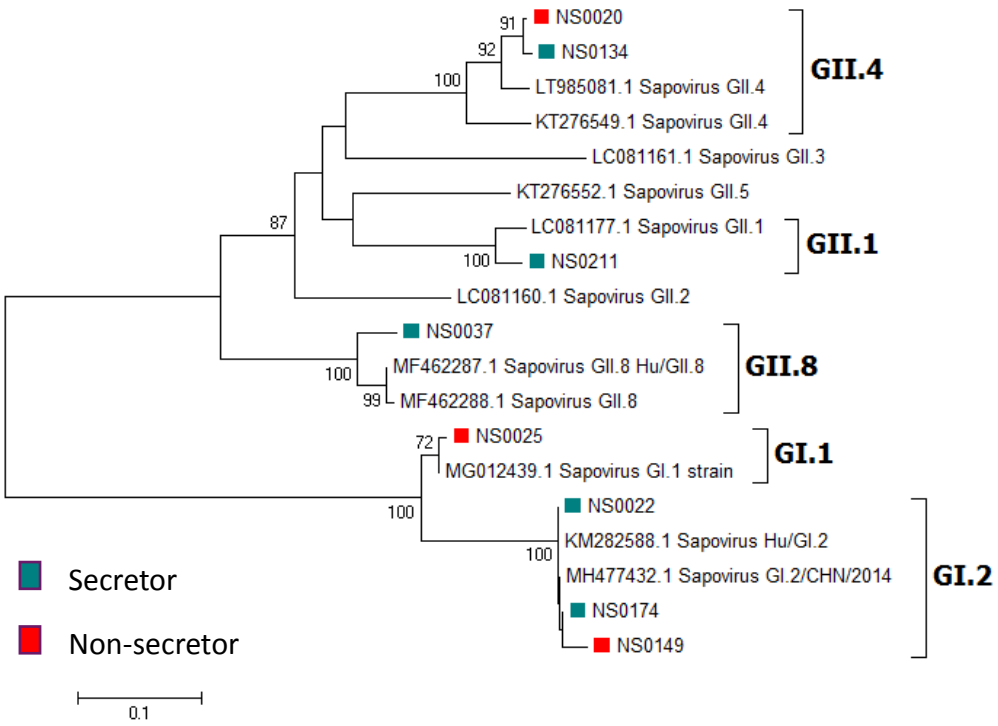


Figure 2.22: Neighbour joining phylogenetic analysis of the partial capsid (≈260 bp) of eight sapovirus strains detected in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.01 nucleotide differences per site, over the indicated region.

2.3.3.3 Rotavirus genotyping

Rotavirus infected children represented 22% of the study population, with a total of 37/46 children being recorded as fully vaccinated against rotavirus. A total of 44/46 rotavirus strains were genotyped, representing 3 P types and 5 G types, with a total of 9 different combinations, of which G3P[4] predominated (Figure 2.23).

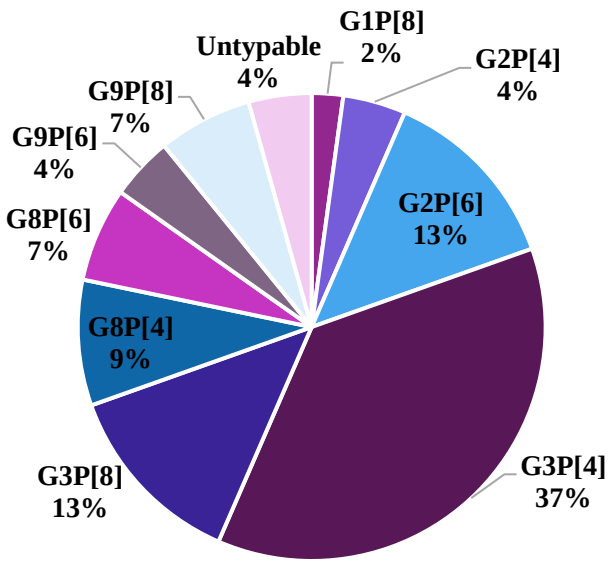


Figure 2.23: Rotavirus genotype distribution in 46/205 children (< 5 years), hospitalised with gastroenteritis between July 2016 and December 2017.

All virus strain sequences that were adequate for analysis (P-type 42/46, and G-type 43/46) were then aligned and a phylogenetic tree was constructed separately for the VP7 and VP4 sequences (Figures 2.24 and 2.25).

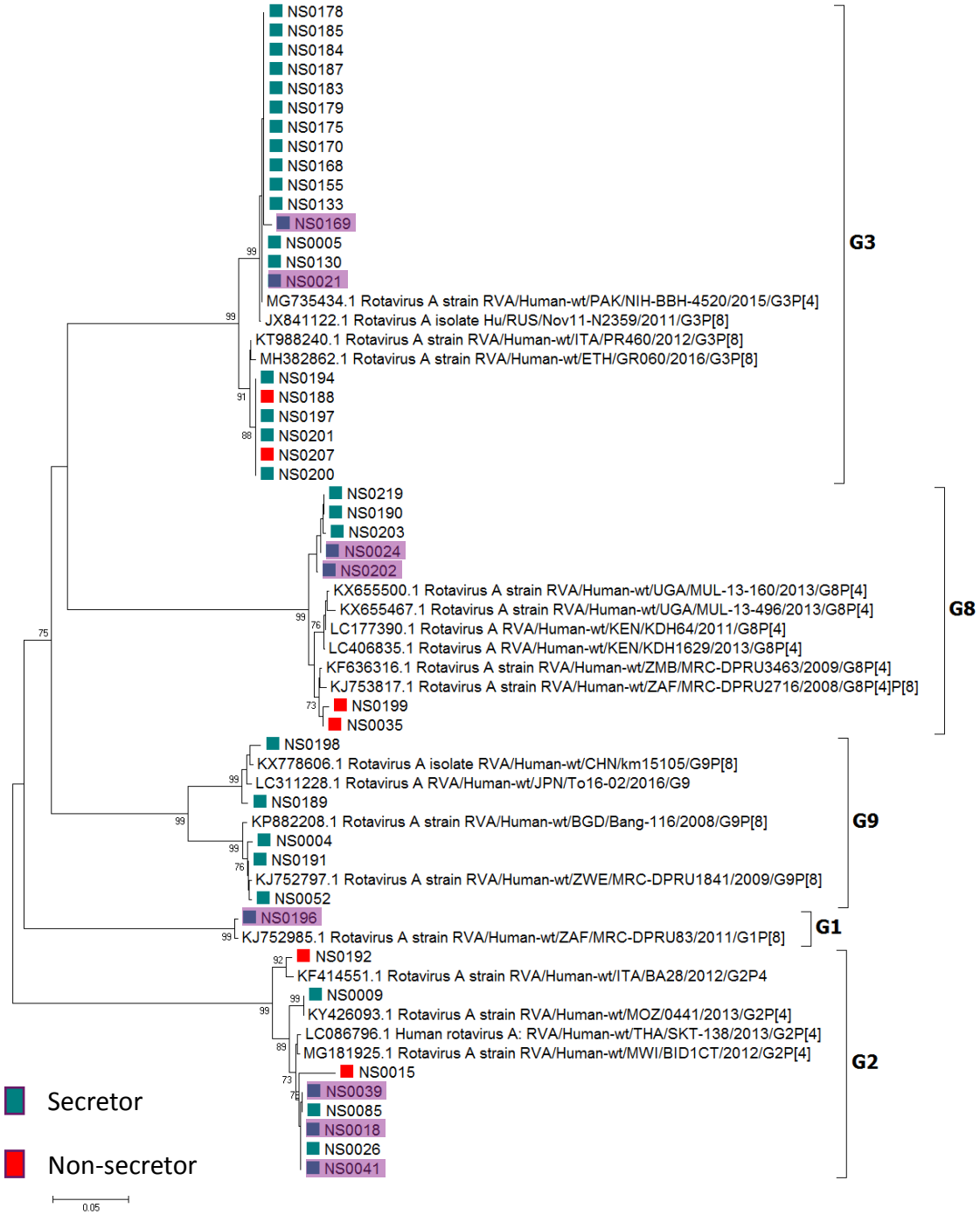


Figure 2.24: Neighbour joining phylogenetic analysis of the VP7 sequence (≈580 bp) of 42 rotavirus strains detected in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region. All unvaccinated participants indicated in purple blocks.

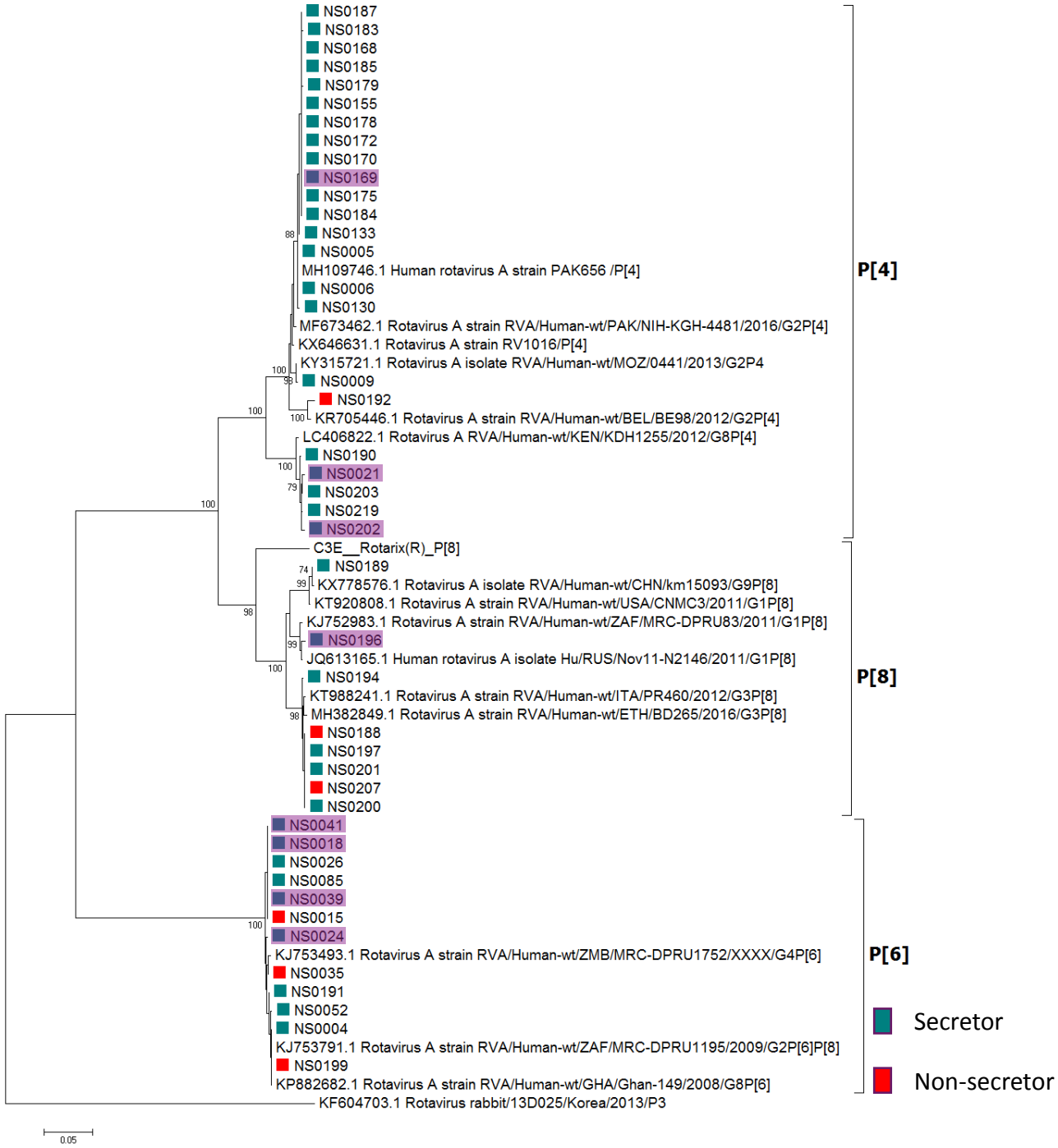


Figure 2.25: Neighbour joining phylogenetic analysis of the VP4 sequence (≈560 bp) of 43 rotavirus strains detected in this study. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region. All unvaccinated participants indicated in purple blocks.

2.4 Discussion

This study was based in a relatively urban setting, in Pretoria, Gauteng, with all enrolments from KPTH. Over an 18-month period (July 2016 – December 2017) a total of 221 children, under the age of five years, hospitalised with gastroenteritis were enrolled, with a maximum enrolment observed in January 2017. This spike could be due to various reasons, but as the study enrolment only required children to be under the age of five years, and hospitalised with gastroenteritis, this peak could possibly be due to a bacterial outbreak, which are more common in summer months (Wiegering *et al.*, 2011) or due to a gastroenteritis causing virus not present in the panel of viruses that were tested for. This theory is re-enforced by the fact that more gastroenteritis bacteria than viruses were observed including three salmonella infections. Most of the enrolments were from Saulsville and Olievenhoutbosch, but this is more likely due to proximity to KPTH, rather than an indication of disease risk in these areas (Figure 2.3). A range of mild to severe disease was calculated through the Vesikari scoring system, with moderate to severe disease being observed more. This was expected as all the patients were hospitalised. Of the 205 children admitted in the study, the majority (119/205) were under the age of 18 months, illustrating that gastroenteritis is more severe in younger children (Liu *et al.*, 2016a, Anwari *et al.*, 2018).

The use of multiplex qPCR assays for the detection of gastroenteritis causing viruses have increased in recent years. These assays have made faster and simultaneous detection of multiple enteric viruses possible, while also increasing sensitivity and specificity. Several studies have indicated that the Allplex gastrointestinal virus panel, along with the Bio-Rad CFX platform has high specificity and sensitivity (Han *et al.*, 2017, Hyun *et al.*, 2018, Hirvonen, 2019). Initially, a few positive controls (especially rotavirus positive controls) did not amplify and consequently tested negative. After an optimisation step was added (see section 4.5.2), these controls were retested, and nucleic acid amplification resulted in consistent curves with lower Ct values. The heating step was added to ensure the separation of all rotavirus double-stranded RNA for adequate amplification.

Pathogens were identified in 106 (52%) patients. Rotavirus (22%), norovirus (15.6%), adenovirus-F (7%) and sapovirus (4%) were the most common pathogens. Astrovirus was mostly observed along with rotavirus infection (2/3), but as there were only three astrovirus

positive patients, no conclusive observations could be made regarding this specific viral infection.

A total of 46 children (22%) were infected with rotavirus (40 single infections (20%), 6 co-infections). This is a very high prevalence when compared to high income countries, with prevalence as low as 1.5% in the United States (Control and Prevention, 2020). In lower income countries such as South Africa, this high prevalence is more common (Lekana-Douki *et al.*, 2015, Ouedraogo *et al.*, 2016), with a recent study indicating rotavirus prevalence at 15% in South Africa (Makgatho *et al*., 2019).

Norovirus was the second most prevalent gastroenteritis virus in this study (32/205; 15.6%), yet this virus is still not routinely tested for in South African hospitals. This is a point of concern, when considering the fact that norovirus is now the leading cause of viral gastroenteritis in developed countries, which may in future become a reality in South Africa (Hemming *et al.*, 2013, Payne *et al.*, 2013). By testing for norovirus and typing positive strains, information regarding the circulating viruses in the population will aid in the development of norovirus vaccines. Norovirus infection was also observed more with severe gastroenteritis (21/32; 66%) based on the Vesikari severity scoring method, which is in line with other studies (Platts-Mills *et al*., 2018, Mans, 2019). In terms of asymptomatic norovirus infection, very limited research has been done in South Africa, which is why it was an objective to obtain follow-up stools to determine any asymptomatic infections. Follow up specimens were received for 22% (46/205) of the study population, all of which were collected from children showing no symptoms of gastroenteritis. When tested, 30% (14/46) of these specimens tested positive for either norovirus GI (3/14) or GII (11/14). These results are comparable to a previous study in South Africa, which determined asymptomatic norovirus infection to be approximately 36% (Kabue *et al*., 2016), as well as other studies in low income settings, such as the Garcia *et al*. study that found almost 30% norovirus positive specimens from asymptomatic patients in Mexico (García *et al*., 2006). The higher frequency of asymptomatic infections in the Kabue *et al*. study is likely due to the more rural setting, which has been known to be associated with a higher prevalence of norovirus transmission (UNICEF, 2012, Mattioli *et al.*, 2013).

Nine viral co-infections were observed (4%), with a final total of 99 (48%) gastroenteritis-pathogen negative children. These ratios are somewhat lower than what has been observed

in other studies, especially in middle- to lower income countries, with co-infections ranging from 1.1% (Zhang *et al.*, 2016a) to 35% or higher (Lekana-Douki *et al*., 2015, Nakamura *et al.*, 2016, Ouedraogo *et al*., 2016, Gupta *et al.*, 2018, Mans, 2019).

Of the 46 rotavirus-positive specimens, rotavirus strains could be genotyped in 44 (96%). Two specimens were untypable, most likely either due to low viral load (Ct>32) or due to mutation at the primer binding site, which inhibited the binding of specific primers, thus preventing amplification. Various G and P type combinations were observed with the most prevalent combination being G3P[4] (17/46; 37%), which is in contradiction to other studies in this setting that have previously observed G3P[8], and G9P[8] to be the most prevalent strains (Page *et al*., 2016b). The G3P[8] strain was only observed in 6/46 specimens. This discrepancy between different studies could be due to change in strain prevalence or an outbreak, as many of the G3P[4] genotyped were very similar or identical. The question of contamination was raised, but following re-extraction, screening and genotyping of the virus strains in question, it was determined that they were all true positive results. Of the 46 children, 37 had been fully vaccinated against rotavirus (the majority with Rotarix (G1P[8]). This would raise concern regarding the effectiveness of the vaccine to control severe rotavirus infection in the South African setting, but when comparing severity of these children infected with rotavirus, through the Vesikari severity scoring, it was found that the predominant subset of this group (25/40; 62.5%) were in the mild to moderate category, indicating that the vaccine could indeed have helped to lessen the severity of the disease, but did not prevent hospitalisation. Most strains were observed in both groups of vaccinated and unvaccinated participants. The only exception is one G1P[8] strain that was only observed in an unvaccinated child (NS0196). This is expected as vaccinated children will have protection against G1P[8] (rotarix strain), but unfortunately the sample size is too small to give any concrete conclusions regarding this. More recent studies have now included whole-genome sequencing of rotaviruses (Matthijnssens *et al.*, 2008, Magagula *et al.*, 2015), to give a better perspective on the specific strains detected, their origin and reassortment, this should be considered for future studies, to improve our understanding of rotavirus epidemiology.

Virus genotyping was performed on all 3 norovirus GI and 29 norovirus GII positive stool specimens, with genotypes based on nucleotide sequences and phylogenetic analysis. For the norovirus GI's, only one virus strain could be typed, as a GI.3, and the other two were

annotated as untypable. This was most likely due to low viral loads, as the virus strains had cycle thresholds of 33 and 39. The majority of the GII strains were determined to be GII.4s (19/29), which was to be expected, as this is the most prevalent and virulent norovirus genotype (Dai *et al*., 2015, Kambhampati *et al*., 2015, Chhabra *et al*., 2018). Two GII.4 variants have circulated throughout the study, namely the Sydney 2012 variant, with three different polymerases, GII.P31 (previously GII.Pe), GII.P4 and GII.P16 and an unassigned GII.4, which is a distant variant of the Sydney 2012 strain (Chhabra *et al*., 2019). Other genotypes identified during this study include emerging types, such as GII.2 (n=2), which has been detected frequently in wastewater, but not yet as commonly in clinical specimens in South Africa (Mabasa *et al*., 2018). In terms of the follow up specimens, 5 different genotypes were observed, with GII.13 being the most prevalent. When comparing the norovirus genotypes detected in the children at initial hospitalisation (GI.3, GII.2, GII.3, GII.4, GII.7, GII.12, GII.21) with the genotypes in asymptomatic infections (GI.3, GI.7, GII.6, GII.13) a different distribution was observed, with only GI.3 and GII.4 detected in both groups. Two of the children initially infected with GII.4 had asymptomatic infections with GI.7 and GII.13, respectively. It is possible that the children with asymptomatic GII.4 infection at follow up may have had earlier symptomatic GII.4 infections. The median age of children asymptomatically infected with norovirus was 11.5 months compared to 10 months for children with symptomatic infection.

A total of 8 out of the 9 sapovirus positive strains could be genotyped, resulting in 5 genotypes being identified, with GI.2 being most prevalent (3/9). A larger number of sapovirus positive specimens would be needed to give a clearer indication of circulation and prevalence. These findings were very different when compared to other studies in South Africa, which reported GIV strains as predominant (Murray *et al*., 2016), since this genotype was not observed during this study. A GII.8 was described for the first time in South Africa in this study and it is most closely related to a GII from South Africa identified in 2014. (GenBank accession number: KP196511.1). Sapovirus infection was associated with a higher amount of moderate than severe gastroenteritis (4/9 moderate infections; 45%), especially when looking at single infections (3/6). This correlates with previous study findings, suggesting that sapovirus infection is usually less severe than noro- or rotavirus

infection (Sakai *et al*., 2001). Despite this, it should still be considered that the virus caused severe enough gastroenteritis as a single infection, for six children to be hospitalised.

One of the objectives of this study was to compare the prevalence and severity of norovirus and rotavirus in HIV-exposed uninfected and HIV unexposed children.

After statistical analysis was performed the data indicated no statistically significant difference between different HIV exposure groups in terms of gastroenteritis virus prevalence (and more specifically noro- and rotavirus prevalence) or severity. This observation was unexpected, as previous studies have indicated that HIV exposure can cause an effect (Feiterna-Sperling *et al*., 2007, Marinda *et al*., 2007, Epalza *et al*., 2010, Slogrove *et al*., 2012, Evans *et al*., 2016). The discrepancy could be due to the small study population size, although it should be considered that HIV exposure might in fact not affect gastroenteritis prevalence and severity. In future, a larger population should be considered to confirm this observation.

Chapter 3 : Fucosyltransferase 2 Genotyping

3.1 Introduction

Preventative interventions to decrease viral gastroenteritis have shown variability in effectiveness. This along with variation in pathogen-specific incidence between different geographical populations suggest that inherited host factors may differentially influence the susceptibility to different enteric infections.

Host factors that may be important are the fucosyltransferase genes, *FUT2* and *FUT3*. Multiple studies have shown that one's secretor status (which is determined by the *FUT2* genes) can influence risk of infections, specifically with some noro- and rotaviruses (Thorven *et al.*, 2005, Ferrer-Admetlla *et al.*, 2009, Lopman *et al.*, 2014a, Payne *et al.*, 2015, Zhang *et al*., 2016b). Population based genetic analysis further motivates this theory, as wide spread selection and non-neutral evolution of the HBGA genes (*FUT2* and *FUT3*) have been recognised to be similar to genes involved in antigen recognition (Fumagalli *et al.*, 2009).

When performing *FUT2* genotyping, several *FUT2* single nucleotide variations (SNV) have been identified. The nonsense variation G428A is representative of the dominating non-secretor allele (se428) in Europeans and Africans (Ferrer-Admetlla *et al*., 2009), but even though the frequencies of non-secretors in most populations are similar, the occurrence of Se enzyme deficiency, and the single-nucleotide variation associated is race specific (Koda *et al.*, 2001, Kindberg and Svensson, 2009, Reid *et al.*, 2012). The G428A SNV appears in approximately 20% of the European population and 30% of the African population (Nordgren *et al.*, 2013, Parker *et al.*, 2018). Histo-blood group antigens are found to affect host susceptibility to rotavirus and norovirus infection, which has been detected at a higher rate in secretors when compared to non-secretors. Multiple studies have shown that rotavirus and norovirus use HBGAs as binding factors in the intestinal epithelium (Tan and Jiang, 2011, Griffin, 2013, Zhang *et al*., 2016b). Feeding habits (specifically breastfeeding) have also been proven to affect the child's microbiome, which in turn may affect susceptibility to noro- and rotavirus infection (Zivkovic *et al*., 2011, Lewis *et al*., 2015, Parker *et al*., 2018). Taking this into account, there are multiple hypotheses of different

pathways by which both the child's *FUT2* and *FUT3* genes as well as those of breastfeeding mother may alter the susceptibility to these enteric infections (Figure 3.1).

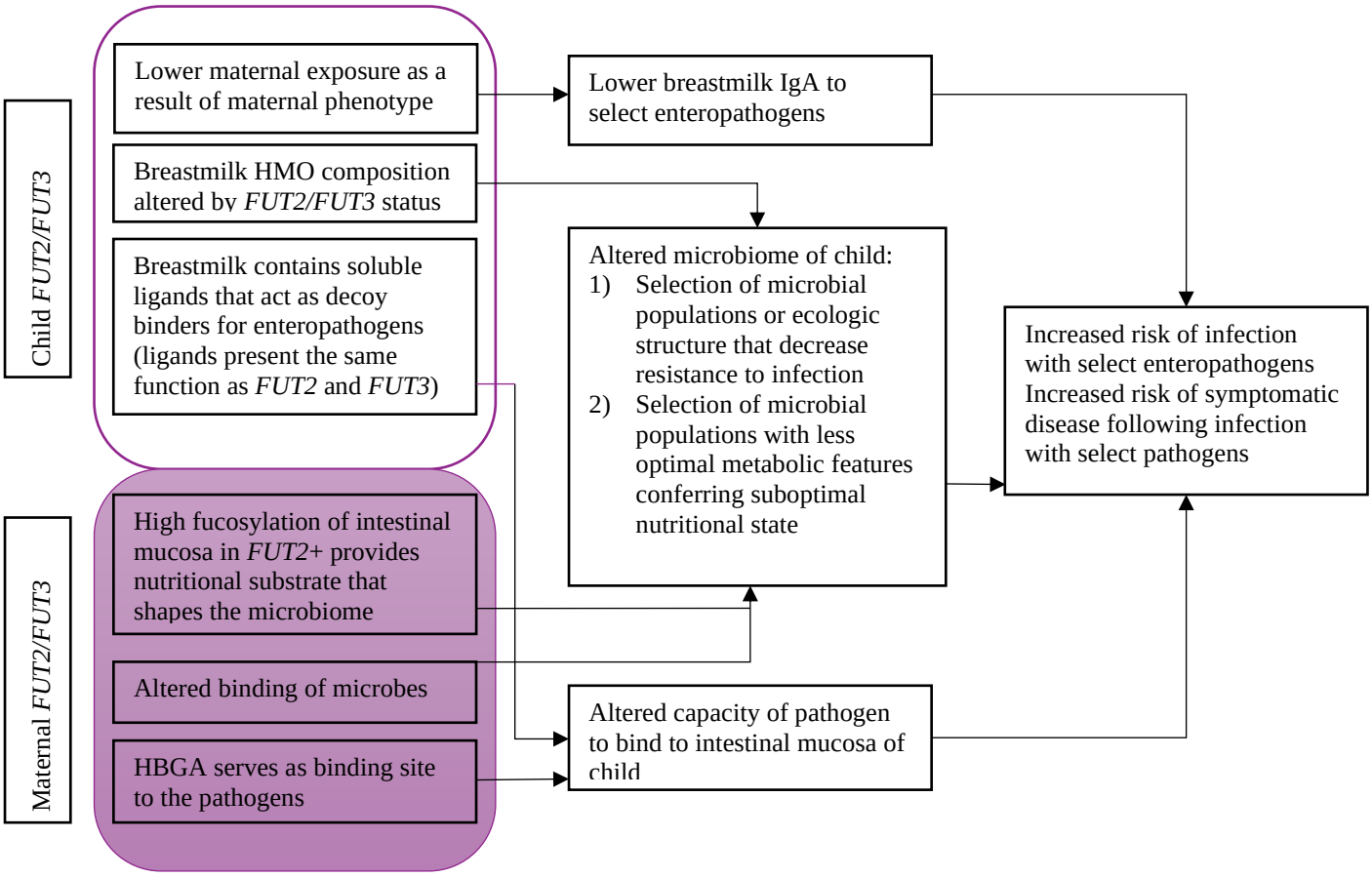


Figure 3.1: Hypothesised pathways through which maternal and child fucosyltransferase (*FUT*) *2* and *FUT3* expression alter susceptibility to enteric infection. HMO: Human milk oligosaccharides. Adapted from Colsten *et al*, 2019.

3.2 Methods and Materials

3.2.1 Total genomic DNA extraction from whole blood samples

Manual genomic DNA extraction was performed on 190/190 (200 µL) of the whole blood specimens using a QIAamp blood kit (Qiagen Inc., Hilden, Germany), to a final elution volume of 200 µl, as per manufacturer's instructions. After extraction DNA was stored at -20˚C until use.

In the case of no blood specimen being available (15/205), total nucleic acids were extracted from stool samples using the NucliSENS® EasyMAG® automated nucleic acid extraction instrument (BioMérieux, Marcy-l'Étoile, France) and stored at -80˚C until use.

3.2.2 Nonsense variation detection using real-time PCR and SNP assay

The TaqMan SNP assay kit (SNP ID: rs601338, Applied Biosystems, Foster City, CA) was used to detect any G428A nonsense variation. Steps included PCR amplification using the Q5® Hot Start High-Fidelity DNA Polymerase, Allelic Discrimination Plate reading and Allelic Discrimination Analysis, as per manufacturer's protocol. The master mix was initially prepared to a final volume of 10 µl, as stated in the protocol, but after encountering discrepant results, the master mix was doubled per sample (final volume = 20 µl).

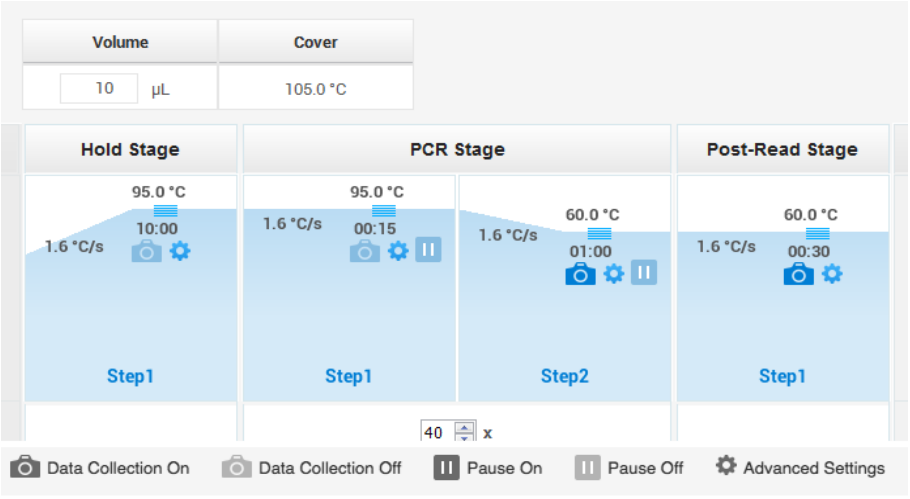


Figure 3.2: *FUT2* genotyping cycling conditions as depicted on the QuantStudio™ Design and Analysis software.

PCR amplification was detected through QuantStudio™ Design and Analysis software (Figure 3.2). Detection of amplicons was done cycle by cycle, by measuring fluorescence during each cycle using fluorescently labelled sequence-specific probes, FAM and VIC.

The TaqMan assay makes use of two minor groove-binding (MGB) probes, each probe is labelled with a different reporter dye. During the real-time PCR, the following labelled probes were used to detect the different alleles with or without the G428A SNV present:

- FAM™- labelled probe detected the presence of the G allele (without the G428A SNV)
- VIC®- labelled probe detected the presence of the A allele (with the G428A SNV)

3.2.3 *FUT2* genotyping using EIA and Lectin

An additional test was performed to confirm the *FUT2* genotype of the children, through the detection of Fucα1-2Gal using a UEA-I lectin EIA and saliva specimens. Lyophilised *Ulex europaeus* agglutinin (UEA-I, Merck – previously Sigma Aldrich) was prepared by adding sterile phosphate buffered saline (PBS) (pH 7.4) to a concentration of 1 mg/mL.

Saliva samples were collected with the SalivaBio Children's Swab (SCS) (Salimetrics, State College, PA).

The parent/caregiver inserted the swab into the infant's mouth and held onto it while the infant chewed and saliva was absorbed. After collection, the swab was inserted in a collection tube and transported to the laboratory. The collection tubes were centrifuged at 3000 x g for 10 minutes to collect all saliva at the bottom of the tube. If no saliva was obtained after centrifugation, 1 ml PBS was added to the swab and the tube was centrifuged. After collection of saliva from tubes, the saliva samples were boiled for 5 minutes and then centrifuged at 10000 x g for two minutes to pellet the debris. The saliva samples were either directly used for coating (1ml eluates), or diluted 1:1000 in PBS (pH 7.4) before coating microtiter plates (Dynex Technologies, Chantilly, VA) in duplicate (100 µl per well). After addition of saliva to plates, they were incubated at 37°C for 2 hours and then at 4°C overnight.

Then following day each well was washed four times with 200 µl PBS containing 0.05% Tween 20 (PBS-T) (Thermo Scientific), and then 100 µl horseradish peroxidase (HRP) conjugated *Ulex europaeus* agglutinin (diluted 1:1600 in PBS-T plus 3% bovine serum albumin (BSA)) was added to each well, and the plate was then incubated at 37°C for 1.5 hours. After incubation the plates were washed a second time with PBS only, and then 100 µL Tetramethylbenzidine (TMB) substrate was added and incubated for 15 min at room temperature. The reaction was stopped with 100 µL 1 M H_2SO_4 and the absorbance was measured at 450 nm using an iMark™ Microplate Absorbance Reader (Bio-Rad Laboratories Inc.).

