ORIGINAL ARTICLE

The occurrence of amino acid substitutions D96N and S242N in VP7 of emergent G2P[4] rotaviruses in Nepal in 2004-2005: a global and evolutionary perspective

Y. H. Doan · T. Nakagomi · N. A. Cunliffe · B. D. Pandey · J. B. Sherchand · O. Nakagomi

Abstract Rotavirus is the leading cause of severe diarrhea among children worldwide. Strains with G2P[4] have captured recent attention because of its abrupt increase or re-emergence in many locations in the world. In Nepal, G2P[4] strains were detected at a rate of 1% in 2003-2004, but increased to 33% in 2004-2005. Thus, the VP7 genes of 45 emergent G2 strains from Nepal were sequenced and analyzed together with a total of 339 G2VP7 sequences detected over the last 34 years that were compiled from the DNA database. We found that all Nepalese VP7 sequences had a substitution from aspartic acid to asparagine at residue 96 (D96N) that was the hallmark of the lineage termed sublineage IVa, which replaced virtually all globally circulating G2 strains during the last decade. Within sublineage IVa, further sublineages emerged, of which a sublineage termed IVa-3 was identified to have another amino acid substitution from serine to asparagine at 242 (S242N). This sublineage, to which all Nepalese sequences belonged, now became the most frequent G2 sequence globally. In conclusion, the G2VP7 gene evolved in a dynamic fashion such that new lineages emerged within the previously dominant lineage, one of which became subsequently dominant.

Y. H. Doan · T. Nakagomi · O. Nakagomi (✉)
Department of Molecular Microbiology and Immunology, Graduate School of Biomedical Sciences, Global Center of Excellence, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
e-mail: onakagom@nagasaki-u.ac.jp

N. A. Cunliffe
Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK

B. D. Pandey
Skura Raj Tropical and Infectious Disease Hospital, Kathmandu, Nepal

J. B. Sherchand
Tropical Disease Research and Prevention Center, Tribhuvan University, Kathmandu, Nepal

Introduction

Members of the species *Rotavirus A*, or group A rotavirus (hereafter referred to rotavirus), within the genus *Rotavirus* of the family *Reoviridae*, has been recognized as a very important etiological agent of severe gastroenteritis in infants and young children worldwide, causing more than 110 million diarrheal episodes, 25 million physician visits, 2 million hospitalizations, and more than half a million deaths among children less than 5 years of age, mainly in developing countries [34, 35]. The rotavirus genome consists of 11 segments of double-stranded RNA encoding six structural viral proteins (VPs) and six nonstructural proteins (NSPs); each genome segment codes for a single viral protein, with the exception of genome segment 11 which encodes two proteins (NSP5 and NSP6) [9]. VP7 and VP4 are structural proteins constituting the outer layer of the virion and the viral spikes, respectively. These two proteins define the serotype of the virus and are therefore considered to be critical for vaccine development because they are the target of neutralizing antibodies that may provide both serotype-specific and, in some instances, cross-protective immunity [18]. However, the correlates of protection were recognized as not being restricted to high levels of neutralizing antibodies [6]. Based on the antigenicity of these

two outer capsid proteins, a dual classification system by G serotypes (G for glycoprotein, coded for by genome segment 7, 8 or 9, depending on the strain) and P serotypes (P for protease-sensitive protein, coded for by genome segment 4) has been adopted [9]. However, serological assays are now being replaced by molecular typing; hence, the term "serotype" is being replaced by "genotype", and these two terms are used almost as synonyms. There is an exact correlation between G serotype and G genotype, thereby allowing the use of the same numbering system. Unfortunately, different numbering systems are adopted to designate P serotype and P genotype, with P genotype being designated within a squared bracket. While more than a dozen of G types and P types have been reported for strains detected in humans and animals and the number of potential G/P combinations is many hundred, only five major G and P combinations commonly occur in human rotaviruses; they are G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] [7, 12, 44].

Recognition of the large global burden of rotavirus disease has propelled vaccine development over the last few decades. As a result, there are currently two live-attenuated oral rotavirus vaccines licensed on the global scale. They are the monovalent, human-rotavirus-based vaccine, Rotarix, (GlaxoSmithKline Biologicals, Belgium) and the pentavalent, bovine-human reassortant vaccine, RotaTeq, (Merck & Co, USA) [42, 49]. Both vaccines have made a remarkable impact on the reduction of the number of rotavirus hospitalizations in high-income and middle-income countries [3, 4, 16, 41]. Recently, a marked increase and dominance of G2P[4] rotavirus strains was observed concurrently with the implementation of a universal mass vaccination program with the monovalent vaccine in Brazil [17], although there was also a simultaneous marked decrease in the detection rate of rotavirus [31]. It was hypothesized that vaccination with the monovalent G1P[8] vaccine created conditions in which G2P[4] rotavirus strains acquired a selective advantage over P[8] strains [31]. However, a similarly high prevalence of G2P[4] rotavirus strains was observed in Latin American countries where no rotavirus immunization program was implemented. For example, Esteban et al. [10] observed that G2 strains were the most prevalent during 2004 (44%) and 2007 (58%) in Argentina. Patel et al. [36] reported that surveillance in El Salvador, Guatemala, and Honduras showed that G2P[4] was the predominant circulating strain in 2006 (68%–81%). It was also reported from Paraguay that G2P[4] was predominant in 2006 (64%) and 2007 (46%) after a 6-year absence of this genotype [29]. Recent studies in countries such as Sierra Leone, Egypt, Jordan, Oman and Yemen have shown that G2P[4] is the most prevalent [22, 24]. Furthermore, during the post-licensure effectiveness study of Rotarix in El Salvador, de Palma et al. did not observe any increase in G2 strains [5]. Thus,

these authors suggested that the predominance of G2P[4] was due to natural genotype variation and was unrelated to the introduction of the monovalent rotavirus vaccine.

However, there are several epidemiological studies in which re-emergence or sharp increase in relative frequency of G2 strains in the population was associated with amino acid substitutions in the antigenic regions of the VP7 protein that might allow such variants to escape from recognition by neutralizing antibodies [28]. In Taiwan, for example, G2 strains re-appeared in 1992 after a five-year interval and caused an epidemic in 1993 [50]. The re-emergence and epidemic of G2 strains in Taiwan was accompanied by an amino acid substitution at position 96 from aspartic acid to asparagine (hereafter abbreviated as D96N) [50]. The amino acid at position 96 was previously described as an epitope recognized by a G2-specific neutralizing monoclonal antibody (RV5:3) [28]. Furthermore, the G2 strains that re-emerged in the UK in 1995-1996 as the second-most common genotype (16-18%) possessed this amino acid substitution and were unable to be serotyped by the enzyme-linked immunosorbent assay using monoclonal antibody RV5:3 [19]. More recently, in Thailand, there was a sharp increase in G2 strains from almost negligible percentages in 2001 and 2002 to 83% in 2003 [23], and these emergent G2 strains possessed the D96N substitution [47].

