Molecular characterization of emerging G9P[4] rotavirus strains possessing a rare E6 NSP4 or T1 NSP3 genotype on a genogroup-2 backbone using a refined classification framework

Gauri N. Pradhan, Atul M. Walimbe and Shobha D. Chitambar

Enteric Viruses Group, National Institute of Virology, Pune, India
Bioinformatics Group, National Institute of Virology, Pune, India

Rotavirus infections associated with unusual strains are an emerging concern in rotavirus vaccination programmes. Recently, an increase in circulation of unusual G9P[4] strains was reported from different regions of India, placing this genotype in third position, after G1P[8] and G2P[4], of the most common rotavirus strains. The aim of the present study was to analyse the complete genomic constellation of three G9P[4] strains (RV09, RV10 and RV11), determine their genetic relatedness to other genogroup-2 strains and understand the evolution of a rare E6 and other NSP4 genotypes. All strains revealed the presence of a genogroup-2 backbone, with RV09 constituting the NSP3 T1 genotype and RV10 and RV11 bearing the NSP4 E6 genotype. A refined criterion adopted to classify the nine internal gene segments of G2P[4] and non-G2P[4] strains with the genogroup-2 backbone into lineages and sub-lineages indicated divergence of >8% (except NSP1: >5.5%) for lineages and >3% for sub-lineages. The VP1 and/or VP3 genes of study strains showed close relationships with animal-like human rotaviruses. The estimated evolutionary rate for the NSP4 E6 genotype was marginally higher (3.78 × 10^(-3) substitutions per site per year) than that of genotypes E1 (2.6 × 10^(-3) substitutions per site per year) and E2 (3.06 × 10^(-3) substitutions per site per year), suggesting a step towards adaptation of E6 on a genogroup-2 backbone. The time and origin of the most recent common ancestor of E6 genotype were estimated to be 1981 and South Asia, respectively. Full-genome and evolutionary analyses performed in this study for G9P[4] strains will help better understand the extent of gene reassortment and origin in unusual rotavirus strains that may remain viable and cause infections in humans.

INTRODUCTION

Group A rotaviruses (RVAs), a major cause of severe gastro-enteritis in infants and young children, belong to the family Reoviridae. The viral genome consists of 11 segments of double-stranded RNA (dsRNA) that encode six structural (VP1–VP4, VP6, VP7) and six non-structural (NSP1–NSP6) proteins enclosed in a triple-layered capsid. The innermost capsid layer, formed mainly of structural protein VP2, surrounds viral dsRNA. Each of the 11 dsRNA segments is associated with one copy of VP1 (the viral polymerase) and VP3 (the mRNA capping enzyme) proteins. The middle layer of the virion is formed of structural protein VP6, that interacts with the inner layer protein, VP2, and with the two outer layer proteins, VP7 and VP4. The non-structural proteins (NSP1–NSP5) interact with viral RNA and are involved in viral replication and interaction with cellular proteins (Estes & Kapikian, 2007).

The attachment of rotavirus to host cells is mediated by VP7 and VP4 proteins that elicit neutralizing antibodies during infection of a host. Both proteins form the basis for a dual classification system that defines, G- and P-genotypes, respectively, of RVA isolates (Estes & Kapikian, 2007). In 2008, a nucleotide-sequence-based complete genome classification was developed for RVAs (Matthijssens et al., 2008). In accordance with this criterion, all 11 gene segments of RVAs, VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6, have been classified into genotypes 28G-39P-14R-14C-13M-24A-14N-L.
16T-21E-16H, respectively, highlighting the vast diversity in rotavirus strains (Matthijnssens et al., 2011; Matthijnssens & Theuns, 2015). Genogroup classification of commonly detected RVAs into genogroups 1 and 2 is based on the nucleotide sequences of the nine internal gene segments. Likewise, a genogroup constellation of I1-R1-C1-M1-A1-N1-T1-E1-H1 (genogroup-1) and I2-R2-C2-M2-A2-N2-T2-E2-H2 (genogroup-2) has been described for G1P[8]/G3P[8]/G4P[8]/G9P[8] and G2P[4] RVAs, respectively (Matthijnssens & Van Ranst, 2012).

Although, 28 G- and 39 P-types have been reported among RVAs (Matthijnssens & Theuns, 2015), only 13 and 15 different G-types (Matthijnssens et al., 2011; Ward et al., 2016) and P-types (Esona et al., 2009; Matthijnssens et al., 2011), respectively, are currently known to cause infections in humans. Of these, only a limited number of G-P genotype combinations such as G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and, in the recent past, G12P[8]/P[6] (Rahman et al., 2001; Arista et al., 1998; Okada et al., 1987; Phan et al., 2007) (Fig. 1a). VP4, encoding gene segments, shows 99.4 % and 93.4 % nucleotide identity, respectively (Table 1). In India, occurrence of G9P[4] strains was rarely detected in strains performed before 2005. However, very low detection rates (0.22–0.86 %) of these strains have been described in hospital-based rotavirus surveillance studies conducted during 2005–2009 (Kang et al., 2009, 2013; Babji et al., 2014; Chitambar et al., 2014; Mullick et al., 2014; Tiku et al., 2014).

