# 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,c], 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,g,*]; 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

Kingdom. Tel: +44 151 795 9614, Fax: +44 151 795 5529, Email:
Khuzwayo.Jere@liverpool.ac.uk

*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 $9.7 \times 10^{-4}$–$4.1 \times 10^{-3}$ 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.

87 **Importance**

88 The error-prone RNA-dependent RNA polymerase and the segmented RNA genome
89 predispose rotaviruses to genetic mutation and genome reassortment, respectively.
90 These evolutionary mechanisms generate novel strains and have the potential to
91 lead to the emergence of vaccine-escape mutants. While multiple African countries
92 have introduced rotavirus vaccine, there are few data describing the evolution of
93 rotaviruses that circulated before and after vaccine introduction. We report the
94 emergence of atypical DS-1-like G1P[8] strains during the post-vaccine era in
95 Malawi. Three distinct G1P[8] lineages circulated chronologically from 1998–2014;
96 mutation and reassortment drove lineage turnover in 2005 and 2013, respectively.
97 Amino acid substitutions within the outer capsid VP7 glycoprotein did not affect the
98 structural conformation of mapped antigenic sites, suggesting limited effect in
99 recognition of G1 specific vaccine-derived antibodies. The genes that constitute the
100 remaining genetic backbone may play important roles in immune evasion, and
101 vaccine effectiveness against such atypical strains needs careful evaluation.
102
103
104
105
106
107
108
109

110 **Introduction**

111 Diarrhoea is a leading cause of mortality in children under the age of five
112 years globally (1, 2). The majority of hospitalisations and deaths in infants due to
113 severe dehydrating diarrhoea are caused by group A rotaviruses (RVA) (3). The
114 World Health Organization (WHO) recommended universal introduction of
115 rotavirus vaccines in 2009 particularly in countries where diarrhoea mortality is
116 high (4). A global decline from 528,000 to 215,000 in rotavirus-associated deaths
117 per year amongst children <5 years of age has been reported between 2000 and
118 2013, and live-attenuated oral rotavirus vaccines (Rotarix™; RV1 and RotaTeq®;
119 RV5) have now been incorporated into national immunization programs of over 60
120 countries worldwide (5).

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

131 G1 – G4, G9 and G12 in association with P[4], P[6] or P[8] are the
132 predominant genotypes associated with human rotavirus infection worldwide (6,

13-15). 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] rotaviruses have recently emerged in southeast Asia. These atypical G1 strains were initially detected in outbreaks of gastroenteritis among Japanese children (21-23), followed by reports from Thailand (24, 25) and then Vietnam (26). To date, there is no evidence that these atypical G1 strains are widespread. Rotavirus strain surveillance conducted in Blantyre, Malawi since 1997, and introduction of the monovalent, Wa-like G1P[8] rotavirus vaccine (Rotarix™ or RV1) into Malawi's childhood immunisation programme in 2012, allows study of the evolution of G1P[8] strains before and after vaccine introduction. We have, for the first time detected DS-1-like G1P[8] rotavirus strains in Africa, which became predominant following vaccine introduction. The

155 evolutionary forces that were associated with the emergence of the atypical G1P[8]
156 rotavirus strains were also determined.
157
158 **Results**
159 **Emergence of reassortant DS-1-like G1P[8] rotavirus strains**
160 All pre- (33.4%: 1,634/4,888) and post-vaccine (22.6%: 477/2,109) rotavirus–
161 positive stools collected from children presenting with acute severe diarrhoea at
162 QECH were genotyped as part of on-going rotavirus surveillance (27-30). Amongst
163 multiple strains were characterised, and G1P[8] rotaviruses were consistently
164 predominant strains that were detected each year before (39.4%: 554/1406) and
165 after (31.4%: 95/303) vaccine introduction (Fig. 1a and S1). Whole genome
166 sequences of 32 pre-vaccine G1P[8] strains (collected between 1998 and 2012),
167 and 18 post-vaccine G1P[8] strains (collected from 2013 to 2014) were successfully
168 generated (see supplementary Table 1 for yearly distribution). Interruption of
169 surveillance in 2010 meant that no G1P[8] strains were available between 2010 and
170 2011. Among the pre-vaccine G1P[8] strains, 31 had the G1-P[8]-I1-R1-C1-M1-A1-
171 N1-T1-E1-H1 genotype constellation, hence designated as Wa-like G1P[8], whereas
172 one was G1-P[8]-I1-R1-C1-M1-A1-N1-T2-E1-H1, hence designated as a mono-
173 reassortant Wa-like G1P[8] strain. In contrast, 16/18 of the post-vaccine G1P[8]
174 strains had a DS-1-like genotype constellation (G1-P[8]-I2-R2-C2-M2-A2-N2-T2-E2-
175 H2) hence were designated as atypical double-reassortant DS-1-like G1P[8] strains.
176 The remaining two post-vaccine G1P[8] strains (BID1LN and BID230) had Wa-like

177 VP1 (R1) and Wa-like NSP3 (T1) genes, respectively, hence were designated as
178 atypical triple-reassortant DS-1-like G1P[8] strains (Fig. 1b and Table S1).
179
180 **Atypical post-vaccine G1P[8] strains emerged through genome reassortment**
181 **between Wa-like and DS-1-like human rotaviruses**
182 Whilst the concatenated sequences of all 11 genome segments of pre- and
183 post-vaccine G1P[8] rotaviruses clustered with prototype Wa- and DS-1-like strains,
184 respectively (Fig. 1b), phylogenetic networks constructed using concatenated whole
185 genome sequences of G1P[8] strains characterised in the revealed frequent
186 reassortment events between strains from the same or different network clusters
187 (Fig. 1c). The Malawi G1P[8] strains were distributed into three main phylogenetic
188 network clusters (L1, L2 and L3), which also contained multiple sub-clusters
189 (shaded within main clusters in Fig. 1c). L1 phylogenetic cluster contained strains
190 that circulated from 1998 to 2004; strains detected from 2005 to 2012 grouped into
191 L2 cluster, while L3 cluster only consisted of post-vaccine G1P[8] strains that
192 circulated from 2013 to at least 2014 in Blantyre.
193 The phylogenetic relationship of Malawi G1P[8] strains was inferred using
194 the Maximum-Likelihood method for which complete nucleotide sequences of RVAs
195 available in the GenBank and sequences of Wa-like (VP4 and VP7 only) and DS-1-
196 like (the other nine genes) strains from Malawi were used. The DS-1-like genes of
197 the post-vaccine G1P[8] reassortants closely clustered with those of G2 strains that
198 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.