3.2.4 FUT2 genotyping of non-secretor samples

Due to discrepancies observed in the pilot study between non-secretor genotype and HBGA phenotype, all virus-positive discrepant non-secretor results were confirmed by nucleotide sequencing of the *FUT2* gene. The 1263 bp exon 2 of the *FUT2* gene was amplified by PCR using Q5® Hot Start High-Fidelity DNA Polymerase and published primers (Table 3.1). The reaction mix consisted of the following: 1×Q5 Reaction buffer, 200 µM dNTPs, 500 nM of the forward and reverse primers respectively and 1 U Q5 Hot Start High-Fidelity DNA polymerase. The reaction mixture was then made up to 50 µL by adding 1 µL of the extracted genomic DNA template and an appropriate quantity of nuclease-free water. The Q5 Hot Start High-Fidelity DNA polymerase does not require a separate activation step. The PCR cycling conditions were as follows: initial denaturation (98 °C, 30 sec) followed by 35 cycles of denaturation (98°C, 30 sec), annealing (66 °C, 30 sec) and extension (72 °C, 90 sec). A final extension cycle (72°C, 2 min) was included to ensure full-length products.

Table 3.1: Primers for amplification of exon 2 of FUT2.

Primer name	Primer sequence	Fragment amplified	Source
FUT2Ex2F FUT2Ex2R	5'-ACACACCCACACTATGCCTGCAC-3' (forward) 3'-ACTTGCAGCCCAACGCATCTT-5' (reverse)	1 263 bp	Published primers (Ferrer-Admetlla *et al*., 2009)

After amplification, the PCR products (10 µL) were analysed by gel electrophoresis, and if a single band was present, the remaining PCR product (40 µL) was purified using the DNA Clean and Concentrator™-25 kit (Zymo Research, Irvine, CA), and eluted in 30 µL. Three µL of this purified sample, along with 3 µL 5X sequencing buffer, 1 µL terminator mix from the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), 1 µL of 3.2 pmol/µL forward or reverse primer respectively, and nuclease-free water were combined to a total of 20 µL. The sequencing reaction was performed with the following cycling conditions: initial denaturation (94˚C, 3 min), followed by 25 cycles of denaturation (94˚C, 30 sec), annealing (50˚C, 10 sec) and extension (60˚C, 4 min). After a sequencing PCR was completed the samples were sent to Inqaba Biotech for purification and analysis

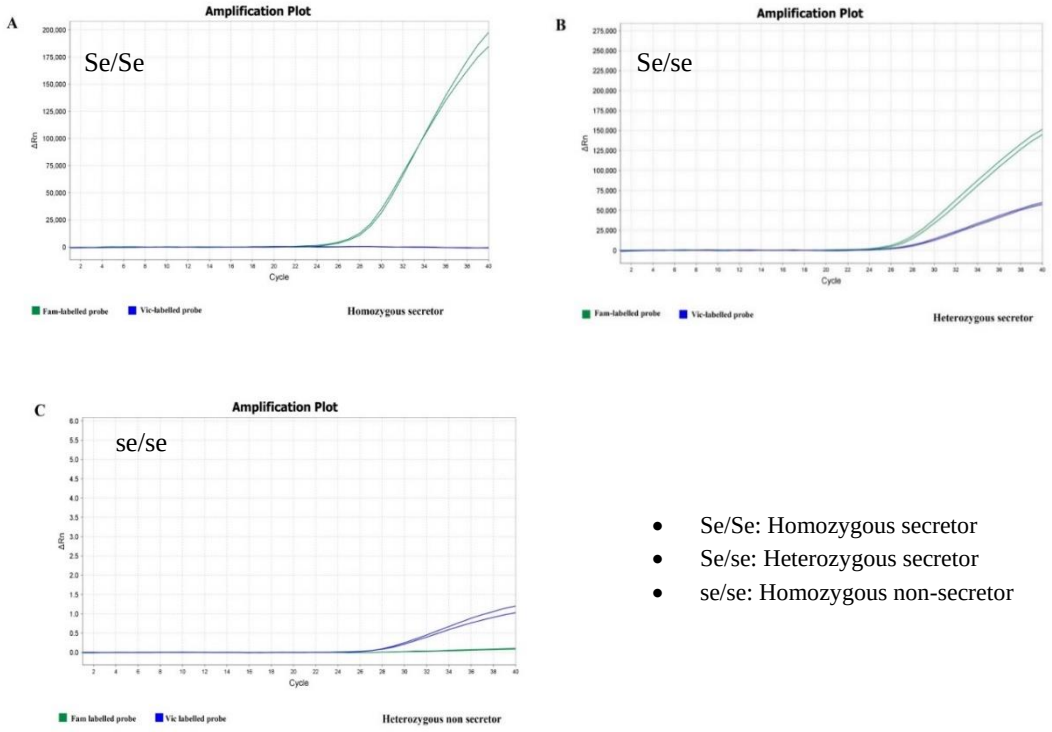
on an ABI 3130 automated analyser. Forward and reverse nucleotide sequences (text and chromatogram files) were received from Inqaba Biotech.

Nucleotide sequences were analysed using the Sequencher DNA Sequence Analysis Software (Gene Codes Corporation, Ann Arbor, MI). Sequences containing the G allele at the 428 bp position were deemed homozygous secretors (*SeSe*), sequences containing an A at this position were homozygous non-secretors (*sese*), and sequences containing both A and G at this position were determined to be heterozygous secretors (*Sese*).

3.3 Results

3.3.1 *FUT2* genotyping

To determine the secretor status of an individual, the *FUT2* gene was amplified to determine the presence of the G428A SNV (Figure 3.3). The detection of the respective dye-labelled probes is dependent on the wavelengths at which the reporter dyes emit fluorescence. The green line illustrates a G allele present (secretor) and the blue line an A allele at the SNV, indicating a non-secretor. *FUT2* genotyping in 205 children resulted in a 71%:29% ratio of secretors vs non secretors. This was comprised of 54 homozygous secretors, 92 heterozygous secretors and 59 homozygous non-secretors.



- Se/Se: Homozygous secretor
- Se/se: Heterozygous secretor
- se/se: Homozygous non-secretor

Figure 3.3: Real-time SNP assay to detect the G428A SNV in three individuals. A: Homozygous secretor, B: Heterozygous secretor and C: Homozygous non-secretor, determined by the amplification of FAM and/or VIC. Both non-secretors and heterozygous secretors, were detected by the VIC®-labelled probe. VIC® dye emits fluorescence at 554 nm. The G allele at the position 428 SNV, present in heterozygous and homozygous secretors, was detected by the FAM™-labelled probe. FAM™ dye emits fluorescence at 518 nm.

Following the completion of each real-time run, an allelic discrimination plot was created by the software supplied with the real-time instrument. The real-time instrument software provided the option of exporting the results into a Word Excel document, which provided the Delta R_n values for each sample. Normalised reporter (R_n) is the ratio of the fluorescence emission intensity of the reporter dye (FAM™ or VIC®) to the fluorescence emission intensity of the passive reference dye (ROX). Delta R_n (ΔR_n) is the normalisation of R_n by subtracting the baseline. The ΔR_n values of each sample were combined into one allelic discrimination plot (Figure 3.4). Therefore, each sample would have a ΔR_n value for the allele with the SNV and the allele without the SNV.

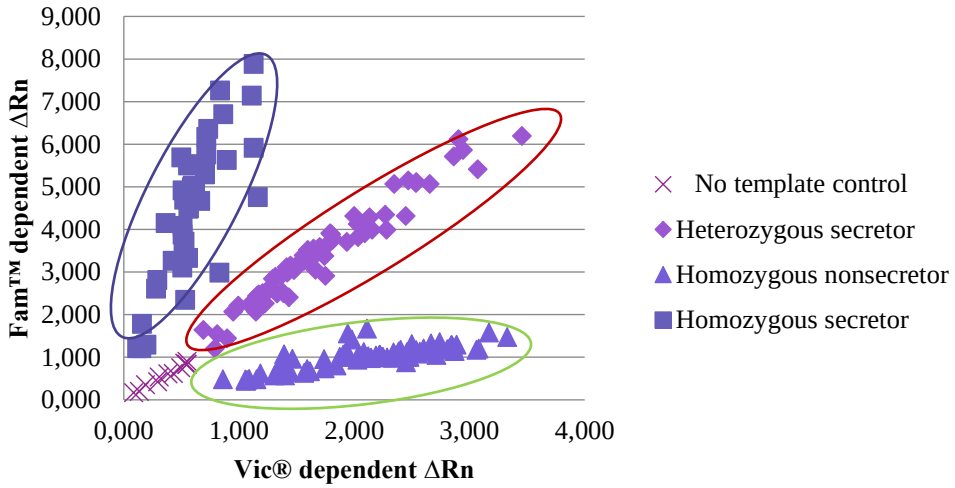


Figure 3.4: Allelic FUT2 discrimination plot of the Delta Rn (ΔRn) values of 205 samples in duplicate. Each symbol represents one of the duplicate reactions, and the position of each symbol is determined by its VIC®-dependent ΔRn (x-axis value) and its FAM™-dependent ΔRn (y-axis value). Each secretor genotype is shown in different colours.

A selection of 23 non-secretor saliva specimens underwent conventional PCR and sequencing, which resulted in an accuracy of 48% (11/23), as 12 specimens were proven to be heterozygous secretors. An additional genotyping method was performed to determine the *FUT2* status of the children via EIA with saliva. This was to verify the *FUT2* genotyping results, based on each child's saliva and DNA, which can differ due to feeding habits (Colston *et al.*, 2019). A comparison between the *FUT2* genotyping results obtained via real-time PCR and EIA is shown in Figure 3.5. After comparing the different methods, it is clear that conventional PCR is still the most accurate in determining the *FUT2* genotype, although the real-time RT-PCR was still found to be more accurate than *FUT2* genotyping through EIA methods using Lectin.

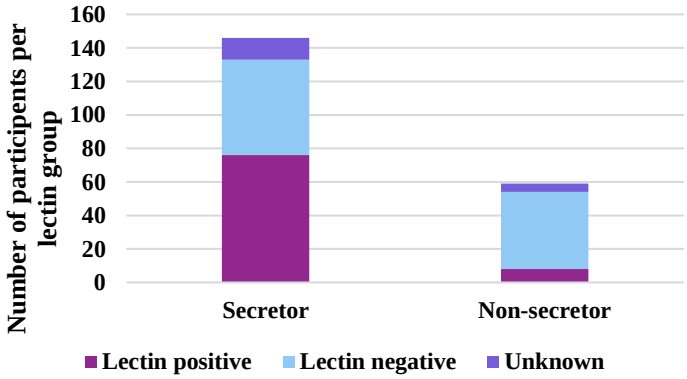


Figure 3.5: Comparison of *FUT2* genotyping through RT-PCR (Secretor/Non-secretor) and EIA (Lectin) from 205 children ≤ 5 years, who were hospitalised with gastroenteritis between July 2016 and December 2017 at KPTH.

3.3.2 *FUT2* and virus genotypes

The virus genotypes identified in the children were correlated with their *FUT2* genotype as determined by real-time as well as conventional PCR. Seventy five percent (24/32) of the children infected with either norovirus GI or GII were secretors, whereas 25% (8/32) were non-secretors (Figure 3.6). In the non-secretor population, four genotypes were identified, while two specimen's strains could not be determined due to low viral load. Between the secretors, a total of four genotypes could also be determined, with one untypable strain observed. In secretors, the GII.4 genotype represented 71% (17/23 typed) of the norovirus strains, whereas in non-secretors it constituted 40% (2/5 typed) of strains. This shows that GII.4 preferentially infect secretors (17/19) in a strain specific manner (p=0.035).

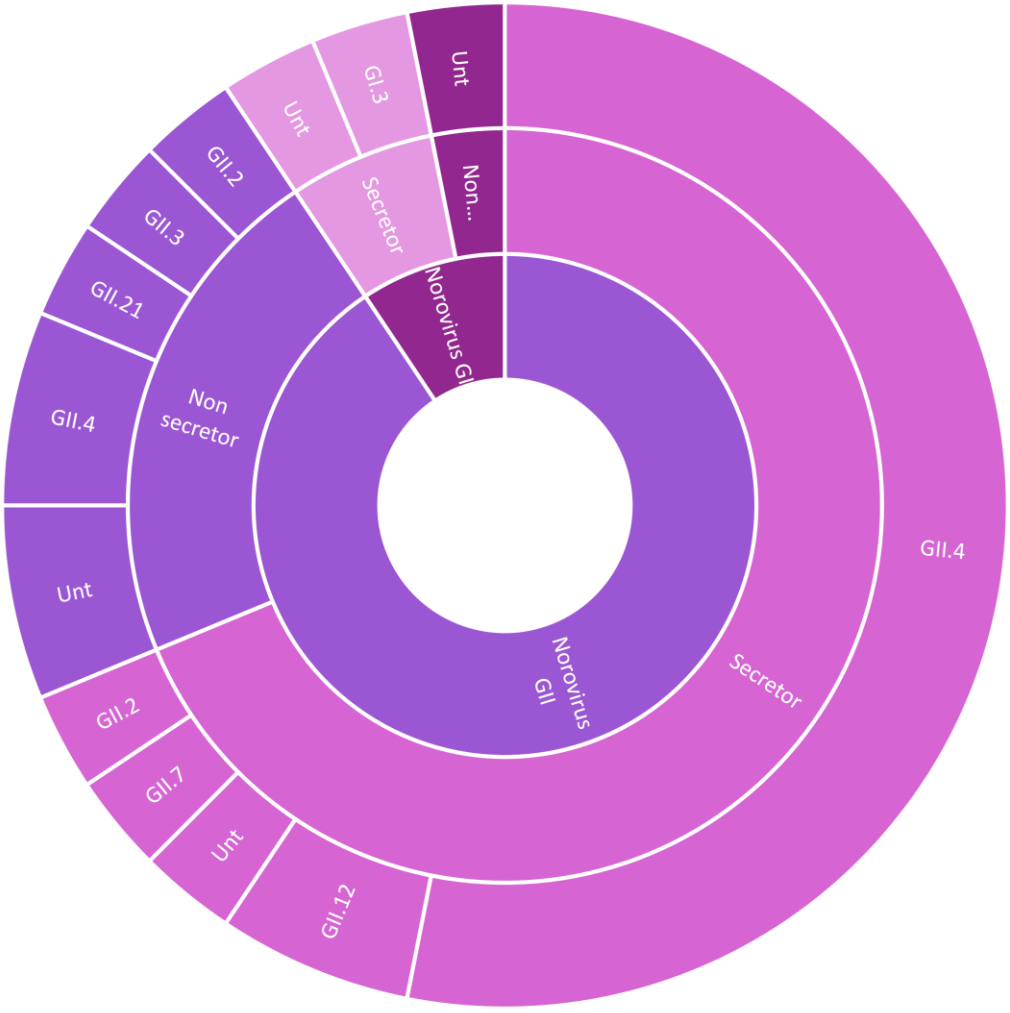


Figure 3.6: Comparison of norovirus infection between secretors and non-secretors as determined from 32 children ≤5 years, who were hospitalised with gastroenteritis. Unt: untypable, Non…: Non-secretor

When comparing rotavirus positive patients, 91% (42/46) of the total population were secretors, whereas 9% (4/46) were non-secretors. A total of three genotypes were identified in the non-secretor population, with G8P[6] (2/4) being most predominant (Figure 3.7). For the secretor positive children infected with rotavirus nine strains were determined, with G3P[4] (17/42) predominating. When the other virus-positive (sapovirus, astrovirus and adenovirus) patients' *FUT2* genotyping analysis was performed, these ratios were closer to 50:50, indicating no specific preference of these viruses for secretors or non-secretors (Figure 3.8)

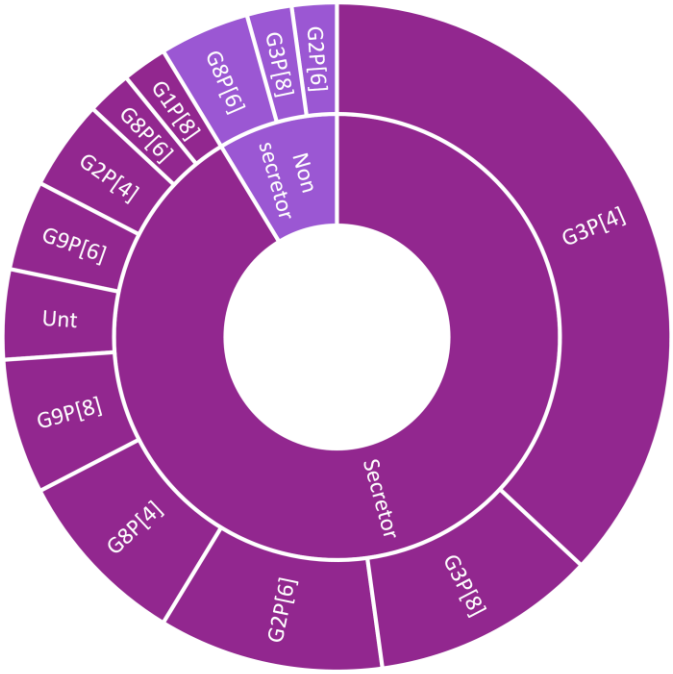


Figure 3.7: Comparison of rotavirus infection between secretors and non-secretors as determined from 46 children ≤ 5 years, who were hospitalised with gastroenteritis. Unt: untypable

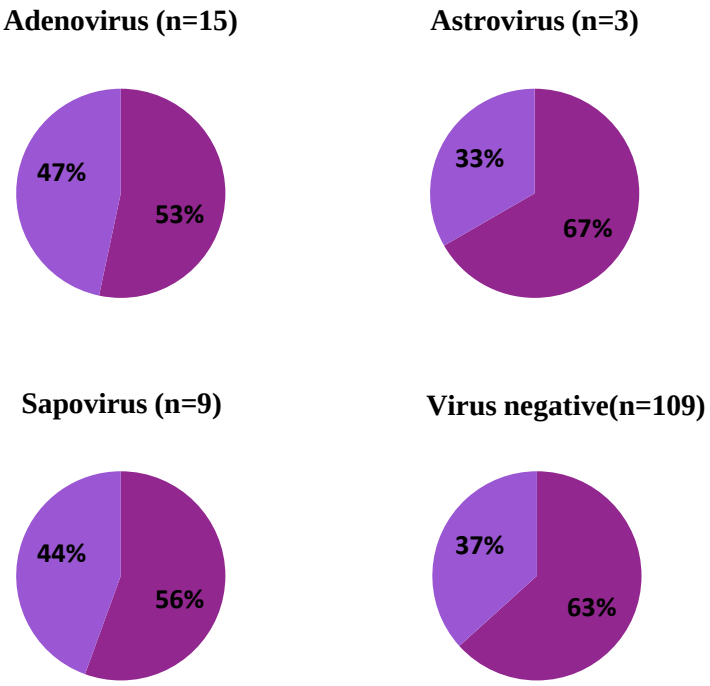


Figure 3.8: Comparison between secretors and non-secretors from adeno-, astro-, sapovirus infected, as well as virus negative children ≤5 years, who were hospitalised with gastroenteritis. Non-secretor Secretor

3.4 Discussion

FUT2 genotypes were determined using three different tests: real-time PCR, conventional PCR combined with nucleotide sequencing, and a lectin EIA. In the beginning of the study, only real-time PCR was planned, with conventional PCR added as a secondary confirmatory method, when discrepant results were observed between secretor genotype and saliva HBGA phenotype. A recent study suggested that breastmilk can affect secretory phenotype in saliva (Colston *et al*., 2019), following which it was decided to do a third confirmatory EIA, with saliva specimens instead of blood, indicating whether breastmilk affected the saliva specifically. This may explain why some *FUT2* genotypes and HBGA phenotypes did not agree.

FUT2 genotyping via real-time PCR resulted in 54 homozygous secretors, 92 heterozygous secretors and 59 homozygous non-secretors, with all virus positive non-secretors verified with conventional PCR and Sanger sequencing. Only virus positive non-secretors were verified due to time and funding constraints. The final results indicated a ratio of 71%:29% for secretors and non-secretors, which is similar to findings of other studies in the African setting (Liu *et al.*, 1998, Nordgren *et al*., 2013, Parker *et al*., 2018).

When comparing real-time PCR non-secretor results with conventional PCR, only 52% (12/23) of the participants were correctly identified as non-secretors, whereas the rest were heterozygous secretors, as shown with conventional PCR. The reasons for this discrepancy are unknown, albeit very worrying. Possible explanations include faulty probes, further variations at the site where the probe attaches, and inhibition of the secretor probe due to the non-secretor probe or other unknown substances present in the DNA. When comparing real-time PCR with the lectin EIA assay, multiple discrepancies were observed (Figure 3.5). While this was expected for FUT2 non-secretors (as breastfeeding can alter their HBGAs), more discrepancies were observed in the secretors, with multiple lectin false negatives, indicating that the EIA assay's accuracy was lower than expected and optimisation would be needed for more conclusive results.

The determined secretor status was compared with norovirus and rotavirus infection detected. Norovirus infection was observed at a ratio of 75%:25% for secretors and non-secretors respectively, indicating a preference for norovirus infection in secretor positive patients, especially when comparing these ratios with the virus negative group (63%:37%), although $p>0.05$, indicating no significant difference. When analysing only norovirus GII.4

infections (n=19), the ratio added up to 89%:11% for secretors and non-secretors, which is significantly different from ratios of the virus negative group (p=0.035). Rotavirus infection was observed at a ratio of 91%:9% for secretors and non-secretors, indicating that rotavirus preferentially infects secretors (p<0.01).

These comparisons were also made for all adeno-, astro, and sapovirus infected children, indicating that secretor status has no significant effect on these viral infections, as has been observed in previous studies (Matussek *et al*., 2015, Oka *et al*., 2015, Colston *et al*., 2019).

The noro- and rotavirus genotypes observed in the study were also compared between secretors and non-secretors. For norovirus, GII.4 is known to preferentially infect secretors (Frenck *et al*., 2012, Nordgren and Svensson, 2019), as was determined in this study as well (17/19). Two non-secretor children were also infected by norovirus GII.4. While most studies have shown that non-secretors are protected against GII.4 infection, exceptions have also been observed (Carlsson *et al.*, 2009, Frenck *et al*., 2012, Jin *et al.*, 2013, Nordgren *et al*., 2013, Liu *et al.*, 2014). The reasons for these exceptions are not yet clear cut, but could include microbiota diversity, such as HBGA-expressing bacteria (Miura *et al.*, 2013), environment or feeding habits (Colston *et al*., 2019), differences between GII.4 variants, general health of the child, weak-secretor phenotype, along with other unidentified host factors, indicating that more studies are needed to gain a better understanding of these mechanisms. Rotavirus P[4] and P[8] genotypes have been found to bind to secretor HBGAs, such as Le(b) and H-type 1, while the P[6] genotype is more likely to infect non-secretors (Liu *et al*., 2012b, Nordgren *et al.*, 2014, Pollock *et al.*, 2018). This was demonstrated in this study, where mostly P[4] and P[8] genotypes infected secretors, whereas mostly P[6] genotypes infected non-secretors, apart from one non-secretor child, who was infected with G3P[8]. It should be noted that this specific child had a co-infection of rotavirus and adenovirus-F, with no clear way to interpret which virus (or if the combination of the two) caused severe gastroenteritis symptoms.

This study re-enforced results from previous studies, indicating that noro- and rotavirus preferentially infect secretors or non-secretors in a strain-specific manner. For norovirus GII.4 specifically was more prevalent in secretors (89%), when compared to non-secretors, while less common genotypes such as GII.2 and GII.3 were observed in non-secretors. In rotavirus infected children, types P[4] and P[8] were observed more frequently in secretors, whereas P[6] was the most common genotype observed in non-secretors. Because the ratios

of secretors are race specific, the strain specificity of noro- and rotavirus interactions with secretor status might have an influence on circulating genotypes in any given population. This study also provided support for the hypothesis that host genetics play an important role in pathogen susceptibility and the circulation of specific noro- and rotavirus strains.

Chapter 4 : Saliva carbohydrate phenotyping

4.1 Introduction

Through retrospective analysis it has been proven in several studies (Lindesmith *et al*., 2003, Hutson *et al*., 2004, Carlsson *et al*., 2009, Liu *et al*., 2012b, Jin *et al*., 2013, Ayouni *et al*., 2015) that *FUT2* has an influence on the risk of infection with norovirus and rotavirus. This is because susceptibility of the host is influenced by the presence of HBGAs on the surfaces of the intestinal epithelium, which allow interaction and binding with the virus. Histo-blood group antigen diversity is a result of the sequential addition of monosaccharides to glycan precursors (Robilotti *et al*., 2015). Histo-blood group antigen type 1 core structures are widely expressed in endodermally derived tissues, such as the lining epithelia and glandular epithelia, where type 2 core structures are found mainly in ecto- and mesodermally derived tissues, which includes skin and erythrocytes (Shirato *et al*., 2008). Type 1 carbohydrates bind more tightly to norovirus VLPs, derived from the prototype strain of norovirus, Norwalk virus, than type 2 carbohydrates. This characteristic may be what allows norovirus tissue specificity (Shirato *et al*., 2008). Although the binding patterns of human norovirus VLPs to HBGAs are variable, strains in the same genotype show a marked tendency to exhibit the same HBGA binding patterns (Shirato, 2011). The large structural diversity found in these carbohydrates are believed to have evolved due to pressure from the environment, which plays an important role in the symbioses, commensalism and parasitism between humans and microorganisms (Henry, 2001, Lewis *et al*., 2015). The divergence of human norovirus is believed to be driven by the different HBGAs found in hosts (Jin *et al*., 2013). Prevalent norovirus genotypes such as GII.4 are most frequently observed in secretor positive patients when compared to secretor negative patients (Frenck *et al*., 2012).

Susceptibility to rotavirus is also found to be much higher in secretors when compared to non-secretors of HBGAs in saliva. It has been shown that the VP8 particle of the VP4 spike on the rotavirus capsid is responsible for the binding to different HBGAs in hosts (Liu *et al*., 2012b). This results in the attachment of rotavirus to the host cells in a P genotype-dependent manner (Huang *et al*., 2012). Higher risk of infection was found among subjects with Lewis (a and b) or Le(x and y) than subjects with Le(a) or Le(x) (Zhang *et al*., 2016b). Two closely related P genotypes, P[8] and P[4] have been found to bind to Le(b) and H-type 1 antigens, while genetically more distant genotypes, such as P[6], only recognise H-type 1 antigens and are more likely to be found in non-secretors (Liu *et al*., 2012b, Nordgren *et al*., 2014, Lee *et al*., 2018, Pollock *et al*., 2018). There is no significant difference when regarding the subjects with the A, B and O secretor blood types

(Zhang *et al*., 2016b). It should be noted that conflicting research, where P[8] rotavirus was observed in both secretor and non-secretor individuals has been published (Ayouni *et al*., 2015), indicating the need for further studies to fully understand the role of the HBGA phenotypes in rotavirus susceptibility and infection. Recent studies have shown that norovirus can bind to H type 2 Lewis antigens (Le(x) and Le(y)) in saliva but not in a synthetic membrane environment (Nasir *et al*., 2017). This suggests that the molecular structure of the HBGAs change in the environment, whether in the gut or in saliva, indicating that Le(x) and Le(y) has little impact on susceptibility to norovirus binding (Carmona-Vicente *et al*., 2016, Nasir *et al*., 2017), which can in turn also be true for rotavirus binding if they do indeed use the same mechanisms. This steered us to look at norovirus and rotavirus, whilst disregarding Le(x) and Le(y).

4.2 Materials and methods

Saliva samples were screened for the presence of A, B and H antigens as well as Lewis antigens through enzyme immunoassay (EIA). Saliva samples were obtained and processed as stated in section 3.2.3. The processed saliva samples (100 µL) were coated onto microtiter plates in triplicate (Dynex Technology), and incubated at 37°C for two hours and then overnight at 4˚C. The following day, after the wells were washed four times with 200 µL PBS-T (Thermo Scientific), 100 µL of mouse-derived monoclonal antibodies (MAbs) specific to human Lewis and ABH antigens were added to the wells. The following MAbs were used for phenotyping: BG-4 anti-H type 1 (17-206), Seraclone® anti-Le(a) and DiaClon® anti-Le(b) (Bio-rad) and anti-A and anti-B (Diagast, Parc Eurasanté, France).

Monoclonal antibodies anti-H1was used at a dilution of 1:500, and anti-Le(a) -Le(b), -type A and –type B at a 1:5000 dilution, with PBS-T with 5% foetal bovine serum (Invitrogen), to a final volume of 100 µL per well. The plates were then incubated for 90 min at 37˚C, after which they were washed with 200 µL of PBS-T four times. Afterwards, 100 µL of horseradish peroxidase conjugated goat anti-mouse IgG (Thermo Fisher) or IgM antibodies (Novus Biologicals, Centennial, Colorado, United States) were added, at a dilution ratio of 1:2000 and 1:13 000 respectively. The plates were incubated at 37°C for another 90 min and then washed with 200 µL PBS for a final time. TMB substrate solution (KPL) was prepared at a ratio of 1:1 of TMB Peroxidase Substrate and Peroxidase Substrate solution B, of which 100 µL was added to each well and incubated with the plate covered in foil for 15 minutes at room temperature. After 15 minutes the reaction was stopped with 100 µL 1 M H_2SO_4 (Merck & Co), and the absorbance was measured

at 450 nm using an iMark™ Microplate Absorbance Reader (Bio-Rad Laboratories Inc., Hercules, CA).

An IgG and IgM conjugate control (saliva + conjugate without the primary antibodies) was included for each sample to eliminate non-specific background binding by the conjugate. The final OD value was calculated by subtracting the average conjugate OD value + 2XSD of the conjugate from the average OD values of each sample. If a sample value was ≥0.1 after subtraction for a specific antibody, the sample was considered positive for that specific antibody binding, indicating that the person possessed the specific antigen.

To determine whether specific virus genotype infections were preferentially associated with ABO and Lewis antigens or secretor and non-secretor genotypes, contingency tables and Fischer's exact tests were performed.

4.3 Results

4.3.1 HBGA distribution

The HBGA phenotypes, comprised of different combinations of H-type 1, type A, type B, Type AB, Le(a) and Le(b) antigens, as well as a HBGA negative group. The HBGA phenotypes could be determined in 174/205 saliva specimens, with 9 specimens insufficient for processing and 22 specimens indicating no HBGAs present. The HBGA negative phenotype was also included as a combination, leading to 26 different HBGA phenotypes observed, with Le(b) being present in most of the specimens (141/174), followed by Le(a) (81/174), Type A (78/174), H-type 1 (47/174) and Type B (43/174). Multiple secretors (Le(b) and/or A/B positive) were also Le(a) positive (Table 4.1), which are referred to as partial secretors or partial non-secretors (Jin *et al*., 2013).

Due to discrepant results observed between non-secretor genotype and HBGA phenotype in 43/57 children, *FUT2* genotyping and lectin phenotyping results from chapter 3 and HBGA phenotyping results from chapter 4 were compared (Table 4.1). Possible confounding factors, such as breastfeeding or discoloured saliva (indicating possible interfering substances), were included in Table 4.1 and Appendix F (detailed table) to allow meaningful analysis of the results.

Table 4.1: Comparison of *FUT2* genotyping, lectin phenotyping and HBGA phenotyping results, with possible confounders (breastfeeding and saliva discolouration). All highlighted numbers indicate discrepant results. All results were obtained from blood (n=190) and stool (n=15) for *FUT2* genotyping, and saliva (n=196) for lectin and HBGA phenotyping, collected from 205 children ≤5 years who were hospitalised with gastroenteritis at KPTH.

FUT2 genotype	Lectin EIA	Breastfed	Saliva	ABH EIA	Le(a) EIA	Le(b) EIA	HBGA negative
Secretor (n=146)	**Lectin positive (n=76)**	Yes = 29	10 Discoloured	8	4	7	0
			11 Normal	9	4	11	0
			8 Unknown	2	3	7	1
		No = 45	11 Discoloured	10	7	11	0
			17 Normal	10	8	15	1
			17 Unknown	14	2	15	1
		Unknown = 2	1 Discoloured	0	1	1	0
			1 Normal	0	0	0	1
	Lectin negative (n=57)	Yes = 16	4 Discoloured	2	2	2	1
			7 Normal	5	5	6	0 (1 NA)
			5 Unknown	1	0	3	1
		No = 40	9 Discoloured	7	4	7	1
			25 Normal	16	7	19	2
			6 Unknown	2	1	6	0
		Unknown = 1	1 Normal	1	0	1	0
	Not available (n=13)	Yes = 7	7 Unknown	4	1	3	2 (1 NA)
		No = 5	5 Unknown	0	0	0	1 (4 NA)
		Unknown = 1	1 Unknown	0	0	0	0 (1 NA)
Non-secretors (n=59)	**Lectin positive (n=8)**	Yes = 2	1 Normal	0	1	1	0
			1 Unknown	0	0	1	0
		No = 5	1 Discoloured	1	1	1	0
			2 Normal	1	1	0	1
			2 Unknown	2	0	1	0
		Unknown = 1	1 Discoloured	1	1	1	0
	Lectin negative (n=46)	Yes = 13	4 Discoloured	1	1	1	2
			6 Normal	2	4	3	2
			3 Unknown	1	2	2	0
		No = 30	7 Discoloured	7	6	3	0
			14 Normal	9	8	5	3
			9 Unknown	4	6	4	1
		3 Unknown	1 Normal	0	0	1	0
			2 Unknown	2	1	1	0
	Not available (n=5)	Yes = 4	4 Unknown	1	0	2	1 (1 NA)
		No = 1	1 Unknown	0	0	0	1 NA

When comparing HBGA phenotypes with *FUT2* genotypes (secretor status), various unexpected HBGAs were observed in non-secretors (Figure 4.1), with various ratios of specific HBGAs found in each group. When comparing each HBGA observed separately, Le(b) was most prevalent in secretors, and Le(a) most prevalent in non-secretors.

