Full genome analysis of group B rotaviruses from western India: genetic relatedness and evolution

Anismrita Lahon,1 Atul M. Walimbe2 and Shobha D. Chitambar1

1Enteric Viruses Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, PO Box 11, Pune 411001, India
2Bionformatics Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, PO Box 11, Pune 411001, India

To date, full-genome sequences of only seven human group B rotavirus (RVBs) strains have been described. Such data on more RVBs are necessary to establish the evolutionary relationship and ecological features of RVBs from different geographical regions. The present study was aimed at determining the full-length sequences of all 11 genes of 13 human RVB strains detected during 1995–2010 in sporadic and outbreak cases of acute gastroenteritis from four different cities of western India. This study also included estimation of evolutionary rates and site-specific selection pressure analysis for all gene segments. Nucleotide/deduced amino acid sequence analyses of structural and non-structural genes showed 95.1–99.8/94.1–100 % identity with the counterparts of RVB strains isolated in India, Bangladesh and Myanmar. Phylogenetic analyses of all gene segments revealed formation of a monophyletic clade of the western Indian RVB strains, reflecting their highly conserved nature. All gene segments were also found to be under negative/purifying selection pressure. These data suggest that RVB is circulating in the natural host as a series of stable viral clones. Estimates of rates of nucleotide substitution in all RVBs ranged from 1.36–4.78 \times 10^{-3} substitutions per site per year. The rate for human RVB VP7 and NSP2 genes were comparable, respectively, with the evolution kinetics of genotype G9/G12 and N1 group A rotavirus strains. The time of the most recent common ancestor of the extant human RVBs was estimated to be during 1915–1974. Evolutionary and genetic analyses carried out in this study provide data that is useful for the elucidation of evolutionary relationship/timescale, stasis or dynamics existing in the RVB population.

INTRODUCTION

Rotavirus, a genus of the family Reoviridae, is a medically very significant cause of acute gastroenteritis in young children and the young of various mammalian and avian species. The viral genome comprises 11 segments of dsRNA encoding six structural (VP1–VP4, VP6 and VP7) and five to six non-structural (NSP1–NSP5/6) proteins (Estes & Kapikian, 2007). The genome is encased within a triple-layer of viral proteins, comprising a core (VP1, VP2 and VP3), an inner capsid (middle layer, VP6) and an outer capsid (outer layer, VP4 and VP7). The core scaffolding protein VP2 contains a transcription complex formed by the RNA-dependent RNA polymerase VP1 and VP3 (a guanylyltransferase and methylase) and genomic RNAs. The inner capsid protein, VP6, carries antigens that classify rotaviruses into at least seven distinct groups A–G (RV A–RV G) and a putative novel group, ADRV-N, proposed as RVH (Kapikian et al., 2001; Jiang et al., 2008; Nagashima et al., 2008; Matthijnssens et al., 2012). The two outer capsid proteins VP4 and VP7 contain neutralizing epitopes. The non-structural proteins are produced in infected cells and are primarily involved in virus replication, morphogenesis and pathogenesis.

Groups A, B, C and H have been detected in human infections (Kapikian et al., 2001; Jiang et al., 2008; Nagashima et al., 2008). Association of human RVBs as enteric pathogens was established in devastating epidemics of diarrhoea in China during 1982–1987 (Hung et al., 1984; Fang et al., 1989). Outside China, RVBs have been identified in sporadic cases of diarrhoea in India, Bangladesh and Myanmar (Krishnan et al., 1999; Sanekata et al., 2003; Kelkar & Zade, 2004; Aung et al., 2009; Saiada et al., 2011; Lahon et al., 2012). In an outbreak of gastroenteritis that occurred in Daman, Union territory of India in 2000, RVB infections predominated (Kelkar et al., 2007). Thereafter, two reports on the association of RVB in outbreak cases were documented from three different localities of western India (Chitambar et al., 2011, 2012).
Full-genome sequences of seven human RVB strains (CAL-1, IDH-084 and IC-008 from eastern India, Bang373 and Bang117 from Bangladesh, MMR-B1 from Myanmar and WH-1 from China) and a murine strain, IDIR have been described to date (Yamamoto et al., 2010). Sequence analysis of the human RVB strains has revealed that all genes were conserved and considered to belong to a single genotype, G2 (Yamamoto et al., 2010). Within the genotype G2, the strains were further categorized into two lineages: Chinese (Chinese strains) and Indian–Bangladeshi (Indian, Bangladeshi and Myanmarese strains) lineages. To understand the ecological and epidemiological features of RVB infections, genetic data on more RVBs from different geographical regions are necessary.

