 doi:10.20944/preprints202010.0470.v1

Article

Multiple introductions and predominance of rotavirus group A genotype G3P[8] in coastal Kenya in 2018, 4 years after nationwide vaccine introduction

Mike J. Mwanga[1], Jennifer R. Verani[2], Richard Omore[3], Jacqueline E. Tate[4], Umesh D. Parashar[4], Nickson Murunga[1], Elijah Gicheru[1], Robert F. Breiman[5], D. James Nokes[1,6], and Charles N. Agoti[1,7,*]

[1] Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Kilifi, Kenya; mikemwanga6@gmail.com (M.J.M.); nmurunga@kemri-wellcome.org (N.M.); egicheru@kemri-wellcome.org (E.G.); jnokes@kemri-wellcome.org (D.J.N.)
[2] Centers for Disease Control and Prevention (CDC), Nairobi, Kenya; qzr7@cdc.gov (J.R.V.)
[3] KEMRI, Center for Global Health Research (KEMRI-CGHR), Kisumu, Kenya; omorerichard@gmail.com (R.O.).
[4] CDC, Atlanta, Georgia, USA; uap2@cdc.gov (U.D.P), jqt8@cdc.gov (J.E.T.)
[5] Emory University, Atlanta, GA, USA; rfbreiman@emory.edu (R.F.B).
[6] University of Warwick, Coventry, United Kingdom.
[7] Pwani University, Kilifi, Kenya.
* Correspondence: cnyaigoti@kemri-wellcome.org; C.N.A)

Abstract: Globally, rotavirus group A (RVA) remains a major cause of severe childhood diarrhea, despite the use of vaccines in > 100 countries. RVA sequencing for local outbreaks facilitates investigation into strain composition, origins, spread, and vaccine failure. In 2018, we collected 248 stool samples from children aged <13 years admitted with diarrheal illness to Kilifi County Hospital, coastal Kenya. Antigen screening detected RVA in 55 samples (22.2%). Of these, VP7 (G) and VP4 (P) segments were successfully sequenced in 48 (87.3%) and phylogenetic analysis based on the VP7 sequences identified seven genetic clusters with six different GP combinations; G3P[8], G1P[8], G2P[4], G2P[8], G9P[8] and G12P[8]. The G3P[8] strains predominated the season (n=37, 67.2%) and comprised three G3 genetic clusters that fell within Lineage I and IX (the latter also known as equine-like G3 lineage). Both two G3 lineages have been recently detected in several countries. Our study is the first to document African children infection with G3 lineage IX. These data highlight the global nature of RVA transmission and the importance of increasing global rotavirus vaccine coverage.

Keywords: Gastroenteritis, rotavirus, G3[P8], phylogenetics, equine-like

1. Introduction

Following progressive introduction of rotavirus vaccines into national immunization programmes (NIP) of >100 countries since 2006, a significant decline of rotavirus group A (RVA) disease burden has occurred [1, 2]. However, despite these successes, RVA remains a leading cause of diarrhea morbidity and mortality [3, 4], resulting in an estimated 128,500 deaths annually among under-5-year-olds, a majority occurring in low-income settings [5]. Consistently, licensed oral RVA vaccines have underperformed in low-income settings compared with high-income settings [6, 7]. After monovalent Rotarix® vaccine was introduced into Kenya's NIP in July 2014, with doses given at 6 and 10 weeks of life, a multi-site case-control study found an overall 2-dose vaccine effectiveness of only 64% (95% confidence interval (CI): 35-80%) in under-5-year-olds [8]. In England, the same vaccine showed effectiveness of 77% (95% CI: 66-85%) [9].

In humans, RVA immunity is partly conferred by neutralizing antibodies directed against the VP4 (protease-sensitive) and VP7 (glycoprotein) viral capsid surface proteins that define P and G

 doi:10.20944/preprints202010.0470.v1

types, respectively [10]. These two viral proteins are highly diverse, with up to 36 different G and 51 different P types recorded to-date [11], some of which predominantly infect non-human animal species [12]. Among other factors, the higher number of co-circulating GP genotypes in low-income settings has been proposed to be a potential contributor to rotavirus vaccine underperformance [6].

Currently, there are four licensed and WHO pre-qualified RVA vaccines; all live attenuated and administered orally, but with different strain compositions. These are monovalent Rotarix®(G1P[8]), pentavalent RotaTeq® (5 reassortant viruses; G1, G2, G3, G4 and G6 genotypes in combination with P[8]), monovalent ROTAVAC® (G9P[11]) and pentavalent ROTASIIL® (5 reassortant viruses; G1, G2, G3, G4 and G9). All four vaccines were shown to be largely cross-protective against heterotypic strains in both clinical trials and following vaccine implementation in several settings [6, 13]. Paradoxically, post-vaccine rollout, outbreaks caused by strains heterotypic to the vaccine in use have been sometimes reported in countries, occurring in patterns seeming to be influenced by the vaccine regimen in use [14-16].

Recent genotyping studies of RVA have found increased proportions of G2P[4], G3P[8] and G12P[8] genotypes in RVA vaccinating countries [14, 16-18]. These genotypes appeared to play only a minor role in the pre-vaccine era; thus, their increasing prevalence is consistent with increased capacity in escaping vaccine immunity [12, 19]. Furthermore, there have been several reports of human infection with equine-like G3 viruses suggestive of greater human vulnerability to antigenically novel RVA strains [20-30]. At the Kenya Medical Research Institute (KEMRI) - Wellcome Trust Research Programme, we have maintained a RVA surveillance at Kilifi County Hospital (KCH), located in rural coastal Kenya since 2009 [31]. The aim of the current analysis was to determine the genetic relatedness of the strains that were in circulation in the 2018 RVA season in Kilifi, their origins, global phylogenetic context, and role in the local sub-optimal vaccine performance.

2. Results

2.1. Study population characteristics.

Between January and December 2018, 384 children aged <13 years were admitted to KCH with diarrhea as one of their illness symptoms. Of these, 208 (54.2%) were Kilifi Health and Demographic surveillance system (KHDSS) area residents. A stool sample was obtained from 248 (64.6%), **Figure 1, panel (a)**. The main reasons for non-sampling were; death (n=13), discharge or transfer before sample collection (n=22), consent refusal (n=52), or other (n=16). Among study eligible children (n=384), the distribution of the sampled and not sampled children differed significantly across age strata (p=0.002) and discharge outcome (p<0.001), **Table 1**. The distribution of the sampled and not sampled children did not differ significantly by gender or rotavirus vaccine eligibility, **Table 1**. The majority of the eligible participants were aged < 2 years (68.2%) and were age-eligible to have received one or two doses of rotavirus vaccine (83.6%). By EIA testing, RVA was detected in 55 children (22.2%), **Figure 1, panel (a),** 32 (58.1%) of which were KHDSS area residents. Fifty-one (92.7%) of the RVA positive children were age-eligible to have received two doses of the RVA vaccine. Of these, the vaccination status was known for 36 (70.6%), of which 29 (80.6%) were confirmed to have received two doses of Rotarix® vaccine while the remainder (19.4%) received one dose, **Table 1**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

