Analysis of genetic diversity and molecular evolution of human group B rotaviruses based on whole genome segments

Dai Yamamoto,1 Souvik Ghosh,1 Balasubramanian Ganesh,2 Triveni Krishnan,2 Mamta Chawla-Sarkar,2 Mohammed Mahbub Alam,3 Tin Sabai Aung4 and Nobumichi Kobayashi1

1Department of Hygiene, Sapporo Medical University School of Medicine, Sapporo, Japan
2National Institute of Cholera and Enteric Diseases, Kolkata, India
3Bangladesh Agricultural University, Mymensingh, Bangladesh
4National Health Laboratory, Yangon, Myanmar

Correspondence
Dai Yamamoto
yamadai@sapmed.ac.jp

INTRODUCTION

Rotavirus, a member of the family Reoviridae, is the most important viral pathogen causing gastroenteritis in humans. Rotavirus has 11 segments of double-stranded RNA (dsRNA) as a genome, and the virus particle is composed of three concentric layers, i.e. the outer capsid, inner capsid and core (Estes & Kapikian, 2007). The outer capsid consists of two structural proteins, VP4 and VP7, which are neutralization antigens. The inner capsid consists of structural protein VP6. Rotavirus is classified into five groups (A–E) and two putative groups (F and G), based on the antigenicity of the inner capsid protein VP6 and genomic characteristics (Kapikian et al., 2001). In humans, groups A, B and C have so far been detected. Group A rotavirus (GAR) is the most prevalent throughout the world and is recognized as the leading viral pathogen of acute gastroenteritis in children.

Group B rotavirus (GBR) is genetically and antigenically distinct from GAR and has been detected in humans, mice, calves, pigs and sheep. In humans, GBR has been noted because it causes severe, cholera-like diarrhoea, mostly in adults (Mackow, 1995). GBR was first identified as adult diarrhoea rotavirus (ADRV) in nationwide outbreaks in China in 1982–1983 (Hung et al., 1983, 1984; Wang et al., 1985), and the detection of this virus has been limited to China (Dai et al., 1987; Fang et al., 1989). They were subsequently detected in sporadic cases in India in 1997 and in Bangladesh in 2000, demonstrating the distribution of GBRs in Asian countries outside China (Krishnan et al., 1999; Sanekata et al., 2003). Thereafter, GBRs have again been detected in these countries in sporadic cases of diarrhoea (Barman et al., 2006; Rahman et al., 2007). Furthermore, a human GBR was detected in Myanmar in 2007 (Aung et al., 2009). In contrast, recently in China, there has been only a 2002 report of detection of GBR in sporadic cases of diarrhoea in Wuhan (Yang et al., 2004).
Despite the extensive epidemiological surveillance of rotaviruses worldwide, the detection of human GBRs has been extremely rare and limited to only the four countries mentioned above.

With the limited genetic information available, human GBRs were known to be quite distinct from bovine, porcine, ovine and murine GBRs. Recently, Kuga et al. (2009) proposed a classification scheme for GBR genotypes in terms of VP7 gene sequence, and GBRs were divided into five genotypes (G1–G5), among which human strains were assigned to a single genotype, G2. The human GBRs detected were classified genetically into two lineages within genotype G2 based on the VP7 gene: the Chinese lineage and the Indian–Bangladeshi lineage (Ahmed et al., 2004; Yang et al., 2004; Aung et al., 2009). The GBR strains of these two lineages were genetically closely related, suggesting that these lineages diverged from a common ancestral origin several decades ago (Yang et al., 2004; Rahman et al., 2007). However, genetic diversity in RNA segments other than the VP7 gene has been scarcely analysed, and thus the accurate status of molecular evolution of the whole genome of GBR is still unknown. Similar to GAR, six RNA segments of GBR encode structural proteins (VP1–VP4, VP6 and VP7) and five segments encode non-structural proteins (NSP1–NSP5). However, it is unique to human GBR that the NSP1 gene contains two open reading frames encoding two putative protein products (Mackow, 1995; Kobayashi et al., 2001).