The present study was conducted to characterize the full-length sequences of all of the 11 gene segments of a total of 13 human RVB strains detected in four different regions of western India during 1995–2010. The data obtained were analysed and compared with those reported previously, to understand the genetic diversity of the individual RNA segments and to assess their conserved and variable features in comparison with other RVB strains. This has yielded fundamental information about the genetic makeup of human RVB strains isolated from western India. This study is further focussed on the estimation of evolutionary rates and the time of the most recent common ancestor (tMRCA) for all RVB gene segments. Site-specific selection pressure acting on all gene segments was also analysed for the prevalent Indian–Bangladeshi lineage RVB strains.

**RESULTS**

Genetic analyses of RVB VP1–VP4, VP6, VP7 and NSP1–NSP5 gene segments

A maximum clade credibility (MCC) tree of VP7 genes of the western Indian human RVB strains representing the period from 1995 to 2010 displayed clustering of the sequences with the corresponding sequences of other strains placed in the Indian–Bangladeshi lineage of genotype G2 (Fig. 1a). Phylogenetic analyses carried out for NSP1 and NSP2 gene sequences of these strains revealed the presence of A2 and N1 genotypes, respectively, while the analyses for the remaining gene segments showed clustering in a single lineage together with the Indian–Bangladeshi RVB strains belonging to VP7 genotype G2 (Fig. 1b–k).

Nucleotide/deduced amino acid sequence analyses of VP1–VP4, VP6, VP7 and NSP1–NSP5 gene segments of the western Indian RVB strains showed 95.1–99.8/94.1–100 % identity with RVB strains of Indian, Bangladeshi and Myanmarese origin (Table 1).

Rotavirus VP1 (1160 aa) contains the RNA-dependent RNA polymerase domain that is located in the central portion (377–843 aa) of the protein (McDonald et al., 2009). The consensus motif (SG, T, N, T and GDD) of this domain which is commonly found in all RNA viruses (Cohen et al., 1989; Nagashima et al., 2008) was also found among western Indian RVB strains (Fig. S1, available in JGV Online). Very recently, aspartic acid residues at positions 520/567, 631/688 and 632/689, and arginine residues at positions 452/494, 457/499 and 460/502 have been described to contribute to the RNA synthesis of RVAs/RVBs (Ogden et al., 2012). These residues were found to be conserved in the RVB strains of this study. However, the N-terminal (376 aa) and/or C-terminal (317 aa) regions of VP1 proteins of these strains showed a total of 10 unique amino acid substitutions (Table S1a). The N terminus amino acid sequence (1–80 aa) of VP2 (934 aa) is known to be highly divergent among RVBs (Yamamoto et al., 2010). In this study, eight of a total of 11 positions carrying unique amino acid substitutions identified in the VP2 protein were located in this region. Although divergent amino acids were noted throughout the VP3 (763 aa) sequence of RVB strains, the region 104–662 in the strains of the present study displayed the presence of 10 unique substitutions (Table S1a). The conserved motif (415ALYSLSXXNX424) of this protein (Ito et al., 2001; Nagashima et al., 2008) found in all rotavirus groups (A–C) was also detected. The putative active site motif (Kx[V/L/I]S) previously identified in the guanylyltransferase of other members of the family Reoviridae (Cook & McCrane, 2004; Mohd Jaafar et al., 2005) was also present in the VP3 RVB amino acid sequence analysed in this study (Fig. S2).

VP4, the product of gene 4, is the viral haemagglutinin and proteolytic cleavage of this protein results in two cleavage products, VP8* and VP5* which enhance rotavirus infectivity (Kapikian et al., 2001). Unlike RVAs, VP8* (1–207 aa) and VP5* (215–750 aa) portions are linked by a connecting peptide (208–214 aa) in RVBs (Kobayashi et al., 2001). The five hydrophobic regions (aa positions 139–160, 375–400, 416–452, 500–530 and 551–573) predicted earlier for this protein (Yamamoto et al., 2010) were found to be conserved with the exception of a position (514) that was replaced by a conservative amino acid (T→S) within one (NIV-0632252) of these strains. Sequence analysis of the VP4 genes of western Indian RVB strains showed 0–6.9 % and 0–1.5 % divergence from other strains of the Indian–Bangladeshi lineage, respectively, in the VP8* and VP5* regions at the amino acid level. Likewise, 5.4–9.9 % and 3.4–5.2 % amino acid divergence from the strains of the Chinese lineage was displayed for the same regions. A total of 17 unique amino acid substitutions (nine and eight positions, respectively, in VP8* and VP5* regions) were found in the VP4 proteins of these strains (Table S1a).