The G9 genotype has been reported to circulate in combination with P[8], P[6], P[4], P[11], P[10] and P[19] in humans in different parts of the world (Clark et al., 1987; Das et al., 1994; Ramachandran et al., 1998; Okada et al., 2000; Ramachandran et al., 2000; Araújo et al., 2001; Arista et al., 2003; Armah et al., 2003; Santos et al., 2002). Prevalence of strain G9P[4], a reassortant between common strains G2P[4] and G9P[8], has been noted to be as high as 66, 36 and 15.3 % in Guatemala, Honduras and Bangladesh, respectively (Quaye et al., 2013). Few of such strains have been shown to possess the rare NSP4 E6 genotype or NSP3 T1 genotype on the genogroup-2 backbone (Quaye et al., 2013; Lewis et al., 2014; Yamamoto et al., 2015). In India, occurrence of G9P[4] strains was rarely reported in studies performed before 2005. However, very low detection rates (0.22–0.86 %) of these strains have been described in hospital-based rotavirus surveillance studies conducted during 2005–2009 (Kang et al., 2009, 2013). More recently, increase in the detection rates of these strains by up to 21.1 % have been documented, assigning this genotype to third position in human infections after the commonly detected G1P[8] and G2P[4] strains (Babji et al., 2014; Chitambar et al., 2014; Mullick et al., 2014; Saluja et al., 2014; Tiku et al., 2014).

The aim of the present study was to investigate the genetic backbone of three G9P[4] rotavirus strains detected in Pune, India during the 2009–2011 rotavirus surveillance period (Chitambar et al., 2014). To determine the genomic relatedness of these strains with other genogroup-2 strains circulating worldwide, phylogenetic analysis of all individual gene segments was carried out and a refined framework was employed for classification of the nine internal gene segments. One of the strains (RV09) displayed inter-genogroup reassortment for gene NSP3 (T1-genogroup), while the remaining two strains (RV10, RV11) showed the presence of a rare genotype, E6 of gene NSP4 detected earlier in Latin American G9P[4] strains. To gain an insight into the evolutionary process of the NSP4 E6 genotype, the study was further focused on estimation of the evolutionary rates and the time of the most recent common ancestor (tMRCA) for commonly (E1, E2) detected and other (E3–E12) NSP4 genotypes.

**RESULTS**

**Genotyping and genogroup constellation of G9P[4] strains**

Nucleotide sequence analysis of all 11 gene segments of G9P[4] strains, RV09, RV10 and RV11 selected respectively from the years 2009, 2010 and 2011 revealed that the genomic constellation of RV09 was G9-P[4]-I2-R2-C2-M2-A2-N2-T1-E2-H2 and that of RV10 and RV11 was G9-P[4]-I2-R2-C2-M2-A2-N2-T6-E-H2 for VP7 (G)-VP4(P)-VP6(I)-VP1(R)-VP2(C)-VP3(M)-NSP1(A)-NSP2(N)-NSP3(T)-NSP4(E) and NSP5(H) gene segments, respectively, thus indicating that the strains were a double gene reassortant on a genogroup-2 (DS-1-like) backbone. The NSP3 gene sequence of RV09 clustered with the sequences of genotype T1 strains displaying inter-genogroup reassortment between genogroup 1 (Wa-like) and 2 (DS-1-like), while a rare E6 genotype of NSP4 gene was detected in strains RV10 and RV11. At the whole-genome level, RV10 and RV11 are closely related to each other with a nucleotide identity of 99.1 % and distantly related to RV09 (95.2–95.4 %). In comparison to the G9P[4] strains from the USA (LB1562) and Japan (S120088) for which sequence data for all 11 gene segments are available in the GenBank database, strain RV09 showed 95.2 and 96.1 % nucleotide identity, respectively, and RV10 and RV11 showed 99.0–99.4 % and 93.4–93.5 % nucleotide identity, respectively (Table 1).

**Phylogenetic analysis of outer capsid VP7 and VP4 encoding gene segments**

Phylogenetic dendrograms were reconstructed for the VP7 and VP4 gene segments of the study strains, along with those of reference strains retrieved from the GenBank database. The VP7 gene that encodes one of the major neutralization antigens of rotavirus is classified into six lineages and eleven sub-lineages in G9 rotaviruses. Phylogenetic analysis clustered the VP7 gene of the three study strains along with the USA strain (LB1562) within G9 lineage III-d, the emerging lineage found globally (Phan et al., 2007) (Fig. 1a). VP4, the viral haemagglutinin encoded by segment 4, comprises two distinct regions (VP8* and VP5*) that are generated following proteolytic cleavage. The VP4 gene of the three study strains clustered along with the USA (LB1562) and...
Japan (S120088) strains within genotype P[4] lineage-5 of the five lineages described for this genotype (Espínola et al., 2008) (Fig. 1b).

