



RESEARCH ARTICLE

Molecular characterization of rotavirus group A strains circulating prior to vaccine introduction in rural coastal Kenya, 2002-2013 [version 1; peer review: 2 approved, 1 approved with reservations]

Betty E. Owor^{1*}, Mike J. Mwangi ^{1*}, Regina Njeru¹, Robert Mugo¹, Mwanajuma Ngama¹, Grieven P. Otieno ¹, D.J. Nokes ^{1,2}, C.N. Agoti ^{1,3}

¹Epidemiology and Demography, KEMRI Wellcome Trust Research Program, Kilifi, Kilifi, 80108, Kenya

²School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology, Warwick University, Coventry, Coventry, CV4 7AL, Kenya

³School of Health and Human Sciences, Pwani University, Kilifi, Kilifi, 80108, Kenya

* Equal contributors

V1 First published: 28 Nov 2018, 3:150
<https://doi.org/10.12688/wellcomeopenres.14908.1>

Latest published: 15 May 2019, 3:150
<https://doi.org/10.12688/wellcomeopenres.14908.2>

Abstract

Background: Kenya introduced the monovalent Rotarix® rotavirus group A (RVA) vaccine nationally in mid-2014. Long-term surveillance data is important prior to wide-scale vaccine use to assess the impact on disease and to investigate the occurrence of heterotypic strains arising through immune selection. This report presents baseline data on RVA genotype circulation patterns and intra-genotype genetic diversity over a 7-year period in the pre-vaccine era in Kilifi, Kenya, from 2002 to 2004 and from 2010 to 2013.

Methods: A total of 745 RVA strains identified in children admitted with acute gastroenteritis to a referral hospital in Coastal Kenya, were sequenced using the di-deoxy sequencing method in the VP4 and VP7 genomic segments (encoding P and G proteins, respectively). Sequencing successfully generated 569 (76%) and 572 (77%) consensus sequences for the VP4 and VP7 genes respectively. G and P genotypes were determined by use of BLAST and the online RotaC v2 RVA classification tool.

Results: The most common GP combination was G1P[8] (51%), similar to the Rotarix® strain, followed by G9P[8] (15%), G8P[4] (14%) and G2P[4] (5%). Unusual GP combinations—G1P[4], G2P[8], G3P[4,6], G8P[8,14], and G12P[4,6,8]—were observed at frequencies of <5%. Phylogenetic analysis showed that the infections were caused by both locally persistent strains as evidenced by divergence of local strains occurring over multiple seasons from the global ones, and newly

Open Peer Review

Reviewer Status   

Invited Reviewers

	1	2	3
version 2 (revision) 15 May 2019			
version 1 28 Nov 2018	 report	 report	 report

- Ernest Wandera Apondi**, KEMRI/Nagasaki University, Nairobi, Kenya
- Mathew D Esona**, Centers for Disease Control and Prevention, Atlanta, USA
- Martin Munene Nyaga**, University of the Free State, Bloemfontein, South Africa

Any reports and responses or comments on the article can be found at the end of the article.

introduced strains, which were closely related to global strains. The circulating RVA diversity showed temporal fluctuations both season by season and over the longer-term. None of the unusual strains increased in frequency over the observation period.

Conclusions: The circulating RVA diversity showed temporal fluctuations with several unusual strains recorded, which rarely caused major outbreaks. These data will be useful in interpreting genotype patterns observed in the region during the vaccine era.

Keywords

Rotavirus, epidemiology, genotype, strains, diversity



This article is included in the [KEMRI | Wellcome Trust gateway](#).

Corresponding author: Mike J. Mwanga (mmwanga@kemri-wellcome.org)

Author roles: **Owor BE:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Review & Editing; **Mwanga MJ:** Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation; **Njeru R:** Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Mugo R:** Formal Analysis, Investigation, Methodology, Software, Writing – Review & Editing; **Ngama M:** Formal Analysis, Investigation, Methodology, Supervision, Writing – Review & Editing; **Otieno GP:** Data Curation, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – Review & Editing; **Nokes DJ:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Review & Editing; **Agoti CN:** Formal Analysis, Investigation, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work was funded by GAVI (under the Rotavirus Immunization Programme Evaluation in Kenya (RIPEK) study; a collaboration between Emory University (USA), Centers for Disease Control and Prevention (CDC) in Atlanta USA, the Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, and the KEMRI Centre for Global Health Research (CGHR), Kisumu, Kenya). This work was also supported by the Wellcome Trust (203077; 102975). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2018 Owor BE *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Owor BE, Mwanga MJ, Njeru R *et al.* **Molecular characterization of rotavirus group A strains circulating prior to vaccine introduction in rural coastal Kenya, 2002-2013 [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2018, 3:150 <https://doi.org/10.12688/wellcomeopenres.14908.1>

First published: 28 Nov 2018, 3:150 <https://doi.org/10.12688/wellcomeopenres.14908.1>

Introduction

Rotavirus group A (RVA) infection is a leading cause of childhood severe dehydrating acute diarrhoea, which can lead to death¹. The 2016 estimates show that, annually, RVA is responsible for 128,500 deaths globally, with the highest burden occurring in sub-Saharan Africa and South-East Asia countries². In 2009, the World Health Organization (WHO) recommended the inclusion of either of the two licensed RVA vaccines (Rotarix® and RotaTeq®) into national immunization programmes (NIPs) of all countries to curb RVA associated disease burden³. Kenya introduced the monovalent Rotarix® vaccine (based on the G1P[8] strain) into its NIP in July 2014⁴.

In Africa, the introduction of the Rotarix® vaccine into the NIPs of several countries has been associated with a marked reduction in hospitalization caused by RVA infection^{5,6}. For instance in Malawi, Burkina Faso and Tanzania, the vaccine effectiveness against hospitalization was estimated at 62%, 58% and 53%, respectively⁷. However, this effectiveness is lower than that observed in developed countries; for example in Belgium, vaccine effectiveness of Rotarix vaccine was estimated at 90%⁸. Furthermore, concerns remain that in time, given the high diversity of RVA strains, vaccine immunity escape variants could emerge which may undermine the gains from the vaccination programmes⁹. Such a scenario was observed in Japan where, a G8P[8] RVA strain appeared to emerge and caused acute gastroenteritis disease in up to 66% (53/80) of children attending a pediatric clinic¹⁰. Similarly, the predominance of non-vaccine type G2P[4] strains was observed in Rotarix® vaccinated populations of Belgium¹¹ and Brazil¹², raising concerns of the effect of the vaccine on circulating non-vaccine strains.

The rotavirus genome is comprised of 11 segments of double-stranded RNA, which encode 12 proteins (VP1-4, VP6, VP7, NSP1-6). The VP7 and VP4 proteins independently elicit neutralizing antibodies and specify the G (glycoprotein) and P (protease-sensitive) genotypes, respectively¹³. Molecular characterization of the VP7 and VP4 proteins encoding regions is commonly used to investigate local and global RVA molecular epidemiology and is the basis of the dual genotype classification of this virus¹⁴. Up to 36 different G and 51 P RVA genotypes have been identified worldwide in humans and animals¹⁵. Globally, G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] (in decreasing order) have been identified as the most common genotypes causing disease in children, although their distribution can vary considerably from region to region and from one season to the next^{16,17}. While these genotypes are similarly dominant in Africa¹⁸, understanding of their local natural seasonal fluctuations and intra-genotype diversity in the pre-vaccine introduction era is incomplete despite importance to vaccine impact evaluation.

The current study presents molecular analysis of historical RVA strains from coastal Kenya detected between 2002–2004, reported in Nokes *et al.*¹⁹, which we refer to as phase I, together with more recent RVA strains detected between 2010–2013, referred to as phase II. We present findings from partial sequence analysis of these longitudinally collected RVA strains identified

at the Kilifi County Hospital (KCH), Kilifi, Kenya, and phylogenetically compare these with those deposited in public databases across the globe. The typing of phase I strains was previously performed by nested multiplex PCR using genotype-specific VP7 and VP4 primers¹⁹. We utilize these extensive sequence data to illuminate on local RVA genotype circulation characteristics and provide baseline information on natural patterns of RVA genotype diversity in coastal Kenya prior to vaccine introduction.

Methods

RVA surveillance in Kilifi County Hospital

RVA surveillance in KCH reported in this analysis was conducted from January 2002 to December 2004, and from January 2010 to December 2013. Study subject recruitment criteria and sample collection methods are as previously described¹⁹. The study targeted children aged less than 13 years admitted with acute diarrhoea defined as three or more watery stools passed during a 24-hour period²⁰. The KEMRI Scientific and Ethics Review Unit (SERU) in Kenya approved the study protocol.

Detection of RVA

Stool samples were screened for RVA using an enzyme immunoassay (EIA) kit, marketed under two different names in the two periods: IDEIA (DAKO Rotavirus IDEIA™, Oxoid, Ely, United Kingdom) in phase I and ProSpect™ (Oxoid, Basingstoke UK) in phase II, following the manufacturer's instructions.

