COMPARISON OF THE DIAGNOSTIC PERFORMANCE OF TAQMAN ARRAY CARDS, ENZYME IMMUNOASSAY, REAL-TIME PCR AND NEXT GENERATION SEQUENCING IN INVESTIGATION OF FIVE COMMON DIARRHOEA-ASSOCIATED ENTERIC VIRUSES IN KILIFI, KENYA

ARNOLD WASIKE LAMBISIA

MASTER OF SCIENCE

(Molecular Biology and Bioinformatics)

JOMO KENYATTA UNIVERSITY OF

AGRICULTURE AND TECHNOLOGY

2021

Comparison of the Diagnostic Performance of TaqMan Array Cards, Enzyme Immunoassay, Real-Time PCR and Next Generation Sequencing in Investigation of Five Common Diarrhoea-Associated Enteric Viruses in Kilifi, Kenya

Arnold Wasike Lambisia

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Molecular Biology and Bioinformatics of the Jomo Kenyatta University of Agriculture and Technology

2021

DECLARATION

This thesis is my original work and has not been presented elsewhere for a degree award

Signature…… ……..Date……

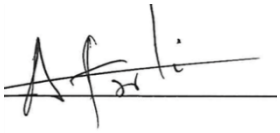
Arnold Wasike Lambisia

The thesis has been submitted for examination with our approval as supervisors

Signature…… ……..Date……

Dr. Steven Ger Nyanjom, PhD

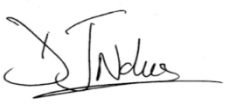
JKUAT, Kenya



Signature…… …… Date……

Dr. Charles Nyaigoti Agoti

KEMRI, Kenya



Signature…… …… Date……

Prof David James Nokes

KEMRI, Kenya

DEDICATION

My thesis is dedicated to my parents and siblings for their unconditional love and support.

ACKNOWLEDGEMENT

First, I would like to acknowledge my supervisors Dr. Charles Agoti Nyaigoti, Professor David James Nokes and Dr. Steven Ger Nyanjom. Many thanks for your unwavering support, guidance and mentorship. Many thanks to Dr. Charles Agoti and Prof. D. James Nokes for conceiving this project and allowing me to do it under your supervision in the Virus Epidemiology and Control (VEC) group at KEMRI-Wellcome Trust Research Programme (KWTRP). The scientific research skills that I have gained will go a long way in helping me succeed in my future career. Many thanks to Dr. Steven Ger Nyanjom for your critical support back in the University and encouraging me to keep going despite the chllenges I have faced throughout the MSc period. I could not do this project without your combined efforts.

Many thanks to the staff at the KWTRP who supported my project both in the laboratory and data analysis. These people include; Sylvia Onchaga, Elijah Gicheru, Clement Lewa, Martin Mutunga, David Collins Owor, John Mutiso and Everline Kamau. I would like to thank the KWTRP training department, Initiative to Develop African Research Leaders (IDeAL), for sponsoring my MSc studentship. Special thanks to Prof. Sam Kinyanjui, Dr. Dorcas Mbuvi and Liz Murabu for being very supportive and encouraging throughout my time at KWTRP. Many thanks to the 2018-2019 MSc cohort at KWTRP for being good friends and office mates for the entire studentship period.

Many thanks to my parents, Aaron Lambisia and Willimina Tabu, for believing in my dreams and sponsoring my tuition fees so that I can make my dreams come true. My siblings, Liz Anjiko and Bonface Wangila, I can not thank you enough for your support, prayers and zeal that makes us want to become better people despite our humble background.

I would also like to thank Jomo Kenyatta University of Agriculture and Technology for providing me with an opportunity to pursue my studies. I would like to end by saying that it has been a great learning opportunity and it could not have happened without the input of every single person who supported me throughout this period.

God bless you all.

“This work was supported through the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769/Z/10/Z] and the UK government. The views expressed in this thesis are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government”.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

C.I Confidence interval

DNA Deoxyribonucleic acid

ELISA Enzyme-Linked Immunoassay

HAdV Human Adenovirus

KCH Kilifi County Hospital

KEMRI Kenya Medical Research Institute

KWTRP KEMRI-Wellcome Trust Research Programme

NGS Next Generation Sequencing

NSP Non-structural Protein

OD Optical Density

PCR Reverse-Transcriptase Polymerase Chain Reaction

RVA Rotavirus group A

RNA Ribonucleic acid

TAC TaqMan Array Card

VEC Virus Epidemiology and Control

WHO World Health Organization

ABSTRACT

Diarrhoea is a major cause of morbidity and mortality in low and middle income countries. Estimates of enteric viral diarrhoea are usually obtained from enzyme immunoassays and PCR-based diagnostic methods. However, the performance of these diagnostic methods can differ due to factors such as their principles, disease prevalence, viral diversity and evolution events. The study was designed as a repeated cross-sectional study to compare the diagnostic performance of an enzyme immunoassay, two PCR-based methods and agnostic Next Generation Sequencing (NGS) used in investigating viral diarrhoeal aetiology during the pre-(2013) and post-rotavirus (2016) vaccine introduction periods in coastal Kenya. The study compared an enzyme immunoassay (EIA) kit and two nucleic acid-based methods for rotavirus group A (RVA) detection in hospitalized children pre-post rotavirus vaccine introduction in Kenya. A total of 489 fecal specimens, 237 from 2013 (48.5%, pre-vaccine) and 252 from 2016 (51.5%, post-vaccine), were analyzed. Although the specificity of the 3 methods was ≥ 96%, the sensitivity of EIA, 68.9% (95% C.I: 53.4-81.8%), differed significantly from that of real-time PCR, 93.3% (95% C.I: 81.7%-98.6%), and taqman array card, 97.8% (95% C.I: 88.2-99.9%), in 2016 (p-value <0.05) unlike in 2013. The study also compared real-time PCR with taqman array card (TAC) in the detection of adenovirus, astrovirus, norovirus genogroup II and sapovirus. A total of 494 fecal specimens, 242 (49%) from 2013 and 252 (51%) from 2016, were analyzed. TAC detected more positives than real-time RT-PCR in all the four viruses. However, the difference was only statistically significant in adenovirus and astrovirus detection (p-value < 0.05). Agnostic NGS was applied to 69 samples that had tested positive for RVA by EIA in 2013 to determine viral coinfections. Detection of the coinfections varied by the bioinformatics workflows i.e. classification of unassembled reads using Kraken detected an enteric virus coinfection in 32 (46.4%) samples compared to Metaphlan which detected a coinfection in 21 (30.4%) samples and classification of contigs using Kraken detected a coinfection in 16 (23.2%) samples. In conclusion, RVA-EIA showed a significant decrease in sensitivity (p-value < 0.05) post-vaccine introduction compared to real-time PCR and TAC. Further, TAC detected significantly more adenovirus and astrovirus positives than real-time RT-PCR. The difference in the diagnostic performance of the methods impacted on the calculated prevalence estimates. Finally, the study also showed the potential of agnostic sequencing in the detection of viral coinfections in clinical samples.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

In 2016, diarrhoea was responsible for approximately ~500,000 deaths in children younger than five years of age and globally, it is considered among the major causes of hospitalizations and mortality (Troeger *et al.*, 2018). Viruses such as Rotavirus Group A (RVA), sapovirus, human adenovirus (HAdV) type 40/41, astrovirus and norovirus genogroup II have been reported among the top ten leading diarrhoeal aetiologies in hospitalized and community cases (Liu *et al.*, 2016; Platts-Mills *et al.*, 2018). RVA is the predominant cause of diarrhoea in countries without rotavirus vaccination programmes, with almost all children experiencing at least one rotavirus infection by the age of five years (Bányai *et al.*, 2018). The heterogeneity of enteric pathogens associated with diarrhoea provides a challenge in the design and use of different diagnostic tools. Further, different methods have their own strengths and weaknesses that limit their applicability in some settings (Vila, 2017). In the surveillance of rotavirus specific diarrhoea and assessment of vaccination impact on rotavirus strains, the ProSpecT^{TM} Rotavirus kit (Oxoid Ltd., Hampshire, UK is a recommended enzyme immunoassays (WHO, 2008). However, suboptimal performance has been reported for ELISA in vaccinated communities (McAuliffe *et al.*, 2019). ELISAs are less sensitive than PCR assays in RVA detection (Tate *et al*., 2013), and this might lead to underestimation of RVA diarrhoea burden. PCR assays vary by the genomic region of the virus targeted by primers/probes, the amount of starting material and assay type (conventional, real-time, singleplex and multiplex). This heterogeneity leads to varied sensitivities across the specific kits or assays (Chhabra *et al*., 2017). Furthermore, the frequency and positions of mismatches in primer and probe annealing sites due to point mutations and reassortment of segments from different serogroups and zoonosis affects the diagnostic performance of molecular diagnostic tools (Lefever *et al.*, 2013; Martella *et al.*, 2010). An additional challenge is the inability to detect novel viruses associated with diarrhoea using existing kits. Viral metagenomics enables detection and description of the viral

content without prior sequence knowledge in specimen from patients presenting with diarrhoea symptoms (Cotten *et al*., 2014; Moore *et al*., 2015; Yinda *et al*., 2019). However, the absence of prior pathogen specific nucleic acid amplification decreases its sensitivity.

In Kilifi, coastal Kenya, the Virus Epidemiology and Control (VEC) Research Group based at KEMRI/Wellcome-Trust Research Programme (KWTRP) has been carrying out routine surveillance of diarrhoea since 2009 in children admitted to the Kilifi County Hospital (KCH). Kenya introduced RVA vaccination, given in two doses at week six and ten, into its National Immunisation Programme (NIP) in July, 2014. Vaccination is likely to reduce replication efficiency of the virus in cases where infection breaks through. Therefore, sensitive methods that have a viral material pre-amplification step before detection e.g. Polymerase Chain Reaction (PCR) are expected to detect even low viral titres during routine surveillance. However, detection by such methods does not always infer causality of diarrhoea but gives an insight of the reservoirs of these pathogens (Bennett *et al.*, 2015; Liu *et al*., 2016; Platts-Mills *et al.*, 2018). As a result, accurate estimation of vaccine effectiveness based on case-definitions of different methods and occurrence of diagnostic errors have become issues of interest prompting the evaluation of the available diagnostic methods used in routine surveillance of diarrhoea aetiology.

This study compared the diagnostic performance of RVA antigen detection by ELISA and two PCR methods: real-time PCR and TaqMan Array Cards (TAC), in the detection of RVA in the pre (2013) and post-vaccine (2016) introduction periods in Kilifi Kenya. Further, TAC and real-time PCR approach were compared in the detection of astrovirus, HAdV 40/41, norovirus genogroup II and sapovirus. A subset of samples from the pre-rotavirus vaccination period in Kenya that were positive for rotavirus A (RVA) by ELISA were sequenced by viral metagenomics approach to determine viral coinfections. The detected viruses were confirmed by either real-time RT-PCR or TAC.

1.2 Problem Statement

The heterogeneity in viral diarrhoeal aetiology is a challenge to the available diagnostic methods. Previously, ELISA which is recommended in surveillance of rotavirus specific diarrhoea (WHO, 2008) has been reported to have suboptimal performance in the routine rotavirus vaccination era (McAuliffe *et al*., 2019). Kenya introduced RVA vaccination, given as two doses at week 6 and 10, into the National Immunization Programme in July 2014 (Wandera *et al*., 2018). The vaccine is expected to reduce the burden of RVA specific diarrhoea and viral replication. However, due to limited vaccine coverage and effectiveness, diarrhoea is still a source of hospitalizations in coastal Kenya (Khagayi *et al.*, 2019; Otieno *et al*., 2019; Wandera *et al*., 2017). Notably, detection does not always mean causation due to coinfections, asymptomatic carriage and microbial gut contamination. PCR methods heterogeneity by type and gene targets affects their sensitivity and specificity across different kits (Chhabra *et al*., 2017). Therefore, continuous surveillance using diagnostic methods with high diagnostic performance such as the PCR is key in describing viral diarrhoeal etiology and in the detection of even low viral titres infections.

1.3 Justification

RVA, adenovirus, norovirus GII, astrovirus and sapovirus are among the top 10 causes of diarrhoea in children < 5 years (Liu *et al*., 2013; Platts-Mills *et al.*, 2018). ELISA, real-time PCR, TAC and viral metagenomics are used in Kilifi by KEMRI Wellcome-Trust for routine surveillance of these five viruses. Comparing the diagnostic performance of these available diagnostic methods used is key in understanding the level of concordance between them and obtaining possible explanations for the equivocal results. The suboptimal performance of previously recommended methods in the routine rotavirus vaccination era (McAuliffe *et al*., 2019) may give incorrect estimates of disease burden, vaccine effectiveness, transmission patterns and misinform vaccine policies. Comparing these diagnostic methods will inform review and ways to improve them to ensure correct inference of viral diarrhoeal disease burden and transmission patterns. Further, viral

metagenomics has enabled examination of enteropathogens beyond the common set of viruses associated with diarrhoea in a sequence independent manner but its sensitivity is not well understood (Moore *et al*., 2015). hence; it will also give a picture of the viral coinfections circulating in the pre-rotavirus vaccination period. Therefore, assessing and reviewing the available testing algorithms used in routine detection of viral diarrhoeal agents in Kilifi is critical for reporting future viral diarrhoeal burden and transmission patterns.

1.4 Hypothesis

1.4.1 Null Hypothesis

1. There is no difference in the diagnostic performance of RVA-ELISA, real-time PCR and TAC in the detection of RVA, HAdV 40/41, astrovirus, sapovirus and norovirus genogroup II.
2. There are no nucleotide mismatches in the real-time PCR and TAC primers/probe binding sites.
3. There are no viral coinfections detectable by viral metagenomics in RVA positive samples by ELISA in the pre-rotavirus vaccine introduction period (2013).

1.5 Objectives

1.5.1 General Objective

To compare the diagnostic performance of four different methods in detection of viral agents in stool samples collected from children admitted with diarrhoea pre- and post-vaccine introduction in Kilifi, Kenya.

1.5.2 Specific Objectives

1 To determine diagnostic disparities between real-time PCR, TAC and RVA-ELISA in detecting RVA, norovirus genogroup II, sapovirus, astrovirus and HAdV type 40/41 in the pre- and post-rotavirus vaccine introduction periods.

2 To examine primer and probes included in the real-time PCR and TAC for nucleotide mismatches in their binding regions for RVA, astrovirus, HAdV type 40/41, norovirus genogroup II and sapovirus.

3 To determine viral coinfections in ELISA rotavirus positives in the pre-rotavirus vaccination period using next generation sequencing.

CHAPTER TWO

LITERATURE REVIEW

2.1 Aetiology of Diarrhoea

In children under the age of five years, diarrhoea is considered the second leading cause of mortality, and globally, it accounts for approximately 500,000 deaths annually (Liu *et al*., 2016; WHO, 2017). In 2017, diarrhoea was estimated to cause more than 10,000 deaths in children under the age of five years in Kenya (Global Burden of Disease Collaborative Network., 2017). Diarrhoea is caused by bacterial, viral and parasitic pathogens. Bacterial causes of diarrhoea include: Shigella, Salmonella, campylobacter, Yersinia, Enterotoxigenic E. coli, Enteropathogenic E. coli, Enteroaggregative E. coli, Vibrio and Clostridium dificile. Protozoa such as Cryptosporidium spp, Entamoeba, Giardia and Cyclospora spp are also associated with diarrhoea. Viral diarrhoea caused by RVA, HAdV type 40/41, astrovirus, norovirus genogroup II and sapovirus, has been reported to have higher diarrhoeal attributable incidence than bacterial and parasitic diarrhoea (Liu *et al*., 2016; Platts-Mills *et al*., 2018).

Rotavirus is recognized as the leading cause of diarrhoea in countries that have not yet adopted the rotavirus vaccine into their national vaccination programmes and nearly all children experience a rotavirus infection before the age of five years (Bányai *et al.*, 2018). Kenya introduced a two-dose oral rotavirus vaccine (Rotarix®, GlaxoSmithKline Biologicals, Belgium) in 2014 with support from the Global Alliance for Vaccine and Immunizations (GAVI) (Wandera *et al*., 2018). The vaccine is administered to all children at six and ten weeks after birth. Rotavivus vaccine effectiveness in Kenya is estimated to be 64% (95% CI: 35-80%) and 54% (95% CI: 20-83%) in children who have received two doses and one dose respectively (Khagayi *et al*., 2019).

2.2 Diagnostic Performance of Methods Used in Investigating Viral Diarrhoeal Aetiology

The heterogeneity of viruses associated with diarrhoea has led to the development of different diagnostic methods such as electron microscopy, ELISA, immunochromatographic tests (ICTs) and molecular assays that are pathogen specific and can be used in viral diarrhoea surveillance (Bennettand Gunson, 2017; Dennehy *et al.*, 1990; Gautam *et al.*, 2013; Tate *et al*., 2013; Vinjé, 2015). These methods employ different principles such as antigen detection, microscopy and nucleic acid amplification in virus detection which affects their sensitivity and specificity. The differences in the sensitivity and specificity of these methods could substantially affect the viral diarrhoeal burden estimates reported in surveillance studies.

Different diagnostic methods have been designed and their diagnostic performance evaluated for detection of specific viruses associated with diarrhoea as discussed below.

2.2.1 Rotavirus A

Rotavirus is a non-enveloped double-stranded RNA virus of the family *Reoviridae.* Its genome is approximately 18.5kb in length and encodes six structural (VP1-4, VP6 and VP7) and six non-structural proteins (NSP1-NSP6). RV is classified into 11 serogroups (A-J) based on the antigenicity of the VP6 protein. RVA is the predominant serogroup among other serogroups (RVB, RVC, RVH) known to infect humans (Sadiq *et al.*, 2018).

The WHO recommends the use of ELISA; Premier Rotaclone® (Meridian Diagnostics, Inc, USA), Abbot Testpack® (Abbot Laboratories, USA), Pathfinder® (Sanofi Diagnostic Pasteur, France) and IDEIA test (Oxoid, UK) replaced by the Oxoid ProSpecT^{TM} kit to monitor trends in rotavirus disease burden (WHO, 2008). The Oxoid ProSpecT^{TM} RVA kit is used in most GAVI sponsored surveillance networks including locally at the KCH for RVA surveillance run by KWTRP.

The diagnostic accuracy of these kits has been previously evaluated (summarised in Table 2.1) using a variety of reference standards (Brooks *et al.,* 1989; Donck *et al.,* 1999; Gautam *et al*., 2013; Steele *et al.,* 1994). The kits recommended by the WHO tend to have varied sensitivities but very high specificities. However, the diagnostic performance evaluation might have been biased by the choice of the reference standard. Electron microscopy was used in rotavirus diagnosis before ELISA and RT-PCR became common pre-rotavirus vaccine introduction but it has a lower sensitivity and specificity than ELISA (Rubensteinand Miller, 1982). On the other hand, RT-PCR has a higher sensitivity than ELISA in RVA detection (Tate *et al*., 2013). Rotascreen II ELISA kit has been reported to have false positives in vaccinated communities (Lopez-Lacort *et al*, 2016; McAuliffe *et al*., 2018) prompting re-evaluation of the testing methods recommended for routine rotavirus surveillance post-vaccine introduction.

Table 0.1: Diagnostic Performance of WHO Recommended Kits

ELISA Kit	Authors	Basis	Sample Size	Reference Method	Sensitivity (%)	Specificity (%)
Premier Rotaclone® (Meridian Diagnostics, Inc, USA)	(Gautam *et al.*, 2013)	Monoclonal antibodies (VP6)	110	Reverse-transcriptase-PCR (VP4 and VP7 and/or VP6)	76.8±7.9	**100**
Abbot Testpack® (Abbott Laboratories, North Chicago, USA)	(Brooks *et al.*, 1989)	Murine monoclonal and bovine polyclonal antibodies	100	Electron microscopy	Frozen samples: 45±9.8 Unfrozen samples: 86±6.8	Frozen samples: 100 Unfrozen samples: 76±11
Abbot Testpack® (Abbott Laboratories, North Chicago, USA)	(Donck *et al.*, 1999)	Murine monoclonal and bovine polyclonal antibodies	100	Premier Rotaclone® kit	100	100
Pathfinder® (Kallestad laboratories, Texas, USA)	(Steele *et al.*, 1994)	Monoclonal and polyclonal antibodies (Antirotavirus IgG)	100	Electron microscopy	97±3.3	94±4.7
ID EIA Rotavirus Test/ ProSpecT^{TM} (Oxoid, Ltd., Basingstoke, Hampshire, UK)	(Gautam *et al.*, 2013)	Polyclonal antibodies	110	Reverse-transcriptase-PCR (VP4 and VP7 and/or VP6)	75±8.1	100

(Brooks *et al.,* 1989; Donck *et al.,* 1999; Gautam *et al*., 2013; Steele *et al.,* 1994)

RVA PCR assays use primers and probes that specifically target some RVA non-structural (NSP2, NSP3 and NSP4) and structural proteins (VP2, VP4, VP6 and VP7). However, continuous point mutations and reassortment of segments from different RV serogroups and also from zoonoses pose a challenge in the detection of these gene targets (Martella *et al.*, 2010).

The NSP3 gene codes for a translational enhancer and it has a highly conserved region near the 3' end compared to other RVA segments, and it is suggested to be an optimal target for detection of RVA genotypes. NSP3 qPCR has a sensitivity and specificity range of 95-100% and 77-100%, respectively (Liu *et al*., 2013; Mijatovic-Rustempasic *et al*., 2013). It has been reported to be more sensitive than ELISA and conventional heminested PCR (one primer used in the initial PCR reaction is

combined with a different primer in the second nested PCR reaction) in detection of wildtype RVA (Pang *et al*., 2004). PCR assays that target VP4 and VP7 regions are used for genotyping of RVA strains into P or G types respectively, with high sensitivity and specificity. It is also notable that NSP2 and VP6 have been established as important gene targets for differentiating between wild type and vaccine strains for Rotarix® and RotaTeq® respectively (Gautam *et al.*, 2016; Joshi *et al.*, 2019). Additionally, the VP6 qPCR has been shown to detect different RotaTeq® vaccine strains with 100% (Gautam *et al*., 2016; Joshi *et al*., 2019).

The diagnostic accuracy of various commercial and non-commercial PCR assays has been previously evaluated as summarized in Table 2.2 (Gautam et al., 2016; Joshi et al., 2019; Liu et al., 2013; Mijatovic-Rustempasic et al., 2013; Chhabra et al., 2017; Tate et al., 2013). It is reported that PCR has enabled detection of RVA with high sensitivity (89%-100%) and specificity (77%-100%), high sample throughput, and a faster turnaround time compared to ELISA which is labour intensive, low throughput and costly. A limitation in these studies is that the primers and probes were designed using the most common RVA G and P type strains and mutations or reassortments from other serogroups or zoonoses may end up undetected.

Table 0.2: Performance Characteristics of RVA PCR Methods

Authors	Assay	Gene-Target	Reference Assay	Sample Size	Sensitivity (%)	Specificity (%)	Location
(Gautam *et al.*, 2016)	Multiplex-qPCR	NSP3 NSP2-Rotarix® VP6-RotaTeq® VP4–P [4,6,8] genotypes VP7–G (1,2,3,4,9,12) genotypes	Conventional PCR and sanger sequencing (NSP2, NSP3, VP4, VP6, VP7)	873	NSP2-100 NSP3-100 VP4-P [4]- 100 VP4-P [6]- 100 VP4-P [8]- 98.8±0.7 VP6 - 100 VP7- G1-100 VP7- G2-100 VP7- G3-99.3±0.6 VP7- G4-100 VP7- G9-99±0.7 VP7-G12-98.8±0.7	NSP2-100 NSP3-100 VP4-P [4]- 100 VP4-P [6]- 100 VP4-P [8]- 100 VP6 - 100 VP7- G1-100 VP7- G2-100 VP7- G3-100 VP7- G4-100 VP7- G9-99.7±0.4 VP7-G12- 100	USA
(Mijatovic-Rustempasic *et al.*, 2013)	Singleplex qPCR	NSP3	ELISA (Premier Rotaclone, Meridian Bioscience, USA)	1830	99±0.5	77±1.9	USA and Ghana
	Singleplex qPCR	NSP3	Genotyping PCR (VP4 and VP7)	1889	100	86±1.6	
(Joshi *et al.*, 2019)	qPCR	VP6	RVA-ELISA (GA Generic assays GmbH, Dahlewitz, Germany)	143	96.2±0.3	100	India
(Tate *et al.*, 2013)	Semiquantitative PCR	NSP3	ELISA (Premier Rotaclone, Meridian Bioscience, USA)	648	99.4±0.3	86.7±1.3	USA
(Chhabra *et al.*, 2017)	Luminex xTAG® Gastrointestinal Pathogen Panel	NM			95.8±2.3	100	
	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	NSP3	Real-Time RT- PCR and conventional PCR followed by Sanger	300	89.6±3.5	100	USA

	Biofire's Gastrointestinal Panel (FilmArray)	NM	sequencing		100	95.6±2.3	
(Liu *et al.*, 2013)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	NSP3	ELISA (Prospect^{TM}) and PCR-Luminex	109	100	94.9±4.1	USA

(Gautam *et al.*, 2016, Mijatovic-Rustempasic *et al.*, 2013, Joshi *et al.*, 2019, Tate *et al.*, 2013, Liu *et al.*, 2013, Chhabra *et al.*, 2017)

2.2.2 Astrovirus

Astrovirus belongs to the *Astroviridae* family and its a linear, positive-sense ssRNA virus that is 6.4-7.7 kb in length with a poly A tail at the 3' end. At the 5'-end, ORF1a and ORF1b encode for non-structural proteins while the ORF2 at the 3' end encodes for the structural proteins. The variability of ORF1b and ORF2 enables typing of different astrovirus strains into different serotypes and genotypes (Vu *et al*., 2017). Astroviruses from the genus *Mamastrovirus* (*MAstV*), specifically Classic HAstv (*MAstV 1), MAstV 6* (MLB1-3), *MAstV 8* (VA2, 4 and 5, BF34) and *MAstV 9* (VA1 and 3) are known to infect and cause disease in humans (Vu *et al*., 2017).

Primers that target the ORF1a, ORF1b and the conserved 5' end of ORF2 are used to detect and type astroviruses (Guix, Bosch, and Pintó, 2005; Noel *et al.*, 1995). The genetic variability of different strains coupled with genomic evolution events that cause point mutations and recombination of different astrovirus strains (Wolfaardt *et al.*, 2011), provide a challenge in the diagnostic performance of different PCR methods. Astrovirus specific PCR assays have been found to be have high sensitivity and specificity compared to ELISA and electron microscopy which are laborious and time consuming (Cubitt *et al.*, 1999). A one-step real-time PCR using primers/probe targeting the ORF1a region was reported to be more sensitive (0.0052 IU/μl vs 261 IU/μl) and less time consuming (one hour vs four and a half hours) compared to conventional PCR (Royuela *et.al.*, 2006). Comparison of real-time PCR targeting the ORF1a region with electron microscopy showed a 9.1% positive percent agreement between the two methods and the real-time PCR detected ten more positives than electron microscopy (Logan *et al.*, 2007). To increase the turn-around time for detecting enteric viruses, several multiplex PCR assays have been developed and

their diagnostic performance evaluated as summarized in Table 2.3 below (Chhabra *et al*., 2017; Feeney *et al*., 2011; Hyun *et al*., 2018; Liu *et al*., 2013).