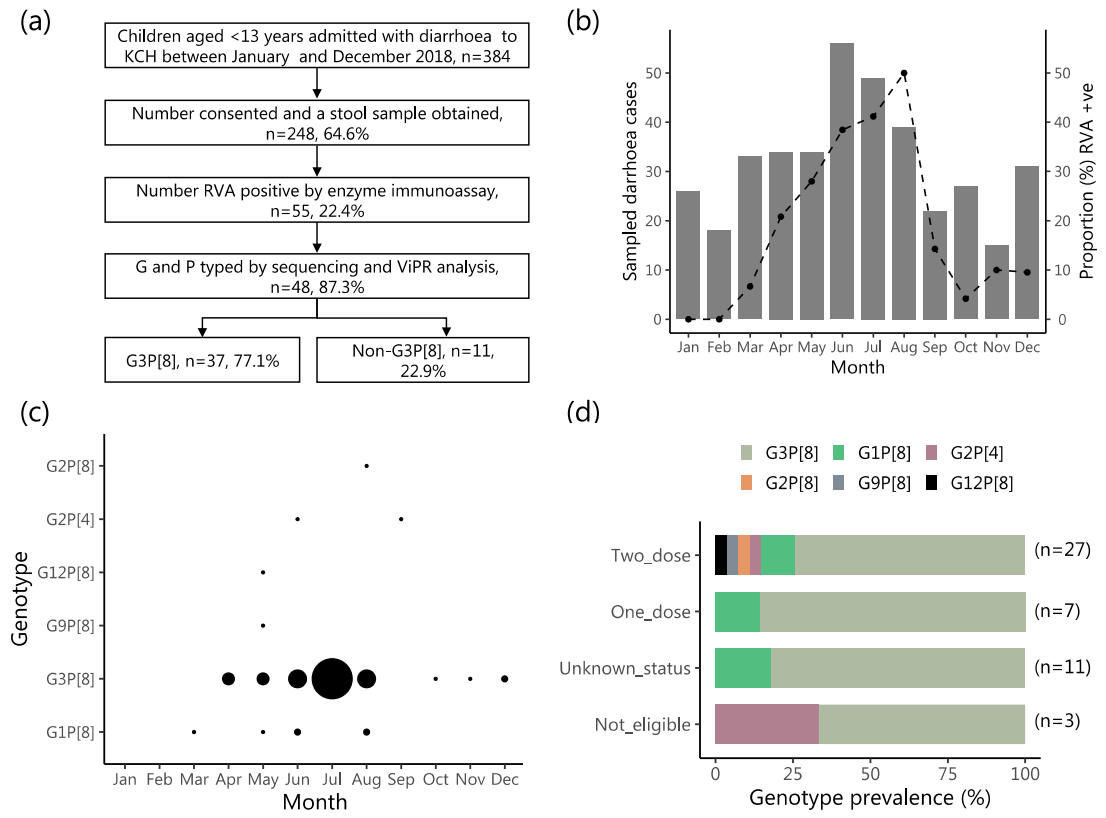

Figure 1. Summary of rotavirus group A (RVA) surveillance in Kilifi County Hospital (KCH) in 2018 and identified genotypes. Panel (a) sample flowgram from patient recruitment to VP4 and VP7 genotyping results of the RVA positives. Panel (b) Monthly cases of diarrhea in children <13 years recorded at KCH in 2018 (grey bars) compared with monthly proportions of RVA positive samples (black dashed line on the secondary axis). Panel (c) the number of RVA positive samples by month in 2018 and by GP genotype. The black circle size is proportional to the number of samples (the smallest indicates one sample and the largest is 13 samples). Panel (d) genotypes identified in children according to rotavirus vaccination status.

2.2. Characteristics of the RVA infections and the infected children

RVA prevalence was higher in female compared to male children admitted with diarrhea (29.8% vs 15.7%, p=0.008), **Table 1**. RVA was detected in all months of 2018 except January and February **Figure 1, panel (b)**. Diarrhea cases peaked in June while RVA prevalence peaked in August (50% of all collected samples were RVA positive). Sequencing and GP typing was successful for 48 (87.3%) of the 55 RVA-positive samples. Five G types (G1, G2, G3, G9 and G12) and two P types (P[4] and P[8]) were identified in the successfully sequenced samples. From these, six GP combinations were identified, namely: G3P[8] (n=37, 77.1%), G1P[8] (n=6, 12.5%), G2P[4] (n=2, 4.2%), G2P[8] (n=1, 2.1%), G9P[8] (n=1, 2.1%) and G12P[8] (n=1, 2.1%). The G3P[8] and G1P[8] strains were the only genotypes detected for > 2 months while the other four genotypes were detected sporadically (1-2 months), **Figure 1, panel (c)**. The distribution of the infecting genotype (summarized as G3P[8] versus non-G3P[8]) did not differ significantly by gender, patient age, vaccination status or discharge outcome, **Table 2 and Figure 1 panel (d).**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

Table 1: A comparison of demographic characteristics of children admitted to KCH sampled versus those whom were not sampled in 2018 and those RVA positive versus those negative.

Characteristic	All (%)	Sampled (%)	Unsampled (%)	p value[$]	RVA +ve (%)	RVA -ve (%)	p value[*]
Number of patients	384	248 (64.6)	136 (35.4)		55 (22.2)	193 (77.8)	
Gender				0.728			0.008
Male	210 (54.7)	134 (54.0)	76 (55.9)		21 (38.2)	113 (58.6)	
Female	174 (45.3)	114 (46.0)	60 (44.1)		34 (61.8)	80 (41.5)	
Age							
Mean (SD[¶])	27.4 (29.9)	26.4 (31.8)	29.3 (26.1)	0.352	19.6 (15.0)	28.3 (35.0)	0.073
Median (IQR[δ])	16.8 (9.8-29.3)	15.1(9.4-24.1)	19.9 (12.0-39.0)	0.025	15.4 (9.9-20.8)	15.1(8.9-24.9)	1.000
Age group				0.002			0.254
0-11 months	126 (32.8)	92 (37.1)	34 (25.0)		19 (34.6)	73 (37.8)	
12-23 months	136 (35.4)	92 (37.1)	44 (32.4)		25 (45.5)	67 (34.7)	
24-59 months	73 (19.0)	34 (13.7)	39 (28.7)		8 (14.6)	26 (14.0)	
>60 months	49 (12.8)	30 (12.1)	19 (14.0)		3 (5.5)	27 (14.0)	
RVA vaccine eligibility				0.327			0.063
Age eligible 2 dose	317 (82.6)	204 (82.3)	113 (83.1)		51 (92.7)	153 (79. 3)	
Age-eligible 1 dose	4 (1.0)	4 (1.6)	0 (0.0)		0 (0.0)	4 (2.1)	
Age ineligible	63 (16.4)	40 (16.1)	23 (16.9)		4 (7.3)	36 (18.7)	
Vaccination status (n=321)				0.273			0.209
Two dose eligible & received 2 doses	165 (51.4)	111 (53.4)	54 (47.8)		29 (56.9)	82 (52.2)	
Two dose eligible & received 1 dose	24 (7.5)	17 (8.2)	7 (6.2)		7 (13.7)	10 (6.2)	
One or 2 dose eligible but received none	6 (1.8)	2 (1.0)	4 (3.5)		0 (0.0)	2 (1.3)	
One or 2 dose eligible but status unknown	126 (39.3)	78 (37.5)	48 (42.5)		15 (29.4)	63 (40.1)	
Outcome (n=379)				<0.001			0.194
Died	38 (10.0)	13 (5.3)	25 (18.9)			12 (6.3)	
Alive	341 (90.0)	234 (94.7)	133 (81.1)			180 (93.8)	

[¶]SD stands for standard deviation; [δ]IQR stands for interquartile range; [$]p value for comparison of sampled and not sampled groups; [*] p value for comparison of RVA positive and negative groups.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

Table 2: Characteristics of children whom were infected with rotavirus G3P[8] versus those whom were infected with non-G3P[8]

Characteristic	Genotyped RVA (%)	G3P[8] (%)	Non-G3P[8] (%)	p value
Number of patients	48	37 (77.1)	11 (22.9)	
Gender				0.248
Male	19 (39.6)	13 (35.1)	6 (55.6)	
Female	29 (60.4)	24 (64.9)	5 (45.5)	
Age				
Mean (SD#)	19.3 (14.8)	19.2 (13.3)	19.5 (18.6)	0.946
Median (IQR)	15.7 (9.9-20.4)	15.9 (9.8-20.4)	15.4 (7.8-23.1)	1.000
Age group				0.770
0-11 months	17 (35.4)	13 (35.1)	4 (36.4)	
12-23 months	22 (45.8)	17 (46.0)	5 (45.6)	
24-59 months	7 (14.6)	6 (16.2)	1 (9.1)	
>60 months	2 (4.2)	1 (2.7)	1 (9.1)	
RVA vaccine eligibility				0.658
Age eligible 2 dose	45 (93.8)	35 (94.6)	10 (90.9)	
Age eligible 1 dose	0 (0.0)	0 (0.0)	0 (0.0)	
Age ineligible	3 (6.3)	2 (5.4)	1 (9.1)	
RVA vaccination status among eligible (n=45)				0.751
Two dose eligible & received two doses	27 (60.0)	20 (57.1)	7 (70.0)	
Two dose eligible & received one dose	7 (15.6)	6 (17.1)	1 (10.0)	
One or 2 dose eligible but received none	0 (0.0)	0 (0.0)	0 (0.0)	
One or 2 dose eligible but status unknown	11 (24.4)	9 (25.7)	2 (20.0)	
Outcome				0.064
Died	1 (2.1)	0 (0.0)	1 (9.1)	
Alive	47 (97.2)	37 (100.0)	10 (90.9)	