Full-genome sequence of rotaviruses has been determined for many GAR strains to date. Accordingly, based on the findings of diversity of individual RNA segments, a full genome-based genotyping system composed of genotypes of 11 RNA segments has been proposed (Matthijnssens et al., 2008). However, genetic data for GBR are extremely limited. Full-genome sequence of GBR has been determined for only three human strains (CAL-1 in India, Bang373 in Bangladesh and WH-1 in China) and a murine strain, IDIR. Genetic information available for human and animal GBRs is mostly for VP7 gene sequences.

In the present study, nearly full-length sequences of all gene segments were determined for human GBRs detected recently in India, Bangladesh and Myanmar. The obtained sequence data were analysed and compared with those reported previously, to understand differences in genetic diversity among the 11 RNA segments and divergent regions in individual RNA segments. The results have provided fundamental information about the genomic evolution of GBR, including the relatedness of genetic diversity to the function of each viral protein.

RESULTS

Phylogenetic analysis of the VP7 gene

VP7 gene sequences of GBRs determined in the present study were analysed phylogenetically with those of human, bovine, porcine and murine GBR strains. Fig. 1(a) represents a phylogenetic tree of the VP7 genes. Sequence identities of VP7 genes among GBRs are shown in Supplementary Table S1 (available in JGV Online). The Indian strains IC-008 and IDH-084 and Bangladeshi strain Bang117 were located in the human Indian–Bangladeshi lineage of the GBR genotype G2. Within this lineage, IC-008 and IDH-084 clustered with MMR-B1 in Myanmar, whereas Bang117 clustered with the Bangladeshi strains (Bang373 and Bang544) reported previously. Strains IC-008 and IDH-084, with 99.5 % nucleotide sequence identity to each other, showed extremely high sequence identities to GBRs in the Indian–Bangladeshi lineage (97.8–99.4 %), with the highest identity to MMR-B1 (99.4 %) (Supplementary Table S1). Strain Bang117 exhibited extremely high sequence identities to Bangladeshi strains Bang373 (99.9 %) and Bang544 (99.6 %), whereas slightly lower identities (97.9–98.3 %) were shown to GBRs in India and Myanmar. All GBRs in the Indian–Bangladeshi lineage showed 91.3–92.6 % identities to Chinese strains ADRV and WH-1, and considerably lower identities to bovine GBRs (62.2–66.6 %), porcine GBRs (62.9–63.9 %) and a murine GBR (58.0–59.8 %).

Genetic analysis of other viral protein genes

Phylogenetic analysis of GBR genes encoding VP1, VP2, VP4, VP6, NSP1, NSP2 and NSP5 revealed the presence of murine, human and bovine GBR clusters corresponding to GBR genotypes G1, G2 and G3, respectively, which had been classified for the VP7 gene (Kuga et al., 2009) (Fig. 1b, c, e–h, k). For the VP3, NSP3 and NSP4 genes, two genotypes (G1 and G2) were discriminated (Fig. 1d, i, j). Furthermore, all of the gene segments of human GBRs were discriminated into two lineages, i.e. the Indian–Bangladeshi and Chinese lineages, within GBR genotype G2. The range of sequence identities of individual gene segments within a lineage and between the two lineages is summarized in Table 1 (sequence identities of each gene segment among GBRs are shown in Supplementary Tables S2–S11, available in JGV Online). Within the same lineage, gene sequences showed extremely high identities, i.e. 95.6–100.0 % (94.7–100.0 % at the amino acid level) in the Indian–Bangladeshi lineage and 98.0–98.9 % (98.0–100.0 % at the amino acid level) in the Chinese lineage, whereas lower identities (89.5–94.9 % at the nucleotide level) were found between the two lineages. Throughout the RNA segments, strain MMR-B1 showed the highest sequence identities (>99 %) to IC-008 and IDH-084. In contrast, within GBRs of the Indian–Bangladeshi lineage, the lowest levels of sequence identity were observed mostly between the recent Indian strains (IC-008 or IDH-084) and CAL-1.