Table 0.3: Performance characteristics of PCR in astrovirus detection

Authors	Method	Gene Target	Reference Method	Sample Size	Sensitivity (%)	Specificity (%)
(Feeney *et al.*, 2011)	Probe-based multiplex TaqMan assay	3' non-coding region	Nested gel-based assay or Probe-based multiplex	137	100	100
(Hyun *et al.*, 2018)	Allplex GI-Virus Assay (Seegene, Seoul, Korea)		Seeplex PCR assay		PPV: 95.7 ± 1.9	NPV: 100
			Genotyping PCR [ORF2]	446	PPV: 100	NPV: 98 ± 1.3
(Liu *et al.*, 2013)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	RdRp	Luminex xTAG® Gastrointestinal Pathogen Panel	109	100%	98.9 ± 2
(Chhabra *et a*l., 2017)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	RdRp	Realtime RT- PCR and conventional PCR followed by Sanger sequencing		92.3± 3	98.9 ± 1.2
	Biofire's Gastrointestinal Panel (FilmArray)				97.4± 1.8	98.9 ± 1.2
				300		

PPV: Positive percent agreement NPV: negative percent agreement RdRp: RNA-dependent RNA polymerase

(*Chhabra et al*., 2017; Feeney *et al.*, 2011; Hyun *et al.*, 2018; Liu *et al.*, 2013)

2.2.3 Sapovirus

Sapoviruses belong to the family *Calciviridae* and are single-stranded positive-sense RNA viruses approximately 7.1-7.7kb in length. Fourteen sapovirus genogroups have been recognized based on the capsid (VP1) region. Genogroups GI, GII, GIV and GV are known to infect humans (Oka *et al.*, 2015).

Different diagnostic methods have been used to detect human sapoviruses in stools samples. In 1976, sapovirus was detected in human stool samples using an electron microscope (Madeleyand Cosgrove, 1976). Real-time PCR has become a routine assay for sapovirus detection from clinical specimens. Primers used in real-time PCR are designed to target the partial VP1 region, partial RdRp region (Liu *et al*., 2013) or the RdRp -VP1 junction (Jiang *et al*., 2014). The RdRp-VP1 junction is conserved

across genogroups GI,GII and GIV and it's a suitable target for primer and probe design (Oka *et al*., 2006).

Singleplex and multiplex PCR assays diagnostic performances have been evaluated before as summarized in Table 2.4. The sensitivity and specificity of real-time PCR varies from 75.6%-100% and 88%-100% respectively depending on the assay and reference standard used (Chhabra *et al*., 2017; Jiang *et al*., 2014; J. Liu *et al*., 2013, 2011; Oka *et al*., 2006). ELISA is affected by cross-reactivity between diverse sapovirus strains and has a lower sensitivity compared to PCR (Hansman *et al*., 2006). Development of broadly reactive monoclonal antibodies based on a common epitope that might exist between GI, GII, GIV and GV sapovirus strains might improve sapovirus ELISA's diagnostic performance.

Table 0.4: Performance characteristics of PCR in sapovirus detection

Authors	Method	Gene Target	Reference Method	Sample Size	Sensitivity (%)	Specificity (%)
(Hansman *et al*., 2007a)	Electron microscopy		Nested PCR (VP1 gene)	11	55.6 ± 29.3	100
(Hansman *et al*., 2006)	ELISA	Rabbit/guinea pig antisera raised against SaV GI VLPs	Single-round PCR (N-terminal region of the capsid gene)	12	60	100
			Nested PCR		25	
(Chhabra *et al*., 2017)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	RdRp	Both RT-qPCR and conventional PCR followed by sequencing	41	75.6 ± 13.1	100
	Biofire FilmArray (FilmArray))				97.6 ± 4.7	99.6 ± 1.9
(Liu *et al*., 2013)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	RdRp	PCR-Luminex	109	100	98.9 ± 2.0
(Liu *et al*., 2011)	Luminex xTAG® Gastrointestinal Pathogen Panel		Singleplex real-time qPCR	229	100	88 ± 4.2
(Jiang *et al*., 2014)	Multiplex real-time PCR	RdRp/capsid junction	Singleplex real-time PCR	812	100	100
(Oka *et al*., 2006)	Singleplex realtime PCR	RdRp/capsid junction	Nested PCR (capsid coding region)	7	100	100

(Hansman *et al*., 2007a, Hansman *et al*., 2006, Chhabra *et al*., 2017, Liu *et al*., 2013, Liu *et al*., 2011, Oka *et al*., 2006, and Jiang *et al*., 2014)

2.2.4 Norovirus Genogroup II

Human norovirus belongs to the *Calciviridae* family. The human norovirus is a linear, positive-sense ssRNA. Its genome is approximately 7.5-7.7 kb in length and contains three opening reading frames (ORF). ORF1 codes for a polyprotein that is processed into seven non-structural proteins (NS1-NS7) while ORF2 and ORF3 encode for structural proteins VP1 and VP2 respectively (Thorneand Goodfellow, 2014). Genetic classification for noroviruses is based on the NS7 (RNA-dependent RNA polymerase) and VP1 (major capsid protein) (Kroneman *et al*., 2013). There are three genogroups (G1, GII and GIV) with several genotypes that are known to infect humans. Norovirus genogroup II, especially GII.4 variants, are responsible for majority of the outbreaks in all age groups globally (Robilotti, Deresinski, and Pinsky, 2015).

Norovirus genogroup II diagnostic testing can be done using different methods such as electron microscopy, serology (ELISA and immunochromatographic tests) and molecular methods (conventional and real-time PCR). The diagnostic performance of these methods have previously been evaluated and reviewed (Robilotti *et al*., 2015). The most commonly used ELISA kits are Ridascreen (R-Biopharm, Darmstadt, Germany) and IDEIA (Oxoid Ltd., Hampshire, United Kingdom). These ELISA assays have been reported to have a wide range of sensitivity (31.6%-92%) and specificity (65.3%-100%) (Robilotti *et al*., 2015). Similarly, immunochromatographic tests have varied sensitivities (17%-92%) and a high specificity (87.5%-100%) (Ambert-Balayand Pothier, 2013; Robilotti *et al*., 2015). The source of variability in the sensitivity and specificity of these methods has been attributed to viral load and genotypes, age, time of stool collection compared to symptom onset, outbreak vs sporadic cases and the choice of reference standard. Due to the varied sensitivities and specificities the positives detected should be confirmed by an additional assay such as an PCR.

PCR is used as the gold standard for detection and genotyping of noroviruses. A conserved region at the ORF1-ORF2 junction has been used to design primer and probes used in real-time PCR (Kageyama *et al*., 2003). The diagnostic performance

of commercial norovirus genogroup II qPCR assays have previously been reviewed and reported to have a high sensitivity (92.5%-100%) and specificity (93.6%-100%) (Dunbar, Bruggink, and Marshall, 2014; Duong *et al*., 2016; Liu *et al*., 2013; Navidad *et al.*, 2013; Wessels *et al.*, 2014). Careful interpretation of norovirus genogroup II real-time PCR results is crucial since detection does not infer clinical disease. The Global Enteric Multicentre Study (GEMS) recommends a norovirus genogroup II TAC Ct cut-off of 23.4 for identifying symptomatic individuals in clinically relevant diarrhoeal cases (Liu *et al*., 2016) while the MAL-ED study recommended a TAC Ct cut-off of 27.2 (1.7×10^8 copies/gram) for community based studies (Platts-Mills *et al*., 2018). Therefore, norovirus genogroup II surveillance studies using qPCR should incorporate a Ct cut-off to differentiate symptomatic and asymptomatic individuals.

2.2.5 Human Adenovirus

Human adenoviruses (HAdVs) belong to the family *Adenoviridae* and genus *Mastadenovirus.* HAdVs are double-stranded, non-enveloped, icosahedral, linear DNA viruses of approximately 36kbp in genome length. HAdVs are classified into seven species (A-G) with 88 different serotypes and more emerging serotypes (Dhingra *et al*., 2019). HAdV F (40, 41) and G (52) are associated with diarrhoea but other HAdV species (A, B and C) have also been isolated from patients with diarrhoea symptoms (La Rosa *et al*., 2015). As a result, molecular detection methods have designed primers that target the hypervariable region seven of the hexon gene to detect all HAdV serotypes (Sarantis *et al.*, 2004). These primers have potential to capture diverse adenovirus species that may have been transmitted due to zoonosis (Hoppe *et al*., 2015). However, for specific detection and genotyping of HAdV type 40/41, primers that target the conserved regions of the fiber gene are used (La Rosa *et al*., 2015).

The performance of some of the commercial multi-pathogen kits has been evaluated before as summarized in Table 2.5.TaqMan Array Cards evaluated use primers that target the hexon gene hence they did not focus specifically on the HAdV type 40/41. The FilmArray and Luminex GPP panel primers and probes are unknown and its

unknown if they specifically target the HAdV type 40/41. The sensitivity and specificity of these kits vary from 68.4-100% and 80.3-100% respectively depending on location and choice of reference standard.

Table 0.5: Performance characteristics of PCR in HAdV detection

Authors	Method	Target	Reference	Sample Size	Sensitivity (%)	Specificity (%)	Location
(Higgins *et al.*, 2011)	Seeplex DV Assay	NA	Electron microscopy	198	97 ± 2.4	99.4 ± 1.1	Canada
			rPCR (hexon)		97± 2.4	99.4± 1.1	
(Chhabra *et al.*, 2017)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	Hexon	qPCR and conventional PCR followed by sequencing	300	68.4 ± 5.3	99.2 ± 1.0	USA
	Biofire FilmArray	NA			97.4 ± 1.8	97.7 ± 1.7	
	Luminex xTAG® Gastrointestinal Pathogen Panel	NA			57.9 ± 5.6	100	
(Liu *et al.*, 2013)	TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA)	Hexon	PCR-Luminex	109	100	80.3 ± 7.5	Tanzania and Bangladesh
(Liu *et al.*, 2011)	Luminex xTAG® Gastrointestinal Pathogen Panel		Singleplex realtime qPCR (hexon)	229	100	100	Tanzania

(Higgins *et al.*, 2011, Chhabra *et al.*, 2017, Liu *et al.*, 2013 and Liu *et al.*, 2011)

2.3 Next Generation Sequencing

Over the past 20 years, DNA sequencing technologies have rapidly evolved with the development of faster, high sequence throughput and cost-effective sequencing platforms. Next generation sequencing, specifically second-generation Illumina platform, utilizes a sequencing by synthesis method which adds fluorescent-labelled deoxynucleotides to a DNA strand and the labelled nucleotides are read using direct imaging. NGS offers high sequencing depth, discovery power, mutation resolution, sample and data throughput compared to conventional sequencing methods such as sanger sequencing (Gu, Miller, and Chiu, 2019).

Advances in the deep sequencing field have provided a platform for characterization of known and novel viruses from fecal samples. Fecal samples have been

documented to contain significant amounts of bacterial, bacteriophage and viral nucleic acids which are important for identifying pathogens that may be associated with diarrhoea (Cotten *et al*., 2014; Yinda *et al*., 2019). Increased sequence recovery due to high mammalian viral titres in fecal material can be very helpful in describing viral coinfections associated with diarrhoea. The sensitivity and specificity of deep sequencing in investigating viral diarrhoeal aetiology has not been extensively explored (Mohammad *et al.,* 2020). Deep sequencing coupled with a custom viral discovery pipeline has been shown to have a similar sensitivity with real-time PCR in the detection of norovirus. A strong negative correlation was observed between the real-time PCR Ct values and the norovirus reads (Cotten *et al*., 2014).

2.4 Effects of Mismatches at the Primer and Probe Binding Sites

Mismatches as a result of genetic drift in the primer and probe binding sites have been reported to affect detection and typing of enteric viruses including RVA (Mitui *et al.*, 2012; Parraand Espinola, 2006). Poor sensitivity in norovirus detection by endpoint PCR assays and HAdV type 40/41 has been correlated to mismatches at the primer binding sites (Jothikumar *et al*., 2005; Rooney *et al*., 2014). Mismatches present in the primers and probes commonly used in real-time PCR assays for the detection of sapovirus have been reported to cause detection failure or causes an increase in the quantification cycle value (Oka *et al*., 2019). Mismatches within 5 bases of the primer's 3' end or an accumulation of more than four mismatches have the greatest effect on the quantification cycle and the effect is also dependant on the whether the substitution is a transition or a transversion (Lefever *et.al*, 2013).

In summary, diarrhoea is one of the leading causes of morbidity and mortality in children under the age of five years globally. The diagnostic methods used to investigate viral diarrhoeal aetiology are important in routine surveillance of these pathogens and estimation of pathogen-specific disease burden. Except for RVA where ELISA is recommended as the gold standard for routine diagnosis and surveillance, other enteric viruses have no specific gold standards for their detection but do have reference standards such as the PCR that are not reliable due to their varied diagnostic performance. The sensitivity and specificity of these methods tends

to differ depending on the virus, viral load, age, location, time of stool collection compared to symptom onset and the choice of the reference standard. The evolution of viruses leads to introduction of point mutations at the primer and probe binding sites and reassortment events from other virus strains or zoonosis also affects the diagnostic performance of these diagnostic methods. Therefore, careful consideration in choosing or designinga diagnostic method to be used in routine viral diarrhoea surveillance is key to ensure correct estimation of diarrhoeal disease burden and understanding of transmission patterns.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Design

The study was a repeated cross-sectional study, comparing the diagnostic performance of RVA ELISA, real-time PCR, TAC and NGS in investigating viral diarrhoeal aetiology in the pre- (2013) and post-rotavirus (2016) vaccine introduction period in coastal, Kenya.

3.2 Study Site and Population

The study utilized stool specimens collected from surveillance of children (0-12 years) admitted to KCH with diarrhoea symptoms. The Kilifi Health and Demographic Surveillance System (KHDSS) was set up in 2000 to capture surveillance data of individuals residing close to KCH (Figure 3.1). The KHDSS area has approximately 260,000 residents and covers an area of 891km^2 (Scott *et al*., 2012). KCH is located within Kilifi town and is the only government facility offering inpatient paediatric services in the KHDSS area. Study participants were either residents of both KHDSS and Kilifi County or just Kilifi County. Stool specimen collection was done between January to December 2013 and January to December 2016. The collected specimens were analysed at KWTRP in Kilifi. A subset of specimens that were positive for RVA by ELISA in the year 2013 were sent for sequencing at the Wellcome Sanger Institute in the United Kingdom and subsequent sequence data analysis was done at the KWTRP Kilifi.

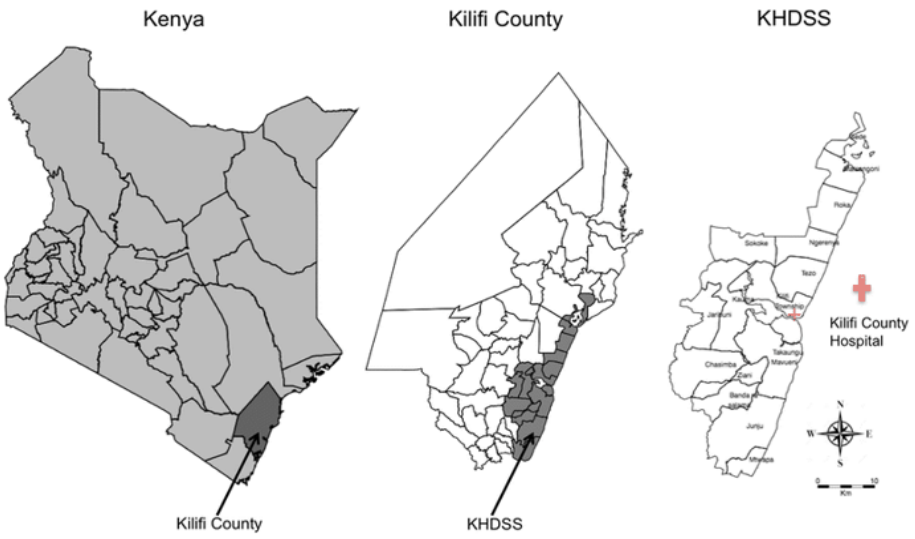


Figure 0.1: Map of Kenya showing Kilifi County and Kilifi Health Demographic and Surveillance System (KHDSS) area, coastal Kenya (Kamau et al., 2017).

3.2.1 Inclusion Criteria

1 Children (0-12 years) presenting with a history of acute diarrhoea defined as three or more watery stools passed during a 24-hour period.

2 Availability of an archived stool sample.

3 Sample tested by the comparable diagnostic methods.

3.2.2 Exclusion Criteria

Failure to consent by the guardian or parent.

3.3 Sample Size

Sample size determination was done with reference to Hajian-Tilaki (2014), for comparing the diagnostic accuracy of two diagnostic methods.

$$n = \frac{[Z_{\frac{\propto}{2}}\sqrt{2\,x\,\bar{P}(1-\bar{P})} + Z\beta\sqrt{P_1(1-P_1)+P_2(1-P_2)}]^2}{(P_1-P_2)^2}$$

P_1 and P_2 represent the expected sensitivity/specificity of the two diagnostic methods as reported elsewhere. P_1= 1.0 (real-time PCR) and P_2 = 0.90 (TAC) or P = 0.75 (Chhabra *et al*., 2017; Gautam *et al*., 2013; Mijatovic-Rustempasic *et al*., 2013).

Þ represents the average of the two sensitivities (0.95).

Margin of error = 0.05 and power 80%.

$Z_{\alpha/2}$ (1.96) and Z_{β} (0.84) are the standard Z values corresponding to probability of type I and II errors respectively.

$$100 = \frac{[1.96\sqrt{2\ x\ 0.825(1-0.825)} + \ 0.84\sqrt{0.75(1-0.75)+0.90(1-0.90)}]^2}{(0.90-0.75\ )^2}$$

$$73 = \frac{[1.96\sqrt{2\ x\ 0.95(1-0.95)} + \ 0.84\sqrt{0.90(1-0.90)+1(1-0)}]^2}{(1.0-0.9)^2}$$

In this study, all the specimens; 275 and 280 collected in 2013 and 2016 respectively were included in the study. This sample size provided a power of 99% and an effect size of 10-15% for comparing the sensitivity of the two PCR methods and or the ELISA method.

3.4 Laboratory Procedures

3.4.1 ELISA

Stool specimen were screened for RVA VP6 antigen using ProSpecT^{TM} kit (Oxoid, Basingstoke, UK) following the manufacturer's protocol. ProSpecT^{TM} kit uses polyclonal antibodies raised against the RVA VP6 antigen. A 10% dilution of the fecal specimen was prepared by mixing 100µl of faeces with 1ml of sample diluents. After 10 minutes, two drops of the enzyme conjugate were mixed with the diluted fecal sample, incubated for one hour at room temperature then washed using 400µl

of wash buffer. Two drops of the substrate were added to each well and incubated for ten minutes. To terminate the enzyme substrate reaction, two drops of stop solution were added to the wells and results read within 30 minutes by visual determination of reactivity and reading of optical density (OD) values. The results were interpreted as follows: Photometrically, specimens having absorbency values higher than the cut off value (0.200 added to the negative control value) were determined as positive. However, a result of within 0.010 absorbency units of the cut off value was interpreted as equivocal and the test repeated.

3.4.2 Singleplex Real-Time PCR

3.4.2.1 Nucleic Acid Extraction

Due to limited availability of extraction kits, fecal specimens from 2013 and 2016 were extracted using the cador pathogen protocol (Qiagen, Venlo, Netherlands) and QIAamp Fast DNA Stool kit (Qiagen, Venlo, Netherlands) respectively as per the manufacturer's instructions. For samples collected in 2016, there was a bead beating pre-treatment step as explained in section 0. For the cador pathogen protocol (Qiagen, Venlo, Netherlands), 100µl fecal sample was mixed with one ml of water and vortexed vigorously. The specimen was then centrifuged for one minute at 14000xg. The supernatant (200µl) was then mixed with buffer VXL (180µl) and proteinase K (20µl). Lysis was completed at a temperature of 70°C with constant agitation for 15 minutes. The remaining steps were performed by the robot on QIAcube HT (Qiagen, Venlo, Netherlands) as per the manufacturer's protocol.

3.4.2.2 cDNA Synthesis and Real-Time PCR

First strand cDNA synthesis was done using the Omniscript Reverse Transcriptase kit (Qiagen, Venlo, Netherlands). A 20µl mixture of 10X buffer RT(2µl), 5mM dNTP mix (2µl), random hexamers (2µl), Omniscript reverse-transcriptase (1µl), RNA free water (7.75µl), RNase inhibitor (0.25µl) and template RNA (5µl) was prepared, vortexed and briefly centrifuged. The mixture was then incubated for 60 minutes at 37ºC and the reverse transcription stopped on ice.

Singleplex real-time PCR assays were performed in a total volume of 20µl containing QuantiFast master mix (12.5µl), primer/probe mix (1.25µl), H_20 (3.75µl), rox dye (0.5µl) and cDNA (2µl). The primers and probes used are adopted from elsewhere (Appendix I*:*) (van Maarseveen *et al.,* 2010). The prepared PCR plates were then run in an Applied Biosystems™ 7500 Real-Time PCR System (ThermoFisher Scientific, Darmstadt, Germany). PCR was done using the following conditions: initial PCR activation and two-step cycling for five minutes at 95°C, denaturation for 30 seconds at 95°C and annealing for 45 seconds at 60°C. Amplification data was analysed using the Ct method on the Applied Biosystems™ 7500 Real-Time PCR v2.0 software (ThermoFisher Scientific, Darmstadt, Germany) and exported for subsequent analysis and manipulation within the R environment.

3.4.3 TaqMan Array Card

3.4.3.1 Total Nucleic Acid Extraction

Fecal specimens (0.2grams/200µl) were lysed with InhibitEX/PhHV/MS2 solution, bead beaten at a maximum speed for three minutes using 370mg of glass beads (Sigma, Merck, Darmstadt, Germany) in a BioSpec bead beating instrument (BioSpec, Bartlesville, USA) and incubated at 95°C for five minutes. The mixture was then vortexed for 15 seconds and centrifuged at 20,000xg for one minute. A total of 600µl of the supernatant was mixed with 25µl proteinase K and 600µl AL buffer, vortexed, incubated at 70°C for ten minutes. The lysate was then mixed with 600µl of ethanol, vortexed and centrifuged briefly to remove drops inside the tube lid. All the 1800µl was transferred into a QIAamp spin column in 600µl aliquots while centrifuging each time at 16000xg for 1.5 minutes and replacing the collection tube. Buffer AW1 was then added to the spin column and centrifuged at 16000xg for 4.5 minutes followed by addition of 500µl of AW2 and centrifugation at 16000xg for 4.5 minutes. Total nucleic acid was eluted using 200µl of ATE buffer and stored at -80°C. MS2 phage was used as an extrinsic control during extraction to evaluate extraction and amplification efficiency.

3.4.3.2 TAC Real-Time PCR Assay

The TAC reaction mix contained 60µl of nucleic acid free water, 25µl of the enzyme mix (TaqMan® Fast Virus 1-Step Master Mix, ThermoFisher Scientific) and 15 µl of the total nucleic acid extract. A 100µl TAC reaction mix was then transferred to the Public Health England Custom TaqMan Array 384-well Card (ThermoFisher Scientific) containing precoated lyophilized primers and probes. The TAC was centrifuged twice at 1200 rpm for two minutes, sealed and placed into a QuantStudio 7-flex thermal cycler (ThermoFisher Scientific, Darmstadt, Germany). The PCR conditions were: 50°C for five minutes for the reverse transcription step, 95°C for 20 seconds for the initial DNA melting and inactivation of RT reaction, followed by 45 cycles of 95°C for one second and 60°C for 20 seconds. Amplification data was analysed using the Ct method using the QuantStudio 7-flex software (ThermoFisher Scientific, Darmstadt, Germany) with a threshold value of 0.2 fluorescence units. The data was then transformed into an excel file and used for subsequent analysis and manipulation within the R environment.

3.4.4 Next Generation Sequencing

A total of 69 samples from the year 2013 that were positive for RVA by ELISA were sequenced using Illumina Hiseq 2500 as described elsewhere (Phan *et al*., 2016). A total of 110µl of a 50% stool suspension in PBS was centrifuged for ten minutes at 10,000xg. Non-encapsidated DNA in the samples was degraded by addition of TURBO DNase (Ambion, ThermoFisher Scientific, Darmstadt, Germany). Virion-protected nucleic acid was extracted with elution in sterile water using the Boom method (Boom *et al*., 1990). Non-ribosomal random hexamers that avoided transcription of rRNA were used for reverse-transcription and second-strand DNA synthesis was done using Klenow fragment 3'–5' exo (New England Biolabs Inc., UK) Extracted nucleic acids were then purified using phenol/chloroform and ethanol precipitation. Illumina libraries were prepared for each sample by shearing the nucleic acids to 400-500 nucleotide in length and indexing the samples separately using adapters. The samples were then pooled and sequenced using Illumina Hiseq 2500 machine that generates 250 bp paired-end reads per sample.

3.4.5 Conventional PCR and Partial Sequencing

ELISA negative but TAC positive samples were retested using VP4 and VP7 PCR followed by dideoxy sanger sequencing. Partial fragments of RVA VP4 and VP7 genes were amplified in a one-step conventional reverse transcriptase PCR reaction using the following primers: VP4F, 5'-TAT- GCTCCAGTNAATTGG-3', VP4R 5'-ATTGCATTTCTTTCCATAATG-3', VP7F, 5'-ATGTATGGTATTGAATATACCAC-3', VP7R 5'-AACTTGCCACCATTTTTTCC-3' (Mwanga *et al*., 2020). A 2% agarose gel electrophoresis was done to confirm successful amplification of VP4 (660bp) and VP7 (881bp) fragments. Samples with the expected band size on the gels were purified using GFX DNA purification kit (GFX-Amersham, UK) following the manufacturer's protocol. They were then sequenced using Big Dye Terminator 3.1 (Applied Biosystems, California, USA) chemistry on an ABI Prism 3130xl Genetic Analyser (Applied Biosystems, California, USA).

3.5 Data Management

ELISA, real-time PCR and TAC results were retrieved from their respective instruments and merged into an excel file according to individual sample identifiers. Samples that were not tested by all the three methods for RVA were eliminated from subsequent analysis. Similarly, samples that were not tested by both real-time PCR and TAC for astrovirus, sapovirus, norovirus genogroup II and adenovirus were also excluded from subsequent analysis. The datasets were then imported into an R project in the R statistical software: R version 3.5.0 (R Core Team, 2018). NGS data was stored on a local high-performance computing cluster at KWTRP Kilifi and access was password restricted. Access to the datasets was restricted to project investigators only.

3.6 Data Analysis

All statistical analysis was done using R version 3.5.0 (R Core Team, 2018) and bioinformatics analysis was done using softwares available on the local server (Figure 3.2).