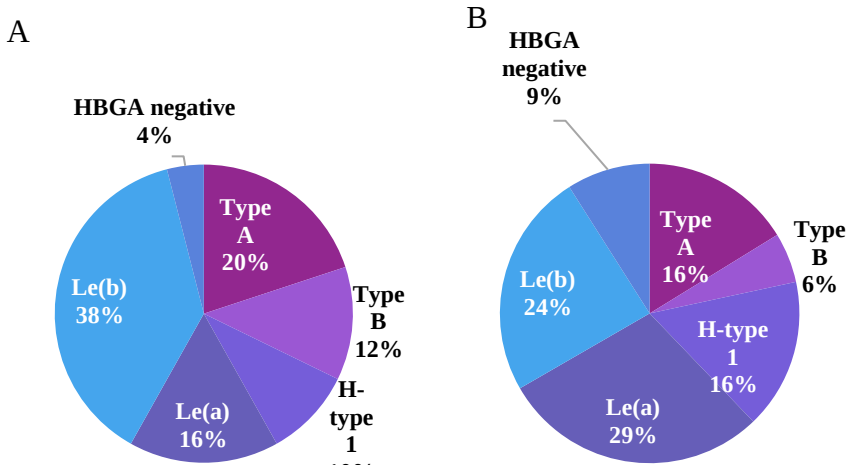


Figure 4.1: HBGA distribution of (A) secretors, (B) non-secretors in 196 children ≤5 years who were hospitalised with gastroenteritis.

4.3.2 HBGA phenotypes and gastroenteritis virus infection

The distribution rates of the HBGAs between virus infected and virus negative participants were similar, apart from H Type 1 and Le(a) (Figure 4.2). No statistical difference could be determined for any specific HBGA phenotype in virus- infected or -negative groups.

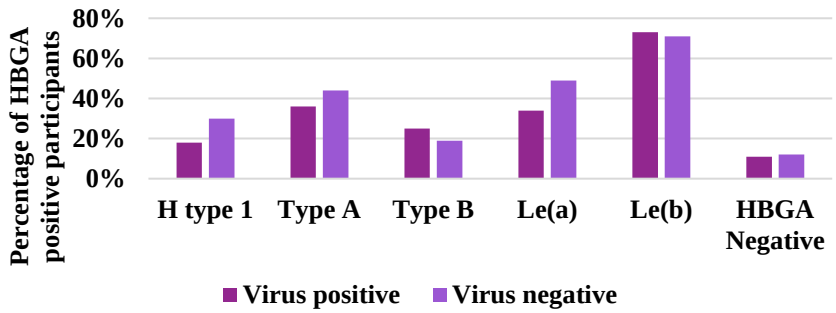


Figure 4.2: Distribution of HBGA antigens in 196 children hospitalised with gastroenteritis either infected with a gastroenteritis virus (n = 95) or uninfected (n = 101).

The distribution of the HBGA phenotypes between adeno-/ astro-/ sapovirus infected and uninfected children were very similar for each virus during this study, with no statistically supported HBGA preference observed (Figure 4.3). The distribution of HBGAs between astrovirus infected and astrovirus uninfected participants must be viewed with caution, as low specimen numbers could give a skewed result. Contingency tables with Fischer's exact testing showed no statistical preference for adeno-/astro-/sapovirus infection to a specific HBGA type. The distribution of HBGAs between norovirus-infected and -uninfected groups as well as rotavirus-infected and -uninfected groups differed marginally for specific HBGA types. Norovirus infection was shown to be less prevalent in A type positive participants (6/32 $p=0.01$), whereas rotavirus infection was linked to Type B positive patients (20/46, $p<0.001$) (Figure 4.3). These results should be viewed with caution, as the statistical analysis was performed on each HBGA specifically, where combinations of HBGAs (HBGA phenotypes) are normally observed.

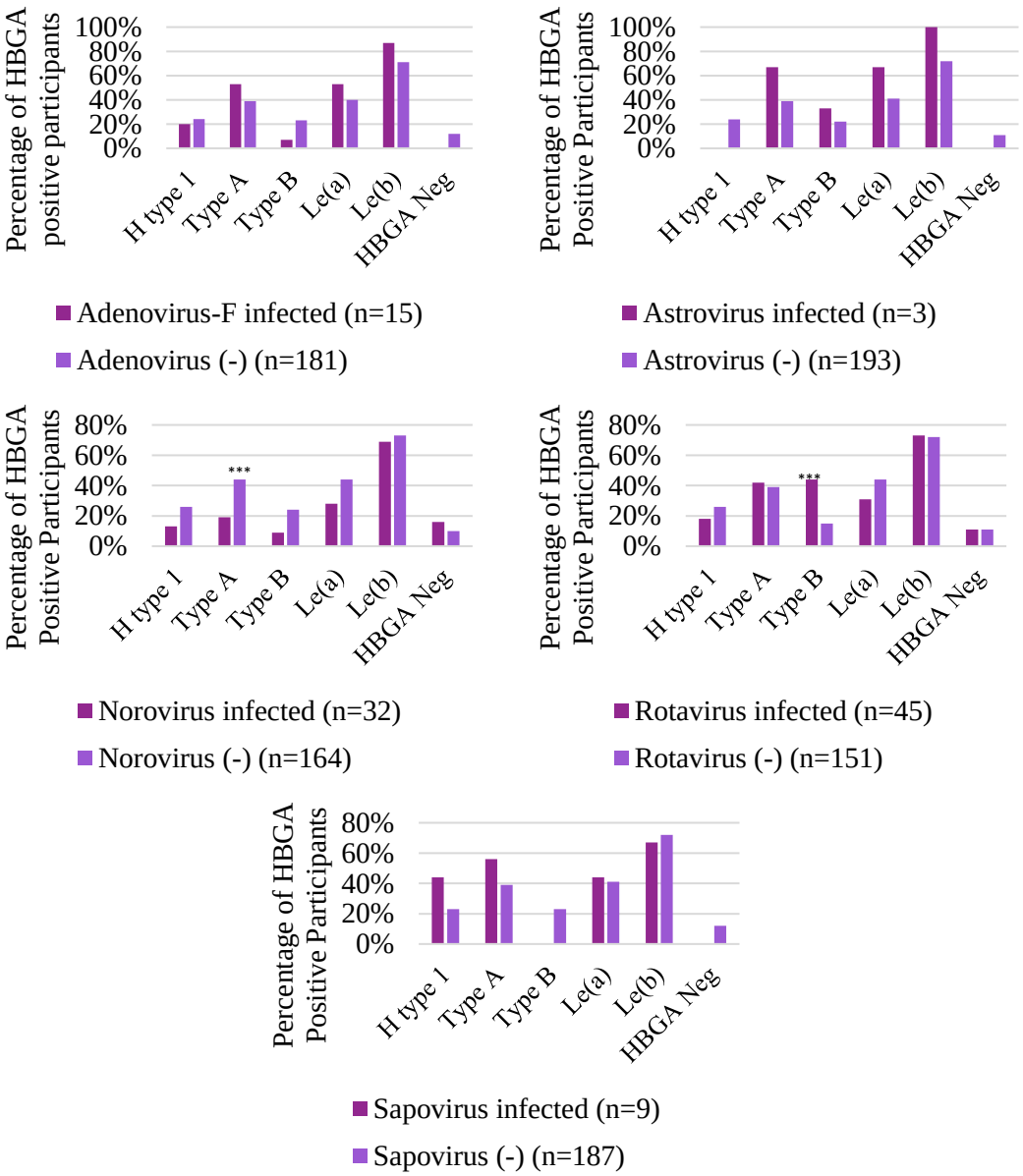


Figure 4.3: Ratios of HBGA phenotypes observed in virus-positive or -negative children (<5 years) hospitalised with gastroenteritis, divided per virus infection. * indicate statistical significance (p≤0.01).**

The number of children per HBGA phenotype within the group of 26 different combinations showed increased susceptibility to norovirus in HBGA phenotype Le(b) only children (p=0.036), and to rotavirus in HBGA phenotypes A, B, Le(b) (p<0.01), A, H1, B, Le(b) (p<0.05), B, Le(b) (p=0.023) or H1, B, Le(a/b) (p<0.05) positive children (Figure 4.4).

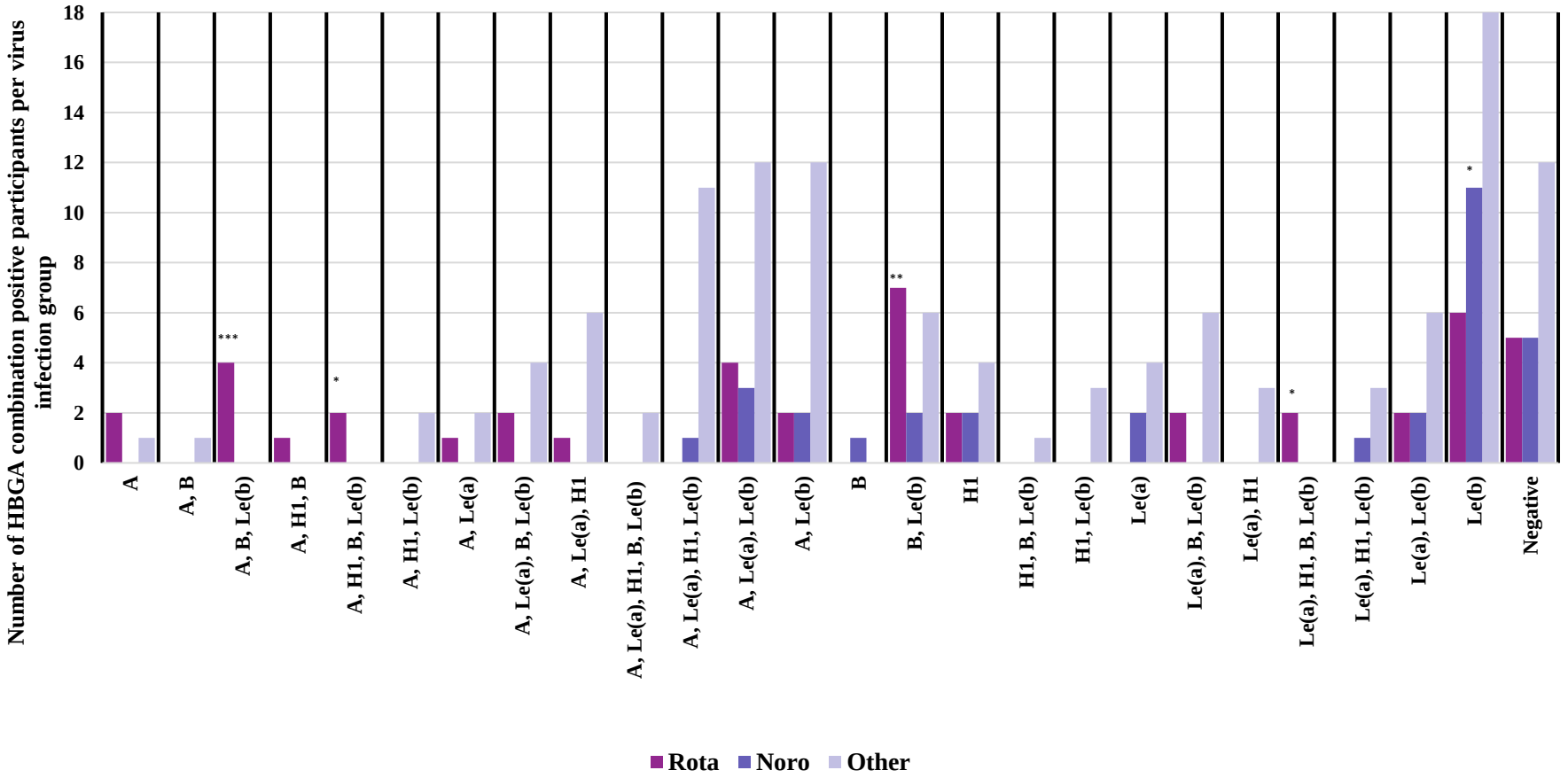


Figure 4.4: Graph depicting the different HBGA combinations observed (n=26), with specific markers for noro- and rotavirus-infected patients. All specimens (n=196) were received from children ≤5 years, who were hospitalised with gastroenteritis at KTPH between June 2016 and December 2017. * indicate statistical significance. (* p≤0.01, **0.4>p>0.01, *p≤0.5)**

Norovirus genotypes were each compared to the specific HBGA observed (Figure 4.5A) as well as with specific HBGA phenotypes (Figure 4.5B), with inconclusive results due to small sample size. An interesting observation was that norovirus type GII.4 was present in each HBGA phenotype of the norovirus infected children, indicating a diverse attachment range.

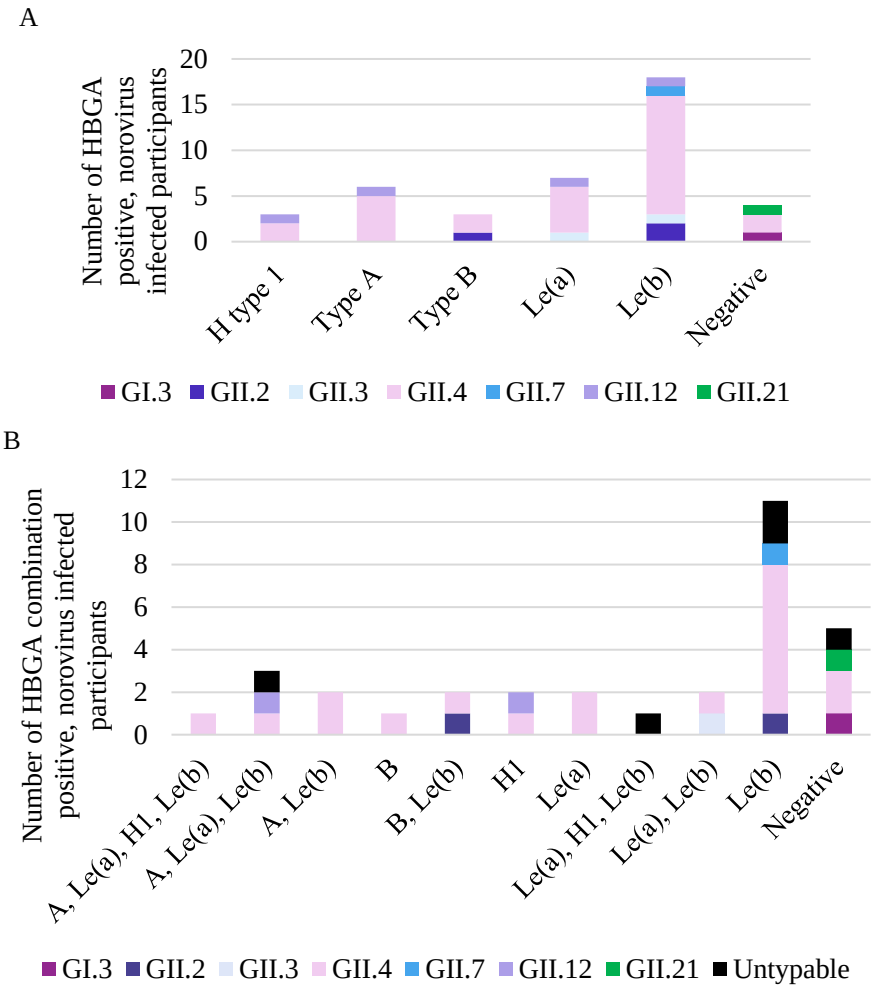


Figure 4.5: A) Ratios of specific HBGA phenotypes as found in 29 genotyped norovirus infected children B) Ratios of HBGA combinations found in 32 norovirus infected children. All children were ≤5 years and hospitalised with gastroenteritis at KPTH.

When exploring further into specific rotavirus strains, specifically the VP4 (P-type), compared to HBGA phenotype distributions, a different picture could be seen (Figure 4.6A) after performing contingency tables with Fischer's exact test, it was observed that P[4] was

statistically linked to HBGA phenotype B ($p<0.01$) and possibly Lewis (b) ($p=0.058$). P[6] infection was less likely to infect Lewis (b) positive patients ($p<0.01$), and P[8] was more likely to infect type B ($p<0.01$) and possibly type H1 ($p=0.053$) positive patients. All data must be interpreted with caution due to small sample sizes. As was mentioned previously, analysing individual HBGA types is not as conclusive as analysis of HBGA combinations and comparing that to rotavirus strains. The combinations compared to specific P-types is shown in Figure 4.6B. P[4] was more likely in patients with the HBGA combination A, B, Le(b) ($p<0.01$) or B, Le(b) ($p<0.01$). P[6] rotavirus infection was more likely in children with HBGA type A only ($p<0.01$) or with no HBGAs present in their saliva ($p=0.018$). Lastly, P[8] infections were more likely in children with the HBGA combinations A, H1, B ($p<0.05$) or A, H1, B, Le(b) ($p<0.01$).

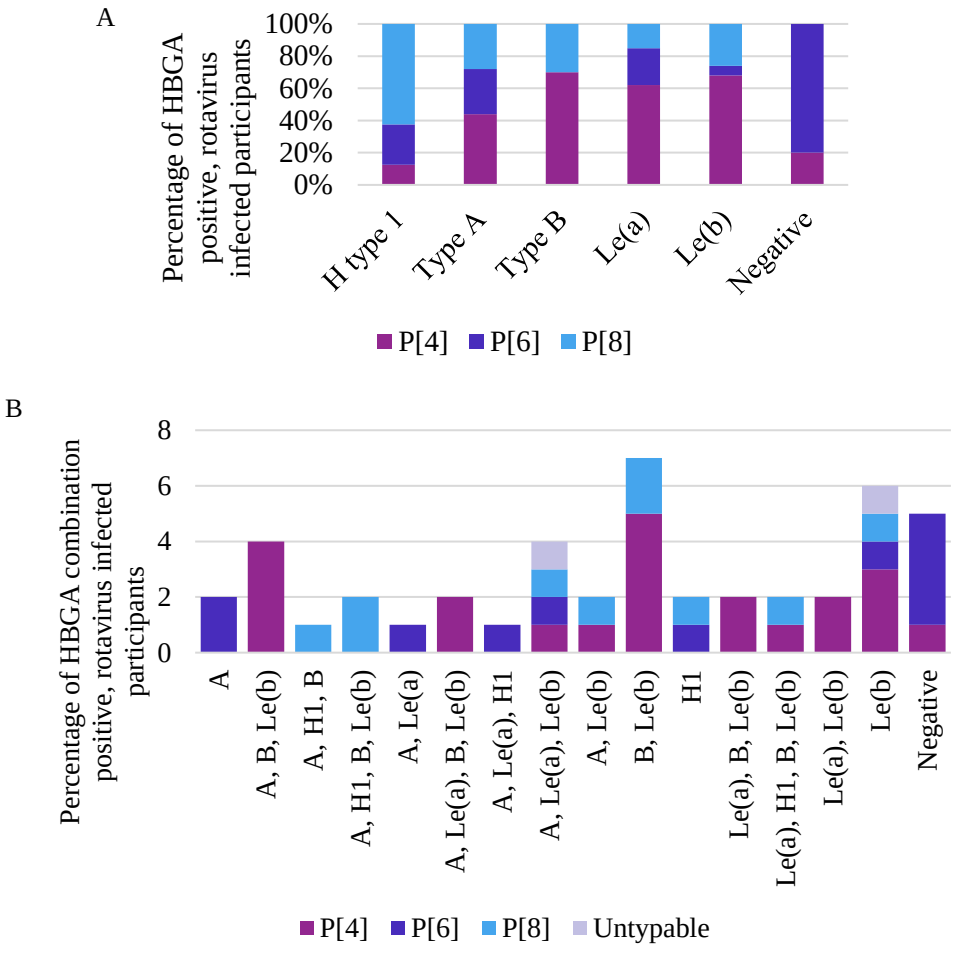


Figure 4.6: A) Ratios of specific HBGA phenotypes as found in 44 genotyped rotavirus infected children B) Ratios of HBGA combinations found in 45 rotavirus infected children. All children were ≤5 years and hospitalised with gastroenteritis at KPTH.

4.4 Discussion

Noro- and rotavirus infections can and have caused worldwide epidemics of acute gastroenteritis. Due to their nature of high levels of transmission and low infectious dose these viruses are difficult to control and prevent. This is why the understanding of host susceptibility to these viruses will help with the fight against them (Jin *et al*., 2013).

Various combinations of HBGA profiles were observed in this study, however with inconsistent results in terms of *FUT2* secretor status and HBGA phenotypes. The goat anti-mouse IgG conjugate was problematic, displaying high levels of non-specific background binding in certain samples. The conjugate control OD value + 2XSD was subtracted from the average OD reading of each sample, however, the high background may have masked lower levels of HBGA antigens in certain saliva specimens. To solve this problem, a new IgG conjugate generated in a different species (rabbit) was obtained. This reduced the non-specific background binding, but discrepant results were still obtained. Children genotyped as non-secretors tested positive for secretor HBGAs in their saliva. Possible explanations for this observation are 1) that their HBGA diversity was affected by breastfeeding (Colston *et al*., 2019), although only 14/45 children were breastfed, 2) another explanation could be non-specific binding to unknown contaminants in the saliva, as many of the specimens obtained varied in colour from pink to dark brown since the children could not rinse their mouths before specimen collection. Another consideration is that the saliva was collected as soon as the participant was enrolled, and not at a consistent time (e.g. 8-10 am), as has been done in previous studies (Carmona-Vicente *et al*., 2016), to minimise the effect of the circadian rhythm. When comparing HBGA phenotypes with *FUT2* genotypes (secretor status), HBGA phenotype Le(b) predominated in secretors, and Le(a) in non-secretors, consistent with previous studies (Liu *et al*., 2014, de Graaf *et al*., 2016).

When comparing HBGA phenotypes with specific gastroenteritis virus infection, HBGA phenotypes appeared to have no effect on the susceptibility of infection for adeno- astro- and sapovirus. This is to be expected as HBGAs are not known to play a role in these viral infections (Matussek *et al*., 2015, Oka *et al*., 2015, Colston *et al*., 2019). Contingency tables comparing rotavirus positivity, norovirus positivity or virus positivity in combination with specific HBGA and Lewis sugars showed that children with HBGA type A are less

susceptible to norovirus infection (p=0.01), and type B more susceptible to rotavirus infection (p<0.01), when using double-sided Fisher's exact test.

However, HBGAs generally occur in combinations and analysis of susceptibility to infection has to take these combinations into account. Covariate analysis was performed on both noro- and rotavirus with the 26 HBGA combinations observed in the study. In terms of norovirus infection, HBGA phenotype Le(b) only children were found to be statistically associated with infection (p=0.036). For rotavirus infection, individuals with the combination of A, B, and Le(b) (p<0.01) A, H1, B, Le(b) (p<0.05), B, Le(b) (p=0.023) or H1, B, Le(a/b) (p<0.05) were more susceptible to rotavirus infection. In terms of specific virus genotypes, too small sample size of any specific norovirus genotype was available to determine any statistically significant results, although the GII.4 type was observed in all norovirus infected children's HBGA phenotypes. For rotavirus, genotypes containing the P[4] VP4 strain infection was more common in children with HBGA combinations A, B, Le(b) (p<0.01) or B, Le(b) (p<0.01). P[6] genotypes were more frequently observed in children with HBGA type A only (p<0.01) or with no HBGAs present in their saliva (p=0.018). As previous studies have also linked P[6] infections with non-secretors (Liu *et al*., 2012b, Nordgren *et al*., 2014, Lee *et al*., 2018, Pollock *et al*., 2018), the observed increase in susceptibility to P[6] rotavirus infections for HBGA negative children is not unexpected. This may indicate that rotavirus P[6] genotypes also use other mechanisms than binding to HBGAs to cause infection or use type 1 precursors specifically to facilitate interaction between the virus and the host (Huang *et al*., 2012, Ma *et al.*, 2015). Lastly, P[8] rotavirus genotype infections were observed more frequently in children with HBGA combinations A, H1, B (p<0.05) or A, H1, B, Le(b) (p<0.01). This was not surprising as P[8] rotavirus strains are known to infect secretors preferentially, with previous studies indicating a link between P[8] and P[4] and HBGA types H1 and Le(b) (Huang *et al*., 2012, Ma *et al*., 2015).

This study supports the concept that susceptibility to noro- and rotavirus infection is affected by the HBGA phenotype of a person. This could in future help with better understanding the viral infection mechanisms and in turn help with vaccine development and treatment.

Chapter 5 : Next Generation Sequencing

5.1 Introduction

Recent years have introduced the scientific community to new ways to cultivate norovirus, although there is still a lack of affordable *in vitro* culture systems. Therefore norovirus whole/partial genome sequencing is still mainly achieved directly from stool samples. Next-generation sequencing (NGS) has become a more frequent method for sequencing, especially in high income countries displacing Sanger sequencing as the preferred method. This new method of sequencing has allowed for new research and diagnostic applications, already revolutionising genomic science (Vinjé, 2015, Pang, 2017). Applications vary from the discovery of novel viruses, characterisation of viral communities, and detection of variability in the genome of viruses (Vinjé, 2015), to whole-genome analysis of individual microbiomes, giving a broader picture in terms of co-infection and gene-environment interactions (Schwarzer *et al.*, 2018).

Various NGS platforms, such as Illumina, Ion torrent and SOLid have become available in recent years, all with variations in sequencing mechanisms, read lengths, accuracy, run time and cost, with a few examples shown in Table 5.1

Table 5.1: A comparison of available NGS platforms.

Platform	Chemistry	Instrument cost	Read length (bp)	Run time	Disadvantage
Ion torrent PGM[a]	Proton Detection	$80 000	200-400	4 - 7h	High error rate
PacBio RS[a]	Single Molecule Sequencing	$695 000	Average 1500 bases	2h	No paired ends, expensive
Illumina MiSeq[b]	Bridge amplification	$128 000	Up to 2x300	3 days	Long run time
Illumina HiSeq[b]	Bridge amplification	$654 000	2x125	6 days	Short reads, long run time, expensive

[a] (Quail *et al.*, 2012)

[b] (Liu *et al.*, 2012a)

The method used by the Illumina platform is sequencing by synthesis chemistry. The Illumina NGS workflow can be explained in five easy steps as shown in Figure 5.1.

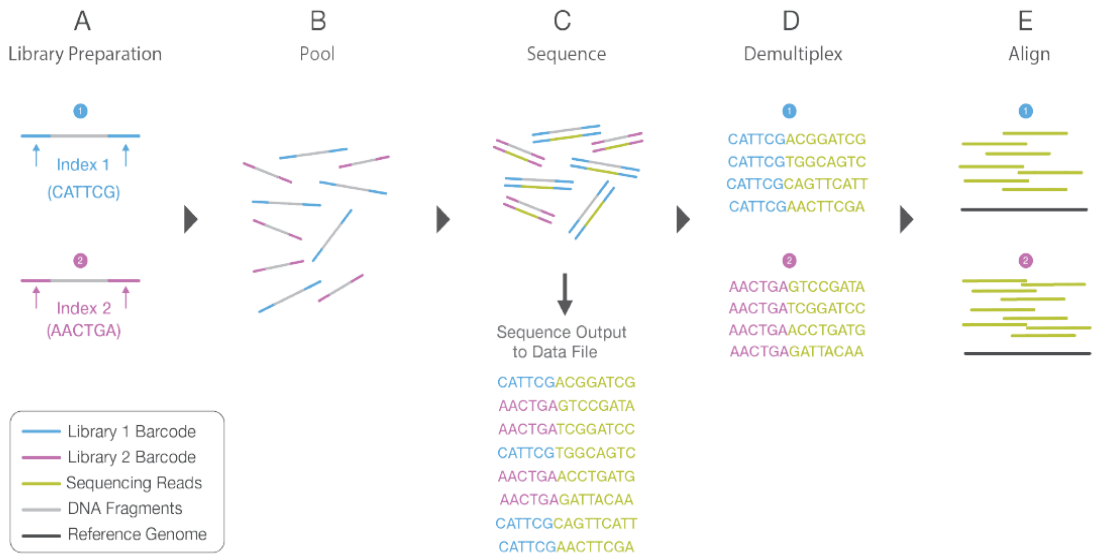


Figure 5.1: Library Multiplexing Overview: (A) Unique index sequences are added to two different libraries during library preparation. (B) Libraries are pooled together and loaded into the same flow cell lane. (C) Libraries are sequenced together during a single instrument run. All sequences are exported to a single output file. (D) A demultiplexing algorithm sorts the reads into different files according to their indexes. (E) Each set of reads is aligned to the appropriate reference sequence. Reproduced from Illumina, 2015.

Along with different platforms, different methods of preparing specimens for NGS are available, all with their advantages and disadvantages, with an example given in Table 5.2.

Table 5.2: Advantages and disadvantages of sequence-dependent and sequence independent genome amplification.

	Sequence-dependent partial genome amplification	**Sequence independent cDNA amplification**
Advantage	Targeted genome amplification, making any other non-targeted sequences insignificant.	Can be used to sequence novel strains for which no primers are available. Allows for universal pathogen detection, regardless of the type of microbe.
Disadvantage	Only genotypes with known sequences can be amplified.	Large diversification of specimen sequences obtained can lead to low amount of reads of specific sequences required.

It has been hypothesised that immunocompromised individuals form part of a reservoir for novel norovirus variants and recombinants (Karst and Baric, 2015). A greater diversity of norovirus and rotavirus infection has also been observed in immunocompromised patients (Parashar *et al*., 2009, Bok and Green, 2012). This is likely due to a higher frequency of persistent diarrhoea in HI and HEU when compared to HU patients, as was seen in a study done in Zaire (Thea *et al.*, 1993). Very few studies have been done which focus on HIV-infected immunocompromised patients co-infected with norovirus. One such study reported prolonged norovirus excretion as well as observed viral evolution (Wingfield *et al.*, 2010), whereas another reported prolonged hospitalisation and increased mortality when compared to HU children (Page *et al.*, 2017). This study included next-generation sequencing on a GII.4 norovirus-positive specimen from one HI, one HEU and one HU child respectively, to assess whether there is a difference in the diversity (quasi species) of the noroviruses detected in HI, HEU and HU children.

5.2 Materials and Methods

5.2.1 Specimen selection

One GII.4 norovirus-positive specimen from each of the HI-, HEU- and HU groups was selected for NGS analysis. All specimens were required to have a Ct below 30, with sufficient stool (≥ 800 µL 10% stool suspension) for processing. Next-generation sequencing of RNA extracted from stool specimens is a sequence-independent method which is ideal for norovirus strains where there is little sequence information available.

5.2.2 Nucleic acid extraction

Ten percent stool suspensions (400 µL) were used in duplicate for viral extraction using the MagMAX Viral RNA Isolation kit (Life Technologies corp., Carlsbad, CA) according to manufacturer's instructions. The two eluted RNA samples (50 µL each) were then pooled together to a total volume of 100 µL for further concentration (see section 5.2.3.3).

5.2.3 cDNA synthesis and amplification

Two different methods were used to try and obtain enough PCR products to undergo NGS, namely sequence-dependent partial genome amplification and sequence independent cDNA amplification (Table 5.2.)

Method 1: Sequence-dependent genome amplification

5.2.3.1 cDNA synthesis

Complimentary DNA was synthesised using Superscript® IV RT (Invitrogen, Carlsbad, CA), extracted RNA and a tagged poly-T primer (Seegene Inc., Seoul, South Korea) in a total volume of 20 µl. For priming, 2.5 µM tagged poly-T primers, 25-50 µg RNA and UltraPure™ Diethyl pyrocarbonate (DEPC)-treated water were added to a 1.5 mL microcentrifuge tube and heated to 65°C for 10 min, followed by placing on ice for 2 min. After briefly centrifuging the tube, 1X First-Strand Reaction Buffer, 10 mM DTT and 500 µM dNTPs (Thermo Scientific), 200 U of SuperScript IV Reverse Transcriptase and 40 U RNaseOUT™ recombinant RNase inhibitor was added. The reverse transcription conditions were as follows: incubation (25°C, 10 min), RT (65°C, 30 min), RT inactivation (80°C, 10 min), and the reactions were cooled to 4°C. The cDNA was used immediately after synthesis and the remainder was stored at - 20°C.

5.2.3.2 Amplification of the complete genome

Amplification of the 7.6 kb norovirus genome was attempted in three overlapping segments (Figure 5.2), using published and in-house primers designed previously (Table 5.3). Segment 3 comprised of two forward and one reverse primer for semi-nested amplification.