Whilst the pre-vaccine G1P[8] strains had a typical Wa-like genetic constellation (8, 16, 34, 35), the backbone of the G1P[8] strains detected after vaccine introduction possessed a DS-1-like genotype constellation, which is most frequently associated with G2P[4] strains, occasionally with G9 and G12 strains, and with G8P[4] or G8P[6] strains in Africa (8, 17, 18, 36). Detection of atypical reassortant DS-1-like G1P[8] strains had not been documented during 15 years of pre-vaccine rotavirus strain surveillance in Malawi or elsewhere in Africa. Similar atypical reassortant DS-1-like G1P[8] strains emerged recently in Southeast Asian countries where rotavirus vaccine use is limited (21-26). Atypical DS-1-like G1P[8] strains from Malawi and Asia exhibited distinct phylogenetic clustering patterns in all 11 genome segments, indicating that these atypical DS-1-like G1P[8] strains most likely emerged independently in Malawi and Asia and did not arise through importation. In addition: (i) tight clustering between G and P genes of the atypical Malawian G1P[8] strains with other G1P[8] Wa-like strains, (ii) monophyletic clusters between non-G and non-P genes of Malawian atypical G1P[8] strains with other Malawian DS-1-like strains, and (iii) evidence of frequent reassortment events

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

337 replacements, similar to influenza (43), dengue (44) and other enteric viruses such
338 as noroviruses (45), where lineage replacement also appears to be an important
339 evolutionary mechanism in response to herd immunity (35, 46, 47). Lineage
340 diversity and replacement coincided temporally for blocks of genes, and can be
341 explained by drift and reassortment events occurring hand in hand.

342 Although the detection of DS-1-like G1P[8] strains coincided with
343 widespread use of a G1P[8] Rotarix™ rotavirus vaccine in Malawi in 2013, it is
344 difficult to ascertain the role the vaccine had on the emergence of these atypical
345 strains considering the short post-vaccine period. The phylodynamic analyses
346 suggested that these strains were derived from those strains circulating in the
347 human population well before vaccine introduction. It was difficult to determine the
348 effect of vaccine introduction on the effective population size of the circulating
349 G1P[8] strains as only post-vaccine G1 strains from 2013 and 2014 were analysed,
350 which was too early to detect genetic variations in the virus population size.
351 Frequent detection of DS-1-like G2 strains just before Rotarix™ introduction and
352 during the post-vaccine era in Malawi (27, 30) may indicate a natural surge of DS-1-
353 like rotaviruses during this period similar to G2 cyclic seasonal epidemic patterns
354 that has been observed in many countries including in Africa (48, 49). However, the
355 predominance of reassortant DS-1-like G1P[8] strains post-vaccine introduction
356 could suggest positive selection for atypical G1P[8] strains, and that such selection
357 may not be driven exclusively by the VP7 and VP4 specificities. The DS-1-like
358 genotype constellation was however found in rotavirus diarrhoea cases regardless
359 of the vaccination status of the children (48% were vaccinated, data not shown).

360 The relatively fewer rotavirus G1P[8] genomes that were analysed is a potential
361 limitation of the current study as such sequencing of additional G1P[8] strains that
362 circulated in Malawi post-2014 is underway to determine whether emergence of
363 DS-1-like G1P[8] strains remain in circulation.

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

383 neutralizing epitope C (Antigenic regions C) of VP7 (52). In order to exclude the
384 potential for this amino acid substitution, further functional studies may be
385 warranted bearing in mind that changes in immunogenicity and neutralisation
386 patterns have been attributed to different glycosylation patterns using mutated
387 laboratory strains (51, 53).
388 The only hydrophobic to hydrophilic change, which would potentially affect the
389 structural conformation and stability of proteins significantly, occurred outside the
390 mapped antigenic regions at position 266 [Alanine (A) to S], and was also present
391 among the strains circulating pre 2009, those in L2. However, it is possible that
392 substitutions in the non-AR could affect the stability of the viral particle or protein
393 assortment specificities, since VP6 serves as an anchor for the outer capsid VP4 and
394 VP7 proteins where the 260 trimers of VP7 lie directly on top of the VP6 trimers
395 (54). Contact with VP6 is facilitated by the arm-like extensions formed by the VP7 N-
396 termini that also forms lattice with other VP7 trimers. This interaction allows
397 gripping of VP7 to the intermediate VP6 layer and reinforces the integrity of the
398 outer-shell (55). Such interactions may drive the selection of particular VP7 and/or
399 VP4 lineages in reassortant strains and explain lineage replacements that may not
400 necessarily be explained exclusively in terms of immune pressure. In a recent
401 analysis, vaccine effectiveness against all G1P[8] strains three years post-Rotarix™
402 introduction in Malawi was 82% (30), suggesting high degree of protection against
403 atypical DS-1-like G1P[8] strains, given that these G1P[8] strains were detected in
404 randomly selected stool samples collected between 2013 – 2014. Further analysis is
405 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.

452 **RNA and cDNA library construction and illumina HiSeq sequencing**

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

462

463 **Sequence assembly and determination of rotavirus genotypes**

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

474

475 **Sequence alignments and Maximum Likelihood phylogeny construction**

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

486

487 **Bayesian inference of phylogenies and population dynamics**

488 Coalescent analyses were performed using BEAUTi v1.7.5 and BEAST v1.8
489 (65, 66) with the following parameter specifications; lognormal relaxed
490 (uncorrelated) clock model (67), constant size coalescent tree prior, Hasegawa-
491 Kishino-Yano (HKY85) nucleotide substitution model with estimated base
492 frequencies (68) and a Gamma (ϒ) site heterogeneity model with 4 rate categories
493 (69) and the prior mutation rate (μ) of ~1.0 x $10^{-3}$ nucleotide
494 substitutions/site/year as previously reported by Zeller et al (35). The maximum
495 likelihood trees generated in the previous section were used as starting trees for the
496 Bayesian analysis in BEAST. We used a Gaussian Markov Random Field (GMRF) tree
497 prior, which also allows for investigation of the population dynamics i.e. effective

498 population size ($N_e\tau$) or relative genetic diversity, over time. A total of 200 million
499 Markov Chain Monte Carlo (MCMC) iterations were performed and sampled every
500 40,000$^{th}$ generation. The first 20 million iterations (10% of the total) from the
501 MCMC analysis, burn-in time, were discarded since these may represent states that
502 the chain explored before reaching the equilibrium state of the target distribution.
503 The mean values and 95% highest posterior densities (HPD) of the mutation rates
504 and the times to the most recent common ancestors (tMRCA) for each rotavirus
505 segment were calculated from the BEAST output using Tracer v1.6.0
506 (http://tree.bio.ed.ac.uk/software/tracer/). The maximum clade credibility (MCC)
507 tree for each viral segment was generated using Tree Annotator v2.1.2
508 (http://beast.bio.ed.ac.uk/treeannotator) and visualized using FigTree v1.4.2
509 (http://tree.bio.ed.ac.uk/software/figtree/) and BioPython scripts (70).

