Emergence of double- and triple-gene reassortant G1P[8] rotaviruses possessing a DS-1-like backbone post rotavirus vaccine introduction in Malawi

Khuzwayo C. Jere a,b,*, Chrispin Chaguza a,b, Naor Bar-Zeev a,b,*, Jenna Lowe a, Chikondi Peno b, Benjamin Kumwenda b, Osamu Nakagomi a,d, Jacqueline E. Tate e, Umesh D. Parashar e, Robert S. Heyderman b,f, Neil French b,c, Nigel A. Cunliffe c,*, Miren Iturriza-Gomara c,e, for the VACSURV Consortium.

Institute of Infection & Global Health, University of Liverpool, Liverpool, UK a;
Malawi-Liverpool-Wellcome Trust Clinical Research Programme/Department of Medical Laboratory Sciences, College of Medicine, University of Malawi, Blantyre, Malawi b; Centre for Global Vaccine Research, Institute of Infection & Global Health, University of Liverpool, UK c; Graduate School of Biomedical Sciences, Nagasaki University, Japan d; Epidemiology Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA e; Division of Infection and Immunity, University College London, UK f; NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, UK f

*Corresponding author: University of Liverpool, Institute of Infection and Global Health, Ronald Ross Building, 8 West Derby Street, Liverpool, L69 7BE, United
*Authors contributed equally.

Running Head: Emergence of atypical G1P[8] rotaviruses in Malawi

Key Words: Rotavirus; Phylodynamics; Genome reassortment; Lineage turnover; whole genome sequencing
Abstract

To combat the high burden of rotavirus gastroenteritis, multiple African countries have introduced rotavirus vaccines into their childhood immunisation programmes. Malawi incorporated a G1P[8] rotavirus vaccine (Rotarix™) into its immunisation schedule in 2012. Utilising a surveillance platform of hospitalised rotavirus gastroenteritis cases, we examined the phylodynamics of G1P[8] rotavirus strains that circulated in Malawi before (1998 – 2012) and after (2013 – 2014) vaccine introduction. Analysis of whole genomes obtained through next generation sequencing revealed that all randomly-selected pre-vaccine G1P[8] strains sequenced (n=32) possessed a Wa-like genetic constellation, whereas post-vaccine G1P[8] strains (n=18) had a DS-1-like constellation. Phylodynamic analyses indicated that post-vaccine G1P[8] strains emerged through reassortment events between human Wa- and DS-1-like rotaviruses that circulated in Malawi from the 1990’s, hence classified as atypical DS-1-like reassortants. The time to the most recent common ancestor for G1P[8] strains was from 1981-1994; their evolutionary rates ranged from 9.7 x 10^{-4}–4.1 x 10^{-3} nucleotide/substitutions/site/year. Three distinct G1P[8] lineages chronologically replaced each other between 1998 and 2014. Genetic drift was the likely driver for lineage turnover in 2005, whereas replacement in 2013 was due to reassortment. Amino acid substitution within the outer glycoprotein VP7 of G1P[8] strains had no impact on the structural conformation of the antigenic regions, suggesting that it is unlikely that they would affect recognition by vaccine-induced neutralizing antibodies. While the emergence of DS-1-like G1P[8] rotavirus reassortants in Malawi was therefore likely due to
natural genotype variation, vaccine effectiveness against such strains needs careful evaluation.
Importance

The error-prone RNA-dependent RNA polymerase and the segmented RNA genome predispose rotaviruses to genetic mutation and genome reassortment, respectively. These evolutionary mechanisms generate novel strains and have the potential to lead to the emergence of vaccine-escape mutants. While multiple African countries have introduced rotavirus vaccine, there are few data describing the evolution of rotaviruses that circulated before and after vaccine introduction. We report the emergence of atypical DS-1-like G1P[8] strains during the post-vaccine era in Malawi. Three distinct G1P[8] lineages circulated chronologically from 1998–2014; mutation and reassortment drove lineage turnover in 2005 and 2013, respectively. Amino acid substitutions within the outer capsid VP7 glycoprotein did not affect the structural conformation of mapped antigenic sites, suggesting limited effect in recognition of G1 specific vaccine-derived antibodies. The genes that constitute the remaining genetic backbone may play important roles in immune evasion, and vaccine effectiveness against such atypical strains needs careful evaluation.
Introduction

Diarrhoea is a leading cause of mortality in children under the age of five years globally (1, 2). The majority of hospitalisations and deaths in infants due to severe dehydrating diarrhoea are caused by group A rotaviruses (RVA) (3). The World Health Organization (WHO) recommended universal introduction of rotavirus vaccines in 2009 particularly in countries where diarrhoea mortality is high (4). A global decline from 528,000 to 215,000 in rotavirus-associated deaths per year amongst children <5 years of age has been reported between 2000 and 2013, and live-attenuated oral rotavirus vaccines (Rotarix™; RV1 and RotaTeq®; RV5) have now been incorporated into national immunization programs of over 60 countries worldwide (5).