3.6.1 Statistical Analysis

Sensitivity, specificity and agreement were measures used to compare the diagnostic performance of the different diagnostic methods. Calculation of sensitivity and specificity was in accordance with the following composite reference standard rules: (*i*) True positives were positive results from ≥ 2 diagnostic methods; (*ii*) True negatives were negative results by ≥ 2 diagnostic methods; (*iii*) False positives were positive results by one method but not confirmed by the other two methods; (*iv*) False negatives were negative results by one method but positive by the other two methods. Percent agreement and unweighted cohen kappa (κ) statistic was used to estimate agreement between TAC and real-time PCR for detection of each of the viruses. Fleiss kappa was used to estimate agreement for more than two methods. Both Cohen and Fleiss' kappa estimates were reported with a 95% confidence interval and p-value, respectively. The kappa statistic was interpreted as follows: The agreement is expected to lie between zero and one whereby for complete agreement k =1; if observed agreement is equal to chance k=0; and if observed agreement is greater than by chance $k > 0$. Agreement is rated as poor, slight, fair, moderate, substantial and almost perfect if $k < 0.0$, $0.0 \leqslant 0.20$, $0.21 \leq 0.40$, $0.41 \leq 0.60$, $0.61 \leq 0.80$ and > 0.80, respectively (Landisand Koch, 1977). Differences in the proportions were calculated using the test for equal proportions. The distribution of TAC and singleplex real-time PCR Ct values was compared using Wilcoxon rank-sum test. Additionally, estimating correlation between real-time PCR and TAC Ct values was done using the Spearman rank correlation test.

3.6.2 Primer and Probe Mismatch Analysis

Available viral sequences for RVA, sapovirus, astrovirus, norovirus genogroup II and adenovirus type 40/41 were downloaded from GenBank. Additional local RVA sequences were also included (Makori *et al*., manuscript). The time period was limited to 2010-2019. Primers and probes used in real-time PCR and TAC methods were aligned to their respective viral sequences to check for mismatches using Geneious prime v.2019.0.4.

3.6.3 Bioinformatics Analysis

NGS short read data was analysed using a modified pipeline adopted from elsewhere (Phan *et al*., 2016) and summarized below (Figure 3.2). The following quality control checks were done on the raw reads for each sample using FASTQC v.0.11.8: basic statistics, per base sequence quality, GC content, sequence length distribution, overrepresented sequences and Kmers, and adapter sequences (Andrews, 2010). The reads were trimmed to remove adapters and low quality reads using QUASR v.7.03 (Watson *et al*., 2013). The trimmed reads were then mapped to a human reference database created from the current version of the human genome GRCh38.p12 using bowtie2 v.2.2.5 (Langmead, 2013). Samtools v.1.7 (Li *et al*., 2009) was used to convert sam files to bam files, recover unmapped reads and sort them. Bedtools v.2.25.0 (Quinlanand Hall, 2010) was used to convert bam files to fastq format. The trimmed reads were *de novo* assembled into contigs using SPAdes v.3.13.0 (Bankevich *et al*., 2012). Viral taxonomic classification was done directly from the reads using Kraken v.0.10.5-beta (Woodand Salzberg, 2014) and MetaPhlan2 v.2.0 (Truong *et al*., 2015). Kraken was also used to assign taxonomic labels to assembled reads (contigs). The identified contigs were checked for open reading frames (ORFs) and compared with reference sequences from GenBank. All the commands and parameters used are listed in the Appendix . Reads from sanger sequencing were assembled into contigs using Sequencher version 5.4.6 (Gene Codes Corp Inc., USA). G and P genotypes were determined using Virus Pathogen Database and Analysis Resource (ViPR) database and or the RotaC version 2.0 classification tool.

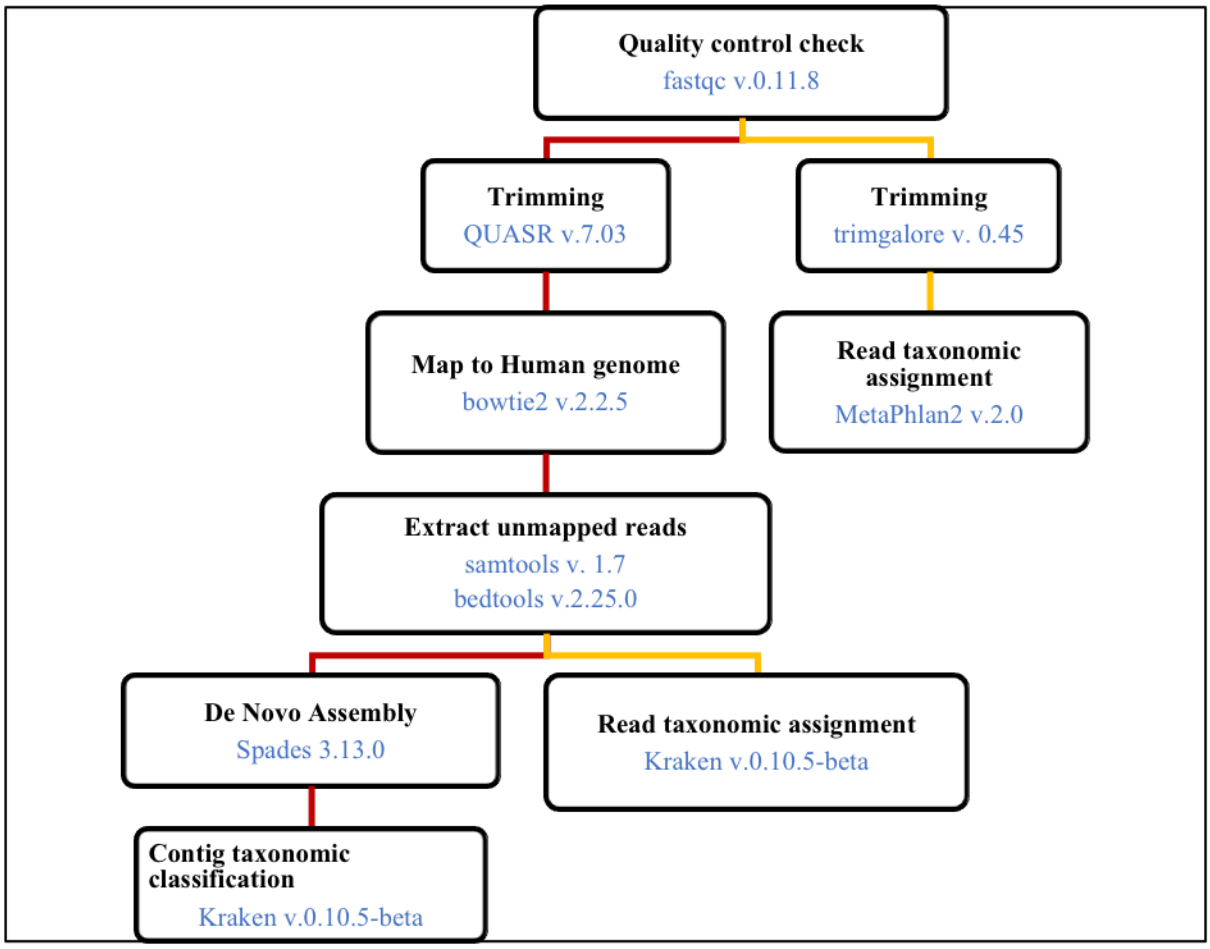


Figure 0.2: Summary of the bioinformatics workflow for analysis of the NGS short read data.

CHAPTER FOUR

RESULTS

4.1 Comparison of ELISA, Real-Time PCR and TAC in RVA Detection

4.1.1 Population Characteristics

Out of 725 chidren aged <12 years presented with diarrhoea at KCH in 2013 (n=363) and 2016 (n=362), only 556 (767%) consented to the study and gave a stool sample (Figure 4.1). The fecal specimen collected were tested for RVA by EIA, real-time PCR and TAC. Only specimens that were tested by all the methods in 2013 (n=237) and 2016 (n=252) were included in subsequent analysis. The age and gender of the children that were included in the study were recorded in Table 4.1. The median age of the study participants in 2013 and 2016 was 11 and 14 months, respectively and the difference was statistically significant (p-value = 0.02).

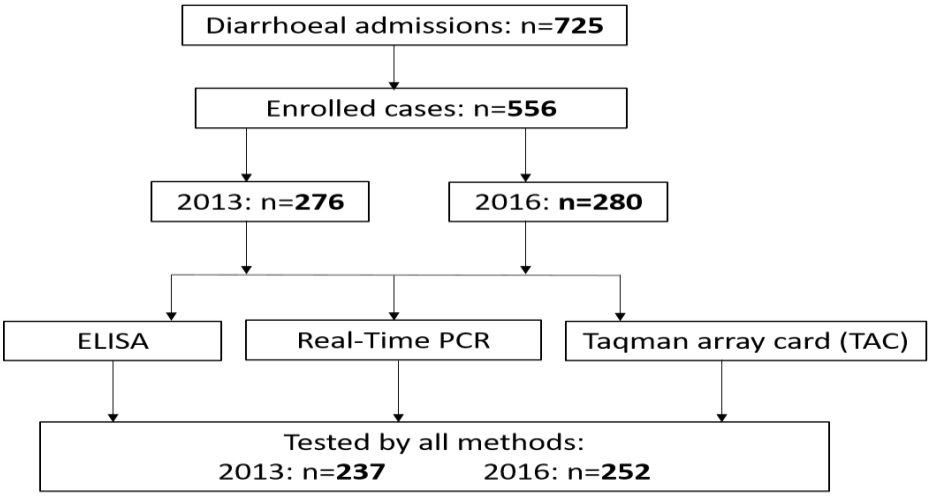


Figure 0.1: Flowchart of enrolled cases in the study in 2013 and 2016, and the samples tested by different methods

Table 0.1: Characteristics of study participants enrolled in the study in 2013 and 2016

Characteristic	2013	2016	p-value
Median age month (IQR)	11 (7-21)	14 (8-25)	0.02
Gender: Female (%)	101 (42.6%)	114 (45.2%)	0.62

4.1.2 Performance of ELISA, TAC and Real-Time PCR in Rotavirus A Detection

The sensitivity and specificity of the ELISA, TAC and Real-Time PCR was estimated using a composite reference method as described in 0. The sensitivity of ELISA, real-time PCR and TAC was 93.8%, 98.5% and 96.9% respectively in 2013, and 68.9%, 93.3% and 97.8% respectively in 2016 (Figure 4.12). The reduction in the sensitivity of the EIA in RVA detection post-vaccine introduction was statistically significant (p-value < 0.05). Further, the EIA method had a significantly lower sensitivity than real-time PCR (p-value: 0.007) and TAC (p-value = 0.0007) in 2016 unlike 2013 (Figure 4.2). The specificities of all the three methods were >96% both in the years 2013 and 2016 (Figure 4.3). The performance of the diagnostic methods for the detection of RVA are summarized in Appendix II.

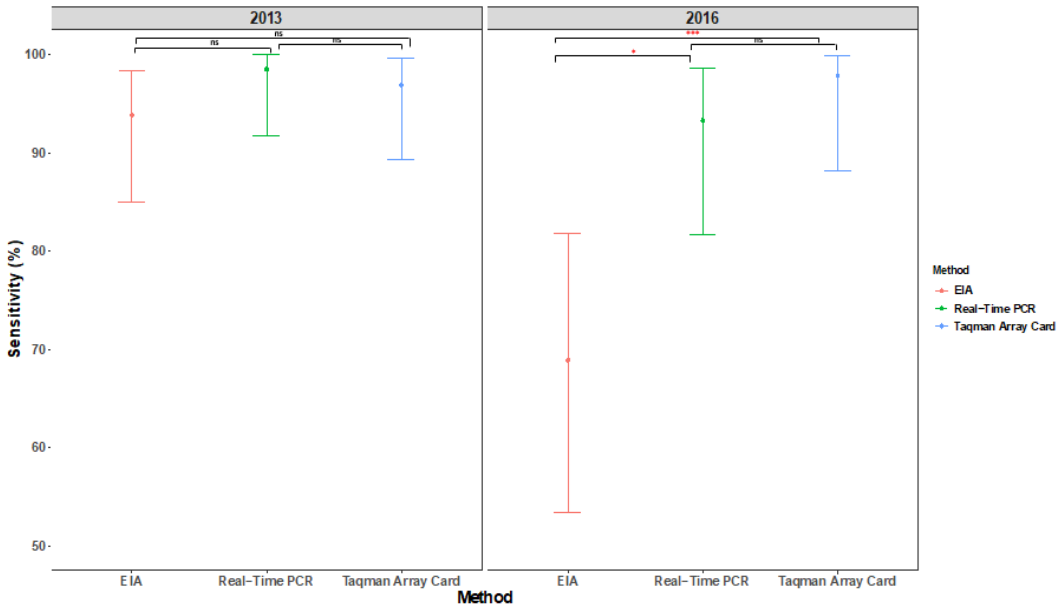


Figure 0.2: A plot of the sensitivity and specificity of the EIA, real-time PCR and TaqMan Array card in RVA detection.

The error bars show the confidence intervals for the calculated sensitivity (clopper-pearson exact method). ns: not significant; *: p-value < 0.05; **: p-value <0.01, ***: p-value < 0.001.

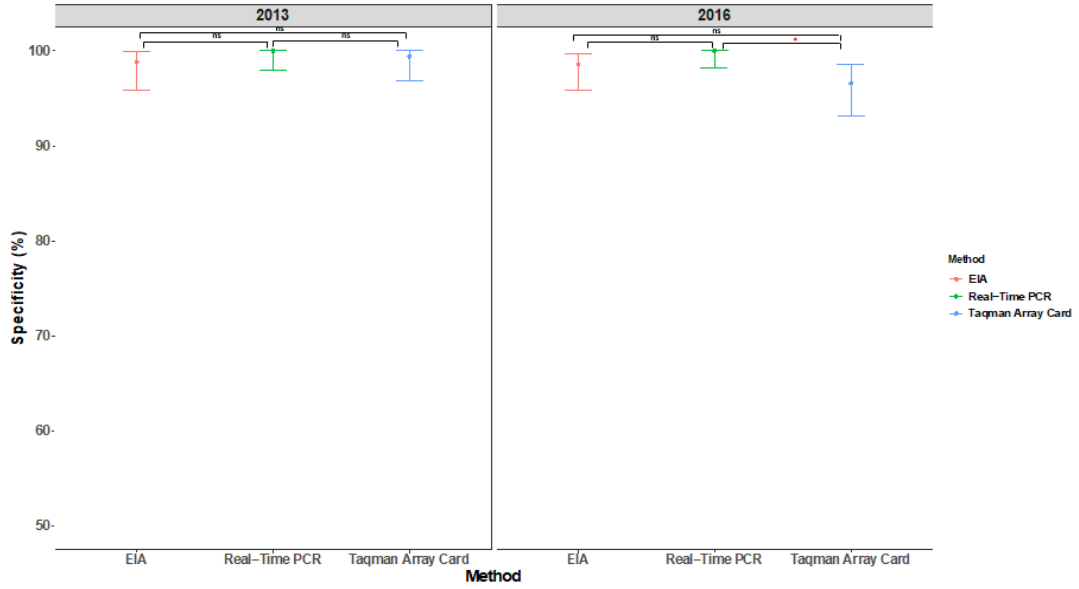


Figure 0.3: A plot of the specificity of EIA, real-time PCR and TaqMan Array card in RVA detection.

The error bars show the confidence intervals for the calculated specificity (clopper-pearson exact method). ns: not significant; *: p-value < 0.05; **: p-value <0.01, ***: p-value < 0.001.

Using Fleiss' kappa statistic to assess the agreement between the three methods, a substantial agreement (defined as $0.61 \leq$ fleiss $k \leq 0.80$) was observed among EIA, real-time PCR and TAC in the detection of RVA in 2016 and it was lower than a very strong agreement (defined as fleiss $k > 0.80$) observed in 2013 (Table 4.2).

Table 0.2: Comparison of RVA detection by ELISA, real-time PCR and TAC in 2013 and 2016

Year	RVA_EIA	RVA_TAC	RVA_PCR	Frequency	Fleiss' kappa statistic
2013	Neg	Neg	Neg	168 (71.2)	Kappa = 0.93 p-value <0.01 n = 237
	Pos	Neg	Neg	3 (1.3)	
	Neg	Pos	Neg	2 (0.8)	
	Pos	Pos	Neg	1 (0.4)	
	Neg	Neg	Pos	0	
	Pos	Neg	Pos	1 (0.4)	
	Neg	Pos	Pos	3 (1.3)	
	Pos	Pos	Pos	59(24.9)	
2016	Neg	Neg	Neg	197 (78.2)	Kappa = 0.74 p-value <0.01 subjects = 252
	Pos	Neg	Neg	4 (1.6)	
	Neg	Pos	Neg	7 (2.8)	
	Pos	Pos	Neg	4 (1.6)	
	Neg	Neg	Pos	0	
	Pos	Neg	Pos	0	
	Neg	Pos	Pos	14 (5.6)	
	Pos	Pos	Pos	26 (10.3)	

Agreement is rated as poor if $0.00 \leq k \leq 0.20$, fair if $0.21 \leq k \leq 0.40$, moderate if $0.41 \leq k \leq 0.60$, substantial if $0.61 \leq k \leq 0.80$ and very strong agreement if $k > 0.80$.

4.1.3 Rotavirus Detection Frequency Pre-Post Vaccine Introduction

Although there were subtle differences in absolute numbers for the year 2013, no significant difference was observed between the prevalence estimates of the three methods ($p = 1$) (Figure 4.4). In 2016, TAC detected more positives than singleplex real-time PCR and EIA. However, the differences in the proportions were not statistically significant (p-value ≥ 0.05) (Figure 4.4).

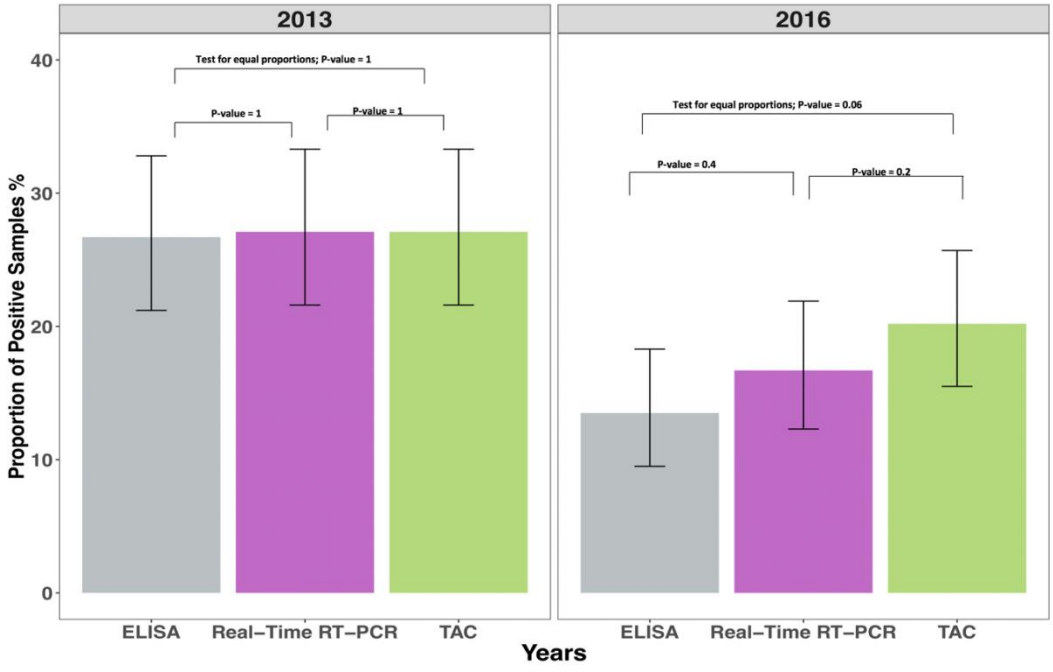


Figure 0.4: Proportion of RVA positive samples by ELISA, singleplex real-time PCR and TaqMan Array Card (TAC).

The error bars represent 95% confidence interval for the proportions (clopper-pearson exact method).

A comparison of the proportions of RVA positive samples detected by the three methods in 2013 versus 2016 showed that the reduction in RVA detected cases was statistically significant by both EIA (p-value = 0.0005) and real-time PCR (p-value = 0.008) but not by TAC (p-value = 0.1), see Figure 4.5. This finding suggests that despite introduction of rotavirus vaccination programme in Kenya in 2014, reduction in the proportion of hospitalised diarrhoea cases positive for RVA cannot be confirmed by TAC unlike EIA and real-time PCR in the instance of the year 2016 versus 2013 (Figure 4.5).

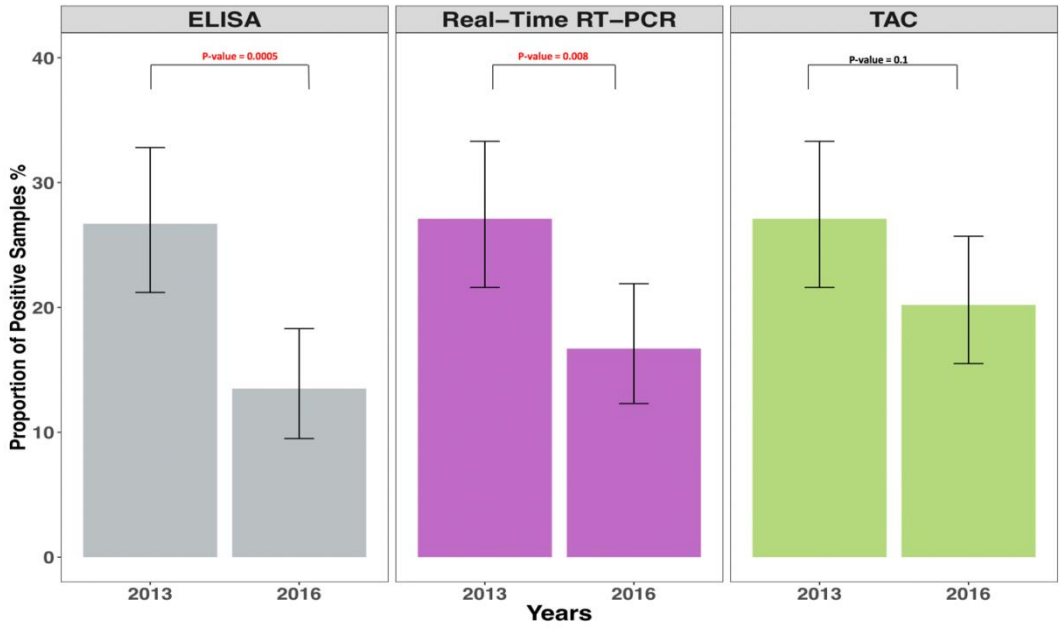


Figure 0.5: Proportion of RVA positive samples by ELISA, singleplex real-time PCR and TaqMan Array Card (TAC)

The error bars represent the confidence intervals for the proportions (clopper-pearson exact method).

4.1.4 Distribution of TAC and Real-Time PCR Ct Values for RVA Positive Samples

A comparison of the distribution of the Ct values, a marker of the virus quantity present in the sample, showed that the difference in the means detected by TAC (25.5) and real-time PCR (27.2) in the year 2013 was statistically significant (Figure 4.6; T-test; p = 0.004). However, in 2016, no significant difference (p-value > 0.05) was observed between the mean Ct values of TAC (24.2) and real-time PCR (24.4). A comparison of the distribution of real-time PCR Ct values in 2013 and 2016 showed that the mean Ct value for the year 2013 was significantly greater than that for 2016 (p-value = 0.003). A comparison of the distribution of TAC Ct values showed that the differences in the means was not statistically significant difference in the years 2013 and 2016 (p-value = 0.1).

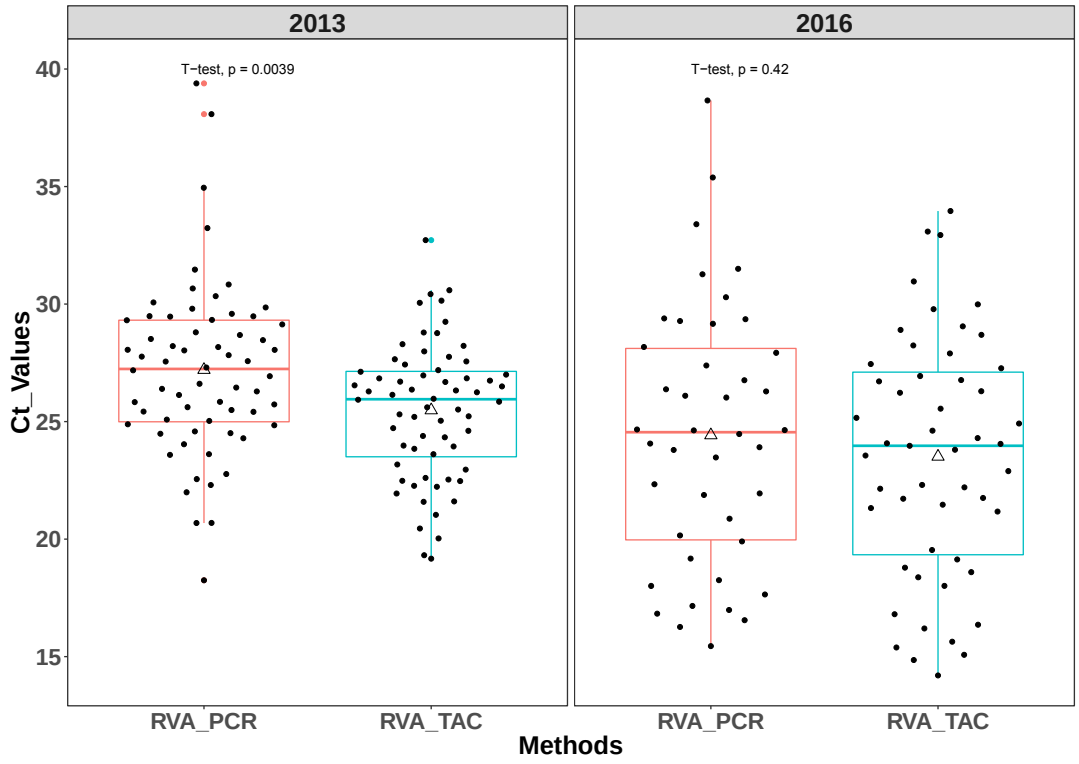


Figure 0.6: Boxplots showing the distribution of TAC and real-time PCR Ct Values of samples positive by either TAC or real-time PCR

4.1.5 Distribution of TAC and Real-Time PCR Ct values based on ELISA Status

No significant difference was observed between real-time PCR or TAC Ct values for samples that were either EIA positive or negative in 2013 and 2016 (p-value > 0.5) Fig. 4.7. The median TAC Ct values for samples that were positive and negative by ELISA were 25.51 (IQR:22.53-26.96) and 23.89 (IQR:21.20-28.15) respectively. The median real-time PCR Ct values for samples that were positive and negative by ELISA were 26.42 (24.43-29.19) and 24.45 (20.98-27.73) respectively. The differences in the distributions of the Ct values was not statistically significant (p-value > 0.5). In both the years 2013 and 2016, a strong positive spearman rank correlation (defined as 0.70-0.89) was observed between TAC and real-time PCR Ct values. Generally, real-time PCR had greater Ct values than TAC , see Appendix 3.