[#]SD stands for standard deviation, [♭]IQR stands for interquartile range

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

2.3. Genetic diversity in the sequenced viruses

For the VP4 segment, a 579 nt long region (~ 25%) was recovered for 47 viruses (87.3%) while for the VP7 segment, a 644 nt long region (~ 65%) was recovered for 48 viruses (87.3%). One virus (KEN/KLF0879/2018), genotyped G9P[8], yielded a significantly shorter VP4 fragment relative to the other viruses (<500 nt) due to low quality sequencing data and was excluded from subsequent analyses. Consistent with the greater number of assigned G types (n=5) compared to P types (n=2) types, the range of pairwise nt differences was much greater in the VP7 (up to 203 nt differences) compared to VP4 segment (up to 87 nt differences), **Figure 2, panel (a)** and **(b),** respectively. A multi-modal distribution of nt differences was observed for both VP4 and VP7 segments. A total of 328 (~51%) and 141 (~24%) SNP positions were identified in the sequenced VP7 and VP4 fragments, respectively. Of the 48 sequenced samples, 22 (45.8%) yielded unique VP7 sequences while 17 (36.2%) gave unique VP4 sequences.

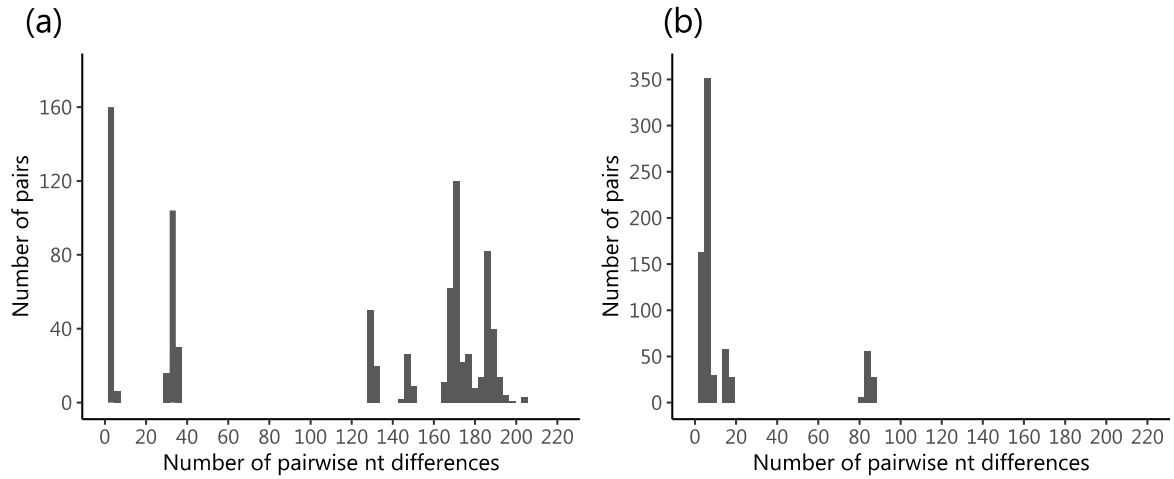

Figure 2. Genetic diversity in the sequenced RVA positives from KCH. Panel (a) shows the distribution of pairwise nt differences in the sequenced portion of VP7 (644 nt long) of 48 RVA positives. Panel (b) shows the distribution of pairwise nt differences in the sequenced portion of VP4 (579 nt long) of 47 RVA positives.

2.4. Molecular genetic clusters.

Using the range of pairwise nt differences observed in first modal distribution for the VP7 (0 to 20 nt differences, i.e. >97 % nt similarity) to define a molecular genetic cluster, seven G clusters were assigned (named Clu_1-7). Members of a cluster were universally of same G type. All G type sequences identified of the same type formed a single cluster except G3P[8] that occurred in three clusters, named Clu_3/G3P[8], Clu_4/G3P[8] and Clu_5/G3P[8]. The temporal pattern of the assigned clusters is shown in **Figure 3, panel (a)**. Most of the high incidence months (April to August) had multiple genetic clusters co-circulating, except for July, which had a single G3P[8] cluster. The reconstructed phylogenetic relationship between strains of the different G and P types sequenced is shown in **Figure 3, panel (b)** and **(c)**. The VP7 phylogeny showed segregation of the seven clusters we identified from the pairwise nt difference analysis. The VP4 phylogeny showed less clear-cut phylogenetic clustering with respect to the assigned genetic clusters. The two phylogenies were not entirely congruent, a feature suggestive of reassortment in the local strains.The minimum spanning networks reconstructed for both the VP7 and VP4 sequences are shown in **Figure 3, panel (d)** and **(e)**. Viruses in the same genetic cluster consistently had 4 or less nt differences to the closest next virus within the same genetic cluster.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