RVA are members of the Reoviridae virus family. They are enveloped icosahedral viruses that contain a triple-layered capsid encasing 11 genome segments of double-stranded RNA (dsRNA). The rotavirus genome encodes six structural (VP1–VP4, VP6 and VP7) and five to six non-structural proteins (NSP1–NSP5/NSP6) (6). Nucleotide homology cut-off values of the open reading frame (ORF) for each genome segment are used to classify rotavirus strains on the basis of the whole genome composition (7, 8). To date, 35 G (VP7), 50 P (VP4), 26 I (VP6), 21 R (VP1), 19 C (VP2), 19 M (VP3), 30 A (NSP1), 20 N (NSP2), 21 T (NSP3), 26 E (NSP4), and 21 H (NSP5) genotypes have been described (8-12).

G1 – G4, G9 and G12 in association with P[4], P[6] or P[8] are the predominant genotypes associated with human rotavirus infection worldwide (6,
Although several G and P genotype combinations have been detected among human rotaviruses, the genotypes for the other nine genes are limited to predominantly genotype 1 (I1-R1-C1-M1-A1-N1-T1-E1-H1; Wa-like) and genotype 2 (I2-R2-C2-M2-A2-N2-T2-E2-H2; DS-1-like) (16). For instance, typically RVAs G1P[8], G3P[8], G4P[8], G9P[8] and G12P[8] have a Wa-like genotype constellation, whereas G2P[4] and G8P[4] or G8P[6] strains usually possess a DS-1-like constellation (16-18). The segmented RNA genome of rotaviruses and their error prone RNA-dependent RNA polymerase, which lacks proof-reading activity (6), allows various evolutionary mechanisms including genetic mutation, recombination and reassortment. This leads to the emergence of distinct lineages within individual genotypes, or reassortant viruses containing segments from different progenitor strains (6, 19, 20).

Novel double-reassortant DS-1-like G1P[8] rotavirus strains in Africa, which became predominant following vaccine introduction. The
evolutionary forces that were associated with the emergence of the atypical G1P[8] rotavirus strains were also determined.

Results

Emergence of reassortant DS-1-like G1P[8] rotavirus strains

All pre- (33.4%: 1,634/4,888) and post-vaccine (22.6%: 477/2,109) rotavirus-positive stools collected from children presenting with acute severe diarrhoea at QECH were genotyped as part of on-going rotavirus surveillance (27-30). Amongst multiple strains were characterised, and G1P[8] rotaviruses were consistently predominant strains that were detected each year before (39.4%: 554/1406) and after (31.4%: 95/303) vaccine introduction (Fig. 1a and S1). Whole genome sequences of 32 pre-vaccine G1P[8] strains (collected between 1998 and 2012), and 18 post-vaccine G1P[8] strains (collected from 2013 to 2014) were successfully generated (see supplementary Table 1 for yearly distribution). Interruption of surveillance in 2010 meant that no G1P[8] strains were available between 2010 and 2011. Among the pre-vaccine G1P[8] strains, 31 had the G1-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1 genotype constellation, hence designated as Wa-like G1P[8], whereas one was G1-P[8]-I1-R1-C1-M1-A1-N1-T2-E1-H1, hence designated as a mono-reassortant Wa-like G1P[8] strain. In contrast, 16/18 of the post-vaccine G1P[8] strains had a DS-1-like genotype constellation (G1-P[8]-I2-R2-C2-M2-A2-N2-T2-E2-H2) hence were designated as atypical double-reassortant DS-1-like G1P[8] strains. The remaining two post-vaccine G1P[8] strains (BID1LN and BID230) had Wa-like
VP1 (R1) and Wa-like NSP3 (T1) genes, respectively, hence were designated as atypical triple-reassortant DS-1-like G1P[8] strains (Fig. 1b and Table S1).

**Atypical post-vaccine G1P[8] strains emerged through genome reassortment between Wa-like and DS-1-like human rotaviruses**

Whilst the concatenated sequences of all 11 genome segments of pre- and post-vaccine G1P[8] rotaviruses clustered with prototype Wa- and DS-1-like strains, respectively (Fig. 1b), phylogenetic networks constructed using concatenated whole genome sequences of G1P[8] strains characterised in the revealed frequent reassortment events between strains from the same or different network clusters (Fig. 1c). The Malawi G1P[8] strains were distributed into three main phylogenetic network clusters (L1, L2 and L3), which also contained multiple sub-clusters (shaded within main clusters in Fig. 1c). L1 phylogenetic cluster contained strains that circulated from 1998 to 2004; strains detected from 2005 to 2012 grouped into L2 cluster, while L3 cluster only consisted of post-vaccine G1P[8] strains that circulated from 2013 to at least 2014 in Blantyre.

The phylogenetic relationship of Malawi G1P[8] strains was inferred using the Maximum-Likelihood method for which complete nucleotide sequences of RVAs available in the GenBank and sequences of Wa-like (VP4 and VP7 only) and DS-1-like (the other nine genes) strains from Malawi were used. The DS-1-like genes of the post-vaccine G1P[8] reassortants closely clustered with those of G2 strains that circulated simultaneously in Malawi (Fig. S2) and not with DS-1-like G1P8] strains.
identified in Southeast Asia and Japan. Sequences within each cluster were 95–99.8% similar (calculated using nucleotide identity matrices, data not shown).