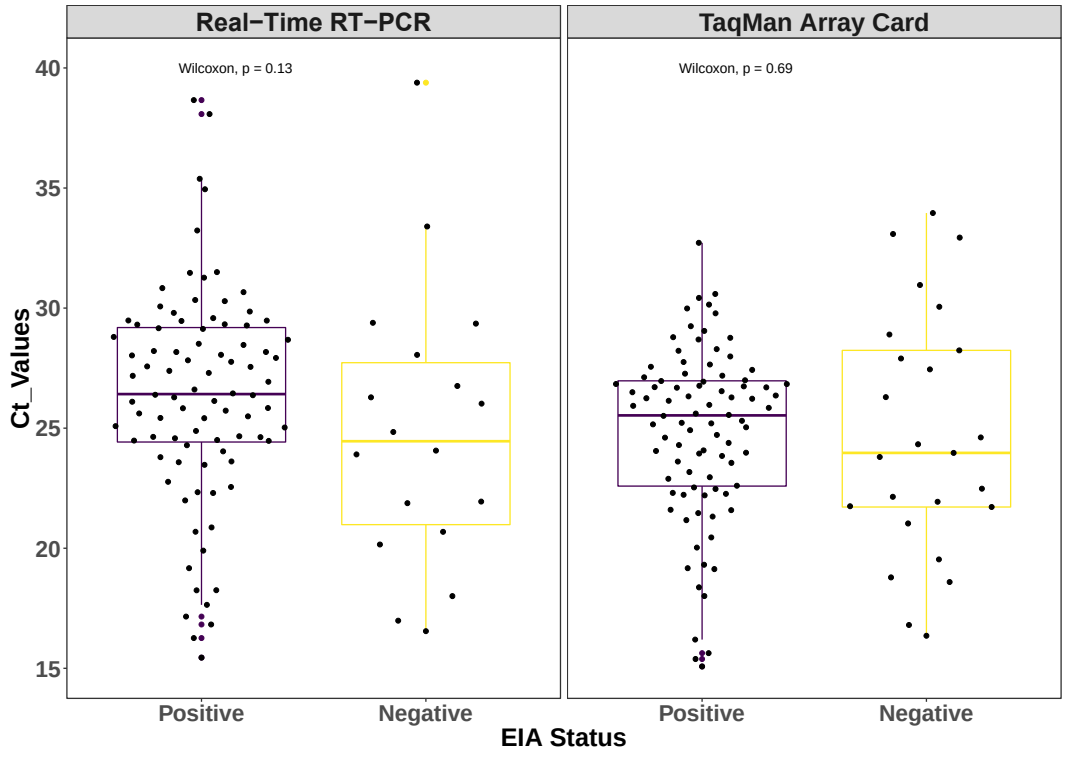


Figure 0.7: Boxplots showing the distribution of TAC and real-time PCR Ct Values in samples that were either positive or negative by EIA.

Positive: EIA positive; Negative: EIA negative

4.2 Comparison of Real-Time PCR and TAC in the Detection of Astrovirus, Adenovirus Type 40/41, Sapovirus And Norovirus Genogroup II

4.2.1 Population Characteristics

Kenya started nationwide RVA vaccination in July 2014 thus 2013 falls in the pre-vaccine period while 2016 falls in the post-vaccine introduction period. Out of the 556 enrolled cases, only 88.8% (n=242 for 2013 and n=252 for 2016) were tested by both real-time PCR and TAC and included in the subsequent analysis.

4.2.2 Real-Time PCR and TAC Viral Detection Frequency

TAC detected more positives than singleplex real-time PCR for the four viruses: Adenovirus (77 vs 52), astrovirus (25 vs 11), norovirus genogroup II (61 vs 42) and sapovirus (29 vs 16). The difference in the prevalence estimate by the two methods

was significant for astrovirus (p-value = 0.03) alone, in comparison to the prevalence estimates of all the other viruses (*Figure 0.8*).

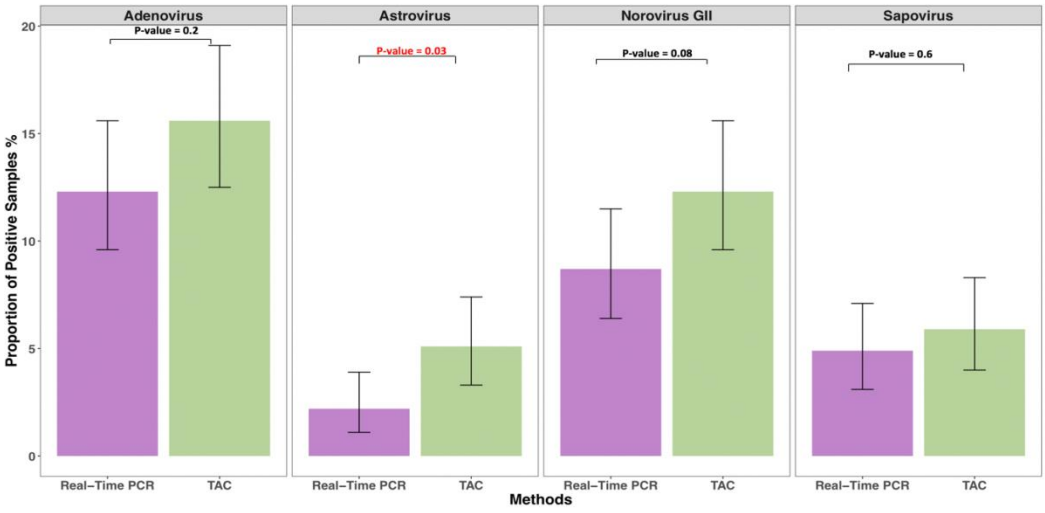


Figure 0.8: Proportions of positive samples detected by TaqMan Array Card and real-time PCR.

The error bars represent 95% confidence intervals for the proportions (clopper-pearson exact method). TAC: TaqMan Array Card

4.2.3 Assessing Agreement Between TAC and Real-Time PCR

Cohen kappa statistic was used to assess agreement between real-time PCR and TAC. A substantial agreement (defined as $0.61 \leq k \leq 0.80$) was seen between real-time PCR and TAC in the detection of adenovirus and norovirus genogroup II. However, in the detection of astrovirus and sapovirus, a moderate agreement (defined as $0.41 \leq k \leq 0.60$) was observed (Table 4.3).

Table 0.3: Agreement Levels Between real-time PCR and TAC in the detection of adenovirus, astrovirus, norovirus GII and sapovirus

Virus	Methods	Cohen kappa statistic (k)	Confidence Intervals (95%)	Overall Percent Agreement	Positive Percent Agreement	Negative Percent Agreement
Adenovirus	TAC~PCR	0.71	0.61, 0.80	93.1%	60.5%	92.3%
Astrovirus	TAC~PCR	0.48	0.28, 0.69	96.4%	33.3%	96.3%
Norovirus GII	TAC~PCR	0.76	0.67, 0.86	95.5%	65.1%	95.1%
Sapovirus	TAC~PCR	0.54	0.38, 0.71	95.3%	39.5%	95.2%

Agreement is rated as poor if $k<0.00$, slight if $0.00 \leq k \leq 0.20$, fair if $0.21 \leq k \leq 0.40$, moderate if $0.41 \leq k \leq 0.60$, substantial if $0.61 \leq k \leq 0.80$ and very strong agreement if $k > 0.80$ (Landisand Koch, 1977); TAC: TaqMan array card; PCR: real-time PCR

4.2.4 Comparison of the Distribution of TAC and real-time PCR Ct Values

From both TAC and real-time PCR methods, we obtained information on the cycle threshold (Ct) which is the PCR cycle at which the amplification signal rises above the background noise. There was no statistically significant difference in distribution of TAC and real-time PCR Ct values in the detection of all viruses except for sapovirus (p = 0.001) (Figure 4.9). A positive spearman correlation was observed between real-time and TAC Ct values for all viruses (

Appendix).

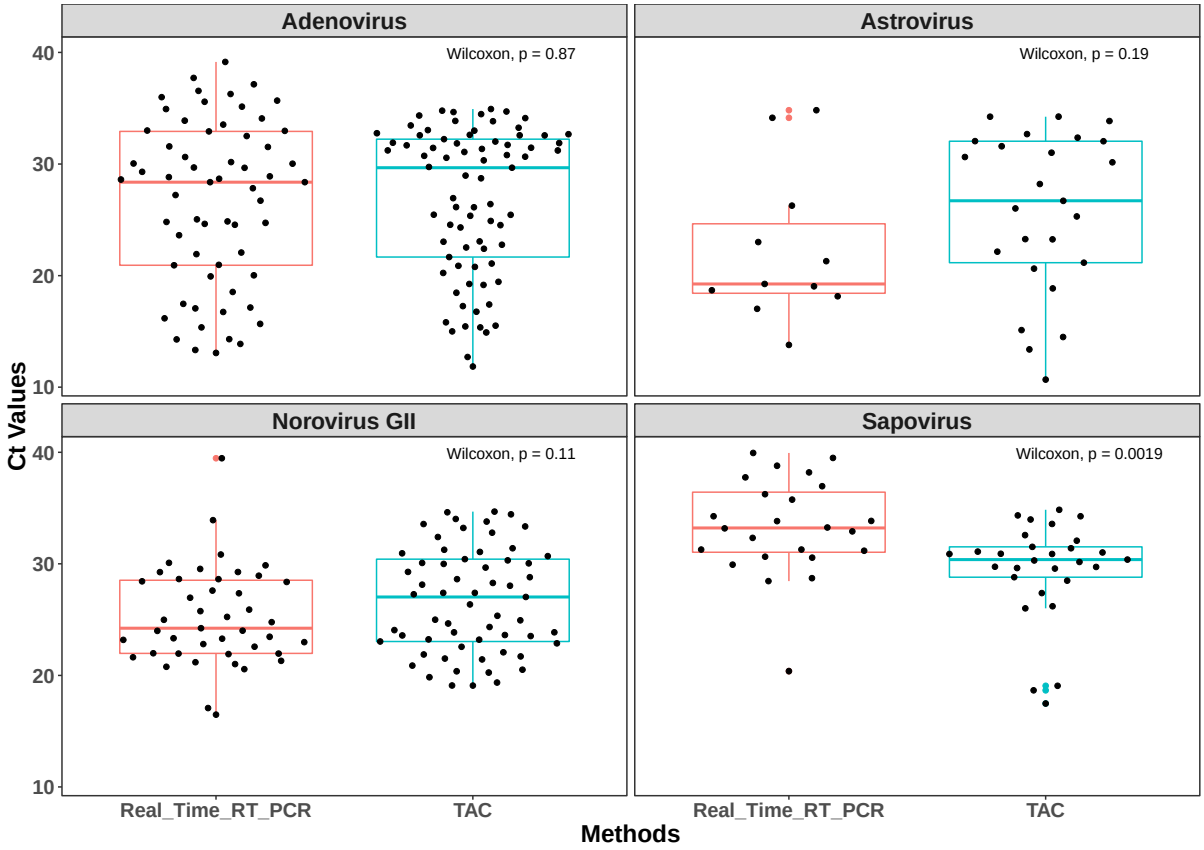


Figure 0.9: Boxplots for the Ct values of real-time PCR and TaqMan Array Card methods for Adenovirus type 401/41, astrovirus and sapovirus.

For each virus, the line inside the boxplot depicts the median and the box indicates the 25th and 75th quartiles.

4.3 Real-Time PCR and TAC Primer/ Probe Mismatches with contemporary sequences

The forward and reverse primers used for adenovirus detection, were identical in both the real-time PCR and TAC assays. However, the probes have only an identical 8 base region that is elongated at the 5'- and 3-'end in real-time PCR and TAC assays respectively. The primers and probes target the hexon gene and can detect all

enteric adenovirus serotypes. Adenovirus real-time PCR assay had two mismatches in the forward primer (G-A and C-G), two mismatches in the reverse primer (C-T and T-C) and three mismatches in the probe (C-T, C-T and T-C) none within 5 bases of the 3'-end (Figure 4.10). On the other hand, the adenovirus TAC assay had only a C-G mismatch in the forward primer (Figure 4.11).

The forward and reverse primers used for norovirus GII detection were identical in both assays. However, the probe in the TAC assay had a 3-base extension at its 5'-end (Appendix 1). The primers and probes targeted the RNA-dependent RNA polymerase (RdRp)-capsid junction and no nucleotide mismatches were observed in both the primers/probes binding sites for both assays.

For sapovirus detection, the primers and probes used in both assays target the RNA-dependant RNA polymerase capsid junction. In the real-time PCR assay, nucleotide mismatches within the sapovirus primer/probe binding sites were observed in sapovirus genogroup V and included six mismatches in the forward primer, two mismatches in the reverse primer and three mismatches in the probe. Mismatches observed were within 5 bases of the 3'-end included; forward primer: C-G, probe: T-C, reverse primer: T-C and A-C, see Figure 4.10. The TAC assay has two primer-probe sets; sapo#1 and sapo#2. The sapo#1 TAC primer set detected sapovirus genogroup I, II and IV, while the sapo#2 TAC primers and probe detect sapovirus genogroup V. No nucleotide mismatches were observed within the primers/probe binding sites in both sets.

The primers/probes used for astrovirus detection in the real-time PCR and TAC assays targeted the ORF1 and ORF2 regions respectively. The primers/probes in both assays could detect different astrovirus strains. Mismatches observed in the real-time PCR primer and probe sequences were in individual sequences rather than pronounced in all the sequences, see Figure 4.10. In the TAC assay, no mismatches were observed in the forward primer and probe binding sites, however, two mismatches (A-G and C-T) were observed in the reverse primer binding site, see Figure 4.11.

Primers/probes used for RVA detection in both the real-time PCR and TAC assays targeted the NSP3 region. No nucleotide mismatches were observed in the reverse primer and probe binding sites for both assays. However, in the real-time PCR assay, RVA forward primer had two mismatches at position 12 (G-A) and 15 (A-G), see Figure 4.10. One mismatch (A-G) was observed in TAC primer/probe set #Liu and three mismatches in set #Cam (A-G, G-A and A-G) , see Figure 4.11

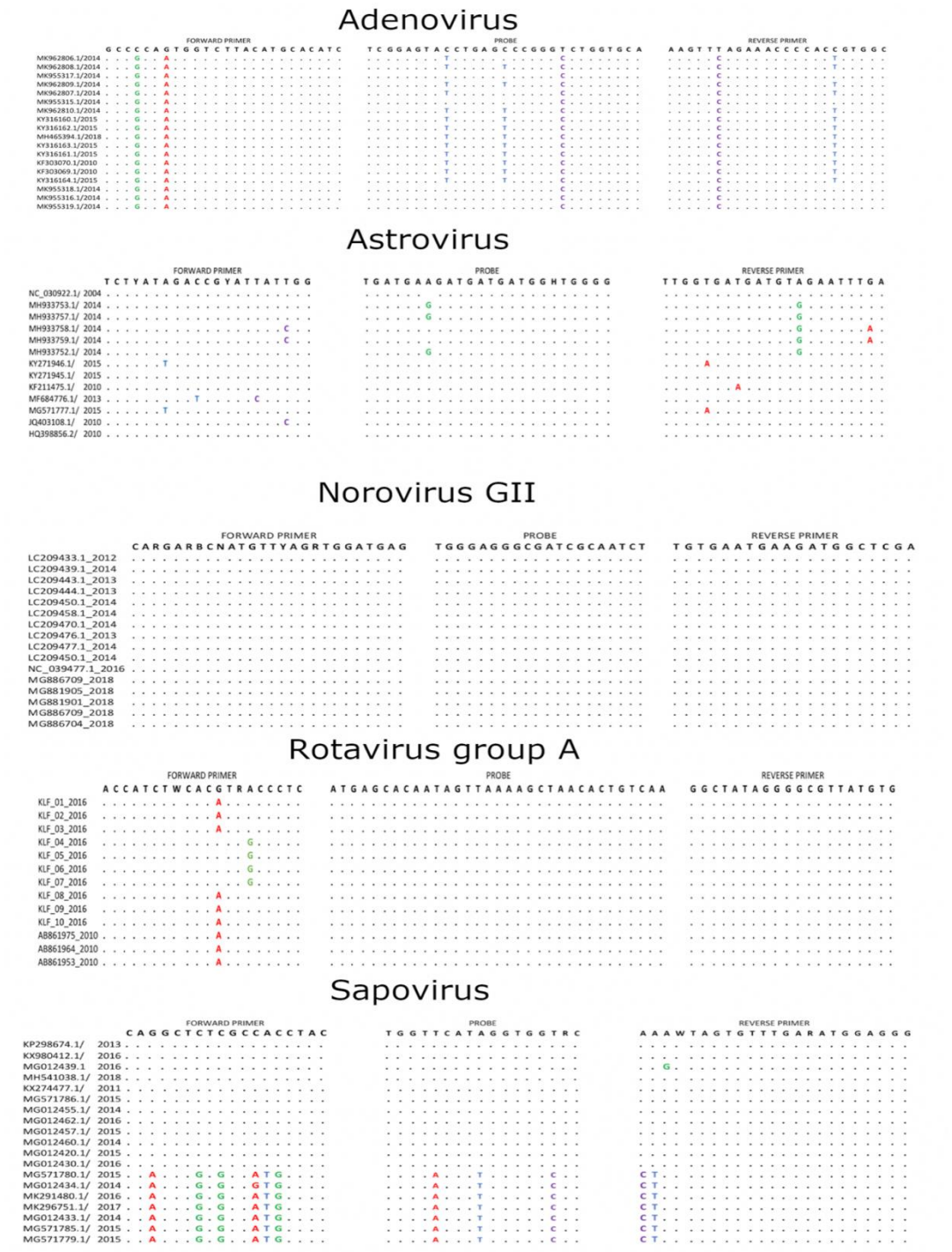


Figure 0.10: The real-time PCR primers and probes target sites for the five viruses were aligned using MAFFT v.7.3.

Nucleotide differences between the primer/ probe target sites and the global viral sequences were detected and highlighted. Dots show 100% identity with primer/

probe sequences (Lambisia *et al.*, 2020). The colors red, green, blue and purple represent the nucleotides A, G, T and C respectively.

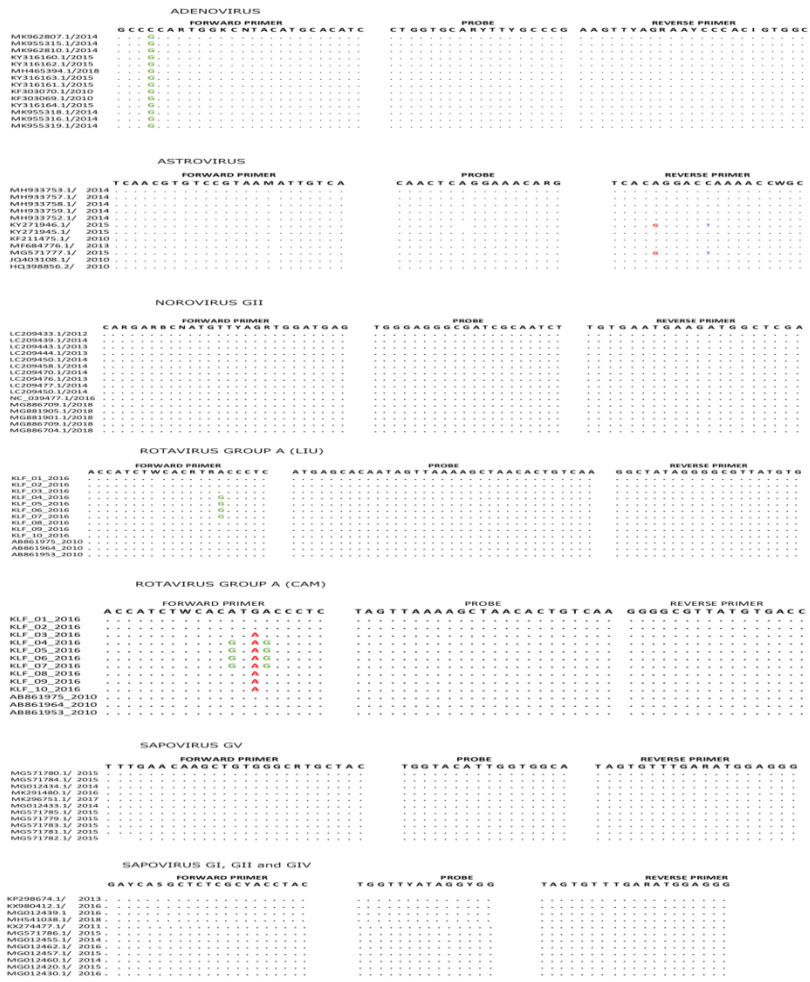


Figure 0.11: The TAC primers and probes target sites for the five viruses were aligned using MAFFT v.7.3.

Nucleotide differences between the primer/ probe target sites and the global viral sequences were detected and highlighted. Dots show 100% identity with primer/ probe sequences. The colors red, green, blue and purple represent the nucleotides A, G, T and C respectively.

4.4 Identification of Viral Coinfections Using Viral Metagenomics

4.4.1 Approach to Taxonomic Classification of NGS Reads

Sixty-nine stool samples from the pre-rotavirus vaccination period (2013) positive for RVA by antigenic detection (EIA) were sequenced by agnostic next generation sequencing (NGS) protocol to determine in an unbiased manner all viral species in the samples. The raw reads were processed as described in the methods section (Figure 0.2). Kraken and MetaPhlan programs were used to assign taxonomic labels to either or both contigs and unassembled sequences. Kraken assigns taxonomic labels by using exact alignment of k-mers (31bp) to its database that contains complete RefSeq genomes (Woodand Salzberg, 2014). Kraken assigned taxonomic labels to both contigs (consensus sequence from a set of overlapping assembled reads) and quality checked trimmed unassembled reads while MetaPhlan only assigned taxonomic labels to quality-checked unassembled reads. MetaPhlan profiles the composition of metagenomic sequencing data by mapping the reads to clade-specific marker genes and estimates the relative abundance of the profiled organisms at the species level or higher taxonomic levels (Truong *et al*., 2015).

4.4.2 Viral Coinfections

4.4.2.1 Coinfections from Taxonomic Classification of Contigs by Kraken

Out of 69 samples that had RVA contigs, 16 samples had enteric viral contigs associated with diarrhoea as coinfections. These viruses included; aichivirus A (n = 3), HAdV A (n = 1), HAdV D (n = 1), HAdV 54 (n=3), sapovirus (n = 1), human enterovirus (n = 2), enterovirus A (n = 2), enterovirus B (n = 3), enterovirus C (n = 5), rhinovirus C (n = 1), parechovirus (n = 5) and cosavirus A (n = 1) (Figure 4.12).

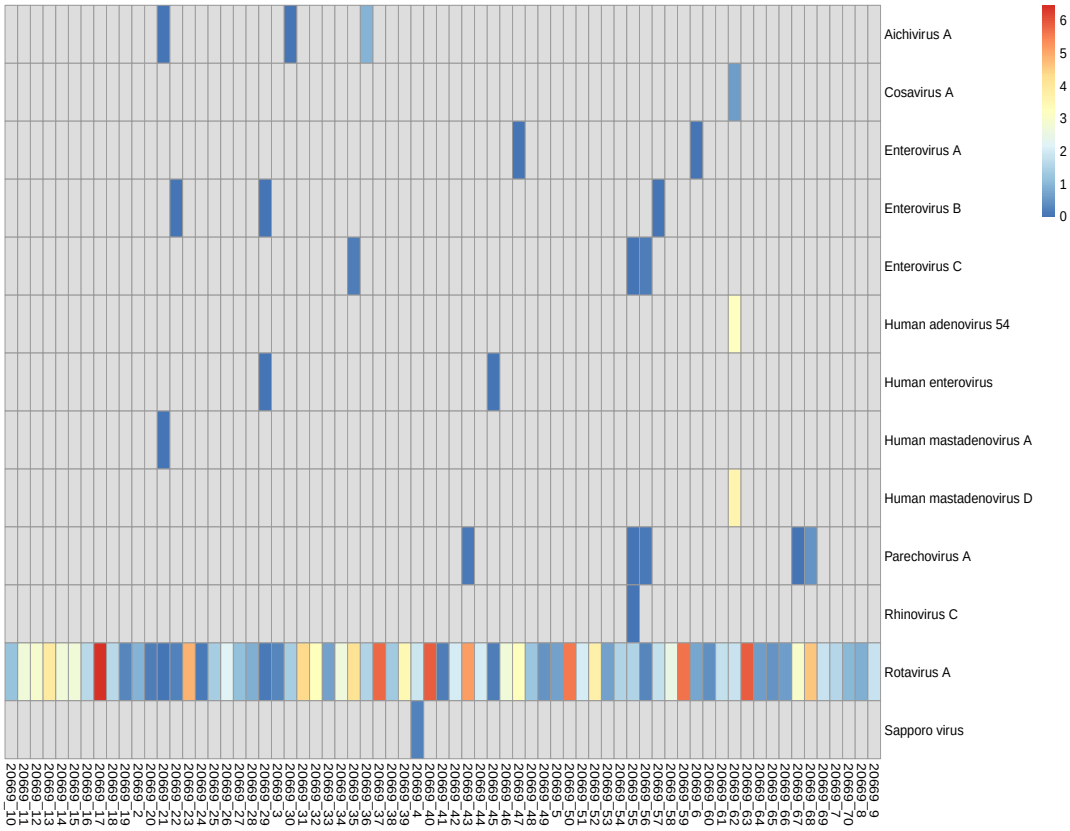


Figure 0.12: Taxonomic distribution of fecal viral contigs as classified by Kraken.

We highlighted contigs from RVA coinfections that had a minimum genome coverage of 10% and checked their average base sequencing depth, TAC or real-time PCR Ct values. A genome coverage of > 90% was observed from sequences that had a TAC Ct value of < 25 suggesting that a high genome coverage is achieved in samples with high viral load. One sample (20669_45) had a 100% enterovirus genome coverage but it was not detected by TAC.

Table 0.4: Data on classified viral contigs

Sample	Virus	TAC Ct Value	Real-time PCR Ct Value	Contig Length ±	Average depth §	Genome coverage	GenBank Accession	Reference Length
20669_4	Sapovirus GII	18.66		7348	428	98.5%	MG012406.1	7463
20669_6	Enterovirus A		NA	4600	12	62.1%	AB501332.1	7411
20669_21	HADV A		33	8653	16	25.6%	JF964962.1	33776
20669_21	HADV A		33	25190	18	74.6%	JF964962.1	33776
20669_22	Enterovirus B		NA	1706	7	23.2%	KC568447	7358
20669_29	Enterovirus B	29.72	NA	1776	124	24.0%	MH005795.1	7400
20669_29	Enterovirus B	29.72	NA	3492	21	46.8%	MH614922.1	7464
20669_43	Parechovirus A	32.63	NA	3537	17	48.5%	GQ183029.1	7294
20669_45	Enterovirus		NA	7444	3842	100%	MH933859.1	7113
20669_55	Parechovirus A	24.06	NA	7397	206	100%	KT879918.1	7077
20669_56	Enterovirus C		NA	3552	217	47.6%	DQ995644.1	7458
20669_56	Enterovirus C		NA	4095	29	54.9%	DQ995644.1	7458
20669_57	Enterovirus B	29.56	NA	743	7	10.3%	HM852755.1	7213
20669_67	Parechovirus A	31.18	NA	1708	14	23.4%	GQ183034.1	7302

§ Calculated by dividing the per-position coverage output by respective contig length

± Contig refers to contiguous length of genomic sequence in which the order of bases is known to a high confidence level.