During surveillance studies in Nepal since 2003 [38, 48], where no rotavirus vaccine has been introduced, we noticed a sudden increase in the detection rate of G2 strains to 39% in 2004-2005 as compared to 1% in the preceding season (2003-2004). This observation prompted us to sequence the VP7 gene of Nepalese G2 strains, for which there had been no sequence information available, in order to examine the presence of the D96N substitution. More recently, a further amino acid substitution (serine to asparagine at 242: S242N) was described based on Brazilian G2 strains to constitute a new sublineage within the lineage hallmarked by the D96N substitution [30]. Thus, we extensively searched the DNA database and extracted relevant G2 VP7 sequences deposited over the last 34 years from across the world. We then conducted a phylogenetic analysis in order to better understand these and other major substitutions in the VP7 antigenic regions from the global and evolutionary perspective and to provide a basis on which to assess the changes in the G2 VP7 genes as rotavirus vaccines penetrate into broader areas of the world.

Material and methods

Specimens

Among the stool specimens that were collected from children with acute diarrhea attending Kanti Children's

Hospital, Kathmandu, Nepal, from August 2004 through July 2005, we previously reported that there were 67 G2 rotavirus-positive specimens identified by the reverse transcription (RT)-PCR genotyping method [38]. For this study, there remained 57 genomic RNAs available, which were extracted using the QIAamp Viral RNA Mini Kit (QIAGEN Sciences, Germantown, MD, USA).

Nucleotide sequencing

Genomic RNAs that were extracted previously using the QIAamp Viral RNA Mini Kit (QIAGEN Sciences, Germantown, MD, USA) were reverse transcribed and amplified with a pair of VP7 consensus primers, Beg9 and End9 [15], using the AcessQuick RT-PCR system (Promega Corporation, Madison, WI, USA). The amplified products were then purified using a QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. Nucleotide sequencing reactions were performed by fluorescent dideoxy chain termination chemistry using the BigDye Terminator Cycle Sequencing Ready Reaction Kit, version 3.1 (Applied Biosystems, Foster City, CA, USA), and nucleotide sequences were determined using an ABI Prism 3730 Genetic Analyzer (Applied Biosystems). The nucleotide sequences thus obtained were aligned using the Megalign program in the Lasergene 8 software package (DNAstar, Inc. Madison, WI, USA).

Compilation of the G2 VP7 sequences in the DNA databases

In order to place the VP7 sequence results from Nepalese strains into a broader evolutionary context, we made an exhaustive DNA database search using the BLAST program (http://blast.ncibi.nlm.nih.gov/) with the DS-1 VP7 sequence as a query sequence and retrieved 500 sequences so that all potential G2 sequences might be included. We then used for our analysis only those G2 VP7 sequences that included at least 615 nucleotide sequences from position 307 to position 921 (corresponding to amino acid residues 87-291) and for which information about the year and the place of isolation were available either in the DNA database entry or in the relevant literature (a list of all sequences used, including names of RV strains, year of isolation and database accession number, is provided in Supplementary Data 2). Accordingly, the number of G2 VP7 sequences we included from the DNA database was 339.

Phylogenetic analysis of the G2 VP7 genes

Phylogenetic analysis was performed using the MEGA4 package [46]. The nucleotide sequences of Nepalese G2 strains as well as those compiled from the DNA database were aligned using the CLUSTALW program, and the genetic distances between sequences were calculated by the Kimura two-parameter method [25]. A phylogenetic tree was then constructed by using the neighbor-joining method [43]. The statistical significance at the branching point was calculated with 1,000 pseudo-replicate datasets. A lineage was defined as a cluster of sequences having a bootstrap probability of 70% or higher at the branching point. Similarly, within the lineage, a sublineage was defined as a cluster of sequences that had a bootstrap probability of 70% or higher. In addition, amino acid substitutions in the antigenic regions of VP7 were analyzed, and their pattern was defined for the substitutions of the amino acids at residues 87, 96, 213, and 242.

Chronological and geographical variations of G2 VP7 genes

All G2 VP7 sequences were categorized according to the place of detection into five groups: Asia, Europe, Africa, Americas and Oceania, and also according to the years of detection into seven quinquennial periods: 1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005 and 2006-2009. The number and the proportion of strains belonging to each lineage or sublineage were counted for each of the abovementioned continents or periods.

Results

Of 67 rotavirus specimens that were typed as G2 in the study performed in Nepal in 2004-2005 [38], we were able to amplify and sequence the entire open reading frame of the VP7 gene from 45 specimens. The nucleotide sequence identity between any pair of those 45 VP7 sequences ranged from 97.6 to 100%, and their deduced amino acid identity ranged from 97.6 to 100%. As for the amino acid residues that were previously identified as important epitopes in the antigenic regions of VP7 [2, 13, 14, 19, 28, 30, 33, 47, 50], all of 45 Nepalese sequences had the same amino acids: threonine at position 87 (antigenic region A), asparagine at 96 (region A), aspartic acid at 213 (region C), and asparagine at 242 (region F).