**Phylogenetic analyses of the nine internal gene segments of genogroup-2 strains**

To understand the genomic relatedness and evolution in the reassortant G9P[4] strains of the study, phylogenetic dendrograms were reconstructed for the individual gene segments of the study strains along with 189 other genogroup-2 [G2P[4], n=151 and non-G2P[4], n=38] strains retrieved from the GenBank database. Each of the nine gene segments of genogroup-2 rotavirus strains were classified into lineages and sub-lineages on the basis of bootstrap probability and nucleotide sequence divergence. Initially, different clusters with high bootstrap support (>70 %) were identified in the tree. Then, the average divergence within and between the different clusters was calculated (Table 2). All the information thus obtained was used to define the cut-off between lineages and sub-lineages for each gene. The nucleotide sequence identities of strains within each sub-lineage ranged from 97.5 to 100 %, with less than 2.5 % genetic difference. The nucleotide sequence divergence between sub-lineages within the same lineage ranged from 3 to 7 %, while that between lineages was in the range of 1–8 %, except for NSP1 (5.5–7.5 %). Thus, a lineage or sub-lineage was defined as a cluster of sequences with >70 % bootstrap support and a cut-off value of more than 8 % (except NSP1: >5.5 %) for lineage and more than 3 % for sub-lineage. The identification of 6–8 and 5–11 different lineages and sub-lineages, respectively, in VP1, VP3 and NSP4 genes demonstrated the existence of higher intragenotypic diversity amongst these genes as compared to the remaining gene segments with 2–4 and 3–4 different lineages and sub-lineages, respectively.

Comparison of the nucleotide/deduced amino acid sequence identities of the nine internal gene segments of the three study strains was made with USA/Japan G9P[4] strains, as described in Table 1. The gene segments VP2, VP6, NSP1, NSP2 and NSP5 of all the study strains clustered in the lineages/sub-lineages similar to those of the USA and Japan strains, while the remaining four gene segments, VP1, VP3, NSP3 and NSP4, differed from the USA and/or Japan strains in their lineages/sub-lineages (Table 3).

Phylogenetic analysis of genotype R2 of VP1, the RNA-dependent RNA polymerase encoded by segment 1 of genogroup-2 strains, showed intra-genotypic variation with seven lineages and eight sub-lineages (Fig. 2a). The three study strains clustered in R2 lineage 7i. The majority of the strains from this lineage were closely related to animal-like human rotavirus strains described previously (Dennis et al., 2014). VP2 is the inner core protein encoded by the gene segment 2. The C2 genotype of this segment detected in genogroup-2 strains was found to be relatively conserved as compared to other gene segments, with only two lineages and four sub-lineages. Most of the strains, along with the three study strains, were found to cluster together in C2-lineage 2-iv (Fig. 2b). VP3, the protein product of the third gene segment, is a guanylyl- and methyl transferase. The M2 genotype of this gene showed clustering of the genogroup-2 strains into six lineages and five sub-lineages. Most of the strains included in the analysis clustered in different sub-lineages of lineage 1. At the sub-lineage level, RV09 clustered in M2 lineage 1 v while RV10 and RV11 clustered in lineage 5, known to share a common ancestor (Dennis et al., 2014) (Fig. 2c). VP6, the major structural protein encoded by gene segment 6, carries the group- and sub-group-specific epitopes. The L2 genotype of this gene in genogroup-2 strains displayed three lineages and three sub-lineages. The

### Table 1. Percentage nucleotide (nt) and amino acid (aa) sequence identities of individual gene segments of G9P[4] strains from Pune as compared to strains USA (LB1562) and Japan (S120088)

<table>
<thead>
<tr>
<th>Gene</th>
<th>RV09 nt/aa sequence identity (%) to:</th>
<th>RV10 nt/aa sequence identity (%) to:</th>
<th>RV11 nt/aa sequence identity (%) to:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LB1562</td>
<td>S120088</td>
<td>LB1562</td>
</tr>
<tr>
<td>VP7</td>
<td>97.6/98.5</td>
<td>92.7/96.6</td>
<td>98.7/99.4</td>
</tr>
<tr>
<td>VP4</td>
<td>98.9/99.0</td>
<td>97.5/99.4</td>
<td>99.4/99.5</td>
</tr>
<tr>
<td>VP6</td>
<td>98.8/99.2</td>
<td>97.0/99.7</td>
<td>99.3/99.7</td>
</tr>
<tr>
<td>VP1</td>
<td>98.4/99.3</td>
<td>93.7/98.7</td>
<td>99.2/99.4</td>
</tr>
<tr>
<td>VP2</td>
<td>97.5/99.7</td>
<td>97.7/100</td>
<td>99.6/99.7</td>
</tr>
<tr>
<td>VP3</td>
<td>86.5/92.9</td>
<td>97/96.9</td>
<td>99.1/99.2</td>
</tr>
<tr>
<td>NSP1</td>
<td>98.8/99.0</td>
<td>96.8/97.5</td>
<td>99.0/99.4</td>
</tr>
<tr>
<td>NSP2</td>
<td>98.3/98.7</td>
<td>97.7/98.4</td>
<td>99.3/99.7</td>
</tr>
<tr>
<td>NSP3</td>
<td>76.2/80.8</td>
<td>93.3/97.1</td>
<td>99.3/99.4</td>
</tr>
<tr>
<td>NSP4</td>
<td>78.2/85.1</td>
<td>87.8/95.4</td>
<td>99.7/100</td>
</tr>
<tr>
<td>NSP5</td>
<td>98.9/100</td>
<td>96.7/100</td>
<td>99.9/100</td>
</tr>
<tr>
<td>Genome</td>
<td>95.2/91.7</td>
<td>96.1/93.4</td>
<td>99.4/98.7</td>
</tr>
</tbody>
</table>
study strains clustered in lineage 3iii along with most of the genogroup-2 strains (Fig. 2d).