Ct value: Cycle threshold value

NA: Not Applicable

All samples that were detected as RVA positive by TAC and real-time PCR had a contig classified as RVA by Kraken. Two contigs classified as adenovirus A were only detected by real-time PCR. No contigs were classified as astrovirus or norovirus genogroup II, but the viruses were detected by either TAC or real-time PCR. Only one sample that was sapovirus positive by TAC (Ct value = 18.7) had two contigs classified as sapovirus (384 basepairs and 7348 basepairs). Contigs

classified as enterovirus (n = 5) and parechovirus (n = 2) were detected in samples with a TAC Ct value of ≤ 30 (Figure 4.13).

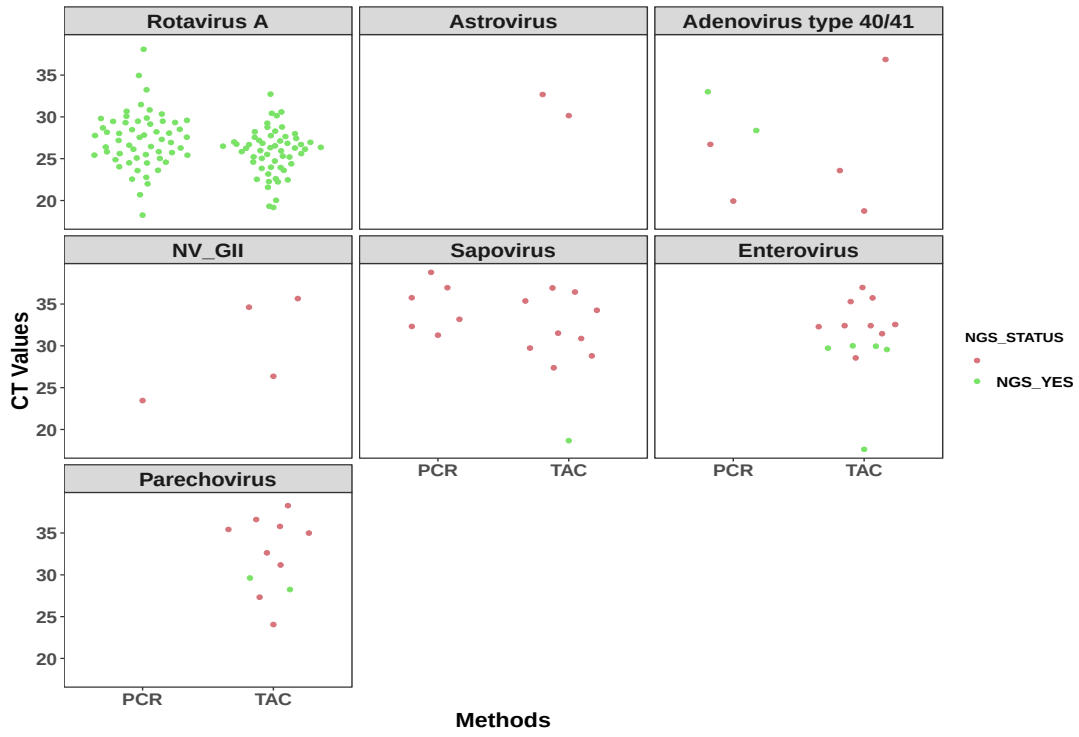


Figure 0.13: Distribution of TAC and real-time PCR Ct Values. Green dots show Ct values of samples also detected by NGS while maroon dots are vice versa

4.4.2.2 Coinfections from Taxonomic Classification of Unassembled Reads by Kraken

Out of 69 samples that had RVA reads, 32 samples had enteric viral reads associated with diarrhoea as coinfections. These viruses included; aichivirus A (n = 10), aichivirus C (n = 1), human cosavirus (n=1), cosavirus A (n = 4), human enterovirus (n = 4), enterovirus A (n = 5), enterovirus B (n = 6), enterovirus C (n = 8), enterovirus D (n = 6), enterovirus G (n = 1), enterovirus H (n = 4), enterovirus J (n = 1), HAdV A (n = 2), HAdV B (n = 1), HAdV D (n = 1), HAdV E (n = 1), HAdV F

(n = 2), parechovirus (n = 8), rhinovirus A (n = 1), rhinovirus C (n = 2) and sapovirus (n = 4) (Figure 4.14).

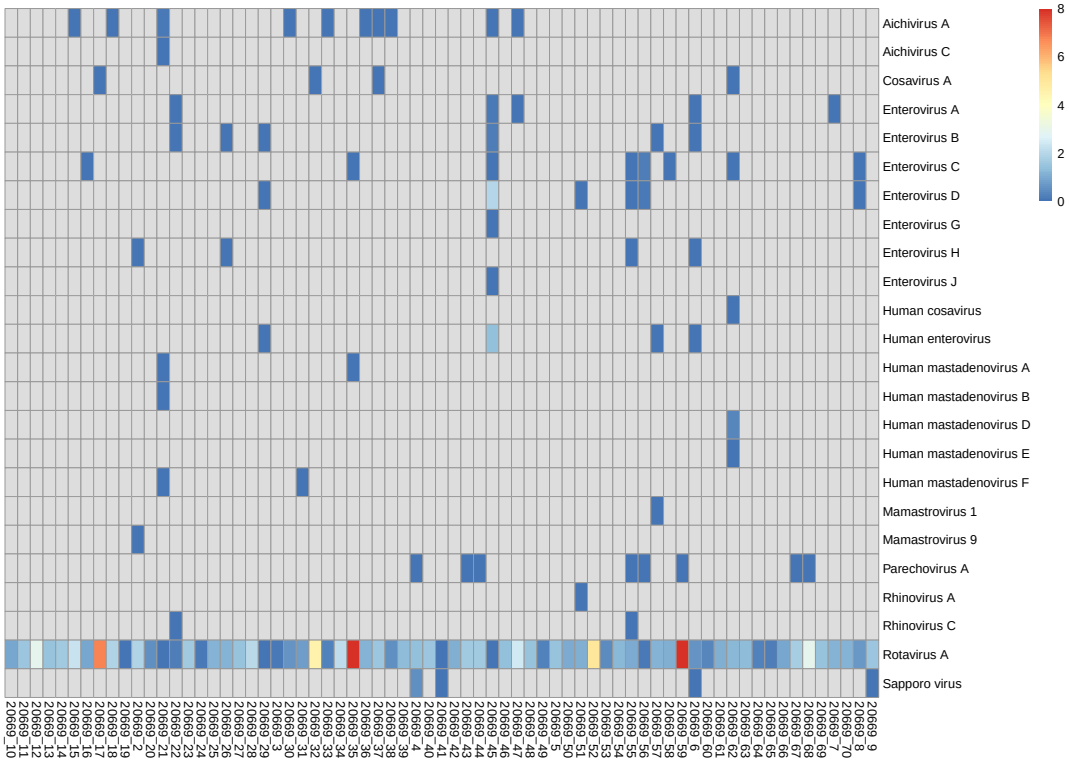


Figure 0.14: Taxonomic distribution of fecal viral reads as classified by Kraken.

Cleaned unassembled reads were assigned taxonomic classifiers by Kraken.

4.4.2.3 Coinfections from Taxonomic Classification of Unassembled Reads by MetaPhlan

Out of 53 samples that had RVA reads, 21 samples had enteric viral reads associated with diarrhoea as coinfections. These viruses included; astrovirus (n = 1), HAdV B (n = 2), HAdV D (n = 3), sapovirus (n = 1), enterovirus A (n = 1), enterovirus C (n = 5), enterovirus E (n = 1), enterovirus G (n = 1), rhinovirus A (n = 1), parechovirus (n = 7), cosavirus A (n = 2) and salivirus A (n = 1) (Figure 4.15).

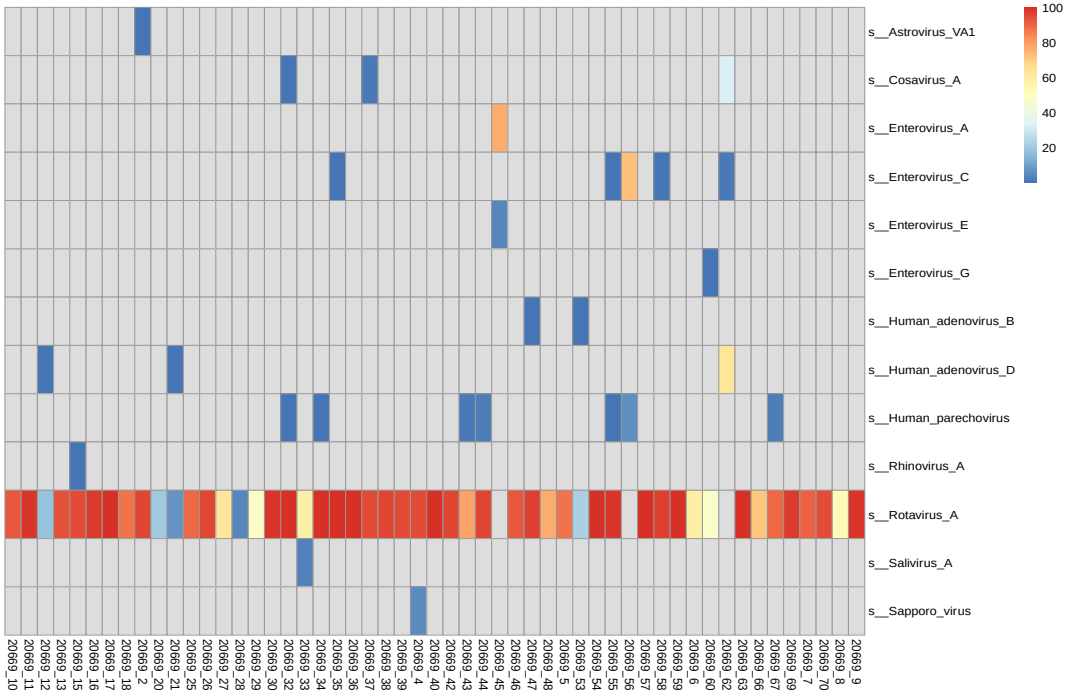


Figure 0.15: Taxonomic distribution of fecal viral reads as classified by MetaPhlan

4.4.3 Comparison of Kraken, MetaPhlan, Real-Time PCR and TAC

Comparison of the two bioinformatics approaches in analysing the NGS data (Kraken and MetaPhlan), real-time PCR and TAC was done for 61 samples that were investigated using all the three methods for RVA, HAdV type F, astrovirus, norovirus genogroup II and sapovirus. For parechovirus and enterovirus, NGS was only compared with TAC for 63 samples. The detection rates of enteric viruses included in the real-time PCR and TAC methods were compared with those of the three bioinformatics approaches. Taxonomic classification of the reads and contigs by Kraken identified RVA in all the samples unlike Metaphlan which identified RVA in 44/61 (72.1%) of the samples. Norovirus genogroup II detected by the real-time PCR and TAC methods, 1/61 (1.6%) and 2/61 (3.3%) respectively, was not detected by NGS. Taxonomic classification of reads by Kraken detected higher proportions of astrovirus and enterovirus than the other two bioinformatics approaches and PCR-based methods (Figure 4.16). For parechovirus and sapovirus,

real-time PCR and TAC detected higher proportions than NGS for all the bioinformatics approaches (Figure 4.16).

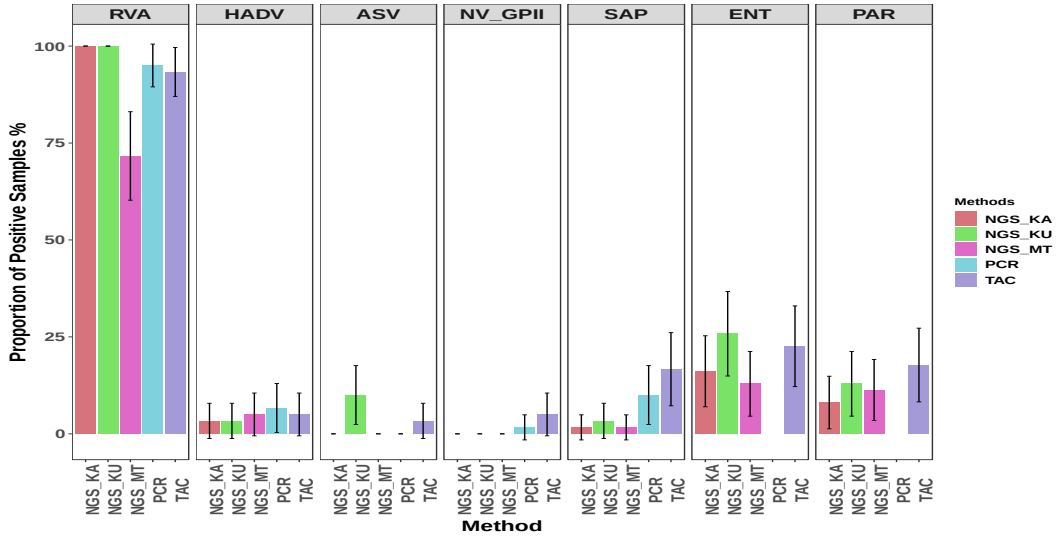


Figure 0.16: Proportion of positive samples by NGS, real-time PCR (PCR) and TaqMan Array Card (TAC).

The error bars represent 95% confidence interval. NGS: Next Generation Sequencing; KA: Kraken assembled; KU: kraken unassembled; MT: MetaPhlan; PCR: real-time PCR; TAC: TaqMan Array Card; RVA: Rotavirus A; HADV: Human adenovirus; ASV: astrovirus; NV_GPII: Norovirus genogroup II; SAP: sapovirus; ENT: enterovirus; PAR: parechovirus

CHAPTER FIVE

DISCUSSION

The current work provides a comparison of RVA-ELISA, real-time PCR, TAC and NGS in the detection of five common enteric viruses associated with diarrhoea. In the detection of RVA, the study reports a significant difference in the sensitivity of our ELISA and PCR-based methods (real-time PCR and TAC) in the the year 2016. The difference in sensitivity was not observed when analysing samples from the pre-vaccine period (2013). It is unsurprising that the sensitivity of ELISA was lower than the PCR-based methods (Gautam *et al.*, 2013), however, it was intruiging that the difference was only significant in 2016. Notably, samples that were negative for RVA by ELISA and positive by real-time PCR and TAC had Ct values (<30) associated with clinical disease (Figure 4.17). These samples were retested using conventional PCR followed by di-deoxy sanger sequencing and 21/26 (80.8%), 3/5 in 2013 and 18/21 in 2016, were positive by the confirmatory assay. There was no significant difference in the distribution of real-time PCR or TAC Ct values in samples that were either ELISA positive or negative post vaccine introduction unlike what has been previously reported in the U.S (Tate *et al*., 2013). This suggest that the discrepancy in this study was not due to RVA viral load as previously reported but may be due to ELISA diagnostic errors.

Despite the significant decrease in the sensitivity of the ELISA (ProsPecT^{TM}, Oxoid ltd., Basingstoke, UK) method, this study reports no significant drop in the specificity (p-value > 0.05). A previous study reported a decrease in ELISA's specificity (Rotascreen II, Microgen, UK) in a vaccinated community in Australia (McAuliffe *et al*., 2019). The study also reports a decrease in the overall agreement of ELISA, real-time PCR and TAC in detection of RVA post-vaccine introduction. The inter-rater agreement by Fleiss' kappa was 0.93 and 0.74 in 2013 and 2016, respectively. The kappa statistic shows how much agreement and reliability the three methods have and a drop in the agreement inversely relates to a drop-in reliability of the results obtained.

The rotavirus vaccine in Kenya has been reported to be effective in reducing RVA associated diarrhoea hospitalizations based on results from commercial ELISA kits (Khagayi *et al.,* 2019). In this study, we report a reduction in rotavirus prevalence estimates as detected by all the three methods in 2016 compared to 2013. However, the RVA reduction rates in diarrhoea hospitalized cases was not significant by TAC unlike ELISA and real-time PCR. This suggests that the diagnostic method used during surveillance is key in explaining the reduction in hospitalized diarrhoea cases positive for RVA, vaccine effectiveness and estimating the disease burden. The reduction in the RVA cases was associated with a shift in median age of diarrhoea cases from 11 months pre-vaccine introduction to 14 months post-vaccine introduction (Wilcoxon: p-value < 0.05). This was also reported by a study in Swaziland looking at the impact of rotavirus vaccine in children under five years and it showed a shift of the median age from 10 months to 13.7 months post-rotavirus vaccine introduction (Maphalala *et al*., 2018).

The study reports no significant difference in the viral proportions detected by real-time PCR and TAC except for adenovirus and astrovirus. The primers and probes used by the two methods in astrovirus detection targeted two different genomic regions and this may partly explain the differences between two assays. In adenovirus detection, the differences observed between the proportions by the two methods may have resulted as a result of the mismatches in the real-time PCR primer and probe binding sites that are absent in the TAC assay.

The performance of qPCR assays has been reported to be impacted significantly by mismatches within the last five bases at a primer's 3' end and the mismatch being either a purine/purine or pyrimidine/pyrimidine (Lefever *et al.,* 2013). The mismatches observed in the primer and probe binding sites of adenovirus, astrovirus and sapovirus may impair the real-time PCR or TAC function by blocking the amplification or increasing the quantification cycles. The magnitude of the nucleotide mismatches on the PCR function could have been shown better using local sequences. However, no sequences from Kenya were found to be deposited in GenBank for astrovirus, norovirus genogroup II and adenovirus type 40/41.

Mismatches in the primer and probe binding sites due to the evolution of virus might cause missed diagnoses and underestimation of the viral diarrhoeal disease burden.

This study reports successful detection of viral coinfections occurring with RVA using viral metagenomics. Rotavirus has been reported to mostly occur as the only aetiology in diarrhoeal cases but it can also occur as the primary or secondary aetiology in mixed infections too (Liu *et al*., 2016; Platts-Mills *et al*., 2018). Detection of the coinfections varied by the bioinformatics workflows i.e. classification of unassembled reads using Kraken detected an enteric virus coinfection in 32 samples compared to Metaphlan which detected a coinifection in 21 samples and classification of contigs using Kraken detected a coinfection in 16 samples. Different workflows have been reported to differ in their sensitivity and specificity when it comes to viral taxonomic classification (Nooij *et al*., 2018). Therefore, it is important to document the parameters used in different bioinformatics softwares clearly to enable reproducibility and explain any differences observed. The common viruses detected as coinfections were either unexpected or part of the panel detected by available diagnostic methods i.e. TAC or real-time PCR panel. Other studies that looked at the diversity of enteric viruses in fecal samples using viral metagenomics also detected these viruses (Mohammad et al., 2020; Moore et al., 2015; Yinda et al., 2019). Succesfull assembly of partial or full genomes for cosavirus, rhinovirus C, salivirus, sapovirus, enterovirus, human adenovirus A and parechovirus was achieved. Genome assembly was inversely proportional with the real-time and TAC Ct values (measure of viral quantity) with more success in samples with a Ct value of <30. These sequences were beneficial in characterizing these viruses associated with diarrhoea and improving the primer/probe sequences used in real-time PCR and TAC methods.

This study had limitations. First, two different extraction protocols were used to obtain nucleic acids used by real-time PCR and TAC methods in 2016. However, extraction of viral nucleic acids of enteropathogens using either QIAamp Stool DNA Mini kit or QIAamp Viral RNA mini kit (Qiagen) has been reported to give comparable viral RNA yield for all common enteric pathogens except for Norovirus GII for which has a higher Ct value with RNA extraction alone (difference within 1

Ct value) (Liu *et al*., 2016) Therefore, the viral concentrations obtained by the two extraction methods may have affected the starting material. Second, the primer targets for TAC and real-time PCR differed for astrovirus and this may have influenced the detection frequency. Third, the samples were subjected to freeze thawing during processing and this may have affected the integrity of the nucleic acids in the samples. Finally, the study only included a single year in both the pre and post-vaccine introduction periods in Kenya and this may have underestimated or overestimated the effect of RVA occurrence due to its fluctuation from year to year.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This study presents the need for re-evaluation of diagnostic kits used in RVA detection in the pre- and post-vaccine introduction periods. The observed differences in the methods diagnostic performance may affect estimation of viral diarrhoeal disease burden and rotavirus vaccine effectiveness. Despite the fact that both real-time PCR and TAC methods are PCR-based assays, TAC generally gives higher prevalence estimates than real-time PCR. The mismatches observed in the primer and probe binding sites, different platforms and different targets for the assays in astrovirus partly explain the differences observed in the two methods. The results from primer/probe mismatch analysis provide insight on how the current primers and probes can be improved to minimize the mismatches that might impair PCR function. The study also shows that agnostic NGS can be used to identify viral coinfections with RVA including those in our real-time PCR or TAC panels. However, the range of coinfections is dependent on the bioinformatics workflow used.

6.2 Recommendations

- The study only compared the performance of diagnostic methods currently available at the KWTRP. Different ELISA kits have been shown to perform differently in RVA detection. Therefore, it would be interesting to know whether the reduction in the sensitivity of our ELISA kit (ProspecT^{TM,} Oxoid Basingstoke, UK) is also present in other ELISA kits in the post-vaccine introduction era. Further, other labs doing surveillance on viral diarrhoeal aetiologies consider re-evaluation of their EIA kits, especially Prospect^{TM} rotavirus kit (Oxoid, UK), post-rotavirus vaccine introduction.
- We also recommend application of agnostic viral NGS to improve detection of viral coinfections associated with diarrhoea. Agnostic viral NGS will not only improve detection of coinfections but it will also help increase the

number of local available sequences for the detected viruses. These sequences will be key in updating the primers and probes used in the available molecular diagnostic methods for local sequence variation and maintain their sensitivity.

- The primers and probes used in the the PCR methods detect different serotypes of adenovirus, sapovirus and astrovirus and it would be interesting to type the different serotypes in the positive samples.

REFERENCES

Ambert-Balay, K., and Pothier, P. (2013). Evaluation of 4 immunochromatographic tests for rapid detection of norovirus in faecal samples. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2012.11.001

Andrews, S. (2010). FastQC. *Babraham Bioinformatics*. https://doi.org/citeulike-article-id:11583827

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., and Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. https://doi.org/10.1089/cmb.2012.0021

Bányai, K., Estes, M. K., Martella, V., and Parashar, U. D. (2018). Viral gastroenteritis. *The Lancet*. https://doi.org/10.1016/S0140-6736(18)31128-0

Bennett, A., Bar-Zeev, N., Jere, K. C., Tate, J. E., Parashar, U. D., Nakagomi, O., and Cunliffe, N. A. (2015). Determination of a viral load threshold to distinguish symptomatic versus asymptomatic rotavirus infection in a high-disease-Burden African population. *Journal of Clinical Microbiology*, *53*(6), 1951–1954. https://doi.org/10.1128/JCM.00875-15

Bennett, S., and Gunson, R. N. (2017). The development of a multiplex real-time RT-PCR for the detection of adenovirus, astrovirus, rotavirus and sapovirus from stool samples. *Journal of Virological Methods*, *242*, 30–34. https://doi.org/10.1016/j.jviromet.2016.12.016

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-Van Dillen, P. M. E., and Van Der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*. https://doi.org/10.1556/AMicr.58.2011.1.7

Brooks, R. G., Brown, L., and Franklin, R. A. Y. B. (1989). *(TestPack Rotavirus) with Standard Enzyme Immunoassay*. *27*(4), 775–777.

Chhabra, P., Gregoricus, N., Weinberg, G. A., Halasa, N., Chappell, J., Hassan, F., and Vinjé, J. (2017). Comparison of three multiplex gastrointestinal platforms for the detection of gastroenteritis viruses. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2017.08.012

Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., and Van Der Hoek, L. (2014). Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0093269

Cubitt, W. D., Mitchell, D. K., Carter, M. J., Willcocks, M. M., and Holzel, H. (1999). Application of electronmicroscopy, enzyme immunoassay, and RT-PCR to monitor an outbreak of astrovirus type 1 in a paediatric bone marrow transplant unit. *Journal of Medical Virology*, *57*(3), 313–321. https://doi.org/10.1002/(SICI)1096-9071(199903)57:3<313::AID-JMV16>3.0.CO;2-A

Dennehy, P. H., Gauntlett, D. R., and Spangenberger, S. E. (1990). Choice of reference assay for the detection of rotavirus in fecal specimens: Electron microscopy versus enzyme immunoassay. *Journal of Clinical Microbiology*.

Dhingra, A., Hage, E., Ganzenmueller, T., Böttcher, S., Hofmann, J., Hamprecht, K., and Heim, A. (2019). Molecular Evolution of Human Adenovirus (HAdV) Species C. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-018-37249-4

Donck, I. Van der, Lemmens, A., Struyf, F., and Ranst, M. Van. (1999). *COMPARISON OF THREE RAPID IMMUNOASSAYS FOR THE DETECTION OF ROTAVIRUS ANTIGEN IN STOOL SAMPLES*. 1999.

Dunbar, N. L., Bruggink, L. D., and Marshall, J. A. (2014). Evaluation of the

RIDAGENE real-time PCR assay for the detection of GI and GII norovirus. *Diagnostic Microbiology and Infectious Disease*. https://doi.org/10.1016/j.diagmicrobio.2014.03.017

Duong, V. T., Phat, V. V., Tuyen, H. T., Dung, T. T. N., Trung, P. D., Minh, P. Van, and Baker, S. (2016). Evaluation of Luminex xTAG Gastrointestinal Pathogen Panel Assay for Detection of Multiple Diarrheal Pathogens in Fecal Samples in Vietnam. *Journal of Clinical Microbiology*, *54*(4), 1094–1100. https://doi.org/10.1128/jcm.03321-15

Feeney, S. A., Armstrong, V. J., Mitchell, S. J., Crawford, L., Mccaughey, C., and Coyle, P. V. (2011). Development and clinical validation of multiplex TaqMan® assays for rapid diagnosis of viral gastroenteritis. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.22162

Gautam, R., Lyde, F., Esona, M. D., Quaye, O., and Bowena, M. D. (2013). Comparison of Premier™ Rotaclone®, ProSpecT™, and RIDASCREEN®rotavirus enzyme immunoassay kits for detection of rotavirus antigen in stool specimens. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2013.06.022

Gautam, R., Mijatovic-Rustempasic, S., Esona, M. D., Tam, K. I., Quaye, O., and Bowen, M. D. (2016). One-step multiplex real-time RT-PCR assay for detecting and genotyping wild-type group A rotavirus strains and vaccine strains (Rotarix® and RotaTeq®) in stool samples. *PeerJ*, *4*, e1560. https://doi.org/10.7717/peerj.1560

Global Burden of Disease Collaborative Network. (2017). Global Burden of Disease Study 2017. *The Lancet*.