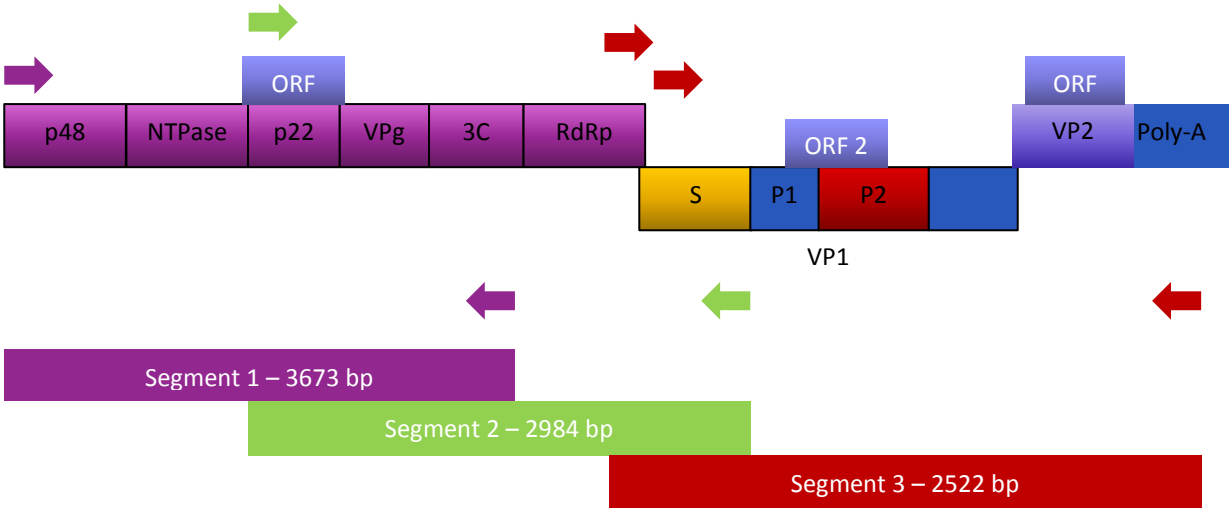


Figure 5.2: A representation of the three overlapping segments amplified and the primer positions (indicated with arrows). *Figure not made to scale.

Table 5.3: Specific primers used for the amplification of the complete norovirus genome.

Primer name	Forward/ reverse	Sequence (5'-3')	Location
G2Seg1-F[a]	Forward	ATGAAGATGGCSTCTAACGAC	1-21
G2Seg1-R[a]	Reverse	TTCATAGGTGCCAGGTGGGAG	3673-3693
G2Seg2-F[a]	Forward	TGCATTTGTCACCACGCGCAT	2466-2486
G2Seg2-R[a]	Reverse	CAGTTGGGAAATTTGGTGGGTC	5450-5471
QNIF2[b]	Forward	ATGTTCAGRTGGATGAGRTTCTCWGA	5008-5033
G2SKF[c]	Forward	CNTGGGAGGGCGATCGCAA	5043-5060
Anchor[d]	Reverse	CTGTGAATGCTGCGACTACGAT	Poly-A tail

[a]In-house primers, [b](Loisy *et al.*, 2005), [c](Kojima *et al.*, 2002), [d]Seegene CapFishing 3'RACE primer

For attempts of complete genome amplification the SequalPrep Long PCR Kit (Invitrogen, California, United States) was used, and reaction mix (20 µL) consisted of: 1X SequalPrep™ 10X Reaction Buffer, dimethyl sulfoxide (DMSO), 0.5X–1X SequalPrep™ 10X Enhancer A, 1.8 UDNA SequalPrep™ Long Polymerase (all included in the PCR kit), dNTPs (Thermo Scientific), 0.5 µM of the forward and reverse primer respectively and nuclease-free water. The amplification protocol was as follows: enzyme activation step (2 min, 94°C), followed by 10 cycles of denaturation (10 sec, 94°C), annealing (30 sec, Primer Tm –5°C) and extension (1 min/Kb, 68°C). Another 20-30 cycles of denaturation (10 sec, 94°C), annealing (30 sec, Primer Tm –5°C) and extension (1 min/Kb +20

sec/cycle, 68°C) which was then finished off by a final extension step (5 min, 72°C) to ensure full length products. The PCR products were stored at - 20°C.

Method 2: Sequence-independent cDNA amplification

5.2.3.3 Concentration of multiple viral nucleic acid extractions

The Agencourt® RNAClean® XP beads (Beckman Coulter, Brea, CA) were used to concentrate the two viral nucleic acid extractions made in section 5.2.2 into smaller volumes. The RNAClean® XP beads were initially vortexed briefly to ensure proper suspension. A total of 1.8 µL beads were added per 1 µL of the sample, along with 270 µL 100% isopropanol, and the bead-isopropanol mixture was gently mixed before being added (and mixed by pipetting the mixture up and down) into the nucleic acid, and incubated at room temperature for five min. This mixture was then placed on a magnetic rack to separate the beads from the supernatant, containing possible contaminants. The supernatant was removed, and the beads were washed twice with 600 µL 80% ethanol, to ensure removal of contaminants. The purified RNA was the eluted to a total volume of 30 µL by adding nuclease-free water (Promega Corp).

5.2.3.4 Double-stranded cDNA synthesis using the Maxima H Minus Double-Stranded cDNA synthesis kit

Double-stranded cDNA was synthesised using the Maxima H Minus Double-Stranded cDNA synthesis kit (Thermo Fischer). First, a Poly(T) primer (100 pmol/µL) and 13 µL of RNA were mixed, incubated at 65°C for 5 min and chilled on ice. The 1st strand was synthesised by adding the 1X First Strand Reaction Mix, First Strand Enzyme Mix and the RNA-primer mixture together and the mixture was incubated at 50°C for 30 min. The reaction was terminated by heating at 85°C for 5 min and then placed on ice.

The 2nd strand was synthesised by adding the 1st strand cDNA synthesis reaction mixture, nuclease-free water, 5X Second Strand Reaction Mix and Second Strand Enzyme Mix together and incubating at 16°C for 60 min. The reaction was stopped by adding EDTA (0.5

M). RNA was removed by adding RNase I (10 U/µL) and incubating at room temperature for 5 min.

5.2.3.5 cDNA Purification

The purification of double-stranded cDNA was performed using the Agencourt AMPure XP beads (Beckman Coulter). Briefly, the AMPure XP beads were vortexed to ensure that they were properly suspended and 1.8 µL AMPure XP beads was added per 1.0 µL of cDNA. The mixture was mixed by pipetting and incubated at room temperature for 5 min. The mixture was then placed on a magnetic rack to separate the beads from the supernatant containing contaminants, the supernatant was discarded and the beads were washed twice with 70% ethanol to further remove contaminants. Finally, the purified double-stranded cDNA was eluted from the beads using 30 µL nuclease-free water (Promega Corp.).

5.2.4. cDNA analysis

For both methods, cDNA analysis was performed in the same way.

The double-stranded cDNA products were separated on a 1% LE agarose gel (Cleaver Scientific Ltd., Warwickshire, United Kingdom). The gel was prepared using 1-1.5 g SeaKem® LE Agarose (FMC Corporation, Philadelphia, PA) in a final volume of 100- 150 mL 1 x TAE buffer. The solution was then boiled and cooled before adding 5-7.5 µL of 10 mg/mL EtBr (Thermo Scientific). Two microliters of LoadingDye (Thermo Scientific) was mixed with 2 µL purified double-stranded cDNA and the samples were loaded on the gel. A 1 kb Plus DNA Ladder (Thermo Scientific) was used as reference to determine the size of the amplification products. The voltage was set between 90 and 120 volts depending on the size of the gel. The Biorad Gel DocTM XR+ System (Bio-Rad Laboratories, Hercules, CA) was used to visualise the gel after electrophoresis.

5.2.5 Next generation sequencing

Sequence dependent amplification proved unsuccessful, so next generation sequencing was performed on the amplicons made form method 2. All three cDNA products were sent to the

National Institute for Communicable Diseases (NICD) to undergo NGS at the NICD Sequencing core facility with the Illumina MiSeq Platform (Illumina, Hayward, CA). Sequencing was performed at a coverage of 2 million reads per sample. The procedure that was followed by the NICD Sequencing Core facility is explained in short in Figure 5.3.

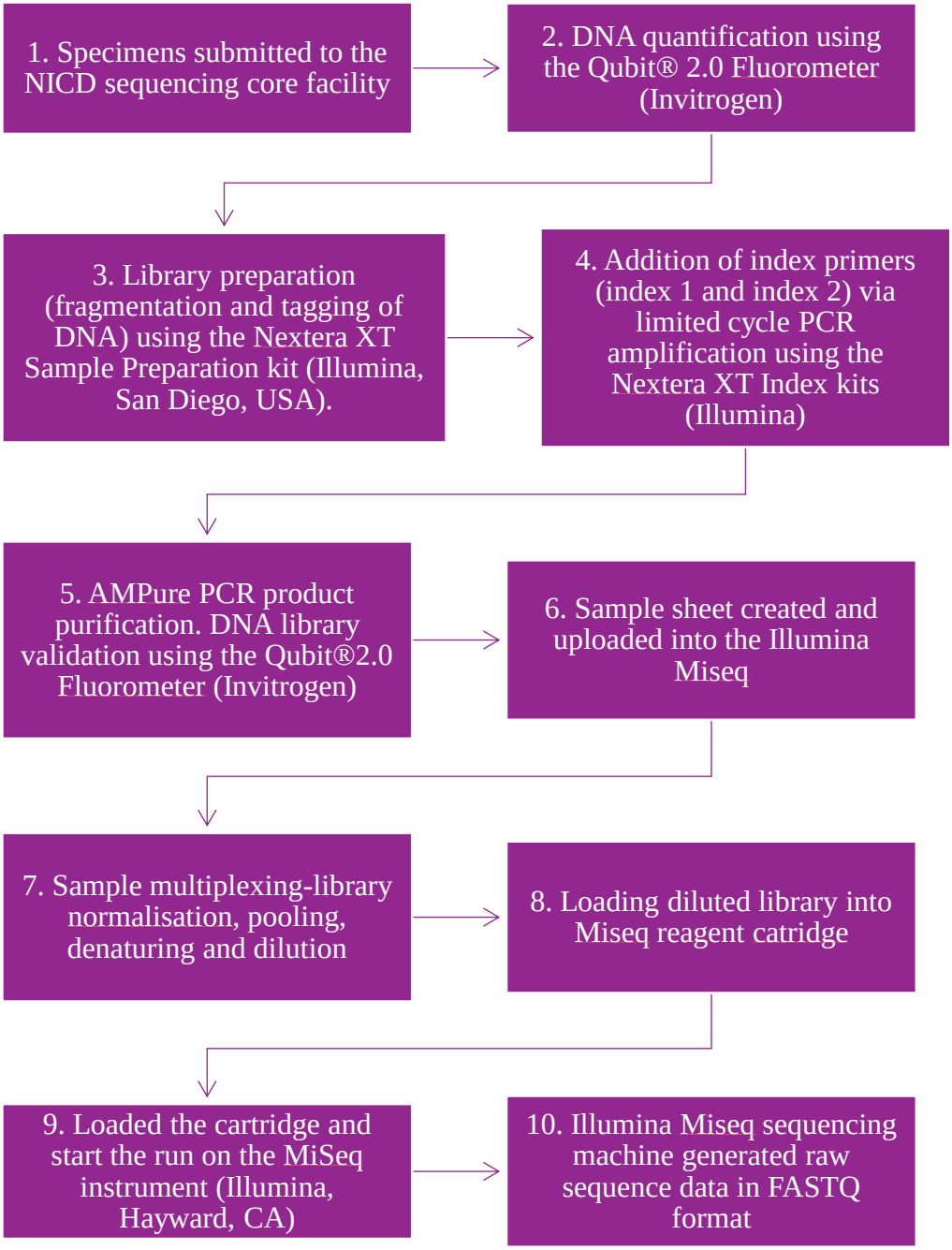


Figure 5.3: Flow diagram depicting the process of NGS followed at the NICD Sequencing Core facility, from sample arrival to data output.

5.2.6 Data analysis

Next-generation sequencing data for the three specimens were received in FASTQ format and analysed as described in Figure 5.4. The CLC Genomics Workbench version 20.0 (Qiagen) was used to perform additional read quality assessment by determining the read length and quality distributions.

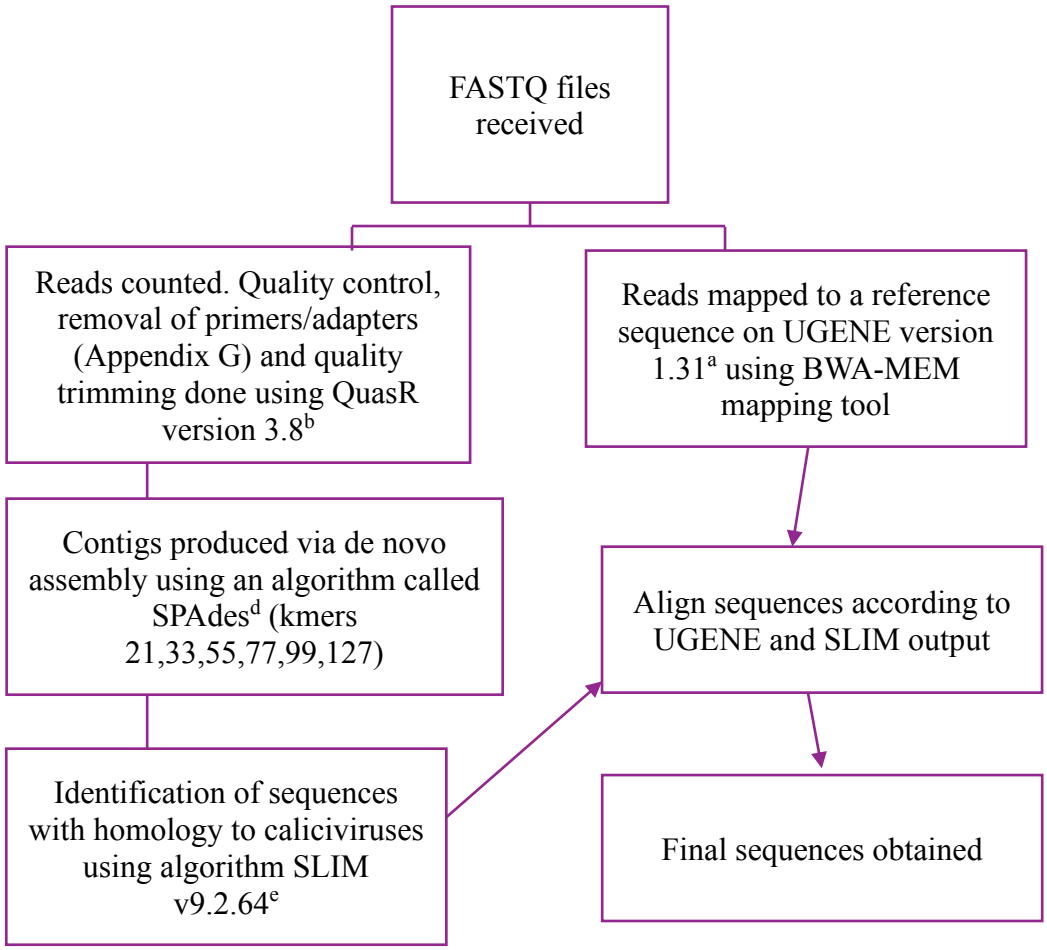


Figure 5.4: Flow diagram of data analysis. (Okonechnikov *et al.*, 2012)[a], (Gaidatzis *et al.*, 2014)[b], (https://sourceforge.net/projects/bbmap/)[c], (Bankevich *et al.*, 2012)[d], (Cotten *et al.*, 2014)[e], (Kroneman *et al.*, 2011)[f], (https://blast.ncbi.nlm.nih.gov/Blast.cgi)[g].

Phylogenetic analysis

BioEdit Sequence Alignment Editor software (V.7.0.9.0) (Hall, 2005) was used for the assembly of norovirus contigs, with the alignment to reference sequences obtained from GenBank as well as the editing of the sequences. Sequences obtained were aligned to reference sequences using MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/) and phylogenetic analysis was performed in MEGA 6. The neighbour-joining clustering method was chosen for the combination of phylogenetic trees, validated by 1000 bootstrap replicates and the evolutionary distances were computed using the Kimura 2-parameter model (Kimura, 1980, Felsenstein, 1985, Saitou and Nei, 1987).

5.3 Results

Method 1: Sequence-dependent partial genome amplification

Sequence-dependent amplification of the three segments of the norovirus genome proved unsuccessful. All further results were obtained via sequence-independent cDNA amplification.

Method 2: Sequence-independent cDNA amplification

In total, three norovirus GII.4 positive specimens underwent NGS. These specimens were selected as there was adequate stool available and all three specimens had Ct-values of below 30 based on real time RT-PCR. Table 5.4 is a summary of the specimens used for analysis.

Table 5.4: Summary of GII.4 strains selected for NGS attempts.

Year	Specimen	Ct-value	HIV status
2016	NS0030	23.24	HEU
2016	NS0063	23.53	HI
2017	NS0157	23.21	HU

Based on personal communication with the staff at the NICD Sequencing core facility, a smear of product had to be observed on the gel to indicate the presence of sufficient

concentrations of cDNA. A smear could be observed for all three specimens (Figure 5.5). The cDNA was then sent to the NICD for NGS analysis.

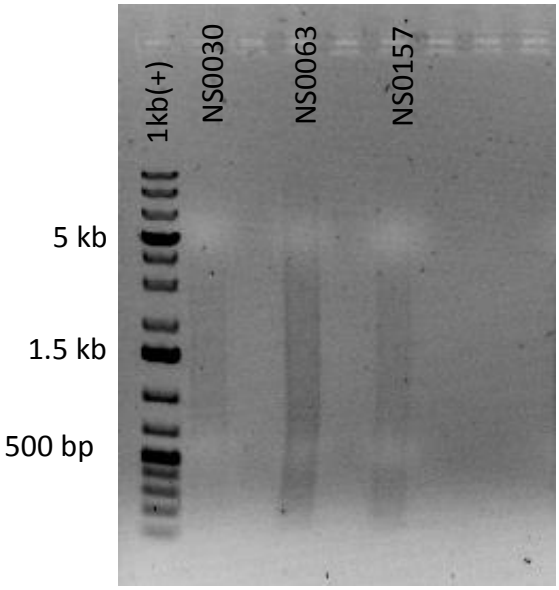


Figure 5.5: Gel electrophoresis analysis of the double-stranded cDNA generated for three GII.4 viruses originating from HI, HEU and HU children. A smear was observed for all three specimens.

5.3.1 Data analysis

Between 282927 and 652861 reads were obtained per sample (Table 5.5). A large proportion of the reads were of low quality as indicated by the drop in reads after the quality control step in QuasR (column 4, Table 5.5).

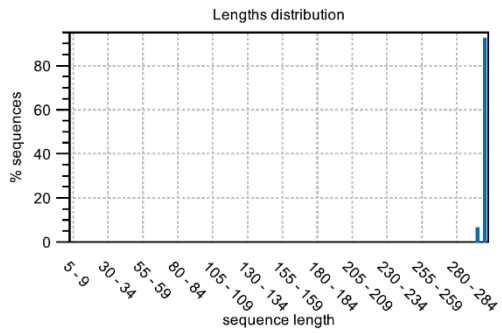
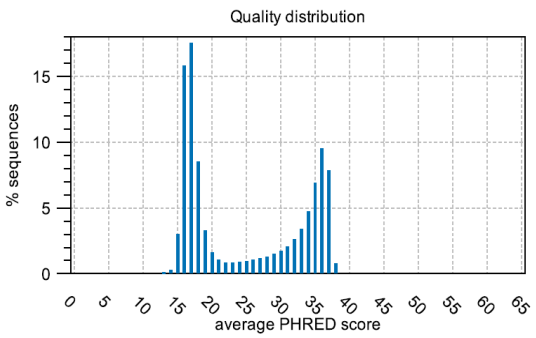
Table 5.5: Summary of the number of reads throughout the data quality control and contig assembly process.

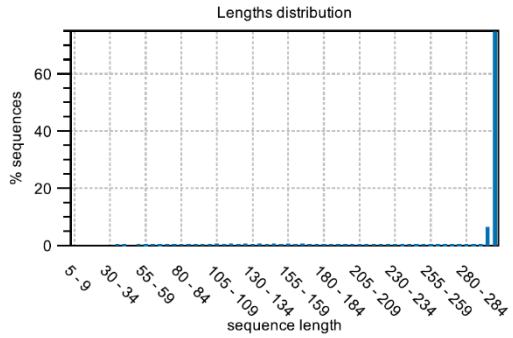
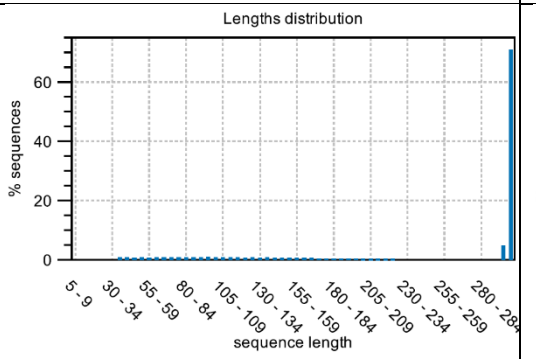
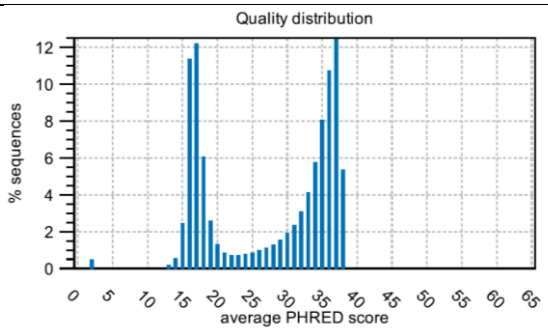
Specimen	Ct-value	Reads before analysis	Reads after quality control	Contigs after analysis (SPAdes)	Range of contig lengths (bp)	Number of norovirus contigs (SLIM)	Range of norovirus contig lengths	BLAST top hit
NS0030 (HEU)	23.24	282927	6895	142	128-15666	1	470 bp	South Africa 2012 (KP784697.1)
NS0063 (HI)	23.51	652861	371976	1494	128-437860	0	-	-
NS0157 (HU)	23.21	430270	38874	2737^ 3197*	128-34104	3	526-762 526-762	USA 2015 (KX354126.1)

^ Quality control parameters: -L 175 –q -l 175 –m 30 *Quality control parameters: -L 150 –q -l 150 –m 25

To confirm this observation the reads were also analysed on the CLC Genomics Workbench 20.0 (Table 5.6). The read and length distribution for each sample showed that although the reads were of adequate length overall, the read quality was suboptimal, with a large proportion of reads with quality scores of between 25 and 15.

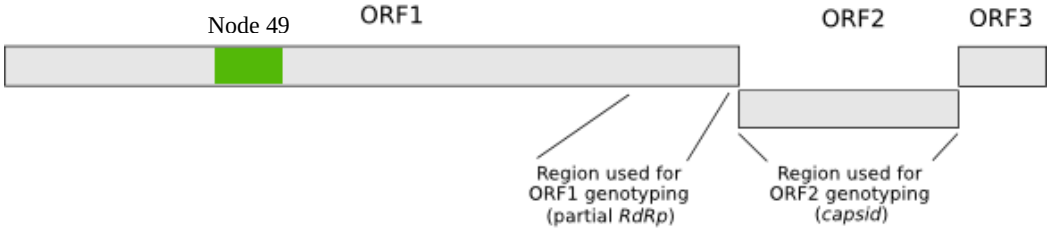
Table 5.6: Summary of results obtained from quality assessment in CLC Genomics Workbench.

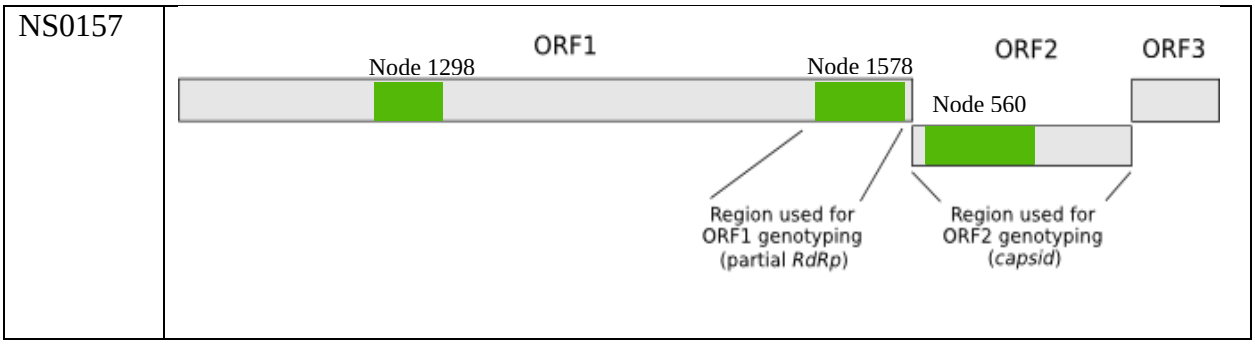
Specimen	Length distribution	Quality distribution
NS0030		

NS0063		
NS0157		

Between 142 and 3197 contigs were generated after assembly, but only four of these (1 in NS0030 and 3 in NS0157) were norovirus specific (Table 5.5). The positions of these contigs on the norovirus genome are shown in Table 5.7. The low quality of the reads did not allow assembly of norovirus genomes for any of the three samples and thus the levels of norovirus GII.4 diversity in the different children could not be compared.

Table 5.7: Contig positions as observed from the norovirus genotyping tool (https://www.rivm.nl/mpf/typingtool/norovirus/).

Specimen	Contig position
NS0030	



Phylogenetic analysis was performed on all sequenced segments obtained, allowing for phylogenetic trees to be constructed for each segment, with reference strains obtained from BLAST (Figure 5.6-5.8).

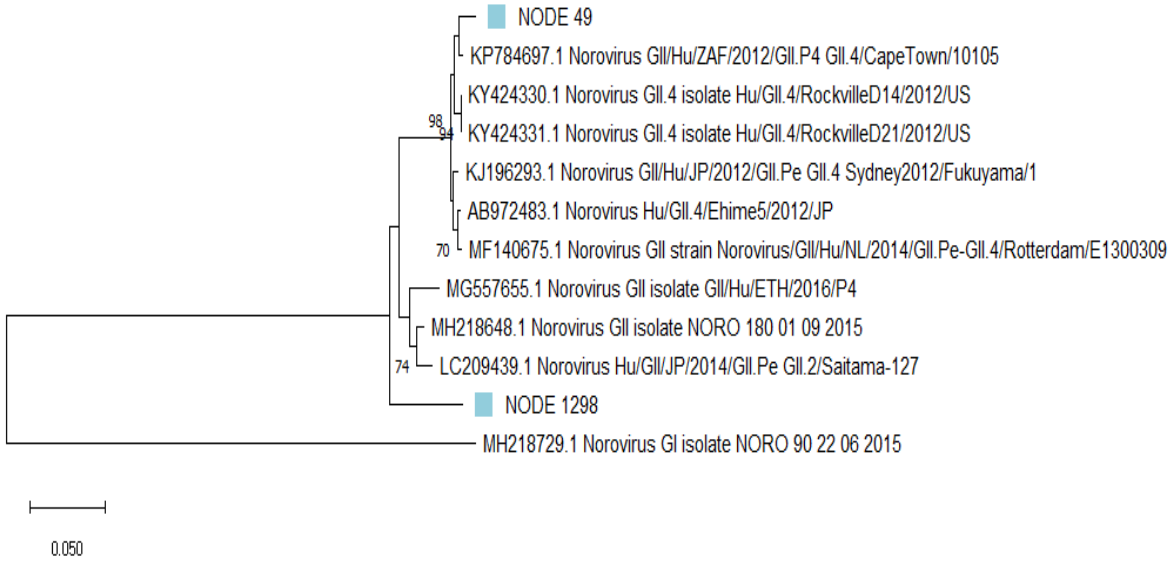


Figure 5.6: Neighbour joining phylogenetic analysis of a partial segment of ORF1 (≈425 bp) of both NS0030 and NS0157 norovirus strains used for NGS. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region.

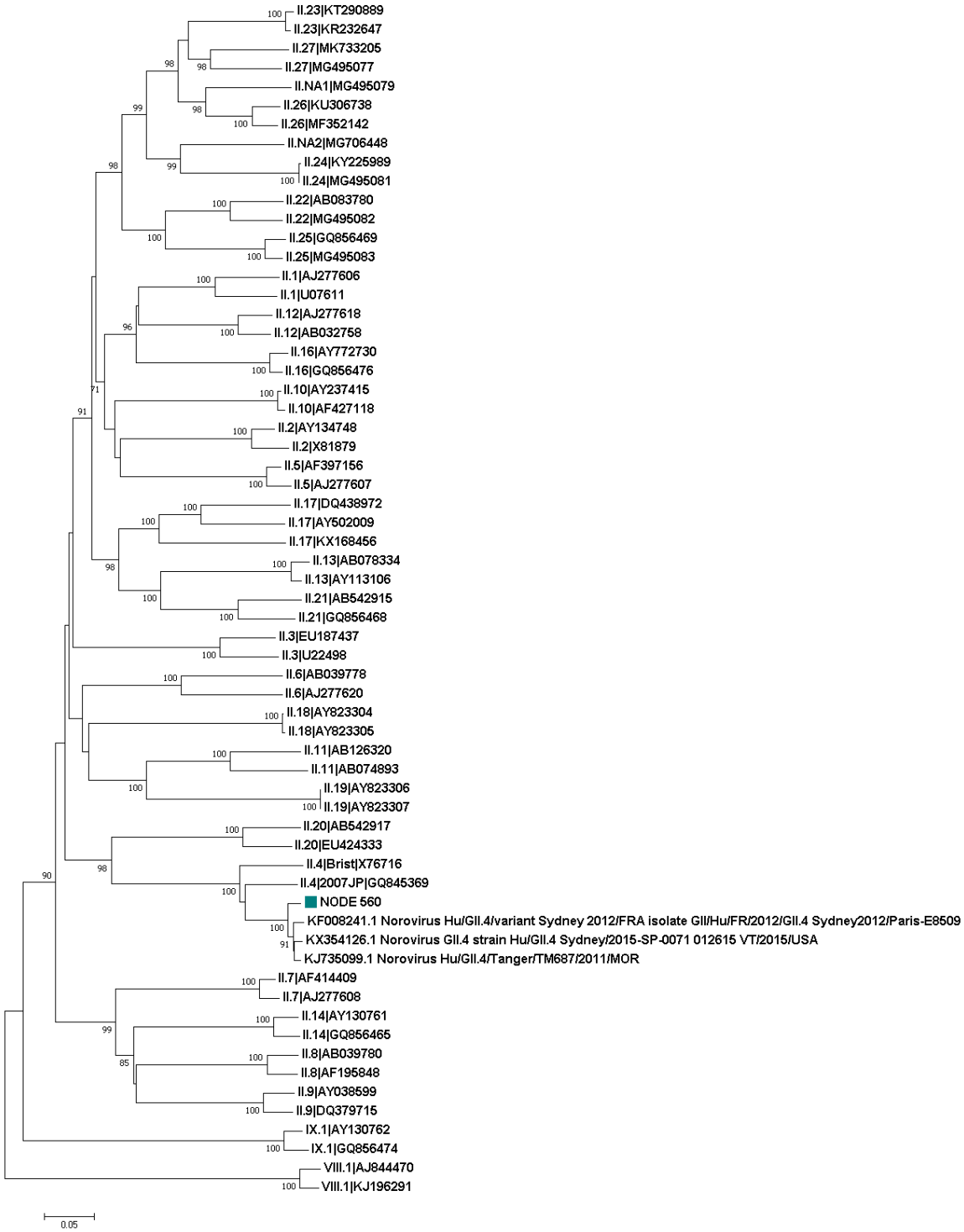


Figure 5.7: Neighbour joining phylogenetic analysis of a partial segment of the capsid region (≈760 bp) of NS0157 norovirus strain used for NGS. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region.

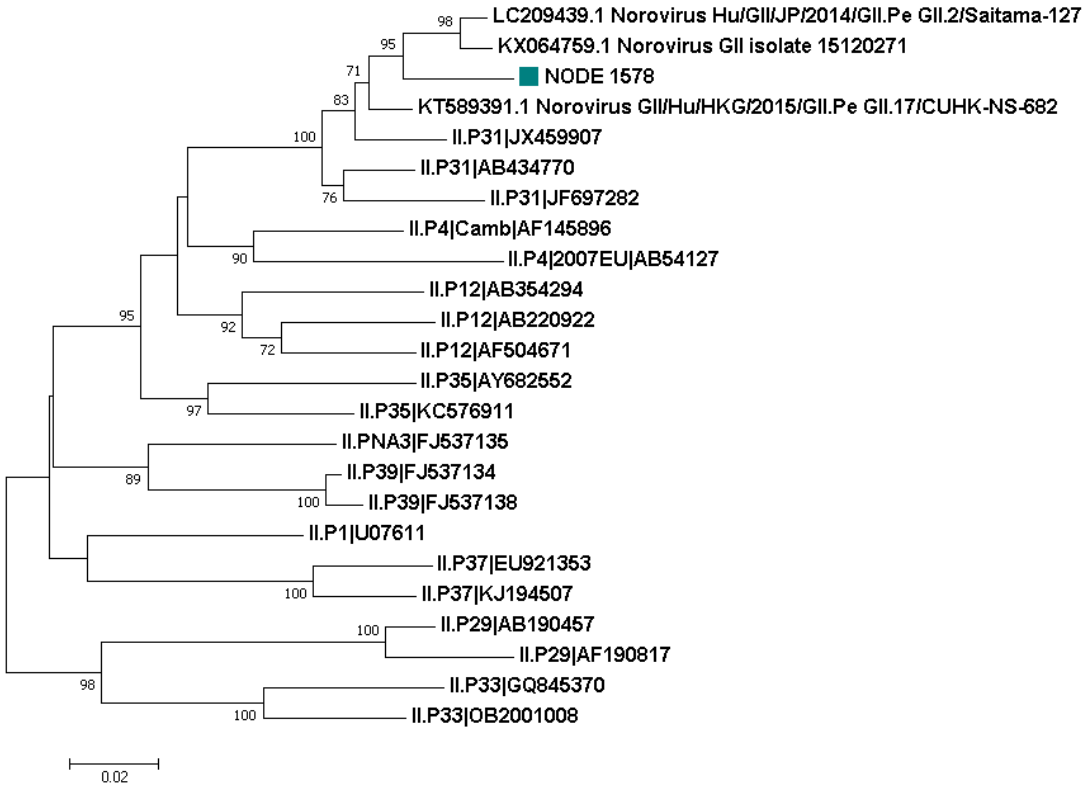


Figure 5.8: Neighbour joining phylogenetic analysis of a partial segment of RdRp region (≈520 bp) of NS0157 norovirus strain used for NGS. Closely related strains from GenBank indicated by accession numbers served as reference. The percentage of replicate trees (cut off value ≥70%) in which the associated strains clustered together in the bootstrap test (1000 replicates) is shown. The scale bar indicates 0.05 nucleotide differences per site, over the indicated region.