All eleven genome segments of the G1P[8] strains were undergoing purifying selection thus potentially resulting in stabilizing selection following purging of deleterious variants arising during error-prone rotavirus replication due to its RNA genome (Table 1). The genetic algorithm for recombination detection (GARD) (31) and the single breakpoint recombination (SBP) (32) did not identify any recombination events within each genome segment of the study strains (Table 1). Thus, the change in genetic constellation of post-vaccine G1P[8] strains were likely generated through exchange of whole gene segment (reassortment) between circulating Wa-like and DS-1-like human rotaviruses.

G1P[8] strains that circulated in Malawi between 1998 and 2014 exhibited distinct replacement dynamic patterns

Bayesian inference of time-measured trees were individually constructed for each of the 11 genome segments of the G1P[8] strains to further determine their evolutionary dynamics. As illustrated by phylogenetic networks, prior to the emergence of the reassortant L3 lineage, both structural and non-structural genes segregated into at least 2 distinct lineages with a common ancestor in the mid-1990's or before (Fig. 2 and S3). For NSP1, NSP2, NSP3, VP1, VP2, VP3 and VP6, a single lineage (L1) was predominant until the mid-2000's (2003-2005), and following its disappearance, strains forming a second lineage L2 circulated until the emergence of the DS-1-like G1P[8] reassortant strains. For the VP4, VP7 and NSP4
genes, two clusters co-circulated until 2004, and replacement with strains in L2 did not occur until later, between 2010 and 2012. In 2013, VP7 and VP4 encoding genes of the emergent reassortant strains also formed a third cluster with likely common ancestor with strains in L2. Both MCC coalescent framework and Maximum-Likelihood phylogenetic approaches showed that G1P[8] strains acquired DS-1-like genome segments (genotype 2) in the nine non-VP4 and non-VP7 genes in 2013, out-competing typical Wa-like L2 variants (genotype 1) during the post-vaccine era (Fig. 2 and S2-S4). These analyses also showed that the post-2013 Malawi DS-1-like G1P[8] reassortant strains clustered with or were derived from Wa-like G1P[8] strains and DS-1-like G2 strains that co-circulated in Malawi, and hence emerged through reassortment among local strains. The genes of Malawi DS-1-like G1P[8] strains clustered away from those of DS-1-like reassortants that emerged recently in southeast Asian countries and Japan indicating that they were likely not imported.

Since the post-vaccine Malawi G1P[8] strains contained a DS-1-like genetic backbone, only sequences for the Wa-like genome segments for G1P[8] strains generated in this study were used for phylodynamic analysis of VP4 and VP7 genes, whereas cognate genes of DS-1-like strains that were assigned various G and P types (Table S1), which also circulated in Malawi from 1997 – 2014 were used to estimate evolutionary dynamics for non-VP4 and non-VP7 genes for post-vaccine G1P[8] strains (Fig. 3 and S4). The calculated mean times per gene to the most recent common ancestor (tMRCA) ranged from 1986 to 1996 (Fig. 4a). When only non-VP4 and non-VP7 genes (NSP1 – NSP4, VP1 – VP3 and VP6) for reassortant DS-1-like G1P[8] and cognate genes of DS-1-like strains collected from Malawi between 1997
and 2014 were used to infer Bayesian time-measured trees, the tMRCA for the atypical G1P[8] strains (L3 cluster) was estimated to range from 2009 – 2011 which was similar to predominantly DS-1-like G2 strains that were detected post-vaccine introduction (Fig 3. and Fig. S4). Marginal differences were observed between the mean evolutionary rates for each genome segment that ranged from $9.7 \times 10^{-4} – 4.1 \times 10^{-3}$ nucleotide substitutions per site per year (Fig. 4b). VP2 had the lowest mutation rates ($9.7 \times 10^{-4}$; 95% Highest Posterior Density interval (HPD): $7.4 \times 10^{-4}$ to $1.2 \times 10^{-3}$ substitutions per site per year), whereas VP3 had the highest ($4.1 \times 10^{-3}$ HPD; $3.1 \times 10^{-3}$ to $5 \times 10^{-3}$ (Fig. 4b Fig. 3 and Fig. S4).

We then utilized GMRF tree prior to investigate whether Rotarix™ introduction had an impact on the relative population size of the circulating G1P[8] strain. There was no evidence to suggest that vaccine introduction affected either the G1P[8] genetic diversity nor population size for VP4, VP7, NSP2, NSP3 and NSP5 genes as the peaks and troughs of their Skygrid plots exhibited similar stable profiles just before (2005 – 2012) and after (2013 – 2014) vaccine use. In contrast, genes encoding VP1 – VP3, VP6 and NSP1 had relatively stable profiles and also smaller effective population size during the post-vaccine era compared to pre-vaccine introduction which could be natural as similar downward trends were already occurring before vaccine introduction (Fig. 4c).

Mutations within VP7 antigenic regions did not affect the structural conformation of neutralising epitopes essential for antigenic recognition by neutralising antibodies.
Mapped amino acid motifs that constitute neutralising epitopes on the outer capsid glycoprotein were compared between Rotarix™ and Malawian G1P[8] strains. In total, 15 lineage defining amino acid substitutions were identified across the entire VP7 sequence, with only five of these being located at the mapped antigenic regions 7-1a (S123N and K291R) and 7-2 (AR C: M202T, M212T and N221S) (Fig. 5a). A single amino acid substitution N221S located at one of these antigenic regions differentiated L3 from the L2 cluster strains (Fig. 5).