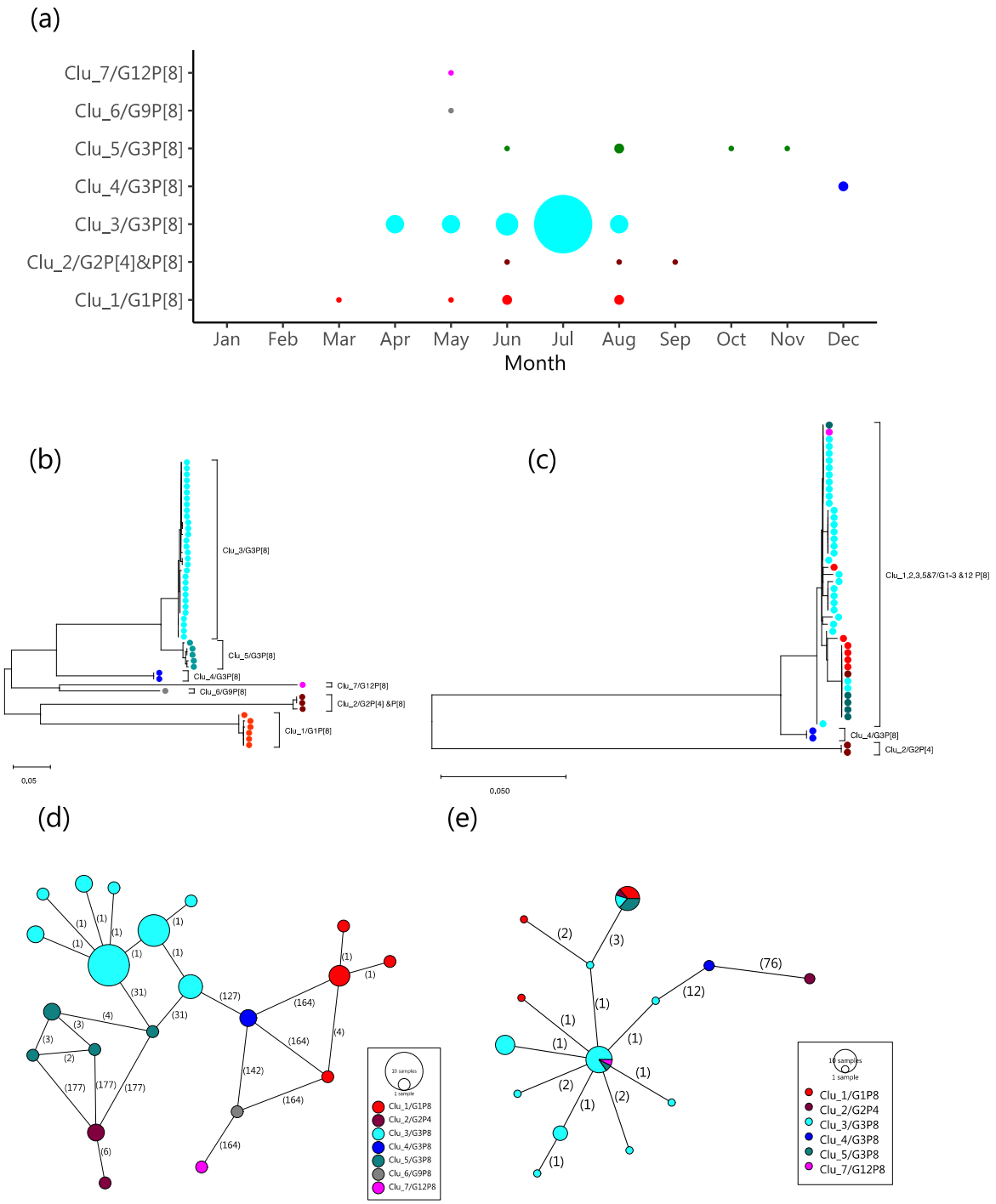

Figure 3. Temporal and genetic relatedness of the sequenced Kilifi rotaviruses. Panel (a) number of RVA positive samples by molecular genetic cluster and month. The circle sizes are proportional to the number of samples (the smallest indicates one sample and the largest is 13 samples). Panel (b) shows a Maximum Likelihood (ML) tree of the Kilifi 48 VP7 sequences. Panel (c) shows an ML tree of the Kilifi 47 VP4 sequences. Panel (d) shows the reconstructed POPART minimum spanning network from the 48 VP7 sequences. The vertexes represent the sequenced VP7 haplotypes. The size of the vertex is proportional to the number of haplotypes (identical sequences) and is colored by the assigned molecular genetic cluster. The numbers shown on the edges represent the number of nucleotide changes from one vertex (haplotype) to the next. Panel (e) same as panel (d) above but for the Kilifi 47 VP4 sequences.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

A few viruses in different VP7-based genetic clusters had identical VP4 sequences. Twenty-eight of the 48 genotyped samples were from KHDSS area residents. The geographical distribution of all diarrhea admissions and the RVA positives by genetic cluster is shown in **Figure 4**. Cases of the predominant Clu_3/G3P[8] strains came from only a few locations although it appeared that road access (especially the Malindi-Mombasa highway) may have played a role in influencing which patients were turning up at KCH due to easier access.

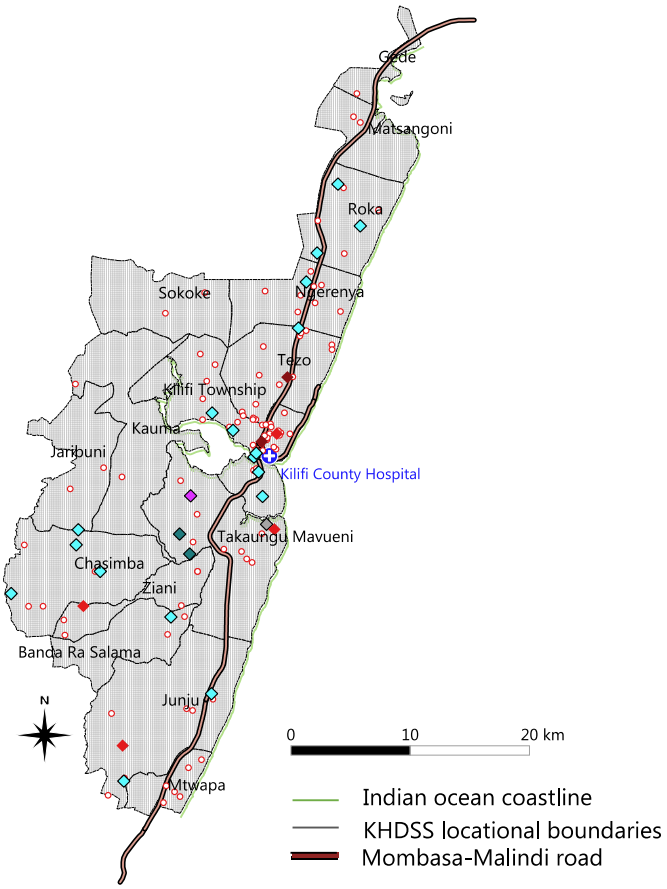

Figure 4. Geographic origin distribution of sampled children who presented with diarrhea symptoms at KCH and were KHDSS residents. The diamonds indicate RVA positives and are colored by genetic cluster (see Figure 3 for details. The red empty circles indicate RVA negative cases.

2.5. Global genetic context of the Kilifi 2018 G3 strains

A total of 308 G3 sequences from 26 countries fully met the criteria for inclusion as comparison data, including two previously collected in Kenya. The phylogeny derived from the combined Kilifi and global G3 viruses is shown in **Figure 5.** A majority of the global viruses fell within two of nine previously identified G3 lineages [21]; Lineage I and equine-like G3 lineage (named Lineage IX). The Kilifi G3 sequences had representation in both these two lineages: Lineage I (n=35, 94.6%) and equine-like G3 Lineage (n=2, 5.6%). Viruses of the genetic cluster Clu_4/G3P[8] clustered with the equine-like G3 Lineage while the Kilifi G3 Lineage I viruses separated into two groups that corresponded to the Clu_3/G3P[8] cluster (n=30) and the Clu_5/G3P[8] cluster (n=5). The distribution of the pairwise nt differences in the compiled global G3 sequences dataset, like for the Kilifi G3 viruses, showed a multi-modal distribution (figure ot shown). The first major trough was observed at 27 nt differences. On applying the threshold used to identify the local molecular genetic clusters (>97% genetic similarity) on the global G3 dataset, 20 clusters were identified, **Table 3**. Of these, ten were singletons, six comprised of between 2 and 5 members and the remaining four clusters had 6, 15, 119 and 179

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

members. All the Kilifi G3 viruses fell in the 3 clusters that had the highest membership overall, **Table 3**.