A phylogenetic tree was constructed using the 45 Nepalese G2 nucleotide sequences determined in this study and 339 G2 sequences that were detected between 1976 and 2009 in 31 countries on five continents and were available in the DNA database. In this global collection, four strains were detected between 1976 and 1980, 21 were detected between 1981 and 1985, 11 were detected between 1986 and 1990, 54 were detected between 1991 and 1995, 43 were detected between 1996 and 2000, 86
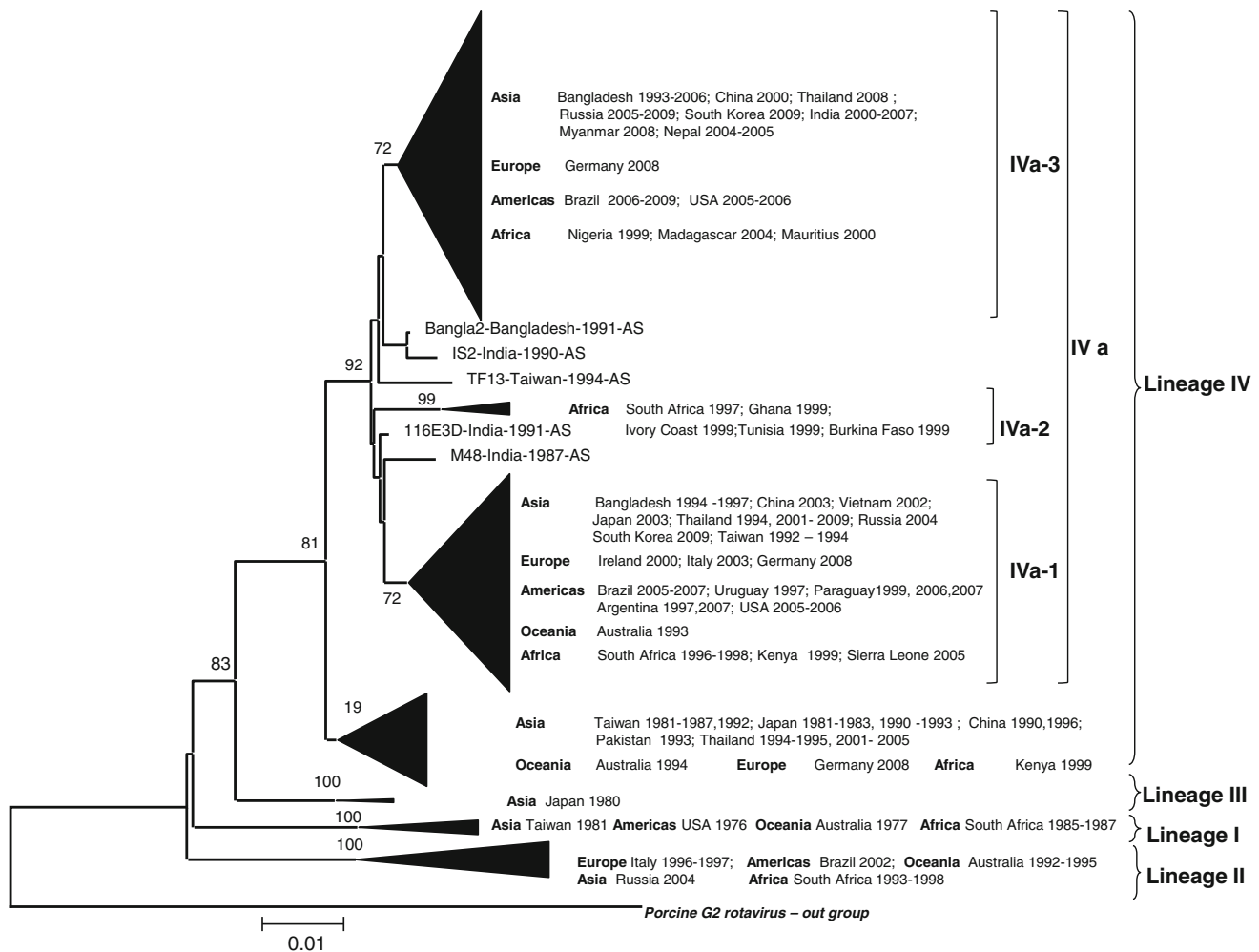
Fig. 1 A simplified phylogenetic tree constructed by the neighbor-joining method based on the VP7 nucleotide sequences of human rotavirus genotype G2 strains (The nucleotide sequences compared spanned a minimum of 615 nucleotides from position 307 to position 921, which corresponded to amino acid residues 87-291). This tree was rooted with porcine G2 rotavirus strains. Since the number of sequences belonging to each lineage or sublineage is too large to be presented in a figure, sequences are condensed into a triangle to the right of which information about the geographic origin and the year of detection of relevant sequences is given. At the right-hand side of the figure is a designation of lineages and sublineages identified in this study. The genetic distance (substitutions per site) is indicated at the bottom. Percent bootstrap support is indicated by the value at each node when the values were 70% or larger. Notice one exception, where there is a triangle with a bootstrap support of 19%, which conglomerates sequences that belonged to lineage IV but did not form a single sublineage. The complete version of this phylogenetic tree is in Supplementary Data 1

were detected between 2001 and 2005, and 120 strains detected between 2006 and 2009.

Four lineages were identified in the phylogenetic tree (Fig. 1). The first lineage (I) contained eight G2 strains from the United States of America, Australia, Taiwan and South Africa that were detected in the 1970s and 1980s. The second lineage (II) contained 16 G2 strains detected between 1992 and 2004 from South Africa, Australia, Italy, Russia and Brazil. The third lineage (III) contained only two strains from Japan detected in a single year, 1980. The fourth lineage (IV) contained 358 G2 strains detected from 1981 to 2009. In this large lineage IV, one sublineage (IVa) that contained 306 strains was identified with a bootstrap probability of 92%. The rest of the lineage IV sequences,

which included 52 sequences detected in Asia, Africa, Australia and Europe, were not considered to form a sublineage due to a low bootstrap probability (19%), and their detection spanned over the period from 1981 in Taiwan to as recently as 2008 in Germany (Fig. 1).

Within sublineage IVa, further sublineages were identified with sufficiently high bootstrap probabilities of 70-100%: sublineage IVa-1, which contained 123 sequences from five continents over the period between 1992 and 2009, the small sublineage IVa-2, which contained seven sequences only from Africa in 1997 and 1999, and sublineage IVa-3, which contained 171 sequences from Asia, Africa, Europe and Americas, but not from Oceania, over the period between 1993 and 2009. All 45 Nepalese G2

Fig. 2 Pairwise sequence identities of VP7 genes: The number of the occurrence of different percentages of identities between lineages is represented by open square columns and that within lineages is represented by filled square columns
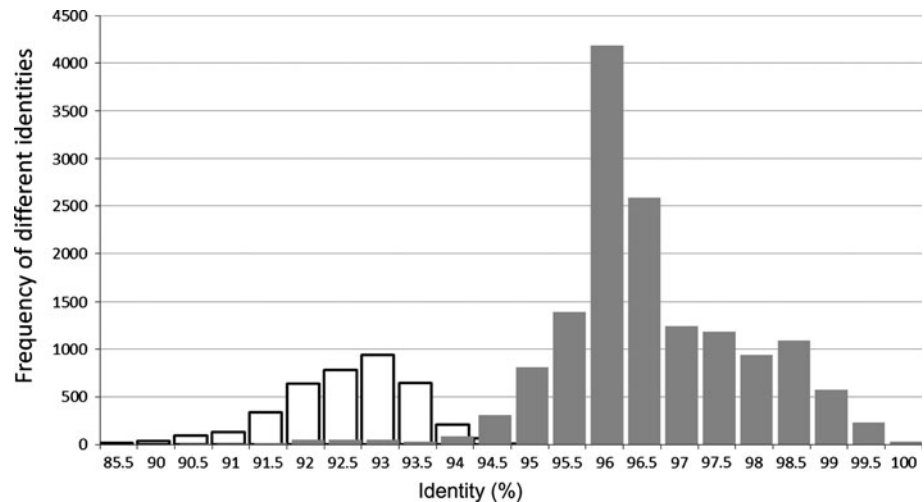
sequences determined in this study were clustered in sublineage IVa-3 (Fig. 1). The absence of sequences with sublineage IVa-3 from Oceania may reflect the lack of G2 sequence deposition in the DNA databases from Australia since the middle of the 1990s. The vast majority (97%) of G2 strains detected in the last decade belonged to sublineage IVa. Exceptions were two strains detected in Brazil in 2002 and one strain detected in Russia in 2004 that belonged to lineage II, three strains detected in Thailand in 2001 and 2005, and one strain from Germany in 2008 that were found among the rest of lineage IV (Fig. 1).