The VP7 structures for L1 – L3 strains and Rotarix™ strain aligned perfectly when superimposed (Fig. 5b) implying a conserved conformation consistent with the conservation in chemical properties of the replacing amino acids. The replacing amino acids did not appear to impact on the structural conformation of the antigenic regions of the glycocapsid protein of the G1P[8] circulating pre and post vaccine introduction (Fig. 5c – e).

**Discussion**

Malawi was one of the first African countries to introduce rotavirus vaccine into its infant immunization schedule in October 2012. By 2015, vaccine coverage had exceeded 95% (30). We enhanced rotavirus surveillance activities in the post-vaccine period in Malawi to primarily assess the impact of vaccine introduction on the burden of rotavirus disease (27, 30). The availability of a rotavirus strain collection from before (1997 – 2012) (28) and after Rotarix™ introduction offered a rare opportunity to assess the early impact of vaccine introduction on the genetic diversity of the circulating strains and their evolutionary patterns over time. In the
current study, phylogenetic analysis and evolutionary history were inferred for G1P[8] rotaviruses, the most prevalent rotavirus strain globally (6). In Malawi, G1P[8] strains were the only rotaviruses consistently detected year on year, from 1998 to 2014 hence enabling a systematic inference of evolutionary patterns over time. Furthermore, G1P[8] strains are homologous with respect to VP7 and VP4 genotype to the Rotarix™ vaccine that is in use in Malawi (33), hence permitting homologous genomic comparison.

among DS-1- and Wa-like strains circulating in Malawi during the study period revealed by phylogenetic networks, together suggest that the atypical Malawian G1P[8] strains emerged locally through genome reassortment among co-circulating Wa- and DS-1-like strains. This is further supported by the detection of high prevalence of G2 strains reported in Malawi from 2012 (27, 30), thus providing the required circulating strains to allow the emergence of Wa/DS-1-like reassortant strains.

The Malawian G1 mutation rate falls in the same range as those of rotavirus G9 and G12 strains (37-41). Zeller et al. (35) recently analysed the phylodynamics of typical Wa-like G1P[8] strains that circulated before and after vaccine introduction in Belgium and Australia. Unlike the Malawian G1P[8] strains, all G1P[8] strains from Belgium and Australia had a Wa-like genetic constellation and their tMRCA ranged from 1846 – 1945, whereas their evolutionary rates which ranged from 6.05 x 10^{-4} – 1.01 x 10^{-3} nucleotide substitutions per site per year were similar to those for Malawian G1P[8] strains. These are relatively slower compared to the known rates of RNA viruses (42) possibly due to the double stranded nature of rotavirus genome.

The emergence of the atypical reassortant DS-1-like G1P[8] strains in Malawi coincided with the programmatic roll-out of Rotarix™. Our data suggest that the atypical Malawian DS-1-like G1P[8] derived from a combination of reassortment and drift of rotaviruses that were circulating locally since the 1990’s. We found that at least three G1P[8] lineages have been circulating in Malawi from 1998 – 2014. The diversity of the circulating G1P[8] variants exhibited periodic lineage
replacements, similar to influenza (43), dengue (44) and other enteric viruses such as noroviruses (45), where lineage replacement also appears to be an important evolutionary mechanism in response to herd immunity (35, 46, 47). Lineage diversity and replacement coincided temporally for blocks of genes, and can be explained by drift and reassortment events occurring hand in hand.

Although the detection of DS-1-like G1P[8] strains coincided with widespread use of a G1P[8] Rotarix™ rotavirus vaccine in Malawi in 2013, it is difficult to ascertain the role the vaccine had on the emergence of these atypical strains considering the short post-vaccine period. The phylodynamic analyses suggested that these strains were derived from those strains circulating in the human population well before vaccine introduction. It was difficult to determine the effect of vaccine introduction on the effective population size of the circulating G1P[8] strains as only post-vaccine G1 strains from 2013 and 2014 were analysed, which was too early to detect genetic variations in the virus population size.

Frequent detection of DS-1-like G2 strains just before Rotarix™ introduction and during the post-vaccine era in Malawi (27, 30) may indicate a natural surge of DS-1-like rotaviruses during this period similar to G2 cyclic seasonal epidemic patterns that has been observed in many countries including in Africa (48, 49). However, the predominance of reassortant DS-1-like G1P[8] strains post-vaccine introduction could suggest positive selection for atypical G1P[8] strains, and that such selection may not be driven exclusively by the VP7 and VP4 specificities. The DS-1-like genotype constellation was however found in rotavirus diarrhoea cases regardless of the vaccination status of the children (48% were vaccinated, data not shown).
The relatively fewer rotavirus G1P[8] genomes that were analysed is a potential limitation of the current study as such sequencing of additional G1P[8] strains that circulated in Malawi post-2014 is underway to determine whether emergence of DS-1-like G1P[8] strains remain in circulation.