NSP1 encoded by segment 5 is involved in evasion from the innate immune response to rotavirus infection. The sequences of NSP1 A2 genotype in genogroup-2 strains were found to be distributed among three lineages and four sub-lineages. Although the NSP1 gene is known to be highly divergent in nature, the diversity was low at the A2 genotype level, with most of the strains, including the three study strains, clustering in lineage 3iii (Fig. 2e). NSP2, encoded by gene segment 8, has a role in replication of rotavirus. The N2 genotype of the NSP2 gene showed clustering of the sequences in four lineages and three sub-lineages, with most of the genogroup-2 strains, including the study strains, in lineage 4 (Fig. 2f). NSP3 is a key player in the assembly of the 11 viral plus-stranded RNAs into the early replication intermediates during rotavirus morphogenesis. Phylogenetic analysis of the NSP3 gene sequences of the three study strains placed RV09 in sub-lineage 1a of the two sub-lineages of genotype T1 (Fig. 2g), and thus showed inter-genogroup reassortment between genogroups 1 and 2. RV10 and RV11 clustered along with most of the genogroup-2 strains in the lineage IIIi of genotype T2 (Fig. 2h).

NSP4 encoded by the gene segment 10 is the viral enterotoxin. Phylogenetic analysis of NSP4 gene sequences of the study strains placed RV09 in lineage 4i of the eight lineages and eleven sub-lineages of the E2 genotype (Fig. 2i), and RV10 and RV11 within a rare genotype, E6, bearing only a single lineage (Fig. 2j). Thus, the E2 genotype exhibited higher intra-genotypic diversity than the E6 genotype that was found to be relatively conserved. NSP5 encoded by gene segment 11 is a phosphoprotein implicated in viroplasm formation. Most of the genogroup-2 strains along with the study strains clustered in lineage IIIi of H2 genotype of this type, except for a single strain, 69M, that clustered in lineage 2 (Fig. 2k).

Amino acid substitutions in antigenic and other functional domains of structural and non-structural proteins of the study strains

Several amino acid substitutions were noted in the VP7-encoding gene of the study strains as compared to other G9-IIIId strains. Out of the 24 amino acid positions defined for lineage classification (Phan et al., 2007), F106L was noted in RV09 while, D100N, which is also a part of

![Fig. 1. Phylogenetic dendrograms reconstructed on the basis of nucleotide sequences of the ORF of rotavirus VP7 (a) and VP4 (b) genes using the maximum-likelihood method included in the MEGA 6.0 software package with 1000 bootstrap replicates. The G9P[4] strains from the study, the USA and Japan are denoted by filled circles, square and triangle, respectively. Scale bar indicates genetic distance.]
Table 2. Gene-wise average nucleotide divergence between lineages and sub-lineages of genogroup-2 RVA strains