Pairwise sequence identities between VP7 sequences were calculated. There were two distinct peaks, one corresponding to identities between lineages and the other corresponding to identities within the same lineages, lending support for the division of global G2 sequences into four lineages (Fig. 2). The peak for the former was 93% and that for the latter was 96%.

The pattern of amino acid substitutions occurring in each lineage (and sublineage) of the G2 VP7 genes was analyzed relative to the DS-1 strain (lineage I) with respect to the amino acids present at four positions in the VP7 antigenic regions previously recognized as potential epitopes for G2 strains: at 87 and 96 in antigenic region A (residues 87-101), 213 in region C (residues 208-221) and 242 in region F (residues 233-242). The lineage I sequences had the same amino acids at all four positions (Fig. 3). The lineage II sequences had three amino acid substitutions from the lineage I sequences: A87T (alanine to threonine), N213S (asparagine to serine) and N242S (asparagine to serine). While the lineage II sequences, with two exceptions, shared the N213S and N242S substitutions, the A87T substitution occurred only in half of the lineage II sequences, and the amino acid remained unchanged at this position in the other half of the sequences (Fig. 3). The lineage III sequences had two substitutions (A87T and N242S) from the lineage I sequences (Fig. 3). The

sequences in lineage IV that did not belong to sublineage IVa (indicated as lineage IV except IVa in Fig. 3) had three amino acid substitutions from the lineage I sequences: A87T, N213D and N242S. While they all shared the N213D and N242S substitutions, the A87T substitution occurred in only one third of the sequences, and the amino acid remained unchanged at this position in the remaining two thirds of the sequences. The salient feature of sublineage IVa sequences is the D96N substitution. What was notable within this sublineage IVa was the amino acid substitution at 242; there was serine in the IVa-1 and IVa-2 sublineages, whereas there was asparagine in the IVa-3 sublineage (Fig. 3).

Figure 4 depicts the temporal distribution of different lineages and sublineages of the G2 VP7 sequences over the last 34 years since 1976. Although the number of sequences available was limited for lineages I-III, there was a successive lineage shift from lineage I to lineage II, then to co-circulation of sublineages IVa-1 and IVa-3. Of note was the fact that the persistence over seven quinquennial periods of the sequences that clustered into lineage IV but that did not belong to sublineage IVa (indicated as sublineage IV except IVa in Fig. 4), although the number of sequences identified in the last quinquennial period was only one (the one detected in Germany in 2008). Lineage I was present only for a limited time period in the last quinquennial period in the 1970s and were never detected thereafter (Fig. 4). The strains belonging to lineage II circulated at low frequencies between 1991 and 2005, yet their circulation was global on the five continents.

Discussion

This study demonstrates that a sharp increase in the relative frequency of G2 strains in Nepal in 2004-2005 was accompanied by the D96N substitution in the VP7 gene, as

Fig. 3 (left panel)

			Antigenic regions A / 87	A / 96	C / 213	F / 242
DS-1	USA	1976	A	D	N	N
Lineage I						
DS-1	USA	1976	A	D	N	N
Hu5	Australia	1977	A	D	N	N
GR410	South Africa	1985	A	D	N	N
GR514	South Africa	1987	A	D	N	N
TW6	Taiwan	1981	A	D	N	N
Lineage II						
92A	Australia	1992	A	D	S	S
1831GR	South Africa	1993	A	D	S	S
95A	Australia	1995	A	D	S	S
H105-Palermo	Italy	1996	T	D	S	S
64SB	South Africa	1996	A	D	S	N
906SB	South Africa	1998	A	D	S	S
Rj5323	Brazil	2002	T	D	S	S
Rj5619	Brazil	2002	T	D	N	S
Nov04-H383	Russia	2004	T	D	S	S
Lineage III						
KUN	Japan	1980	T	D	N	S
Lineage IV (except IVa)						
S2	Japan	1981	A	D	D	S
TA81	Taiwan	1981	T	D	D	S
TA20	Taiwan	1982	A	D	D	S
Japan0022	Japan	1983	A	D	D	S
TA28	Taiwan	1983	T	D	D	S
TA34	Taiwan	1983	A	D	D	S
TW348	Taiwan	1984	T	D	D	S
TW424	Taiwan	1985	T	D	D	S
TW786	Taiwan	1987	A	D	D	S
CHIN5	China	1990	A	D	D	S
90H115	Japan	1990	A	D	D	S
TD7	Taiwan	1992	A	D	D	S
JAPAN085	Japan	1993	A	D	D	S
PAK426	Pakistan	1993	A	D	D	S
94E	Australia	1994	T	D	D	S
TWCH105	Thailand	1994	A	D	D	S
TB-Chen	China	1996	A	D	D	S
KY3303	Kenya	1999	A	D	D	S
CMH268	Thailand	2001	A	D	D	S
CU100G2	Thailand	2005	A	D	D	S
GER177	Germany	2008	A	N	D	S

Fig. 3 (middle panel)

			Antigenic regions A / 87	A / 96	C / 213	F / 242
DS-1	USA	1976	A	D	N	N
Sublineage IVa- except IVa (1 -3)						
M48	India	1987	T	N	D	S
IS2	India	1990	T	N	D	S
116E3D	India	1991	T	N	D	S
TF13	Taiwan	1994	T	N	D	S
Bangla2	Bangladesh	1991	T	N	D	N
Sublineage IVa -1						
TD64	Taiwan	1992	T	N	D	S
E212	Australia	1993	T	N	D	S
Bangla84	Bangladesh	1994	T	N	D	S
TF85	Taiwan	1994	T	N	D	S
Bangla33	Bangladesh	1995	T	N	D	S
356PT	South Africa	1996	T	N	D	S
Bangla164	Bangladesh	1997	T	N	D	S
3958GR	South Africa	1997	T	N	D	S
Mvd9707	Uruguay	1997	T	N	N	S
Arg83	Argentina	1997	T	N	N	S
4372JB	South Africa	1998	T	N	D	S
KY3103	Kenya	1999	T	N	D	S
Py40699	Paraguay	1999	T	N	D	S
CIT-220RV	Ireland	2000	T	N	N	S
CMH277	Thailand	2001	T	N	D	S
VN-11	Vietnam	2002	T	N	D	S
T-160	Thailand	2003	T	D	D	S
26-Palermo	Italy	2003	T	N	D	S
J-4708	Japan	2003	T	N	D	S
CMH017	Thailand	2003	T	N	D	S
CH-86	China	2003	T	N	D	S
Nov04-H391	Russia	2004	T	N	D	S
11257 05MS	Brazil	2005	T	N	D	S
11531 05AC	Brazil	2005	T	N	N	S
SL244	Sierra Leone	2005	T	N	D	S
CU90G2	Thailand	2005	T	N	D	S
CU127G2	Thailand	2005	T	D	D	S
11837 06AC	Brazil	2006	T	N	N	S
Py1033ASR06	Paraguay	2006	T	N	D	S
LB2772	USA	2006	T	N	D	S
14322 07MG	Brazil	2007	T	N	D	S
Arg2395	Argentina	2007	T	N	D	S
Py1138ASR07	Paraguay	2007	T	N	D	S
CU128-BK	Thailand	2008	T	N	D	S
GER84	Germany	2008	T	N	D	S
Seoul-556	South Korea	2009	T	N	D	S
CU497-BK	Thailand	2009	T	N	D	S