Partial sequencing of RVA positive samples in VP4 and VP7 segments

Sequencing was conducted on 272 (46%) of 558 positive samples detected in phase I, and all positive samples identified in phase II (n=473). The phase I samples were selected to represent common RVA genotypes (>70%) observed throughout the surveillance period from each year. Partial fragments of the VP4 and VP7 genes, were amplified in a one-step reverse transcriptase PCR reaction using the following primer pairs: VP4F, 5'-TATGCTCCAGTNAATTGG-3', VP4R 5'-ATTGCATTCTTTC-CATAATG-3', VP7F, 5'-ATGTATGGTATTGAATATACCAC-3', VP7R 5'—AACTTGCCACCATTTTTTCC-3', as previously described by Simmonds *et al.*²¹ and Gomara *et al.*²². To confirm successful amplification of the targeted genomic area the products were checked (VP7, 881 bp; VP4, 660 bp) by electrophoresis in a 2% agarose gel. Products of samples that showed presence of the expected band size on gels were purified using GFX DNA purification kit (GFX-Amersham, UK) following the manufacturer's instructions. These were then sequenced using Big Dye Terminator 3.1 (Applied Biosystems, Foster City, California, USA) chemistry and the same primers as in PCR amplification on an ABI Prism 3130xl Genetic Analyser (Applied Biosystems, Foster City, California, USA).

RVA genotyping and sequence analysis

The sequence reads were assembled into contigs using Sequencher version 5.4.6 (Gene Codes Corp Inc., Ann Arbor, MI, USA). The nucleotide sequences were aligned using MAFFT version 7.222²³ and visualized in Aliview version 1.8 and further trimmed to remove sequence overhangs, resulting in

contigs of lengths between 480–660 bp (coordinates; 184–748) covering ~23% of the VP4 gene, and 486–854 bp (coordinates; 460–824 of the VP7 gene) covering ~67% of the VP7 gene. G and P genotypes were determined using **NCBI BLAST** for sequences <500 bp (n=13 for VP4, n=5 for VP7) and the **RotaC** version 2.0 classification tool²⁴ for sequences >500 bp. **MEGA** v7.0.26 was used to select the best maximum likelihood evolution models based on the Bayesian Information Criterion²⁵ (**Supplementary Table 1**) and reconstruction of maximum likelihood phylogenetic trees with 500 bootstrap replicates. Global contemporaneous sequences (2002–2013) (accession numbers in **Supplementary File 1**, lists 1 and 2) together with the Rotarix® vaccine strain sequences were retrieved from GenBank database and phylogenetically compared with the local sequences. Duplicate sequences from the same country and non-overlapping sequences were removed. Clusters were identified based on high bootstrap values of >70% and high nt sequence similarity of >98%. Nucleotide and amino acid pairwise distances between the sequences were determined in **MEGA** v7.0.26. The trees were drawn to scale indicating nucleotide substitution rates per site.

Results

The prevalence of the genotypes and the circulation patterns described in this study were determined by use of all the data collected in 2002–2004 (phase I) surveillance period and all data collected between 2010–2013 (phase II). Data are available under restriction on Harvard Dataverse⁴⁰.

RVA prevalence in KCH pediatric diarrhoea admissions

Over the 7-year surveillance period, a total of 3,779 stool samples were screened for RVA using EIA, of which 27.3% (n=1,031) tested positive. In phase I, the prevalence of RVA in the study population was 27.4% (n=558) while in phase II the prevalence was 27.2% (n=473) (**Table 1**). Sequencing was successful for 569 (76%) and 572 (77%) samples for the VP4 and VP7 segments respectively (**Table 1**).

RVA genotypes in the study populations

The G genotypes identified in patients admitted at the KCH were G1–G3, G8–G10, G12, G29, while the P genotypes were P[4], P[6], P[8] and P[14] genotypes. Overall, G1P[8] was the dominant strain at 51% followed by G9P[8] (15%), G8P[4] (14%) and G2P[4] (5%) as shown in **Table 2** and **Supplementary Figure 1**. Strains G1P[4], G2P[8], G3P[4,6,8], G8P[6,8,14], G9P[4,6], G12P[4,6,8] were also detected, albeit in low frequency (<5%).

The previous use of genotype-specific primers (rather than sequencing) to identify RVA strains in phase I enabled the detection of mixed infections in 8.2% of the cases; however, mixed infections were not seen when using sequence based methods. Additionally, 9.2% of the samples were typed for only one of the two genes due to failure in sequencing and/or contig assembly (**Table 2**). G1P[8] predominated in all the years of phase I, while in phase II, this strain was dominant only in 2011 and 2013 (**Figure 1A, B**). Whilst strain G8P[4] was observed in low frequency in the whole of phase I, it was observed as the most common strain in 2010 (46%) and 2012 (40%). Strain G9P[8] circulated in moderate proportions, and was observed in all the years except in 2013. Strain G2P[4] which was observed in low frequency in phase I (1.1%), was seen to increase in proportions in years 2010 (12%) and 2012 (23%). The rare strains, G1P[6], G2P[8], G8P[6] and G9P[6] were only observed in phase I, while the rare strains G3P[4,6], G12P[4,6,8], G10P[8] and G8P[14] were observed only in phase II. None of the rare strains observed in phase I became common in phase II.

Genetic relationship between RVA strains

Figure 2 shows the temporal frequency of genotypes G1, G2, G8, G9, P[4] and P[8] with their corresponding phylogenetic trees, while nucleotide pairwise difference within each genotype is shown in **Supplementary Figure 2**. The time period is split into 2002–2004, 2010–2011 and 2012–13 (shown by different colours) to facilitate temporal comparison. The G1 strains which were observed in all the years, formed clusters containing strains from both phase I and II, showing an overall sequence homology of >92% at the nucleotide level. Additionally, minor distinct clusters containing strains observed in phase II were also observed. The occurrence of G8 strains fluctuated with high prevalence observed in 2002, 2004 and 2010, and less prevalent in 2003 and 2009. Majority of the G8 strains showed high sequence homology of 96–100%, forming a common cluster including strains from both phases. However, a single sequence showed a decrease in homology up to 84% at nt level, and formed distinct clusters. The infrequently occurring G2 strains, formed two distinct clusters, where one cluster contained strains observed in phase II while the other had strains from both phases. Nevertheless, a high sequence homology of >95% at the nucleotide level was observed within G2 strains. Such high sequence homology was also observed in G9 strains, which were observed in high frequencies in all epidemic years except 2013.

Table 1. A summary of diarrhoea cases, the number of samples tested, the proportion of RVA cases observed in the entire surveillance period and the number of samples sequenced from each phase from childhood admissions to KCH, Kenya, between 2002–2004 and 2010–2013.

Period	Admissions, n	Diarrhoea, n	Samples tested, n	RVA cases, n	Proportion, %	Sequenced, n (%)	Successfully assembled, n (%)	
							VP4	VP7
2002–2004 (Phase I)	15347	3296	2039	558	27.2	272 (48)	192 (71)	218 (80)
2010–2013 (Phase II)	11579	2260	1740	473	27.4	473 (100)	377 (80)	354 (75)
Total	26926	5556	3779	1031	27.3	745 (72%)	569 (76)	572 (77)

Table 2. Frequency and proportions of RVA strains observed in Kilifi County Hospital between 2002–2004 and 2010–2013.

	Type	2002		2003		2004		2010		2011		2012		2013		Total	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
GP Genotypes	G1P[8]	90	(59)	79	(59)	63	(38)	19	(21)	101	(85)	16	(21)	28	(93)	396	51
	G9P[8]	23	(15)	36	(27)	25	(15)	12	(13)	12	(10)	10	(13)	0	(0)	118	15
	G8P[4]	3	(2)	7	(5)	28	(17)	42	(46)	0	(0)	30	(40)	0	(0)	110	14
	G2P[4]	1	(1)	4	(3)	0	(0)	11	(12)	1	(1)	17	(23)	1	(3)	35	5
	G8P[6]	17	(11)	1	(1)	15	(9)	0	(0)	0	(0)	0	(0)	0	(0)	33	4
	G8P[8]	8	(5)	1	(1)	13	(8)	0	(0)	0	(0)	2	(3)	0	(0)	24	3
	G1P[4]	0	(0)	0	(0)	10	(6)	3	(3)	0	(0)	0	(0)	0	(0)	13	2
	G9P[6]	3	(2)	4	(3)	5	(3)	0	(0)	0	(0)	0	(0)	0	(0)	12	2
	G1P[6]	3	(2)	0	(0)	6	(4)	0	(0)	0	(0)	0	(0)	0	(0)	9	1
	G9P[4]	0	(0)	2	(1)	1	(1)	2	(2)	0	(0)	0	(0)	0	(0)	5	1
	G2P[8]	4	(3)	1	(1)	1	(1)	0	(0)	0	(0)	0	(0)	0	(0)	6	1
	G3P[8]	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(3)	1	0
	G12P[8]	0	(0)	0	(0)	0	(0)	1	(1)	3	(3)	0	(0)	0	(0)	4	1
	G12P[4]	0	(0)	0	(0)	0	(0)	1	(1)	1	(1)	0	(0)	0	(0)	2	0
	G3P[4]	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	0
	G12P[6]	0	(0)	0	(0)	0	(0)	0	(0)	1	(1)	0	(0)	0	(0)	1	0
	G3P[6]	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	0
	G10P[8]	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	0
	G8P[14]	0	(0)	0	(0)	0	(0)	1	(1)	0	(0)	0	(0)	0	(0)	1	0
	Total	152	(100)	135	(100)	167	(100)	92	(100)	119	(100)	75	(100)	30	(100)	770	100
Mixed & Non Typable	Mixed	7	(27)	24	(71)	26	(59)	0	(0)	0	(0)	0	(0)	0	(0)	57	55
	GNTP[4]	0	(0)	0	(0)	6	(14)	0	(0)	0	(0)	0	(0)	0	(0)	6	6
	GNTP[6]	2	(8)	0	(0)	3	(7)	0	(0)	0	(0)	0	(0)	0	(0)	5	5
	GNTP[8]	8	(31)	0	(0)	1	(2)	0	(0)	0	(0)	0	(0)	0	(0)	9	9
	G1PNT	5	(19)	5	(15)	2	(5)	0	(0)	0	(0)	0	(0)	0	(0)	12	12
	G8PNT	1	(4)	1	(3)	1	(2)	0	(0)	0	(0)	0	(0)	0	(0)	3	3
	G9PNT	2	(8)	2	(6)	2	(5)	0	(0)	0	(0)	0	(0)	0	(0)	6	6
	GNTPNT	1	(4)	2	(6)	3	(7)	0	(0)	0	(0)	0	(0)	0	(0)	6	6
	Total	26	100	34	100	44	100	0	0	0	0	0	0	0	0	104	100
	G1P[x]	0	(0)	0	(0)	0	(0)	0	(0)	7	(27)	2	(13)	7	(22)	16	16
Failed sequencing	G3P[x]	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	3	(9)	3	3
	G8P[x]	0	(0)	0	(0)	0	(0)	7	(26)	0	(0)	1	(6)	0	(0)	8	8
	G9P[x]	0	(0)	0	(0)	0	(0)	0	(0)	2	(8)	1	(6)	0	(0)	3	3
	G12P[x]	0	(0)	0	(0)	0	(0)	3	(11)	2	(8)	2	(13)	0	(0)	7	7
	G29P[x]	0	(0)	0	(0)	0	(0)	1	(4)	0	(0)	0	(0)	0	(0)	1	1
	GxP[8]	0	(0)	0	(0)	0	(0)	7	(26)	15	(58)	5	(31)	19	(59)	46	46
	GxP[4]	0	(0)	0	(0)	0	(0)	9	(33)	0	(0)	5	(31)	3	(9)	17	17
	Total	0	0	0	0	0	0	27	100	26	100	16	100	32	100	101	100