<table>
<thead>
<tr>
<th>Gene</th>
<th>Lineage</th>
<th>Within (%)</th>
<th>Between (%)</th>
<th>Cut-off (%)</th>
<th>Sub-lineage</th>
<th>Within (%)</th>
<th>Between (%)</th>
<th>Cut-off (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>Inner</td>
<td>0.1–5.8</td>
<td>7.1–14.1</td>
<td>8</td>
<td>0.1–1.9</td>
<td>3.7–6.8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VP2</td>
<td>Intermediate</td>
<td>2.5</td>
<td>15.1–15.3</td>
<td>8</td>
<td>1.6–2.3</td>
<td>3.5–6.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VP3</td>
<td>Outer</td>
<td>0.6–3.1</td>
<td>8.4–17.2</td>
<td>8</td>
<td>1.4–2.2</td>
<td>4.0–7.2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>VP6</td>
<td>E154K</td>
<td>1.7–2.3</td>
<td>11.8–14.2</td>
<td>8</td>
<td>2.2</td>
<td>6.7–7.4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSP1</td>
<td>within</td>
<td>1.0–3.2</td>
<td>5.5–7.5</td>
<td>5.5</td>
<td>2.3–2.5</td>
<td>4.0–5.3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSP2</td>
<td>core</td>
<td>2.3–3.2</td>
<td>8.7–13.9</td>
<td>8</td>
<td>0.8–2.3</td>
<td>3.5–4.4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSP3</td>
<td>VP2</td>
<td>2.1</td>
<td>8.5–9.5</td>
<td>8</td>
<td>1.6–2.0</td>
<td>3.2–4.6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSP4</td>
<td>VP3</td>
<td>0.4–6.1</td>
<td>9.2–16.8</td>
<td>8</td>
<td>0.5–2.6</td>
<td>4.4–7.3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSP5</td>
<td>VP1</td>
<td>1.4</td>
<td>8.9–9.6</td>
<td>8</td>
<td>1.2–1.9</td>
<td>3.2–4.6</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Lineage and sub-lineage constellation of all 11 gene segments of G9P[4] strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>VP7</th>
<th>VP4</th>
<th>VP1</th>
<th>VP2</th>
<th>VP3</th>
<th>VP6</th>
<th>NSP1</th>
<th>NSP2</th>
<th>NSP3</th>
<th>NSP4</th>
<th>NSP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV09</td>
<td>G9-III</td>
<td>d</td>
<td>P[4]-Va</td>
<td>R2-7i</td>
<td>C2-iv</td>
<td>M2-1iv</td>
<td>I2-3ii</td>
<td>A2-3ii</td>
<td>N2-5</td>
<td>T1 E2-1v</td>
<td>H2-1ii</td>
</tr>
<tr>
<td>RV10</td>
<td>G9-III</td>
<td>d</td>
<td>P[4]-Va</td>
<td>R2-7i</td>
<td>C2-iv</td>
<td>M2-5</td>
<td>I2-3ii</td>
<td>A2-3ii</td>
<td>N2-5</td>
<td>T2-1ii</td>
<td>E6</td>
</tr>
<tr>
<td>RV11</td>
<td>G9-III</td>
<td>d</td>
<td>P[4]-Va</td>
<td>R2-7i</td>
<td>C2-iv</td>
<td>M2-5</td>
<td>I2-3ii</td>
<td>A2-3ii</td>
<td>N2-5</td>
<td>T2-1ii</td>
<td>E6</td>
</tr>
<tr>
<td>LB1562 (USA)</td>
<td>G9-III</td>
<td>d</td>
<td>P[4]-Va</td>
<td>R2-7i</td>
<td>C2-iv</td>
<td>M2-5</td>
<td>I2-3ii</td>
<td>A2-3ii</td>
<td>N2-5</td>
<td>T2-1ii</td>
<td>E6</td>
</tr>
<tr>
<td>S120088 (Japan)</td>
<td>G9-IV</td>
<td>P[4]-Vb</td>
<td>R2-7ii</td>
<td>C2-iv</td>
<td>M2-1iv</td>
<td>I2-3ii</td>
<td>A2-3ii</td>
<td>N2-5</td>
<td>T1</td>
<td>E2-1v</td>
<td>H2-1ii</td>
</tr>
</tbody>
</table>
Fig. 2. Phylogenetic dendrogram reconstructed on the basis of the ORFs of (a) VP1 R2 genotype, (b) VP2 C2 genotype, (c) VP3 M2 genotype, (d) VP6 I2 genotype, (e) NSP1 A2 genotype, (f) NSP2 N2 genotype, (g) NSP3 T1 genotype, (h) NSP3 T2 genotype, (i) NSP4 E2 genotype, (j) NSP4 E6 genotype and (k) NSP5 H2 genotype using the maximum-likelihood method included in the MEGA 6.0 software package with 1000 bootstrap replicates. The G9P[4] strains from this study, the USA and Japan are denoted by filled circles, square and triangle, respectively. Scale bar indicates genetic distance. The countries (ISO 3166-1-alpha-2 Codes) and sampling times of the strains included in the collapsed branches are also indicated.
for the VP2 gene of G3P[8] strains (Wang et al., 2014). Strain-specific amino acid substitutions in the N-terminal domain were recorded in RV10 and RV11 (Table S1).

The VP3 protein of the study strains showed conserved proline residues at fifteen positions and cysteine residues at four positions that have been implicated as having structural importance (Cook & McCrae, 2004). Motifs KXTAMDXE XP and KXXGNNH, around amino acids 385 and 545, respectively, regarded as the putative enzymatic active sites, were conserved in all the study strains. The conserved motif (415ALYSLSNXXN424) of this protein, found in all rotavirus groups (Ito et al., 2001; Nagashima et al., 2008), was also detected in the study strains. Strain-specific amino acid substitutions were found in the VP3 gene of all study strains (Table S1).

Immunodominant sites on VP6 of RVAs, mapped to the regions 32–64, 155–167, 208–294 and 380–397 (Holmes, 1983), were found to be conserved in all of the study strains.
except for one unique amino acid substitution detected in RV10 and RV11 at position 291 (S–L). It has been stated that Ala172, Arg296, Asn299, Ala305 and Asn310 contribute to reactivity to SGI MAb. (Tang et al., 1997). These residues were found to be conserved in all study strains.