Deduced amino acid sequences of the inner capsid gene, VP6 of western Indian RVB strains showed only 0–1.8 % divergence from other strains of the Indian–Bangladeshi lineage. Unique amino acid substitutions were observed at four positions in this protein (Table S1a). Within the VP7 gene sequences of all of the human RVB strains including the strains of this study, the intra-lineage (Indian–Bangladeshi lineage) divergence in the nucleotide/amino acid sequences ranged from 0 to 3.1 %/0 to 2.9 %, while the inter-lineage (between the Chinese and Indian–Bangladeshi lineage)
The divergence of these sequences ranged from 7.0 to 8.4% to 6.6% Within the strains investigated in this study 0–2.5% divergence was noted in the deduced amino acid sequences. Only a single strain, NIV-0924341 recovered from a sporadic case showed two unique amino acid substitutions (V29M and T162K) in the VP7 protein as compared to other human RVB strains (Table S1a).

The NSP1 gene of human RVB strains is known to contain two overlapping ORFs, peptide 1 (107 aa) and peptide 2 (321 aa). Analysis of this gene in this study revealed 96.9–99.7% nucleotide identity and 95.3–100% (peptide 1) and 96.9–100% (peptide 2) amino acid identity with those from the Indian–Bangladeshi lineage strains. Although unique amino acid substitutions were
located throughout the sequence (Table S1b), cysteine-rich (aa 57–66) and cysteine and histidine-rich (64–76) regions were found conserved in peptide 1 and peptide 2, respectively, of all the RVB strains analysed.

The NSP2 gene sequences from RVB strains of the present study were comparatively closer to their counterparts from the Indian–Bangladeshi and Chinese lineage strains and also within themselves (Table 1). Unique amino acids were only noted in three strains, NIV-005623/NIV-005626 (A52S) and NIV-0924341 (D288E and I295N) (Table S1b). As described earlier (Vende et al., 2000; Deo et al., 2002; Yamamoto et al., 2010), the N-terminal 120 aa region carrying the motif RNXXW located in the α-helix 4 (H4) known to be corresponding to the RNA-binding region for RVA was found to be conserved in the NSP3 gene segment of RVB strains of the present study (Fig. S3). The remaining part of this segment associated with eIF4G binding showed six of a total of seven unique amino acid substitutions detected in NSP3, indicating relative divergence (Table S1b).

Two hydrophobic regions (aa residues 6–20 and 41–55) and two putative enterotoxin regions (aa residues 106–127 and 191–219) in NSP4 (Ishino et al., 2006) were also highly conserved among all of the RVB strains in this present study. In the NSP5 sequence, the C-terminal 60 aa sequence was more conserved than the remaining 50 aa sequence.
N-terminal portion, comprising 110 aa residues as corroborated earlier (Yamamoto et al., 2010). The presence of unique amino acids was noted at different positions in both the NSP4 and NSP5 gene products, respectively, in six and four of the 13 RVB strains (Table S1b).

**Rates of molecular evolution for the RVB gene segments**

The relaxed uncorrelated exponential clock was found to fit the data better than other clock models for all gene segments (Table S2). The coefficient of variation and the credible intervals, clearly excluded zero, which also justified the use of a relaxed molecular clock (data not shown).

The gene wise evolutionary rate for all of the human and animal RVB strains was in the range of $1.36 \times 10^{-3}$–$4.78 \times 10^{-3}$ nucleotide substitutions per site per year (Table 2). The mean rates of nucleotide substitutions per site per year for NSP2 ($1.36 \times 10^{-3}$; higher posterior density (HPD): $0.37–2.55 \times 10^{-3}$) and VP7 ($1.38 \times 10^{-3}$; HPD: $0.36–2.72 \times 10^{-3}$) gene segments were observed to be the lowest, while the same for VP1–VP4, NSP3 and NSP4 segments were found to
be $>3.0 \times 10^{-3}$ with a maximum rate of evolution ($4.78 \times 10^{-3}$; HPD: 1.19–9.09 $10^{-3}$) for the VP3 gene segment. The mean rates of nucleotide substitutions per site per year for the VP6 and NSP5 gene segments were found to be between $2.0 \times 10^{-3}$ and $3.0 \times 10^{-3}$ (Table 2).

The tMRCA of the extant RVBs for the VP7 gene was calculated to be around the year 565 with a 95% HPD limit of 1318 BC to 1756 (node A, Fig. 1a). The same was estimated to be around the year 1300 with a 95% HPD limit ranging from 341 to 1852 for the NSP1 and NSP2.
genes (node A, Fig. 1g, h). For the remaining genes, the tMRCA was found to lie around 1600 (1044–1945) (node A, Fig. 1b–f, i–k).

The evolutionary rates for different genes exclusively from human RVB strains were estimated to be in the range of 1.15–2.30 × 10^{-3} (HPD: 0.52–3.67 × 10^{-3}) nucleotide substitutions per site per year (Table 2). The rates were found to be less than 1.7 × 10^{-3} and above 2.0 × 10^{-3}, respectively, for the two groups of gene segments VP6–VP7, NSP1, NSP2 and VP1–VP4, NSP3–NSP5. The lowest mean rate of nucleotide substitution (1.15 × 10^{-3}) was observed for the NSP2 gene segment and the highest rate (2.30 × 10^{-3}) of the same was estimated for the VP4 gene segment.