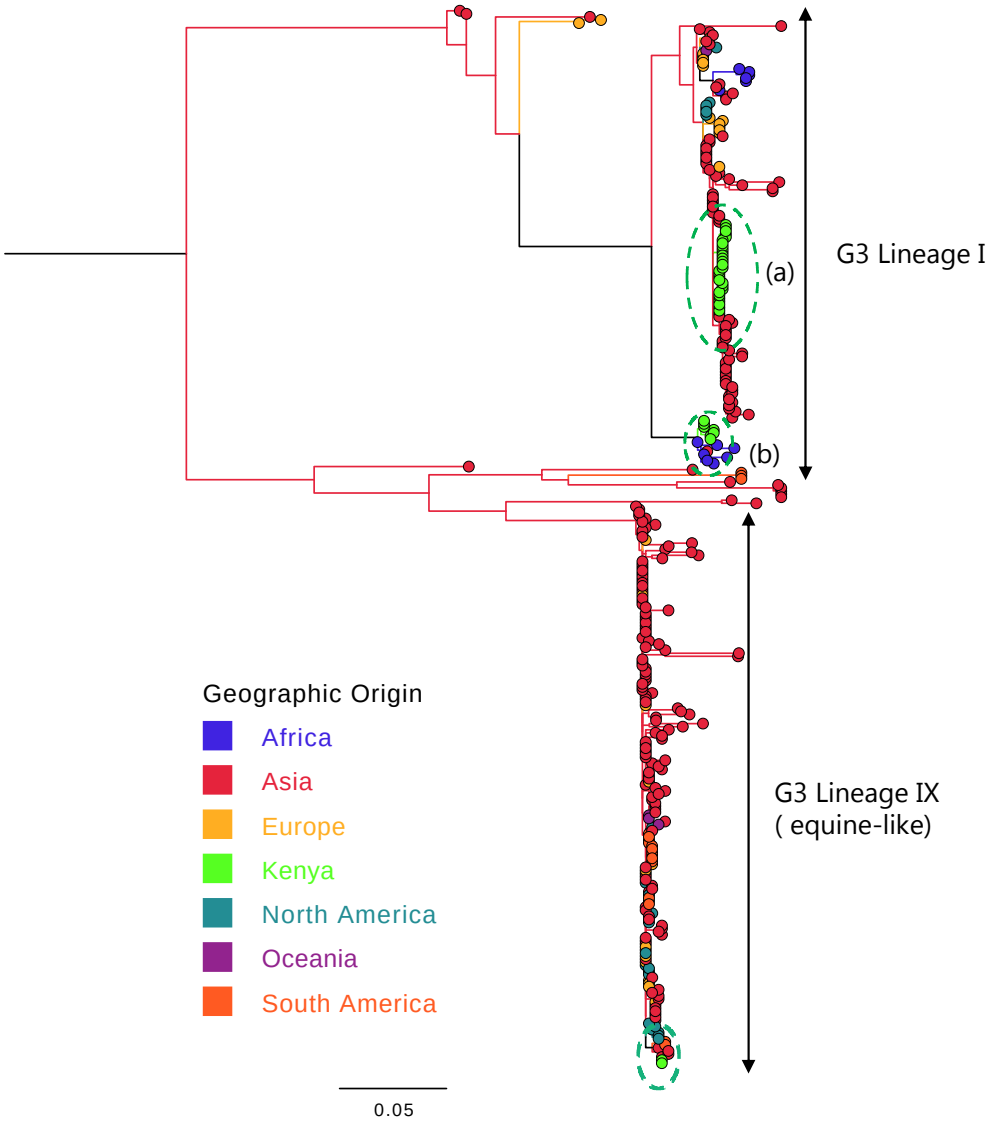

Figure 5. Global maximum likelihood phylogeny of nucleotide sequences G3 strains sampled between 2013-2018. The phylogenetic tree was reconstructed from 345 VP7 sequences of G3 type (308 collated from GenBank sampled across 26 countries and 37 G3 viruses sequenced in the current study) to determine the lineage and global context of the Kilifi sequences. The countries included were Australia, Belarus, Brazil, China, Dominican Republic, Ethiopia, Hungary, India, Indonesia, Italy, Japan, Kenya, South Korea, Kuwait, Nigeria, Pakistan, Peru, Russia, Spain, Taiwan, Thailand, USA, Uganda and Viet Nam. The filled circles show the taxa and are colored differently by the continent of sampling. The taxa for Kenya are colored green and highlighted with an oval dashed green.

For each of the three Kilifi G3 genetic clusters we explored their closest genetic relative in the global dataset by network reconstructions, **Figure 6**. For the Kilifi Clu_3/G3P[8] the closest similar sequences were from India (G3P[8] collected in 2016) and Singapore (G3P[8] collected in 2016) that had 2 nucleotide differences **Figure 6, panel (a)**. For the Kilifi Clu_4/G3P[8] (the equine-like G3 Lineage) the closest relative was from Taiwan (G3P[8] collected in 2016) with zero nucleotide difference in the sequenced region **Figure 6, panel (b)**. For the Kilifi Clu_5/G3P[8] the closest relatives

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

were from Kenya (G3P[6] collected in 2014) and Uganda (G3P[6] collected in 2013) that had zero and 2 nucleotide difference, respectively, **Figure 6, panel (c)**. Overall, within these 3 major global G3 genetic clusters, clustering by country was common.

Table 3: The global distribution of the identified G3 global genetic clusters

Global cluster name	Number of sequences	Includes Kilifi cluster	Years detected	Countries detected
Global/Clu_1	6	-	2015, 2016	Ethiopia, Kuwait
Global/Clu_2	15	Clu_5/G3P[8]	2013, 2014, 2016, 2017, 2018	Ethiopia, Kenya, Nigeria, Uganda, Indonesia
Global/Clu_3	1	-	2013	
Global/Clu_4	5	-	2013	
Global/Clu_5	2	-	2013	China
Global/Clu_6	119	Clu_3/G3P[8]	2013, 2014, 2015, 2016, 2018	India, Pakistan, China, Taiwan, Korea, Japan, Singapore, Australia, Italy, Spain, Russia, USA, Dominican, Ethiopia, Kenya
Global/Clu_7[#]	**179**	**Clu_4/G3P[8]**	**2013, 2014, 2015, 2016, 2017, 2018**	**Australia, Taiwan, Japan, Indonesia, Thailand, USA, Dominican, Brazil, Italy, Germany, Hungary, Spain, Kenya**
Global/Clu_8	2	-	2013, 2014	Belarus
Global/Clu_9	1	-	2016	India
Global/Clu_10	**1[#]**	**-**	**2016**	**Thailand**
Global/Clu_11	**1[#]**	**-**	**2015**	**Thailand**
Global/Clu_12	1	-	2013	China
Global/Clu_13	2	-	2015, 2016	India
Global/Clu_14	2	-	2014, 2016	South Korea, Vietnam
Global/Clu_15	1	-	2016	Kuwait
Global/Clu_16	1	-	2015	Thailand
Global/Clu_17	**1[#]**	**-**	**2016**	**Thailand**
Global/Clu_18	**1[#]**	**-**	**2016**	**Thailand**
Global/Clu_19	3	-	2014	Peru
Global/Clu_20	1	-	2014	China

[#]Members of the equine-like G3 strain (Lineage IX)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