Among the 11 RNA segments of GBR strains from India, Bangladesh and Myanmar, the VP6 and NSP2 genes showed the highest identities (>98 %), whereas the lowest identities were observed in the NSP4 (96.1 %) and NSP5 (95.6 %) genes. The sequence identities of the VP8*
encoding region of the VP4 gene were lower than those of the VP5* region. Similarly, between the two lineages, VP6 and NSP2 gene sequences were most conserved (93.1–94.9 % identity); in contrast, the NSP3, NSP5 and VP8* genes exhibited the highest diversity (88.4–92.4 % identity). Between strains IC-008 and IDH-084, derived from a child and an adult, respectively, sequence identity was 98.9–99.7 % throughout the gene segments.

Divergent regions in viral proteins

To investigate whether genetic diversity among GBR genes occurred randomly, the presence of divergent or conserved region(s) within a viral protein was analysed by sequence alignment of the deduced amino acid sequences. Amino acid sequences of GBR VP1 and NSP3, and partial VP3 and NSP2 sequences were aligned with those of GAR and group C rotavirus (GCR) (Figs 2–4). Except for VP1 and NSP3, primary amino acid sequence alignments of human GBRs are shown in Supplementary Figs S1–S9 (available in JGV Online). Due to the extremely conserved nature of VP7 (Supplementary Fig. S5) and VP6 (Supplementary Fig. S4), the divergent region of these proteins was not specified.

The RNA polymerase domain of VP1 (finger I, II, palm I, II, and thumb) located in the central portion (469 aa) (McDonald et al., 2009), including functionally critical motifs shared by different rotavirus groups as well as other RNA viruses (Cohen et al., 1989; Nagashima et al., 2008), was highly conserved among GBRs (Fig. 2). In the catalytic region (182 aa from position 546 to 727), 54 aa (29.7 %) were conserved among GAR, GBR and GCR, and the consensus motif of RNA polymerase (SG, T, N, T, GDD) was commonly found in these three rotavirus groups. Most of the amino acid differences in GBR VP1 were found in the N-terminal region (376 aa) and C-terminal region (317 aa), where GAR and GCR had quite distinct sequences.

Table 1. Sequence identities (%) of individual gene segments among human GBRs

<table>
<thead>
<tr>
<th>Gene segment</th>
<th>Indian, Bangladeshi, Myanmarese strains† (A)</th>
<th>Chinese strains‡ (B)</th>
<th>Between A and B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nucleotide</td>
<td>Amino acid</td>
<td>Nucleotide</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP1 gene</td>
<td>97.4–99.8</td>
<td>98.7–99.9</td>
<td>98.0</td>
</tr>
<tr>
<td>VP2 gene</td>
<td>97.2–99.4</td>
<td>98.9–99.9</td>
<td>98.4</td>
</tr>
<tr>
<td>VP3 gene</td>
<td>97.5–99.6</td>
<td>98.6–100.0</td>
<td>98.6</td>
</tr>
<tr>
<td>VP4 gene</td>
<td>97.3–99.4</td>
<td>97.5–99.6</td>
<td>98.4</td>
</tr>
<tr>
<td>VP5* region</td>
<td>97.6–99.7</td>
<td>98.7–99.8</td>
<td>98.6</td>
</tr>
<tr>
<td>VP8* region</td>
<td>95.9–99.5</td>
<td>93.2–99.5</td>
<td>98.1</td>
</tr>
<tr>
<td>VP6 gene</td>
<td>98.0–99.9</td>
<td>98.7–100.0</td>
<td>98.6</td>
</tr>
<tr>
<td>VP7 gene</td>
<td>97.7–99.9</td>
<td>98.0–100.0</td>
<td>98.6</td>
</tr>
<tr>
<td>NSP1 gene</td>
<td>97.6–99.6</td>
<td>98.8</td>
<td></td>
</tr>
<tr>
<td>Peptide 1 region</td>
<td>97.2–100.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Peptide 2 region</td>
<td>97.2–99.4</td>
<td>98.8</td>
<td></td>
</tr>
<tr>
<td>NSP2 gene</td>
<td>98.0–99.4</td>
<td>98.0–99.7</td>
<td>98.4</td>
</tr>
<tr>
<td>NSP3 gene</td>
<td>97.0–99.7</td>
<td>96.8–99.7</td>
<td>98.1</td>
</tr>
<tr>
<td>NSP4 gene</td>
<td>96.1–99.9</td>
<td>97.3–100.0</td>
<td>98.4</td>
</tr>
<tr>
<td>NSP5 gene</td>
<td>95.6–100.0</td>
<td>94.7–100.0</td>
<td>98.9</td>
</tr>
</tbody>
</table>