The tMRCA for the VP7 gene of human RVB strains was 1915 (HPD: 1798–1978) (node B, Fig. 1a). For the remaining genes, it was between 1942 and 1974 (HPD: 1882–1995) (node B, Fig. 1b–k).

The evolutionary rates estimated for the individual gene segments (VP2, VP4, VP6, VP7 and NSP1–NSP5) of the Chinese (0.91–2.08 × 10^{-3} nucleotide substitutions per year) were

\[ \text{NSP2} \]

\[ 1978 (1562, 1945) \]

\[ 1976 (1564, 1945) \]

\[ 1964 (1528, 1955) \]

\[ 1992 (1986, 1995) \]

\[ 1796 (1562, 1945) \]

\[ 1964 (1564, 1945) \]

\[ 1992 (1986, 1995) \]
site per year) and Indian–Bangladeshi (1.14–2.26 × 10⁻³ nucleotide substitutions per site per year) lineage strains were found to be closer to each other (Table 2). The same could not be calculated for the VP1 and VP3 gene segments of the Chinese lineage strains due to unavailability of sequences for more than one strain. The tMRCA estimated for all genes except VP7 was found to be in existence since the 1960s for the Chinese lineage strains, while the same for the Indian–Bangladeshi lineage strains was found since the 1980s. The tMRCA inferred from VP7 gene revealed the existence of the Chinese lineage strains since the 1940s and that of the Indian–Bangladeshi lineage strains since the 1960s.

As evident from the MCC trees of all gene segments, the murine strain IDIR shared a common ancestor with the human RVB strains with a minimum posterior support of 0.61 in all of the gene segments (Fig. 1b–d, f–k) except the
two outer capsid protein genes, VP4 and VP7. The divergence time of the IDIR strain from the human RVB strains was found to be during 1679–1867 (node E, Fig. 1b–d, f–k). The VP4 and VP7 gene segments of the IDIR strain shared a common ancestor with other animal (bovine and bovine/porcine) strains around the year 1836 and 1882, respectively, (node F, Fig. 1a, e). A common ancestor for all, except the VP3, NSP3 and NSP4 genes (which lacked relevant data), was identified in the RVB strains from one or more of the animal species (bovine, caprine and porcine) during 1557–1994 (node G, Fig. 1a–c, e–h, k).

The evolutionary rate and tMRCA were also estimated for the RVB VP7 gene (43 sequences) from animal strains separately. The RVB strains from G3, G5 and G6–G9 genotypes, respectively, from the bovine and porcine species diverged from the human RVB strains around the year 1279 (512–1837) (node H, Fig. 1a), while the porcine RVB strains belonging to genotypes G4, G11–G16 and G18–G20 formed a different clade and shared a common ancestor that was found to arise around the year 1229 (400–1832) (node I, Fig. 1a). The evolutionary rate for the VP7 gene sequences from these strains was estimated to be $1.58 \times 10^{-3}$; HPD: 0.35–3.27 $\times 10^{-3}$ (G3, G5), 1.98 $\times 10^{-3}$; HPD: 0.64–3.84 $\times 10^{-3}$ (G4, G11–G16 and G18–G20) and 2.30 $\times 10^{-3}$ HPD: 0.62–4.60 $\times 10^{-3}$ (G6–G9) nucleotide substitutions per site per year.

### Selection pressure analyses

Site-specific selection pressure analyses revealed the existence of purifying selection pressure in all of the gene segments of the Indian–Bangladeshi lineage strains (Table S3). The mean $d_{S}/d_{A}$ values for the gene segments VP1, VP2, VP6, VP7 and NSP2 were relatively low (0.050–0.099) in comparison to the remaining genes (0.112–0.323). The VP6 gene showed the presence of highest proportion (~11%) of negatively selected codon sites (Table 3).