5.4 Discussion

Next generation sequencing has brought a whole new level to investigating viruses and viral communities, with multiple advantages (Luciani *et al.*, 2012, Bavelaar *et al.*, 2015). Unfortunately, it is also an expensive, time consuming and difficult process, which can fail depending on the quality of the input material as well as the virus concentration in the specimen. This study serves as an example of problems that can be encountered, as although many bacterial reads were obtained, very few norovirus specific reads were generated, and complete or even partial genomes could not be assembled from the data. Although the aim was to obtain two million reads per sample, the maximum number of reads obtained in sample NS0063 was ~ 650 000. Therefor a combination of low-quality reads and low numbers of reads prevented determination of the complete genomes of the three GII.4

strains. The lack of reads for analysis could be due to multiple reasons, including the age of the stool specimens (between two and three years old), as fresh stool would be optimal for analysis (Vermeulen, 2019), insufficient sequencing depth and suboptimal preparation of the ds cDNA template. Due to time and funding constraints, the optimisation and resubmission of ds cDNA could not be performed but should be considered for future study. Phylogenetic trees could be constructed for segments from 2/3 specimens, with only one tree containing both specimen strains (Figure 5.6). This tree showed a large difference between the two GII.4 norovirus strains in the ORF1 region, indicating diversity between the two strains. Unfortunately, this was the only segment that could be amplified for both specimen strains. Due to the lack of adequate sequence data, no observation could be made as to how HIV exposure affects gastroenteritis virus infections in children in terms of pathogen diversity. One positive note is that a rarely amplified segment of the norovirus genome could be observed for both sample NS0030 and NS0157, which could help with prospective primer design and primer walking to analyse ORF1 of that genome.

In future, fresh specimens should be considered, with increased sequencing coverage for sequence-independent methods. Another consideration would be to initially pre-treat the stool specimens (e.g. filtration) to remove any bacterial or host nucleic acid, to reduce non-viral background sequences. In addition, the diversity of noroviruses in individuals with chronic versus acute gastroenteritis may be considered, rather than only studying acute infections in HIV-infected individuals.

Chapter 6 : Concluding discussion

Viral gastroenteritis is still a major cause of illness, in the face of improvements to provision of safe water and sanitation, oral rehydration therapy, the rotavirus vaccine and reductions in hospitalisations and mortality (Kotloff *et al.*, 2013, Liu *et al.*, 2016b, UNICEF, 2019). In high-income countries, the rotavirus vaccine has led to a rapid decrease in hospitalisations due to rotavirus, with norovirus taking the lead as the most prevalent virus found to cause hospitalisations in children under the age of five due to viral gastroenteritis (Tate *et al*., 2012, Griffin, 2013, Burnett *et al*., 2017). In low- and middle-income countries such as South Africa, however, despite routine rotavirus vaccination, rotavirus is still the leading cause of viral gastroenteritis hospitalisations (Enquist *et al*., 2015, Liu *et al*., 2016a, Operario *et al.*, 2017), as was illustrated by this study. Possible reasons for this include higher burden of pathogen exposure in early life, with repeated exposures to multiple pathogens when compared to high income countries, and the effect that this might have on individual microbiomes and susceptibility. It has also been suggested that inherited host factors such as secretor status have an effect on the susceptibility to different enteric infections in different populations (Fumagalli *et al*., 2009). As the reasons for this discrepancy are not yet clear cut, studies to further understand the epidemiology and host susceptibility of these viruses are crucial to help with the fight against these severe viral infections.

This study combined virus epidemiology with specific host factors in hopes of determining specific links to severity and susceptibility of these viral infections. One of the main objectives was to determine which pathogens caused gastroenteritis hospitalisations and to determine what the diversity of rotavirus, norovirus and sapovirus pathogens circulating in this environment was, as well as to investigate norovirus asymptomatic infection. Detection of gastroenteritis viruses was done by using the Allplex™ Gastrointestinal Virus Panel (Seegene) and multiplex qPCR. The kit was shown to be highly effective, with accurate results obtained. Almost half of the participants (96/205; 47%) were infected by enteric viruses, although high, this is a common ratio observed in previous studies (Sdiri-Loulizi *et al.*, 2008, Iturriza-Gómara and Cunliffe, 2020). Rotavirus predominated (46/205), followed by norovirus (32/205), adenovirus-F (15/205), sapovirus (9/205) and astrovirus (3/205).

This highlights the large spread and impact of these viruses in terms of gastroenteritis hospitalisations, with rotavirus still the most predominant virus detected in infected children (48%; 46/96), despite 83% (38/46) of the rotavirus infected children being fully vaccinated against the virus. Rotavirus infection was however observed more frequently with mild to moderate symptoms, suggesting that the vaccine decreased the severity of gastroenteritis but not the incidence of hospitalisations. This is in line with the characteristic of the vaccine to prevent severe rotavirus gastroenteritis, but not necessarily infection (Groome *et al.*, 2014).

Bacteria and parasite infections were determined by the NHLS through MC&S for 183/205 children, with the remaining specimens being undetermined due to rejection of specimen (due to leakage or insufficient amounts) or test failure. Testing resulted in 10 enteric bacterial infections (4 *E. coli*, 4 salmonella and 2 shigella infections), and five parasite infections (1 *Ascaris lubricoides*, 2 cryptosporidium, 1 entamoeba and 1 *Giardia lamblia* infection). In total, 52% (106/205) of the study population was found to be infected with at least one gastroenteritis causing pathogen, with 3 virus and bacteria-, 2 virus and parasite- and 9 viral co-infections observed. No bacteria and parasite co-infections were observed. Various genotypes of noro-, rota- and sapovirus were observed, indicating a large diversity of strains circulating in the population. Norovirus and rotavirus strains were similar to those circulating globally, with full reference list in Appendix D.

Only 22% (46/205) of children provided a follow up stool specimen. This low follow up rate is most likely due to the fact that the parents/caregivers had to collect a soiled diaper and bring it to the hospital, and transport may have been problematic for many of them. All follow up specimens were screened with a singleplex assay for norovirus GI and GII specifically. A total of 28% (14/46) of the children were infected with norovirus, with all the children being asymptomatic at the time of stool collection. These infections were considered asymptomatic, although one cannot completely rule out virus shedding after symptomatic infection since norovirus shedding has been described to last up to 60 days in some cases (Teunis *et al.*, 2015) and the follow up samples were collected after 6 weeks. Comparison of genotypes detected in asymptomatic and symptomatic cases showed different viruses in the two groups except for GI.3 and GII.4 which were present in both. Norovirus GII.13 predominated in the follow up norovirus positive specimens. This is contradictory to other studies done on asymptomatic norovirus infection in Africa, where

norovirus GII.13 is either not present at all (Bucardo *et al.*, 2010), or detected at very low frequency (Hungerford *et al.*, 2020). The GII.13 strains clustered with other South African strains, observed at low frequencies (Mans *et al.*, 2016) as early as 2011 (KC495668.1) in children with diarrhoea. Only one norovirus strain sequence (NS0173) was identical in the original and follow up specimen obtained, indicating prolonged shedding. The high burden of asymptomatic norovirus infection may be an important source of transmission in this population, with norovirus circulation being largely underrepresented.

Another main objective of the study was to determine the secretor status of all 205 children, and to determine if secretor status had an impact on the susceptibility to any enteric virus infections as secretor status is known to affect the susceptibility to noro- and rotavirus infection. The secretor status of all 205 children was determined with the TaqMan SNP assay kit. Blood was mainly used to determine the FUT2 genotype, but in cases where blood was not available (n=15), stool specimens were used. This study showed that stool specimens of infants contain high enough amounts of the child's DNA to determine the FUT2 genotype. This could prove to be a useful method, as stool is the most frequently collected specimen for diagnosis of diarrhoeal disease and collection is less intrusive than the drawing of blood, especially in infants. After genotyping of all 205 specimens, a 71%:29% ratio of secretors and non-secretors were observed, which is similar to observations in other studies in the African population (Moores and Brain, 1968, Liu *et al.*, 1998). No link was observed between secretor status and adeno- astro or sapovirus. This correlated with previous studies (Matussek *et al.*, 2015, Oka *et al.*, 2015, Colston *et al.*, 2019). Rotavirus ($p<0.01$) and norovirus GII.4 ($p=0.035$), however, were found to be significantly associated with secretors, as was observed in previous studies (Frenck *et al.*, 2012, Nordgren and Svensson, 2019). No statistically significant preference could be observed for any virus genotype and non-secretors, although rotavirus P[6] strains were more prevalent in this group. This correlates with previous studies as well (Liu *et al.*, 2012b, Nordgren *et al.*, 2014, Pollock *et al.*, 2018). As African countries have a higher ratio of P[6] strains circulating (Todd *et al.*, 2010, Nyaga *et al.*, 2018), and a higher ratio of non-secretors when compared to European and American countries (Liu *et al.*, 1998), this may be a factor to consider when looking at vaccinations in Africa specifically (Patel *et al.*, 2009).

A third main objective was to determine the HBGA phenotypes of the 205 children and compare these phenotypes with the FUT2 genotypes and virus infections observed in each

individual. Various combinations of HBGAs were determined for secretors and non-secretors, with various discrepancies observed when comparing the secretor status with expected HBGAs. This could be due to multiple factors, all mentioned in chapter 4, but one thing that is certain, is that characterisation of saliva HBGA is not as straight forward as expected, and many factors should be considered when initiating a study that includes these tests. When comparing the HBGA phenotypes with the observed virus infection, there was no association between adeno-/astro-/sapovirus and specific HBGA types. This is to be expected, as stated above, because secretor status has no effect on susceptibility to these viruses. Therefor comparison of HBGA phenotypes was mainly done with the noro- and rotavirus infected population. Statistical analysis through contingency tables and double sided Fischer's exact testing was performed to determine if any HBGA combinations were linked to increased susceptibility to noro- and rotavirus due to the large diversity of HBGA combinations. In terms of norovirus infection, HBGA type Le(b) on its own was found to be statistically linked to an increased susceptibility of infection ($p=0.036$). Individuals with the combination of A, B, and Le(b) ($p<0.01$) A, H1, B, Le(b) ($p<0.05$), B, Le(b) ($p=0.023$) or H1, B, Le(a/b) ($p<0.05$) were seen to be more susceptible to rotavirus infection. The study then analysed specific genotypes of noro-and rotavirus, to see if different virus genotypes are linked to different HBGA phenotypes. The norovirus sample size was too small to make any definitive conclusions, but rotavirus P[4] was statistically linked to HBGA phenotypes A, B, Le(b) ($p<0.01$) and B, Le(b) ($p<0.01$). P[6] was linked to HBGA type A only ($p<0.01$) or to no HBGAs present in there saliva ($p=0.018$). Lastly P[8] was associated with A, H1, B ($p<0.05$) and A, H1, B, Le(b) ($p<0.01$). These findings are in agreement with other studies that state that P[4] and P[8] bind to Le(b) and H1, whereas P[6] is more commonly found in non-secretors (or in this case, children with no HBGAs present) (Liu *et al*., 2012b, Nordgren *et al*., 2014, Lee *et al*., 2018, Pollock *et al*., 2018). These findings were very interesting, but because there was only a small number of virus infected children (specifically noro- and rotavirus infected children), links to HBGAs should be interpreted with caution, and more research with larger groups of infected individuals should be considered.

The last main objective of this study was to compare the different HIV status groups (HI, HEU and HU), to determine if HIV status had an effect on the prevalence and severity of enteric viruses, and to determine if HIV exposure affected norovirus diversity within the

host. This was done by comparing the HIV status groups with susceptibility to virus infection and with severity of symptoms (as determined by the Vesikari clinical severity scoring system). These comparisons showed that HIV-status had no statistically significant effect on the prevalence and severity of enteric viruses in this study specifically, although a general trend in the increase of the number of co-infections was observed from HU, to HEU, to HI.

Next generation sequencing was also attempted to assess norovirus GII.4 diversity in different HIV exposure groups, to look at the hypothesis that immunocompromised individuals form part of a reservoir for novel norovirus variants and recombinants (Karst and Baric, 2015). Inability to amplify large regions of the GII.4 genomes precluded NGS based on amplicons. Sequence-independent cDNA sequencing was then attempted but very limited norovirus reads were obtained that did not allow construction of the complete genomes or diversity analysis. Thus, no information on the diversity of norovirus groups could be obtained, as aimed.

The overall aim of this study was to investigate the relationship between gastroenteritis virus infections and host secretor status, as well as to determine whether HIV exposure affects gastroenteritis virus infections in children in terms of pathogen diversity and severity. The aims were attained for the most part, producing results that further the understanding of the distribution of gastroenteritis viruses, the impact on HIV exposure on viral gastroenteritis and the burden of diarrhoeal disease in South Africa. This contributes to a better understanding of the global burden of gastroenteritis virus infection and illness as well as genotype diversity, dominant strains, and strain replacement patterns, which are essential for infection control and effective vaccine development. The secretor status of this study population could also give a representative of the ratios of secretors and non-secretors in South Africa. The *FUT2* secretor status ratios have only been done in limited studies in Africa, with even less data on South Africa itself, with most research performed more than 20 years ago (Moores and Brain, 1968, Liu *et al*., 1998, Ferrer-Admetlla *et al*., 2009).

The data from this study will hopefully prove useful for future research, in the continued effort to understand and decrease the burden of gastroenteritis in the future.

Limitations and future considerations

This study was performed to the best of our abilities, with all results being given as thoroughly and accurately as possible. As with any study, the research usually provides more questions than answers. This was also true for this study, with the limitations and future considerations listed below.

Limitations:

1. Small sample size. Ideally, a larger sample size would have helped to give more accurate and significant results. A total of 205 participants may be sufficient in a study where all patients are virus infected, but as only 96/205 specimens were virus positives, with only 46/205 rota-, 32/205 noro-, 15/205 adeno-, 9/205 sapo- and 3/205 astrovirus positive cases. These small ratios made statistical analysis difficult, with most of the analysis only performed on noro- and rotavirus positive patients.
2. Small HIV infected subset. In the broader spectrum of things, only identifying 10/205 HI patients should be a good thing, but as with the virus infections, this did make statistical analysis complicated, with most analysis being performed on HEU and HU patients.
3. Saliva collection. Saliva collection resulted in multiple complications. The fact that infants could not rinse their mouths before saliva collection. This led to confusing and discrepant results, each being sample specific. Feeding habits such as drinking breastmilk were recorded, but as breastmilk can affect HBGA diversity in children's sputum, ideally having a sample or an indication of the mothers HBGA phenotype could be able to help to compare and disregard contaminating HBGAs in breastfed children. Saliva volumes were also inconsistent with collection, resulting in some saliva specimens to be insufficient for testing.
4. Next generation sequencing. NGS was performed on three participants stool specimens. In theory this is a good plan, and if NGS is successful, diversity could be determined, but true diversity would have been seen in HI children with chronic infection, allowing the norovirus to mutate over time in the participant, as has been observed in previous studies. Unfortunately, even when taking this in consideration, the age of the specimens collected had a large impact on the quality of norovirus

sent for sequencing, resulting in low quality reads, of varying lengths, with no full genomes being obtained from NGS, thereby, no results could be given in terms of pathogen diversity between the three HIV subsets.

5. FUT2 genotyping with the TaqMan SNP assay kit proved unreliable, as only 52% (12/23) of the non-secretor specimens that underwent conventional PCR, were true non-secretors, this indicates that the ratio of non-secretors should be viewed with caution.

Future considerations:

1. Including a control subset, to increase the number of the sample population, and give a better indication as to which viruses are spreading asymptomatically.
2. Ensuring the same number of HI, HEU and HU patients, to give more statistically significant results. This may be the ideal setup, but realistically, with a transmission rate of only 9% for mother to child transmission (UNAIDS, 2019), a very large subset would be needed to make this possible.
3. Collect saliva at a specific time of the day, noting any medicine or food ingested in the last 2 hours before specimen collection. If the child is breastfed, it should be considered to ask the mother (if available) for a saliva specimen to compare. Adding a HBGA negative as a control, along with the conjugate control has been a consideration, but as each specimen was unique in the problems that it gave, it is unlikely that such a control will be effective.
4. For comparison of the diversity of norovirus strains between HI, HEU and HU, set up a study on children with chronic diarrhoea from each subset.
5. Following up with caregivers, to ensure that follow up stool specimens could be collected. As only 46/205 follow up specimens were obtained, little could be concluded from their results.

These considerations require substantial funding and dedicated surveillance officers, which limits their implementation but will hopefully be feasible within larger, collaborative studies.

References

AHMED, S. M., HALL, A. J., ROBINSON, A. E., VERHOEF, L., PREMKUMAR, P., PARASHAR, U. D., KOOPMANS, M. & LOPMAN, B. A. 2014. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *The Lancet Infectious Diseases,* 14**,** 725-730.

ALSVED, M., FRAENKEL, C.-J., BOHGARD, M., WIDELL, A., SÖDERLUND-STRAND, A., LANBECK, P., HOLMDAHL, T., ISAXON, C., GUDMUNDSSON, A. & MEDSTRAND, P. 2019. Sources of airborne norovirus in hospital outbreaks. *Clinical Infectious Diseases*.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-3402.

ANWARI, P., SAFI, N., PAYNE, D. C., JENNINGS, M. C., RASIKH, S., WACIQI, A. S. & PARWIZ, S. M. 2018. Rotavirus is the leading cause of hospitalizations for severe acute gastroenteritis among Afghan children< 5 years old. *Vaccine,* 36**,** 7765-7768.

ATMAR, R. L. 2010. Noroviruses: state of the art. *Food and environmental virology,* 2**,** 117-126.

ATMAR, R. L., OPEKUN, A. R., GILGER, M. A., ESTES, M. K., CRAWFORD, S. E., NEILL, F. H. & GRAHAM, D. Y. 2008. Norwalk virus shedding after experimental human infection. *Emerging Infectious Diseases,* 14**,** 1553.

AYOUNI, S., SDIRI-LOULIZI, K., DE ROUGEMONT, A., ESTIENNEY, M., AMBERT-BALAY, K., AHO, S., HAMAMI, S., AOUNI, M., NEJI-GUEDICHE, M. & POTHIER, P. 2015. Rotavirus P [8] Infections in Persons with Secretor and Nonsecretor Phenotypes, Tunisia. *Emerging Infectious Diseases,* 21**,** 2055-2058.

BARTSCH, S. M., LOPMAN, B. A., OZAWA, S., HALL, A. J. & LEE, B. Y. 2016. Global economic burden of norovirus gastroenteritis. *PloS one,* 11**,** e0151219.

BAVELAAR, H. H., RAHAMAT-LANGENDOEN, J., NIESTERS, H. G., ZOLL, J. & MELCHERS, W. J. 2015. Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. *Journal of Clinical Virology,* 72**,** 122-125.

BISHOP, R., DAVIDSON, G., HOLMES, I. & RUCK, B. 1973. Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *The Lancet,* 302**,** 1281-1283.

BLACKLOW, N. R. & GREENBERG, H. B. 1991. Viral gastroenteritis. *New England Journal of Medicine,* 325**,** 252-264.

BOK, K., ABENTE, E. J., REALPE-QUINTERO, M., MITRA, T., SOSNOVTSEV, S. V., KAPIKIAN, A. Z. & GREEN, K. Y. 2009. Evolutionary dynamics of GII. 4 noroviruses over a 34-year period. *Journal of virology,* 83**,** 11890-11901.

BOK, K. & GREEN, K. Y. 2012. Norovirus gastroenteritis in immunocompromised patients. *New England Journal of Medicine,* 367**,** 2126-2132.

BOSCH, A., GUIX, S., KRISHNA, N., MÉNDEZ, E., MONROE, S., PANTIN-JACKWOOD, M. & SCHULTZ-CHERRY, S. 2010. Nineteen new species in the genus Mamastrovirus in the Astroviridae family. ICTV.

BROWN, J. R., SHAH, D. & BREUER, J. 2016. Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015. *Journal of Clinical Virology,* 84**,** 1-6.

BUCARDO, F., NORDGREN, J., CARLSSON, B., KINDBERG, E., PANIAGUA, M., MÖLLBY, R. & SVENSSON, L. 2010. Asymptomatic norovirus infections in Nicaraguan children and its association with viral properties and histo-blood group antigens. *The Pediatric Infectious Disease Journal,* 29**,** 934-939.

BULL, R. A., TANAKA, M. M. & WHITE, P. A. 2007. Norovirus recombination. *Journal of General Virology,* 88**,** 3347-3359.

BUNDERS, M. J., VAN HAMME, J. L., JANSEN, M. H., BOER, K., KOOTSTRA, N. A. & KUIJPERS, T. W. 2014. Fetal exposure to HIV-1 alters chemokine receptor expression by CD4+ T cells and increases susceptibility to HIV-1. *Scientific Reports,* 4**,** 1-8.

BURNETT, E., JONESTELLER, C. L., TATE, J. E., YEN, C. & PARASHAR, U. D. 2017. Global impact of rotavirus vaccination on childhood hospitalizations and mortality from diarrhea. *The Journal of Infectious Diseases,* 215**,** 1666-1672.

CANNON, J. L., LOPMAN, B. A., PAYNE, D. C. & VINJÉ, J. 2019. Birth cohort studies assessing norovirus infection and immunity in young children: a review. *Clinical Infectious Diseases,* 69**,** 357-365.

CARLSSON, B., KINDBERG, E., BUESA, J., RYDELL, G. E., LIDON, M. F., MONTAVA, R., MALLOUH, R. A., GRAHN, A., RODRÍGUEZ-DÍAZ, J. & BELLIDO, J. 2009. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII. 4 Norovirus infection. *PloS one,* 4**,** e5593.

CARMONA-VICENTE, N., ALLEN, D. J., RODRÍGUEZ-DÍAZ, J., ITURRIZA-GÓMARA, M. & BUESA, J. 2016. Antibodies against Lewis antigens inhibit the binding of human norovirus GII. 4 virus-like particles to saliva but not to intestinal Caco-2 cells. *Virology Journal,* 13**,** 82.

CARROLL, K. C., BUTEL, J. & MORSE, S. 2015. *Jawetz Melnick and Adelbergs Medical Microbiology 27 E*, McGraw-Hill Education.

CARTER, J. & SAUNDERS, V. A. 2007. *Virology: principles and applications*, John Wiley & Sons.

CHANG, K.-O., SOSNOVTSEV, S. S., BELLIOT, G., WANG, Q., SAIF, L. J. & GREEN, K. Y. 2005. Reverse genetics system for porcine enteric calicivirus, a prototype sapovirus in the Caliciviridae. *Journal of Virology,* 79**,** 1409-1416.

CHEUNG, S. K., KWOK, K., ZHANG, L.-Y., MOHAMMAD, K. N., LUI, G. C., LEE, N., NELSON, E. A. S., LAI, R. W., LEUNG, T. F. & CHAN, P. K. 2019. Higher viral load of emerging norovirus GII. P16-GII. 2 than pandemic GII. 4 and epidemic GII. 17, Hong Kong, China. *Emerging Infectious Diseases,* 25**,** 119.

CHHABRA, P., ASWATH, K., COLLINS, N., AHMED, T., OLÓRTEGUI, M. P., KOSEK, M., CEBELINSKI, E., COOPER, P. J., BUCARDO, F. & LOPEZ, M. R. 2018. Near-complete genome sequences of several new norovirus genogroup II genotypes. *Genome Announcements.,* 6**,** e00007-18.

CHHABRA, P., DE GRAAF, M., PARRA, G. I., CHAN, M. C.-W., GREEN, K., MARTELLA, V., WANG, Q., WHITE, P. A., KATAYAMA, K. & VENNEMA, H. 2019. Updated classification of norovirus genogroups and genotypes. *Journal of General Virology,* 100**,** 1393-1406.

CHIBA, S., NAKATA, S., NUMATA-KINOSHITA, K. & HONMA, S. 2000. Sapporo virus: history and recent findings. *Journal of Infectious Diseases,* 181**,** S303-S308.

CLAIRE THORNE, D. P., SIMONA FIORE, CATHERINE PECKHAM, AND MARIE-LOUISE NEWELL 2005. Mother-to-child transmission of HIV infection in the era of highly active antiretroviral therapy. *Clinical Infectious Diseases,* 40**,** 458-465.

CLARKE, I. N. & LAMBDEN, P. R. 2000. Organization and expression of calicivirus genes. *The Journal of Infectious Diseases,* 181**,** S309-S316.

COLSTON, J. M., FRANCOIS, R., PISANIC, N., PEÑATARO YORI, P., MCCORMICK, B. J., OLORTEGUI, M. P., GAZI, M. A., SVENSEN, E., AHMED, M. M. M. & MDUMA, E. 2019. Effects of Child and Maternal Histo-Blood Group Antigen Status on Symptomatic and Asymptomatic Enteric Infections in Early Childhood. *The Journal of Infectious Diseases,* 220**,** 151-162.

CONTROL, C. F. D. & PREVENTION 1999. Rotavirus vaccine for the prevention of rotavirus gastroenteritis among children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morbidity and Mortality Weekly Report,* 48**,** 1-20.

CONTROL, C. F. D. & PREVENTION. 2020. *The National Respiratory and Enteric Virus Surveillance System (NREVSS)* [Online]. Available:

https://www.cdc.gov/surveillance/nrevss/rotavirus/natl-trend.html [Accessed 20/01/2020 2020].

COSTANTINI, V., MORANTZ, E. K., BROWNE, H., ETTAYEBI, K., ZENG, X.-L., ATMAR, R. L., ESTES, M. K. & VINJÉ, J. 2018. Human norovirus replication in human intestinal enteroids as model to evaluate virus inactivation. *Emerging Infectious Diseases,* 24**,** 1453.

DAI, Y.-C., ZHANG, X.-F., XIA, M., TAN, M., QUIGLEY, C., LEI, W., FANG, H., ZHONG, W., LEE, B. & PANG, X. 2015. Antigenic relatedness of Norovirus GII. 4 variants determined by human challenge sera. *PloS one,* 10**,** e0124945.

DAS, B. K., GENTSCH, J. R., CICIRELLO, H. G., WOODS, P. A., GUPTA, A., RAMACHANDRAN, M., KUMAR, R., BHAN, M. & GLASS, R. I. 1994. Characterization of rotavirus strains from newborns in New Delhi, India. *Journal of Clinical Microbiology,* 32**,** 1820-1822.

DE GRAAF, M., VAN BEEK, J. & KOOPMANS, M. P. 2016. Human norovirus transmission and evolution in a changing world. *Nature Reviews Microbiology,* 14**,** 421-433.

DE GRAAF, M., VAN BEEK, J., VENNEMA, H., PODKOLZIN, A., HEWITT, J., BUCARDO, F., TEMPLETON, K., MANS, J., NORDGREN, J. & REUTER, G. 2015. Emergence of a novel GII. 17 norovirus–End of the GII. 4 era? *Euro surveillance: bulletin europeen sur les maladies transmissibles= European communicable disease bulletin,* 20.

DONALDSON, E. F., LINDESMITH, L. C., LOBUE, A. D. & BARIC, R. S. 2010. Viral shape-shifting: norovirus evasion of the human immune system. *Nature Reviews Microbiology,* 8**,** 231.

EMBREE, J., DATTA, P., STACKIW, W., SEKLA, L., BRADDICK, M., KREISS, J., PAMBA, H., WAMOLA, I., NDINYA-ACHOLA, J. & LAW, B. 1992. Increased risk of early measles in infants of human immunodeficiency virus type 1-seropositive mothers. *Journal of Infectious Diseases,* 165**,** 262-267.

ENQUIST, L. W., SKALKA, A.-M., FLINT, S. J., RALL, G. F. & RACANIELLO, V. R. 2015. Principles of Virology, 2-vol-set. ASM (American Society for Microbiology).

EPALZA, C., GOETGHEBUER, T., HAINAUT, M., PRAYEZ, F., BARLOW, P., DEDISTE, A., MARCHANT, A. & LEVY, J. 2010. High incidence of invasive group B streptococcal infections in HIV-exposed uninfected infants. *Pediatrics,* 126**,** e631-e638.

ESTES, M. & KAPIKIAN, A. 2007. Rotaviruses, p 1917–1974. *Fields Virology,* 2.

ETTAYEBI, K., CRAWFORD, S. E., MURAKAMI, K., BROUGHMAN, J. R., KARANDIKAR, U., TENGE, V. R., NEILL, F. H., BLUTT, S. E., ZENG, X.-L. & QU, L. 2016. Replication of human noroviruses in stem cell–derived human enteroids. *Science,* 353**,** 1387-1393.

ETTAYEBI, K. & HARDY, M. E. 2003. Norwalk virus nonstructural protein p48 forms a complex with the SNARE regulator VAP-A and prevents cell surface expression of vesicular stomatitis virus G protein. *Journal of Virology,* 77**,** 11790-11797.

EVANS, C., JONES, C. E. & PRENDERGAST, A. J. 2016. HIV-exposed, uninfected infants: new global challenges in the era of paediatric HIV elimination. *The Lancet Infectious Diseases,* 16**,** e92-e107.

FARKAS, T., ZHONG, W., JING, Y., HUANG, P., ESPINOSA, S., MARTINEZ, N., MORROW, A., RUIZ-PALACIOS, G., PICKERING, L. & JIANG, X. 2004. Genetic diversity among sapoviruses. *Archives of Virology,* 149**,** 1309-1323.

FEITERNA-SPERLING, C., WEIZSAECKER, K., BÜHRER, C., CASTELEYN, S., LOUI, A., SCHMITZ, T., WAHN, V. & OBLADEN, M. 2007. Hematologic effects of maternal antiretroviral therapy and transmission prophylaxis in HIV-1-exposed uninfected newborn infants. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 45**,** 43-51.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* 39**,** 783-791.

FERNANDEZ-VEGA, V., SOSNOVTSEV, S. V., BELLIOT, G., KING, A. D., MITRA, T., GORBALENYA, A. & GREEN, K. Y. 2004. Norwalk virus N-terminal nonstructural protein is associated with disassembly of the Golgi complex in transfected cells. *Journal of Virology,* 78**,** 4827-4837.

FERRER-ADMETLLA, A., SIKORA, M., LAAYOUNI, H., ESTEVE, A., ROUBINET, F., BLANCHER, A., CALAFELL, F., BERTRANPETIT, J. & CASALS, F. 2009. A natural history of FUT2 polymorphism in humans. *Molecular Biology and Evolution,* 26**,** 1993-2003.

FREEDMAN, S. B., ELTORKY, M., GORELICK, M. & GROUP, P. E. R. C. G. S. 2010. Evaluation of a gastroenteritis severity score for use in outpatient settings. *Pediatrics,* 125**,** e1278-e1285.

FRENCK, R., BERNSTEIN, D. I., XIA, M., HUANG, P., ZHONG, W., PARKER, S., DICKEY, M., MCNEAL, M. & JIANG, X. 2012. Predicting susceptibility to norovirus GII. 4 by use of a challenge model involving humans. *Journal of Infectious Diseases***,** jis514.

FUMAGALLI, M., CAGLIANI, R., POZZOLI, U., RIVA, S., COMI, G. P., MENOZZI, G., BRESOLIN, N. & SIRONI, M. 2009. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Research,* 19**,** 199-212.

GALLO RC, M. L. 1987. The chronology of AIDS research. *Nature,* 326**,** 435-436.

GARCÍA, C., DUPONT, H. L., LONG, K. Z., SANTOS, J. I. & KO, G. 2006. Asymptomatic norovirus infection in Mexican children. *Journal of Clinical Microbiology,* 44**,** 2997-3000.

GASTAÑADUY, P. A., HALL, A. J. & PARASHAR, U. D. 2013. Rotavirus. *Foodborne Infections and Intoxications.* Elsevier.

GAULT, E., CHIKHI-BRACHET, R., DELON, S., SCHNEPF, N., ALBIGES, L., GRIMPREL, E., GIRARDET, J.-P., BEGUE, P. & GARBARG-CHENON, A. 1999. Distribution of human rotavirus G types circulating in Paris, France, during the 1997–1998 epidemic: high prevalence of type G4. *Journal of Clinical Microbiology,* 37**,** 2373-2375.

GENTSCH, J., GLASS, R., WOODS, P., GOUVEA, V., GORZIGLIA, M., FLORES, J., DAS, B. & BHAN, M. 1992. Identification of group A rotavirus gene 4 types by polymerase chain reaction. *Journal of Clinical Microbiology,* 30**,** 1365-1373.

GLASS, R. I., PARASHAR, U. D. & ESTES, M. K. 2009. Norovirus gastroenteritis. *New England Journal of Medicine,* 361**,** 1776-1785.

GOUVEA, V., GLASS, R. I., WOODS, P., TANIGUCHI, K., CLARK, H. F., FORRESTER, B. & FANG, Z.-Y. 1990. Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. *Journal of Clinical Microbiology,* 28**,** 276-282.

GREEN, K. Y. 2007. Caliciviridae: the noroviruses. *Fields Virology,* 5**,** 949-979.

GRIFFIN, D. Fields virology. Wolters Kluwer Health Adis (ESP), 2013.