A conserved N-terminal zinc-binding motif, along with eleven conserved cysteines and nine conserved prolines (Xu et al., 1994), were noted in the protein encoded by the NSP1 gene of the study strains. Two unique amino acid substitutions, V62I and N377D, were detected in the RV09 strain. Amino acid residue (V), at position 75, known to be under positive selection pressure (Donker et al., 2012) in genotype N2 of the NSP2 gene, was found to be conserved in all three study strains. A few strain-specific amino acid substitutions were noted in RV10 and RV11 (Table S1). Cysteines located at positions 123 and 139 and critical for formation of oligomeric structures (Piron et al., 1999) of NSP3, were found to be conserved in all of the study strains. The consensus sequence (I/L)XXM(I/L)(S/T)XXG present at positions 104–112, essential for sequence-specific single-stranded RNA binding (Mattion et al., 1992; Rao et al., 1995), was found to be conserved with LRLLLSSKG in RV09 (genotype T1) and LRMSSLSSKG in RV10 and RV11 (genotype T2), suggesting that residues at positions 106–107 showed a difference with respect to the genotype, although were conserved within a genotype. Amino acid substitutions were found in the hydrophobic heptad repeat in RV09 (V42I, Q183L) and in the basic region near the putative ssRNA binding motif of RV10 (N/S125R, R128I) as in RV09 (V42I, Q183L) and in the basic region near the putative ssRNA binding motif of RV10 (N/S125R, R128I) as in RV09 (V42I, Q183L) and in the basic region near the putative ssRNA binding motif of RV10 (N/S125R, R128I) as in RV09 (V42I, Q183L).

Phylogenetic analysis of NSP4 genotypes (E1–E12)

Phylogenetic analysis of NSP4 genotypes (E1–E12)

Based on the Maximum Clade Credibility (MCC) tree, the posterior support for most of the major nodes corresponding to the NSP4 genotypes was greater than 0.85, except for genotypes E1 (0.66) and E2 (0.55) (Fig. 3a). The overall estimated rate of nucleotide substitution for the NSP4 gene was 4.52 × 10⁻³, with 95% highest posterior density (HPD) limits between 3.9 × 10⁻³ and 5.08 × 10⁻³ substitutions per site per year. The mean rate of nucleotide substitution for the common E1 and E2 genotypes was 2.6 × 10⁻³ (HPD: 2.2–3.17 × 10⁻³) and 3.06 × 10⁻³ (HPD: 2.5–3.6 × 10⁻³) substitutions per site per year, respectively. That for the rare E6 genotype was 3.78 × 10⁻³ (HPD: 2.7–4.8 × 10⁻³) substitutions per site per year. Based on these results, the ancestral root of NSP4 gene dates back to the 19th century (1845). Genotypes E1–E4 and E8 showed old ancestral history (1952–1976), while genotypes E5–E7 and E9–E12 showed relatively recent ancestral history (1970–1999) (Table 4). The most probable ancestral geographical locations for genotypes E1 and E6 were found to be North America and South Asia, respectively (Fig. 3b), while that of genotype E2 could not be predicted (Table 4).

DISCUSSION

The segmented nature of the rotavirus genome enables reassortment of virus genes during co-infection in a single host with more than one rotavirus strain. Although genes encoding outer capsid proteins VP7 and VP4 have a high tendency to reassort, such reassortment events were considered uncommon in the other nine internal gene segments (Ward et al., 1990; Heiman et al., 2008). However, the complete genomic analysis carried out for three unusual G9P[4] rotavirus strains in the present study highlights the existence of reassortment of non-structural gene NSP3 or NSP4 on a genogroup-2 backbone. The occurrence of NSP3 T1 genotype on the genogroup-2 backbone in one of the study strains (RV09) is in concurrence with the results reported earlier for Japanese G9P[4] strain (S120088) (Yamamoto et al., 2015). NSP3 gene, described as interacting, extensively with host cell factors rather than with any other viral proteins, has been considered non-essential for replication in cell culture, and therefore may have a tendency to reassort more frequently as compared to other internal genes (Poncet et al., 1993; Piron et al., 1998; Heiman et al., 2008). Such reassortment involving the NSP2 N1 genotype on the genogroup-2 backbone over a period of six years has been reported for RVA strains in Japan (Doan et al., 2012). In view of this, persistence of reassortment of the NSP3 gene needs to be checked in RVA strains causing infections in humans.

Genomic analysis of G9P[4] strains from 2010 and 2011 identified a rare NSP4 genotype, E6, on a genogroup-2 backbone. The E6 genotype was first identified in G12P[6] strains from Bangladesh and was predicted to be distantly related to NSP4 genotype C, which included AU-1-like strains (Rahman et al., 2014). Recent studies from Latin America and the USA have reaffirmed the existence of this genotype in G9P[4] strains (Quaye et al., 2013; Lewis et al., 2014). The emergence of this rare genotype in G9P[4] strains investigated in this study could be attributed either to a possible selective pressure of host population immunity to common E2 genotype, or to the slightly higher evolutionary rate of the E6 genotype as also detected in this study, thus facilitating circulation of the E6 genotype over the commonly detected E2 genotype.

Human G9 lineage III rotavirus strains that have emerged since the mid-1990s are believed to have resulted from reassortment events involving human and porcine G9 rotaviruses over a period of time (Clark et al., 2004). Of the four
Fig. 3. (a) Maximum clade credibility (MCC) tree for rotavirus NSP4 genotypes, with most probable ancestral time for different nodes. Collapsed branches correspond to NSP4 genotypes. The numbers correspond to the posterior supports in square brackets and the tMRCA (time to most recent common ancestor) estimate of key nodes. (b) MCC tree for rotavirus NSP4
Genotypes, with most probable ancestral location for different nodes. Number of sequences included in the analysis: 320. The countries (ISO 3166-1-alpha-2 Codes) and sampling times of the strains included in the collapsed branches are also indicated.