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annual Review of Pathology: Mechanisms of Disease*, *14*(1), 319–338. https://doi.org/10.1146/annurev-pathmechdis-012418-012751

Guix, S., Bosch, A., and Pintó, P. M. (2005). Human astrovirus diagnosis and typing: Current and future prospects. *Letters in Applied Microbiology*. https://doi.org/10.1111/j.1472-765X.2005.01759.x

Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics*, *48*, 193–204. https://doi.org/10.1016/j.jbi.2014.02.013

Hansman, G. S., Guntapong, R., Pongsuwanna, Y., Natori, K., Katayama, K., and Takeda, N. (2006). Development of an antigen ELISA to detect sapovirus in clinical stool specimens. *Archives of Virology*. https://doi.org/10.1007/s00705-005-0630-x

Hansman, Grant S., Ishida, S., Yoshizumi, S., Miyoshi, M., Ikeda, T., Oka, T., and Takeda, N. (2007). Recombinant sapovirus gastroenteritis, Japan [6]. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid1305.070049

Higgins, R. R., Beniprashad, M., Cardona, M., Masney, S., Low, D. E., and Gubbay, J. B. (2011). Evaluation and verification of the seeplex diarrhea-V ACE assay for simultaneous detection of adenovirus, rotavirus, and norovirus genogroups I and II in clinical stool specimens. *Journal of Clinical Microbiology*, *49*(9), 3154–3162. https://doi.org/10.1128/JCM.00599-11

Hoppe, E., Pauly, M., Gillespie, T. R., Akoua-Koffi, C., Hohmann, G., Fruth, B., and Calvignac-Spencer, S. (2015). Multiple cross-species transmission events of human adenoviruses (HAdV) during hominine evolution. *Molecular Biology and Evolution*, *32*(8), 2072–2084. https://doi.org/10.1093/molbev/msv090

Hyun, J., Ko, D. H., Lee, S. K., Kim, H. S., Kim, J. S., Song, W., and 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*(3), 220–225. https://doi.org/10.3343/alm.2018.38.3.220

Jiang, Y., Fang, L., Shi, X., Zhang, H., Li, Y., Lin, Y., and Hu, Q. (2014). Simultaneous detection of five enteric viruses associated with gastroenteritis by

use of a PCR assay: A single real-time multiplex reaction and its clinical application. *Journal of Clinical Microbiology*, *52*(4), 1266–1268. https://doi.org/10.1128/JCM.00245-14

Joshi, M. S., Deore, S. G., Walimbe, A. M., Ranshing, S. S., and Chitambar, S. D. (2019). Evaluation of different genomic regions of Rotavirus A for development of real time PCR. *Journal of Virological Methods*, *266*, 65–71. https://doi.org/10.1016/j.jviromet.2019.01.017

Jothikumar, N., Cromeans, T. L., Hill, V. R., Lu, X., Sobsey, M. D., and Erdman, D. D. (2005). Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.71.6.3131-3136.2005

Kageyama, T., Kojima, S., Shinohara, M., Uchida, K., Fukushi, S., Hoshino, F. B., and 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*. https://doi.org/10.1128/JCM.41.4.1548-1557.2003

Kamau, A., Nyaga, V., Bauni, E., Tsofa, B., Noor, A. M., Bejon, P., and Hammitt, L. L. (2017). Trends in bednet ownership and usage, and the effect of bednets on malaria hospitalization in the Kilifi Health and Demographic Surveillance System (KHDSS): 2008-2015. *BMC Infectious Diseases*, *17*(1), 1–9. https://doi.org/10.1186/s12879-017-2822-x

Khagayi, S., Omore, R., Otieno, G. P., Ogwel, B., Ochieng, J. B., Juma, J., and Verani, J. R. (2019). *Effectiveness of monovalent rotavirus vaccine against hospitalization with acute rotavirus gastroenteritis in Kenyan children Authors*.

Kroneman, A., Vega, E., Vennema, H., Vinjé, J., White, P. A., Hansman, G., and Koopmans, M. (2013). Proposal for a unified norovirus nomenclature and genotyping. *Archives of Virology*. https://doi.org/10.1007/s00705-013-1708-5

La Rosa, G., Della Libera, S., Petricca, S., Iaconelli, M., Donia, D., Saccucci, P., and Divizia, M. (2015). Genetic Diversity of Human Adenovirus in Children with Acute Gastroenteritis, Albania, 2013–2015. *BioMed Research International*, *2015*, 1–7. https://doi.org/10.1155/2015/142912

Lambisia, A. W., Onchaga, S., Murunga, N., Lewa, C. S., Nyanjom, S. G., and Agoti, C. N. (2020). Epidemiological trends of five common diarrhea-associated enteric viruses pre-and post-rotavirus vaccine introduction in coastal Kenya. *Pathogens*, *9*(8), 1–14. https://doi.org/10.3390/pathogens9080660

Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*. https://doi.org/10.2307/2529310

Langmead. (2013). Bowtie2. *Nature Methods*. https://doi.org/10.1038/nmeth.1923.Fast

Lefever, S., Pattyn, F., Hellemans, J., and Vandesompele, J. (2013). Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clinical Chemistry*, *59*(10), 1470–1480. https://doi.org/10.1373/clinchem.2013.203653

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp352

Liu, J., Gratz, J., Amour, C., Kibiki, G., Becker, S., Janaki, L., and Taniuchi, M. (2013). *A Laboratory-Developed TaqMan Array Card for Simultaneous Detection of 19 Enteropathogens*. *51*(2), 472–480. https://doi.org/10.1128/JCM.02658-12

Liu, J., Gratz, J., Amour, C., Nshama, R., Walongo, T., Maro, A., and Houpt, E. R. (2016). Optimization of quantitative PCR methods for enteropathogen detection. *PLoS ONE*, *11*(6), 1–11. https://doi.org/10.1371/journal.pone.0158199

Liu, J., Kibiki, G., Maro, V., Maro, A., Kumburu, H., Swai, N., and Houpt, E. (2011). Multiplex reverse transcription PCR Luminex assay for detection and quantitation of viral agents of gastroenteritis. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2010.12.009

Liu, J., Platts-Mills, J. A., Juma, J., Kabir, F., Nkeze, J., Okoi, C., and Houpt, E. R. (2016). Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *The Lancet*. https://doi.org/10.1016/S0140-6736(16)31529-X

Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., and Black, R. E. (2016). 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*. https://doi.org/10.1016/S0140-6736(16)31593-8

Logan, C., O'Leary, J. J., and O and apos;Sullivan, N. (2007). Real-time reverse transcription PCR detection of norovirus, sapovirus and astrovirus as causative agents of acute viral gastroenteritis. *Journal of Virological Methods*, *146*(1–2), 36–44. https://doi.org/10.1016/j.jviromet.2007.05.031

Lopez-Lacort, M., Collado, S., Díez-Gandía, A., and Díez-Domingo, J. (2016). Rotavirus, vaccine failure or diagnostic error? *Vaccine*. https://doi.org/10.1016/j.vaccine.2016.10.032

Madeley, C. R., and Cosgrove, B. P. (1976). CALICIVIRUSES IN MAN. *The Lancet*. https://doi.org/10.1016/S0140-6736(76)91309-X

Maphalala, G., Phungwayo, N., Masona, G., Lukhele, N., Tsegaye, G., Dube, N., and Weldegebriel, G. (2018). Early impact of rotavirus vaccine in under 5 year old children hospitalized due to diarrhea, Swaziland. *Vaccine*. https://doi.org/10.1016/j.vaccine.2017.07.072

Martella, V., Bányai, K., Matthijnssens, J., Buonavoglia, C., and Ciarlet, M. (2010). Zoonotic aspects of rotaviruses. *Veterinary Microbiology*, *140*(3–4), 246–255. https://doi.org/10.1016/j.vetmic.2009.08.028

McAuliffe, G. N., Taylor, S. L., Moore, S., Hewitt, J., Upton, A., Howe, A. S., and Best, E. J. (2019). Suboptimal performance of rotavirus testing in a vaccinated community population should prompt laboratories to review their rotavirus testing algorithms in response to changes in disease prevalence. *Diagnostic Microbiology and Infectious Disease*. https://doi.org/10.1016/j.diagmicrobio.2018.10.003

Mijatovic-Rustempasic, S., Tam, K. I., Kerin, T. K., Lewis, J. M., Gautam, R., Quaye, O., and Bowen, M. D. (2013). Sensitive and specific quantitative detection of rotavirus a by one-step real-time reverse transcription-PCR assay without antecedent double-stranded-RNA denaturation. *Journal of Clinical Microbiology*, *51*(9), 3047–3054. https://doi.org/10.1128/JCM.01192-13

Mitui, M. T., Chandrasena, T. N., Chan, P. K., Rajindrajith, S., Nelson, E. A. S., Leung, T. F., and Ahmed, K. (2012). Inaccurate identification of rotavirus genotype G9 as genotype G3 strains due to primer mismatch. *Virology Journal*, *9*(1), 1. https://doi.org/10.1186/1743-422X-9-144

Mohammad, H. A., Madi, N. M., and Al-Nakib, W. (2020). Analysis of viral diversity in stool samples from infants and children with acute gastroenteritis in Kuwait using Metagenomics approach. *Virology Journal*. https://doi.org/10.1186/s12985-020-1287-5

Moore, N. E., Wang, J., Hewitt, J., Croucher, D., Williamson, D. A., Paine, S., and Hall, R. J. (2015). Metagenomic analysis of viruses in feces from unsolved outbreaks of gastroenteritis in humans. *Journal of Clinical Microbiology*. https://doi.org/10.1128/JCM.02029-14

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

Navidad, J. F., Griswold, D. J., Gradus, M. S., and Bhattacharyya, S. (2013).

Evaluation of luminex xTAG gastrointestinal pathogen analyte-specific reagents for high-throughput, simultaneous detection of bacteria, viruses, and parasites of clinical and public health importance. *Journal of Clinical Microbiology*, *51*(9), 3018–3024. https://doi.org/10.1128/JCM.00896-13

Noel, J. S., Lee, T. W., Kurtz, J. B., Glass, R. I., and Monroe, S. S. (1995). Typing of human astroviruses from clinical isolates by enzyme immunoassay and nucleotide sequencing. *Journal of Clinical Microbiology*.

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M. P. G. (2018). Overview of virus metagenomic classification methods and their biological applications. *Frontiers in Microbiology*, *9*(APR). https://doi.org/10.3389/fmicb.2018.00749

Oka, T., Iritani, N., Yamamoto, S. P., Mori, K., Ogawa, T., Tatsumi, C., and Wu, F. T. (2019). Broadly reactive real-time reverse transcription-polymerase chain reaction assay for the detection of human sapovirus genotypes. *Journal of Medical Virology*, *91*(3), 370–377. https://doi.org/10.1002/jmv.25334

Oka, T., Katayama, K., Hansman, G. S., Kageyama, T., Ogawa, S., Wu, F. T., and Takeda, N. (2006). Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.20699

Oka, T., Wang, Q., Katayama, K., and Saifb, L. J. (2015). Comprehensive review of human sapoviruses. *Clinical Microbiology Reviews*, *28*(1), 32–53. https://doi.org/10.1128/CMR.00011-14

Otieno, G. P., Bottomley, C., Khagayi, S., Adetifa, I., Ngama, M., Omore, R., and Nokes, D. J. (2019). Impact of the Introduction of Rotavirus Vaccine on Hospital Admissions for Diarrhea Among Children in Kenya: A Controlled Interrupted Time-Series Analysis. *Clinical Infectious Diseases*, (Cdc), 1–8. https://doi.org/10.1093/cid/ciz912

Pang, X. L., Lee, B., Boroumand, N., Leblanc, B., Preiksaitis, J. K., and Ip, C. C. Y.

(2004). Increased Detection of Rotavirus Using a Real Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Assay in Stool Specimens from Children with Diarrhea. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.20009

Parra, G. I., and Espinola, E. E. (2006). Nucleotide mismatches between the VP7 gene and the primer are associated with genotyping failure of a specific lineage from G1 rotavirus strains. *Virology Journal*. https://doi.org/10.1186/1743-422X-3-35

Phan, M. V. T., Kellam, P., Cuong, N. Van, Rabaa, M. A., My, P. T., Cotten, M., and Tri, T. N. (2016). Unbiased whole-genome deep sequencing of human and porcine stool samples reveals circulation of multiple groups of rotaviruses and a putative zoonotic infection. *Virus Evolution*, *2*(2), vew027. https://doi.org/10.1093/ve/vew027

Platts-Mills, J. A., Liu, J., Rogawski, E. T., Kabir, F., Lertsethtakarn, P., Siguas, M., and Houpt, E. 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*(12), e1309–e1318. https://doi.org/10.1016/S2214-109X(18)30349-8

Preeti Chhabra, P., Gregoricus, N., Weinberg, G. A., Halasa, N., Chappell, J., Hassan, F., and Vinjé, J.. (2018). Comparison of three multiplex gastrointestinal platforms for the detection of gastroenteritis viruses. *Journal of Clinical Virology*, *95*, 5. https://doi.org/10.1016/j.jcv.2017.08.012.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq033

R Core Team, R. C. (2018). R: A Language and Environment for Statistical Computing.

Robilotti, E., Deresinski, S., and Pinsky, B. A. (2015). Norovirus. *Clinical Microbiology Reviews*, *28*(1), 134–164. https://doi.org/10.1128/CMR.00075-14

Rooney, B. L., Pettipas, J., Grudeski, E., Mykytczuk, O., Pang, X. L., Booth, T. F., and Leblanc, J. J. (2014). Detection of circulating norovirus genotypes: Hitting a moving target. *Virology Journal*, *11*(1), 20–25. https://doi.org/10.1186/1743-422X-11-129

Royuela, E., Negredo, A., and Sánchez-Fauquier, A. (2006). Development of a one step real-time RT-PCR method for sensitive detection of human astrovirus. *Journal of Virological Methods*, *133*(1), 14–19. https://doi.org/10.1016/j.jviromet.2005.10.012

Rubenstein, A. S., and Miller, M. F. (1982). Comparison of an enzyme immunoassay with electron microscopic procedures for detecting rotavirus. *Journal of Clinical Microbiology*.

Sadiq, A., Bostan, N., Yinda, K. C., Naseem, S., and Sattar, S. (2018). Rotavirus: Genetics, pathogenesis and vaccine advances. *Reviews in Medical Virology*, *28*(6), 1–13. https://doi.org/10.1002/rmv.2003

Sarantis, H., Johnson, G., Brown, M., Petric, M., and Tellier, R. (2004). Comprehensive detection and serotyping of human adenoviruses by PCR and sequencing. *Journal of Clinical Microbiology*, *42*(9), 3963–3969. https://doi.org/10.1128/JCM.42.9.3963-3969.2004

Scott, J. A. G., Bauni, E., Moisi, J. C., Ojal, J., Gatakaa, H., Nyundo, C., and Williams, T. N. (2012). Profile: The Kilifi health and demographic surveillance system (KHDSS). *International Journal of Epidemiology*, *41*(3), 650–657. https://doi.org/10.1093/ije/dys062

Steele, A. D., Williams, M. M., Bos, P., and Peenze, I. (1994). Comparison of two rapid enzyme immunoassays with standard enzyme immunoassay and latex agglutination for the detection of human rotavirus in stools. *J Diarrhoeal Dis Res*.

Tate, J. E., Mijatovic-Rustempasic, S., Tam, K. I., Lyde, F. C., Payne, D. C., Szilagyi, P., and Parashar, U. D. (2013). Comparison of 2 assays for diagnosing rotavirus and evaluating vaccine effectiveness in children with gastroenteritis. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid1908.130461

Thorne, L. G., and Goodfellow, I. G. (2014). Norovirus gene expression and replication. *Journal of General Virology*, *95*(PART 2), 278–291. https://doi.org/10.1099/vir.0.059634-0

Troeger, C., Blacker, B. F., Khalil, I. A., Rao, P. C., Cao, S., Zimsen, S. R., and Reiner, R. C. (2018). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Infectious Diseases*, *18*(11), 1211–1228. https://doi.org/10.1016/S1473-3099(18)30362-1

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*. https://doi.org/10.1038/nmeth.3589

van Maarseveen, N. M., Wessels, E., de Brouwer, C. S., Vossen, A. C. T. M., and Claas, E. C. J. (2010). Diagnosis of viral gastroenteritis by simultaneous detection of Adenovirus group F, Astrovirus, Rotavirus group A, Norovirus genogroups I and II, and Sapovirus in two internally controlled multiplex real-time PCR assays. *Journal of Clinical Virology*. https://doi.org/10.1016/j.jcv.2010.07.019

Vila, J. (2017). New molecular diagnostic tools in traveller's diarrhea. *Journal of Travel Medicine*, *24*(1), S23–S28. https://doi.org/10.1093/jtm/taw071

Vinjé, J. (2015). Advances in laboratory methods for detection and typing of norovirus. *Journal of Clinical Microbiology*. https://doi.org/10.1128/JCM.01535-14

Vu, D.-L., Bosch, A., Pintó, R., and Guix, S. (2017). Epidemiology of Classic and Novel Human Astrovirus: Gastroenteritis and Beyond. *Viruses*, *9*(2), 33.

https://doi.org/10.3390/v9020033

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

Wandera, E. A., Mohammad, S., Ouko, J. O., Yatitch, J., Taniguchi, K., and Ichinose, Y. (2017). Variation in rotavirus vaccine coverage by sub-counties in Kenya. *Tropical Medicine and Health*. https://doi.org/10.1186/s41182-017-0051-z

Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., and Kellam, P. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2012.0205

Wessels, E., Rusman, L. G., van Bussel, M. J. A. W. M., and Claas, E. C. J. (2014). Added value of multiplex Luminex gastrointestinal pathogen panel (xTAG® GPP) testing in the diagnosis of infectious gastroenteritis. *Clinical Microbiology and Infection*, *20*(3), O182–O187. https://doi.org/10.1111/1469-0691.12364

Whiley, D. M., and Sloots, T. P. (2005). Sequence variation in primer targets affects the accuracy of viral quantitative PCR. *Journal of Clinical Virology*, *34*(2), 104–107. https://doi.org/10.1016/j.jcv.2005.02.010

Wolfaardt, M., Kiulia, N. M., Mwenda, J. M., and Taylor, M. B. (2011). Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. *Journal of Clinical Microbiology*, *49*(2), 728–731. https://doi.org/10.1128/JCM.01093-10

Wood, D. E., and Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*.

https://doi.org/10.1186/gb-2014-15-3-r46

World Health organization - WHO. (2008). Generic protocol for monitoring impact of rotavirus vaccination on gastroenteritis disease burden and viral strains. *Who/Ivb/08.16*, 157.

World Health Organization (WHO). (2017). Diarrhoeal disease: Key facts. *Who*, *2017*(15 June), 1–5. https://doi.org/10.5897/AJAR2013.7983

Yinda, C. K., Shi, C., Deboutte, W., Conceição-Neto, N., Van Ranst, M., Beller, L., and Matthijnssens, J. (2019). Gut Virome Analysis of Cameroonians Reveals High Diversity of Enteric Viruses, Including Potential Interspecies Transmitted Viruses. *MSphere*, *4*(1). https://doi.org/10.1128/msphere.00585-18

APPENDICES

Appendix I: Primer and Probes included in the real-time PCR and TAC assays

Virus	Target gene	Strand	Nucleotide position	Sequence: Real-time PCR	Sequence: Taqman array card
Sapovirus	Polymerase/Capsid Junction	Primer-F	5077-5094	CAGGCTCTCGCCACCTAC	GAYCASGCTCTCGCYACCTAC
		Primer-R	5155-5177	CCCTCCATYTCAAACACTAWTTT	CCCTCCATYTCAAACACTA
		Probe	5101-5117	TGGTTCATAGGTGGTRC	TGGTTYATAGGYGG[b]
Sapo#2	Polymerase-capsid junction	Primer-F	5071-5094		TTTGAACAAGCTGTGGCRTGCTAC
		Primer-R	5155-5173		CCCTCCATYTCAAACACTA
		Probe	5101-5117		TGGTTYATAGGYGG[b]
Astrovirus	ORF1 (real-time PCR) ORF2 (TAC)	Primer-F	2209-2229	TCTYATAGACCGYATTATTGG	TCAACGTGTCCGTAAMATTGTCA
		Primer-R	2322-2301	TCAAAATTCTACATCATCACCA	GCWGGTTTTGGTCCTGTGA
		Probe	2295-2272	CCCCADCCATCATCATCTTCATCA	CAACTCAGGAAACARG
Rotavirus A	NSP3	Primer-F	963-982	ACCATCTWCACGTRACCCTC	ACCATCTWCACRTRACCCTCTATGAG
		Primer-R	1046-1027	CACATAACGCCCCTATAGCC	GGTCACATAACGCCCCTATAGC
		Probe	984-1016	ATGAGCACAATAGTTAAAAGCTAACACTGTCAA	AGTTAAAAGCTAACACTGTCAAA
Rota#cam	NSP3	Primer-F	963-982		ACCATCTWCACATGACCCTC
		Primer-R	1049-1026		GGTCACATAACGCCCC
		Probe	994-1016		TAGTTAAAAGCTAACACTGTCAA
Norovirus	RdRp/Capsid	Primer-F	5003-5080	CARGARBCNATGTTYAGRTGGATGAG	CARGARBCNATGTTYAGRTGGATGAG

GII	junction				
		Primer-R	5100-5080	TCGACGCCATCTTCATTCACA	TCGACGCCATCTTCATTCACA
		Probe	5067-5051	GAGGGSGATCGCRATCT	TGGGAGGGCGATCGCAATCT
Adenovirus	Hexon	Primer-F	17794-17772	GCCCCAGTGGTCTTACATGCACATC	GCCCCARTGGKCNTACATGCACATC
		Primer-R	17663-17687	GCCACGGTGGGGTTTCTAAACTT	GCCACIGTGGGRTTYCTRAACTT
		Probe	17703-17731	TCGGAGTACCTGAGCCCGGGTCTGGTGCA[b]	CTGGTGCARTTYGCCCG[b]

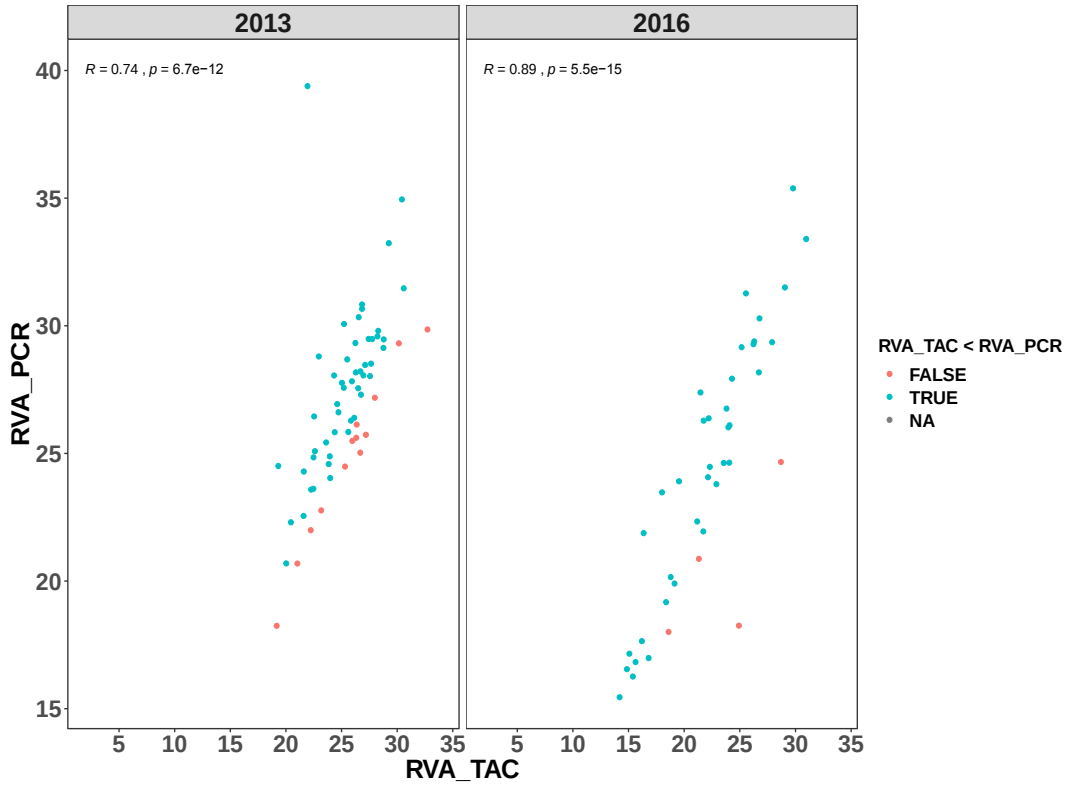
Green characters show regions that are identical between real-time PCR and TAC. [b]: reverse complemented sequences

Appendix II: Sensitivity and Specificity for ELISA, real-time PCR and TAC in RVA detection

	ProsPecT™ EIA Kit	**In-house real-time PCR**	**In-house TaqMan Array Card**	**Year**
Number of TP	61	64	63	2013
Number of FN	4	1	2	
Number of FP	2	0	1	
Number of TN	170	172	171	
Sensitivity	93.8% (85.0%-98.3%)	98.5%% (91.7%-100%)	96.9% (89.3%-99.6%)	
Specificity	98.8% (95.8%-99.9%)	100% (97.9%-100%)	99.4%(96.8%-100%)	
Number of TP	31	42	44	2016
Number of FN	14	3	1	
Number of FP	3	0	7	
Number of TN	204	207	200	
Sensitivity	68.9% (53.4%-81.8%)	93.3% (81.7%-98.6%)	97.8 (88.2%-99.9%)	
Specificity	98.6% (95.8%-99.7%)	100% (98.2%-100%)	96.6 (93.2%-98.6%)	

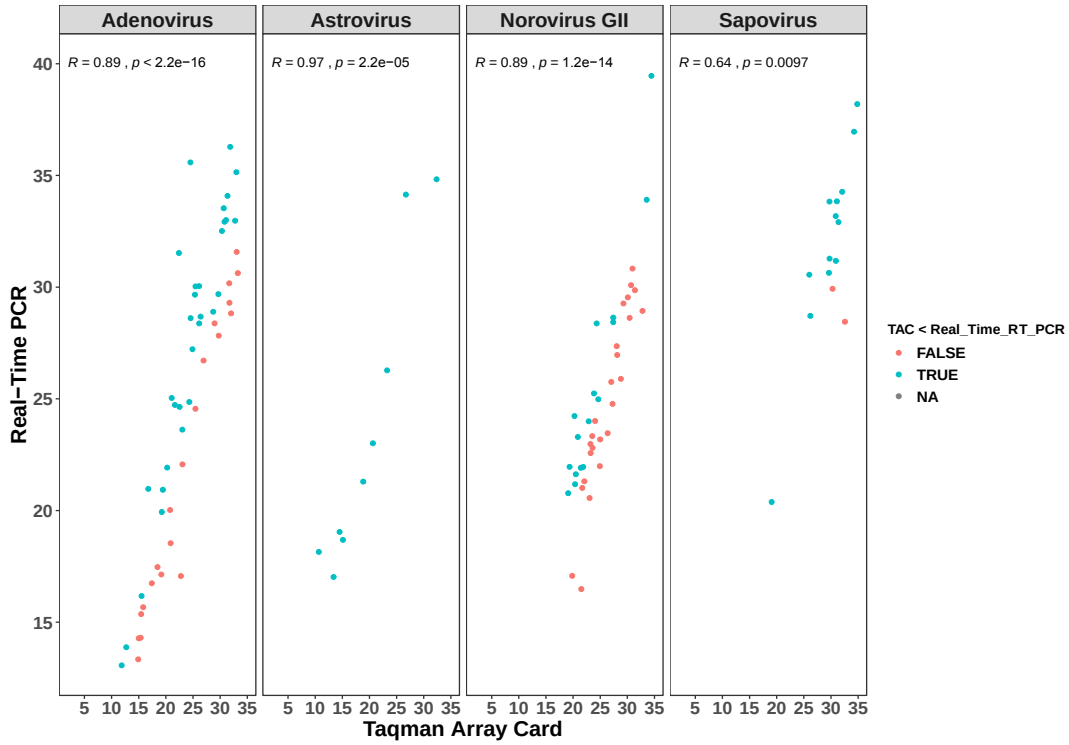
Appendix III: Correlation plots for TAC and real-time PCR Ct values from detection of RVA.