†CAL-1, Bang373, IC-008, IDH-084, Bang117, MMR-B1.
‡ADRV, WH-1 (VP1 and VP3 sequences are available only for WH-1).

http://vir.sgmjournals.org
Sequence divergence of VP2 among GBRs was found in the 80 aa sequence from the N terminus (Supplementary Fig. S1). Amino acid differences of VP3 sequences among the GBR strains were found throughout the sequence. However, a conserved motif (ALYSLSNXXN) (Ito et al., 2001) was found in all GBRs, as well as in GAR and GCR (Fig. 3a). Some sequences similar to the possible active sites of guanylyltransferase in VP3 (Cook & McCrae, 2004) were
conserved among human GBRs (Supplementary Fig. S2). In the VP4 sequence, some hydrophobic regions located mostly in the VP5* portion were highly conserved, whilst the sequence in the N-terminal hydrophilic region in the VP8* portion had more amino acid diversity than any other region in VP4 (Supplementary Fig. S3).

In peptide 2 of the NSP1 gene, whilst the divergent amino acids were located throughout the sequence, a cysteine- and histidine-rich region was highly conserved (Supplementary Fig. S6). A cysteine-rich region was found in peptide 1. Although NSP2 sequences of human GBRs showed much less diversity, they were noticeably distinct from those of GAR and GCR (Supplementary Fig. S7; Fig. 3b). However, some amino acids that had been known to be required for nucleoside triphosphatase activity in GAR, including the conserved histidine (H225 for GAR) (Kumar et al., 2007), were also conserved among human GBR, as well as GCR (Fig. 3b).

Among the NSP3 sequences of GBRs, the N-terminal 120 aa region was highly conserved and corresponded to the RNA-binding region revealed for GAR NSP3 (Deo et al., 2002) (Fig. 4). Remarkably, amino acids in a motif (RNXXW) in the z-helix 4 (H4), which are essential for RNA binding by GAR NSP3 (Vende et al., 2000), were conserved in NSP3 of GBR, as well as GCR. In contrast, the remaining portion (two-thirds) was divergent and included a region associated with eIF4G binding.

Two hydrophobic regions and two putative enterotoxin regions in NSP4 (Ishino et al., 2006) were also highly conserved among the GBR strains (Supplementary Fig. S8). In the NSP5 sequence, the C-terminal 60 aa sequence was more conserved than the remaining N-terminal portion, comprising 110 aa (Supplementary Fig. S9).