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

511 ORFs that encode both the structural and non-structural proteins were
512 identified and extracted from the multiple sequence alignments of each rotavirus
513 genomic segment. We used BioPython (70) scripts to extract and manipulate the
514 sequence alignments. The ORFs alignments for each protein were then converted to
515 corresponding codon alignments using CodonAlign (http://www.hiv.lanl.gov/cgi-
516 bin/CodonAlign). The alignments were then used to calculate the global ratio of
517 synonymous and non-synonymous substitutions (dN/dS) for each ORF using single
518 likelihood ancestor counting (SLAC) (71) and a Fast, Unconstrained Bayesian
519 AppRoximation for Inferring Selection (FUBAR)(72). To identify specific sites under
520 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 50 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 C Crampin, Carina King, Sonia Lewycka, Hazzie Mvula, Tambosi Phiri, Jennifer R 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). MIG is partly supported by the NIHR HPRU in gastrointestinal Infections.

K.C.J., J.E.T., U.D.P., N.F., R.S.H., N.A.C. and M.I-G. conceived and designed the study. N.A.C and M.I-G. supervised the study. K.C.J., N.B-Z., C.P. and N.A.C. collected the stool samples. K.C.J., J.L., and C.P. performed the laboratory work. K.C.J. and C.C. carried out the bioinformatic and statistical analyses and drafted the manuscript. B.K. and K.C.J. conducted protein modelling. K.C.J., C.C., N.B-Z., J.L., B.K., C.P., J.E.T., U.D.P., O.N., R.S.H., N.F., N.A.C. and M.I-G. contributed to the discussions, interpretation of the results and reviewing of the manuscript.

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 onrotavirus 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.

588 **Reference**


589 1. **Bryce J, Boschi-Pinto C, Shibuya K, Black RE.** 2005. WHO estimates of the
590 causes of death in children. Lancet **365:**1147-1152.

591 2. **Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers**
592 **C, Black RE.** 2016. Global, regional, and national causes of under-5 mortality
593 in 2000-15: an updated systematic analysis with implications for the
594 Sustainable Development Goals. Lancet **388:**3027-3035.

595 3. **Estes MK, Greenberg HB.** 2013. Rotaviruses, p 1347-1401. *In* Knipe DM HP,
596 Cohen JI, Griffin DE, Lamb RA, Martin MA, et al. (ed), Fields Virology, 6th ed
597 ed. Wolters Kluwer/Lippincott, Williams and Wilkins, Philadelphia, PA.

598 4. **WHO.** 2007. Rotavirus vaccines. Wkly Epidemiol Rec **82:**285-295.

599 5. **Tate JE, Burton AH, Boschi-Pinto C, Parashar UD.** 2016. Global, Regional,
600 and National Estimates of Rotavirus Mortality in Children <5 Years of Age,
601 2000-2013. Clin Infect Dis **62 Suppl 2:**S96-S105.

602 6. **Estes MK, Kapikian AZ.** 2007. Rotaviruses, p 1917–1974. *In* Knipe DM,
603 Howley, P.M., Griffin,, D.E. L, R.A., Martin, M.A., Roizman, B., Straus, S.E. (ed),
604 Fields Virology, fifth edition ed. Lippincott Williams & Wilkins, Philadelphia.

605 7. **Maes P, Matthijnssens J, Rahman M, Van Ranst M.** 2009. RotaC: a web-
606 based tool for the complete genome classification of group A rotaviruses.
607 BMC Microbiol **9:**238.

608 8. **Matthijnssens J, Ciarlet M, McDonald SM, Attoui H, Banyai K, Brister JR,**
609 **Buesa J, Esona MD, Estes MK, Gentsch JR, Iturriza-Gomara M, Johne R,**
610 **Kirkwood CD, Martella V, Mertens PP, Nakagomi O, Parreno V, Rahman**

**M, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K, Patton JT, Desselberger U, Van Ranst M.** 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Arch Virol **156:**1397-1413.

9. **Papp H, Al-Mutairi LZ, Chehadeh W, Farkas SL, Lengyel G, Jakab F, Martella V, Szucs G, Banyai K.** 2012. Novel NSP4 genotype in a camel G10P[15] rotavirus strain. Acta Microbiol Immunol Hung **59:**411-421.

10. **Trojnar E, Sachsenroder J, Twardziok S, Reetz J, Otto PH, Johne R.** 2013. Identification of an avian group A rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses. J Gen Virol **94:**136-142.

11. **Guo D, Liu J, Lu Y, Sun Y, Yuan D, Jiang Q, Lin H, Li C, Si C, Qu L.** 2012. Full genomic analysis of rabbit rotavirus G3P[14] strain N5 in China: identification of a novel VP6 genotype. Infect Genet Evol **12:**1567-1576.

12. **Jere KC, Esona MD, Ali YH, Peenze I, Roy S, Bowen MD, Saeed IK, Khalafalla AI, Nyaga MM, Mphahlele J, Steele D, Seheri ML.** 2014. Novel NSP1 genotype characterised in an African camel G8P[11] rotavirus strain. Infect Genet Evol **21:**58-66.