NT, non-typable; Gx and Px, undetermined G and P genotypes, respectively.

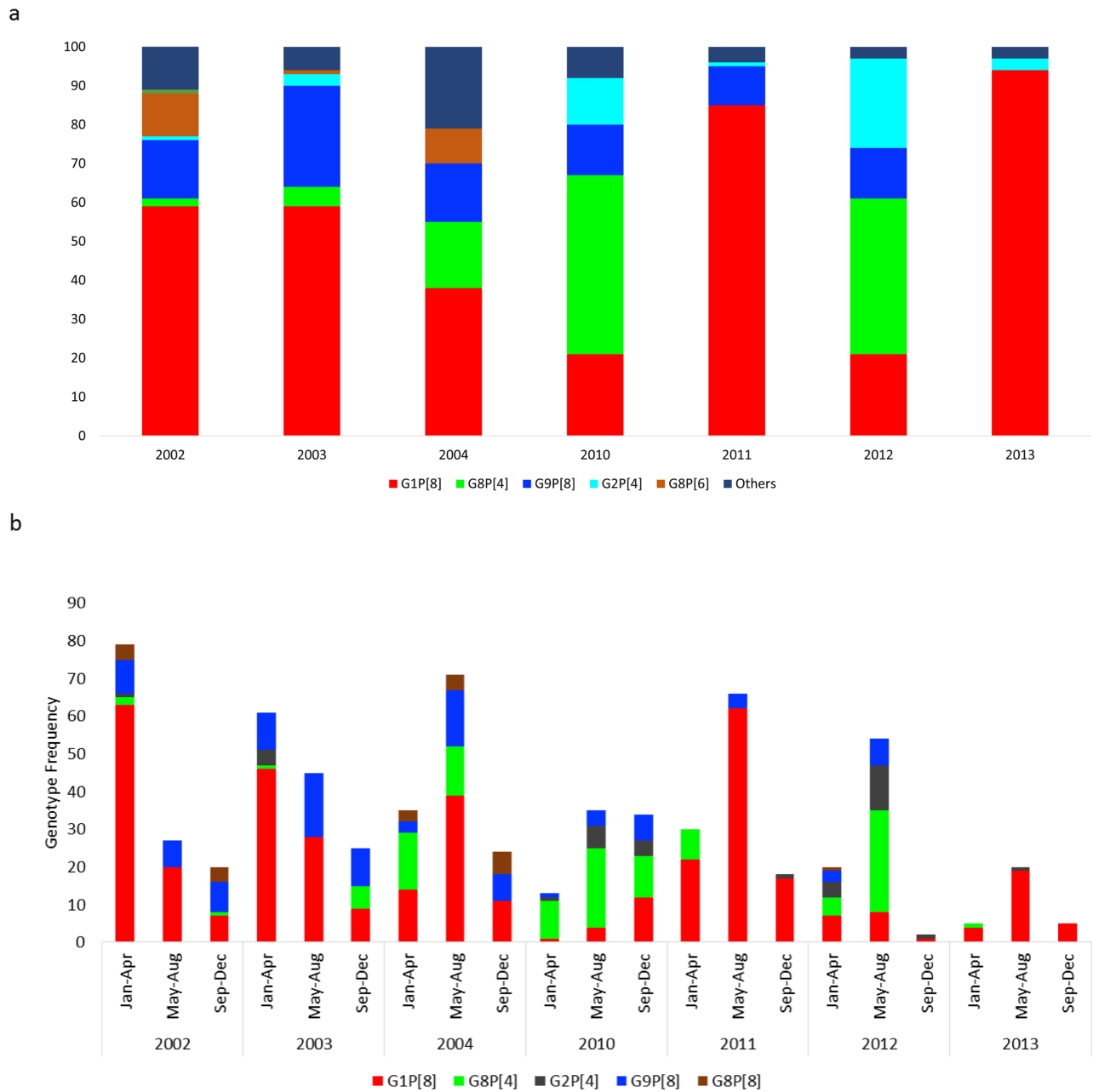


Figure 1. Temporal distribution of RVA genotypes from positive RVA cases isolated from Kilifi County Hospital from 2002–2004 and 2010–2013. (A) Annual proportions of the common genotypes. **(B)** The 4-month frequency of the commonly occurring genotypes. The colors represent the GP combinations as indicated on the legend of the plots. Genotypes that fall in the “Others” category in **(A)** represents those that occurred in low frequency <5%: G8P[8, 14], G1P[4, 6], G9P[4, 6], G2P[8], G12P[4, 6, 8], G3P[4, 6], G10P[8].

Phylogenetically, the P[8] strains showed a close association among themselves with high sequence homology of >92% at nt level. Despite, the high homology, majority of the P[8] strains observed in phase I formed separate clusters from those observed in phase II. Unlike P[8] strains, P[4] strains occurred less frequently, with high prevalence observed in 2004 in phase I and 2010 and 2012 in phase II. These strains formed three clusters, with one cluster containing both phase I and II strains while the

other two clusters containing only phase II strains. Despite the distinct clustering, P[4] strains showed a high sequence similarity of 95–100% at the nucleotide level.

Phylogenetic placement of Kilifi strains in the global context

The placement of Kilifi strains in the global context is shown in [Figure 3](#) and [Figure 4](#). Tree clusters leading to Kilifi strains are