Green dots show interaction points where the Ct values for TAC were less than those for real-time PCR and red dots are vice versa.



Appendix IV: Correlation plots for TAC and real-time PCR Ct values in the detection of adenovirus, astrovirus, norovirus genogroup II and sapovirus.

Green dots show interaction points where the Ct values for TAC were less than those for real-time PCR and red dots are vice versa.



Appendix V: Summary of the steps and codes used in the bioinformatics workflows

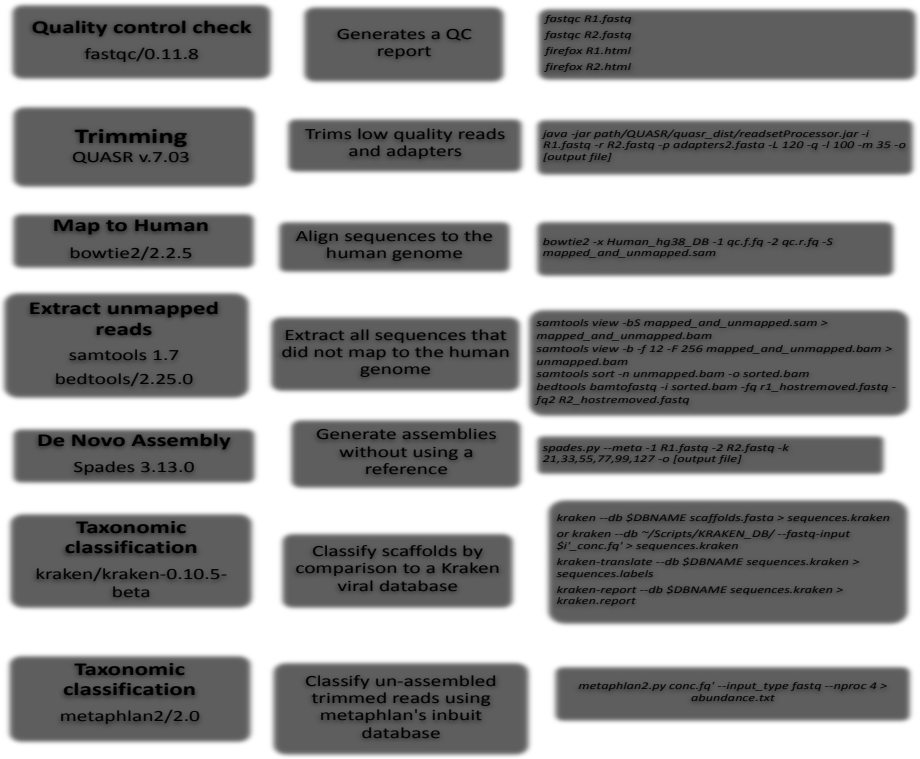


Figure 1: Bioinformatics workflow

Summary of the steps/codes performed during the bioinformatics pipeline for viral taxonomic classification

Appendix VI: Ethical Approval

KENYA MEDICAL RESEARCH INSTITUTE

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

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

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

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

Dear Sir,

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

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

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

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

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

Yours faithfully,

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

Appendix VII: Published Manuscript

pathogens

MDPI

Article

Epidemiological Trends of Five Common Diarrhea-Associated Enteric Viruses Pre- and Post-Rotavirus Vaccine Introduction in Coastal Kenya

Arnold W. Lambisia [1,2,*,†], Sylvia Onchaga [1,†], Nickson Murunga [1], Clement S. Lewa [1], Steven Ger Nyanjom [2] and Charles N. Agoti [1,3]

1 Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi 230-80108, Kenya; onchagasylvia@gmail.com (S.O.); nmurunga@kemri-wellcome.org (N.M.); clewa@kemri-wellcome.org (C.S.L.); cnyaigoti@kemri-wellcome.org (C.N.A.)
2 Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Juja 62000-00200, Kenya; snyanjom@jkuat.ac.ke
3 School of Health and Human Sciences, Pwani University, Kilifi 195-80108, Kenya
* Correspondence: ALambisia@kemri-wellcome.org; Tel.: +254-708-164-077
† Joint First Authors.

Received: 8 July 2020; Accepted: 10 August 2020; Published: 15 August 2020

check for updates

Abstract: Using real-time RT-PCR, we screened stool samples from children aged <5 years presenting with diarrhea and admitted to Kilifi County Hospital, coastal Kenya, pre- (2003 and 2013) and post-rotavirus vaccine introduction (2016 and 2019) for five viruses, namely rotavirus group A (RVA), norovirus GII, adenovirus, astrovirus and sapovirus. Of the 984 samples analyzed, at least one virus was detected in 401 (40.8%) patients. Post rotavirus vaccine introduction, the prevalence of RVA decreased (23.3% vs. 13.8%, $p < 0.001$) while that of norovirus GII increased (6.6% vs. 10.9%, $p = 0.023$). The prevalence of adenovirus, astrovirus and sapovirus remained statistically unchanged between the two periods: 9.9% vs. 14.2%, 2.4% vs. 3.2 %, 4.6% vs. 2.6%, ($p = 0.053$, 0.585 and 0.133), respectively. The median age of diarrhea cases was higher post vaccine introduction (12.5 months, interquartile range (IQR): 7.9–21 vs. 11.2 months pre-introduction, IQR: 6.8–16.5, $p < 0.001$). In this setting, RVA and adenovirus cases peaked in the dry months while norovirus GII and sapovirus peaked in the rainy season. Astrovirus did not display clear seasonality. In conclusion, following rotavirus vaccine introduction, we found a significant reduction in the prevalence of RVA in coastal Kenya but an increase in norovirus GII prevalence in hospitalized children.

Keywords: viral diarrhea; real-time PCR; rotavirus vaccination; Kenya

1. Introduction

In the year 2016 alone, approximately 300,000 children aged <5 years succumbed to diarrhea in sub-Saharan Africa [1]. Viral pathogens including rotavirus group A (RVA), adenovirus (type 40/41), astrovirus, norovirus (genogroup GI and GII) and sapovirus are among the top causative agents of severe diarrhea globally [2,3]. Understanding their epidemiological patterns such as prevalence, incidence, seasonality, clinical severity and infection age distribution in local settings is essential for designing and prioritizing interventions. Historically, RVA has been the single most important cause of severe childhood diarrhea, responsible for ~38% (95% CI: 4.8–73.4%) of hospital cases (<5 years) pre-vaccine introduction [4]. However, RVA prevalence has been rapidly declining since 2009 and was approximately 23% (95% CI: 0.7–57.7%) in 2016, in settings where the rotavirus vaccine was in use [4]. Due to the shared ecological niche and the apparent decline of all-cause gastroenteritis-associated

hospital admissions, it has been hypothesized that rotavirus vaccination has likely impacted the epidemiology of the other enteric viruses [5]. However, there are contradicting reports on the specific impact of rotavirus vaccination on the prevalence of the individual enteric viruses—for example, norovirus [6,7]. This has not been adequately examined in African populations where diarrhea burden is highest. Kenya began rotavirus vaccination in July 2014 using the monovalent Rotarix® (RV1), derived from G1P[8] strain, administered at 6 and 10 weeks of life. RV1 vaccine coverage in Kenya has increased over time since 2014 but is varied by age group, number of doses and geographic region in Kenya [8]. Within Kilifi County, coastal Kenya, coverage in 2017 in <1-year-olds was 73% (at least one dose) vs. 65% (complete two doses), while in <12–24 month-olds, it was 86% (at least one dose) vs. 84% (complete two doses) [9].

The KEMRI/Wellcome Trust Research Programme (KWTRP) has been running surveillance of RVA since 2002 in children admitted to the Kilifi County Hospital (KCH). The current study screened archived diarrheal samples from KCH, spanning both the pre- and post-rotavirus vaccine introduction periods in Kenya for RVA, astrovirus, adenovirus (all serotypes), sapovirus and norovirus (only GII) using real-time reverse-transcription polymerase chain reaction (RT-PCR) approach. We update on the prevalence of these viral diarrheal agents and their seasonal patterns pre and post introduction of the rotavirus vaccination program in Kenya.

2. Results

2.1. Study Population Characteristics

Out of 2156 children aged <5 years who presented with diarrhea at KCH during the four selected years (2003, 2013, 2016 and 2019), 1397 (64.8%) provided a stool sample; see Table 1. Overall, the demographic characteristics of the eligible children sampled, and eligible children not sampled, differed in age strata distribution ($p = 0.001$) and discharge outcome ($p < 0.001$); see Table 1. The main reasons for failure to sample eligible children were as follows: death ($n = 21$, 2.8%), discharge or transfer before sample collection ($n = 296$, 40.0%), consent refusal ($n = 315$, 41.5%) or other ($n = 127$, 16.7%). Among the sampled cases, 984 (70.4%) had a specimen available and tested by real-time RT-PCR for the five enteric viruses, and these were included in subsequent analysis. The median age of the sampled participants was significantly higher for the post-vaccine introduction period compared to pre-vaccine introduction period ($p < 0.001$); see Table 2.

Table 1. Characteristics of children under 5 years of age admitted to Kilifi County Hospital (KCH), coastal Kenya, with diarrhea symptoms that were sampled versus those who were not sampled in the study.

Characteristics	All Subjects	Sampled (%)	Not Sampled (%)	*p*-Value
Total Admissions	2156	1397 (64.8)	759 (35.2)	
Admissions Per Year				
2003	1007 (46.7)	587 (42.0)	420 (55.3)	
2013	332 (15.4)	254 (18.2)	78 (10.3)	
2016	334 (15.5)	257 (18.4)	77 (10.1)	
2019	483 (22.4)	299 (21.4)	184 (24.2)	
Gender				0.838
Male	1262 (58.5)	815 (58.3)	447 (58.9)	
Female	894 (41.5)	582 (41.7)	312 (41.1)	
Age				
Median (IQR)	12.4 (7.7–20.5)	11.7 (7.4–19.7)	13.8 (8.5–22.1)	**<0.001**
Mean (SD)	15.7 (11.4)	15.0 (11.1)	16.9 (12.0)	**<0.001**
Age Group				**0.001**
0–11 Months	1045 (48.4)	718 (51.4)	326 (43.0)	
12–23 Months	716 (33.2)	444 (31.8)	272 (35.8)	
24–59 Months	396 (18.4)	235 (16.8)	161 (21.2)	

Table 1. *Cont.*

Characteristics	All Subjects	Sampled (%)	Not Sampled (%)	*p*-Value
Discharge Outcome (*n* = 2153) #				**<0.001**
Alive	1918 (88.9)	1306 (93.5)	612 (80.5)	
Dead	235 (10.9)	89 (6.4)	146 (19.3)	

SD means standard deviation; IQR means interquartile range. Not sampled: sample was not collected due to lack of consent, time-up, death and others. # Discharge outcome data for three subjects were missing.

Table 2. Characteristics of children under 5 years of age admitted to KCH, coastal Kenya, with diarrhea symptoms and tested pre-vaccine introduction versus those tested post-vaccine introduction.

Characteristics	Total	Pre-Vaccine Introduction (%)	Post-Vaccine Introduction (%)	*p*-Value
Number of Samples Tested	984	454 (46.1)	530 (53.9)	
Samples Tested (Year)				
2003	223	223	-	
2013	231	231	-	
2016	239	-	239	
2019	291	-	291	
Gender				0.847
Male	570 (57.9)	261 (57.5)	309 (58.3)	
Female	414 (42.1)	193 (42.5)	221 (41.7)	
Age				
Mean (SD)	15 (11.2)	13.4 (9.9)	16.3 (12)	<0.001
Median (IQR)	11.7 (7.3–19.3)	11.2 (6.8–16.5)	12.5 (7.9–21)	<0.001
Age group				0.003
0–11 Months	505 (51.3)	252 (55.5)	253 (47.7)	
12–23 Months	323 (32.8)	148 (32.6)	175 (33.0)	
24–59 Months	156 (15.9)	54 (11.9)	102 (19.3)	
Disease Severity in RVA Cases = *n* (139)				
Mild	12 (8.6)	7 (10.6)	5 (6.8)	0.441
Moderate	50 (36.0)	26 (39.4)	24 (32.9)	
Severe	77 (55.4)	33 (50)	44 (60.3)	
Discharge Outcome = *n* (982) #				0.556
Alive	925 (94.2)	425 (93.6)	500 (94.7)	
Dead	57 (5.8)	29 (6.4)	28 (5.3)	

SD means standard deviation; IQR means interquartile range; RVA means rotavirus group A. Values given are the counts and percentages are provided in brackets. # Discharge outcome for two subjects was missing. Disease Severity Was Calculated Using the Vesikari Clinical Severity Scoring System Manual [10].

2.2. *Overall Virus Detection*

Of the 984 samples analyzed, at least one of the viruses was detected in 401 samples (40.8%) at the real-time RT-PCR cycle threshold (Ct) value of <35.0. The lower the Ct value, the higher the virus titer in the sample. The detection frequency differed significantly for adenovirus ($p = 0.001$) and sapovirus ($p < 0.001$) pre- and post-rotavirus vaccine introduction when the Ct cut-off value was gradually lowered (<30, <35, <40), unlike for RVA, astrovirus and norovirus GII; see Figure 1. All our subsequent analyses were undertaken at Ct value <35.0 Single infections were detected in 354 specimens (36.0%) and included RVA ($n = 149$, 42.1%), adenovirus ($n = 91$, 25.7%), norovirus GII ($n = 75$, 21.2%), sapovirus ($n = 20$, 5.7%) and astrovirus ($n = 18$, 5.1%).

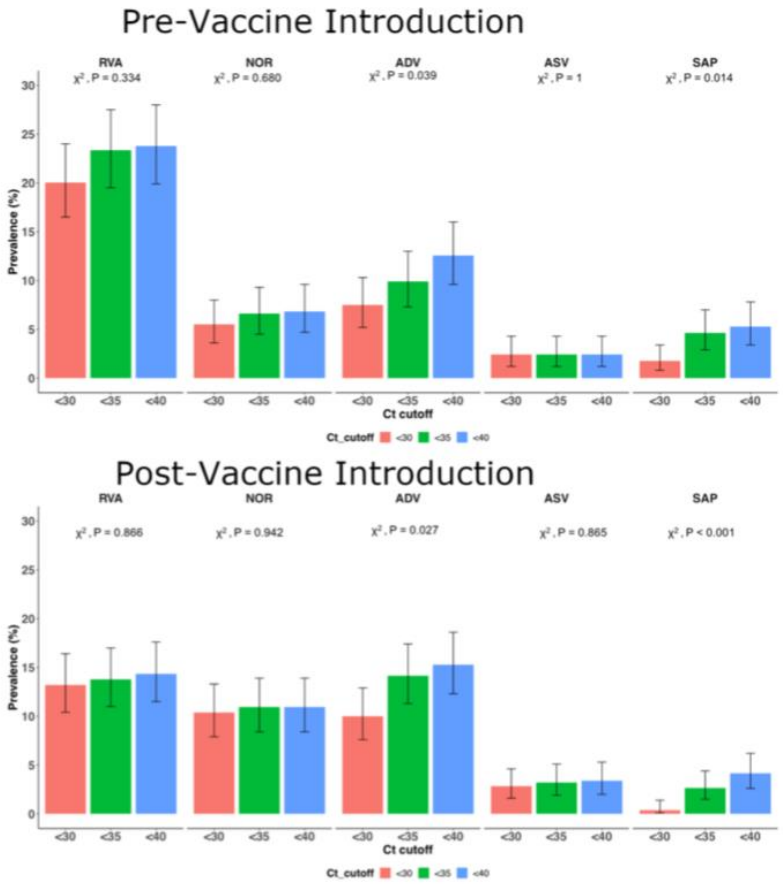


Figure 1. Detection frequency of RVA, adenovirus, norovirus GII, astrovirus and sapovirus at different cycle threshold (Ct) cutoffs for children under 5 years of age admitted to KCH Kenya with diarrhea symptoms. The error bars represent 95% confidence interval for the proportions. Proportions were compared using chi-square test. RVA stands for rotavirus group A, ADV stands for adenovirus, NOR stands for norovirus GII, ASV stands for astrovirus and SAP stands for sapovirus.

2.3. *Patterns Pre-Post Vaccine Introduction*

RVA showed a significant decrease (23.3% vs. 13.8%, $p < 0.001$) in prevalence while norovirus GII showed a significant increase (6.6% vs. 10.9%, $p = 0.02$) post-vaccine introduction compared to pre-vaccine introduction; see Table 3. There were no significant changes in the prevalence of astrovirus ($p = 0.585$), adenovirus ($p = 0.053$) and sapovirus ($p = 0.133$) pre- and post-RVA vaccine introduction (chi-squared (χ^2) test); see Table 3. Notably, norovirus GII had a gradual increase in prevalence across the four years, from 6.7% (95% CI: 3.8–10.9%) to 12.4% (95% CI: 8.8–16.7%); see Figure 2. RVA was

the most commonly detected virus across all years, except in year 2019, in which adenovirus had the highest prevalence; see Figure 2.

Table 3. Comparison of the prevalence of viral detection in children under 5 years of age admitted to KCH Kenya with diarrhea symptoms pre- and post-rotavirus vaccine introduction.

Viruses Detected	Total	Pre-Vaccine Introduction (%)	Post-Vaccine Introduction (%)	*p*-Value
Samples Tested	984	454 (46.1)	530 (53.9)	
Rotavirus Group A	179 (18.2)	106 (23.3)	73 (13.8)	<0.001
Adenovirus	120 (12.2)	45 (9.9)	75 (14.2)	0.053
Norovirus GII	88 (8.9)	30 (6.6)	58 (10.9)	0.023
Astrovirus	28 (2.8)	11 (2.4)	17 (3.2)	0.585
Sapovirus	35 (3.6)	21 (4.6)	14 (2.6)	0.133

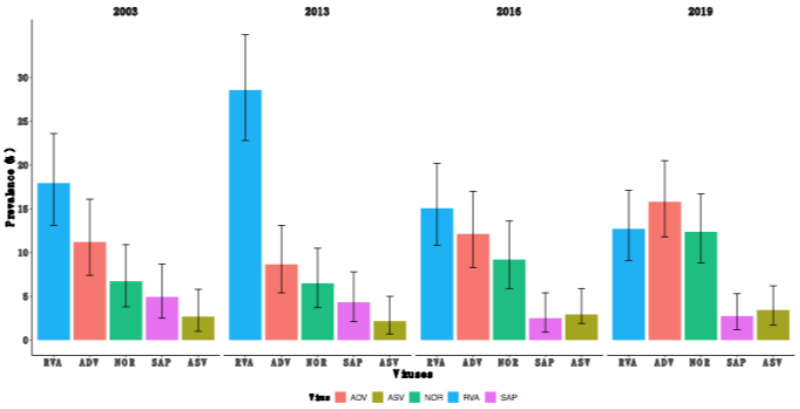


Figure 2. Prevalence of RVA, adenovirus, norovirus GII, astrovirus and sapovirus in 2003, 2013, 2016 and 2019 in children under 5 years of age admitted to KCH Kenya with diarrhea symptoms. The error bars represent 95% confidence interval for the proportions. Proportions were compared using chi-square test. Abbreviations used for viruses as in Figure 1.

Notably, RVA and sapovirus cases in the post-vaccine introduction period had statistically significant lower and higher median Ct values, respectively, compared to the pre-vaccine period (Wilcoxon, p value < 0.001); see Figure 3. This was not observed for the other three screened viruses pre- and post-rotavirus vaccine introduction. The median age of the RVA positive cases was significantly higher for the post-vaccine introduction period (14.0 months) compared to the pre-vaccine introduction period (10.4 months) (Wilcoxon, $p < 0.001$). A similar shift was not observed for the other viruses; see Figure 4.

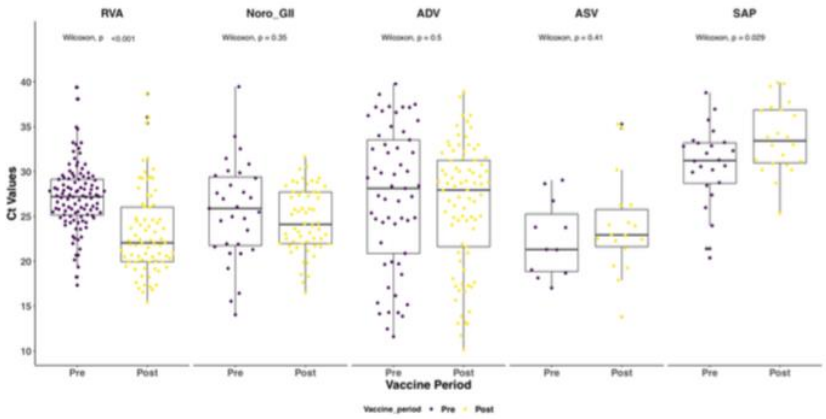


Figure 3. Distribution of Ct values among cases under 5 years of age admitted to KCH Kenya with diarrhea symptoms pre- and post-vaccine introduction. RVA stands for rotavirus group A, ADV stands for adenovirus, NOR GII stands for norovirus GII, ASV stands for astrovirus and SAP stands for sapovirus.

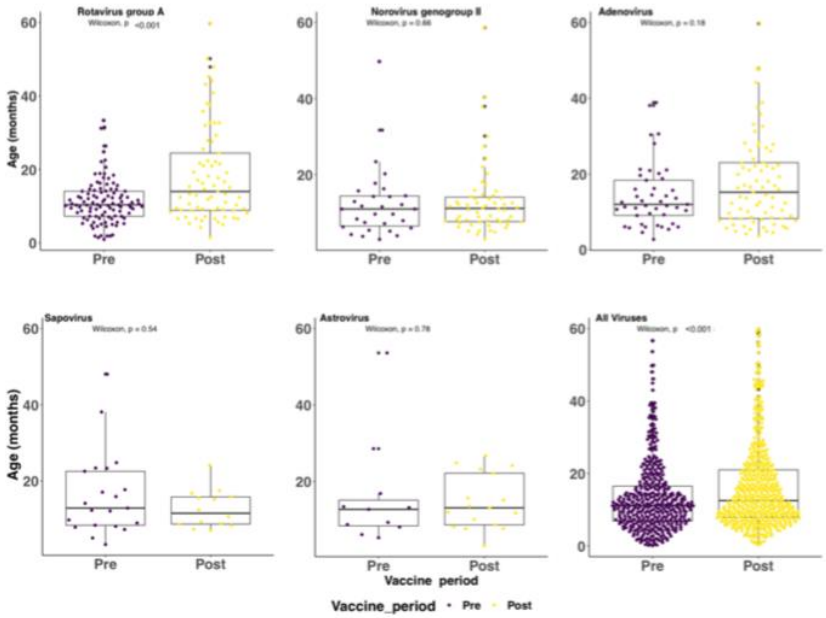


Figure 4. Distribution of age in months among cases under 5 years of age admitted to KCH Kenya with diarrhea symptoms pre- and post-vaccine introduction.

2.4. Virus Coinfections (i.e., Two or More Viruses in a Single Specimen)

These were detected in 47 specimens (4.8%). In 583 specimens (59.2%), none of the targeted viruses was detected. The prevalence of coinfections pre-vaccine was 4.4% (95% CI: 2.7–6.7%), while in the post-vaccine introduction period, this value was 5.7% (95% CI: 3.9–8.0%), $p = 0.454$. RVA and astrovirus were the most common coinfections in the pre-vaccine introduction period ($n = 6$), while in the post-vaccine introduction period, it was RVA and adenovirus ($n = 15$); see Table 4.

Table 4. Coinfections pre- and post-rotavirus vaccine introduction. RVA stands for rotavirus group A, ADV stands for adenovirus, NOR GII stands for norovirus GII, ASV stands for astrovirus and SAP stands for sapovirus.

PATHOGEN COINFECTION	PRE-VACCINE INTRODUCTION	POST-VACCINE INTRODUCTION
RVA & NOR GII	1	2
RVA & ADV	2	15
RVA & ASV	3	0
RVA & SAP	6	1
NOR GII& ADV	3	4
NOR GII & ASV	0	1
NOVGII & SAP	2	1
ADV & ASV	1	3
ADV & SAP	1	1
ASV & SAP	1	2

Abbreviations used for viruses as in Figure 3.

2.5. Circulating RVA Genotypes Pre- and Post-Vaccine Introduction

G1P[8] was the predominant RVA genotype pre vaccine introduction. However, in the post-vaccine introduction period, the predominant genotypes were G2P[4] (2016) and G3P[8] (2019); see Table 5.

Table 5. Frequency of RVA genotypes detected in coastal Kenya pre- (2003 and 2013) and post- (2016 and 2019) vaccine introduction.

Year	2003		2013		2016		2019	
	No. of Cases	%	No. of Cases	%	No. of Cases	%	No. of Cases	%
RVA Positive	40		66		36		37	
Genotyped	2	5.0	48	72.7	34	94.4	36	97.3
Genotypes								
G1P[8]	1	50.0	43	89.6	5	14.7	1	2.8
G2P[4]	-	-	2	4.2	29	85.3	-	-
G3P[8]	-	-	1	2.1	-	-	34	94.4
G9P[8]	1	50.0	1	2.1	-	-	-	-
G10P[8]	-	-	1	2.1	-	-	-	-
G8P[8]	-	-	-	-	-	-	1	2.8

2.6. Seasonality of the Detected Viruses

We constrained this analysis to the years 2013, 2016 and 2019, where >70% of the eligible patients had been analyzed. Pre-vaccine introduction (in 2013), for RVA, there were two peak months, in June and September. However, post-vaccine introduction (in 2016 and 2019), there was only a single peak month for RVA in September and August, respectively. For norovirus GII, cases were observed throughout the year, with peak months varying from year-to-year, in July, April and June in 2013, 2016 and 2019, respectively. Similarly, adenovirus cases appeared to occur throughout the year, with two peak months in 2013 (June and September) and one peak month in 2016 and 2019 (August for both). For sapovirus and astrovirus, we observed less than five cases monthly between January and August and no cases in the last quarter of each the three years; see Figure 5.