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

14. **Banyai K, Laszlo B, Duque J, Steele AD, Nelson EA, Gentsch JR, Parashar UD.** 2012. Systematic review of regional and temporal trends in global

634 rotavirus strain diversity in the pre rotavirus vaccine era: insights for
635 understanding the impact of rotavirus vaccination programs. Vaccine **30**
636 **Suppl 1:**A122-130.
637 15. **Esona MD, Gautam R.** 2015. Rotavirus. Clin Lab Med **35:**363-391.
638 16. **Matthijnssens J, Van Ranst M.** 2012. Genotype constellation and evolution
639 of group A rotaviruses infecting humans. Current opinion in virology **2:**426-
640 433.
641 17. **Jere KC, Mlera L, O'Neill HG, Potgieter AC, Page NA, Seheri ML, van Dijk**
642 **AA.** 2011. Whole genome analyses of African G2, G8, G9, and G12 rotavirus
643 strains using sequence-independent amplification and 454(R)
644 pyrosequencing. J Med Virol **83:**2018-2042.
645 18. **Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM,**
646 **Palombo EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van**
647 **Ranst M.** 2008. Full genome-based classification of rotaviruses reveals a
648 common origin between human Wa-Like and porcine rotavirus strains and
649 human DS-1-like and bovine rotavirus strains. J Virol **82:**3204-3219.
650 19. **Blackhall J, Fuentes A, Magnusson G.** 1996. Genetic stability of a porcine
651 rotavirus RNA segment during repeated plaque isolation. Virology **225:**181-
652 190.
653 20. **Gouvea V, Brantly M.** 1995. Is rotavirus a population of reassortants?
654 Trends Microbiol **3:**159-162.
655 21. **Kuzuya M, Fujii R, Hamano M, Kida K, Mizoguchi Y, Kanadani T,**
656 **Nishimura K, Kishimoto T.** 2014. Prevalence and molecular

characterization of G1P[8] human rotaviruses possessing DS-1-like VP6, NSP4, and NSP5/6 in Japan. J Med Virol **86:**1056-1064.

22. **Fujii Y, Nakagomi T, Nishimura N, Noguchi A, Miura S, Ito H, Doan YH, Takahashi T, Ozaki T, Katayama K, Nakagomi O.** 2014. Spread and predominance in Japan of novel G1P[8] double-reassortant rotavirus strains possessing a DS-1-like genotype constellation typical of G2P[4] strains. Infect Genet Evol **28:**426-433.

23. **Yamamoto SP, Kaida A, Kubo H, Iritani N.** 2014. Gastroenteritis outbreaks caused by a DS-1-like G1P[8] rotavirus strain, Japan, 2012-2013. Emerg Infect Dis **20:**1030-1033.

24. **Komoto S, Tacharoenmuang R, Guntapong R, Ide T, Haga K, Katayama K, Kato T, Ouchi Y, Kurahashi H, Tsuji T, Sangkitporn S, Taniguchi K.** 2015. Emergence and Characterization of Unusual DS-1-Like G1P[8] Rotavirus Strains in Children with Diarrhea in Thailand. PLoS One **10:**e0141739.

25. **Komoto S, Tacharoenmuang R, Guntapong R, Ide T, Tsuji T, Yoshikawa T, Tharmaphornpilas P, Sangkitporn S, Taniguchi K.** 2016. Reassortment of Human and Animal Rotavirus Gene Segments in Emerging DS-1-Like G1P[8] Rotavirus Strains. PLoS One **11:**e0148416.

26. **Nakagomi T, Nguyen MQ, Gauchan P, Agbemabiese CA, Kaneko M, Do LP, Vu TD, Nakagomi O.** 2017. Evolution of DS-1-like G1P[8] double-gene reassortant rotavirus A strains causing gastroenteritis in children in Vietnam in 2012/2013. Arch Virol **162:**739-748.

27. **Bar-Zeev N, Kapanda L, Tate JE, Jere KC, Iturriza-Gomara M, Nakagomi O, Mwansambo C, Costello A, Parashar UD, Heyderman RS, French N, Cunliffe NA.** 2015. Effectiveness of a monovalent rotavirus vaccine in infants in Malawi after programmatic roll-out: an observational and case-control study. Lancet Infect Dis **15:**422-428.

28. **Cunliffe NA, Gondwe JS, Broadhead RL, Molyneux ME, Woods PA, Bresee JS, Glass RI, Gentsch JR, Hart CA.** 1999. Rotavirus G and P types in children with acute diarrhea in Blantyre, Malawi, from 1997 to 1998: predominance of novel P[6]G8 strains. J Med Virol **57:**308-312.

29. **Cunliffe NA, Gentsch JR, Kirkwood CD, Gondwe JS, Dove W, Nakagomi O, Nakagomi T, Hoshino Y, Bresee JS, Glass RI, Molyneux ME, Hart CA.** 2000. Molecular and serologic characterization of novel serotype G8 human rotavirus strains detected in Blantyre, Malawi. Virology **274:**309-320.

30. **Bar-Zeev N, Jere KC, Bennett A, Pollock L, Tate JE, Nakagomi O, Iturriza-Gomara M, Costello A, Mwansambo C, Parashar UD, Heyderman RS, French N, Cunliffe NA.** 2016. Population Impact and Effectiveness of Monovalent Rotavirus Vaccination in Urban Malawian Children 3 Years After Vaccine Introduction: Ecological and Case-Control Analyses. Clin Infect Dis **62 Suppl 2:**S213-219.

31. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD.** 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics **22:**3096-3098.

32. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol **23:**1891-1901.

33. **Ward RL, Bernstein DI, Shukla R, McNeal MM, Sherwood JR, Young EC, Schiff GM.** 1990. Protection of adults rechallenged with a human rotavirus. J Infect Dis **161:**440-445.

34. **Magagula NB, Esona MD, Nyaga MM, Stucker KM, Halpin RA, Stockwell TB, Seheri ML, Steele AD, Wentworth DE, Mphahlele MJ.** 2015. Whole genome analyses of G1P[8] rotavirus strains from vaccinated and non-vaccinated South African children presenting with diarrhea. J Med Virol **87:**79-101.

35. **Zeller M, Donato C, Trovao NS, Cowley D, Heylen E, Donker NC, McAllen JK, Akopov A, Kirkness EF, Lemey P, Van Ranst M, Matthijnssens J, Kirkwood CD.** 2015. Genome-Wide Evolutionary Analyses of G1P[8] Strains Isolated Before and After Rotavirus Vaccine Introduction. Genome Biol Evol **7:**2473-2483.

36. **Heiman EM, McDonald SM, Barro M, Taraporewala ZF, Bar-Magen T, Patton JT.** 2008. Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. J Virol **82:**11106-11116.

37. **Matthijnssens J, Heylen E, Zeller M, Rahman M, Lemey P, Van Ranst M.** 2010. Phylodynamic analyses of rotavirus genotypes G9 and G12 underscore their potential for swift global spread. Mol Biol Evol **27:**2431-2436.