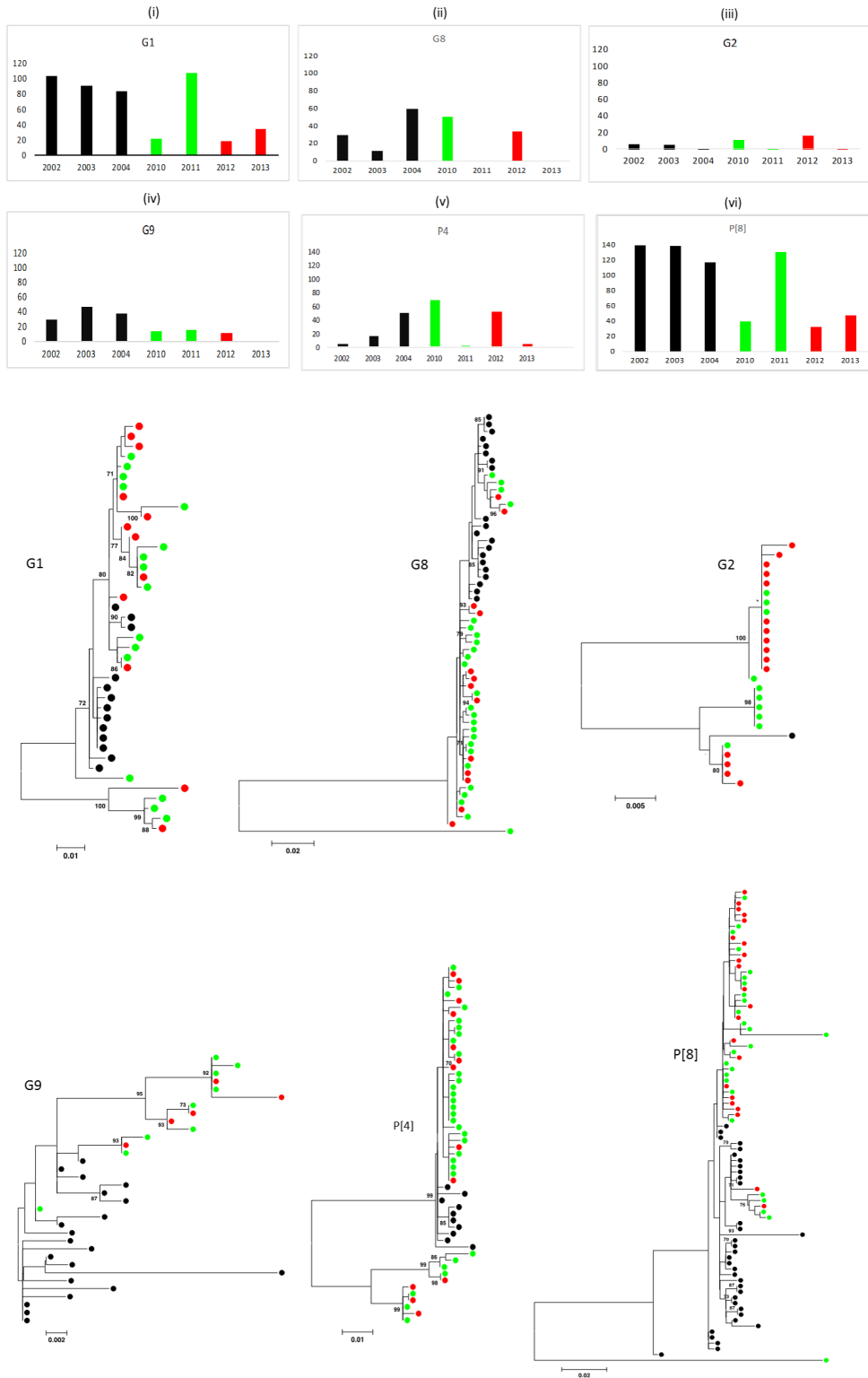


Figure 2. Maximum likelihood phylogenetic trees of G1, G2, G8, G9, P[4] and P[8] genotypes inferred in MEGA v7, with taxa stratified in 3 groups, black for 2002–2004, green for 2010–2011 and red for 2012–2013, from viruses detected in childhood diarrhea admissions to KCH, Kenya. The bar graphs represent frequency of the same genotypes between 2002–2004 and 2010–2013. Only bootstrap values $\geq 70\%$ are shown. This figure excludes the infrequent genotypes G3, G10, G12, G29 and P[6]. The scale bars indicate nucleotide substitutions per site.

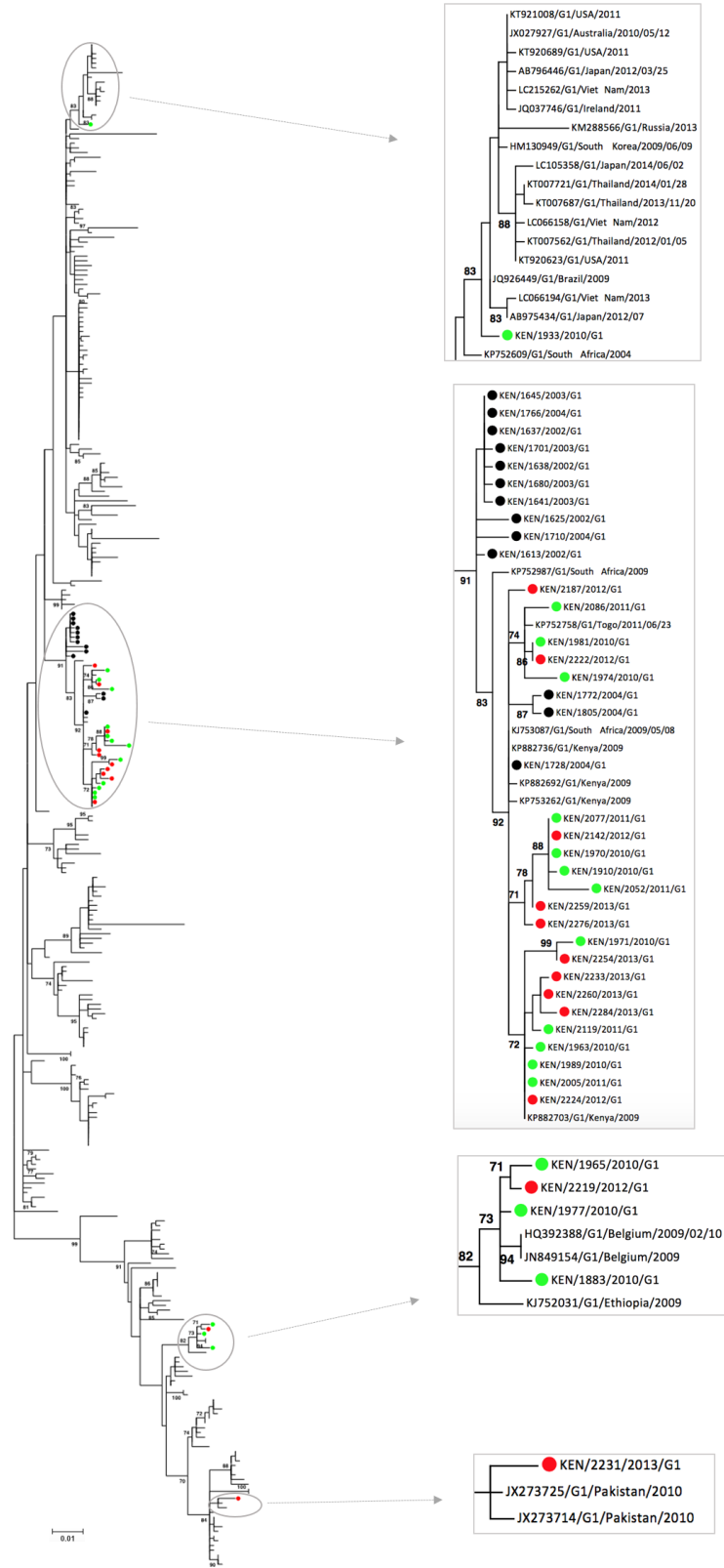


Figure 3. Maximum likelihood tree for VP7 G1 genotype showing the relationship between G1 genotypes from viruses detected in childhood admissions to KCH, Kenya, and to other global G1 genotypes detected between 2002 and 2013. Tree clusters (branches) including Kilifi strains are shown in the expanded boxes. Taxa for Kilifi strains are stratified in three groups, black for 2002–2004, green for 2010–2011 and red for 2012–2013. Only bootstrap values $\geq 70\%$ are shown. The scale bars indicate nucleotide substitution per site.