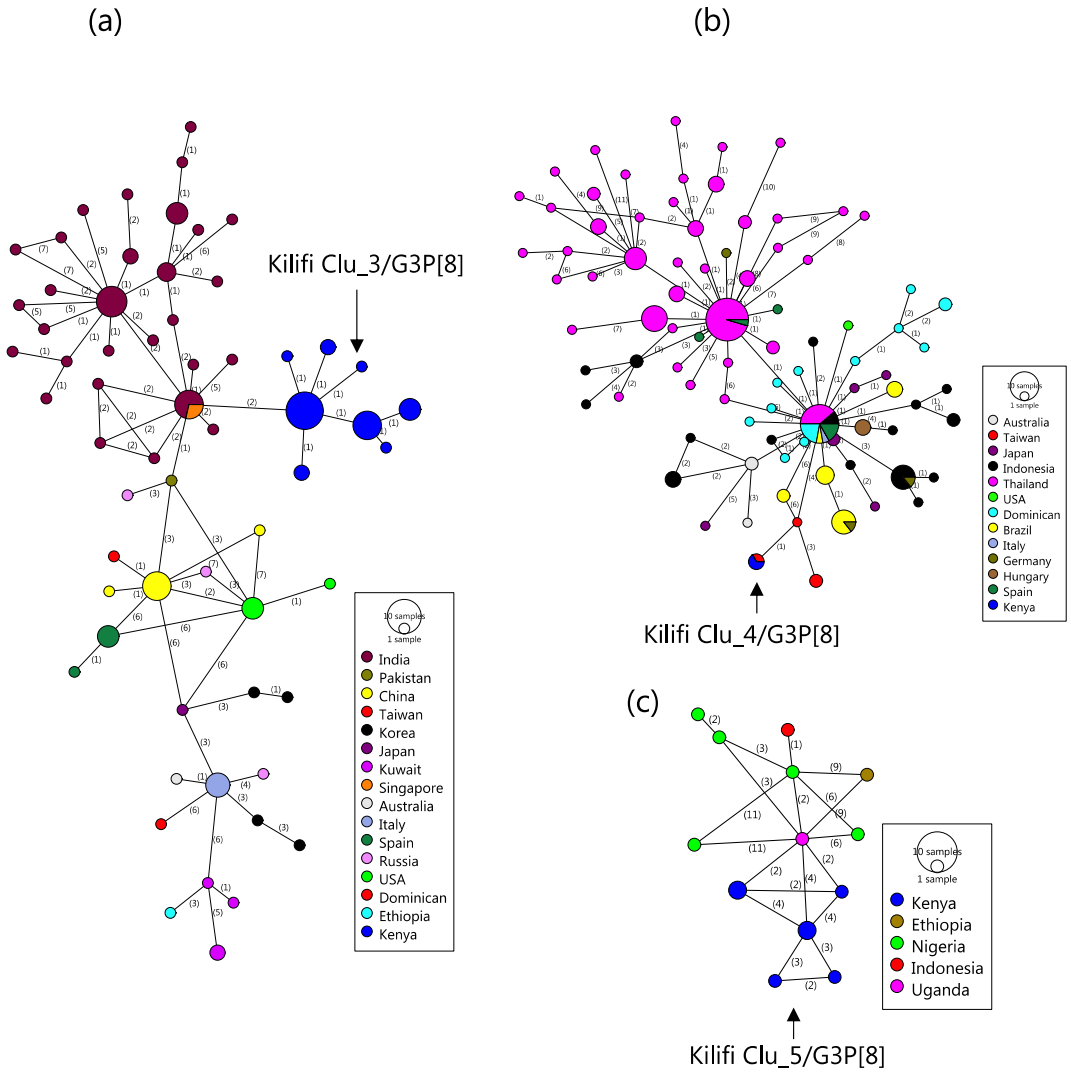

Figure 6. Haplotype network showing relationships of the identified global G3 lineages that included Kilifi viruses. Panel (a) shows the network for Lineage I cluster viruses that included the Kilifi Clu_3/G3P[8] strains. The vertices represent the VP7 haplotypes. The size of the vertex is proportional to the number of haplotypes (identical sequences) and is colored by the country of sampling. The numbers shown on the edges represent the number of nucleotide changes from one vertex (haplotype) to the next. Panel (b) and (c) have the same description as panel (a) above but represent Lineage IX (equine-like G3) cluster that included Kilifi Clu_4 G3P[8] and the Lineage I cluster that included Kilifi Clu_5 G3P[8] sequences, respectively.

3. Discussion

Four years after Kenya introduced Rotarix® vaccine into its NIP, multiple RVA GP genotypes circulated during the 2018 season in Kilifi, Kenya, with the G3P[8] genotype predominating at 67.2%. At this study site, the preceding two years (2016 and 2017) were dominated by the G2P[4] and G1P[8] genotypes, respectively, with only six cases of G3P[8] detected from September 2009 to December 2017 [31, 32]. The G3P[8] strains are partially heterotypic to the monovalent Rotarix® vaccine, which is comprised of an attenuated G1P[8] strain. During 2018, this local G3P[8] predominance is consistent with the previously documented season-to-season spatial-temporal fluctuations in the prevalence of RVA genotypes [12], hypothesized to be driven by the prevailing population-level immunity derived from natural infections and the use of vaccines [14].

Vaccination records were available for 70.6% of the children with an RVA positive test. Of these, 92.7% were age-eligible to have received the two doses of Rotarix® vaccine and, in that subgroup,

the vast majority (80.6%) had indeed received the full 2-dose series. However, overall, the vaccination status of these children did not appear to predict either their RVA diagnosis result or the infecting GP genotype. These findings, albeit from a single season and site, suggest that for these children who acquired an RVA infection despite one or two-dose vaccination, host factors rather than viral characteristics or vaccine composition may explain the vaccine failures. A follow-up study is planned.

At least seven distinct genetic clusters constituted the 2018 coastal Kenya RVA season. The VP7 sequences showed greater genetic diversity and provided a better phylogenetic resolution compared to the VP4 sequences. Each of the identified G types corresponded to a single genetic cluster except G3 viruses that segregated into three genetically distinct clusters. Strikingly, some samples with different G types yielded identical VP4 sequences, indicating that some of the children may have been infected by reassortant viruses or harbored mixed infections [21]. Our analyses improve understanding on the recent composition and transmission patterns of local RVA seasons, providing insight into the design of final stretch RVA control strategies following vaccine introduction.

Several recent studies have reported the increased proportion of G3P[8] strains e.g. in Australia [14], Japan [33], Thailand [22], Indonesia [24], Pakistan [34], Dominican Republic [21], Brazil [27], Spain [30] and Mozambique [35], Malawi [36], and Botswana [37]. The global G3 sequences available from GenBank showed extensive genetic diversity. The significance of this diversity in relation to human immune recognition should be investigated. Notably, recent years have also observed the emergence and global spread of a new G3 lineage named equine-like G3, of putative equine origin, assigned G3 Lineage IX [21]. Strains of G3 Lineage IX were first detected in 2013 in Japan and have since been widely detected in several other countries (Australia [29], Taiwan (unpublished data in GenBank), Indonesia [24], Thailand [22], USA [26], Dominican Republic [21], Brazil [27], Italy [23], Germany [25], Hungary [28] and Spain [30]). Our study is the first to document African children infection with the G3 Lineage IX. Continued surveillance to monitor whether this particular strain becomes endemic in Kenya and the wider Africa continent in the face of increased RVA vaccine coverage is important to optimize RVA vaccine-mediated control. Notably, recent studies in Botswana [37], Mozambique [35], Malawi [36]and Ethiopia [38] reported increased prevalence of G3 type viruses but sequencing data from these studies are not yet available.

Based on sequence data deposited in GenBank, the predominant Kilifi G3 cluster (Clu_3/G3P[8]) was the second most common genetic cluster globally. The closest sequences were from Singapore and India, both countries that did not yet have RVA vaccine in their NIP in 2018. The second most prevalent Kilifi G3 genetic cluster was Clu_5/G3P[8]. Notably, this cluster has not been detected frequently around the globe and the closest genetic links were Kenyan strains collected in Kiambu County (Central province) in July and August 2014 [39], and strains from Ethiopia (collection date: April 2016 [40]) and Uganda (collection date: Jan 2013 [41]), neighboring countries which included RVA vaccines in their NIP in 2013 and 2018, respectively. Although the Kilifi Clu_4/G3P[8] (equine-like G3 Lineage) was the least prevalent locally, it was the most prevalent globally. The closest relatives to the Kenyan strains were from Taiwan, a country yet to introduce RVA vaccination.

This study had some limitations. First, the sequence data from the cohort represents a single site and one season. Second, we only sequenced portions of the VP4 and VP7 segments. Whereas these data were adequate to assign genotypes, lineages and estimate the number of genetic clusters, whole genome sequences provide a better resolution in examining reassortment events, evolution in internal genes and studying genetic clusters [18, 21, 42]. Third, to determine the origin and pathways of spread of the imported genetic clusters, background sequence data from more countries and including populations neighboring coastal Kenya would have been ideal. Unfortunately, sequence data in public sequence databases to facilitate such phylogeographic analysis are currently limited. Fourth, the absence of significant epidemiological data for some variables e.g. vaccine status for ~30% of the RVA positive children and geographic origin for children from outside the KHDSS area limited our analyses

In conclusion, the finding that >20% of diarrheal stools from children admitted to KCH with diarrhea in 2018 were RVA positive highlights that RVA is still a significant contributor to severe childhood diarrhea in coastal Kenya, despite the introduction of Rotarix® into Kenya's NIP in 2014.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2020 doi:10.20944/preprints202010.0470.v1

The cross-continent detection of the emerging equine-like G3 viruses and other typical human G3 strains demonstrates the global nature of RVA transmission. Strikingly, strains found circulating in the Kilifi population were most closely related to strains circulating in countries that were yet to introduce RVA vaccines into their NIP. This observation reminds of the global connectedness with regard to pathogen movement and emphasizes the importance of vaccinating all eligible populations across the world, as failure to do so builds a reservoir for strains that continue to seed transmission in vaccinated populations. Identifying factors responsible for RVA vaccine underperformance in low-income settings is a priority research area that may support efforts to further reduce RVA burden. Our study did not ascertain that viral genetic diversity is a contributor to the vaccine underperformance in this setting. Studies investigating the relationship between RVA vaccine immunogenicity and infant characteristics such as malnutrition, age at first RVA dose, concomitant receipt of oral polio vaccine (OPV), enteric co-infections and enteric dysbiosis may provide better insight into RVA vaccine performance characteristics.

4. Materials and Methods

4. 1 Study population and location

KCH is the main referral hospital in Kilifi County (population size ~1.5 million people). The major economic activities in the county are subsistence farming, fishing and tourism [43]. An area around KCH (~900 km^2 with a population of ~300,000 people) is monitored by the KWTRP and is known as the KHDSS area [43], **Figure 4**. A high proportion of the patients seeking care at the KCH are KHDSS area residents [43]. Vaccination data of admitted children is collected using an electronic registry [8, 44, 45].