724 38. **Nagaoka Y, Tatsumi M, Tsugawa T, Yoto Y, Tsutsumi H.** 2012.
725 Phylogenetic and computational structural analysis of VP7 gene of group a
726 human rotavirus G1P[8] strains obtained in Sapporo, Japan from 1987 to
727 2000. J Med Virol **84:**832-838.
728 39. **Wentzel JF, Yuan L, Rao S, van Dijk AA, O'Neill HG.** 2013. Consensus
729 sequence determination and elucidation of the evolutionary history of a
730 rotavirus Wa variant reveal a close relationship to various Wa variants
731 derived from the original Wa strain. Infect Genet Evol **20:**276-283.
732 40. **Afrad MH, Matthijnssens J, Afroz SF, Rudra P, Nahar L, Rahman R,**
733 **Hossain ME, Rahman SR, Azim T, Rahman M.** 2014. Differences in lineage
734 replacement dynamics of G1 and G2 rotavirus strains versus G9 strain over a
735 period of 22 years in Bangladesh. Infect Genet Evol **28:**214-222.
736 41. **Trang NV, Yamashiro T, Anh le TK, Hau VT, Luan le T, Anh DD.** 2012.
737 Genetic variation in the VP7 gene of rotavirus G1P[8] strains isolated in
738 Vietnam, 1998-2009. Virus Res **165:**190-196.
739 42. **Jenkins GM, Cowart LA, Signorelli P, Pettus BJ, Chalfant CE, Hannun YA.**
740 2002. Acute activation of de novo sphingolipid biosynthesis upon heat shock
741 causes an accumulation of ceramide and subsequent dephosphorylation of
742 SR proteins. J Biol Chem **277:**42572-42578.
743 43. **Huang K, Zhu H, Fan X, Wang J, Cheung CL, Duan L, Hong W, Liu Y, Li L,**
744 **Smith DK, Chen H, Webster RG, Webby RJ, Peiris M, Guan Y.** 2012.
745 Establishment and lineage replacement of H6 influenza viruses in domestic
746 ducks in southern China. J Virol **86:**6075-6083.

747 44. **Martin E, Chirivella M, Co JKG, Santiago GA, Gubler DJ, Munoz-Jordan JL,**
748 **Bennett SN.** 2016. Insights into the molecular evolution of Dengue virus type
749 4 in Puerto Rico over two decades of emergence. Virus Res **213:**23-31.
750 45. **Donaldson EF, Lindesmith LC, Lobue AD, Baric RS.** 2010. Viral shape-
751 shifting: norovirus evasion of the human immune system. Nat Rev Microbiol
752 **8:**231-241.
753 46. **Zeller M, Patton JT, Heylen E, De Coster S, Ciarlet M, Van Ranst M,**
754 **Matthijnssens J.** 2012. Genetic analyses reveal differences in the VP7 and
755 VP4 antigenic epitopes between human rotaviruses circulating in Belgium
756 and rotaviruses in Rotarix and RotaTeq. J Clin Microbiol **50:**966-976.
757 47. **Zhang S, McDonald PW, Thompson TA, Dennis AF, Akopov A, Kirkness**
758 **EF, Patton JT, McDonald SM.** 2014. Analysis of human rotaviruses from a
759 single location over an 18-year time span suggests that protein coadaption
760 influences gene constellations. J Virol **88:**9842-9863.
761 48. **Page NA, Steele AD.** 2004. Antigenic and genetic characterization of
762 serotype G2 human rotavirus strains from the African continent. J Clin
763 Microbiol **42:**595-600.
764 49. **Page NA, Steele AD.** 2004. Antigenic and genetic characterization of
765 serotype G2 human rotavirus strains from South Africa from 1984 to 1998. J
766 Med Virol **72:**320-327.
767 50. **Coulson BS, Kirkwood C.** 1991. Relation of VP7 amino acid sequence to
768 monoclonal antibody neutralization of rotavirus and rotavirus monotype. J
769 Virol **65:**5968-5974.

770 51. **Caust J, Dyall-Smith ML, Lazdins I, Holmes IH.** 1987. Glycosylation, an
771 important modifier of rotavirus antigenicity. Arch Virol **96:**123-134.
772 52. **Estes MK, Cohen J.** 1989. Rotavirus gene structure and function. Microbiol
773 Rev **53:**410-449.
774 53. **Lazdins I, Coulson BS, Kirkwood C, Dyall-Smith M, Masendycz PJ, Sonza**
775 **S, Holmes IH.** 1995. Rotavirus antigenicity is affected by the genetic context
776 and glycosylation of VP7. Virology **209:**80-89.
777 54. **Trask SD, McDonald SM, Patton JT.** 2012. Structural insights into the
778 coupling of virion assembly and rotavirus replication. Nat Rev Microbiol
779 **10:**165-177.
780 55. **Settembre EC, Chen JZ, Dormitzer PR, Grigorieff N, Harrison SC.** 2011.
781 Atomic model of an infectious rotavirus particle. EMBO J **30:**408-416.
782 56. **Potgieter AC, Page NA, Liebenberg J, Wright IM, Landt O, van Dijk AA.**
783 2009. Improved strategies for sequence-independent amplification and
784 sequencing of viral double-stranded RNA genomes. J Gen Virol **90:**1423-
785 1432.
786 57. **Iturriza-Gomara M, Elliot AJ, Dockery C, Fleming DM, Gray JJ.** 2009.
787 Structured surveillance of infectious intestinal disease in pre-school children
788 in the community: 'The Nappy Study'. Epidemiol Infect **137:**922-931.
789 58. **Chen C, Khaleel SS, Huang H, Wu CH.** 2014. Software for pre-processing
790 Illumina next-generation sequencing short read sequences. Source Code Biol
791 Med **9:**8.

792 59. **Martin M.** 2011. Cutadapt removes adapter sequences from high-throughput
793 sequencing reads EMBnet **17:**10-12.
794 60. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S,**
795 **Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes**
796 **P, Drummond A.** 2012. Geneious Basic: an integrated and extendable
797 desktop software platform for the organization and analysis of sequence
798 data. Bioinformatics **28:**1647-1649.
799 61. **Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Rapp B, Wheeler D.**
800 2002. GenBank. Nucleic Acids Res **30:**17 - 20.
801 62. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy
802 and high throughput. Nucleic Acids Res **32:**1792-1797.
803 63. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6:
804 Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-
805 2729.
806 64. **Felsenstein J.** 1988. Phylogenies from molecular sequences: inference and
807 reliability. Annu Rev Genet **22:**521-565.
808 65. **Drummond A, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by
809 sampling trees. BMC Evol Biol **7:**214.
810 66. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian
811 phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **29:**1969-1973.
812 67. **Drummond AJ, Ho SY, Phillips MJ, Rambaut A.** 2006. Relaxed
813 phylogenetics and dating with confidence. PLoS Biol **4:**e88.