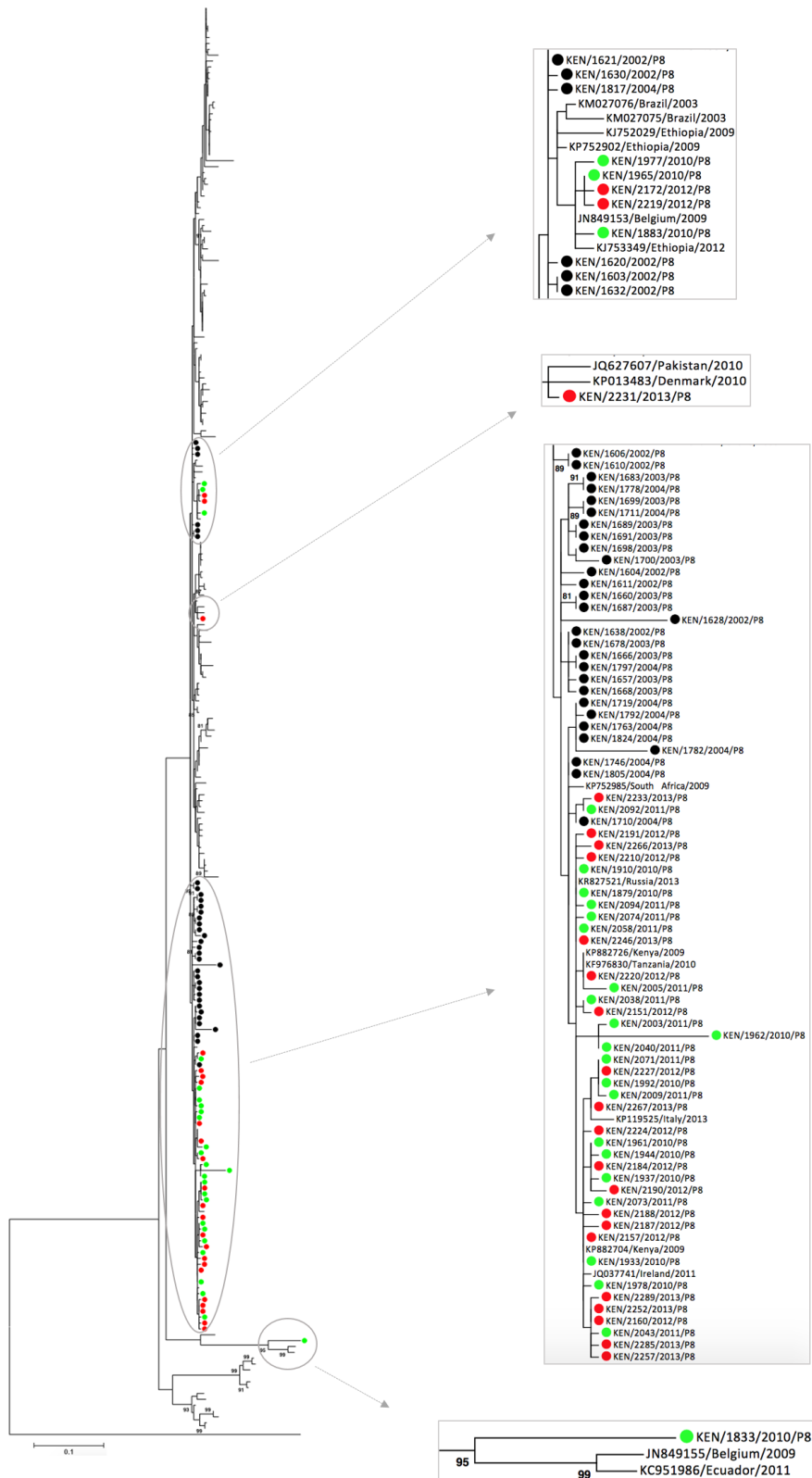


Figure 4. Maximum likelihood tree for VP4 P[8] genotype showing the relationship between P[8] genotypes detected in childhood admissions to KCH, Kenya, and to other global P[8] genotypes detected between 2002–2013. Tree clusters (branches) including Kilifi strains are shown in the expanded boxes. Taxa for Kilifi strains are stratified in three groups, black for 2002–2004, green for 2010–2011 and red for 2012–2013. Only bootstrap values $\geq 70\%$ are shown. Scale bar represents nucleotide substitutions per site.

shown in the expanded boxes. A majority (85%) of the observed G1 strains, clustered away from the other global strains, clustering closely to strains detected in Africa, specifically in Kenya, South Africa and Togo. The second cluster comprised only strains from 2010–2012 which distinctively clustered with strains from Belgium and Ethiopia. The last clusters which had only single strains grouped together with strains from Japan and Pakistan. The Kilifi P[8], (Figure 5) strains were placed into four clusters, where the largest group comprised of Kilifi strains observed in both phase I and II, with external strains observed in Kenya, S. Africa, Tanzania, Ireland and Russia. The second cluster included Kilifi strains from phase I and II with strains from Belgium, Brazil and Ethiopia. The last minor clusters, each made of a single virus, showed a close similarity to strains isolated in Pakistan, Denmark, Ecuador and Belgium.

Phylogenetic analysis of the rare G8P[14] strain

Whilst several rare GP combinations (not commonly detected) were observed during the study period, G8 associated with a P[14] genotype has overall been rarely detected in human population²⁶. Additionally, there has been an increasing number of human P[14] rotavirus strains globally, which are associated with rabbits, cattle, sheep and guanacos²⁷. We therefore sought to investigate the probable origin of the observed G8P[14] detected in a 14 months old infant in 2010. All cognate sequences for genotypes G8 (n=71) and P[14] (n=47) isolated by 2013 were retrieved from GenBank and phylogenetically compared to the observed genotypes. Duplicate sequences from strains isolated from the same country were removed. The G type in this samples (G8) clustered closely to other G8 strains isolated from humans with a nucleotide and amino acid (aa) identity of 95% and 99%, respectively, and G8 strains isolated from camel showing a nucleotide and aa identity of 94% and 98%, respectively (Figure 5A). The P[14] genotype showed a high sequence similarity to other P[14] strains isolated from humans and bovine with a nucleotide similarity of 96% and 93% respectively and aa identity of 98% (Figure 5B).

Discussion

The present study provides insight into the molecular epidemiology and phylogenetic relatedness of distant (7 years or more) and recent pre-vaccine introduction RVA strains detected in Kilifi, Coastal Kenya. This is the first detailed study on prevalence of RVA genotypes causing diarrhoea in children in rural coastal Kenya spanning over a decade before introduction of the nationwide routine RVA vaccination programme. The work builds on a previous study¹⁹, which highlighted the importance of genotypes G1, G8 and G9 in sub-Saharan Africa during the pre-vaccine introduction period. During phase I surveillance period, genotype-specific primers were used to characterize the strains into different G and P genotypes. In the present analysis, a fraction of phase I (46%) and all phase II RVA samples were sequenced and GP annotations assigned as per the guidelines of the Rotavirus Classification Working Group²⁸.

In this analysis, strains G1P[8], G9P[8], G8P[4] and G2P[4] were the most common RVA strains, accounting for over 70% of the infections. These strains have also been observed in

studies conducted elsewhere in Kenya^{29–31} and the world^{32,33}. Genotype G8P[4] was the third most important strain after G1P[8] and G9P[8] accounting for 15% of RVA infections. The G8 genotype is mostly found in combination with P[4], P[6] and P[8] VP4 specificities³⁴. In this study, the majority (83%) of the G8 strains combined with P[4] types, while only 16% combined with P[8] types. The increase in prevalence of this strain in phase II supports the notion of G8 strain regarded as an unusual and newly emerging strain in the world^{29,31}. Genotype G3P[8] is also among the common genotypes causing infections in children, and is the second most important strain in Africa and fifth most important globally³². Here, G3P[8] was detected at a low frequency, accounting for only 1% of all the cases. Genotype G12 detection has increased in Africa and has also been observed in Kenya and for the first time in Kilifi (2010–2013).

The detection of atypical GP combinations; G1P[4, 6], G2P[6], G3P[4,6], G10P[8], and G8P[14], albeit at low levels, raises interest in their origins. Despite such atypical strains being less frequent, strains G3P[4] and G2P[6] were found to be the most important causes of diarrhoea in the late 1990s in Ghana³⁵. Genotype G10 has long been reported to infect calves, pigs or cattle but recently has sporadically been reported in humans in several studies^{36,37}. Similarly strain G8P[14] has recently been detected in humans and is thought to have originated from animals^{26,38}. In this study, the close association of the observed strain G8P[14] with strains from both humans and animal origins shows a possibility of zoonotic transmission. The increase in diversity of RVA in this setting could be attributed to the emergence of such unusual strains which might have arisen due to zoonotic transmission or re-assortment cases within and between RVA genotypes.

Post-vaccine surveillance studies have reported shifts in the prevalence of RVA genotypes. Recent post-vaccine introduction reports from Kenya have indicated an increase in prevalence of uncommon genotypes G3, G9 and G12³⁹. Similarly, data from the USA depicted an increase in prevalence of G3P[8] in post-vaccine era relative to G1P[8] in the pre-vaccine period. In contrast, surveillance studies in Australia and Belgium revealed the dominance of G2 strains in post-vaccine period, relative to G1P[8] in pre-vaccine period. Similar studies in Ghana reported an increase in prevalence of G12P[8] and G10P[6] in the post-vaccine era. This shift in distribution of genotypes post-vaccine introduction might be associated with either selective vaccine pressure or the natural fluctuations of RVAs, although these are not evidently supported. The emergence of uncommon genotypes and increased prevalence of non-vaccine strains warrants close monitoring to determine their circulation in the post-vaccine introduction period and their probable effect on performance of the vaccine.

Overall, the observed strains showed a high nucleotide sequence homology of up to 100%, as observed in the different genotypes. The close genetic relationship of strains observed in phase I and phase II suggest a persistence in circulation of these RVA strains to continuously cause the observed epidemics. In addition, the exclusive clustering of majority of Kilifi strains