Fig. 3 (right panel)

			Antigenic regions A / 87	A / 96	C / 213	F / 242
DS-1	USA	1976	A	D	N	N
Sublineage IVa-2						
4476PT	South Africa	1997	T	N	D	S
BF 3767	Burkino Faso	1999	T	N	D	S
GH1803	Ghana	1999	T	N	D	S
CI-1735	Ivory Coast	1999	T	N	D	S
TN 1529	Tunisia	1999	T	N	D	S
Sublineage IVa-3						
Bangla66	Bangladesh	1993	T	N	D	N
NG5113	Nigeria	1999	T	N	D	N
MR 4717	Mauritius	2000	T	N	D	N
SC4	India	2000	T	N	D	N
XJ00-486	China	2000	T	N	D	N
Agroj5	Bangladesh	2001	T	N	D	N
ISO-8	India	2001	T	N	D	N
Agroj50	Bangladesh	2002	T	N	D	N
04N596	Nepal	2004	T	N	D	N
CRI 21113	India	2004	P	N	D	N
CRI 21162	India	2004	T	N	D	N
DR0017	Madagascar	2004	T	N	D	N
SK138	Bangladesh	2004	T	N	D	N
05N011	Nepal	2005	T	N	D	N
mani-4	India	2005	T	N	D	N
MMC6	Bangladesh	2005	T	N	Y	N
Nov05-202	Russia	2005	T	N	D	N
SK424	Bangladesh	2005	T	N	D	N
12842 06ES	Brazil	2006	T	N	D	N
LB2764	USA	2006	T	N	D	N
Nov06-885	Russia	2006	T	N	D	N
PSAL1920-C/PA	Brazil	2006	T	N	D	N
SK464	Bangladesh	2006	T	N	D	N
14157 07RS	Brazil	2007	T	N	N	N
mani-257	India	2007	T	H	D	N
mani-384	India	2007	T	N	D	N
14919 08RS	Brazil	2008	T	N	D	N
CU304-NR	Thailand	2008	T	N	D	N
GER31	Germany	2008	T	N	D	N
MMA08-5	Myanmar	2008	T	N	D	N
Omsk08-408	Russia	2008	T	N	D	N
16054 09MA	Brazil	2009	T	N	D	N
Nov09-D152	Russia	2009	T	N	D	N
Seoul-397	South Korea	2009	T	N	D	N

Fig. 3 The amino acids at residues 87, 96, 213 and 242 in the antigenic regions A (residues 87-101), C (residues 208-221) and F (residues 233-242) of the VP7 sequences that belong to each of the lineages and sublineages identified in this study. While the amino acids present at these four residues were examined for all 384 G2 VP7 sequences, including 45 Nepalese sequences (and provided in Supplementary Data 2), sequences were selected to present in the figure such that there was no duplication of the sequences from the same geographical location and in the same year except in cases where amino acids at any of these four positions were different

was observed in the studies conducted in Taiwan [50], the United Kingdom [19], South Africa [32, 33], Italy [2], and Thailand [47]. Moreover, although the presence of the D96N substitution was not explicitly described in other reports in association with the re-emergence or an increase of G2 strains, this association was confirmed by our analysis. For example, in Paraguay, 39% and 64% of circulating strains in 2005 and 2006, respectively, were G2, whereas no G2 strains were reported from 2000 to 2004 [29]. The VP7 genes of all seven strains reported in the paper by Martinez et al. [29] contained the D96N substitution, and they belong to sublineage IVa according to our classification. In Bangladesh, there was a constant circulation of G2 strains at low frequencies during the early years of the 2000s, and then there was a sharp increase to 43% in 2005-2006 [40]. In the same or nearby regions in Bangladesh, other investigators also reported a high detection rate of G2 strains (43-54%) during the same period [1, 8, 37]. The nucleotide sequence data from these

studies, as far as they were available in the databases, were analyzed, and it was shown that there was the same D96N substitution and that all strains analyzed belonged to sublineage IVa.

However, it is important to understand such substitutions of G2 sequences in the global and evolutionary context. For this purpose, we did an extensive database search that included 339 G2 sequences deposited over the last three decades, analyzed the amino acid substitutions having occurred at the four potential neutralization epitopes in the VP7 antigenic regions, and examined whether there was any distinct relationship between the phylogenetic lineages of G2 VP7 genes and the observed amino acid substitutions. We found that each lineage or sublineage had essentially a set (or two sets) of amino acid combinations at four previously recognized amino acid positions in the G2 VP7 antigenic regions (Fig. 3). Most notably, this study showed that a single D96N substitution was associated with the emergence of sublineage IVa. The first detection of strains
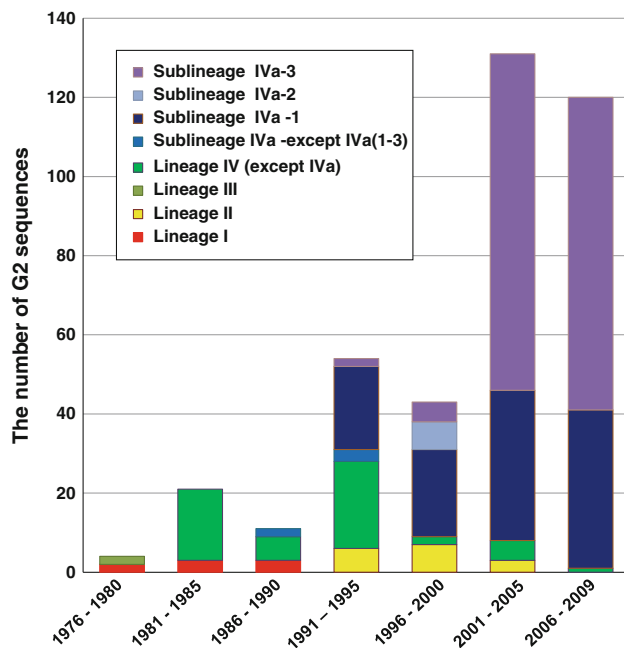
Fig. 4 The number of G2 VP7 sequences belonging to each lineage and sublineage during seven quinquennial periods: 1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005 and 2006-2009. Note, however, that the bars indicated as lineage IV except IVa and sublineage IVa-except IVa1-3 did not actually form a single lineage but a collection of sequences that did not belong to IV or IVa

carrying this sublineage sequence was in India in 1987. It then increased in the following years and rapidly dominated globally in the 2000s (Fig. 4). It is therefore tempting to speculate that current, globally predominant G2 rotavirus strains originated in Asia.