Sub-lineages of lineage III, sub-lineage IIId has been reported in many regions around the world (Phan et al., 2007). Although genetic relatedness to this sub-lineage was exhibited by the VP7 gene of the study strains, amino acid substitutions identified in antigenic region 7-1 and B-cell and CTL epitopic regions as compared to other IIId strains suggest possible changes in the antigenicity of these strains. Likewise, several amino acid substitutions were identified in the antigenic regions (8–1, 8–3, 8–4 and 5–1) of lineage 5 of VP4 P[4] genotype of the study strains as compared to other P[4]–5 strains. It may be noted that among the licensed rotavirus vaccines, only Rotavac launched in India contains the G9 (lineage II) component, and that the P[4] component has not been included in any of these vaccines. In such circumstances, it would be important to track the impact of monovalent/pentavalent vaccine on unusual G9P[4] genotypes in the post-vaccination period.

Genotype constellation and evolutionary patterns of all 11 gene segments with the genogroup-2 backbone have been described recently (Doan et al., 2015). This study was limited by the inclusion of only G2P[4] strains, and defined the genetic lineages by considering the clustering pattern of all genes simultaneously without taking into account gene-wise nucleotide divergence between clusters. In the present study, a classification system for internal gene segments was updated by using the criterion for genetic lineages by considering the clustering pattern of all genes and intra-genotypic diversity in VP1 and VP3 genes. This was also reflected in the genotype-specific amino acid residues described for these genes (Heiman et al., 2008). It is worth noting that the VP1 gene of all three study strains and the VP3 gene of the strains from 2010 and 2011 have an origin common to animal-like human strains, being similar to those of the USA strain reported in 2010. These data thus suggest that a distinct genetic constellation of G9P[4] strains has been circulating in different parts of world. An insertion in the N-terminal part of the VP2 gene, in the form of a 3-nucleotide duplication similar to that reported earlier for the genogroup-1 VP2 gene (Wang et al., 2014), was noted in RV09. The N-terminus of VP2, carrying variations and sites interactive with VP1/VP3, has been

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Most likely ancestral location (probability)</th>
<th>tMRCA (Time to the Most Recent Common Ancestor) with 95% HPD (Highest Posterior Density) interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>Unpredictable</td>
<td>167 (120, 214)</td>
</tr>
<tr>
<td>E1</td>
<td>North America (0.51)</td>
<td>51 (44, 60)</td>
</tr>
<tr>
<td>E2*</td>
<td>North America (0.24)</td>
<td>45 (40, 51)</td>
</tr>
<tr>
<td>E3</td>
<td>East Asia (0.59)</td>
<td>44 (38, 51)</td>
</tr>
<tr>
<td>E4*</td>
<td>North and West Europe (0.27)</td>
<td>42 (36, 49)</td>
</tr>
<tr>
<td>E5†</td>
<td>South, North-west Europe (0.67)</td>
<td>22 (17, 27)</td>
</tr>
<tr>
<td>E6</td>
<td>South Asia (0.55)</td>
<td>31 (24, 42)</td>
</tr>
<tr>
<td>E7</td>
<td>North America (0.94)</td>
<td>31 (30, 34)</td>
</tr>
<tr>
<td>E8†</td>
<td>South, North-west Europe (0.54)</td>
<td>43 (39, 47)</td>
</tr>
<tr>
<td>E9*</td>
<td>South-east and West Asia (0.29)</td>
<td>19 (15, 26)</td>
</tr>
<tr>
<td>E10</td>
<td>South and Central America (0.47)</td>
<td>18 (13, 25)</td>
</tr>
<tr>
<td>E11</td>
<td>North and West Europe (0.78)</td>
<td>35 (33, 38)</td>
</tr>
<tr>
<td>E12</td>
<td>South and Central America (0.89)</td>
<td>27 (23, 33)</td>
</tr>
</tbody>
</table>

*The most likely ancestral location is unclear, hence cannot be predicted.
†The most likely ancestral location was obtained by clubbing the corresponding probabilities of the regions in geographical proximity.

http://jgv.microbiologyresearch.org
considered important for viral genome replication and transcription. Insertion and variations in this domain of VP2 may influence its binding with VP1/VP3 (McDonald & Patton, 2008) and, finally, viral protein–protein interactions. While specific functional regions of VP1–VP3 and VP6 proteins were found to be conserved, a few strain-specific substitutions were noted in the study strains.

In the present study, amongst the non-structural genes of genogroup-2 strains, NSP1–NSP3 and NSP5 were conserved while NSP4 exhibited diversity. The NSP1 gene, known to be the most divergent rotavirus gene, showed low (7%) nucleotide sequence divergence between the three lineages of A2 genotype as compared to the NSP4 gene, which showed the highest (10–17%) intra-genotypic diversity of the E2 genotype with eight lineages. Inter-genotypic diversity between the three study strains was displayed by the genogroup-1-like NSP3 gene in RV09 and a rare E6 NSP4 genotype in RV10 and RV11. In accordance with the data obtained for the four structural genes (VP1–VP3, VP6), most of the functional domains of the non-structural genes (NSP1–NSP5) of the study strains were conserved.