GROOME, M. J., PAGE, N., CORTESE, M. M., MOYES, J., ZAR, H. J., KAPONGO, C. N., MULLIGAN, C., DIEDERICKS, R., COHEN, C. & FLEMING, J. A. 2014. Effectiveness of monovalent human rotavirus vaccine against admission to hospital for acute rotavirus diarrhoea in South African children: a case-control study. *The Lancet Infectious Diseases,* 14**,** 1096-1104.

GUPTA, S., KRISHNAN, A., SHARMA, S., KUMAR, P., ANEJA, S. & RAY, P. 2018. Changing pattern of prevalence, genetic diversity, and mixed infections of viruses associated with acute gastroenteritis in pediatric patients in New Delhi, India. *Journal of Medical Virology,* 90**,** 469-476.

HALE, MURRAY, C., BARBER, R., FOREMAN, K., OZGOREN, A., ABD-ALLAH, F., ABERA, S., ABOYANS, V., ABRAHAM, J. & ABUBAKAR, I. 2015. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition.

HALL, T. 2005. Bioedit v 7.0. 5. *Ibis Therapeutics, a division of Isis Pharmaceuticals, Carlsbad*.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series, 1999. [London]: Information Retrieval Ltd., c1979-c2000., 95-98.

HAN, J.-H., KIM, S.-H. & PAIK, S.-Y. 2017. Evaluation of an Assay Based On Multiple Detection Temperature Technique for Simultaneous Detection of Viral Gastroenteritis-Causing Pathogens. *Jundishapur Journal of Microbiology,* 10.

HANSMAN, G. S. 2007. Antigenic Diversity of Human Sapoviruses-Volume 13, Number 10—October 2007-*Emerging Infectious Disease Journal*-CDC.

HANSMAN, G. S., JIANG, X. J. & GREEN, K. Y. 2010. *Caliciviruses: Molecular and Cellular Virology*, Horizon Scientific Press.

HANSMAN, G. S., OKA, T., KATAYAMA, K. & TAKEDA, N. 2007. Human sapoviruses: genetic diversity, recombination, and classification. *Reviews in Medical Virology,* 17**,** 133-141.

HARRIS, V. C., ARMAH, G., FUENTES, S., KORPELA, K. E., PARASHAR, U., VICTOR, J. C., TATE, J., DE WEERTH, C., GIAQUINTO, C. & WIERSINGA, W. J. 2016. Significant correlation between the infant gut microbiome and rotavirus vaccine response in rural Ghana. *The Journal of Infectious Diseases,* 215**,** 34-41.

HAVELAAR, A. H., KIRK, M. D., TORGERSON, P. R., GIBB, H. J., HALD, T., LAKE, R. J., PRAET, N., BELLINGER, D. C., DE SILVA, N. R. & GARGOURI, N. 2015. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLoS Medicine,* 12**,** e1001923.

HEMMING, M., RÄSÄNEN, S., HUHTI, L., PALONIEMI, M., SALMINEN, M. & VESIKARI, T. 2013. Major reduction of rotavirus, but not norovirus, gastroenteritis in children seen in hospital after the introduction of RotaTeq vaccine into the National Immunization Programme in Finland. *European Journal of Pediatrics,* 172**,** 739-746.

HENRY, S. 2001. Molecular diversity in the biosynthesis of GI tract glycoconjugates. A blood group related chart of microorganism receptors. *Transfusion Clinique et Biologique,* 8**,** 226-230.

HIRVONEN, J. J. 2019. Comparison of three multiplex real-time PCR assays for detection of enteric viruses in patients with diarrhea. *European Journal of Clinical Microbiology & Infectious Diseases,* 38**,** 241-244.

HUANG, P., XIA, M., TAN, M., ZHONG, W., WEI, C., WANG, L., MORROW, A. & JIANG, X. 2012. Spike protein VP8* of human rotavirus recognizes histo-blood group antigens in a type-specific manner. *Journal of Virology,* 86**,** 4833-4843.

HUNGERFORD, D., JERE, K. C., BAR-ZEEV, N., HARRIS, J. P., CUNLIFFE, N. A. & ITURRIZA-GÓMARA, M. 2020. Epidemiology and genotype diversity of norovirus infections among children aged< 5 years following rotavirus vaccine introduction in Blantyre, Malawi. *Journal of Clinical Virology,* 123**,** 104248.

HUTSON, A. M., ATMAR, R. L. & ESTES, M. K. 2004. Norovirus disease: changing epidemiology and host susceptibility factors. *Trends in Microbiology,* 12**,** 279-287.

HYUN, J., KO, D.-H., LEE, S.-K., KIM, H.-S., KIM, J.-S., SONG, W. & KIM, H. S. 2018. Evaluation of a new multiplex real-time PCR assay for detecting gastroenteritis-causing viruses in stool samples. *Annals of Laboratory Medicine,* 38**,** 220-225.

IHME. 2018. *Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results* [Online]. United States: Institute for Health Metrics and Evaluation (IHME). Available: http://ghdx.healthdata.org/gbd-results-tool [Accessed 12 November 2019 2019].

ITURRIZA-GÓMARA, M. & CUNLIFFE, N. A. 2020. Viral Gastroenteritis. *Hunter's Tropical Medicine and Emerging Infectious Diseases.* Elsevier.

JIANG, X., ZHONG, W., KAPLAN, M., PICKERING, L. K. & MATSON, D. O. 1999. Expression and characterization of Sapporo-like human calicivirus capsid proteins in baculovirus. *Journal of Virological Methods,* 78**,** 81-91.

JIN, M., HE, Y., LI, H., HUANG, P., ZHONG, W., YANG, H., ZHANG, H., TAN, M. & DUAN, Z.-J. 2013. Two gastroenteritis outbreaks caused by GII Noroviruses: host susceptibility and HBGA phenotypes. *PloS one,* 8**,** e58605.

JONESTELLER, C. L., BURNETT, E., YEN, C., TATE, J. E. & PARASHAR, U. D. 2017. Effectiveness of rotavirus vaccination: a systematic review of the first decade of global postlicensure data, 2006–2016. *Clinical Infectious Diseases,* 65**,** 840-850.

KABUE, J. P., MEADER, E., HUNTER, P. R. & POTGIETER, N. 2016. Norovirus prevalence and estimated viral load in symptomatic and asymptomatic children from rural communities of Vhembe district, South Africa. *Journal of Clinical Virology,* 84**,** 12-18.

KAGEYAMA, T., KOJIMA, S., SHINOHARA, M., UCHIDA, K., FUKUSHI, S., HOSHINO, F. B., TAKEDA, N. & KATAYAMA, K. 2003. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *Journal of Clinical Microbiology,* 41**,** 1548-1557.

KAMBHAMPATI, A., PAYNE, D. C., COSTANTINI, V. & LOPMAN, B. A. 2015. Host genetic susceptibility to enteric viruses: a systematic review and metaanalysis. *Clinical Infectious Diseases,* 62**,** 11-18.

KAPIKIAN, A. Z. 2000. The Discovery of the 27-nm Norwalk Virus: An Historic Perspective. *The Journal of Infectious Diseases,* 181**,** S295-S302.

KAPIKIAN, A. Z., WYATT, R. G., DOLIN, R., THORNHILL, T. S., KALICA, A. R. & CHANOCK, R. M. 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *Journal of Virology,* 10**,** 1075-1081.

KARST, S. M. & BARIC, R. S. 2015. What is the reservoir of emergent human norovirus strains? *Journal of Virology,* 89**,** 5756-5759.

KARST, S. M., ZHU, S. & GOODFELLOW, I. G. 2015. The molecular pathology of noroviruses. *The Journal of Pathology,* 235**,** 206-216.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* 16**,** 111-120.

KINDBERG, E. & SVENSSON, L. 2009. Genetic basis of host resistance to norovirus infection. *Future Virology,* 4**,** 369-382.

KIRKWOOD, C. D., BONIFACE, K., BARNES, G. L. & BISHOP, R. F. 2011. Distribution of rotavirus genotypes after introduction of rotavirus vaccines, Rotarix® and RotaTeq®, into the National Immunization Program of Australia. *The Pediatric Infectious Disease Journal,* 30**,** S48-S53.

KIULIA, N., HOFSTRA, N., VERMEULEN, L., OBARA, M., MEDEMA, G. & ROSE, J. 2015. Global occurrence and emission of rotaviruses to surface waters. *Pathogens,* 4**,** 229-255.

KOBAYASHI, S., FUJIWARA, N., YASUI, Y., YAMASHITA, T., HIRAMATSU, R. & MINAGAWA, H. 2012. A foodborne outbreak of sapovirus linked to catered box lunches in Japan. *Archives of Virology***,** 1-3.

KODA, Y., TACHIDA, H., PANG, H., LIU, Y., SOEJIMA, M., GHADERI, A. A., TAKENAKA, O. & KIMURA, H. 2001. Contrasting patterns of polymorphisms at the ABO-secretor gene (FUT2) and plasma α (1, 3) fucosyltransferase gene (FUT6) in human populations. *Genetics,* 158**,** 747-756.

KOJIMA, S., KAGEYAMA, T., FUKUSHI, S., HOSHINO, F. B., SHINOHARA, M., UCHIDA, K., NATORI, K., TAKEDA, N. & KATAYAMA, K. 2002. Genogroup-specific PCR primers for detection of Norwalk-like viruses. *Journal of Virological Methods,* 100**,** 107-114.

KOTLOFF, K. L. 2017. The burden and etiology of diarrheal illness in developing countries. *Pediatric Clinics,* 64**,** 799-814.

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. 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. *The Lancet,* 382**,** 209-222.

KRONEMAN, A., VEGA, E., VENNEMA, H., VINJÉ, J., WHITE, P. A., HANSMAN, G., GREEN, K., MARTELLA, V., KATAYAMA, K. & KOOPMANS, M. 2013. Proposal for a unified norovirus nomenclature and genotyping. *Archives of Virology,* 158**,** 2059-2068.

KRONEMAN, A., VENNEMA, H., DEFORCHE, K., AVOORT, H., PENARANDA, S., OBERSTE, M., VINJÉ, J. & KOOPMANS, M. 2011. An automated genotyping tool for enteroviruses and noroviruses. *Journal of Clinical Virology,* 51**,** 121-125.

KUMTHIP, K., KHAMRIN, P. & MANEEKARN, N. 2019. Epidemiology and genotypic distribution of noroviruses in patients with acute gastroenteritis in developing and developed countries. *Annals of Research Hospitals,* 3.

LEE, B., DICKSON, D. M., DECAMP, A. C., ROSS COLGATE, E., DIEHL, S. A., UDDIN, M. I., SHARMIN, S., ISLAM, S., BHUIYAN, T. R. & ALAM, M. 2018. Histo–Blood Group Antigen Phenotype Determines Susceptibility to Genotype-Specific Rotavirus Infections and Impacts Measures of Rotavirus Vaccine Efficacy. *The Journal of Infectious Diseases,* 217**,** 1399-1407.

LEE, L. E., CEBELINSKI, E. A., FULLER, C., KEENE, W. E., SMITH, K., VINJÉ, J. & BESSER, J. M. 2012. Sapovirus outbreaks in long-term care facilities, Oregon and Minnesota, USA, 2002–2009. *Emerging Infectious Diseases,* 18**,** 873.

LEFKOWITZ, E. J., DEMPSEY, D. M., HENDRICKSON, R. C., ORTON, R. J., SIDDELL, S. G. & SMITH, D. B. 2017. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research,* 46**,** D708-D717.

LEKANA-DOUKI, S. E., KOMBILA-KOUMAVOR, C., NKOGHE, D., DROSTEN, C., DREXLER, J. F. & LEROY, E. M. 2015. Molecular epidemiology of enteric viruses and genotyping of rotavirus A, adenovirus and astrovirus among children under 5 years old in Gabon. *International Journal of Infectious Diseases,* 34**,** 90-95.

LEROUX-ROELS, G., CRAMER, J. P., MENDELMAN, P. M., SHERWOOD, J., CLEMENS, R., AERSSENS, A., DE COSTER, I., BORKOWSKI, A., BAEHNER, F. & VAN DAMME, P. 2017. Safety and immunogenicity of different formulations of norovirus vaccine candidate in healthy adults: A randomized, controlled, double-blind clinical trial. *The Journal of Infectious Diseases,* 217**,** 597-607.

LEW, E. A., POLES, M. A. & DIETERICH, D. T. 1997. Diarrheal diseases associated with HIV infection. *Gastroenterology Clinics of North America,* 26**,** 259-290.

LEWIS, Z. T., TOTTEN, S. M., SMILOWITZ, J. T., POPOVIC, M., PARKER, E., LEMAY, D. G., VAN TASSELL, M. L., MILLER, M. J., JIN, Y.-S. & GERMAN, J. B. 2015. Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome,* 3**,** 13.

LINDESMITH, L., MOE, C., MARIONNEAU, S., RUVOEN, N., JIANG, X., LINDBLAD, L., STEWART, P., LEPENDU, J. & BARIC, R. 2003. Human susceptibility and resistance to Norwalk virus infection. *Nature Medicine,* 9**,** 548-553.

LINHARES, A. C., STUPKA, J. A., CIAPPONI, A., BARDACH, A. E., GLUJOVSKY, D., ARUJ, P. K., MAZZONI, A., RODRIGUEZ, J. A. B., REARTE, A. & LANZIERI, T. M. 2011. Burden and typing of rotavirus group A in Latin America and the Caribbean: systematic review and meta-analysis. *Reviews in Medical Virology,* 21**,** 89-109.

LIU, J., PLATTS-MILLS, J. A., JUMA, J., KABIR, F., NKEZE, J., OKOI, C., OPERARIO, D. J., UDDIN, J., AHMED, S. & ALONSO, P. L. 2016a. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *The Lancet,* 388**,** 1291-1301.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012a. Comparison of next-generation sequencing systems. *BioMed Research International,* 2012.

LIU, L., OZA, S., HOGAN, D., CHU, Y., PERIN, J., ZHU, J., LAWN, J. E., COUSENS, S., MATHERS, C. & BLACK, R. E. 2016b. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet,* 388**,** 3027-3035.

LIU, P., WANG, X., LEE, J.-C., TEUNIS, P., HU, S., PARADISE, H. T. & MOE, C. 2014. Genetic Susceptibility to Norovirus GII. 3 and GII. 4 Infections in Chinese Pediatric Diarrheal Children. *The Pediatric Infectious Disease Journal,* 33**,** e305.

LIU, Y., HUANG, P., TAN, M., LIU, Y., BIESIADA, J., MELLER, J., CASTELLO, A. A., JIANG, B. & JIANG, X. 2012b. Rotavirus VP8*: phylogeny, host range, and interaction with histo-blood group antigens. *Journal of Virology,* 86**,** 9899-9910.

LIU, Y., KODA, Y., SOEJIMA, M., PANG, H., SCHLAPHOFF, T., DU TOIT, E. D. & KIMURA, H. 1998. Extensive polymorphism of the FUT2 gene in an African (Xhosa) population of South Africa. *Human Genetics,* 103**,** 204-210.

LOISY, F., ATMAR, R., GUILLON, P., LE CANN, P., POMMEPUY, M. & LE GUYADER, F. 2005. Real-time RT-PCR for norovirus screening in shellfish. *Journal of Virological Methods,* 123**,** 1-7.

LOPMAN, B. 2015. Global burden of norovirus and prospects for vaccine development. *Centers for Disease Control and Prevention*.

LOPMAN, B. A., TRIVEDI, T., VICUÑA, Y., COSTANTINI, V., COLLINS, N., GREGORICUS, N., PARASHAR, U., SANDOVAL, C., BRONCANO, N. & VACA, M. 2014a. Norovirus infection and disease in an Ecuadorian birth cohort: association of certain norovirus genotypes with host FUT2 secretor status. *The Journal of Infectious Diseases,* 211**,** 1813-1821.

LOPMAN, B. A., VINJÉ, J. & GLASS, R. I. 2014b. Noroviruses, Sapoviruses, and Astroviruses. *Viral Infections of Humans.* Springer.

LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R. & AHN, S. Y. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet,* 380**,** 2095-2128.

LUCIANI, F., BULL, R. A. & LLOYD, A. R. 2012. Next generation deep sequencing and vaccine design: today and tomorrow. *Trends in Biotechnology,* 30**,** 443-452.

LÜTHI, T. M. 2018. Epidemiology of foodborne viral infections. *Foodborne Disease Handbook: Volume II: Viruses, Parasites, Pathogens, and HACCP*.

MA, X., DAN-DI LI, X.-M. S., GUO, Y.-Q., XIANG, J.-Y., WANG, W.-H., ZHANG, L.-X., GU, Q.-J. & DUAN, Z.-J. 2015. Binding patterns of rotavirus genotypes P [4], P [6], and P [8] in China with histo-blood group antigens. *PloS one,* 10.

MABASA, V., MENO, K., TAYLOR, M. & MANS, J. 2018. Environmental surveillance for noroviruses in selected South African wastewaters 2015–2016: Emergence of the novel GII. 17. *Food and Environmental Virology,* 10**,** 16-28.

MADELEY, C. 1979. Comparison of the features of astroviruses and caliciviruses seen in samples of feces by electron microscopy. *Journal of Infectious Diseases,* 139**,** 519-523.

MADELEY, C. & COSGROVE, B. 1976. Caliciviruses in man. *The Lancet,* 307**,** 199-200.

MAGAGULA, N. B., ESONA, M. D., NYAGA, M. M., STUCKER, K. M., HALPIN, R. A., STOCKWELL, T. B., SEHERI, M. L., STEELE, A. D., WENTWORTH, D. E. & MPHAHLELE, M. J. 2015. Whole genome analyses of G1P [8] rotavirus strains from vaccinated and non-vaccinated South African children presenting with diarrhea. *Journal of Medical Virology,* 87**,** 79-101.

MAKGATHO, E., PATEL, F., SOLOMON, F., GROOME, M. J., LALA, S. G., VALLABH, P. & DANGOR, Z. 2019. The Burden of Acute Diarrheal Disease in Young Hospitalized Urban South African Children Five Years After Rotavirus Vaccine Introduction: A Retrospective Descriptive Study. *The Pediatric Infectious Disease Journal,* 38**,** 752-756.

MANS, J. 2019. Norovirus Infections and Disease in Lower-Middle-and Low-Income Countries, 1997–2018. *Viruses,* 11**,** 341.

MANS, J., ARMAH, G. E., STEELE, A. D. & TAYLOR, M. B. 2016. Norovirus Epidemiology in Africa: A Review. *PloS one,* 11**,** e0146280.

MARINDA, E., HUMPHREY, J. H., ILIFF, P. J., MUTASA, K., NATHOO, K. J., PIWOZ, E. G., MOULTON, L. H., SALAMA, P., WARD, B. J. & GROUP, Z. S. 2007. Child mortality according to maternal and infant HIV status in Zimbabwe. *The Pediatric Infectious Disease Journal,* 26**,** 519-526.

MATTHIJNSSENS, J., CIARLET, M., HEIMAN, E., ARIJS, I., DELBEKE, T., MCDONALD, S. M., PALOMBO, E. A., ITURRIZA-GÓMARA, M., MAES, P. & PATTON, J. T. 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**,** 3204-3219.

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**,** 1177-1182.

MATTIOLI, M. C., PICKERING, A. J., GILSDORF, R. J., DAVIS, J. & BOEHM, A. B. 2013. Hands and water as vectors of diarrheal pathogens in Bagamoyo, Tanzania. *Environmental Science & Technology,* 47**,** 355-363.

MATUSSEK, A., DIENUS, O., DJENEBA, O., SIMPORE, J., NITIEMA, L. & NORDGREN, J. 2015. Molecular characterization and genetic susceptibility of sapovirus in children with diarrhea in Burkina Faso. *Infection, Genetics and Evolution,* 32**,** 396-400.

MCGOVERN, D. P., JONES, M. R., TAYLOR, K. D., MARCIANTE, K., YAN, X., DUBINSKY, M., IPPOLITI, A., VASILIAUSKAS, E., BEREL, D. & DERKOWSKI, C. 2010. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Human Molecular Genetics,* 19**,** 3468-3476.

MELLOU, K., POTAMITI-KOMI, M., SIDEROGLOU, T., KALLIMANI, A., BAKA, A., ILIOPOULOS, D., LIAPIS, E., RIGAKOS, G., KARAGEORGOU, A. & TSERONI, M. 2012. Detection and management of a norovirus gastroenteritis outbreak, Special Olympics World Summer Games, Greece, June 2011. *International Journal of Public Health Epidemiology,* 1**,** 20-25.

MIURA, F., MATSUYAMA, R. & NISHIURA, H. 2018. Estimating the asymptomatic ratio of norovirus infection during foodborne outbreaks with laboratory testing in Japan. *Journal of Epidemiology***,** JE20170040.

MIURA, T., SANO, D., SUENAGA, A., YOSHIMURA, T., FUZAWA, M., NAKAGOMI, T., NAKAGOMI, O. & OKABE, S. 2013. Histo-blood group antigen-like substances of human enteric bacteria as specific adsorbents for human noroviruses. *Journal of Virology,* 87**,** 9441-9451.

MOHANTY, E., DWIBEDI, B., KAR, S. & ACHARYA, A. 2017. Epidemiological features and genetic characterization of virus strains in rotavirus associated gastroenteritis in children of Odisha in Eastern India. *Infection, Genetics and Evolution,* 53**,** 77-84.

MONAVARI, S. H. R., HADIFAR, S., MOSTAFAEI, S., MIRI, A., KESHAVARZ, M., BABAEI, F. & MOGHOOFEI, M. 2017. Epidemiology of rotavirus in the Iranian children: A systematic review and meta-analysis. *Journal of Global Infectious Diseases,* 9**,** 66.

MOORES, P. & BRAIN, P. 1968. Lewis groups and secretor status in Natal Bantu. *Transfusion,* 8**,** 283-288.

MURPHY, T. V., GARGIULLO, P. M., MASSOUDI, M. S., NELSON, D. B., JUMAAN, A. O., OKORO, C. A., ZANARDI, L. R., SETIA, S., FAIR, E. & LEBARON, C. W. 2001. Intussusception among infants given an oral rotavirus vaccine. *New England Journal of Medicine,* 344**,** 564-572.

MURRAY, T. Y., NADAN, S., PAGE, N. A. & TAYLOR, M. B. 2016. Diverse sapovirus genotypes identified in children hospitalised with gastroenteritis in selected regions of South Africa. *Journal of Clinical Virology,* 76**,** 24-29.

NADAN, S., TAYLOR, M. B., GROOME, M. J., COHEN, C., MADHI, S. A. & PAGE, N. A. 2019. Epidemiology of human astroviruses among children younger than 5 years: Prospective hospital-based sentinel surveillance in South Africa, 2009-2014. *Journal of Medical Virology,* 91**,** 225-234.

NAKAMURA, N., KOBAYASHI, S., MINAGAWA, H., MATSUSHITA, T., SUGIURA, W. & IWATANI, Y. 2016. Molecular epidemiology of enteric viruses in patients with acute gastroenteritis in Aichi Prefecture, Japan, 2008/09− 2013/14. *Journal of Medical Virology,* 88**,** 1180-1186.

NASIR, W., FRANK, M., KUNZE, A., BALLY, M., PARRA, F., NYHOLM, P.-G., HÖÖK, F. & LARSON, G. R. 2017. Histo-blood group antigen presentation is critical for binding of norovirus VLP to glycosphingolipids in model membranes. *ACS Chemical Biology,* 12**,** 1288-1296.

NORDGREN, J., NITIEMA, L. W., OUERMI, D., SIMPORE, J. & SVENSSON, L. 2013. Host genetic factors affect susceptibility to norovirus infections in Burkina Faso. *PloS one,* 8**,** e69557.

NORDGREN, J., SHARMA, S., BUCARDO, F., NASIR, W., GÜNAYDıN, G., OUERMI, D., NITIEMA, L. W., BECKER-DREPS, S., SIMPORE, J. & HAMMARSTRÖM, L. 2014. Both Lewis and secretor status mediate susceptibility to rotavirus infections in a rotavirus genotype–dependent manner. *Clinical Infectious Diseases,* 59**,** 1567-1573.

NORDGREN, J. & SVENSSON, L. 2019. Genetic susceptibility to human norovirus infection: an update. *Viruses,* 11**,** 226.

NYAGA, M. M., TAN, Y., SEHERI, M. L., HALPIN, R. A., AKOPOV, A., STUCKER, K. M., FEDOROVA, N. B., SHRIVASTAVA, S., STEELE, A. D. & MWENDA, J. M. 2018. Whole-genome sequencing and analyses identify high genetic heterogeneity, diversity and endemicity of rotavirus genotype P [6] strains circulating in Africa. *Infection, Genetics and Evolution,* 63**,** 79-88.

O'RYAN, M. 2009. The ever-changing landscape of rotavirus serotypes. *The Pediatric Infectious Disease Journal,* 28**,** S60-S62.

OKA, T., IRITANI, N., OKADA, M., OGAWA, T., IIZUKA, S., TATSUMI, C., HARADA, S., HAGA, K. & DOAN, Y. H. 2018. First complete genome sequences of human sapovirus strains classified as GI. 3, GI. 4, GI. 6, GI. 7, and GII. 7. *Genome Announcement.,* 6**,** e00168-18.

OKA, T., MORI, K., IRITANI, N., HARADA, S., UEKI, Y., IIZUKA, S., MISE, K., MURAKAMI, K., WAKITA, T. & KATAYAMA, K. 2012. Human sapovirus classification based on complete capsid nucleotide sequences. *Archives of Virology,* 157**,** 349-352.

OKA, T., WANG, Q., KATAYAMA, K. & SAIF, L. J. 2015. Comprehensive review of human sapoviruses. *Clinical Microbiology Reviews,* 28**,** 32-53.

OPERARIO, D. J., PLATTS-MILLS, J. A., NADAN, S., PAGE, N., SEHERI, M., MPHAHLELE, J., PRAHARAJ, I., KANG, G., ARAUJO, I. T. & LEITE, J. P. G. 2017. Etiology of severe acute watery diarrhea in children in the global rotavirus surveillance network using quantitative polymerase chain reaction. *The Journal of Infectious Diseases,* 216**,** 220-227.

OUEDRAOGO, N., KAPLON, J., BONKOUNGOU, I. J. O., TRAORE, A. S., POTHIER, P., BARRO, N. & AMBERT-BALAY, K. 2016. Prevalence and genetic diversity of enteric viruses in children with diarrhea in Ouagadougou, Burkina Faso. *PloS one,* 11.

PAGE, N., GROOME, M., NADAN, S., NETSHIKWETA, R., KEDDY, K., POONSAMY, B., MOYES, J., WALAZA, S., KAHN, K. & MADHI, S. 2017. Norovirus epidemiology in South African children< 5 years hospitalised for diarrhoeal illness between 2009 and 2013. *Epidemiology & Infection,* 145**,** 1942-1952.

PAGE, N., GROOME, M. J., MURRAY, T., NADAN, S., NETSHIKWETA, R., KEDDY, K. H., POONSAMY, B., MOYES, J., WALAZA, S. & KAHN, K. 2016a. Sapovirus prevalence in children less than five years of age hospitalised for diarrhoeal disease in South Africa, 2009–2013. *Journal of Clinical Virology,* 78**,** 82-88.

PAGE, N., KRUGER, T., SEHERI, M., PEENZE, I., QUAN, V., GROOME, M. & MADHI, S. 2015. Rotavirus surveillance report, South Africa 2014-2015: A comparison of previous rotavirus seasons. *Communicable Diseases Bulletin,* 14**,** 119-125.

PAGE, N., KRUGER, T., SEHERI, M., PEENZE, I., QUAN, V., GROOME, M. & MADHI, S. 2016b. ROTAVIRUS SURVEILLANCE REPORT, SOUTH AFRICA, 2014-2015: A COMPARISON WITH PREVIOUS ROTAVIRUS SEASONS. *FOREWORD CONTENTS***,** 126.

PANG, X. 2017. Detection and laboratory diagnosis of noroviruses. *The Norovirus.* Elsevier.

PARASHAR, U. D., BURTON, A., LANATA, C., BOSCHI-PINTO, C., SHIBUYA, K., STEELE, D., BIRMINGHAM, M. & GLASS, R. I. 2009. Global mortality associated with rotavirus disease among children in 2004. *Journal of Infectious Diseases,* 200**,** S9-S15.

PARKER, E. P., RAMANI, S., LOPMAN, B. A., CHURCH, J. A., ITURRIZA-GOMARA, M., PRENDERGAST, A. J. & GRASSLY, N. C. 2018. Causes of impaired oral vaccine efficacy in developing countries. *Future Microbiology,* 13**,** 97-118.

PATEL, M. M., HALL, A. J., VINJÉ, J. & PARASHAR, U. D. 2009. Noroviruses: a comprehensive review. *Journal of Clinical Virology,* 44**,** 1-8.

PAYNE, D. C., CURRIER, R. L., STAAT, M. A., SAHNI, L. C., SELVARANGAN, R., HALASA, N. B., ENGLUND, J. A., WEINBERG, G. A., BOOM, J. A. & SZILAGYI, P. G. 2015. Epidemiologic association between FUT2 secretor status and severe rotavirus gastroenteritis in children in the United States. *JAMA pediatrics,* 169**,** 1040-1045.

PAYNE, D. C., VINJÉ, J., SZILAGYI, P. G., EDWARDS, K. M., STAAT, M. A., WEINBERG, G. A., HALL, C. B., CHAPPELL, J., BERNSTEIN, D. I. & CURNS, A. T. 2013. Norovirus and medically attended gastroenteritis in US children. *New England Journal of Medicine,* 368**,** 1121-1130.

PECENKA, C., DEBELLUT, F., BAR-ZEEV, N., ANWARI, P., NONVIGNON, J., SHAMSUZZAMAN, M. & CLARK, A. 2018. Re-evaluating the cost and cost-effectiveness of rotavirus vaccination in Bangladesh, Ghana, and Malawi: A comparison of three rotavirus vaccines. *Vaccine,* 36**,** 7472-7478.

PHILLIPS, G., TAM, C. C., RODRIGUES, L. & LOPMAN, B. 2010. Prevalence and characteristics of asymptomatic norovirus infection in the community in England. *Epidemiology & Infection,* 138**,** 1454-1458.

PIRES, S. M., FISCHER-WALKER, C. L., LANATA, C. F., DEVLEESSCHAUWER, B., HALL, A. J., KIRK, M. D., DUARTE, A. S., BLACK, R. E. & ANGULO, F. J. 2015. Aetiology-specific estimates of the global and regional incidence and mortality of diarrhoeal diseases commonly transmitted through food. *PloS one,* 10**,** e0142927.

PLATTS-MILLS, J. A., BABJI, S., BODHIDATTA, L., GRATZ, J., HAQUE, R., HAVT, A., MCCORMICK, B. J., MCGRATH, M., OLORTEGUI, M. P. & SAMIE, A. 2015. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *The Lancet Global Health,* 3**,** e564-e575.

PLATTS-MILLS, J. A., LIU, J., ROGAWSKI, E. T., KABIR, F., LERTSETHTAKARN, P., SIGUAS, M., KHAN, S. S., PRAHARAJ, I., MUREI, A. & NSHAMA, R. 2018. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *The Lancet Global Health,* 6**,** e1309-e1318.

POLLOCK, L., BENNETT, A., JERE, K. C., DUBE, Q., MANDOLO, J., BAR-ZEEV, N., HEYDERMAN, R. S., CUNLIFFE, N. A. & ITURRIZA-GOMARA, M. 2018. Nonsecretor Histo–blood Group Antigen Phenotype Is Associated With Reduced Risk of Clinical Rotavirus Vaccine Failure in Malawian Infants. *Clinical Infectious Diseases*.

QI, R., HUANG, Y.-T., LIU, J.-W., SUN, Y., SUN, X.-F., HAN, H.-J., QIN, X.-R., ZHAO, M., WANG, L.-J. & LI, W. 2018. Global prevalence of asymptomatic norovirus infection: a meta-analysis. *EClinicalMedicine,* 2**,** 50-58.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

RAMANI, S. & GIRI, S. 2019. Influence of histo blood group antigen expression on susceptibility to enteric viruses and vaccines. *Current Opinion in Infectious Diseases*.

RCWG. 2017. *Rotavirus Classification Working Group* [Online]. Available: https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification/rcwg [Accessed].

REID, M. E., LOMAS-FRANCIS, C. & OLSSON, M. L. 2012. *The Blood Group Antigen Factsbook*, Academic Press.

RICHARDS, G. P., WATSON, M. A., FANKHAUSER, R. L. & MONROE, S. S. 2004a. Genogroup I and II noroviruses detected in stool samples by real-time reverse transcription-PCR using highly degenerate universal primers. *Applied and Environmental Microbiology,* 70**,** 7179-7184.

RICHARDS, G. P., WATSON, M. A., FANKHAUSER, R. L. & MONROE, S. S. 2004b. Genogroup I and II noroviruses detected in stool samples by real-time reverse transcription-PCR using highly degenerate universal primers. *Applied Environmental Microbiology.,* 70**,** 7179-7184.

RICHARDSON, S., GRIMWOOD, K., GORRELL, R., PALOMBO, E., BARNES, G. & BISHOP, R. 1998. Extended excretion of rotavirus after severe diarrhoea in young children. *The Lancet,* 351**,** 1844-1848.

ROBILOTTI, E., DERESINSKI, S. & PINSKY, B. A. 2015. Norovirus. *Clinical Microbiology Reviews,* 28**,** 134-164.

ROLLINS, N. C., NDIRANGU, J., BLAND, R. M., COUTSOUDIS, A., COOVADIA, H. M. & NEWELL, M.-L. 2013. Exclusive breastfeeding, diarrhoeal morbidity and all-cause mortality in infants of HIV-infected and HIV uninfected mothers: an intervention cohort study in KwaZulu Natal, South Africa. *PloS one,* 8.