68. **Hasegawa M, Kishino H, Yano T.** 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol **22:**160-174.

69. **Yang Z.** 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol **39:**306-314.

70. **Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ.** 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics **25:**1422-1423.

71. **Pond SL, Frost SD.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21:**2531-2533.

72. **Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K.** 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol **30:**1196-1205.

73. **Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL.** 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet **8:**e1002764.

74. **Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K.** 2011. A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. Molecular Biology and Evolution **28:**3033-3043.

75. **Yang Z, Nielsen R, Goldman N, Pedersen A.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431 - 449.

76. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676-679.

77. **Sali A, Blundell TL.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:**779-815.

78. **Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE.** 2000. The Protein Data Bank. Nucleic Acids Res **28:**235-242.

79. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:**W244-248.

858 **Figure Legends**

859 **Fig. 1. Rotavirus strains, genetic constellation and phylogenetic networks of**
860 **G1P[8] rotaviruses detected from Malawian children at QECH from 1997 –**
861 **2015.** (a) Schematic representation of the proportions of all genotypes detected in
862 rotavirus-positive stool samples. The size of the circle is directly proportion to the
863 detection frequency of G and P genotypes. There were no rotavirus
864 surveillance activities in 2010 hence samples were not collected. (b) Bayesian
865 Maximum Clade Credibility tree for concatenated whole rotaviruses genome
866 sequences illustrating genetic constellation and reassortment patterns for 32 pre-
867 and 18 post-vaccine G1P[8] strains from Malawi as well as the prototype Wa and
868 DS-1 strains for comparison. Genotype and genogroups assignment for each
869 segment were based on nucleotide sequence identities, assigned using RotaC.
870 Genome segment that were assigned the same genotype are shown with the same
871 colour and genotype numbers. Green represents Wa-like P[8] for VP4 and genotypes
872 1 for the rest of the other 10 gene segments; Red represents all DS-1-like genotype 2
873 for all 11 gene segments. (C) Phylogenetic network of complete concatenated whole
874 genome sequences of G1P[8] rotavirus strains detected in Malawi from 1998 –
875 2014. Branches are drawn to scale and splits in the network indicate reassortments.
876 Network clusters are colour-coded and named in accordance with their phylogenic
877 lineages (L1 – L3) that correlated with time of strain isolation before or after
878 rotavirus vaccine introduction. Cluster L1 (green) and L2 (blue) contained pre-
879 vaccine G1P[8] strains, whereas L3 (red) contained post-vaccine strains. Network
880 sub-clusters within each main cluster are shaded in blue (L1), red (L2) or orange

(L3).

**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 tMRCAs for each genome segment of the

902 Malawian atypical G1P[8] strains shown together with their 95% Highest Posterior
903 Density (HPD) intervals. (b) Absolute values for the mean and range of the
904 evolutionary rates and tMRCA at 95% HPD intervals are shown. (c) Phylogenies
905 and relative genetic diversity were estimated using the Gaussian Markov Random
906 Field (GMRF) model represents Bayesian Skygrid plots for VP1 – VP4, VP6, for NSP1
907 – NSP5 encoding genome segments. Solid lines in the GMRF plot represent the mean
908 relative genetic diversity through time.

909

910 **Fig. 5. Amino acid substitutions and structural conformation of the outer**
911 **capsid glycoprotein of Malawian G1P[8] strains.** (a) Complete VP7 sequence of
912 representative pre- and post-vaccine G1P[8] strains aligned to that of RV1
913 exhibiting amino acid substitutions that occurred within the variable regions (VR)
914 and mapped antigenic regions (AR) over time. Lineage defining amino acid amino
915 acid substitutions are highlighted in green, blue and yellow for L1, L2 and L3
916 lineages, respectively. Pre- and post-vaccine strains are shown with vertical green
917 and red bars on the right, respectively. Strains belonging to the L1, L2 and L3
918 phylogenetic clusters are shown with green, blue and red bars respectively on the
919 right. (b) Perfect alignment of superimposed VP7 structures exhibiting few
920 differences between RV1 and L1 – L3 strains. Antigenic regions A, B and C are
921 shown in white. L1 – L3 and RV1 strains are shown in yellow, green, blue and red,
922 respectively. (c - e) Surface visualisation of VP7 from the outside of the virion on 3-
923 fold axis displaying amino acid differences when structures for L1 (c), L2 (d) and L3

924 (e) G1P[8] strains were superimposed on the outer capsid glycoprotein of RV1.

925 Numbers correspond to the positions where mutations occurred.

926
927
928
929
930
931
932
933
934

**Table 1. Evolutionary selective forces and recombination in all eleven proteins of the Malawian G1P[8] rotavirus strains.**

| Protein | SLAC | FUBAR[a] | Consensus Selective | Recombination | |
|---|---|---|---|---|---|
| | dN/dS | ω (β/α) | Force | GARD | SBP |
| **VP1** | 0.0502 | 2.22 | Purifying Selection | - | - |
| **VP2** | 0.0538 | 2.67 | Purifying Selection | - | - |
| **VP3** | 0.0965 | 2.47 | Purifying Selection | - | - |
| **VP4** | 0.1052 | 3.15 | Purifying Selection | - | - |
| **VP6** | 0.0310 | 4.66 | Purifying Selection | - | - |
| **VP7** | 0.2033 | 6.01 | Purifying Selection | - | - |
| **NSP1** | 0.2239 | 4.21 | Purifying Selection | - | - |
| **NSP2** | 0.0920 | 5.62 | Purifying Selection | - | - |
| **NSP3** | 0.0870 | 5.44 | Purifying Selection | - | - |
| **NSP4** | 0.1097 | 7.10 | Purifying Selection | - | - |
| **NSP5** | 0.1097 | 7.47 | Purifying Selection | - | - |

[a]This summary table reports the means of posterior distribution of synonymous (α) and non-synonymous (β) substitution rates over **sites**, as well as the mean posterior probability for ω (=β/α) < 1 at a site.