In the current analysis, stool samples were collected from eligible and consented pediatric patients admitted to KCH between January and December 2018 (the surveillance period), as part of the ongoing rotavirus surveillance program [8, 32, 44]. All children aged <13 years old admitted with diarrhea (defined as passing three or more watery stools in the last 24-hours) were eligible for inclusion [8, 32, 44]. Following a review of demographic and clinical data collected by a clinical staff, parents or caregivers of eligible children were approached for consent, and a single stool sample was collected. The samples were immediately transferred into a cool box with 'ice' blocks before transportation to the KWTRP for RVA testing and long-term storage at -80^o C.

4.2 Specimen laboratory processing

RVA in the stool samples was detected using ProSpecT™ enzyme immunoassay (EIA) kit (Oxoid, Basingstoke, UK) following the manufacturer's instructions. RVA positive samples were amplified in the VP4 and VP7 segments using One-step Reverse Transcriptase PCR Kit (Qiagen, California, USA) using previously published primers [32]. Successful amplification of the target regions was confirmed by presence of expected bands (VP4: 660bp and VP7: 881bp) following gel electrophoresis of the PCR products. Products from successful PCRs were purified using GFX DNA purification kit (GFX-Amersham, Amersham, UK) and sequenced using Big Dye Terminator 3.1 (Applied Biosystems, Foster City California, USA) chemistry. The primers used during PCR amplification were used for sequencing on an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Foster City, California, USA).

4.3. Genotyping and phylogenetic analysis

The sequence reads were assembled using Sequencher v5.4.6 (Gene Codes Corp Inc., Ann Arbor, MI, USA). Nucleotide (nt) sequence alignments were prepared using MAFFT v7.222 and visualized using Aliview v1.8. G and P genotypes were determined using Virus Pathogen Resource (ViPR) online classification tool [46]. The best nt substitution model for the alignments were determined IQ-Tree v1.6.6 [47]. Phylogenetic trees were reconstructed using the maximum likelihood (ML) method in RaxML v8.2.12 [48] and MEGA v7 [49]. Support for the tree branching patterns was evaluated by 1000 bootstrap iterations.

4.4. Genetic clusters

Molecular genetic clusters were defined from the distribution of pairwise nt differences of VP7 segment sequences. Pairwise nt differences were determined using pairsnp (https://github.com/gtonkinhill/pairsnp/). Viruses within the same molecular genetic clusters were those which pairwise nt differences occurred within the first modal distribution. Using this threshold, clusters were identified using the USEARCH algorithm [50]. Single nucleotide polymorphic (SNP) positions in alignments were assessed using parseSNP [51]. The minimum spanning networks between the RVA positive patients were reconstructed using POPART v1.70 program [52].

4.5. Comparison dataset

The phylogenetic context of the locally predominant genotype in global RVA populations was investigated by co-analysis with similar G type strains sequence data deposited in GenBank. The search in GenBank was conducted in September 2019. The criteria for comparison data inclusion were (i) detection in a human stool/rectal swab specimen, (ii) sequence fully overlapping with the VP7 region sequenced for the Kilifi viruses, (iii) information on country and date of sampling available and (iv) sample collected between 2013-2018.

4.6. Statistical analysis

Numerical data were analyzed in STATA v15.1. Continuous variables were summarized using various measures of dispersion. Differences between groups were assessed using a t-test or Wilcoxon rank-sum test. Binary data were summarized using proportions and comparison between groups made using χ^2 or Fisher's exact test. The 95% CI were presented for proportions and standard deviation for means. A p-value of <0.05 was considered significant.

4.7. Ethical statement

Before enrollment into the study, all eligible subjects' parent or guardian were requested for informed written consent to allow them participate in the study. The Scientific Ethics Review Unit (SERU) board that sits at KEMRI, Nairobi, approved the study protocols (SERU#3049). The study was conducted in accordance with the Declaration of Helsinki.

Data availability: Partial sequences for the VP7 and VP4 segments reported in this work have been deposited to GenBank database under the sequence accession numbers MN194408-MN194485 for VP7 and MN194325-MN194364 for VP4.

Author Contributions: Conceptualization, C.N.A, D.J.N, M.J.M and R.F.B; methodology, D.J.N, C.NA and M.J.M.; software, M.M & C.N.A.; validation, C.NA , E.G. and M.M.; formal analysis, C.N.A and M.J.M.; investigation, C.N.A, M.J.M and D.J.N and E.G., resources, R.F.B, J.E.T, U.D.P and D.J.N.; data curation, N.M, M.J.M and C.N.A.; writing—original draft preparation, M.J.M and C.N.A.; writing—review and editing, R.O, J.R.V, R.F.B, J.E.T, U.D.P and D.J.N.; visualization, C.N.A and M.J.M.; supervision, D.J.N and C.N.A.; project administration, E.G. R.O., M.J.M., N.M and D.J.N..; funding acquisition, D.J.N, R.O., J.R.V, J.E.T, U.D.P, R.F.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Gavi, The Vaccine Alliance through Emory University, to the Rotavirus Immunization Program and Evaluation in Kenya (RIPEK) and the Wellcome Trust (grant numbers 203077 and 102975). CNA is 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. Views expressed in this publication are those of the authors and not necessarily those of Gavi, AAS, NEPAD Agency, Wellcome Trust or the UK government.

Acknowledgments: We thank the participants who provided samples for analysis, and the Clinical and Laboratory staff at Virus Epidemiology and Control group, KEMRI Wellcome Trust Research Programme for collection and processing the samples. This work is published with permission from Director KEMRI.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention (CDC) or the authors' affiliated institutions.

References

1. Burnett E, Jonesteller CL, Tate JE, Yen C, Parashar UD. Global Impact of Rotavirus Vaccination on Childhood Hospitalizations and Mortality From Diarrhea. J Infect Dis **2017**; 215:1666-72.

2. Steele AD, Groome MJ. Measuring Rotavirus Vaccine Impact in sub-Saharan Africa. Clin Infect Dis **2019**.

3. Operario DJ, Platts-Mills JA, Nadan S, et al. Etiology of Severe Acute Watery Diarrhea in Children in the Global Rotavirus Surveillance Network Using Quantitative Polymerase Chain Reaction. J Infect Dis **2017**; 216:220-7.

4. Iturriza-Gomara M, Jere KC, Hungerford D, et al. Etiology of Diarrhea Among Hospitalized Children in Blantyre, Malawi, Following Rotavirus Vaccine Introduction: A Case-Control Study. J Infect Dis **2019**; 220:213-8.

5. 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**; 172:958-65.

6. Steele AD, Victor JC, Carey ME, et al. Experiences with rotavirus vaccines: can we improve rotavirus vaccine impact in developing countries? Hum Vaccin Immunother **2019**; 15:1215-27.

7. Willame C, Vonk Noordegraaf-Schouten M, Gvozdenovic E, et al. Effectiveness of the Oral Human Attenuated Rotavirus Vaccine: A Systematic Review and Meta-analysis-2006-2016. Open Forum Infect Dis **2018**; 5:ofy292.

8. Khagayi S, Omore R, Otieno GP, et al. Effectiveness of monovalent rotavirus vaccine against hospitalization with acute rotavirus gastroenteritis in Kenyan children. Clin Infect Dis **2019**.

9. Walker JL, Andrews NJ, Atchison CJ, et al. Effectiveness of oral rotavirus vaccination in England against rotavirus-confirmed and all-cause acute gastroenteritis. Vaccine X **2019**; 1:100005.

10. Nair N, Feng N, Blum LK, et al. VP4- and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. Sci Transl Med **2017**; 9.

11. RCWG. Rotavirus Classification Working Group: Newly assigned genotypes. Available at: https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification/rcwg. Accessed 07-Jan-2020 2020.

12. Sadiq A, Bostan N, Yinda KC, Naseem S, Sattar S. Rotavirus: Genetics, pathogenesis and vaccine advances. Rev Med Virol **2018**; 28:e2003.