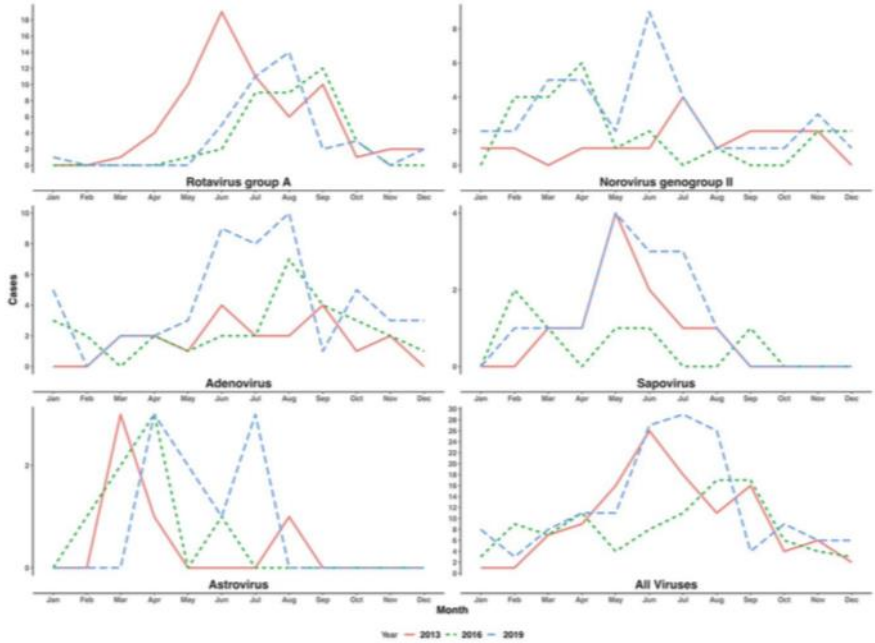


Figure 5. The frequency of detection of RVA, adenovirus, norovirus GII, astrovirus and sapovirus by month in children under 5 years of age admitted to KCH Kenya with diarrhea in 2013, 2016 and 2019.

2.7. *Primer/Probe Mismatches with Contemporary Sequences*

Nucleotide mismatches were observed in either or both the primers and probes and the viral target sequences for all the viruses except for norovirus GII; see Figure 6. The RVA forward primer had a G-A and A-G mismatches at positions 12 and 15, respectively. Adenovirus had two mismatches in the forward primer (C-G and G-A), three mismatches in the probe (C-T, C-T and T-C) and two mismatches in the reverse primer (T-C and C-T), and none of them were within five bases of the 3′ end. Mismatches within the sapovirus primer/probe binding sites were pronounced in sapovirus genogroup V and included six mismatches in the forward primer, three mismatches in the probe and two mismatches in the reverse primer. Some of the mismatches were within five bases of the 3′ end (forward primer: C-G, probe: T-C, reverse primer: A-C and T-C). Astrovirus primers and probe did not have pronounced mismatches present in all the sequences—rather, they had mismatches in individual sequences; see Figure 6.

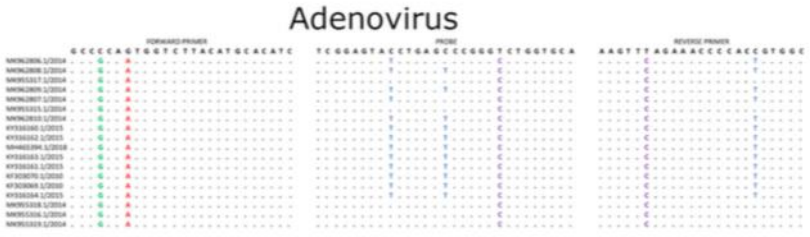


Figure 6. *Cont.*

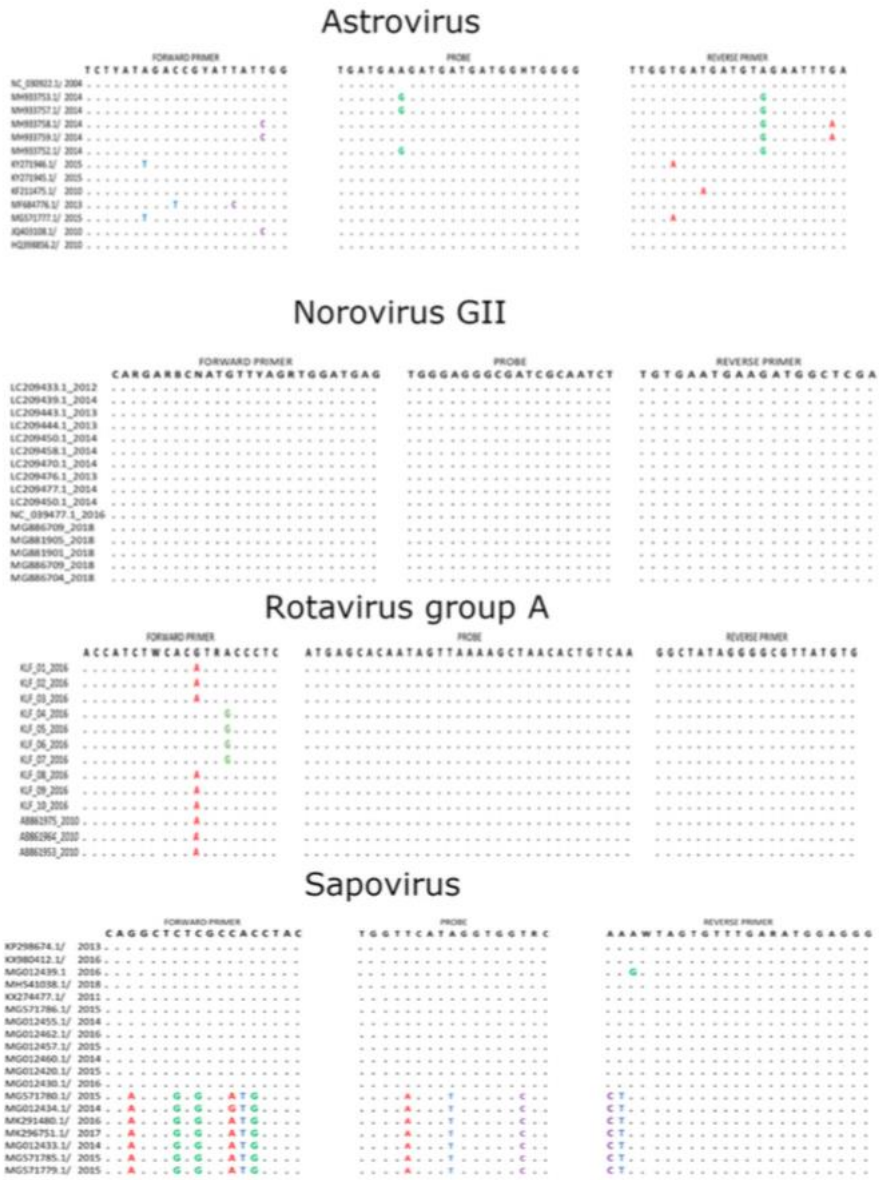


Figure 6. The primers and probes target sites for RVA, adenovirus and norovirus GII, sapovirus and astrovirus were aligned using MAFFT v.7.31313 and the alignments were trimmed to the region of the primer and probe target sites. Nucleotide differences between the expected primer and probe target sites and the viral sequences were identified and highlighted. Dots indicate identity with primer or probe sequences.

3. Discussion

We observed a significant decrease in the prevalence of RVA in the post-vaccine introduction period in KCH, concurring with findings of a recent multi-site study in Kenya that reported RVA vaccine effectiveness of ~64% (95% CI: 35–80%) and a reduction in rotavirus-associated hospital admissions two years post-vaccine introduction of ~80% (95% CI: 46–93%) [9,11]. Note that Kenya rotavirus vaccine coverage was considered medium in 2018 (70–79%) [12]. Our pre- and post-vaccine introduction analysis observed a significant increase in the prevalence of norovirus GII in KCH post-rotavirus vaccine introduction, as similarly observed in the United States, Nicaragua and Bolivia following RVA vaccine introduction [13–15]. It is unclear if this has been driven by an established biological interaction between these two viruses or that this reflects natural norovirus GII fluctuation in prevalence across multiple years.

The shift in the predominant genotypes pre- and post-vaccine introduction from G1P[8] to G2P[4] in 2016 and G3P[8] in 2019 in our setting has also been described elsewhere, e.g., in Belgium, Madagascar and Ethiopia [16–18]. G3P[8] was the predominant genotype in this setting in 2018 [19] and it continued being the dominant genotype in 2019. Although these dominant post-vaccine genotypes are either partially or fully heterotypic to the Rotarix G1P[8] strain, in their surface exposed immunodominant proteins, there is not enough evidence yet to directly attribute their increased incidence to vaccine introduction [20]. Additional analysis will help to bring better understanding on the reason behind their dominance.

Despite RV vaccine introduction in Kilifi, Kenya, no significant difference was observed in the discharge outcome for all causes of diarrhea pre- and post-rotavirus vaccine introduction. We suggest two explanations for this. Firstly, the majority of the children who were eligible to be in this study and died did not have a sample collected to determine their RVA and other enteric pathogens' status. Secondly, inpatient mortality of children treated for diarrhea in Kilifi County Hospital has been previously found to be predicted by a positive HIV test, bacteremia and poor nutritional status [21]. This may have not changed pre- or post-introduction of rotavirus vaccination.

RVA Ct values were decreased in post-vaccine samples compared to pre-vaccination years. This was despite RVA disease severity remaining unchanged between the two periods. Different extraction methods were used to process the samples between 2003, 2013 and 2016, 2019. However, according to Liu et al., the difference in the extraction methods for enteric pathogen studies is not significant, except for norovirus GII, which showed a higher Ct value with kits targeting RNA purification alone compared to those targeting total nucleic acid (TNA) (difference within 1 Ct value). Different extraction kits were used in this study because raw stool samples from 2003 to 2016 were already destroyed following a directive by the WHO in 2016 that was part of the larger global polio eradication effort.

It has been previously noted the introduction to rotavirus vaccines may result in the shift of diarrhea disease burden to slightly older age groups [20]. Our study found a significant increase in the median age of diarrhea cases post-vaccine introduction (12.5 months) compared 11.2 months pre-introduction. This in part may be explained by the higher immunity at both individual and population levels against rotavirus that wanes as children grow older.

On local seasonality patterns, in each year, a peak month(s) of occurrence was observed for RVA, norovirus GII, sapovirus and adenovirus but not astrovirus. The Kilifi area has a tropical climate with two rainy seasons; the main rains usually peak in May (up to July) while the short rains usually peak in November (can run from October to December). RVA and adenovirus appeared to peak in the dry months while norovirus GII and sapovirus peaked in the rainy season. Similar patterns in the seasonality of RVA, adenovirus, norovirus GII and sapovirus have been observed elsewhere [22–25]. The seasonality of astrovirus is not well described.

The performance of qPCR assays can be impacted by mismatches within the last five bases at the 3′ end of primers and probe or/and the number of mismatches being more than five in the primers and probe [26,27]. The mismatches observed in the primer and probe binding sites of adenovirus,

astrovirus and sapovirus may have impaired the real-time PCR function by blocking the amplification or increasing the quantification cycles. Consequently, this may have impacted the estimated frequency of detection of these viruses. Unlike for RVA, the magnitude of the mismatches in qPCR function could have been shown better using recent local sequences of the other viruses.

This study had limitations: firstly, we did not analyze healthy children in the community to inform on the background prevalence of the five viruses in our study population. Secondly, the adenovirus assay was not specific to type 40/41 alone; thus, some of the adenoviruses detected may not be associated with diarrhea. Thirdly, a significant number of eligible cases were not sampled, including those who died before sampling. This potentially biased prevalence of the screened pathogens in the study population. Fourthly, extracting TNA from samples after many years of storage could lead to lower Ct values due to deterioration. Finally, the seasonality of examined pathogens will be best described if we examine more years.

In conclusion, we found a significant decline in the prevalence of rotavirus in hospitalized children in coastal Kenya after rotavirus vaccine introduction. This finding reinforces evidence of the continued benefit of rotavirus vaccination in this setting. Concomitantly, there has been a surge in norovirus GII prevalence, but the factors driving this increase are unclear and will require future investigation. The observation that the screened viruses peak at different times of the year also would benefit further investigation in order to understand drivers of their transmission and inform the design of effective intervention measures.

4. Materials and Methods

4.1. Study Site and Population

This study was undertaken at KCH, a referral hospital serving the Kilifi County population, which is majorly a rural population. We utilized stool specimens collected during routine surveillance of rotavirus in children with diarrhea as one of their illness symptoms, aged below five years and admitted to KCH [9,11]. Diarrhea was defined as observation of three or more loose stools in the preceding 24-h period. In this study, we selected two pre-vaccine years (2003 and 2013) and two post-vaccine years (2016 and 2019) for analysis. A stool specimen was collected from children who met the diarrhea case-definition following parental or guardian consent. The study protocol was approved by the Scientific and Ethics Review Unit (SSC#2861 and SERU#CGMRC/113/3624) based at KEMRI, Nairobi, Kenya.

4.2. Laboratory Methods

Irrespective of their previously determined rotavirus status, TNA were extracted from 0.2 g of 2003 and 2013 specimens (or 200 µL if liquid) using the cador Pathogen 96 QIAcube HT Kit (Qiagen, Manchester, UK). For 2016 and 2019 specimens, TNA were extracted using QIAamp Fast DNA Stool Mini kit (Qiagen, Manchester, UK) as per the manufacturer's instructions. Fecal specimens from the post-vaccine period (0.2 mg or 200 µL) were subjected to bead beating prior to TNA extraction and collected in a 200 µL of elution buffer [28].

The TNA extracts were screened for the five viruses by a two-step real-time RT-PCR assay [29]. First, cDNA was synthesized in a total volume of 20 µL using random hexamers and 5µL of TNA using the Omniscript Reverse Transcriptase kit (Qiagen, Manchester, UK), as per the manufacturer's instructions. Two µL of the cDNA was henceforth used for real-time RT-PCR in a total volume of 20 µL using the QuantiFast RT-PCR Kit (Qiagen, Manchester, UK) and run on the ABI 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Primers and probes were adopted from previously published work [30]. The presence of nucleotide mismatches in the primer and probe binding sites was investigated by aligning the primers/probes to genomic sequences deposited in GenBank from 2010 to 2019, using MAFFT v.7.313 [31]. The adenovirus probe/primer pair used in this study detected adenovirus serotypes beyond type 40/41. We used three Ct cut-off values (<40.0, <35.0

and <30.0) to define positive samples. Samples that were positive for RVA in 2003, 2013, 2016 and 2019 were processed for RVA genotyping using VP4 and VP7 RT-PCR, followed by either dideoxy sanger sequencing, as described elsewhere [19], or next-generation sequencing on the Illumina Miseq platform [32].

4.3. Statistical Analysis

All statistical analyses were performed using R version 3.6.1 [33]. Prevalence was defined as the proportion of these viruses in a hospital-admitted diarrhea patient population during the study period in Kilifi, Kenya. Means and medians of continuous variables were compared using a Kruskal Wallis and Wilcoxon rank-sum test, respectively. Binary data were summarized using proportions and comparisons between groups made using χ^2 statistics. A p value of <0.05 was considered statistically significant. Diarrhea severity in RVA positive cases pre- (year 2013) and post- (years 2016 and 2019) was assessed using the Vesikari Clinical Severity Scoring System Manual [10], with a modification in the treatment parameter. If the participant was given oral rehydration therapy or intravenous fluid therapy, they received a score of one or two, respectively.

Author Contributions: Conceptualization, A.W.L. and C.N.A.; methodology, A.W.L. and C.N.A.; formal analysis, A.W.L., N.M. and C.N.A.; investigation, A.W.L., S.O., C.S.L. and C.N.A.; resources, D.J.N. and C.N.A.; data curation, A.W.L. and N.M.; writing—original draft preparation, A.W.L.; writing—review and editing, A.W.L., S.O., N.M., C.S.L., S.G.N. and C.N.A.; visualization, A.W.L. supervision, S.G.N. and C.N.A., project administration, C.N.A.; funding acquisition, C.N.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Wellcome Trust (102975,203077) The authors Arnold Lambisa, Sylvia Onchaga and Charles Agoti were supported by the Initiative to Develop African Research Leaders (IDeAL) through the DELTAS Africa Initiative (DEL-15-003). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (107769/Z/10/Z) and the UK government. The views expressed in this publication are those of the authors and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government. This paper is published with the permission of the Director of KEMRI.

Acknowledgments: We thank all the study participants for their contribution of study samples, their parents/guardians, members of the viral epidemiology and control research group (http://virec-group.org/) and colleagues at the KEMRI Wellcome Trust Research Programme for their useful discussions during the preparation of the manuscript. We are grateful to James Nokes of KEMRI-Wellcome Trust for his comments and suggestions on the presentation of this work This paper is published with the permission of the Director of KEMRI.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Troeger, C.; Blacker, B.F.; Khalil, I.A.; Rao, P.C.; Cao, S.; Zimsen, S.R.; Albertson, S.B.; Stanaway, J.D.; Deshpande, A.; Abebe, Z.; et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **2018**, *18*, 1211–1228. [CrossRef]
2. Platts-Mills, J.A.; Liu, J.; Rogawski, E.T.; Kabir, F.; Lertsethtakarn, P.; Siguas, M.; Khan, S.S.; Praharaj, I.; Murei, A.; Nshama, R.; et al. 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. *Lancet Glob. Health* **2018**, *6*, e1309–e1318. [CrossRef]
3. Liu, J.; Platts-Mills, J.A.; Juma, J.; Kabir, F.; Nkeze, J.; Okoi, C.; Operario, D.J.; Uddin, J.; Ahmed, S.; Alonso, P.L.; et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: A reanalysis of the GEMS case-control study. *Lancet* **2016**. [CrossRef]
4. Aliabadi, N.; Antoni, S.; Mwenda, J.M.; Weldegebriel, G.; Biey, J.N.M.; Cheikh, D.; Fahmy, K.; Teleb, N.; Ashmony, H.A.; Ahmed, H.; et al. Global impact of rotavirus vaccine introduction on rotavirus hospitalisations among children under 5 years of age, 2008–2016: Findings from the Global Rotavirus Surveillance Network. *Lancet Glob. Health* **2019**, *7*, e893–e903. [CrossRef]

5. Yu, W.J.; Chen, S.Y.; Tsai, C.N.; Chao, H.C.; Kong, M.S.; Chang, Y.J.; Chiu, C.H. Long-term impact of suboptimal rotavirus vaccines on acute gastroenteritis in hospitalized children in Northern Taiwan. *J. Formos. Med. Assoc.* **2018**, *117*, 720–726. [CrossRef] [PubMed]
6. Muhsen, K.; Kassem, E.; Rubenstein, U.; Goren, S.; Ephros, M.; Shulman, L.M.; Cohen, D. No evidence of an increase in the incidence of norovirus gastroenteritis hospitalizations in young children after the introduction of universal rotavirus immunization in Israel. *Hum. Vaccines Immunother.* **2019**, *15*, 1284–1293. [CrossRef]
7. Halasa, N.; Piya, B.; Stewart, L.S.; Rahman, H.; Payne, D.C.; Woron, A.; Thomas, L.; Constantine-Renna, L.; Garman, K.; McHenry, R.; et al. The Changing Landscape of Pediatric Viral Enteropathogens in the Post-Rotavirus Vaccine Era. *Clin. Infect. Dis.* **2020**, *53*, 1689–1699. [CrossRef]
8. Wandera, E.A.; Mohammad, S.; Ouko, J.O.; Yatitch, J.; Taniguchi, K.; Ichinose, Y. Variation in rotavirus vaccine coverage by sub-counties in Kenya. *Trop. Med. Health* **2017**. [CrossRef]
9. Otieno, G.P.; Bottomley, C.; Khagayi, S.; Adetifa, I.; Ngama, M.; Omore, R.; Ogwel, B.; Owor, B.E.; Bigogo, G.; Ochieng, J.B.; et al. Impact of the Introduction of Rotavirus Vaccine on Hospital Admissions for Diarrhea Among Children in Kenya: A Controlled Interrupted Time-Series Analysis. *Clin. Infect. Dis.* **2019**, 1–8. [CrossRef]
10. Lewis, K. Vesikari Clinical Severity Scoring System Manual. Path. 2011, pp. 1–50. Available online: https://www.path.org/publications/files/VAD_vesikari_scoring_manual.pdf (accessed on 22 July 2020).
11. Khagayi, S.; Omore, R.; Otieno, G.P.; Ogwel, B.; Ochieng, J.B.; Juma, J.; Apondi, E.; Bigogo, G.; Onyango, C.; Ngama, M.; et al. Effectiveness of monovalent rotavirus vaccine against hospitalization with acute rotavirus gastroenteritis in Kenyan children. *Clin. Infect. Dis.* **2020**, *70*, 2298–2305. [CrossRef]
12. VIEW-hub, International Vaccine Access Center (IVAC), Johns Hopkins Bloomberg School of Public Health. Available online: https://view-hub.org/map/?set=wuenic-coverage&group=vaccine-coverage&category=rv (accessed on 31 July 2020).
13. Bucardo, F.; Reyes, Y.; Svensson, L.; Nordgren, J. Predominance of norovirus and sapovirus in nicaragua after implementation of universal rotavirus vaccination. *PLoS ONE* **2014**, *9*, e98201. [CrossRef] [PubMed]
14. McAtee, C.L.; Webman, R.; Gilman, R.H.; Mejia, C.; Bern, C.; Apaza, S.; Espetia, S.; Pajuelo, M.; Saito, M.; Challappa, R.; et al. Burden of norovirus and rotavirus in children after rotavirus vaccine introduction, Cochabamba, Bolivia. *Am. Trop. Med. Hyg.* **2016**. [CrossRef] [PubMed]
15. 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.; et al. Norovirus and medically attended gastroenteritis in US children. *N. Engl. J. Med.* **2013**. [CrossRef] [PubMed]
16. Gelaw, A.; Pietsch, C.; Liebert, U.G. Molecular epidemiology of rotaviruses in Northwest Ethiopia after national vaccine introduction. *Infect. Genet. Evol.* **2018**. [CrossRef] [PubMed]
17. Rahajamanana, V.L.; Raboba, J.L.; Rakotozanany, A.; Razafindraibe, N.J.; Andriatahirintsoa, E.J.P.R.; Razafindrakoto, A.C.; Mioramalala, S.A.; Razaiarimanga, C.; Weldegebriel, G.G.; Burnett, E.; et al. Impact of rotavirus vaccine on all-cause diarrhea and rotavirus hospitalizations in Madagascar. *Vaccine* **2018**. [CrossRef] [PubMed]
18. Zeller, M.; Rahman, M.; Heylen, E.; De Coster, S.; De Vos, S.; Arijs, I.; Novo, L.; Verstappen, N.; Van Ranst, M.; Matthijnssens, J. Rotavirus incidence and genotype distribution before and after national rotavirus vaccine introduction in Belgium. *Vaccine* **2010**. [CrossRef]
19. Mwanga, M.J.; Owor, B.E.; Ochieng, J.B.; Ngama, M.H.; Ogwel, B.; Onyango, C.; Juma, J.; Njeru, R.; Gicheru, E.; Otieno, G.P.; et al. Rotavirus group A genotype circulation patterns across Kenya before and after nationwide vaccine introduction, 2010–2018. *BMC Infect. Dis.* **2020**, *20*, 504. [CrossRef]
20. Pitzer, V.E.; Bilcke, J.; Heylen, E.; Crawford, F.W.; Callens, M.; De Smet, F.; Van Ranst, M.; Zeller, M.; Matthijnssens, J. Did Large-Scale Vaccination Drive Changes in the Circulating Rotavirus Population in Belgium? *Sci. Rep.* **2015**, *5*, 1–14. [CrossRef]
21. Talbert, A.; Ngari, M.; Bauni, E.; Mwangome, M.; Mturi, N.; Otiende, M.; Maitland, K.; Walson, J.; Berkley, J.A. Mortality after inpatient treatment for diarrhea in children: A cohort study. *BMC Med.* **2019**. [CrossRef]
22. Ahmed, S.M.; Lopman, B.A.; Levy, K. A Systematic Review and Meta-Analysis of the Global Seasonality of Norovirus. *PLoS ONE* **2013**, *8*, e75922. [CrossRef]
23. Dey, S.K.; Phathammavong, O.; Nguyen, T.D.; Thongprachum, A.; Chan-It, W.; Okitsu, S.; Mizuguchi, M.; Ushijima, H. Seasonal pattern and genotype distribution of sapovirus infection in Japan, 2003–2009. *Epidemiol. Infect.* **2012**, *140*, 74–77. [CrossRef] [PubMed]

24. Omore, R.; Tate, J.E.; O'Reilly, C.E.; Ayers, T.; Williamson, J.; Moke, F.; Schilling, K.A.; Awuor, A.O.; Jaron, P.; Ochieng, J.B.; et al. Epidemiology, seasonality and factors associated with rotavirus infection among children with moderate-to-severe diarrhea in rural western Kenya, 2008–2012: The Global Enteric Multicenter Study (GEMS). *PLoS ONE* **2016**, *11*, e0160060. [CrossRef] [PubMed]
25. Vetter, M.R.; Staggemeier, R.; Vecchia, A.D.; Henzel, A.; Rigotto, C.; Spilki, F.R. Seasonal variation on the presence of adenoviruses in stools from non-diarrheic patients. *Braz. J. Microbiol.* **2015**, *46*, 749–752. [CrossRef] [PubMed]
26. Stadhouders, R.; Pas, S.D.; Anber, J.; Voermans, J.; Mes, T.H.; Schutten, M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. *Mol. Diagn.* **2010**, *12*, 109–117. [CrossRef]
27. Lefever, S.; Pattyn, F.; Hellemans, J.; Vandesompele, J. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clin. Chem.* **2013**, *59*, 1470–1480. [CrossRef]
28. Liu, J.; Gratz, J.; Amour, C.; Nshama, R.; Walongo, T.; Maro, A.; Mduma, E.; Platts-Mills, J.; Boisen, N.; Nataro, J.; et al. Optimization of quantitative PCR methods for enteropathogen detection. *PLoS ONE* **2016**, *11*, e0158199. [CrossRef]
29. Bennett, S.; Gunson, R.N. The development of a multiplex real-time RT-PCR for the detection of adenovirus, astrovirus, rotavirus and sapovirus from stool samples. *Virol. Methods* **2017**, *242*, 30–34. [CrossRef]
30. Van Maarseveen, N.M.; Wessels, E.; de Brouwer, C.S.; Vossen, A.C.; Claas, E.C. Diagnosis of viral gastroenteritis by simultaneous detection of Adenovirus group F, Astrovirus, Rotavirus group A, Norovirus genogroups I and II, and Sapovirus in two internally controlled multiplex real-time PCR assays. *Clin. Virol.* **2010**. [CrossRef]
31. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**. [CrossRef]
32. 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. Whole genome analyses of G1P[8] rotavirus strains from vaccinated and non-vaccinated South African children presenting with diarrhea. *Med. Virol.* **2015**, *87*, 79–101. [CrossRef]
33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: https://www.R-project.org/ (accessed on 22 October 2019).