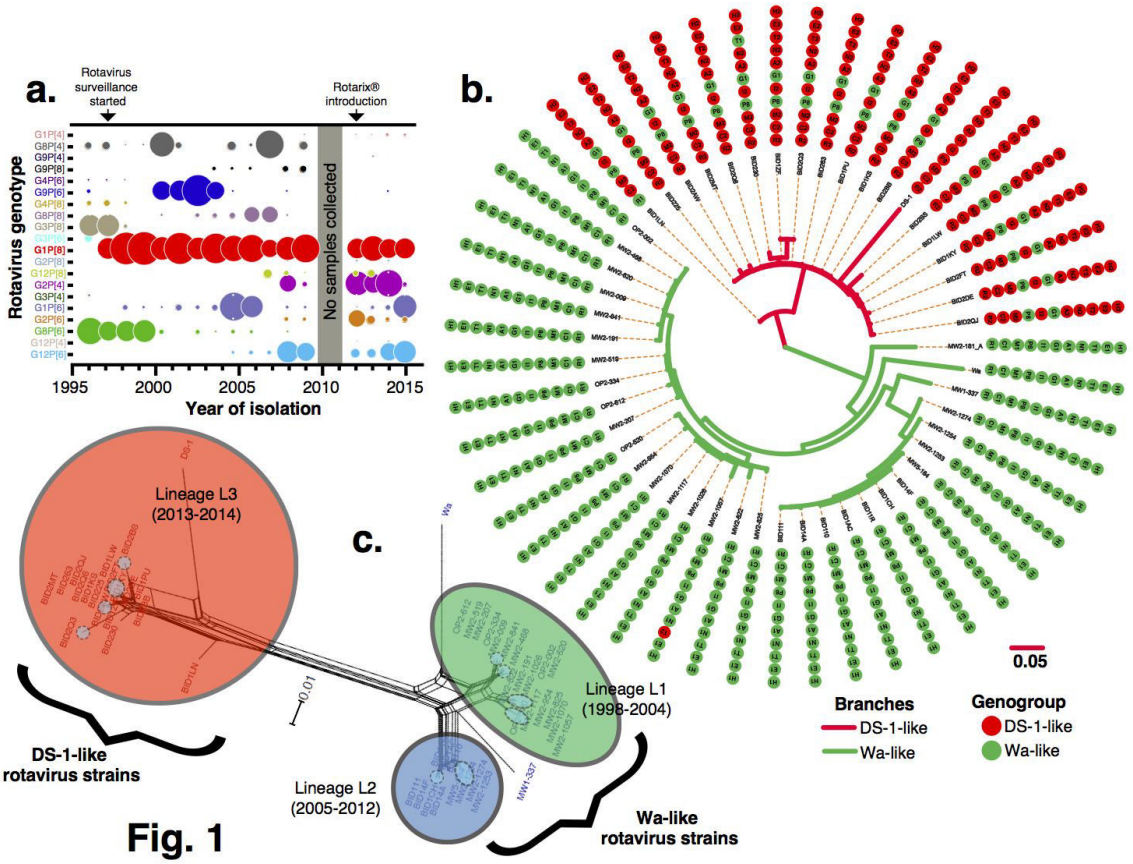


Fig. 1

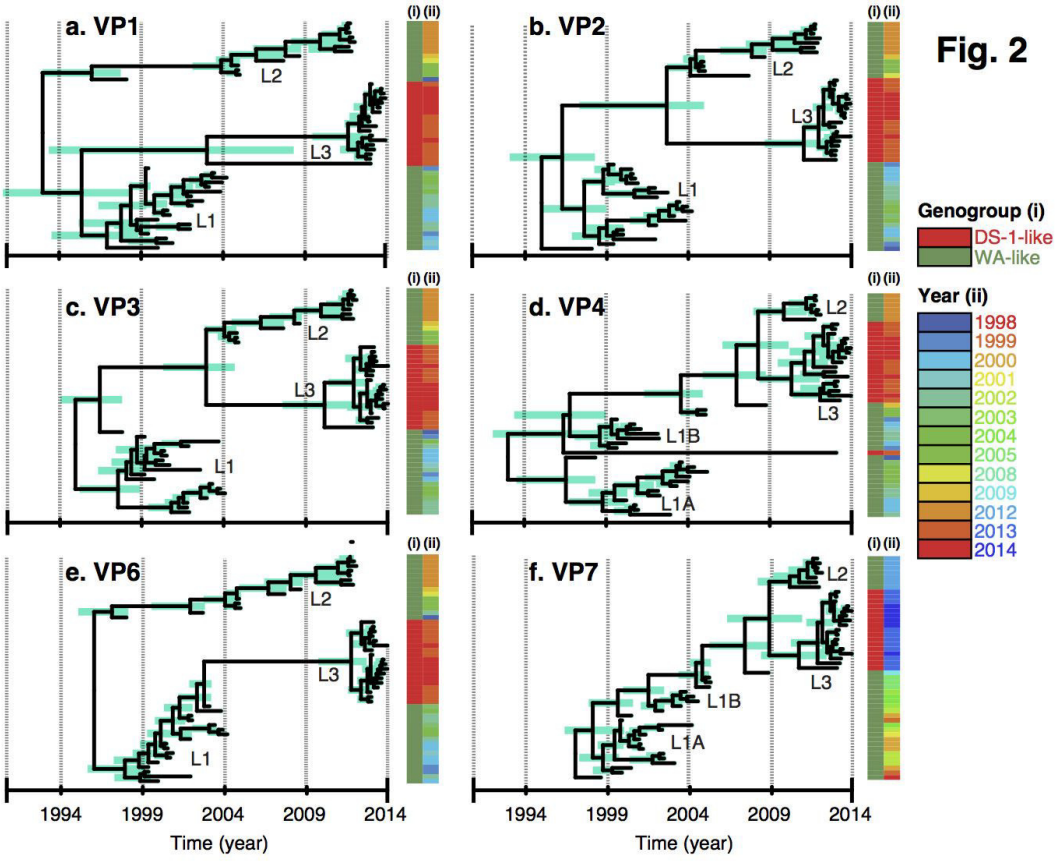


Fig. 2