**DISCUSSION**

To date, human GBRs have been detected only in China, India, Bangladesh and Myanmar, and have been classified genetically into two lineages, i.e. the Chinese and Indian–Bangladeshi lineages. However, antibodies to human GBRs have been detected at low prevalence in humans from several other countries, such as Kenya, Thailand, Canada, the USA and the UK, suggesting wide distribution of this virus (Brown et al., 1987; Nakata et al., 1987). Also, GBRs have been detected in rats, pigs, cows and goats, and have been suggested to be highly prevalent in these animals through seroepidemiological evidence (Brown et al., 1987; Tsunemitsu et al., 2005). Animal GBRs are genetically distinct from human GBRs and contain divergent populations of viruses. According to the classification of GBR in terms of the VP7 gene by Kuga et al. (2009), GBRs were divided into five genotypes (G1–G5), comprising a single genotype (G2) of human strains and the other four genotypes of murine, porcine and bovine strains. Although genetic and molecular epidemiological study of GBR has been based primarily on the VP7 gene, genetic information of other viral gene segments is extremely limited. So far, full-genome information has been available for only a murine strain and three human strains, and eight gene segments have been sequenced for a bovine strain (Ghosh et al., 2007, 2010). For ovine and porcine GBRs, sequence data of only a few gene segments are available (Shen et al., 1999; Kuga et al., 2009). In the present study, nearly full-length sequences of four human GBRs were determined, which enabled us to analyse substantially the genetic diversity of GBR at the full-genome level.

Sequence analysis of whole gene segments of the human GBR provided two epidemiologically significant findings. First, the Myanmarese strain MMR-B1 is more closely related genetically to the recent Indian strains than the old Indian strain (CAL-1) and Bangladeshi strains for all gene segments, suggesting that the GBR in Myanmar and current Indian GBRs may have been derived from the same origin. Second, the two latest GBR strains from a child (IC-008) and an adult (IDH084) detected in Kolkata, India, were genetically almost identical, which
indicated distribution of the same GBR in a child and an adult. Similarly, distribution of genetically identical GAR in both children and adults was reported in China and Bangladesh (Wang et al., 2007; Paul et al., 2008). Together with the report that GBR was detected by RT-PCR in 18.5% of diarrheal specimens from children in Kolkata (Barman et al., 2006), GBR is believed to infect both children and adults and may be maintained among them.

Although the human GBRs analysed in the present study were genetically highly similar, sequence identities were different depending on gene segments. The VP6 and NSP2 genes showed the highest identities, whilst NSP3–NSP5 genes were more divergent. Similarly, within a single genotype of GAR, the VP6 gene showed the highest identity among structural proteins (Matthijnssens et al., 2008). However, sequence variation in non-structural proteins seems to be different from GAR; that is, in GAR, NSP5 is...
the most conserved, whilst NSP1 is the most divergent. Such a difference in GBR may be due in part to insufficient numbers of strains analysed in the present study.

In some RNA segments, genetic divergence or conservation was detected in specific regions that are correlated with the function of the protein. The RNA polymerase domain in VP1 and N-terminal RNA-binding domain in NSP3 are evidently conserved compared with other regions in individual proteins, as also found in cognate proteins of GAR (Rao et al., 1995; Heiman et al., 2008; McDonald et al., 2009). In these domains, some amino acids/motifs are commonly found among GAR, GBR and GCR, despite high sequence diversity among them. Furthermore, some functionally important motifs and structural features of other rotavirus proteins known for GAR were also conserved in GBR, including the conserved motif in VP3 of which the function is unknown (Ito et al., 2001; Nagashima et al., 2008), and that in NSP2 related to nucleoside triphosphatase activity (Kumar et al., 2007). The cysteine- and histidine-rich region noted for GAR NSP1 (Hua et al., 1993) was found also in peptides 1 and 2 of GBR NSP1. The C-terminal region of NSP5, which is conserved among GAR and critical for its function for viroplasm-like structure formation in cells (Sen et al., 2007), was also conserved in GBR NSP5.

In contrast, the VP8* portion in VP4 and the N-terminal region in VP2 were more divergent than other regions in these proteins in both GAR and GBR. VP8* is located as the outermost portion of the VP4 spike protein on the rotavirus virion and is associated with the antigenic specificity and genotype of VP4 (Dormitzer et al., 2004; Kapikian et al., 2001). The N termini of GAR VP2 are predicted to lie inside the core shell and to bind the viral enzyme–RNA complex (VP1–VP3–RNA) (McDonald & Patton, 2008). It can be suggested that the interaction with protein and/or RNA causes the sequence variation in the N-terminal region of VP2. The above findings suggest that viral proteins of GAR and GBR have similar structural and functional features and are subjected to similar molecular evolution, despite a considerable genetic divergence between the two rotavirus groups. Although the degree of genetic evolution may be dependent on each RNA segment, human GBR genes are suggested to have evolved in association with functional roles of individual viral proteins.