ROTH, A. N. & KARST, S. M. 2016. Norovirus mechanisms of immune antagonism. *Current Opinion in Virology,* 16**,** 24-30.

RUIZ-PALACIOS, G. M., PÉREZ-SCHAEL, I., VELÁZQUEZ, F. R., ABATE, H., BREUER, T., CLEMENS, S. C., CHEUVART, B., ESPINOZA, F., GILLARD, P. & INNIS, B. L. 2006. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *New England Journal of Medicine,* 354**,** 11-22.

RUUSKA, T. & VESIKARI, T. 1990. Rotavirus disease in Finnish children: use of numerical scores for clinical severity of diarrhoeal episodes. *Scandinavian Journal of Infectious Diseases,* 22**,** 259-267.

SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* 4**,** 406-425.

SAKAI, Y., NAKATA, S., HONMA, S., TATSUMI, M., NUMATA-KINOSHITA, K. & CHIBA, S. 2001. Clinical severity of Norwalk virus and Sapporo virus gastroenteritis in children in Hokkaido, Japan. *The Pediatric Infectious Disease Journal,* 20**,** 849-853.

SANO, D., PÉREZ-SAUTU, U., GUIX, S., PINTÓ, R. M., MIURA, T., OKABE, S. & BOSCH, A. 2011. Quantification and genotyping of human sapoviruses in the Llobregat river catchment, Spain. *Applied and Environmental Microbiology,* 77**,** 1111-1114.

SARVESTANI, S. T., COTTON, B., FRITZLAR, S., O'DONNELL, T. B. & MACKENZIE, J. M. 2016. Norovirus Infection: Replication, Manipulation of Host, and Interaction with the Host Immune Response. *Journal of Interferon & Cytokine Research,* 36**,** 215-225.

SCHNADOWER, D., TARR, P. I., GORELICK, M. H., O'CONNELL, K., ROSKIND, C. G., POWELL, E. C., RAO, J., BHATT, S. & FREEDMAN, S. B. 2013. Validation of the modified Vesikari score in children with gastroenteritis in 5 US emergency departments. *Journal of Pediatric Gastroenterology and Nutrition,* 57**,** 514.

SCHWARZER, M., STRIGINI, M. & LEULIER, F. 2018. Gut microbiota and host juvenile growth. *Calcified Tissue International,* 102**,** 387-405.

SDIRI-LOULIZI, K., GHARBI-KHÉLIFI, H., DE ROUGEMONT, A., CHOUCHANE, S., SAKLY, N., AMBERT-BALAY, K., HASSINE, M., GUÉDICHE, M. N., AOUNI, M. & POTHIER, P. 2008. Acute infantile gastroenteritis associated with human enteric viruses in Tunisia. *Journal of Clinical Microbiology,* 46**,** 1349-1355.

SEAH, E. L., GUNESEKERE, I. C., MARSHALL, J. A. & WRIGHT, P. J. 1999. Variation in ORF3 of genogroup 2 Norwalk-like viruses. *Archives of Virology,* 144**,** 1007-1014.

SHIRATO, H. 2011. Norovirus and histo-blood group antigens. *Japan Journal of Infectious Disease,* 64**,** 95-103.

SHIRATO, H., OGAWA, S., ITO, H., SATO, T., KAMEYAMA, A., NARIMATSU, H., XIAOFAN, Z., MIYAMURA, T., WAKITA, T. & ISHII, K. 2008. Noroviruses distinguish between type 1 and type 2 histo-blood group antigens for binding. *Journal of Virology,* 82**,** 10756-10767.

SIMMONDS, M. K., ARMAH, G., ASMAH, R., BANERJEE, I., DAMANKA, S., ESONA, M., GENTSCH, J. R., GRAY, J. J., KIRKWOOD, C. & PAGE, N. 2008. New oligonucleotide primers for P-typing of rotavirus strains: Strategies for typing previously untypeable strains. *Journal of Clinical Virology,* 42**,** 368-373.

SLOGROVE, A., REIKIE, B., NAIDOO, S., DE BEER, C., HO, K., COTTON, M., BETTINGER, J., SPEERT, D., ESSER, M. & KOLLMANN, T. 2012. HIV-exposed uninfected infants are at increased risk for severe infections in the first year of life. *Journal of Tropical Pediatrics,* 58**,** 505-508.

SOMEYA, Y. 2002. Genomic organization of Norwalk-like viruses and functions of viral gene products. *Nihon rinsho. Japanese Journal of Clinical Medicine,* 60**,** 1155-1164.

TAN, E. M., CAWCUTT, K. A., ZOMOK, C. D., GO, R. S. & SIA, I. G. 2017. Activity of nitazoxanide against viral gastroenteritis: a systematic review. *International Journal of Travel Medicine and Global Health,* 5**,** 107-112.

TAN, M., HEGDE, R. S. & JIANG, X. 2004. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *Journal of Virology,* 78**,** 6233-6242.

TAN, M. & JIANG, X. 2011. Norovirus–host interaction: multi-selections by human histo-blood group antigens. *Trends in Microbiology,* 19**,** 382-388.

TATE, J. E., BURTON, A. H., BOSCHI-PINTO, C., PARASHAR, U. D., NETWORK, W. H. O. C. G. R. S., AGOCS, M., SERHAN, F., DE OLIVEIRA, L., MWENDA, J. M. & MIHIGO, R. 2016. Global, regional, and national estimates of rotavirus mortality in children< 5 years of age, 2000–2013. *Clinical Infectious Diseases,* 62**,** S96-S105.

TATE, J. E., BURTON, A. H., BOSCHI-PINTO, C., STEELE, A. D., DUQUE, J. & PARASHAR, U. D. 2012. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *The Lancet,* 12**,** 136-141.

TEUNIS, P., SUKHRIE, F., VENNEMA, H., BOGERMAN, J., BEERSMA, M. & KOOPMANS, M. 2015. Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiology & Infection,* 143**,** 1710-1717.

THEA, D. M., ST. LOUIS, M. E., ATIDO, U., KANJINGA, K., KEMBO, B., MATONDO, M., TSHIAMALA, T., KAMENGA, C., DAVACHI, F. & BROWN, C. 1993. A prospective study of diarrhea and HIV-1 infection among 429 Zairian infants. *New England Journal of Medicine,* 329**,** 1696-1702.

THORNE, L. G. & GOODFELLOW, I. G. 2014. Norovirus gene expression and replication. *Journal of General Virology,* 95**,** 278-291.

THORVEN, M., GRAHN, A., HEDLUND, K.-O., JOHANSSON, H., WAHLFRID, C., LARSON, G. & SVENSSON, L. 2005. A homozygous nonsense mutation (428G→ A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *Journal of Virology,* 79**,** 15351-15355.

TODD, S., PAGE, N. A., STEELE, A. D., PEENZE, I. & CUNLIFFE, N. A. 2010. Rotavirus strain types circulating in Africa: review of studies published during 1997–2006. *Journal of Infectious Diseases,* 202**,** S34-S42.

TRAN, T. H., TRAINOR, E., NAKAGOMI, T., CUNLIFFE, N. A. & NAKAGOMI, O. 2013. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII. 4 variants. *Journal of Clinical Virology,* 56**,** 269-277.

TREANOR, J. J., ATMAR, R. L., FREY, S. E., GORMLEY, R., CHEN, W. H., FERREIRA, J., GOODWIN, R., BORKOWSKI, A., CLEMENS, R. & MENDELMAN, P. M. 2014. A novel intramuscular bivalent norovirus virus-like particle vaccine candidate—reactogenicity, safety, and immunogenicity in a phase 1 trial in healthy adults. *The Journal of Infectious Diseases,* 210**,** 1763-1771.

TROEGER, C., FOROUZANFAR, M., RAO, P. C., KHALIL, I., BROWN, A., REINER JR, R. C., FULLMAN, N., THOMPSON, R. L., ABAJOBIR, A. & AHMED, M. 2017. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet,* 17**,** 909-948.

TROEGER, C., KHALIL, I. A., RAO, P. C., CAO, S., BLACKER, B. F., AHMED, T., ARMAH, G., BINES, J. E., BREWER, T. G. & COLOMBARA, D. V. 2018. Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years. *JAMA pediatrics,* 172**,** 958-965.

TROJNAR, E., SACHSENRÖDER, J., TWARDZIOK, S., REETZ, J., OTTO, P. H. & JOHNE, R. 2013. Identification of an avian group A rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses. *Journal of General Virology,* 94**,** 136-142.

UHNOO, I., WADELL, G., SVENSSON, L. & JOHANSSON, M. 1984. Importance of enteric adenoviruses 40 and 41 in acute gastroenteritis in infants and young children. *Journal of Clinical Microbiology,* 20**,** 365-372.

UNAIDS. 2019. *UNAIDS data 2019 | UNAIDS* [Online]. UNAIDS. Available: https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf [Accessed 12 November 2019 2019].

UNICEF 2019. Estimates of child cause of death, Diarrhoea 2018.

UNICEF, W. 2012. Pneumonia and diarrhoea-Tackling the deadliest diseases for the world's poorest children. *New York: UNICEF,* 2013.

UNICEF, W. 2016. Diarrhoea remains a leading killer of young children, despite the availability of a simple treatment solution. *UNICEF data: monitoring the situation of children and women, Updated: Jun*.

VAN BEEK, J., DE GRAAF, M., AL-HELLO, H., ALLEN, D. J., AMBERT-BALAY, K., BOTTELDOORN, N., BRYTTING, M., BUESA, J., CABRERIZO, M. & CHAN, M. 2018. Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network. *The Lancet,* 18**,** 545-553.

VERMEULEN, A. 2019. *RE: Genome characterisation of non-GII.4 norovirus strains using next-generation sequencing. MSc Medical Virology Dissertation, University of Pretoria.*

VINJÉ, J. 2015. Advances in laboratory methods for detection and typing of norovirus. *Journal of Clinical Microbiology,* 53**,** 373-381.

VINJÉ, J., ESTES, M. K., ESTEVES, P., GREEN, K. Y., KATAYAMA, K., KNOWLES, N. J., L'HOMME, Y., MARTELLA, V., VENNEMA, H. & WHITE, P. A. 2019. ICTV Virus Taxonomy Profile: Caliciviridae. *Journal of General Virology,* 100**,** 1469-1470.

VONGPUNSAWAD, S., PRASAD, B. V. & ESTES, M. K. 2013. Norwalk virus minor capsid protein VP2 associates within the VP1 shell domain. *Journal of Virology,* 87**,** 4818-4825.

WALKER, C. L. F., ARYEE, M. J., BOSCHI-PINTO, C. & BLACK, R. E. 2012. Estimating diarrhea mortality among young children in low and middle income countries. *PloS one,* 7**,** e29151.

WHO. 2017. *Number of people (all ages) living with HIV. Estimates by WHO region.* [Online]. Available: http://apps.who.int/gho/data/view.main.22100WHO?lang=en [Accessed 18/07 2017].

WIEGERING, V., KAISER, J., TAPPE, D., WEIßBRICH, B., MORBACH, H. & GIRSCHICK, H. J. 2011. Gastroenteritis in childhood: a retrospective study of 650 hospitalized pediatric patients. *International Journal of Infectious Diseases,* 15**,** e401-e407.

WINGFIELD, T., GALLIMORE, C. I., XERRY, J., GRAY, J. J., KLAPPER, P., GUIVER, M. & BLANCHARD, T. J. 2010. Chronic norovirus infection in an HIV-positive patient with persistent diarrhoea: a novel cause. *Journal of Clinical Virology,* 49**,** 219.

WOBUS, C. E., KARST, S. M., THACKRAY, L. B., CHANG, K.-O., SOSNOVTSEV, S. V., BELLIOT, G., KRUG, A., MACKENZIE, J. M., GREEN, K. Y. & VIRGIN IV, H. W. 2004. Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS biology,* 2.

XU, S., LIU, Y., TAN, M., ZHONG, W., ZHAO, D., JIANG, X. & KENNEDY, M. A. 2019. Molecular basis of P [6] and P [8] major human rotavirus VP8* domain interactions with histo-blood group antigens. *BioRxiv***,** 512301.

YAMASHITA, Y., OOTSUKA, Y., KONDO, R., OSETO, M., MIYAMOTO, T., UEDA, T., KONDO, H., TANAKA, T., WAKITA, T. & KATAYAMA, K. 2010. Molecular characterization of Sapovirus detected in a gastroenteritis outbreak at a wedding hall. *Journal of Medical Virology,* 82**,** 720-726.

YAN, H., YAGYU, F., OKITSU, S., NISHIO, O. & USHIJIMA, H. 2003. Detection of norovirus (GI, GII), Sapovirus and astrovirus in fecal samples using reverse transcription single-round multiplex PCR. *Journal of Virological Methods,* 114**,** 37-44.

YINDA, C. K., CONCEIÇÃO-NETO, N., ZELLER, M., HEYLEN, E., MAES, P., GHOGOMU, S. M., VAN RANST, M. & MATTHIJNSSENS, J. 2017. Novel highly divergent sapoviruses detected by metagenomics analysis in straw-colored fruit bats in Cameroon: Divergent bat sapoviruses. *Emerging Microbes & Infections,* 6**,** 1-7.

ZAHORSKY, J. 1929. Hyperemesis hiemis or the winter vomiting disease. *Arch Pediatr,* 46**,** 391-395.

ZHANG, S.-X., ZHOU, Y.-M., XU, W., TIAN, L.-G., CHEN, J.-X., CHEN, S.-H., DANG, Z.-S., GU, W.-P., YIN, J.-W. & SERRANO, E. 2016a. Impact of co-infections with enteric pathogens on

children suffering from acute diarrhea in southwest China. *Infectious Diseases of Poverty,* 5**,** 64.

ZHANG, X.-F., LONG, Y., TAN, M., ZHANG, T., HUANG, Q., JIANG, X., TAN, W.-F., LI, J.-D., HU, G.-F. & TANG, S. 2016b. P [8] and P [4] rotavirus infection associated with secretor phenotypes among children in South China. *Scientific Reports,* 6**,** 34591.

ZHENG, D.-P., ANDO, T., FANKHAUSER, R. L., BEARD, R. S., GLASS, R. I. & MONROE, S. S. 2006. Norovirus classification and proposed strain nomenclature. *Virology,* 346**,** 312-323.

ZIVKOVIC, A. M., GERMAN, J. B., LEBRILLA, C. B. & MILLS, D. A. 2011. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proceedings of the National Academy of Sciences,* 108**,** 4653-4658.

Appendices

Appendix A: Study questionnaires

Appendix A1: Patient information

PATIENT OR PARTICIPANT'S INFORMATION & INFORMED CONSENT DOCUMENT: FIRST VISIT OR RE-ADMISSIONS

Researcher's name: **Dr Marieke Brauer and Dr Janet Mans**

Department of Medical Virology

University of Pretoria

Dear Parent / Legal guardian

MOLECULAR EPIDEMIOLOGY AND CHARACTERISATION OF NOROVIRUS INFECTIONS IN HIV-INFECTED AND HIV-EXPOSED UNINFECTED CHILDREN IN GAUTENG, SOUTH AFRICA

We are researchers in the Department of Medical Virology, University of Pretoria. You are invited to volunteer to participate in our research project on **Molecular epidemiology and characterisation of norovirus infections in HIV-infected and HIV-exposed uninfected children in Gauteng, South Africa.**

This letter gives information to help you to decide if you want your child to take part in this study. Before you agree you should fully understand what is involved. If you do not understand the information or have any other questions, do not hesitate to ask us. You should not agree to take part unless you are completely happy about what we expect of you.

The purpose of the study is to test for the germ called norovirus. This germ infects people of all ages and causes a runny tummy (diarrhoea). Many children in South Africa become sick with this germ every year and many of these children will need to be in hospital on an intravenous drip. Some researchers have found that people with poor immune systems (such as those who are infected with a germ called human immunodeficiency virus or HIV) may have more severe norovirus infections, or may carry the norovirus germ for longer periods than other people. Because of these potential complications in HIV-infected people, we would like to count the number of HIV-infected and HIV-exposed children who get sick from this germ norovirus, to compare with those who are HIV negative. It is known that a few people are seldom infected with the norovirus germ if a certain gene does not

work in their bodies. We would also like to test for this gene in the children who participate in the study. We are talking with you about this project because your child has been admitted to the hospital with diarrhoea which may be caused by this germ. This information will be used to try to stop these germs causing illness in other children.

We would like you to complete a questionnaire. This may take about 15 to 20 minutes. A person (the surveillance officer) will complete the questionnaire with you before you leave the ward / the clinic. It will be kept in a safe place to ensure confidentiality. We will not write your name or your child's name on the questionnaire. This will ensure confidentiality.

Some of the questions in the questionnaire may be of a sensitive nature and ask about your child's HIV status. If you feel uncomfortable about some questions, you do not need to answer them.

The Research Ethics Committee of the University of Pretoria, Faculty of Health Sciences, telephone numbers 012 3541677 / 012 3541330, granted written approval for this study.

Your participation in this study is voluntary. You can refuse to participate or stop at any time without giving any reason. As we do not write your name on the questionnaire, you give us the information anonymously. Once we have completed the questionnaire with you, you cannot recall your consent. We will not be able to trace your information. Therefore, your child will also not be identified as a participant in any publication that comes from this study.

In the event of questions asked, which will cause emotional distress, then the researcher is able to refer you to a competent counselling.

Note: The implication of completing the questionnaire is that informed consent has been obtained from you. Thus, any information derived from your form (which will be totally anonymous) may be used for e.g. publication, by the researchers.

We sincerely appreciate your help.

Yours truly,

Dr Marieke Brauer and Dr Janet Mans

Patient Questionnaire and Data Capture Sheet

Was the participant enrolled in this particular study during a previous admission for diarrhoea?

☐ Y ☐N *If "Yes", please use previously allocated study number, if "No" allocate new study number*

Unique study number: _____

Please document the unique study number in the child's Road-to-Health Booklet.

Hospital information

Date of admission: (dd/mm/yyyy) ___ / ___ / _____

Hospital Number: _____

Date of interview: (dd/mm/yyyy) ___ / ___ / _____

Ward: _____

What is the interviewee's relationship to the participant?

☐ Mother ☐ Father ☐ Legal guardian ☐ Other (Specify): _____

Check list for surveillance:

☐ Y ☐N Consent form for questionnaire signed?

☐ Y ☐N Consent form for stool specimen collection signed?

☐ Y ☐N Consent form for blood specimen collection signed?

☐ Y ☐N Does patient meet diarrhoea case definition (see checklist)?

☐ Y ☐N Questionnaire completed?

☐ Y ☐N Stools taken?

☐ Y ☐N Blood specimen taken?

DEMOGRAPHIC DATA:

1. Date of birth DOB: (dd/mm/yyyy) ___ / ___ / _____ ☐ unknown

   If DOB unknown: Age in months: _____

2. Gender: ☐ Male ☐ Female

3. Race: ☐Asian ☐Black ☐Coloured ☐White ☐ Other

4. Suburb/area of residence _____

5. Where do you get water from at home?

☐Indoor tap ☐ Outdoor tap ☐ Communal tap

☐Rain water tank ☐ Well ☐ River

6. What type of toilet do you have at home?

☐ In the home ☐ Communal

☐ Flush toilet ☐ Pit latrine ☐No facilities

CLINICAL DETAILS:

SIGNS AND SYMPTOMS

1. Date of onset of symptoms (beginning of diarrhoea/vomiting/fever): (dd/mm/yyyy) ___ / ___ / _____

2.1 **Signs and symptoms on admission:**

Sign/symptom	**Yes**	**No**	**Unknown**	**Number in 24hrs**	**Maximum duration (days)**
Diarrhoea					
If yes Watery diarrhoea				---	---
Bloody diarrhoea				---	---
Vomiting					
Severe abdominal pain					

General danger symptoms, signs & assessment of dehydration	**Yes**	**No**	**Unknown**	**Duration (in days)**

Unable to drink or drinking poorly				
Vomits everything				
Sunken eyes				
Sunken fontanelle (if applicable)				
Decreased skin turgor (> 2 seconds)				
Lethargy/listless/moves only when stimulated				
Symptoms of shock: capillary refill, heart rate				
Unconsciousness				
Medical doctor evaluated level of dehydration <5%, 5-10% >10%				

2.2 Max recorded temperature > 38°C: ☐ Y ☐N ☐ Unknown

If yes, record temperature: _____°C (##. #°C)

How the temperature was measured: ☐ Axillary ☐ Rectal ☐ Oral

Number of days with fever: _____

Did the child have any convulsions? ☐ Y ☐N ☐ Unknown

Weight: _____ kg (##. #kg)

Height: _____cm

PREVIOUS HOSPITALISATIONS:

1. Diarrhoea admissions in last 3 months? ☐ Y ☐N ☐ Unknown
2. If yes, number of admissions _____(##)
3. Date of last diarrhoeal admission (dd/mm/yyy)___/___/_____
4. Date of last diarrhoeal clinic visit (dd/mm/yyy)___/___/_____

VACCINATION HISTORY:

***ROTAVIRUS* VACCINATION** *(Ask to see the road to health card.)*

1. Has the child ever received the vaccine against **rotavirus**?

☐ Y ☐N ☐ Unknown

Dose	**Dose given?**	**Date given (dd/mm/yyyy) if exact date unknown give month and year**
6 weeks	☐ Y ☐N ☐ Unknown	
14 weeks	☐ Y ☐N ☐ Unknown	

Source of vaccination status information:

☐ RTHC card seen by surveillance officer ☐ Drs notes from RTHC

☐ Verbal report from caregiver ☐ Drs notes from verbal report

☐ Directly from clinic ☐ other (specify): _____

FEEDING:

Is the child breast feeding? ☐ Y ☐N ☐ Unknown

If yes, how is child fed? Breastfed only ☐ Y ☐N ☐ Unknown

Breastfed + Other ☐ Y ☐N ☐ Unknown

If no, how is the child fed? Bottle only ☐ Y ☐N ☐ Unknown

Bottle + water ☐ Y ☐N ☐ Unknown

Mixed diet ☐ Y ☐N ☐ Unknown

Is the water which is used to prepare the bottle boiled before use?

☐ Y ☐N ☐ Unknown

PATIENT MANAGEMENT:

Treatment in hospital

Oral rehydration therapy: ☐ Y ☐N

IVI fluids: ☐ Y ☐N

Antibiotics: ☐ Y ☐N

FINAL OUTCOME OF PATIENT:

HOSPITAL PATIENT:	**CLINIC PATIENT:**
☐ Died	Died in clinic: ☐Y ☐N Date of death (dd/mm/yyy)___/___/_____
☐ Discharged	☐ Discharged
☐ Absconded /RHT	☐ Referred to HOSPITAL, if so name of facility;_____
Date of discharge/death/RHT /absconded/referred: (dd/mm/yyyy) ___ / ___ / ____	

DATA ON HIV STATUS:

1. Prior diagnosis of HIV infection: ☐ Y ☐N

   1.1. If yes, date diagnosed (dd/mm/yyyy) ___ / ___ / _____

   1.2. How was diagnosis made?

   1.2.1. PCR ☐ Positive ☐Negative ☐Not done

   1.2.2. ELISA ☐ Positive ☐Negative ☐Not done

2. Current antiretroviral use for HIV treatment: ☐ Y ☐N ☐ Unknown

   If yes, date initiated (dd/mm/yyyy) ___ / ___ / _____

3. CD4 count _____ Date determined (dd/mm/yyyy) ___ / ___ / _____

4. HIV viral load _____ Date determined (dd/mm/yyyy) ___ / ___ / _____

5. New diagnosis of HIV infection: ☐ Positive ☐Negative

   5.1 PCR ☐ Positive ☐Negative ☐Not done

   5.2 ELISA ☐ Positive ☐Negative ☐Not done

6. HIV Exposure:

   ☐ HIV-negative exposed *(mother is HIV-seropositive)*

   ☐HIV-negative unexposed *(mother is HIV negative)*

Appendix A2: Consent form

PARENT OR GUARDIAN INFORMATION & INFORMED CONSENT DOCUMENT

TITLE OF STUDY: Molecular epidemiology and characterisation or norovirus infections in HIV-infected and HIV-exposed uninfected children in Gauteng, South Africa

Principal Investigators:

Dr Janet Mans

Dr Marieke Brauer

Institution:

University of Pretoria

DAYTIME AND AFTER HOURS TELEPHONE NUMBER(S):

Daytime numbers: 012 319 2618/60

Afterhours: 084 526 0581 or 072 091 6500

DATE AND TIME OF FIRST INFORMED CONSENT DISCUSSION:

dd	**mmm**	**ivy**

:
Time

One copy will stay with the surveillance officer, and be filed with the clinical case report form. The other copy will be given to the parent or legal guardian.

Dear Parent/Legal guardian

1) INTRODUCTION

Hello. My name is _____ (name of surveillance officer) and I would like to ask you for some of your time to explain the work that we do and to invite you and your child to please assist us by participating in a research study. This information leaflet will help you to decide if you want to participate. Before you agree to take part you should fully understand what is involved. If you have any questions that this leaflet does not fully explain, please do not hesitate to ask the investigator / surveillance officer.

2) THE NATURE AND PURPOSE OF THIS STUDY

In South Africa (like elsewhere in the world), laboratories and government health departments look at the germs that cause disease and where these germs occur. For this study we are interested in the germ called norovirus. This germ infects people of all ages and causes a runny tummy (diarrhoea). Many children in South Africa become sick with this germ every year and many of these children will need to be in hospital on an intravenous drip. Some researchers have found that people with poor immune systems (such as those who are infected with a germ called human immunodeficiency virus or HIV) may have more severe norovirus infections, or may carry the norovirus germ for longer periods than other people. Because of these potential complications in HIV-infected people, we would like to count the number of HIV-infected and HIV-exposed children who get sick from this germ norovirus, to compare with those who are HIV negative. It is known that a few people are seldom infected with the norovirus germ if a certain gene does not work in their bodies. We would also like to test for this gene in the children who participate in the study. We are talking with you about this project because your child has been admitted to the hospital with diarrhoea which may be caused by this germ. This information will be used to try to stop these germs causing illness in other children.

3) EXPLANATION OF PROCEDURES TO BE FOLLOWED

If you choose to have your child join the project, we will ask you a number of questions as part of the study. During the study, we will ask about your child's present and past history, including questions about vaccines s/he received. We will also ask questions about the child's home and other children living with your child. You may choose not to answer any question. We will also look at your child's hospital records and his/ her records in the laboratory to complete the form.

In addition, you will be asked if you will allow us to take a stool specimen from your child to test for the germs which cause diarrhoea, as part of the study. The stool specimen will be taken from the child's nappy or bed pan. If the stool is very watery a piece of the nappy will be cut or the whole nappy will be taken and sent to the laboratory for detection of the germs. We will also need to take additional stool specimens at follow-up visits (see section 4 below).

Also, as part of the study, you will be asked if you are willing for us to take a specimen from your child to be tested for HIV infection. This will be done taking a small tube of blood by inserting a needle into your child's veins to collect the blood. The results of your HIV test will not be known by me or the other study nurses. If you want to know your HIV result we can refer you to your local VCT clinic for testing and follow up.

We also need to take two specimens to test for the gene that does not work in people who are seldom infected with the norovirus germ. The first specimen will be a small tube of blood to be taken by inserting a needle into your child's veins to collect the blood. If permission was also given for an HIV test, this blood specimen will be collected at the same time as the tube for HIV testing. The second specimen will be a small amount of spit, which is taken by gently inserting a swab into your child's mouth.

If you agree to the questionnaire but not to take a stool, blood or saliva specimen, this is also okay.

4) EXPECTED DURATION OF PARTICIPATION

Completing the form with questions should take about 15-20 minutes. Taking the stool specimen will take about 2 minutes. Taking the blood sample should take only about 5-10 minutes. Taking the saliva specimen should take about a minute. If the child is not having diarrhoea right now we may need to come back to collect the stool specimen. We will also visit your child regularly while s/he stays in hospital to collect information about his/ her hospital stay. S/he will be visited only while in hospital.

We will also ask you to return to the hospital's outpatients department for follow-up visits. At each follow-up visit we will complete another form with a few questions (which should take about 5 minutes) and we will collect another stool specimen, regardless of whether or not your child has diarrhoea at the time of the visit. You will be asked to return with your child 6 weeks after the first visit to the hospital, and after that every 3 months. You may choose to allow your child to participate in the study during the current admission but not in any follow-up visits if you so wish.

5) RISK AND DISCOMFORT INVOLVED

There should be no risks to you or your child if you agree to take part in the study. The questions will be asked are general and not personal. The collection of blood and saliva may cause mild discomfort. The questionnaire may take some of your time to complete.

6) POSSIBLE BENEFITS OF THIS STUDY

Your child will not directly benefit from this study. His/ her treatment in the hospital will not change if you agree that he/ she may take part in the study. The information we collect will be used to try to stop the germ that caused his/ her illness from causing illness in other children. You may ask the study nurse questions about your child's health and progress in hospital which may make you feel more informed.

7) WHAT ARE YOUR RIGHTS AS A PARTICIPANT?

It is your right to choose to have your child be part or not be a part of the project. If your child does not join, s/he will continue to receive the treatment needed for this infection and s/he will not lose any health care services. You may choose to leave the study at any time you like. If you decide this, your child will be treated in the same way in hospital. He/ she will not lose any benefits to which he/ she may be entitled to if you stop participation in the study.

8) STORAGE OF SAMPLES FOR FUTURE TESTING

Once we have completed the testing for this study we may store your child's sample for testing in the future. These tests would be done if new viruses (germs) are identified or if new test become available. The sample will not have a name on it so that it cannot be linked to you or your child.

9) HAS THE STUDY RECEIVED ETHICAL APPROVAL?

This study has received written approval from the Research Ethics Committee of the Faculty of Health Sciences at the University of Pretoria, telephone numbers 012 3541677 / 012 3541330.

10) INFORMATION AND CONTACT PERSON

If you have any questions about the study, about your child's rights while taking part in the study or about any injury that may have happened because of the study, you may contact any of the people listed below:

Janet Mans: **012 319 2660 or 072 091 6500**

Marieke Brauer: **012 319 2618 or 084 526 0581**

11) COMPENSATION

You or your child will not have to pay any costs, if s/he takes part in the study. Your participation is voluntary. A contribution of R50 towards your transport expenses will be given for your participation if we ask you return to the hospital to give follow-up specimens for the purpose of this study.

12) CONFIDENTIALITY

All information that you give will be kept strictly confidential. The forms with your personal information will be posted in sealed envelopes to other people who work on the study and will always be kept in locked cabinets or offices. All your information from this study will be put under a special number (study number) and not your name. Therefore, none of the study personnel will know which information/ HIV result belongs to you. Once we have analysed the information no one will be able to identify you. Research reports and articles in scientific journals will not include any information that may identify you. You will receive a piece of paper with your unique study number and the date of your follow-up visit (if you have agreed to this), which you will need to bring with you to the follow-up visit.

CONSENT TO PARTICIPATE IN THIS STUDY

I confirm that the person asking my consent for my child to take part in this study has told me about nature, process, risks, discomforts and benefits of the study. I have also received, read and understood the above written information (Information and Informed Consent) regarding the study. I am aware that the results of the study, including personal details, will be anonymously processed into research reports. I am participating willingly. I have had time to ask questions and have no objection to my child's participation in the study. I understand that there is no penalty should I wish to discontinue with the study and my withdrawal will

not affect any treatment / access to treatment in any way.

I, _____ (name of parent/guardian), acknowledge that the study questionnaire and specimen collection have all been explained to me and that I agree for my child to participate in the following study procedures:

☐ Y ☐ N I agree to be interviewed and allow data from my child hospital records to be collected to answer the questions from the study questionnaire.

☐ Y ☐ N I agree to the test for germs from his/her stools

☐ Y ☐ N I agree to his/her stool sample(s) being stored for future testing.

☐ Y ☐ N I agree to the test for HIV on his/her blood specimen.

☐ Y ☐ N I agree to the test for the gene that does not work in people who are less likely to become infected with the norovirus germ on his/her blood and saliva specimens.

☐ Y ☐ N I agree to bring my child for follow-up visits where another questionnaire will be completed and further stool specimens will be collected to be tested for germs.

I have received a signed copy of this informed consent agreement.

The signature of the parent or legal guardian below means that the study has been explained to the parent or legal guardian and that he/ she agrees that his/ her child or ward may participate.

Participant's name …….....(Please print)

Parent or legal guardian's name.....(Please print)

Parent or legal guardian's signature: .....…… Date.....

Investigator/Surveillance officer's name..... (Please print)

Investigator/Surveillance officer's signature .....….. Date.….....

The signature of the witness below means that another person has observed the consenting of the parent or legal guardian. The witness must be impartial and not part of the study staff.

Witness's Name .....…….....(Please print)

Witness's signature .....…….. Date.….....

VERBAL INFORMED CONSENT

I, the undersigned, have read and have fully explained the participant information leaflet, which explains the nature, process, risks, discomforts and benefits of the study to the participant's parent or legal guardian whom I have asked to participate in the study.

The participant's parent or legal guardian indicates that s/he understands that the results of the study, including personal details regarding the interview will be anonymously processed into a research report. The participant's parent or legal guardian indicates that s/he has had time to

ask questions and has no objection to participate in the interview. S/he understands that there is no penalty should s/he wish to discontinue with the study and his/her withdrawal will not affect any treatment / access to treatment in any way. I hereby certify that the parent or legal guardian has agreed to his/her child's participation in this study.