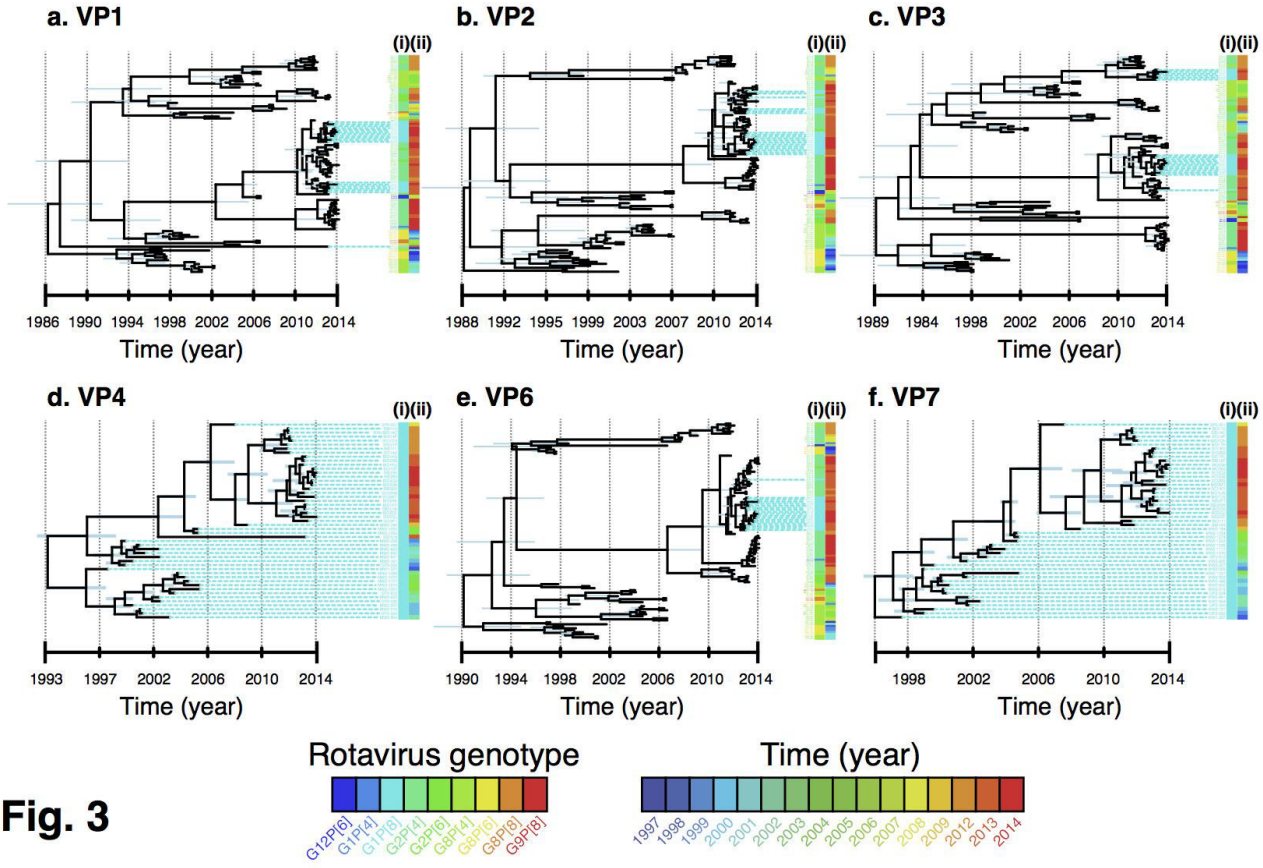


**Fig. 3**

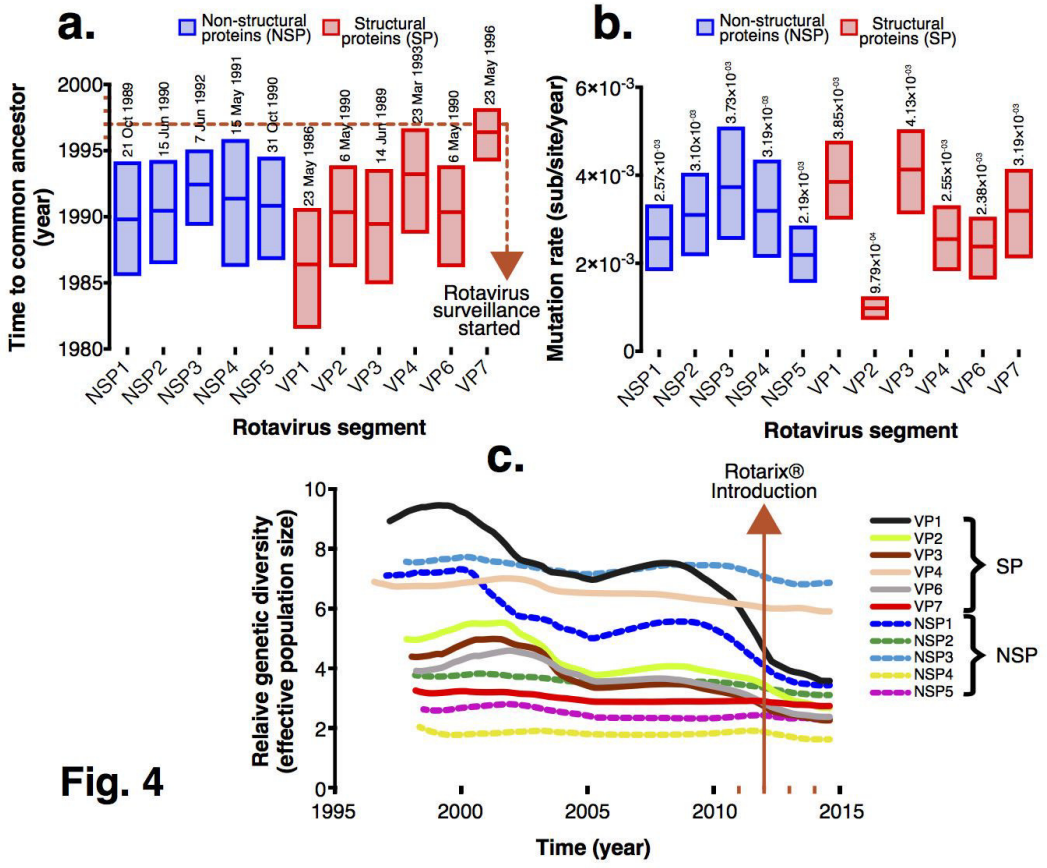


**Fig. 4**

**Fig. 5**