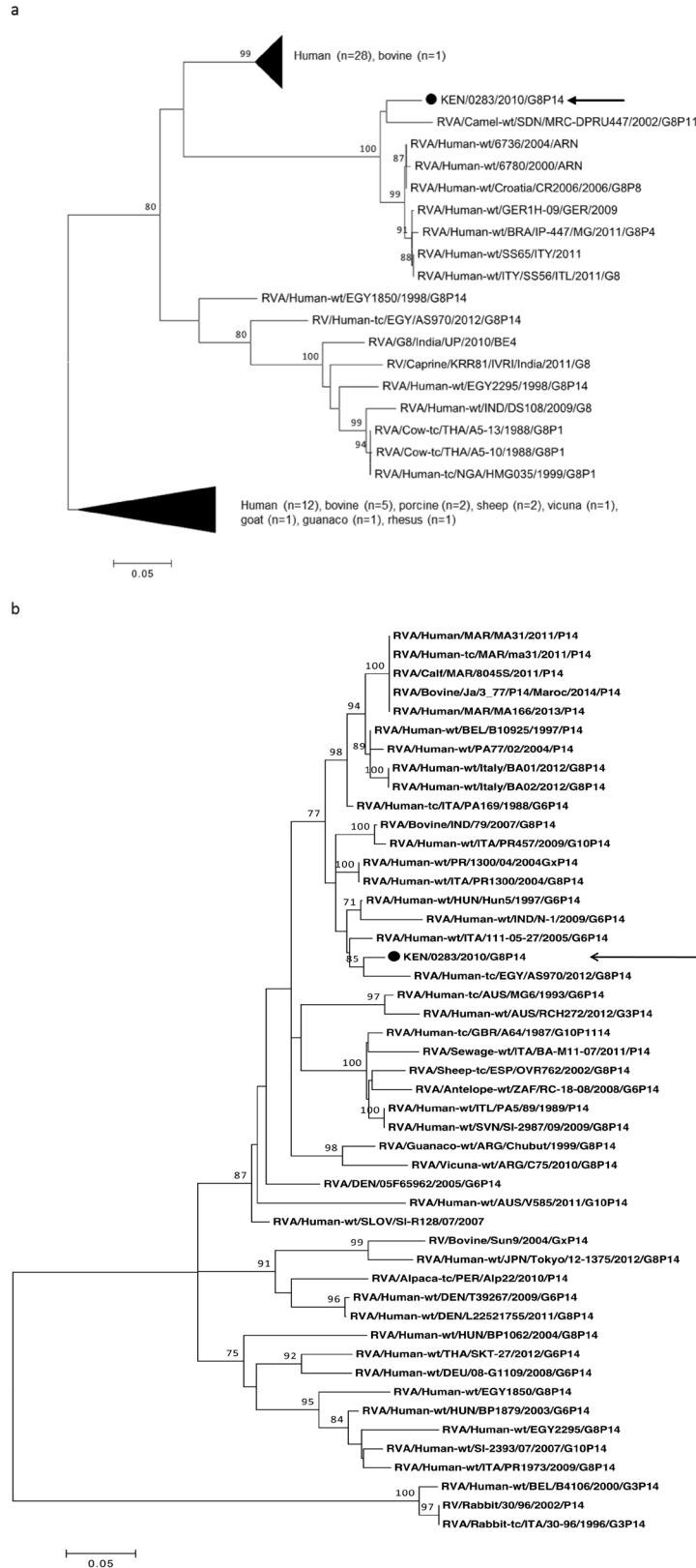


Figure 5. Maximum likelihood phylogenetic tree showing the relationship of the rare G8P[14] strain detected in a child admitted to KCH, Kenya, and, other similar strains detected in humans and animals retrieved from GenBank. (A) shows the phylogenetic relationship of the VP7 G8 genotype to other G8 genotypes. (B) Phylogenetic relationship of the VP4 P[14] genotype to other P[14] genotypes. Sequences for the strains identified in this study are marked by the black filled circle and the arrows. Only bootstrap values $\geq 70\%$ are shown. Scale bar represents nucleotide substitution per site.

from the global strains shows that these strains might have been localized in Kilifi over a long period of time. However, few strains that formed three distinct clusters in both G1 and P[8] global trees, supports the notion of separate introductions and persistence of possibly foreign strains in this setting. Although cases of re-assortment and possible introductions is evident, partial data from only two genes is insufficient in providing a complete understanding of the genetic diversity of such common and not common genotypes. Full genome sequencing will thus illuminate on the complete genomic constellations of these strains and provide data on their evolutionary dynamics. The marked seasonal and longer-term changes in genotype distribution observed in this pre-vaccine surveillance should be considered when interpreting changes to genotype patterns that may follow the introduction of rotavirus vaccine in any setting.

This study had several limitations, e.g. firstly, by use of partial sequencing method, we were unable to identify mixed infections in phase II, which were previously identified in phase I using other primer-based methods. Partial sequencing only identifies the dominant genotype in mixed infections resulting to one genotype. The sequencing chromatograms of samples identified as mixed infections in phase I, appeared clean and mono-infected, with no background indicators of co-infections. Secondly, the classification of the strains into lineages and sub-lineages was limited due to the short consensus sequences, since only ~23% and ~67% of the VP4 and VP7 genes were sequenced, respectively. Thirdly, it was not possible to perform comparative analysis of the rare genotype G29 due to unavailability of cognate sequences in GenBank. Only a single reference sequence for genotype G29 had been deposited in GenBank by the time of this analysis.

In conclusion, this study shows that most of the pre-vaccine RVA infections and epidemics have been caused by a diverse range of RVA strains which fluctuated in prevalence from season to season, with some persistent in circulation for a long period. Additionally, new strains might have been introduced in this population and contributed significantly to the epidemics experienced in the pre-vaccine period. The recommendation by WHO for countries to vaccinate infants against rotavirus infection led to the inclusion of Rotarix™ vaccine in the childhood immunization programme in Kenya. In addition to reducing hospitalization caused by RVA diarrhoea, the vaccine has been reported to offer protection against both homotypic and heterotypic RVA strains⁹. With the increase in diversity of circulating strains and emergence of rare strains in Kilifi, continuous monitoring will help evaluate the performance of this vaccine against the circulating strains.

Data availability

The replication data and analysis data for this manuscript are available from the Harvard Dataverse: <https://doi.org/10.7910/DVN/LVGYW40>.

Owing to data personal protection concerns, these data are restricted, but will be made available to researchers who meet the criteria for access to confidential data. Details of the criteria for sharing data and the conditions under which data are made available can be found in the KEMRI-Wellcome [data sharing guidelines](#). Users who wish to use the data should send a request to the KEMRI Wellcome Trust Research Programme data governance committee, which can be contacted by emailing: dgc@kemri-wellcome.org.

Nucleotide sequence accession numbers

Partial sequences for the VP7 and VP4 genes reported in this work were deposited in the GenBank database under the sequential accession numbers MH402005-MH402781 and MH402782-MH403560 for the VP7 and VP4 genes, respectively.

Grant information

The work was funded by GAVI (under the Rotavirus Immunization Programme Evaluation in Kenya (RIPEK) study; a collaboration between Emory University (USA), Centers for Disease Control and Prevention (CDC) in Atlanta USA, the Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, and the KEMRI Centre for Global Health Research (CGHR), Kisumu, Kenya). This work was also supported by the Wellcome Trust (203077; 102975).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank all study participants for their contribution of samples and data, and the KEMRI Wellcome Trust Research Programme colleagues for the useful discussions in preparation of the manuscript. We thank the laboratory staff at KEMRI Wellcome Trust, [Virus Epidemiology and Control Research Group](#) for processing of samples and initial lab analysis. We are grateful to Emory University which supported the work as part of the Rotavirus Immunization Program Evaluation in Kenya (RIPEK) led by Rob Breiman and funded by the Global Alliance for Vaccines and Immunization (GAVI). This paper is published with the permission of the Director KEMRI.

Supplementary material

Supplementary Figure 1. Pie charts showing proportions of the most common genotypes circulating in Kilifi between 2002–2004 and 2010–2013. (a) Proportions of the most common VP7 G genotypes. (b) Proportions of the most common VP4 P genotypes. (c) Proportions of the most common G/P combinations. The colour coding for each specific genotype is shown below the

pie charts. The others category represents genotypes that circulated in low proportions. This included G3, 10, G12, G29 for panel a, P[6] and P[14] for panel b and G1P[4], G2P[8], G3P[4,6,8], G8P[6,8,14], G9P[4,6], G12P[4,6,8] for panel c.

[Click here to access the data](#)

Supplementary Figure 2. Pairwise distances within the common genotypes indicating the level of identity within genotypes. The y axis indicates genetic distances determined in MEGA v7 by calculating the proportions of pairwise nucleotide differences between sequences.

[Click here to access the data](#)

Supplementary Table 1. Evolutionary models used in construction of phylogenetic trees for Kilifi sequences and both Kilifi and Global sequences. The models were tested using the maximum likelihood method in MEGA v7.

[Click here to access the data](#)

Supplementary File 1. Accession numbers of VP7 G1 global sequences (list 1) and VP4 P[8] global sequences (list 2) used in phylogenetic comparison with the local strains.

[Click here to access the data](#)