It has been shown that mutations along the three main mapped neutralising epitopes of VP7 can generate vaccine-escape mutants (50). For instance, amino acid substitutions at positions 94, 97, 147 and 291 significantly affect antigenic recognition of human G1 strains (50). When VP7 of post-vaccine G1P[8] strains (L3) were compared to that of Rotarix™, only K291R (7-1a) substitution occurred within the sites associated with antibody escape mutants, and this substitution was already present among the strains circulating pre-vaccine introduction since 1998, hence not selected due to potential vaccine pressure. Furthermore, both Lysine (K) and Arginine (R) are positively charged polar proteins hence this substitution is unlikely to produce significant changes to the biochemical properties of VP7. The only substitution present among the post-vaccine G1 strains was N221S. As both asparagine (N) and serine (S) are small non-charged residues, it does not appear to have a significant impact on the overall conformation and structure of the protein surface. However, the loss of an asparagine residue may have resulted in the loss of a potential glycosylation site which could affect VP7 antigenic determinants (51). This change does not appear to be a universal glycosylation position as Serine also occurs naturally at position 221 for some non-G1 strains like S2 (G2), RV-5 (G3) and ST5 (G4) (52). Whilst this change is outside of the currently proven glycosylation sites (69 – 71, 238 – 240 and 318 – 320), the N221S change occurred within the
neutralizing epitope C (Antigenic regions C) of VP7 (52). In order to exclude the potential for this amino acid substitution, further functional studies may be warranted bearing in mind that changes in immunogenicity and neutralisation patterns have been attributed to different glycosylation patterns using mutated laboratory strains (51, 53).

The only hydrophobic to hydrophilic change, which would potentially affect the structural conformation and stability of proteins significantly, occurred outside the mapped antigenic regions at position 266 [Alanine (A) to S], and was also present among the strains circulating pre 2009, those in L2. However, it is possible that substitutions in the non-AR could affect the stability of the viral particle or protein assortment specificities, since VP6 serves as an anchor for the outer capsid VP4 and VP7 proteins where the 260 trimers of VP7 lie directly on top of the VP6 trimers (54). Contact with VP6 is facilitated by the arm-like extensions formed by the VP7 N-termini that also forms lattice with other VP7 trimers. This interaction allows gripping of VP7 to the intermediate VP6 layer and reinforces the integrity of the outer-shell (55). Such interactions may drive the selection of particular VP7 and/or VP4 lineages in reassortant strains and explain lineage replacements that may not necessarily be explained exclusively in terms of immune pressure. In a recent analysis, vaccine effectiveness against all G1P[8] strains three years post-Rotarix™ introduction in Malawi was 82% (30), suggesting high degree of protection against atypical DS-1-like G1P[8] strains, given that these G1P[8] strains were detected in randomly selected stool samples collected between 2013 – 2014. Further analysis is underway in order to assess the extent of the spread of the DS-1-like G1P[8] strains
and to calculate vaccine effectiveness against various G and P types possessing Wa- and DS-1-like genetic backbone.

In conclusion, genome reassortment and mutation are the major evolutionary mechanism that influenced the genetic diversity of G1P[8] strains that circulated in Malawi from 1998 – 2014. Atypical DS-1-like G1P[8] strains emerged in 2013 through genome reassortment events between Wa- and DS-1-like human strains that can be traced back in Malawi to the 1990’s. Mutations within the outer capsid VP7 of Malawian G1P[8] strains compared to RV1 had no impact on the structural conformation of antigenic regions, suggesting little or no effect on the recognition of vaccine-induced antibodies. Thus, the remaining genome segments (non-G or -P) might also play an important role in immune evasion. It is likely that the atypical DS-1-like G1P[8] strains emerged through natural strain evolutionary pressure which is unrelated to vaccine use. However, the predominance of atypical reassortant DS-1-like G1P[8] strains, which coincided with vaccine introduction, could suggest positive selection of atypical G1P[8] strains that were undergoing purifying selection. Vaccine effectiveness against such atypical strains needs careful investigation.

Materials and methods

Rotavirus strains

Stool samples collected from children aged <5 years presenting with acute gastroenteritis at Queen Elizabeth Central Hospital (QECH) in Blantyre, Malawi from 1998 – 2014 were utilised (27-29). Diarrhoea-case definition, rotavirus screening,
VP4 and VP7 genotyping methods and strains that circulated in Malawi have been published elsewhere (27-29). In total, 4,888 stool specimens were collected before vaccine introduction (1997 – 2012), whereas 2,109 were collected after vaccine introduction (2013 – 2014). Rotaviruses with G1P[8] outer capsid proteins were the only strains that were detected every year from 1998 – 2015 (comprising of 554 pre- and 95 post-vaccine introduction strains). Therefore, only G1P[8] strains from each surveillance year, and that had sufficient faecal material for dsRNA re-extraction were selected for further examination. Where available, a single faecal sample from each month in each year was randomly selected for whole genome sequencing; only samples from which whole genome data were obtained were included in the analysis. Ethical approval was obtained from the National Health Sciences Research Committee, Lilongwe, Malawi (# 867) and the Research Ethics Committee of the University of Liverpool, Liverpool, UK (# 000490).