### DISCUSSION

Although serological findings indicate the global distribution of human RVBs, documentation of molecular detection has only revealed their presence in China, India, Bangladesh and Myanmar (Brown et al., 1987; Nakata et al., 1987; Krishnan et al., 1999; Sanekata et al., 2003; Aung et al., 2009; Saiada et al., 2011). RVB has also been isolated from other species of animals that include lambs, pigs, cows, goats and rats (Vonderfecht et al., 1984; Chasey & Banks, 1984; Theil et al., 1985; Eiden et al., 1991; Parwani et al., 1996; Tsunemitsu et al., 1999; Barman et al., 2004). According to a classification scheme based on the VP7 gene, five genotypes (G1–G5) of RVB have been proposed by Kuga et al. (2009). Among these, human RVBs are confined to the genotype G2. The other four genotypes are found in murine (G1), bovine/porcine (G3) and porcine (G4 and G5) RVB strains. Human RVBs are further classified genetically into two lineages within genotype G2 based on all of the 11 gene segments: the Chinese lineage and the Indian–Bangladeshi lineage (Yamamoto et al., 2010). Additionally, new classification systems comprising of seven [murine (A1), human (A2), ovine (A3), bovine (A4) and porcine (A5–A7)] and four [murine and human (N1), porcine (N2 and N4), bovine and porcine (N3)] genotypes have been proposed on the
basis of the NSP1 and NSP2 gene segments (Suzuki et al., 2011, 2012). Very recently, using 80% as a cut-off for nucleotide identity, 20 genotypes (G1–G20) of RVB have been proposed on the basis of the VP7 gene by Marthaler et al. (2012).

Complete understanding of locally circulating RVB strains is essential for the development of diagnostic assays and preventive measures of RVB infections. Genetic characterization of RVB strains has been carried out from time to time. To date, analysis of the full genome sequences is only available for seven human and one murine RVB strains. In the present study, analyses of nucleotide/deduced amino acid sequences carried out for all 11 gene segments of the western Indian RVB strains revealed a high level of relatedness (95.1–99.8/94.1–100 %) with other RVB strains from India, Bangladesh and Myanmar. The VP6 and NSP2 gene segments are known to be highly conserved in RVA (Matthijssens et al., 2008), Along with these two segments, VP1–VP3 were also found to be conserved in the RVB strains used in this study at the amino acid level. In addition, specific regions such as the RNA polymerase domain (VP1), the N-terminal RNA-binding domain (NSP3), motif 415ALYSLSNXXN424 (VP3), the cysteine-rich (aa 57–66) and cysteine and histidine-rich (64–76) regions (NSP1) and the C-terminal RNA-binding domain (NSP2) remained conserved. These data together with phylogenetic analysis of all strains of the present study indicated conserved genome constellation of human RVB strains. Despite this observation, it was also noted that all of the 11 gene segments carried different unique amino acid substitutions in different strains. The VP8* portion in VP4 and the N-terminal region in VP2 were found to be more divergent as has been described for RVAs occurring in nature (Gorziglia et al., 1988; Sereno & Gorziglia, 1994; McDonald & Patton, 2008). These findings are in agreement with earlier findings (Yamamoto et al., 2010), suggesting that the viral proteins of RVB and RVA have similar structural and functional characteristics and that they evolve in a similar manner at the molecular level.

The mean rate of nucleotide substitutions for all RVBs and exclusively for the human RVBs was in the overlapping range (1.36–4.78 × 10^{-3}; HPD: 0.36–9.09 × 10^{-3} vs 1.15–2.30 × 10^{-3}; HPD: 0.52–3.67 × 10^{-3}). This rate estimated only for the RVB genotype G2 VP7 genes (1.32 × 10^{-3} nucleotide substitutions per site per year) in the present study was comparable to the rate (1.57 × 10^{-3} nucleotide substitutions per site per year) reported earlier for human RVB strains (Rahman et al., 2007). It may be noted that the estimation for the latter was made on the limited gene sequence data using Path-O-Gen software that used the strict molecular clock model. In the present study, the coefficient of variation and the credible intervals invalidated the use of the strict molecular clock model. Therefore, RVB genome sequence data available to date was analysed by the relaxed molecular clock model. The evolutionary rate obtained was comparable to those of G9 (1.87 × 10^{-3} substitutions per site per year) and G12 (1.66 × 10^{-3} substitutions per site per year).

### Table 2: Estimates of mean rate of nucleotide substitution and tMRCA under relaxed uncorrelated exponential clock for different gene segments of the RVB strains