References

- GBD Diarrhoeal Diseases Collaborators: **Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015.** *Lancet Infect Dis.* 2017; **17**(9): 909–948.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Troeger C, Khalil IA, Rao PC, *et al.*: **Rotavirus Vaccination and the Global Burden of Rotavirus Diarrhea Among Children Younger Than 5 Years.** *JAMA Pediatr.* 2018; **172**(10): 958–965.
[PubMed Abstract](#) | [Publisher Full Text](#)
- World Health Organization: **Rotavirus vaccines: an update.** *Wkly Epidemiol Rec.* 2009; **84**(50): 533–540.
[PubMed Abstract](#)
- Wandera EA, Mohammad S, Bundi M, *et al.*: **Impact of rotavirus vaccination on rotavirus hospitalisation rates among a resource-limited rural population in Mbita, Western Kenya.** *Trop Med Int Health.* 2018; **23**(4): 425–432.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ngabo F, Tate JE, Gatera M, *et al.*: **Effect of pentavalent rotavirus vaccine introduction on hospital admissions for diarrhoea and rotavirus in children in Rwanda: a time-series analysis.** *Lancet Glob Health.* 2016; **4**(2): e129–e136.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Weldegebriel G, Mwenda JM, Chakauya J, *et al.*: **Impact of rotavirus vaccine on rotavirus diarrhoea in countries of East and Southern Africa.** *Vaccine.* 2018; **36**(47): 7124–7130.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mwenda JM, Parashar UD, Cohen AL, *et al.*: **Impact of rotavirus vaccines in Sub-Saharan African countries.** *Vaccine.* 2018; **36**(47): 7119–7123.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Braeckman T, Van Herck K, Meyer N, *et al.*: **Effectiveness of rotavirus vaccination in prevention of hospital admissions for rotavirus gastroenteritis among young children in Belgium: Case-control study.** *BMJ.* 2012; **345**(7872): e4752.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jain S, Vashist J, Changotra H: **Rotaviruses: is their surveillance needed?** *Vaccine.* 2014; **32**(27): 3367–3378.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hoque SA, Kobayashi M, Takanaishi S, *et al.*: **Role of rotavirus vaccination on an emerging G8P[8] rotavirus strain causing an outbreak in central Japan.** *Vaccine.* 2018; **36**(1): 43–49.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zeller M, Rahman M, Heylen E, *et al.*: **Rotavirus incidence and genotype distribution before and after national rotavirus vaccine introduction in Belgium.** *Vaccine.* 2010; **28**(47): 7507–7513.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Almeida TN, Fiaccadori FS, Souza M, *et al.*: **Molecular characterization of group A rotavirus before and after the introduction of vaccines in Brazil.** *Rev Soc Bras Med Trop.* 2015; **48**(5): 599–602.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Desselberger U: **Rotaviruses.** *Virus Res.* 2014; **190**: 75–96.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gentsch JR, Glass RI, Woods P, *et al.*: **Identification of group A rotavirus gene 4 types by polymerase chain reaction.** *J Clin Microbiol.* 1992; **30**(6): 1365–1373.
[PubMed Abstract](#) | [Free Full Text](#)
- Metagenomics V: **Virus Classification.** 2018.
- Bányai K, László B, Duque J, *et al.*: **Systematic review of regional and temporal trends in global rotavirus strain diversity in the pre rotavirus vaccine era: insights for understanding the impact of rotavirus vaccination programs.** *Vaccine.* 2012; **30** Suppl 1: A122–130.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Van Damme P, Giaquinto C, Maxwell M, *et al.*: **Distribution of rotavirus genotypes in Europe, 2004-2005: the REVEAL Study.** *J Infect Dis.* 2007; **195** Suppl 1: S17–S25.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Seheri LM, Magagula NB, Peenze I, *et al.*: **Rotavirus strain diversity in Eastern and Southern African countries before and after vaccine introduction.** *Vaccine.* 2017; **36**(47): 7222–7230.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nokes DJ, Peenze I, Netshifhehe L, *et al.*: **Rotavirus genetic diversity, disease association, and temporal change in hospitalized rural Kenyan children.** *J Infect Dis.* 2010; **202** Suppl: S180–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- World Health Organization: **The Treatment of Diarrhoea. A Manual for physicians and other senior health workers.** 2005; 1–50.
[Reference Source](#)
- Simmonds MK, Armah G, Asmah R, *et al.*: **New oligonucleotide primers for P-typing of rotavirus strains: Strategies for typing previously untypeable strains.** *J Clin Virol.* 2008; **42**(4): 368–373.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gómara MI, Cubitt D, Desselberger U, *et al.*: **Amino Acid Substitution within the VP7 Protein of G2 Rotavirus Strains Associated with Failure To Serotype.** *J Clin Microbiol.* 2001; **39**(10): 3796–3798.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Katoh K, Rozewicki J, Yamada KD: **MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.** *Brief Bioinform.* 2017; **18**: 108.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Maes P, Matthijssens J, Rahman M, *et al.*: **RotaC: a web-based tool for the complete genome classification of group A rotaviruses.** *BMC Microbiol.* 2009; **9**: 238.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cavanaugh JE: **Model Selection: Bayesian Information Criterion.** *Wiley StatsRef: Statistics Reference Online.* 2016; 1–3.
[Publisher Full Text](#)
- Jamnikar-Ciglenecki U, Kuhar U, Steyer A, *et al.*: **Whole genome sequence and a phylogenetic analysis of the G8P[14] group A rotavirus strain from roe deer.** *BMC Vet Res.* 2017; **13**(1): 353.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Matthijssens J, Potgieter CA, Ciarlet M, *et al.*: **Are human P[14] rotavirus strains**

- the result of interspecies transmissions from sheep or other ungulates that belong to the mammalian order *Artiodactyla*? *J Virol.* 2009; **83**(7): 2917–2929.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Maes P, Matthijssens J, Rahman M, *et al.*: **RotaC: a web-based tool for the complete genome classification of group A rotaviruses.** *BMC Microbiol.* 2009; **9**: 238.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Nyangao J, Page N, Esona M, *et al.*: **Characterization of human rotavirus strains from children with diarrhea in Nairobi and Kisumu, Kenya, between 2000 and 2002.** *J Infect Dis.* 2010; **202** Suppl: S187–S192.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Gikonyo JN, Nyangao J, Mbae C, *et al.*: **Molecular characterization of group A rotaviruses in Mukuru slums Kenya: Detection of novel strains circulating in children below 5 years of age.** *BMC Res Notes.* 2017; **10**(1): 290.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Kiulia NM, Nyaga MM, Seheri ML, *et al.*: **Rotavirus G and P types circulating in the eastern region of Kenya: predominance of G9 and emergence of G12 genotypes.** *Pediatr Infect Dis J.* 2014; **33** Suppl 1(1): S85–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Esona MD, Armah GE, Steele AD: **Rotavirus VP4 and VP7 genotypes circulating in Cameroon: Identification of unusual types.** *J Infect Dis.* 2010; **202** Suppl: S205–S211.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Dereci S, Çopur Çiçek A, Savaş Acar S, *et al.*: **Prevalence and genotype distribution of rotaviruses in children with gastroenteritis in rize province.** *Bosn J Basic Med Sci.* 2015; **15**(3): 35–39.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Matthijssens J, Rahman M, Yang X, *et al.*: **G8 rotavirus strains isolated in the Democratic Republic of Congo belong to the DS-1-like genogroup.** *J Clin Microbiol.* 2006; **44**(5): 1801–1809.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Asmah RH, Green J, Armah GE, *et al.*: **Rotavirus G and P genotypes in rural Ghana.** *J Clin Microbiol.* 2001; **39**(5): 1981–1984.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Green MJ, Peterson SK, Baker MW, *et al.*: **Use of an educational computer program before genetic counseling for breast cancer susceptibility: effects on duration and content of counseling sessions.** *Genet Med. NIH Public Access.* 2005; **7**(4): 221–229.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Dunn SJ, Greenberg HB, Ward RL, *et al.*: **Serotypic and genotypic characterization of human serotype 10 rotaviruses from asymptomatic neonates.** *J Clin Microbiol.* 1993; **31**(1): 165–169.
[PubMed Abstract](#) | [Free Full Text](#)
38. Delogu R, Ianiro G, Morea A, *et al.*: **Molecular characterization of two rare human G8P[14] rotavirus strains, detected in Italy in 2012.** *Infect Genet Evol.* 2016; **44**: 303–312.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Agutu MT, Ongus J, Kombich J, *et al.*: **Prevalence and genetic diversity of rotavirus infection in children with acute gastroenteritis in a hospital setting, Nairobi Kenya in post vaccination era: a cross-sectional study.** *Pan Afr Med J.* 2017; **26**: 38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Mwanga, MJ, Owor BE, Njeru RW, *et al.*: **Replication Data for: Molecular characterization of rotavirus group A strains circulating prior to vaccine introduction in rural coastal Kenya, 2002-2013.** Harvard Dataverse, V1. 2018.
<http://www.doi.org/10.7910/DVN/LVGYW>

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 10 April 2019

<https://doi.org/10.21956/wellcomeopenres.16252.r34548>

© 2019 Nyaga M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Martin Munene Nyaga

Next Generation Sequencing Unit, Department of Medical Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa

This is the first detailed study on the prevalence of RVA genotypes causing diarrhoea in children in rural coastal Kenya spanning over a decade to investigate the evolutionary dynamics of human rotaviruses pre-vaccine introduction, which present original findings as per the problem statement of the study.

While this reviewer finds the study appropriate for indexing, we do suggest the authors should consider the following suggestions before final acceptance as noted by the reviewer:

Introduction:

1. The authors need to clarify that the dominance of G2P[4]s in Belgium and Brazil were not linked to rotavirus vaccine pressure. Recent data in different parts of the world indicate that the vaccine provides heterotypic immunity against a variety of rotavirus strains including G2P[4]s. Furthermore, G2P[4]s are known to have a cyclic trend where they disappear and reappear in four-year cycles which could have been the case at the time.
2. Encodes 11 proteins, sometimes 12 (NSP5/6).
3. Consider rephrasing to "G1-4, G9, and G12 in combination with P,[4], P[6] and P[8] are the most common genotypes..."
4. 37P types and 27 G types have been described globally. Another 13P and 7 G types were proposed as of 2018 and await rectification by the RCWG and ICTV.

Methods:

1. Indicate the ethics number as approved by SERU.

Results:

1. Add a comma before the word 'respectively' throughout the document.
2. Uncommon strains bearing G1P[4]...
3. How many sequence-based methods were applied? To my understanding only Sanger sequencing, right? If you used NGS you could have detected the mixed infections. Rectify accordingly.
4. Homologous meaning completely similar. "High homology" is not scientifically sound. Again 92% is very diverse for different genotypes. Check proposed cut off values.
5. There are partial sequences (Incomplete CDs) but there is no partial sequencing method unless if refereeing to primer based sequencing based on targeted regions.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Medical Virology and Next Generation Sequencing**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 06 May 2019

Mike Mwangi, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

The authors need to clarify that the dominance of G2P[4]s in Belgium and Brazil were not linked to rotavirus vaccine pressure. Recent data in different parts of the world indicate that the vaccine provides heterotypic immunity against a variety of rotavirus strains including G2P[4]s.

Furthermore, G2P[4]s are known to have a cyclic trend where they disappear and reappear in four-year cycles which could have been the case at the time.

Our response: This has been revised in the updated version. (Refer to the Introduction section).

Encodes 11 proteins, sometimes 12 (NSP5/6),

Our response: This has been revised in the updated version. (Refer to the Introduction section).

Consider rephrasing to "G1-4, G9, and G12 in combination with P,[4], P[6] and P[8] are the most common genotypes..."