**Preparation of purified dsRNA and cDNA for rotavirus whole genome**

Rotavirus dsRNA was extracted and purified as previously described (17, 56). An additional DNase I treatment step following a lithium chloride precipitation step was included to remove contaminating DNA (Sigma-Aldrich, Dorset, UK). Purified dsRNA was quantified using a Qubit® 3.0 Fluorometer (Life Technologies, CA, USA). Sequence-independent cDNA synthesis and PCR amplification procedures described previously (17, 56) were used to amplify cDNA for rotavirus whole genomes from samples with ≥ 2ng/µl dsRNA.
RNA and cDNA library construction and illumina HiSeq sequencing

After denaturing dsRNA at 95°C for 5 min, ScriptSeq RNA-Seq Library Preparation Kit V2 was used to generate Illumina sequencing libraries for samples that had <2 ng/µl dsRNA (Illumina Inc., CA, USA). Purified cDNA generated from samples with >2 ng/µl dsRNA was subjected to standard bar-coding and library construction for illumina sequencing using Nextera XT DNA Library Preparation Kit (Illumina Inc., CA, USA). Rotavirus VP6-specific qPCR (57) and 2100 Bioanalyzer (Agilent Technologies Inc., CA, USA) were used to quality control the DNA libraries followed by sequencing using HiSeq 2000 platform (Illumina Inc., CA, USA) at the Centre for Genomic Research (CGR), University of Liverpool, UK.

Sequence assembly and determination of rotavirus genotypes

Illumina adapter sequences were trimmed from the raw Fastq sequence data using Cutadapt v1.2.1 (58) and Sickle v1.2 software (59). Complete consensus nucleotide sequences were generated by mapping trimmed illumina sequence reads to various complete nucleotide sequences of prototype rotavirus genogroup strains using both de novo and Reference DNA sequence assembler algorithms implemented in Geneious software v8 (60). Rotavirus genotypes were assigned to each of the 11 genome segments using the web-based automated rotavirus genotyping tool, RotaC v2.0 (http://rotac.regatools.be) (7). All complete nucleotide and deduced amino acid sequences generated in this study were deposited into the GenBank (61) under the accession numbers MG181227 - MG181941.
Sequence alignments and Maximum Likelihood phylogeny construction

Reference nucleotide sequences for each rotavirus genome segment were retrieved from the Rotavirus resource in the GenBank database (61). This was followed by multiple sequence alignment of the assembled sequences for the study strains using Muscle v3.8.31 (62) included in MEGA, v6.0 (63). Initial phylogenetic trees for each segment were inferred using Maximum Likelihood approach implemented in MEGA by selecting the DNA model that best fitted the data according to the corrected Akaike Information Criterion (AICc) as described previously (12). We used a generalized time reversible (GTR) model with Gamma (Y) heterogeneity across nucleotide sites while the reliability of the branching order and partitioning were assessed by performing 1000 bootstrap replicates (64).

Bayesian inference of phylogenies and population dynamics

Coalescent analyses were performed using BEAUTi v1.7.5 and BEAST v1.8 (65, 66) with the following parameter specifications; lognormal relaxed (uncorrelated) clock model (67), constant size coalescent tree prior, Hasegawa-Kishino-Yano (HKY85) nucleotide substitution model with estimated base frequencies (68) and a Gamma (Y) site heterogeneity model with 4 rate categories (69) and the prior mutation rate (μ) of ~1.0 x 10^{-3} nucleotide substitutions/site/year as previously reported by Zeller et al (35). The maximum likelihood trees generated in the previous section were used as starting trees for the Bayesian analysis in BEAST. We used a Gaussian Markov Random Field (GMRF) tree prior, which also allows for investigation of the population dynamics i.e. effective
population size ($N_{e\tau}$) or relative genetic diversity, over time. A total of 200 million Markov Chain Monte Carlo (MCMC) iterations were performed and sampled every 40,000th generation. The first 20 million iterations (10% of the total) from the MCMC analysis, burn-in time, were discarded since these may represent states that the chain explored before reaching the equilibrium state of the target distribution. The mean values and 95% highest posterior densities (HPD) of the mutation rates and the times to the most recent common ancestors (tMRCA) for each rotavirus segment were calculated from the BEAST output using Tracer v1.6.0 (http://tree.bio.ed.ac.uk/software/tracer/). The maximum clade credibility (MCC) tree for each viral segment was generated using Tree Annotator v2.1.2 (http://beast.bio.ed.ac.uk/treeannotator) and visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and BioPython scripts (70).

**Detection of natural selection in proteins encoded by rotavirus segments**

ORFs that encode both the structural and non-structural proteins were identified and extracted from the multiple sequence alignments of each rotavirus genomic segment. We used BioPython (70) scripts to extract and manipulate the sequence alignments. The ORFs alignments for each protein were then converted to corresponding codon alignments using CodonAlign (http://www.hiv.lanl.gov/cgi-bin/CodonAlign). The alignments were then used to calculate the global ratio of synonymous and non-synonymous substitutions ($dN/dS$) for each ORF using single likelihood ancestor counting (SLAC) (71) and a Fast, Unconstrained Bayesian AppRoximation for Inferring Selection (FUBAR) (72). To identify specific sites under selection in the ORFs, we also used the mixed effects model of episodic selection.
(MEME) (73), random effects likelihood (REL) (74) and fixed effects likelihood (FEL) (75) methods implemented in DataMonkey, a webserver for the HyPhy package (76). Occurrence of genetic recombination was checked using the genetic algorithm for recombination detection (GARD) (31) and the single breakpoint recombination (SBP) (32). We used the following default significance levels i.e. p-value, Bayes Factor or posterior probability of 0.1 for SLAC, MEME and FEL, 0.9 for FUBAR and 0.5 for REL. The Hasegawa-Kishino-Yano 85 (HKY85) nucleotide substitution model, Beta-Gamma site-to-site rate variation and Neighbour-Joining trees were used for the selection analysis. All the analyses were done using the DataMonkey webserver (71).