<table>
<thead>
<tr>
<th>Gene</th>
<th>All RVB strains</th>
<th>Human RVB strains</th>
<th>Chinese lineage</th>
<th>Indian–Bangladeshi lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>3.93 (1.15, 7.17)</td>
<td>1686 (1332, 1925)</td>
<td>2.23 (1.22, 3.22)</td>
<td>2.17 (1.12, 3.21)</td>
</tr>
<tr>
<td>VP2</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>VP3</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>VP4</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>VP5</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>VP6</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>VP7</td>
<td>4.78 (1.90, 6.20)</td>
<td>1679 (1221, 1947)</td>
<td>2.14 (1.07, 3.32)</td>
<td>2.09 (1.07, 3.27)</td>
</tr>
<tr>
<td>NSP1</td>
<td>1.91 (1.09, 3.34)</td>
<td>1198 (341, 1798)</td>
<td>1.66 (0.89, 2.49)</td>
<td>1.56 (0.88, 2.32)</td>
</tr>
<tr>
<td>NSP2</td>
<td>1.91 (1.09, 3.34)</td>
<td>1198 (341, 1798)</td>
<td>1.66 (0.89, 2.49)</td>
<td>1.56 (0.88, 2.32)</td>
</tr>
<tr>
<td>NSP3</td>
<td>1.91 (1.09, 3.34)</td>
<td>1198 (341, 1798)</td>
<td>1.66 (0.89, 2.49)</td>
<td>1.56 (0.88, 2.32)</td>
</tr>
<tr>
<td>NSP4</td>
<td>1.91 (1.09, 3.34)</td>
<td>1198 (341, 1798)</td>
<td>1.66 (0.89, 2.49)</td>
<td>1.56 (0.88, 2.32)</td>
</tr>
<tr>
<td>NSP5</td>
<td>1.91 (1.09, 3.34)</td>
<td>1198 (341, 1798)</td>
<td>1.66 (0.89, 2.49)</td>
<td>1.56 (0.88, 2.32)</td>
</tr>
</tbody>
</table>
Table 3. Results of purifying selection pressure acting on human RVBs

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. codons analysed (position in the gene)</th>
<th>Percentage of codon sites under purifying selection pressure (no. codons)</th>
<th>Mean $d_S/d_S$ for the gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>1153 (6–1158)</td>
<td>0.69 (8)</td>
<td>0.0842852</td>
</tr>
<tr>
<td>VP2</td>
<td>929 (6–934)</td>
<td>2.15 (20)</td>
<td>0.0606228</td>
</tr>
<tr>
<td>VP3</td>
<td>757 (7–763)</td>
<td>2.77 (21)</td>
<td>0.111764</td>
</tr>
<tr>
<td>VP4</td>
<td>747 (4–750)</td>
<td>1.87 (14)</td>
<td>0.137615</td>
</tr>
<tr>
<td>VP6</td>
<td>391 (1–391)</td>
<td>10.99 (43)</td>
<td>0.0503588</td>
</tr>
<tr>
<td>VP7</td>
<td>244 (6–249)</td>
<td>1.23 (3)</td>
<td>0.0982552</td>
</tr>
<tr>
<td>NSP1 (peptide 1)</td>
<td>107 (1–107)</td>
<td>3.74 (4)</td>
<td>0.322724</td>
</tr>
<tr>
<td>NSP1 (peptide 2)</td>
<td>321 (1–321)</td>
<td>0.93 (3)</td>
<td>0.229579</td>
</tr>
<tr>
<td>NSP2</td>
<td>301 (1–301)</td>
<td>0.33 (1)</td>
<td>0.0991883</td>
</tr>
<tr>
<td>NSP3</td>
<td>347 (1–347)</td>
<td>0.29 (1)</td>
<td>0.246812</td>
</tr>
<tr>
<td>NSP4</td>
<td>217 (3–219)</td>
<td>0.92 (2)</td>
<td>0.142937</td>
</tr>
<tr>
<td>NSP5</td>
<td>170 (1–170)</td>
<td>2.35 (4)</td>
<td>0.16774</td>
</tr>
</tbody>
</table>

Genotypes of RVAs that are known to carry genetically different variants (Matthijssens et al., 2010). Similar estimates for the G3–G9, G11–G16 and G18–G20 genotypes of animal RVB VP7 genes ($1.58–2.30 \times 10^{-3}$) were slightly high as compared with those measured for the human RVB VP7 gene. The substitution rate estimated for the human RVB VP4 gene ($2.30 \times 10^{-3}$ substitutions per site per year) was found to be higher than that of the human RVA VP4 gene ($0.58 \times 10^{-3}$ substitutions per site per year) (Jenkins et al., 2002). However, the RVB NSP2 ($1.15 \times 10^{-3}$ substitutions per site per year) and the human RVA NSP2 genes from genotype N1 ($0.87 \times 10^{-3}$ substitutions per site per year) were comparable (Donker & Kirkwood, 2012). A negative correlation between genome length and rate of nucleotide substitution has been described earlier for RNA viruses (Jenkins et al., 2002; Holmes, 2003; Donker & Kirkwood, 2012). However, such an observation was not recorded for the RVB strains of the present study.

It has been described that the Chinese lineage has a much lower rate ($7.9 \times 10^{-4}$ nucleotide substitutions per site per year) of evolution as compared with that of the Indian–Bangladeshi lineage (Yang et al., 2004; Rahman et al., 2007). Interestingly, our study revealed that the evolutionary rates estimated for the individual gene segments from the strains of both lineages were closer to each other ($0.91–2.04 \times 10^{-3}$ vs $1.14–2.26 \times 10^{-3}$ nucleotide substitutions per site per year).