Similarly notable was the observation that there were two amino acid substitution patterns within sublineage IVa, where the difference was whether serine or asparagine was present at position 242; one was serine at 242 and the other asparagine. Serine at 242 was the hallmark substitution of sublineages IVa-1 and IVa-2 (in addition to the A87T, D96N and N213D substitutions), whereas asparagine at 242 was characteristic of sublineage IVa-3. However, it was not clarified in this study whether the asparagine at 242 in the sublineage IVa-3 sequence represented a back mutation from serine that was present in lineage II and later lineages or the original amino acid, which had remained unchanged from the lineage I sequence. While both sublineage IVa-1 (which had serine at 242) and sublineage IVa-3 (which had asparagine at 242) were abundantly detected at similar frequencies in different geographic locations (mostly Asia, the Americas, and Africa), there was a significant difference in their detection rates in the respective quinquennial periods over the last two decades. The rate of detection of sublineage IVa-1 remained rather constant, meaning its proportion among sublineage IVa

sequences was diminishing. In sharp contrast, there was an increasing number as well as an increasing proportion of sublineage IVa-3 sequences detected, from a few sequences (5-10%) in the 1990s to 80 sequences (70%) in the 2000s. In this context, the recent observation made about 15 Brazilian G2 sequences by Mascarenhas et al. [30] is noteworthy: local Brazilian G2 sequences with the D96N substitutions were clustered into two sublineages IIa and IIc, which correspond to sublineage IVa-1 and IVa-3 in this study, respectively. They described the presence of the former sequences (serine at 242) in 2006 and that of the latter (asparagine at 242) in 2006 (only one sequence) and 2008. Also in Brazil, Gomez et al. [13] reported that, while all 54 G2 sequences detected between 2005 and 2009 had the D96N substitution (belonging to sublineage IVa), serine at 242 (sublineage IVa-1) was observed in the first three years (2005-2007) and asparagine at 242 (sublineage IVa-3) was observed in the last four years (2006-2009). Thus, it appears that there was a shift of sublineage from IVa-1 to IVa-3 in Brazil. Whether or not this sublineage shift is global is yet to be determined.

Also worthy of mention is the persistence of the sequences in lineage IV that did not belong to sublineage IVa (indicated as lineage IV except IVa in Figs. 3 and 4). All of these sequences had two amino acid substitutions from lineage I (N213D and N242S) and they might have contributed to the predominance of these "lineage IV except IVa" sequences in the three quinquennial periods between 1981 and 1995 (Fig. 4). Thereafter, these sequences decreased substantially, but they persisted for another three quinquennial periods. Such persistence may possibly be related to the abundance of genetic diversity, as indicated by the fact that they did not form a single lineage, and the small minor sublineages may have represented surviving sequences that found their niches.

Lineage II sequences were possessed only by G2 strains from five countries that were detected during the 1990s, with the exception of two G2 strains from Brazil in 2002 and one strain in Russia in 2004. It is of particular interest that only one G2 strain in lineage II was detected in Asia despite the fact that two-thirds of strains analyzed in this study were from Asian countries.

Reviewing the evolutionary history of G2 VP7 sequences over the last three decades with reference to lineage diversification in the phylogenetic tree and amino acid substitutions occurring in the VP7 antigenic regions, it is remarkable that virtually 100% of the VP7 sequences of contemporary G2 strains belong to sublineage IVa and that this shift was accompanied by the D96N substitution. In this regard, it should be mentioned that the amino acid at residue 96 was previously identified as a neutralizing epitope in an experiment using neutralization monoclonal antibody escape mutants [28]. It is therefore very likely

that the D96N substitution, from a negatively charged polar amino acid (aspartic acid) to an uncharged polar amino acid (asparagine), contributed to a major antigenic change that may have helped the emergence and substantial increase of G2 strains observed in various geographic locations, including Nepal. It should be noted, however, that the high prevalence of G2 strains of sublineage IVa may not only be due to the D96N substitution in the VP7 gene but also to ongoing reassortment events within G2 strains involving various genes, as observed by Gomez et al. [13]; in this way, emerging variants may gain a competitive advantage (fitness) in particular population environments to become dominant strains. It would be of interest to investigate whether similar reassortment events were involved in the emergent Nepalese G2 strains by examining the other genome segments.

There are some limitations in this study. First, the unavailability of G2 rotaviruses before the abrupt increase in 2004-2005 in Nepal made less conclusive the association between the G2 increase and the D96N substitution. Second, uneven availability of sequence information may have produced a distorted picture of the evolutionary history of the G2 VP7 genes; most notably, the VP7 sequences of G2 strains detected in Europe and Australia in the most recent quinquennial period are lacking. Overcoming the latter limitation is considered feasible, since rotavirus strain surveillance has been actively performed in such countries as Australia, UK, and Italy for long periods of time [11, 20, 21, 26, 45].

In conclusion, a sharp increase in the relative frequency of G2 strains in Nepal in 2004-2005 occurred concurrently with the D96N substitution that was the hallmark of strains belonging to sublineage IVa, by which virtually all globally circulating G2 strains were replaced. However, within sublineage IVa, further variants emerged belonging to sublineage IVa-3 that had an additional S242N substitution and then became predominant. All Nepalese strains analyzed in this study belonged to this sublineage. Since there is continuous emergence of variants that have amino acid substitutions in the four potential epitopes in the antigenic regions of VP7, it is likely that rotavirus evolves by escaping the immunity formed in the population. Given the limited spectrum of amino acid substitutions observed in the G2 VP7 antigenic regions, directional selection rather than diversifying selection is likely to have operated on these amino acid residues. Since an increasing number of molecular evolutionary methods for detecting natural selection are available [27, 39], it would be interesting to determine the kind of natural selection forces that drive the evolution of the G2VP7 genes. In addition, it would also be interesting to investigate whether reassortment events were involved in the emergent Nepalese G2 strains, as was documented for the G2 strains from Brazil [13].