Structure comparison between the outer capsid glycoprotein of RV1 and G1P[8] strains to predict changes in antibody binding

To investigate the likely impact of amino acid substitutions on anti-RV1 antibody recognition due to mutations that occurred over time within the antigenic regions of Malawian G1P[8] rotaviruses, the structural conformation of the outer capsid glycoprotein of pre- and post-vaccine strains were compared to that of RV1. Representative VP7 sequences for each G1P[8] lineage were utilised for protein structure modeling using Modeler Version 9.17 (77). Templates were searched in Protein Data bank (78) using an integrated web based HHpred program (79). The best model with highest zdope score was selected for analysis from the one hundred models that were generated for each sequence.
Acknowledgements

The authors wish to thank the collaborating members of the VacSurv Consortium (Anthony Costello, Charles Mwansambo, James Beard, Amelia Crampin, Carina King, Sonia Lewycka, Hazzie Mvula, Tambosi Phiri, Jennifer Verani, Cynthia G Whitney, Louisa Pollock and Aisleen Bennett). We acknowledge the support of the laboratory staff at the Malawi-Liverpool-Wellcome Trust Clinical Research Programme and the sequencing and informatics teams at the Centre for Genomic Research (CGR), University of Liverpool, UK.

This work was supported by a research grant from GlaxoSmithKline Biologicals and the Wellcome Trust (Programme grant number 091909/Z/10/Z and the MLW Programme Core Award). K.C.J. is a Wellcome Trust Training Fellow (grant number: 201945/Z/16/Z). C.C. acknowledges funding from the Commonwealth Scholarship Commission (PhD Studentship). M.I-G is partly supported by the NIHR HPRU in gastrointestinal Infections.

The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication. GlaxoSmithKline Biologicals SA was provided the opportunity to review a preliminary version of this manuscript for factual accuracy but the authors are solely responsible for final content and interpretation. The authors received no financial support or other form of compensation related to the development of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US Centers for Disease Control and Prevention.

The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), University of East Anglia, University of Oxford and the Institute of Food Research. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

K. C. J. N. B.-Z. and N. F. have received investigator-initiated research grant support from GlaxoSmithKline Biologicals. M. I.-G. has received investigator-initiated research grant support from GlaxoSmithKline Biologicals and Sanofi Pasteur Merck Sharpe & Dohme. O. N. has received research grant support and honoraria from Japan Vaccine and MSD for delivering lectures on rotavirus vaccines. N. A. C. has received research grant support and honoraria for participation in rotavirus vaccine advisory board meetings from GlaxoSmithKline Biologicals. All other authors report no potential conflicts.


rotavirus strain diversity in the pre rotavirus vaccine era: insights for understanding the impact of rotavirus vaccination programs. Vaccine 30 Suppl 1:A122-130.


Phylogenetic and computational structural analysis of VP7 gene of group a 

sequence determination and elucidation of the evolutionary history of a 
rotavirus Wa variant reveal a close relationship to various Wa variants 
derived from the original Wa strain. Infect Genet Evol 20:276-283.

Afrad MH, Matthijnssens J, Afroz SF, Rudra P, Nahar L, Rahman R, 
Hossain ME, Rahman SR, Azim T, Rahman M. 2014. Differences in lineage 
replacement dynamics of G1 and G2 rotavirus strains versus G9 strain over a 
period of 22 years in Bangladesh. Infect Genet Evol 28:214-222.

Genetic variation in the VP7 gene of rotavirus G1P[8] strains isolated in 

Jenkins GM, Cowart LA, Signorelli P, Pettus BJ, Chalfant CE, Hannun YA. 
2002. Acute activation of de novo sphingolipid biosynthesis upon heat shock 
causes an accumulation of ceramide and subsequent dephosphorylation of 