Except VP7, all genes showed that the human RVBs have originated in the 20th century (1901–1995). However, the data on the VP7 gene could trace back to the origin of human RVBs around 1600 (Table 2). This could be the result of more sequence data analysed for the VP7 gene in comparison to other genes. The existence of the Chinese lineage strains was found after the 1960s; however, the Indian–Bangladeshi lineage strains were found to have a comparatively recent origin (after 1980) as evident from the data on all genes. Thus, our estimates suggest that the human RVBs must have been around since the beginning of the 19th century in their natural host, but have not received full attention from researchers.

Evolution of the gene segments of RVB from the Indian–Bangladeshi lineage is driven by negative/purifying (conserved) selection pressure. For all genes, the mean $d_S/d_S$ ratio was considerably low ($<1$), indicating that the synonymous substitutions were very high in comparison to non-synonymous substitutions. The VP6 gene displayed a high percentage ($\pm 11\%$) of negatively selected codon sites as compared with the other gene segments (Table 3). This finding should be assessed on the basis that the RNA encoding VP6 is a highly conserved gene in nature and expected to have functional constraints. Considering the major role of VP6 in rotavirus grouping, viral replication and structural integrity, a detailed study is needed to know how negative selection pressure affects the RVB maintenance and its biological function.

Although adequate sampling of the western Indian human RVB genetic pool was carried out at different time points from 1995 to 2010 [13 strains, equivalent to nine sampling events (from three outbreaks and six independent sporadic cases)], a highly conserved nature of the virus was observed in all 11 gene segments, which formed a monophyletic branch along with other human RVB strains of the Indian–Bangladeshi lineage. This pattern of lineage constellation is consistent with the long-term, independent evolution. Identification of a single lineage may be attributed to the nucleotide sequences that were derived from the strains recovered within a narrower geographical base. However, such data suggest that the virus is circulating in the natural host as a series of stable viral clones.

Genetic diversity allows a virus to adapt dynamically to its host and environment. Although the rate of evolution for the RVB VP7 gene was similar to genetically variable G9/G12 RVA strains, human RVBs are found to be highly conserved. Multiple factors may contribute towards the genetic stability. Narrow geographical confinement, single
host–pathogen relationship and shedding of RVBs in low titres may be the indirect barriers of RVB diversity. In India, RVB strains may have been introduced from China or emerged from indigenous reservoirs. It could be the founder effect of the environment that allows the virus to occupy a relatively stable fitness peak. It is well established that fitness gains in one environment often lead to fitness losses in an alternative environment (Domingo & Holland, 1997) and this may be the reason for the lack of the cosmopolitan distribution pattern of RVBs. Vaccines are available for RVAs. Vaccination against one RVA genotype may lead to increased relative prevalence of the other RVA genotypes (Pitzer et al., 2011). Such vaccine pressure has not been encountered by human RVBs since there is no licensed RVB vaccine.

Our study coupled with the previous studies (Yamamoto et al., 2010; Ghosh & Kobayashi, 2011; Lahon & Chitambar, 2011) provides the evidence of a highly conserved nature of human RVB strains recovered from the outbreak or sporadic cases of acute gastroenteritis, thus suggesting that the RVB strains have the potential to perpetuate in the environment like genetically diverse RVAs. Overall, the evolutionary inference analyses together with genetic analyses carried out in this study will be useful in enhancing the understanding of the evolutionary timescale and stasis or dynamics existing in the RVBs.

METHODS

Specimen selection. Faecal specimens were collected during 1994–1995 and 2004–2010 from sporadic cases of acute gastroenteritis to study the molecular epidemiology of RVB infections in Pune, western India. Faecal specimens were also collected during investigation of gastroenteritis outbreaks that occurred in three different regions of western India. RVB was detected in these specimens by RT-PCR targeting the NSP2 gene (Kelkar et al., 2007; Chitambar et al., 2011; Lahon et al., 2012). Six (NIV-957971/1995, NIV-0632252/2006, NIV-076222/2007, NIV-0924341/2009, NIV-0948756/2009 and NIV-1048101/2010) and seven (Daman: NIV-005623/2000, NIV-005626/2000; Surat: NIV-04623/2004, NIV-04624/2004 and Sholapur: 9222/2010, 10913/2010 and 11037/2010) RVB strains recovered, respectively, in sporadic and outbreak infections of gastroenteritis were selected for sequence analysis of all RNA segments. All these strains were isolated from adults (25–78 years) who suffered from mild to moderate gastroenteritis.

RNA extraction and RT-PCR. Viral RNA was extracted from faecal specimens using TRizol LS reagent as per manufacturer’s instructions (Invitrogen). RT-PCR was carried out for amplification of all structural and non-structural genes by using the one step RT-PCR kit (Qiagen) and gene-specific primers published earlier (Ahmed et al., 2004; Chen et al., 1990; Chen et al., 1991; Lahon & Chitambar, 2011) and designed for this study (Table S4). The full-length gene segments were amplified and sequenced. The sequences from the primer-binding regions at the 5′- and 3′-ends were excluded from the analysis.