Our response: We also received similar comments from the other reviewers and this sentence has since been revised.

37P types and 27 G types have been described globally. Another 13P and 7 G types were proposed as of 2018 and await rectification by the RCWG and ICTV.

Our response: This has been revised in the updated version.

Methods:

Indicate the ethics number as approved by SERU.

Our response: The SERU number is now indicated.

Results:

Add a comma before the word 'respectively' throughout the document.

Our response: This has been checked and corrected throughout the manuscript.

Uncommon strains bearing G1P[4].

Our response: This has been rectified in the Results section.

How many sequence-based methods were applied? To my understanding only Sanger sequencing, right? If you used NGS you could have detected the mixed infections. Rectify accordingly.

Our response: One sequencing method, i.e, di-deoxy sequencing was used in this study. Mixed infections identified in phase I (2002-2004) were determined by genotype-specific primer-based PCR method and previously published (Nokes *et al.*, 2010) This has been made clear in our revised version.

Homologous meaning completely similar. "High homology" is not scientifically sound. Again 92% is very diverse for different genotypes. Check proposed cut off values.

Our response: This has been checked and reworded to read "with sequence similarities of between 92-100% at nt level". (Refer to the Results section).

There are partial sequences (Incomplete CDs) but there is no partial sequencing method unless if refereeing to primer-based sequencing based on targeted regions.

Our response: This has been changed from "**partial sequencing method**" to "**di-deoxy sequencing method**" (Refer to the Discussion section).

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 February 2019

<https://doi.org/10.21956/wellcomeopenres.16252.r34355>

© 2019 Esona M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Mathew D Esona

Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

Summary of article:

This article by Owor et al., described the genotypes circulating pre-vaccination in Kilifi, Kenya. They reported G1P[8] as the most predominant strain in circulation during the study periods.

1) Is the work clearly and accurately presented and does it cite the current literature? With the exception of citing genotype G12P[8] as the sixth most common genotype in circulation, the authors have used recent and current citations. I would recommend that authors edit the statement in paragraph 3 to include G12P[8].

2) Authors use sequencing method to determine these genotypes. With sequencing you can not call mixed genotypes. Authors should explain how mixed genotypes were determined and confirmed by sequencing reaction.

3). The above data adds very little to the information we have on rotavirus genotypes circulating in Kenya. I recommend that authors take this one step further, by performing a next generation sequencing on selected genotypes. This will give the public more information on the interaction between the vaccine strains and the wild type strains in Kenya.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Virology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 09 Apr 2019

Mike Mwangi, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

- **Reviewer comment:** Is the work clearly and accurately presented and does it cite the current literature? *With the exception of citing genotype G12P[8] as the sixth most common genotype in circulation, the authors have used recent and current citations. I would recommend that authors edit the statement in paragraph 3 to include G12P[8].*
- **Our response:** We appreciate this correction and have modified the sentence (starting line #31) to read "Globally, G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] have been identified as the most common genotypes (in decreasing order) while G12P[6] and G12P[8] have recently been reported as emerging genotypes."
- **Reviewer comment:** *Authors use a sequencing method to determine these genotypes. With sequencing, you cannot call mixed genotypes. Authors should explain how mixed genotypes were determined and confirmed by sequencing reaction.*
- **Our response:** We agree that our sequencing strategy could not support calling mixed infections. This is a limitation we highlight in the discussion section of the manuscript. As explained in the manuscript, the study was divided into two phases (I & II). In phase I (2002-2014), genotyping was previously done using the PCR primer-based strategy <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2923076/>. Phase II of the study used sequencing to genotype. We have made this clear throughout the manuscript in this revised version.

- **Reviewer comment:** *The above data adds very little to the information we have on rotavirus genotypes circulating in Kenya.*
- **Our response:** We point out that our RVA genotypic and sequence data from Kenya represents a large dataset in the period leading up to the vaccine introduction that spans a wide time period. We feel this offers additional information than previously available by which to explore patterns in one part of the country in a period spanning a decade to show natural fluctuations and the extent of genetic diversity in circulating strains in the absence of a vaccination programme. We further examined the global context of the local strains to understand the nature of their source year-in-year-out. Additionally, to the best of our knowledge, this is the first study in Kenya, to generate vast rotavirus partial sequence data from a large sample set collected before vaccine introduction, hence an addition to the limited sequence data for tracking RVA transmission. We believe that this study provides an important baseline for future studies especially those attempting to interpret post-vaccine introduction strain patterns.
- **Reviewer comment:** *I recommend that authors take this one step further, by performing a next-generation sequencing on selected genotypes. This will give the public more information on the interaction between the vaccine strains and the wild type strains in Kenya.*
- **Our response:** We thank the reviewer for this suggestion. We are developing an in-house NGS protocol but the results will form the basis of a separate publication.

Competing Interests: No competing interests.

Author Response 06 May 2019

Mike Mwangi, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

Is the work clearly and accurately presented and does it cite the current literature?

Reviewer comment: *With the exception of citing genotype G12P[8] as the sixth most common genotype in circulation, the authors have used recent and current citations. I would recommend that authors edit the statement in paragraph 3 to include G12P[8].*

Our response: We appreciate this correction and have modified the sentence to read “Globally, G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] have been identified as the most common genotypes (in decreasing order) while G12[P6] and G12P[8] have recently been reported as emerging genotypes.”

Authors use sequencing method to determine these genotypes. With sequencing you cannot call mixed genotypes. Authors should explain how mixed genotypes were determined and confirmed by sequencing reaction.

Our response: We agree that our sequencing strategy could not support calling mixed infections. This is a limitation we highlight in the discussion section of the manuscript. As

explained in the manuscript, the study was divided into two phases (I & II). In phase I (2002-2014), genotyping was previously done using the PCR primer-based strategy <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2923076/>. Phase II of the study used sequencing to genotype. We made this clear through the manuscript in this revised version.

The above data adds very little to the information we have on rotavirus genotypes circulating in Kenya.

Our response: We appreciate this concern but would also like to point out that RVA genotypic and sequence data from Kenya in the period leading to the vaccine introduction is sparse. We examine and present a reasonably large dataset to explore patterns in one part of the country in a period spanning a decade to show natural fluctuations and the extent of genetic diversity in circulating strains in the absence of a vaccination programme. We further examined the global context of the local strains to understand the nature of their source year-in year-out. We believe that this study provides an important baseline for future studies especially those attempting to interpret post-vaccine introduction strain patterns.

I recommend that authors take this one step further, by performing a next-generation sequencing on selected genotypes. This will give the public more information on the interaction between the vaccine strains and the wild type strains in Kenya.

Our response: We thank the reviewer for this suggestion and we totally agree with it. We actually have already started generating some sequence data using the NGS approach but are convinced that this will be best presented as a separate report in which the focus will not be genotype prevalence and long-term patterns.

Competing Interests: No competing interests were disclosed.

Reviewer Report 21 December 2018

<https://doi.org/10.21956/wellcomeopenres.16252.r34354>

© 2018 Wandera Apondi E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ernest Wandera Apondi

Kenya Research Station, Institute of Tropical Medicine, KEMRI/Nagasaki University, Nairobi, Kenya

Owor *et al* conducted a longitudinal epidemiological study of rotavirus genotype distribution patterns in Coastal Kenya before the nationwide introduction of rotavirus vaccine. The authors employed methods such as enzyme immunoassay (EIA) for the detection of group A rotavirus

(RVA) and partial sequencing of RVA positive samples in VP4 and VP7 segments for G and P genotyping. Data analysis reveals remarkable genetic diversity of RVA strains circulating in this area, characterized by substantial frequencies of unusual, mixed and emerging genotypes. Temporal fluctuation in RVA genotypes was observed, with major shifts in G-P predominance involving G1P[8] and G8P[4].

The study was well conducted and the manuscript well written. The findings of this study are timely in light of the recent introduction of rotavirus vaccine in Kenya and provide the baseline data necessary for the assessment of vaccine effectiveness. This baseline data will also allow monitoring of RVA G and P genotype changes that may alter vaccine effectiveness or that may be a result of vaccination, such as possible breakthrough events under vaccine immune selective pressure.

Of noteworthy, a rare G8P[14] strain was detected in this study and the partial sequencing of this strain indicated that its VP7 segment is closely related to humans and animals while its VP4 segment clustered closely to that of human and bovine origin. Due to the unconventional nature of this and many other uncommon strains detected in this study, it will be useful to sequence and characterize the full genomes of the representative strains in order to provide important insights into their evolutionary dynamics.

Furthermore, since the uncommon strains, such as the ones detected in this study are either partially or fully heterotypic to the currently licensed RVA vaccines (RV1 and RV5), vaccine effectiveness against these strains needs to be closely monitored.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Author Response 09 Apr 2019

Mike Mwanga, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

We thank the reviewer for this appraisal of our work. We confirm that we are developing a whole genome sequencing protocol to sequence a select group of positive samples. This will include unusual strains and a longitudinal set spanning before to after introduction of the vaccine. The results will be presented in a separate publication.

Competing Interests: No competing interests were disclosed.

Author Response 06 May 2019

Mike Mwanga, KEMRI Wellcome Trust Research Program, Kilifi, Kenya

We thank the reviewer for applauding the work. We can also confirm that whole genome sequencing of the positive samples and including those collected in the post-vaccine introduction period has commenced.

Competing Interests: No competing interests were disclosed.