References

1. Ahmed K, Ahmed S, Mitui MT, Rahman A, Kabir L, Hannan A, Nishizono A, Nakagomi O (2010) Molecular characterization of VP7 gene of human rotaviruses from Bangladesh. Virus Genes 40:347–356
2. Arista S, Giammanco GM, De Grazia S, Colomba C, Martella V, Cascio A, Iturriza-Gòmara M (2005) G2 rotavirus infections in an infantile population of the South of Italy: variability of viral strains over time. J Med Virol 77:587–594
3. Boom JA, Tate JE, Sahni LC, Rench MA, Hull JJ, Gentsch JR, Patel MM, Baker CJ, Parashar UD (2010) Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. Pediatrics 125:199–207
4. Correia JB, Patel MM, Nakagomi O, Montenegro FM, Germano EM, Correia NB, Cuevas LE, Parashar UD, Cunliffe NA, Nakagomi T (2010) Effectiveness of monovalent rotavirus vaccine (Rotarix) against severe diarrhea caused by serotypically unrelated G2P[4] strains in Brazil. J Infect Dis 201:363–369
5. de Palma O, Cruz L, Ramos H, de Baires A, Villatoro N, Pastor D, de Oliveira LH, Kerin T, Bowen M, Gentsch J, Esposito DH, Parashar U, Tate J, Patel M (2010) Effectiveness of rotavirus vaccination against childhood diarrhoea in El Salvador: case–control study. Br Med J 340:c2825
6. Desselberger U, Huppertz HI (2011) Immune responses to rotavirus infection and vaccination and associated correlates of protection. J Infect Dis 203:188–195
7. Desselberger U, Iturriza-Gómara M, Gray JJ (2001) Rotavirus epidemiology and surveillance. Novartis Found Symp 238: 125–147 (discussion 147–152)
8. Dey SK, Hayakawa Y, Rahman M, Islam R, Mizuguchi M, Okitsu S, Ushijima H (2009) G2 strain of rotavirus among infants and children, Bangladesh. Emerg Infect Dis 15:91–94
9. Estes M, Kapikian A (2007) Rotaviruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology. Wolters Kluwer Health/Lippincott, Williams and Wilkins, Philadelphia, pp 1917–1974
10. Esteban LE, Rota RP, Gentsch JR, Jiang B, Esona M, Glass RI, Glikmann G, Castello AA (2010) Molecular epidemiology of group A rotavirus in Buenos Aires, Argentina 2004–2007: reemergence of G2P[4] and emergence of G9P[8] strains. J Med Virol 82:1083–1093
11. Finamore E, Vitiello M, Kampanaraki A, Rao M, Galdiero M, Galdiero E, Bevilacqua P, Gallo MA, Galdiero M (2011) G2 as an emerging rotavirus strain in pediatric gastroenteritis in southern Italy. Infection 39:113–119
12. Gentsch JR, Gentsch JR, Laird AR, Bielfelt B, Griffin DD, Banyai K, Ramachandran M, Jain V, Cunliffe NA, Nakagomi O, Kirkwood CD, Fischer TK, Parashar UD, Bresee JS, Jiang B,

Glass RI (2005) Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. J Infect Dis 192:146–159

13. Gómez MM, de Mendonça MC, Volotão Ede M, Tort LF, da Silva MF, Cristina J, Leite JP (2011) Rotavirus A genotype P[4]G2: Genetic diversity and reassortment events among strains circulating in Brazil between 2005 and 2009. J Med Virol 83:1093–1106

14. Gómez MM, Tort LF, Volotao Ede M, Recarey R, Moratorio G, Musto H, Leite JP, Cristina J (2011) Analysis of human P[4]G2 rotavirus strains isolated in Brazil reveals codon usage bias and strong compositional constraints. Infect Genet Evol 11:580–586

15. Gouvea V, Glass RI, Woods P, Taniguchi K, Clark HF, Forrester B, Fang ZY (1990) Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. J Clin Microbiol 28:276–282

16. Gurgel RG, Bohland AK, Vieira SC, Oliveira DM, Fontes PB, Barros VF, Ramos MF, Dove W, Nakagomi T, Nakagomi O, Correia JB, Cunliffe N, Cuevas LE (2009) Incidence of rotavirus and all-cause diarrhea in northeast Brazil following the introduction of a national vaccination program. Gastroenterology 137:1970–1975

17. Gurgel RQ, Cuevas LE, Vieira SC, Barros VC, Fontes PB, Salustino EF, Nakagomi O, Nakagomi T, Dove W, Cunliffe N, Hart CA (2007) Predominance of rotavirus P[4]G2 in a vaccinated population, Brazil. Emerg Infect Dis 13:1571–1573

18. Hoshino Y, Kapikian AZ (2000) Rotavirus serotypes: classification and importance in epidemiology, immunity, and vaccine development. J Health Popul Nutr 18:5–14

19. Iturriza-Gómara M, Cubitt D, Desselberger U, Gray J (2001) Amino acid substitution within the VP7 protein of G2 rotavirus strains associated with failure to serotype. J Clin Microbiol 39:3796–3798

20. Iturriza-Gómara M, Dallman T, Bányai K, Böttiger B, Buesa J, Diedrich S, Fiore L, Johansen K, Korsun N, Kroneman A, Lappalainen M, László B, Maunula L, Matthinjnssens J, Midgley S, Mladenova Z, Poljsak-Prijatelj M, Pothier P, Ruggeri FM, Sanchez-Fauquier A, Schreier E, Steyer A, Sidaraviciute I, Tran AN, Usonis V, Van Ranst M, de Rougemont A, Gray J (2009) Rotavirus surveillance in Europe, 2005–2008: web-enabled reporting and real-time analysis of genotyping and epidemiological data. J Infect Dis 200:215–221

21. Iturriza-Gómara M, Dallman T, Bányai K, Böttiger B, Buesa J, Diedrich S, Fiore L, Johansen K, Koopmans M, Korsun N, Koukou D, Kroneman A, László B, Lappalainen M, Maunula L, Marques AM, Matthijnssens J, Midgley S, Mladenova Z, Nawaz S, Poljsak-Prijatelj M, Pothier P, Ruggeri FM, Sanchez-Fauquier A, Steyer A, Sidaraviciute-Ivaskeviciene I, Syriopoulou V, Tran AN, Usonis V, Van Ranst M, De Rougemont A, Gray J (2011) Rotavirus genotypes co-circulating in Europe between 2006 and 2009 as determined by EuroRotaNet, a pan-European collaborative strain surveillance network. Epidemiol Infect 139:895–909

22. Jere KC, Sawyerr T, Seheri LM, Peenze I, Page NA, Geyer A, Steele AD (2011) A first report on the characterization of rotavirus strains in Sierra Leone. J Med Virol 83:540–550

23. Khamrin P, Peerakome S, Tonusin S, Malasao R, Okitsu S, Mizuguchi M, Ushijima H, Maneekarn N (2007) Changing pattern of rotavirus G genotype distribution in Chiang Mai, Thailand from 2002 to 2004: decline of G9 and reemergence of G1 and G2. J Med Virol 79:1775–1782

24. Khoury H, Ogilvie I, El Khoury AC, Duan Y, Goetghebeur MM (2011) Burden of rotavirus gastroenteritis in the Middle Eastern and North African pediatric population. BMC Infect Dis 11:9

25. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

26. Kirkwood CD, Boniface K, Bogdanovic-Sakran N, Masendycz P, Barnes GL, Bishop RF (2009) Rotavirus strain surveillance-an Australian perspective of strains causing disease in hospitalised children from 1997 to 2007. Vaccine 27:102–107

27. Kosakovsky Pond SL, Poon AFY, Frost SDW (2009) Estimating selection pressures on alignments of coding sequences. In: Lemey P, Salemi M, Vandamme A-M (eds) The phylogenetic handbook, 2nd edn. Cambridge University Press, Cambridge, pp 419–490

28. Lazdins I, Coueson BS, Kikwood C, Dyall-Smith M, Masendycz P, Sonza S, Holmes IH (1995) Rotavirus antigenicity is affected by the genetic context and glycosylation of VP7. Virology 209:80–89

29. Martínez M, Amarilla AA, Galeano ME, Aquino VH, Fariña N, Russomando G, Parra GI (2010) Predominance of rotavirus G2P[4] and emergence of G12P[9] strains in Asunción, Paraguay, 2006–2007. Arch Virol 155:525–533

30. Mascarenhas JD, Lima CS, de Oliveira DS, Guerra Sde F, Maestri RP, Gabbay YB, de Lima IC, de Menezes EM, Linhares Ada C, Bensabath G (2010) Identification of two sublineages of genotype G2 rotavirus among diarrheic children in Parauapebas, Southern Pará State, Brazil. J Med Virol 82:712–719

31. Nakagomi T, Cuevas LE, Gurgel RG, Elrokhsi SH, Belkhir YA, Abugalia M, Dove W, Montenegro FM, Correia JB, Nakagomi O, Cunliffe NA, Hart CA (2008) Apparent extinction of non-G2 rotavirus strains from circulation in Recife, Brazil, after the introduction of rotavirus vaccine. Arch Virol 153:591–593

32. Page NA, Steele AD (2004) Antigenic and genetic characterization of serotype G2 human rotavirus strains from the African continent. J Clin Microbiol 42:595–600

33. Page NA, Steele AD (2004) Antigenic and genetic characterization of serotype G2 human rotavirus strains from South Africa from 1984 to 1998. J Med Virol 72:320–327

34. Parashar UD, Burton A, Lanata C, Boschi-Pinto C, Shibuya K, Steele D, Birmingham M, Glass RI (2009) Global mortality associated with rotavirus disease among children in 2004. J Infect Dis 200:S9–S15

35. Parashar UD, Givson CJ, Bresee JS, Glass RI (2006) Rotavirus and severe childhood diarrhea. Emerg Infect Dis 12:304–306

36. Patel MM, Oliveira LH, Bispo AM, Gentsch J, Parashar UD (2008) Rotavirus P[4]G2 in a vaccinated population, Brazil. Emerg Infect Dis 14:863–865

37. Paul SK, Kobayashi N, Nagashima S, Ishino M, Watanabe S, Alam MM, Ahmed MU, Hossain MA, Naik TN (2008) Phylogenetic analysis of rotaviruses with genotypes G1, G2, G9 and G12 in Bangladesh: evidence for a close relationship between rotaviruses from children and adults. Arch Virol 153:1999–2012

38. Pun SB, Nakagomi T, Sherchand JB, Pandey BD, Cuevas LE, Cunliffe NA, Hart CA, Nakagomi O (2007) Detection of G12 human rotaviruses in Nepal. Emerg Infect Dis. 13:482–484

39. Pybus OG, Shapiro B (2009) Natural selection and adaptation of molecular sequences. In: Lemey P, Salemi M, Vandamme A-M (eds) The phylogenetic handbook, 2nd edn. Cambridge University Press, Cambridge, pp 407–418

40. Rahman M, Sultana R, Ahmed G, Nahar S, Hassan ZM, Saiada F (2007) Prevalence of G2P[4] and G12P[6] rotavirus, Bangladesh. Emerg Infect Dis 13:18–24

41. Richardson V, Hernandez-Pichardo J, Quintanar-Solares M, Esparza-Aguilar M, Johnson B, Gomez-Altamirano CM, Parashar U, Patel M (2010) Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. N Engl J Med 362:299–305

42. Ruiz-Palacios GM, Pérez-Schael I, Velázquez FR, Abate H, Breuer T, Clemens SC, Cheuvart B, Espinoza F, Gillard P, Innis BL, Cervantes Y, Linhares AC, López P, Macías-Parra M, Ortega-Barría E, Richardson V, Rivera-Medina DM, Rivera L, Salinas B, Pavía-Ruz N, Salmerón J, Rüttimann R, Tinoco JC, Rubio P, Nuñez E, Guerrero ML, Yarzábal JP, Damaso S,

Tornieporth N, Sáez-Llorens X, Vergara RF, Vesikari T, Bouckenooghe A, Clemens R, De Vos B, O'Ryan M, Human Rotavirus Vaccine Study Group (2006) Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. N Engl J Med 354:11–22

43. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

44. 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

45. Snelling TL, Andrews RM, Kirkwood CD, Culvenor S, Carapetis JR (2011) Case-control evaluation of the effectiveness of the G1P[8] human rotavirus vaccine during an outbreak of rotavirus G2P[4] infection in central Australia. Clin Infect Dis 52:191–199

46. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

47. Trinh QD, Pham NT, Nguyen TA, Phan TG, Yan H, Hoang le P, Khamrin P, Maneekarn N, Li Y, Okitsu S, Mizuguchi M, Ushijima H (2010) Sequence analysis of the VP7 gene of human rotaviruses G2 and G4 isolated in Japan, China, Thailand, and Vietnam during 2001–2003. J Med Virol 82:878–885

48. Uchida R, Pandey BD, Sherchand JB, Ahmed K, Yokoo M, Nakagomi T, Cuevas LE, Cunliffe NA, Hart CA, Nakagomi O (2006) Molecular epidemiology of rotavirus diarrhea among children and adults in Nepal: detection of G12 strains with P[6] or P[8] and a G11P[25] strain. J Clin Microbiol 44:3499–3505

49. Vesikari T, Matson DO, Dennehy P, Van Damme P, Santosham M, Rodriguez Z, Dallas MJ, Heyse JF, Goveia MG, Black SB, Shinefield HR, Christie CD, Ylitalo S, Itzler RF, Coia ML, Onorato MT, Adeyi BA, Marshall GS, Gothefors L, Campens D, Karvonen A, Watt JP, O'Brien KL, DiNubile MJ, Clark HF, Boslego JW, Offit PA, Heaton PM, Rotavirus Efficacy Safety Trial (REST) Study Team (2006) Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. N Engl J Med 354:23–33

50. Zao CL, Yu WN, Kao CL, Taniguchi K, Lee CY, Lee CN (1999) Sequence analysis of VP1 and VP7 genes suggests occurrence of a reassortant of G2 rotavirus responsible for an epidemic of gastroenteritis. J Gen Virol 80:1407–1415