Briefly, the extracted dsRNA was denatured at 95 °C for 5 min. PCR conditions involved an initial reverse transcription at 50 °C for 30 min, followed by PCR activation at 95 °C for 15 min, 35 cycles of amplification (1 min at 94 °C, 30 s at 45–55 °C and 1 min at 72 °C) with a final extension at 72 °C for 7 min.

All PCR products were analysed on a 2% agarose gel containing 0.5 µg ethidium bromide ml⁻¹ and visualized under the UV transilluminator.

Nucleotide sequencing and genetic analyses. The excised PCR products were purified using a QIAquick gel extraction kit (Qiagen). The sequencing of the products was carried out by using a ABI Prism Big Dye Terminator cycle sequencing ready reaction kit V3.1 (Applied Biosystems) on an automated DNA sequencer (ABI Prism 3100 Genetic Analyzer; Applied Biosystems).

The nucleotide/deduced amino acid sequences of the VP1–VP4, VP6, VP7 and NSP1–NSP5 gene segments were aligned with the sequences of the RVB strains available in GenBank by using CLUSTAL W program (Thompson et al., 1994). Per cent identity of the sequences was calculated using p-distance available in MEGA version 5 software package (Tamura et al., 2011).

GenBank accession numbers. The nucleotide sequences derived in this study were deposited in GenBank under the accession numbers JQ904147–JQ904289 (Table S5). GenBank accession numbers of the reference strains utilized in this study are provided in Table S6.

Estimation of evolutionary rates and tMRCA. The rates of nucleotide substitution and the divergence times for RVBs were estimated on the basis of temporal information of the sequences on the 11 gene segments isolated from a variety of host species by using Bayesian Markov Chain Monte Carlo (MCMC) approach available in the BEAST 1.7.1 package (Drummond & Rambaut, 2007). Especially for the VP7 gene, a total of 69 sequences were utilized that included 26 from human (19 from India and seven from other countries), four from bovine, 38 from porcine and one from murine RVB strains, while the number of sequences varied from 21 to 31 for the remaining gene segments. The best model of nucleotide substitution was selected on the basis of the Akaike information criterion, as implemented in MEGA5. The evolutionary model, TN93 + G (Tamura & Nei, 1993; model with gamma-distributed rates of variation among sites) was found to be best suitable for the dataset of each gene. The data were analysed by employing the strict molecular and relaxed clock (uncorrelated exponential and uncorrelated lognormal) (Drummond et al., 2006) with Bayesian Skyline tree prior. The Bayes factors and posterior probability were used for selecting the best model for the data (Suchard et al., 2001).

Three independent MCMC chains were run each for 30,000,000 generations, sampling every 1000th generation. The results were combined by using LogCombiner program available in BEAST with a burn-in of 10%. The results were examined for convergence of the MCMC using Tracer 1.5 (http://tree.bio.ed.ac.uk/software/tracer) (Drummond & Rambaut, 2007). The MCC tree was generated by using the Tree Annotator program available in BEAST and FigTree 1.2.3 (http://tree.bio.ed.ac.uk/software/figtree) was used for the visualization of the annotated trees. The 95% HPD intervals were used to ascertain the uncertainty in the parameter estimates. Mean clade attribute available in TreeStat program of the BEAST package was used to estimate the evolutionary rate for individual lineages.

Selection pressure analysis. Site-specific selection pressure was analysed on all 11 gene segments separately by using maximum-likelihood procedures including single-likelihood ancestor counting (SLAC), random-effects likelihood (REL) and fixed-effects likelihood
(FEL) available at the web server http://www.datamonkey.org. The analysis was carried out only for the sequences of the Indian–Bangladeshi lineage strains using the neighbouring-join tree and TN93 model of nucleotide substitution. The strength of the selection pressure was determined on the basis of the ratio of non-synonymous (\(d_N\)) to synonymous (\(d_S\)) substitutions per site (ratio \(d_N/d_S\)). Sites were considered to be under positive/negative selection if at least two of the three methods supported the hypothesis of selection pressure with high statistical significance (SLAC and FEL: \(P<0.1/\)REL: Bayes factor \(>50\)).

ACKNOWLEDGEMENTS

Authors would like to thank the Indian Council of Medical Research (ICMR), New Delhi for supporting the first author (A.L.) with a Junior Research Fellowship and Dr A. C. Mishra, Director, National Institute of Virology, Pune for his constant support during this study.

REFERENCES


Kellner, S. D. & Zade, J. K. (2004). Group B rotaviruses similar to strain CAL-1, have been circulating in Western India since 1993. Epidemiol Infect 132, 745–749.