Establishment and lineage replacement of H6 influenza viruses in domestic 


Figure Legends

Fig. 1. Rotavirus strains, genetic constellation and phylogenetic networks of G1P[8] rotaviruses detected from Malawian children at QECH from 1997 - 2015. (a) Schematic representation of the proportions of all genotypes detected in rotavirus-positive stool samples. The size of the circle is directly proportion to the detection frequency of G and P genotypes. There were no rotavirus surveillance activities in 2010 hence samples were not collected. (b) Bayesian Maximum Clade Credibility tree for concatenated whole rotaviruses genome sequences illustrating genetic constellation and reassortment patterns for 32 pre- and 18 post-vaccine G1P[8] strains from Malawi as well as the prototype Wa and DS-1 strains for comparison. Genotype and genogroups assignment for each segment were based on nucleotide sequence identities, assigned using RotaC. Genome segment that were assigned the same genotype are shown with the same colour and genotype numbers. Green represents Wa-like P[8] for VP4 and genotypes 1 for the rest of the other 10 gene segments; Red represents all DS-1-like genotype 2 for all 11 gene segments. (C) Phylogenetic network of complete concatenated whole genome sequences of G1P[8] rotavirus strains detected in Malawi from 1998 - 2014. Branches are drawn to scale and splits in the network indicate reassortments. Network clusters are colour-coded and named in accordance with their phylogenic lineages (L1 – L3) that correlated with time of strain isolation before or after rotavirus vaccine introduction. Cluster L1 (green) and L2 (blue) contained pre-vaccine G1P[8] strains, whereas L3 (red) contained post-vaccine strains. Network sub-clusters within each main cluster are shaded in blue (L1), red (L2) or orange
Fig. 2. Bayesian Maximum Clade Credibility (MCC) time tree based on complete nucleotide sequences illustrating lineage replacement within the genome segments encoding structural proteins of the G1P[8] strains that circulated in Malawi from 1998 – 2014. With the exception of VP4 and VP7 genes that had L2 and L3 genes sharing close ancestry, the rest had three distinct G1P[8] lineages. L1, L2, and L3 represents lineage 1, 2 and 3, respectively.

Fig. 3. Bayesian Maximum Clade Credibility (MCC) time tree based on complete nucleotide sequences of the structural proteins for G1P[8] strains from Malawi. Only DS-1-like genome segments for typical DS-1-like strains that were assigned G2P[4], G2P[6], G8P[4], G8P[6] and G12P[6] outer capsid genotypes from Malawi were included to calculate evolutionary dynamics for VP1 – VP4, VP6 and VP7 encoding genome segments for the atypical DS-1-like G1P[8] strains (L3 cluster). The summary for their evolutionary rates and tMRCA are presented in Table 2.

Fig. 4. Time to the most recent common ancestor (TMRCA), evolutionary rates for each genome segment of Malawian G1P[8] strains and comparative population dynamics of G1P[8] rotavirus strains circulating in Malawi, 1998 – 2014. (a) The evolutionary rates and tMRCA for each genome segment of the
Malawian atypical G1P[8] strains shown together with their 95% Highest Posterior Density (HPD) intervals. (b) Absolute values for the mean and range of the evolutionary rates and tMRCA at 95% HPD intervals are shown. (c) Phylogenies and relative genetic diversity were estimated using the Gaussian Markov Random Field (GMRF) model represents Bayesian Skygrid plots for VP1 – VP4, VP6, for NSP1 – NSP5 encoding genome segments. Solid lines in the GMRF plot represent the mean relative genetic diversity through time.

**Fig. 5. Amino acid substitutions and structural conformation of the outer capsid glycoprotein of Malawian G1P[8] strains.** (a) Complete VP7 sequence of representative pre- and post-vaccine G1P[8] strains aligned to that of RV1 exhibiting amino acid substitutions that occurred within the variable regions (VR) and mapped antigenic regions (AR) over time. Lineage defining amino acid amino acid substitutions are highlighted in green, blue and yellow for L1, L2 and L3 lineages, respectively. Pre- and post-vaccine strains are shown with vertical green and red bars on the right, respectively. Strains belonging to the L1, L2 and L3 phylogenetic clusters are shown with green, blue and red bars respectively on the right. (b) Perfect alignment of superimposed VP7 structures exhibiting few differences between RV1 and L1 – L3 strains. Antigenic regions A, B and C are shown in white. L1 – L3 and RV1 strains are shown in yellow, green, blue and red, respectively. (c-e) Surface visualisation of VP7 from the outside of the virion on 3-fold axis displaying amino acid differences when structures for L1 (c), L2 (d) and L3
(e) G1P[8] strains were superimposed on the outer capsid glycoprotein of RV1. Numbers correspond to the positions where mutations occurred.
Table 1. Evolutionary selective forces and recombination in all eleven proteins of the Malawian G1P[8] rotavirus strains.

<table>
<thead>
<tr>
<th>Protein</th>
<th>SLAC dN/dS</th>
<th>FUBAR(^a) (\omega (\beta/\alpha))</th>
<th>Consensus Selective Force</th>
<th>Recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>0.0502</td>
<td>2.22</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>VP2</td>
<td>0.0538</td>
<td>2.67</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>VP3</td>
<td>0.0965</td>
<td>2.47</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>VP4</td>
<td>0.1052</td>
<td>3.15</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>VP6</td>
<td>0.0310</td>
<td>4.66</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>VP7</td>
<td>0.2033</td>
<td>6.01</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>NSP1</td>
<td>0.2239</td>
<td>4.21</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>NSP2</td>
<td>0.0920</td>
<td>5.62</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>NSP3</td>
<td>0.0870</td>
<td>5.44</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>NSP4</td>
<td>0.1097</td>
<td>7.10</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
<tr>
<td>NSP5</td>
<td>0.1097</td>
<td>7.47</td>
<td>Purifying Selection</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\)This summary table reports the means of posterior distribution of synonymous (\(\alpha\)) and non-synonymous (\(\beta\)) substitution rates over sites, as well as the mean posterior probability for \(\omega (=\beta/\alpha) < 1\) at a site.
Fig. 3
Fig. 4