I, the undersigned, acknowledge that the study questionnaire and specimen collection have all been explained to the participant's parent or legal guardian and that s/he agrees for his/her child to participate in the following study procedures:

☐ Y ☐ N S/he agrees to be interviewed and allow data from his/her child's hospital records to be collected to answer the questions from the study questionnaire.

☐ Y ☐ N S/he agrees to the test for germs from the child's stools

☐ Y ☐ N S/he agrees to a stool sample being stored for future testing.

☐ Y ☐ N S/he agrees to the test for HIV on the child's blood specimen.

☐ Y ☐ N to the test for the gene that does not work in people who are less likely to become infected with the norovirus germ on the child's blood and saliva specimens.

☐ Y ☐ N S/he agrees to bring his/her child for follow-up visits where another questionnaire will be completed and further stool specimens will be collected to be tested for germs.

Participant's Name .....(Please print)

Parent or legal guardian's name.....(Please print)

Person seeking consent .....(Please print)

Signature .....…….....Date**.....**

The signature of the witness below means that another person has observed the consenting of the parent or legal guardian. The witness must be impartial and not part of the study staff.

Witness's name .....……..….....(Please print)

Signature .....……Date….....

Appendix A3: Follow-up questionnaire

PATIENT OR PARTICIPANT'S INFORMATION & INFORMED CONSENT DOCUMENT: FOLLOW-UP VISITS

Researcher's name: **Dr Marieke Brauer and Dr Janet Mans**

Department of Medical Virology

University of Pretoria

Dear Parent / Legal guardian

MOLECULAR EPIDEMIOLOGY AND CHARACTERISATION OF NOROVIRUS INFECTIONS IN HIV-INFECTED AND HIV-EXPOSED UNINFECTED CHILDREN IN GAUTENG, SOUTH AFRICA

We are researchers in the Department of Medical Virology, University of Pretoria. You are invited to volunteer to participate in our research project on **Molecular epidemiology and characterisation of norovirus infections in HIV-infected and HIV-exposed uninfected children in Gauteng, South Africa.**

This letter gives information to help you to decide if you want your child to take part in this study. Before you agree you should fully understand what is involved. If you do not understand the information or have any other questions, do not hesitate to ask us. You should not agree to take part unless you are completely happy about what we expect of you.

The purpose of the study is to test for the germ called norovirus. This germ infects people of all ages and causes a runny tummy (diarrhoea). Many children in South Africa become sick with this germ every year and many of these children will need to be in hospital on an intravenous drip. Some researchers have found that people with poor immune systems (such as those who are infected with a germ called human immunodeficiency virus or HIV) may have more severe norovirus infections, or may carry the norovirus germ for longer periods than other people. Because of these potential complications in HIV-infected people, we would like to count the number of HIV-infected and HIV-exposed children who get sick from this germ norovirus, to compare with those who are HIV negative. It is known that a few people are seldom infected with the norovirus germ if a certain gene does not work in their bodies. We would also like to test for this gene in the children who participate in the study. We are talking with you about this project because your child has been admitted to the hospital with diarrhoea which may be caused by this germ. This information will be used to try to stop these germs causing illness in other children.

We would like you to complete a follow-up questionnaire. This may take about 5 minutes. A person (the surveillance officer) will complete the questionnaire with you before you leave the ward / the clinic. It will be kept in a safe place to ensure confidentiality. We will not write your name or your child's name on the questionnaire. This will ensure confidentiality.

Some of the questions in the questionnaire may be of a sensitive nature and ask about your child's HIV status. If you feel uncomfortable about some questions, you do not need to answer them.

The Research Ethics Committee of the University of Pretoria, Faculty of Health Sciences, telephone numbers 012 3541677 / 012 3541330, granted written approval for this study.

Your participation in this study is voluntary. You can refuse to participate or stop at any time without giving any reason. As we do not write your name on the questionnaire, you give us the information anonymously. Once we have completed the questionnaire with you, you cannot recall your consent. We will not be able to trace your information. Therefore, your child will also not be identified as a participant in any publication that comes from this study.

In the event of questions asked, which will cause emotional distress, then the researcher is able to refer you to a competent counselling.

<u>Note:</u> The implication of completing the questionnaire is that informed consent has been obtained from you. Thus any information derived from your form (which will be totally anonymous) may be used for e.g. publication, by the researchers.

We sincerely appreciate your help.

Yours truly,
Dr Marieke Brauer and Dr Janet Mans

Follow-Up Patient Questionnaire and Data Capture Sheet

Unique study number: _____ *(Use previously allocated study number as documented on Road-to-Health Booklet)*

Hospital information

Date of follow-up visit: (dd/mm/yyyy) ___ / ___ / _____

Hospital Number: _____

Date of interview: (dd/mm/yyyy) ___ / ___ / _____

Ward: _____

What is the interviewee's relationship to the participant?

☐ Mother ☐ Father ☐ Legal guardian ☐ Other (Specify): _____

Check list for surveillance officer:

☐ Y ☐N Questionnaire completed?

☐ Y ☐N Stools taken?

CLINICAL DETAILS:

SIGNS AND SYMPTOMS

2. Does the child still have ongoing diarrhoea since the last visit? ☐ Y ☐N
   (If "Yes", please complete section 2.1)

   Does the child have diarrhoea at this moment? ☐ Y ☐N

   If answered "Yes", what is the date of onset of symptoms (beginning of diarrhoea):

   (dd/mm/yyyy) ___ / ___ / _____ *(also complete section 2.1)*

2.3 **Signs and symptoms on visit:**

Sign/symptom	**Yes**	**No**	**Unknown**	**Number in 24hrs**	**Maximum duration (days)**
Diarrhoea					
If yes Watery diarrhoea				---	---

Bloody diarrhoea				---	---
Vomiting					
Severe abdominal pain					

General danger signs & assessment of dehydration	**Yes**	**No**	**Unknown**	**Duration (days)**
Unable to drink or drinking poorly				---
Vomits everything				---
Sunken eyes				---
Sunken fontanelle if applicable				---
Decreased skin turgor (> 2 seconds)				---
Lethargy/listless/moves only when stimulated				
Symptoms of shock: capillary refill, heart rate				
Unconsciousness				
Medical doctor evaluated level of dehydration <5%, 5-10% >10%				

PLEASE NOTE: KINDLY REFER THE PATIENT FOR ASSESSMENT BY A MEDICAL DOCTOR OR NURSE AT CASUALTY IF THE CHILD HAS ANY GENERAL DANGER SIGNS OF DEHYDRATION

FEEDING:

Has the child's feeding changed since the last visit? ☐ Y ☐N

If "Yes", please complete the questions below on feeding:

Is the child breast feeding? ☐ Y ☐N ☐ Unknown

If yes, how is child fed? Breastfed only ☐ Y ☐N ☐ Unknown

Breastfed + Other ☐ Y ☐N ☐ Unknown

If no, how is the child fed? Bottle only ☐ Y ☐N ☐ Unknown

Bottle + water ☐ Y ☐N ☐ Unknown

Mixed diet ☐ Y ☐N ☐ Unknown

Is the water which is used to prepare the bottle boiled before use?

☐ Y ☐N ☐ Unknown

PATIENT MANAGEMENT:

Current treatment:

Oral rehydration therapy: ☐ Y ☐N

Antibiotics: ☐ Y ☐N

Antiretroviral therapy (for HIV): ☐ Y ☐N

Appendix B: Ethical approval

Appendix B1: Specimen collection

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 20 Oct 2016.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 22/04/2017.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee

27/08/2015

Approval Certificate
New Application

Ethics Reference No.: 362/2015

Title: Molecular epidemiology and characterisation of norovirus infections in HIV-infected, HIV-exposed uninfected and HIV-unexposed children with gastroenteritis in Gauteng, South Africa

Dear Marieke Brauer

The **New Application** as supported by documents specified in your cover letter dated 19/08/2015 for your research received on the 19/08/2015, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 26/08/2015.

Please note the following about your ethics approval:
- Ethics Approval is valid for 3 years
- Please remember to use your protocol number (**362/2015**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the receipt of 6 monthly written Progress Reports, and
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

Additional Conditions:

- The REC notes that parents/caregivers/legal guardians will give consent according to the DOH's new guideline.

We wish you the best with your research.

Yours sincerely

Dr R Sommers, MBChB; MMed (Int); MPharMed.
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes 2004 (Department of Health).

☎ 012 354 1677 🖷 0866516047 deepeka.behari@up.ac.za http://www.healthethics-up.co.za
✉ Private Bag X323, Arcadia, 0007 - 31 Bophelo Road, HW Snyman South Building, Level 2, Room 2.33, Gezina, Pretoria

Appendix B2: 2017 Ethical approval

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee

30/03/2017

Approval Certificate
New Application

Ethics Reference No.: 90/2017

Title: Determination of secretor genotype and saliva carbohydrate phenotype of children with gastroenteritis

Dear Esmari Rossouw

The **New Application** as supported by documents specified in your cover letter dated 13/03/2017 for your research received on the 13/03/2017, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 29/03/2017.

Please note the following about your ethics approval:
- Ethics Approval is valid for 1 year
- Please remember to use your protocol number (**90/2017**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the receipt of **6 monthly written Progress Reports**, and
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers; MBChB; MMed (Int); MPharMed,PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health).

☎ 012 356 3084 deepeka.behari@up.ac.za / fhsethics@up.ac.za http://www.up.ac.za/healthethics
✉ Private Bag X323, Arcadia, 0007 - Tswelopele Building, Level 4, Room 60, Gezina, Pretoria

Appendix B3: 2018 Ethical approval

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee

31/05/2018

Approval Certificate
New Application

Ethics Reference No: 182/2018

Title: The molecular epidemiology and diversity of gastroenteritis viruses in HIV-infected, -exposed and -unexposed children under the age of five years in Pretoria, South Africa.

Dear Esmari Rossouw

The **New Application** as supported by documents specified in your cover letter dated 17/04/2018 for your research received on the 17/04/2018, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 30/05/2018.

Please note the following about your ethics approval:
- Ethics Approval is valid for 2 years
- Please remember to use your protocol number (**182/2018**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the receipt of **6 monthly written Progress Reports**, and
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers; MBChB; MMed (Int); MPharMed,PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health).

☎ 012 356 3084 deepeka.behari@up.ac.za / fhsethics@up.ac.za http://www.up.ac.za/healthethics
✉ Private Bag X323, Arcadia, 0007 - Tswelopele Building, Level 4, Room 60 / 61, 31 Bophelo Road, Gezina, Pretoria

Appendix B4: 2019 Ethical renewal

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

10 April 2019

Approval Certificate
Annual Renewal

Ethics Reference No.: 182/2018
Title: The molecular epidemiology and diversity of gastroenteritis viruses in HIV-infected, -exposed and -unexposed children under the age of five years in Pretoria, South Africa

Dear Miss E Rossouw

The **Annual Renewal** as supported by documents received between 2019-03-19 and 2019-04-10 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 2019-04-10.

Please note the following about your ethics approval:

- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2020-04-10.
- Please remember to use your protocol number (182/2018) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers
MBChB MMed (Int) MPharmMed PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag X323
Arcadia 0007, South Africa
Tel +27 (0)12 356 3084
Email deepeka.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo

Appendix C: All co-infections detected in this study

Table C1: Viral co-infections in children (< 5 years), hospitalised with gastroenteritis at KPTH between July 2016 and December 2017.

Virus	Co-infection type
Norovirus GI	None
Norovirus GII	1 with rotavirus 1 with adenovirus 1 with sapovirus
Sapovirus	1 with rotavirus 1 with adenovirus 1 with Norovirus GII
Rotavirus	2 with adenovirus 1 with norovirus GII 2 with astrovirus 1 with sapovirus
Adenovirus	2 with rotavirus 1 with sapovirus 1 with norovirus GII
Astrovirus	2 with rotavirus

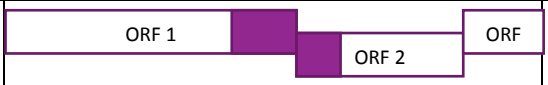
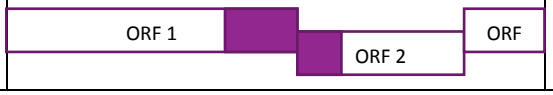
Table C2: Bacterial and parasite infections with or without viral co-infection.

Sample Nr	Parasite	Bacteria	Virus
NS0026	*Ascaris lumbricoides*	Not isolated	rotavirus
NS0040	*Giardia lamblia*	Not isolated	-
NS0128	*Cryptosporidium parvum*	Not isolated	-
NS0163	*Cryptosporidium Sp*	Not isolated	-
NS0176	*Entamoeba*	Not isolated	norovirus G2
NS0052	Not observed	*E.coli*	rotavirus
NS0087	Not observed	*Salmonella Group C*	-
NS0091	Not observed	*Salmonella Group D*	-
NS0096	Not observed	*Salmonella Group D*	-
NS0106	Not observed	*E.coli*	-
NS0122	Not observed	*Shigella Sonei*	-
NS0141	Not observed	*Salmonella*	-
NS0159	Not observed	*E.coli*	adenovirus-F
NS0192	Not observed	*E.coli*	astrovirus, rotavirus
NS0218	Not observed	*Shigella flexneri*	-

Appendix D: All norovirus strains region amplified and product size

Table D1: The norovirus genome regions that were characterised for the 27/32 norovirus strains detected in 205 children under the age of 5 years who were hospitalised with gastroenteritis between July 2016 and December 2017 at KPTH.

Specimen ID	Region amplified	Product size (bp)	Genotype	Closely related strain (Accession nr)
NS0010		676 bp (cloned)		South Africa 2012 (KP784697.1)
NS0013		383 bp		Taizhou 2018 (MH842245.1)
NS0017		398 bp		South Africa 2012 (KP784697.1)
NS0030		538 bp		Brazil 2013 (MG023178.1)
NS0038		676 bp (cloned)		South Korea 2019 (MN461113.1)
NS0046		430 bp		United Kingdom 2014 (MH218718.1)
NS0050		540 bp		Thailand 2011 (MK396776.1)
NS0053		586 bp		United States 2019 (MT031988.1)
NS0062		363 bp		Italy 2008 (MH279830.1)
NS0063		569 bp		Indonesia 2015 (MK408523.1)
NS0067		533 bp		United States 2017 (MK762639.1)
NS0068		559 bp		Indonesia 2015 (MK408523.1)
NS0069		540 bp		South Africa 2012 (KP784697.1)
NS0114		948 bp		United states 2016 (MK753029.1)

NS0119		515 bp		United States 2016 (MK753031.1)
NS0126		554 bp		South Africa 2012 (KP784697.1)
NS0138		908 bp		South Africa 2010 (KC962458.2)
NS0156		973 bp		United states 2018 (MT028542.1)
NS0157		580 bp		Brazil 2014 (MH271655.1)
NS0158		371 bp		South Africa 2012 (KJ710246.1)
NS0164		568 bp		United states 2016 (MK753031.1)
NS0173		471 bp		Japan 2012 (AB972484.1)
NS0176		975 bp		Cameroon 2013 (MN294769.1)
NS0208		477 bp		South Africa 2012 (KJ710246.1)
NS0211		319 bp		Cameroon 2012 (MN294768.1)
NS0215		590 bp		Brazil 2013 (MG023178.1)
NS0220		493 bp		Cameroon 2012 (MN294769.1)

Appendix E: Comparison of FUT 2 genotype and HBGA phenotype

Table E1: All specimens with secretor status, HBGA results and saliva information

Sample Name	Secretor Status	Lectin results	Breastfed?	Saliva colour	HBGA Phenotype
NS0001	Secretor	Positive	Yes	Normal	A, Le(b), Le(y)
NS0002	Secretor	Positive	No	Slightly pink	A, Le(a), Le(b), Le(y)
NS0003	Non-secretor	negative	Yes	Slightly pink	Le(b)
NS0004	Secretor	Positive	No	Normal	A, H1, B, Le(b), Le(y)
NS0005	Secretor*	Positive	No	Empty	A, Le(b), Le(y)

NS0006	Secretor	Unknown	Yes	Diluted	B, Le(b), Le(x), Le(y)
NS0007	Secretor	negative	No	Normal	Le(b), Le(y)
NS0008	Secretor	negative	No	Empty	Le(b), Le(y)
NS0009	Secretor	Positive	Yes	Slightly pink	A, Le(a), Le(b), Le(y)
NS0010	Secretor	Positive	No	Normal	Le(b), Le(y)
NS0011	Non-secretor	Positive	No	Pink	Le(a), H1, Le(b)
NS0012	Secretor	negative	No	Normal	Le(a), Le(b), Le(y)
NS0013	Secretor*	negative	No	Yellow	Le(b), Le(y)
NS0014	Secretor	Positive	No	Very brown	B, Le(b)
NS0015	Non-secretor*	negative	No	Brown	Le(y)
NS0016	Non-secretor	negative	No	Diluted	Negative
NS0017	Secretor*	Positive	No	Normal	Le(b), Le(y)
NS0018	Secretor	negative	Yes	Pink	A, Le(y)
NS0019	Non-secretor	negative	No	Milky	A, Le(a), B, Le(b)
NS0020	Non-secretor*	Unknown	Yes	Unknown	Le(b), Le(x), Le(y)
NS0021	Secretor	Positive	Yes	Very Pink	A, B, Le(b), Le(x), Le(y)
NS0022	Secretor*	Positive	Yes	Slightly pink	A, Le(a), H1, Le(y)
NS0023	Secretor	Positive	No	Extremely Pink	A, Le(a), B, Le(b), Le(y)
NS0024	Secretor	Positive	No	Empty	Le(y)
NS0025	Non-secretor*	negative	No	Normal	Le(a)
NS0026	Secretor*	Positive	No	Normal	A, Le(a), Le(y)
NS0027	Secretor	Positive	No	Normal	A, Le(a), Le(b), Le(y)
NS0028	Non-secretor	negative	Yes	Slightly pink	Le(y)
NS0029	Non-secretor	negative	Yes	Normal	Le(a)
NS0030	Secretor	Unknown	Yes	Unknown	B, Le(b), Le(x), Le(y)
NS0031	Secretor	Positive	Yes	Empty	Le(a), H1, Le(b), Le(y)
NS0032	Non-secretor	negative	No	Normal	A, Le(a), H1, Le(b), Le(y)
NS0033	Non-secretor	negative	No	Normal	H1
NS0034	Secretor	negative	Yes	Normal	A, Le(a), H1, B, Le(b), Le(y)
NS0035	Non-secretor*	negative	No	Brown	A, Le(a), H1
NS0036	Secretor	Positive	No	Very brown	A, Le(a), Le(b), Le(y)
NS0037	Secretor	Positive	No	Unknown	A, Le(a), H1, Le(b), Le(x), Le(y)
NS0038	Secretor	Positive	Yes	Brown	A, Le(a), Le(b), Le(y)
NS0039	Secretor	Unknown	Yes	Unknown	Le(y)
NS0040	Secretor	negative	No	Normal	A, Le(y)
NS0041	Secretor	Unknown	Yes	Unknown	Le(y)
NS0042	Secretor	Positive	Yes	Yellow	A, Le(a), B, Le(b), Le(y)
NS0043	Non-secretor	negative	No	Yellow	Le(a), H1, Le(b)
NS0044	Secretor	Positive	Yes	Normal	A, Le(b), Le(y)

NS0045	Secretor	negative	No	Yellow	A, Le(a), H1, Le(b), Le(y)
NS0046	Secretor	Positive	Yes	Unknown	Le(y)
NS0047	Secretor	Unknown	No	Unknown	Unknown
NS0049	Secretor	Positive	No	Yellow	A, Le(a), H1, Le(b), Le(y)
NS0050	Non-secretor*	negative	Yes	Normal	Le(y)
NS0051	Secretor	Positive	No	Normal	Le(x), Le(y)
NS0052	Secretor	Unknown	Yes	Unknown	A, Le(y)
NS0053	Secretor	Positive	No	Unknown	Le(x), Le(y)
NS0054	Non-secretor	negative	No	Normal	A, Le(a), Le(b), Le(y)
NS0055	Non-secretor	negative	No	Empty	Le(a), Le(b)
NS0056	Non-secretor	negative	No	Normal	A, Le(A), h1
NS0058	Secretor	negative	No	Normal	A, Le(a), Le(b)
NS0059	Non-secretor	Positive	No	Normal	Negative
NS0060	Non-secretor	Unknown	Yes	Unknown	Unknown
NS0061	Secretor	negative	Yes	Normal	A, Le(a), Le(b), Le(y)
NS0062	Secretor	Positive	Yes	Normal	Le(b), Le(x), Le(y)
NS0063	Secretor	Unknown	Yes	Unknown	B, Le(y)
NS0064	Non-secretor*	Unknown	Yes	Unknown	Le(x), Le(y)
NS0065	Non-secretor	negative	No	Extremely Brown	H1, Le(a)
NS0066	Non-secretor	negative	Yes	Normal	Le(a)
NS0067	Secretor*	Positive	Yes	Unknown	Le(b), Le(x), Le(y)
NS0068	Secretor	negative	Yes	Unknown	Le(y)
NS0069	Secretor	Positive	No	Unknown	Le(b), Le(y)
NS0070	Secretor	negative	Yes	Normal	Le(a), Le(b), Le(y)
NS0071	Non-secretor	negative	No	Normal	A, Le(a), H1, Le(y)
NS0072	Secretor	Positive	No	Normal	Le(b), Le(y)
NS0073	Secretor	negative	No	Normal	Le(b), Le(y)
NS0074	Secretor	negative	No	Yellow	H1, Le(b), Le(y)
NS0075	Secretor	negative	Yes	Diluted	Le(b), Le(y)
NS0076	Secretor	Positive	No	Normal	A, Le(a), B, Le(b), Le(y)
NS0077	Non-secretor	negative	Yes	Empty	Le(a), Le(b)
NS0078	Secretor	Positive	No	Milky	Le(b), Le(y)
NS0079	Secretor	negative	Yes	Normal	Unknown
NS0080	Secretor	Unknown	No	Diluted	Le(y)
NS0081	Non-secretor	Unknown	No	Unknown	Unknown
NS0082	Non-secretor	negative	No	Normal	Negative
NS0083	Secretor	Unknown	Yes	Unknown	Unknown
NS0084	Non-secretor	Positive	Yes	Unknown	Le(b), Le(y)
NS0085	Secretor*	Positive	Yes	Empty	Le(b), Le(y)
NS0086	Secretor	Unknown	No	Unknown	Unknown
NS0087	Secretor	negative	No	Normal	Le(b), Le(y)
NS0088	Secretor	Unknown	No	Unknown	Unknown

NS0090	Secretor	negative	No	Normal	Negative
NS0091	Secretor	negative	No	Normal	Le(b)
NS0092	Secretor	negative	No	Normal	A, Le(b), Le(y)
NS0093	Non-secretor	negative	No	Normal	A, Le(a), Le(b)
NS0094	Secretor	negative	No	Normal	Le(b), Le(y)
NS0095	Secretor	Unknown	Unknown	Unknown	Unknown
NS0096	Non-secretor	negative	Unknown	Normal	Le(b), Le(y)
NS0097	Non-secretor	negative	No	Normal	Le(y)
NS0099	Non-secretor	negative	No	Normal	Negative
NS0101	Secretor	Positive	Yes	Normal	A, Le(a), B, Le(b), Le(x), Le(y)
NS0102	Secretor	negative	No	Normal	Le(b)
NS0103	Non-secretor	negative	No	Normal	B, Le(b), Le(x), Le(y)
NS0104	Secretor	negative	No	Normal	Le(a), B, Le(b), Le(x), Le(y)
NS0105	Secretor	Positive	No	Normal	Le(a), Le(b), Le(y)
NS0106	Secretor	negative	No	Extremely Pink	A, Le(a), H1, Le(b), Le(y)
NS0107	Non-secretor	negative	No	Normal	A, Le(a)
NS0108	Secretor	Positive	No	Normal	A, Le(a), Le(b), Le(y)
NS0109	Secretor	negative	No	Yellow	A, Le(a), Le(b), Le(y)
NS0110	Non-secretor	negative	No	Normal	Negative
NS0112	Secretor	Positive	No	Diluted	A, Le(b), Le(y)
NS0113	Secretor	negative	No	Empty	Le(a), H1, Le(b), Le(y)
NS0114	Secretor*	Positive	No	Normal	Le(a), H1, Le(b), Le(y)
NS0115	Non-secretor	negative	Yes	Pink	A, Le(a), H1
NS0116	Secretor	negative	No	Normal	A, Le(a), H1, Le(b)
NS0117	Non-secretor	negative	No	Normal	A, Le(a), H1, Le(b)
NS0118	Secretor	negative	No	Normal	Negative
NS0119	Non-secretor	Unknown	Yes	Unknown	B, Le(b)
NS0120	Non-secretor	Positive	No	Diluted	A, B, Le(y)
NS0121	Non-secretor*	negative	Yes	Diluted	Le(b), Le(y)
NS0122	Secretor	negative	Unknown	Normal	A, Le(b)
NS0123	Non-secretor	negative	Yes	Normal	B, Le(b), Le(y)
NS0124	Secretor	Positive	Unknown	Normal	Le(y)
NS0125	Secretor	negative	No	Yellow	A, Le(a), H1, Le(b), Le(y)
NS0126	Secretor	negative	Yes	Slightly pink	Le(y)
NS0127	Secretor	Positive	No	Yellow	Le(b)
NS0128	Secretor	negative	Yes	Normal	A, Le(a), H1, Le(b), Le(y)
NS0129	Secretor	negative	Yes	Diluted	Negative
NS0130	Secretor	Unknown	No	Unknown	Unknown
NS0131	Secretor	negative	No	Normal	A, Le(b)
NS0132	Secretor	Positive	No	Normal	A, Le(a), B, Le(b), Le(y)

NS0133	Secretor	Positive	Yes	Normal	A, Le(a), B, Le(b), Le(x), Le(y)
NS0134	Secretor	Positive	No	Diluted	A, Le(b), Le(y)
NS0135	Non-secretor	negative	Yes	Normal	A, Le(a), Le(b), Le(y)
NS0136	Secretor	negative	No	Normal	B, Le(b)
NS0137	Secretor	Positive	No	Empty	A, Le(b), Le(y)
NS0138	Non-secretor*	negative	Yes	Normal	Le(a), Le(b), Le(x)
NS0139	Secretor	Positive	No	Normal	Le(b), Le(y)
NS0141	Secretor	Positive	Yes	Diluted	A, H1, Le(b), Le(y)
NS0142	Secretor	Positive	Yes	Normal	A, Le(a), Le(b), Le(y)
NS0143	Secretor	Positive	No	Normal	Le(a), H1, Le(b), Le(y)
NS0145	Secretor	Positive	Yes	Normal	A, Le(a), H1. Le(b)
NS0146	Non-secretor	negative	No	Diluted	A, Le(a)
NS0147	Secretor	Positive	No	Pink	A, Le(a), B, Le(b), Le(y)
NS0149	Non-secretor	negative	No	Yellow	A, Le(a), H1
NS0150	Secretor	Positive	Unknown	Orange	Le(a), Le(b), Le(y)
NS0151	Secretor	negative	No	Normal	H1
NS0152	Secretor	negative	No	Pink	B, Le(b), Le(y)
NS0153	Secretor	Positive	Yes	Extremely yellow	Le(b)
NS0154	Non-secretor	Positive	Unknown	Orange	A, Le(a), H1, Le(b), Le(y)
NS0155	Secretor	Positive	No	Normal	A, B, Le(b), Le(x), Le(y)
NS0156	Secretor	Positive	Yes	Brown	A, Le(a), Le(b), Le(y)
NS0157	Non-secretor	Positive	Yes	Normal	Le(a), Le(b), Le(y)
NS0158	Secretor	negative	No	Normal	A, Le(a), Le(b), Le(y)
NS0159	Secretor	Positive	Yes	Normal	B, Le(b), Le(x), Le(y)
NS0160	Secretor	Positive	Yes	Normal	H1, B, Le(b)
NS0161	Secretor	Positive	No	Normal	A, Le(a), Le(b), Le(y)
NS0162	Secretor	Positive	No	Slightly pink	Le(a), B, Le(b), Le(x), Le(y)
NS0163	Secretor	Positive	No	Unknown	B, Le(b), Le(y)
NS0164	Secretor*	Positive	No	Diluted	Le(b), Le(y)
NS0165	Secretor	Positive	No	Unknown	A, Le(a), Le(b), Le(y)
NS0167	Non-secretor	negative	Unknown	Unknown	A, Le(a), H1
NS0168	Secretor	negative	Yes	Pink	Le(a), Le(b), Le(y)
NS0169	Secretor	negative	No	Normal	Le(a), B, Le(b), Le(y)
NS0170	Secretor	negative	Yes	Normal	B, Le(b), Le(x), Le(y)
NS0172	Secretor	Positive	No	Diluted	A, B, Le(b), Le(y)
NS0173	Secretor*	Positive	No	Diluted	A, Le(b), Le(y)
NS0174	Secretor	negative	No	Diluted	Le(b)
NS0175	Secretor	negative	No	Normal	Le(a), B, Le(b), Le(x), Le(y)
NS0176	Secretor	negative	Yes	Diluted	Le(b), Le(y)
NS0178	Secretor	Positive	No	Normal	Le(b), Le(y)

NS0179	Secretor	negative	No	Diluted	Le(b), Le(y)
NS0183	Secretor	Positive	No	Diluted	B, Le(b), Le(y)
NS0184	Secretor	negative	Yes	Diluted	Negative
NS0185	Secretor	Positive	No	Diluted	B, Le(b), Le(y)
NS0186	Non-secretor	negative	No	Unknown	Le(a), Le(y)
NS0187	Secretor	negative	No	Normal	A, Le(b), Le(y)
NS0188	Secretor*	Positive	Yes	Slightly pink	B, Le(b), Le(y)
NS0189	Secretor	negative	No	Normal	Le(y)
NS0190	Secretor	negative	No	Normal	B, Le(b), Le(y)
NS0191	Secretor	negative	No	Milky	Le(y)
NS0192	Secretor*	Positive	Yes	Diluted	Le(a), Le(b), Le(y)
NS0193	Secretor	Positive	No	Diluted	H1, Le(b)
NS0194	Secretor	negative	No	Diluted	Le(b)
NS0195	Secretor	negative	No	Diluted	A, Le(b), Le(y)
NS0196	Secretor	negative	No	Normal	Le(b), Le(y)
NS0197	Secretor	Positive	Yes	Normal	A, Le(b), Le(y)
NS0198	Secretor	negative	No	Pink	A, Le(a), H1, B, Le(y)
NS0199	Non-secretor	negative	Yes	Extremely orange	Negative
NS0200	Secretor	negative	Yes	Normal	A,H1, B, Le(b), Le(y)
NS0201	Secretor	negative	Yes	Slightly pink	Le(a), H1, B, Le(b), Le(y)
NS0202	Secretor	Positive	No	Pink	Le(a), Le(b), Le(y)
NS0203	Secretor	Positive	Yes	Pink	A, Le(a), H1, B, Le(b), Le(y)
NS0204	Secretor	Positive	Yes	Unknown	Le(a), Le(b), Le(y)
NS0205	Non-secretor	negative	No	Unknown	Le(b), Le(y)
NS0206	Secretor	negative	No	Normal	Le(y)
NS0207	Non-secretor*	negative	No	Very white	H1, Le(b), Le(y)
NS0208	Secretor	Positive	Yes	Milky	H1, Le(y)
NS0209	Non-secretor	Positive	No	Normal	Negative
NS0210	Non-secretor	negative	No	Diluted	Le(b), Le(y)
NS0211	Secretor	Positive	No	Very Pink	A, Le(b)
NS0212	Non-secretor	negative	No	Unknown	Le(a), H1
NS0214	Secretor	Positive	No	Diluted	A, Le(b), Le(y)
NS0215	Secretor	Positive	Yes	Unknown	Le(b), Le(y)
NS0216	Non-secretor	negative	Unknown	Unknown	H1, Le(b), Le(y)
NS0217	Non-secretor	Positive	No	Unknown	A, Le(b), Le(y)
NS0218	Non-secretor	negative	No	Diluted	A, Le(a), H1
NS0219	Secretor	Positive	Yes	Normal	Le(b), Le(y)
NS0220	Non-secretor*	negative	Yes	Unknown	Le(a), Le(y)
NS0221	Non-secretor*	negative	No	Unknown	A, Le(a), Le(b), Le(y)

*Sequences resulting in non-secretors in SNV allele SNP assay, that were retested with conventional PCR.

Appendix F: Next generation sequencing

Primer	Sequences
MiSeq adaptors primers	1-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG 2-CTGTCTCTTATACACATCTGACGCTGCCGACGA 3-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG 4-CTGTCTCTTATACACATCTCCGAGCCCACGAGAC
Amplification primers	1-CTGTGAATGCTGCGACTACGAT 2-TCAGAGAACCTCATCCACCTGAACAT 3-TCAGAGAATCTCATCCATCTGAACATATGTTCAGATGGATGAGATTCTCAGA 4-TCTGAGAATCTCATCCATCTGAACATATGTTCAGATGGATGAGGTTCTCAGA 5-TCTGAGAACCTCATCCATCTGAACATATGTTCAGATGGATGAGGTTCTCTGA 6-TCAGAGAACCTCATCCATCTGAACATATGTTCAGGTGGATGAGATTCTCAGA 7-TCTGAGAATCTCATCCACCTGAACATATGTTCAGGTGGATGAGATTCTCTGA 8-TCAGAGAATCTCATCCACCTGAACATATGTTCAGGTGGATGAGGTTCTCAGA 9-TCTGAGAACCTCATCCACCTGAACATATGTTCAGGTGGATGAGGTTCTCTGA 10- ATCGTAGTCGCAGCATTCACAGATGTTCAGATGGATGAGATTCTCTGA