13. Leshem E, Lopman B, Glass R, et al. Distribution of rotavirus strains and strain-specific effectiveness of the rotavirus vaccine after its introduction: a systematic review and meta-analysis. Lancet Infect Dis **2014**; 14:847-56.

14. Roczo-Farkas S, Kirkwood CD, Cowley D, et al. The Impact of Rotavirus Vaccines on Genotype Diversity: A Comprehensive Analysis of 2 Decades of Australian Surveillance Data. J Infect Dis **2018**; 218:546-54.

15. Burke RM, Tate JE, Barin N, et al. Three Rotavirus Outbreaks in the Postvaccine Era - California, 2017. MMWR Morb Mortal Wkly Rep **2018**; 67:470-2.

16. Pitzer VE, Bilcke J, Heylen E, et al. Did Large-Scale Vaccination Drive Changes in the Circulating Rotavirus Population in Belgium? Sci Rep **2015**; 5:18585.

17. Ianiro G, Micolano R, Di Bartolo I, Scavia G, Monini M. Group A rotavirus surveillance before vaccine introduction in Italy, September 2014 to August 2017. Euro Surveill **2019**; 24.

18. Ogden KM, Tan Y, Akopov A, et al. Multiple Introductions and Antigenic Mismatch with Vaccines May Contribute to Increased Predominance of G12P[8] Rotaviruses in the United States. J Virol **2019**; 93.

19. Santos N, Hoshino Y. Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. Rev Med Virol **2005**; 15:29-56.

20. Tacharoenmuang R, Komoto S, Guntapong R, et al. High prevalence of equine-like G3P[8] rotavirus in children and adults with acute gastroenteritis in Thailand. J Med Virol **2019**.

21. Katz EM, Esona MD, Betrapally NS, et al. Whole-gene analysis of inter-genogroup reassortant rotaviruses from the Dominican Republic: Emergence of equine-like G3 strains and evidence of their reassortment with locally-circulating strains. Virology **2019**; 534:114-31.

22. Tacharoenmuang R, Komoto S, Guntapong R, et al. High prevalence of equine-like G3P[8] rotavirus in children and adults with acute gastroenteritis in Thailand. J Med Virol **2020**; 92:174-86.

23. Esposito S, Camilloni B, Bianchini S, et al. First detection of a reassortant G3P[8] rotavirus A strain in Italy: a case report in an 8-year-old child. Virol J **2019**; 16:64.

24. Utsumi T, Wahyuni RM, Doan YH, et al. Equine-like G3 rotavirus strains as predominant strains among children in Indonesia in 2015-2016. Infect Genet Evol **2018**; 61:224-8.

25. Pietsch C, Liebert UG. Molecular characterization of different equine-like G3 rotavirus strains from Germany. Infect Genet Evol **2018**; 57:46-50.

26. Perkins C, Mijatovic-Rustempasic S, Ward ML, Cortese MM, Bowen MD. Genomic Characterization of the First Equine-Like G3P[8] Rotavirus Strain Detected in the United States. Genome Announc **2017**; 5.

27. Guerra SFS, Soares LS, Lobo PS, et al. Detection of a novel equine-like G3 rotavirus associated with acute gastroenteritis in Brazil. J Gen Virol **2016**; 97:3131-8.

28. Doro R, Marton S, Bartokne AH, et al. Equine-like G3 rotavirus in Hungary, 2015 - Is it a novel intergenogroup reassortant pandemic strain? Acta Microbiol Immunol Hung **2016**; 63:243-55.

29. Cowley D, Donato CM, Roczo-Farkas S, Kirkwood CD. Emergence of a novel equine-like G3P[8] inter-genogroup reassortant rotavirus strain associated with gastroenteritis in Australian children. J Gen Virol **2016**; 97:403-10.

30. Arana A, Montes M, Jere KC, Alkorta M, Iturriza-Gomara M, Cilla G. Emergence and spread of G3P[8] rotaviruses possessing an equine-like VP7 and a DS-1-like genetic backbone in the Basque Country (North of Spain), 2015. Infect Genet Evol **2016**; 44:137-44.

31. Mwanga MJ, Owor BE, Ochieng JB, et al. Rotavirus group A genotype circulation patterns across Kenya before and after nationwide vaccine introduction, 2010-2018. **Under review**.

32. 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. Wellcome Open Res **2018**; 3:150.

33. Thongprachum A, Chan-it W, Khamrin P, et al. Reemergence of new variant G3 rotavirus in Japanese pediatric patients, 2009-2011. Infect Genet Evol **2013**; 13:168-74.

34. Umair M, Abbasi BH, Sharif S, et al. High prevalence of G3 rotavirus in hospitalized children in Rawalpindi, Pakistan during 2014. PLoS One **2018**; 13:e0195947.

35. João ED, Munlela B, Chissaque A, et al. Molecular Epidemiology of Rotavirus A Strains Pre- and Post-Vaccine (Rotarix®) Introduction in Mozambique, 2012–2019: Emergence of Genotypes G3P[4] and G3P[8]. Pathogens **2020**; 9:671.

36. Mhango C, Mandolo JJ, Chinyama E, et al. Rotavirus Genotypes in Hospitalized Children with Acute Gastroenteritis Before and After Rotavirus Vaccine Introduction in Blantyre, Malawi, 1997 – 2019. The Journal of Infectious Diseases **2020**.

37. Mokomane M, Esona MD, Bowen MD, et al. Diversity of Rotavirus Strains Circulating in Botswana before and after introduction of the Monovalent Rotavirus Vaccine. Vaccine **2019**; 37:6324-8.

38. Abebe A, Getahun M, Mapaseka SL, et al. Impact of rotavirus vaccine introduction and genotypic characteristics of rotavirus strains in children less than 5years of age with gastroenteritis in Ethiopia: 2011-2016. Vaccine **2018**; 36:7043-7.

39. Wandera EA, Komoto S, Mohammad S, et al. Genomic characterization of uncommon human G3P[6] rotavirus strains that have emerged in Kenya after rotavirus vaccine introduction, and pre-vaccine human G8P[4] rotavirus strains. Infect Genet Evol **2019**; 68:231-48.

40. Gelaw A, Pietsch C, Liebert UG. Molecular epidemiology of rotaviruses in Northwest Ethiopia after national vaccine introduction. Infect Genet Evol **2018**; 65:300-7.

41. Bwogi J, Jere KC, Karamagi C, et al. Whole genome analysis of selected human and animal rotaviruses identified in Uganda from 2012 to 2014 reveals complex genome reassortment events between human, bovine, caprine and porcine strains. PLoS One **2017**; 12:e0178855.

42. Jere KC, Chaguza C, Bar-Zeev N, et al. Emergence of Double- and Triple-Gene Reassortant G1P[8] Rotaviruses Possessing a DS-1-Like Backbone after Rotavirus Vaccine Introduction in Malawi. J Virol **2018**; 92.

43. Scott JA, Bauni E, Moisi JC, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). Int J Epidemiol **2012**; 41:650-7.

44. Otieno GP, Bottomley C, Khagayi S, et al. Impact of the introduction of rotavirus vaccine on hospital admissions for diarrhoea among children in Kenya: A controlled interrupted time series analysis. Clin Infect Dis **2019**.

45. Adetifa IMO, Bwanaali T, Wafula J, et al. Cohort Profile: The Kilifi Vaccine Monitoring Study. Int J Epidemiol **2017**; 46:792-h.

46. Pickett BE, Greer DS, Zhang Y, et al. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses **2012**; 4:3209-26.

47. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268-74.

48. Stamatakis A. Using RAxML to Infer Phylogenies. Curr Protoc Bioinformatics **2015**; 51:6 14 1-.

49. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **2016**; 33:1870-4.

50. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **2010**; 26:2460-1.

51. Taylor NE, Greene EA. PARSESNP: A tool for the analysis of nucleotide polymorphisms. Nucleic Acids Res **2003**; 31:3808-11.

52. Leigh JW, Bryant D. POPART: full-feature software for haplotype network construction Methods in Ecology and Evolution **2015**.