In the present study, GBR was characterized for its genomic diversity and evolution at the full-genome level. Further accumulation of genetic data with more GBRs may be necessary to understand the ecological features and epidemiological dynamics of GBRs.

METHODS

Rotavirus strains. Four human GBR strains, IC-008, IDH-084, Bang117 and MMR-B1, were analysed. GBR strains IC-008 and IDH-084 were detected as sole pathogens of diarrhoea in stool specimens from a child (12-month-old female) and an adult (35-year-old male), respectively, who visited the Infectious Disease Hospital in Kolkata, India, in January 2008 and November 2007, respectively. Strain Bang117 was found in a 32-year-old male patient with severe diarrhoea admitted to SK hospital in Mymensingh, Bangladesh, in March 2002. Strain MMR-B1 was detected in the diarrhoea of an adult patient in Yangon, Myanmar, and its VP7, VP4, VP6 and NSP4 genes were sequenced and reported previously (Aung et al., 2009). Therefore, in the present study, sequences of the remaining seven viral genes were determined for this strain. The presence of GBRs in stool specimens was determined by detection of the typical migration pattern (4-2-1-1-1-1-1) of 11 dsRNA segments in PAGE, and confirmed by RT-PCR as described previously (Gouveia et al., 1991). Stool specimens collected from patients were stored at −80 °C until analysed.

Sequencing and phylogenetic analysis. Nucleotide sequences of GBR genes were determined directly from the amplified cDNA products by RT-PCR. As a template for RT-PCR, dsRNA was extracted from stool suspensions with a commercially available kit (RNAid kit; Bio 101) according to the manufacturer’s instructions. RT-PCR was performed with avian myeloblastosis virus reverse transcriptase (Seikagaku) and thermostable DNA polymerase (Expand High Fidelity PCR system; Roche) with the primers described previously for rotavirus genes encoding VP2, VP4, VP6, VP7 and NSP1–NSP5 (Ahmed et al., 2004). The cDNAs for VP1 and VP3 genes were amplified by primers prepared in the present study based on the sequences of strain Bang573 (Supplementary Table S12). For all gene segments, full-length sequences except for primer-binding regions at the 5’ and 3’ ends were amplified and sequenced.

PCR products were purified by using the Wizard SV Gel and PCR Clean-Up system (Promega). Sequencing reactions were performed with fluorescent deoxyxide chain-termination chemistry using a BigDye Terminator v. 3.1 cycle sequencing kit (Applied Biosystems). Sequence was determined by an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Genetyx-Win v. 5.1 (Software Development) was used to perform pairwise alignment and to calculate the identity of gene segments among GBRs. Multiple alignment of GBR sequences was performed by the neighbour-joining method using the CLUSTAL W program. Phylogenetic analysis was performed with MEGA software v. 4.1 based on the neighbour-joining method and the Kimura two-parameter model. Phylogenetic trees were supported statistically by bootstrapping with 1000 replicates.

GenBank accession numbers of sequences. The nucleotide sequences of GBR strains determined in this study were deposited in GenBank/EMBL/DDBJ under the following accession numbers: GU377213–GU377223 (IC-008), GU377224–GU377234 (IDH-084), GU391301–GU391311 (Bang117) and GU370054–GU370060 (MMR-B1).

ACKNOWLEDGEMENTS

This study was supported in part by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases (Okayama University – National Institute of Cholera and Enteric Diseases, India) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and a Grant-in-Aid of the Japan Medical Association.

REFERENCES