It is worth pointing out that in our earlier studies performed to assess genetic diversity in genogroup-2 rotavirus strains during 2009–2013, 10/30 (33.3%) belonged to E6 NSP4 genotype while 20/30 (66.6%) clustered with E2 NSP4 genotype, indicating frequent association of a rare NSP4 E6 genotype with genogroup-2 strains. This observation, together with those reported by others (Rahman et al., 2007; Sharma et al., 2009; Lewis et al., 2014), encouraged us to determine the evolutionary pattern of this genotype with respect to common and other NSP4 genotypes. Overall, the mean rate of evolution for the NSP4 gene was comparable to that of the group B rotavirus NSP4 gene (Lahon et al., 2012). The estimated evolutionary rates for common E1 and E2 NSP4 genotypes were higher than that reported by Tatsumi et al. (2014) for E1 genotype [1.4×10−3 sub per site per year; HPD: (0.65, 2.31)×10−3] and by Doan et al. (2015) for E2 genotype [1.05×10−3 sub per site per year; HPD: (0.76, 1.36)×10−3]. The higher rate of evolution estimated in this study as compared to those of the other studies (Tatsumi et al., 2014; Doan et al., 2015) may be attributed to the different evolutionary models, tree priors and clock models used for the analysis. Moreover, the time frames and region-specific data may also have contributed to the difference. The evolutionary rate reported for E6 genotype was marginally higher than that for the common E1 and E2 genotypes. It is to be noted that most of the E6 genotype strains included in the analysis were found in association with genogroup-2 gene segments. These data may suggest adaptation of E6 on a genogroup-2 backbone, were noted in the study. Additional insight was obtained by analysis of the VP1 and VP3 genes of the study strains, which displayed relatedness to animal-like human rotavirus strains. These data, together with data from earlier studies (Rahman et al., 2007; Lewis et al., 2014), indicate that the genomic constellation of G9 strains carrying animal-derived VP1 and/or VP3 gene(s), along with a rare E6 NSP4 genotype or a T1 NSP3 genotype on a genogroup-2 backbone, has been in circulation. Taking into account the complexity in genetic make-up of the G9P[4] strains analysed in this study, it be crucial to monitor the survival of such strains in human infections in the post-vaccination era.

**METHODS**

**Specimen selection.** Faecal specimens were collected from children <5 years of age hospitalized for acute gastroenteritis in Pune, India for surveillance of rotavirus disease and strains during 2009–2012 (Chitambar et al., 2014). Three faecal specimens detected positive for rotavirus by ELISA, and were identified as having G9P[4] specificity by multiplex RT-PCR, were selected due to the higher detection rates of such strains and their availability in adequate amounts to perform full-genome sequence analysis.

**Viral RNA extraction and RT-PCR.** Viral RNA was extracted from 30% (w/v) stool suspension using Trizol, LS reagent (Invitrogen) according to the manufacturer’s instructions. RT-PCR was performed for amplification of all structural and nonstructural genes using primers described earlier (Gouvea et al., 1990; Matthijnssens et al., 2006; Chitambar et al., 2009). Briefly, extracted RNA was denatured at 95°C for 5 min, followed by snap-chilling on ice, and RT-PCR was carried out using the Qiagen One-step RT-PCR kit (Qiagen). This involved an initial reverse transcription step of 30 min at 45°C, followed by PCR activation at 95°C for 15 min, 40 cycles of amplification (1 min at 94°C, 1 min at 45/50°C, and 2.5 min at 70°C), with a final extension of 7 min at 70°C. PCR products were electrophoresed in 2% agarose gels containing ethidium bromide (0.5 µg ml−1) and visualized under a UV transilluminator.

**Determination of the 5′ and 3′ terminal sequences.** To obtain the complete nucleotide sequence of each segment, the 5′ and 3′ terminal sequences of the 11 gene segments were determined using a modified
version of the single-primer amplification method, as described previously (Matthijnssens et al., 2006). Gene-specific forward and reverse primers respectively for 5’ and 3’ end amplification of all 11 gene segments were designed for this study (Table S2).

Nucleotide sequencing and phylogenetic analyses. PCR products excised from the agarose gel were purified using a QAquick gel extraction kit (Qiagen) according to the manufacturer’s instructions, and sequences in both directions were determined using a Bigdyte Terminator cycle sequencing reaction kit V3.1 (Applied Biosystems) in an automated sequencer ABI 3130 XL (Applied Biosystems).

To determine the genetic relatedness of the three G9P[4] strains with other strains having genogroup-2 internal gene constellation, the nucleotide sequence data of nine internal gene segments of genogroup-2 rotavirus strains were retrieved from the GenBank database using appropriate keywords. A total of up to 189 strains were found to carry genogroup-2 constellation. The nucleotide sequences of individual gene segments were aligned using the MUSCLE program available in the MEGA version 6.0 software package (Tamura et al., 2013). For each gene segment, a maximum-likelihood tree was reconstructed with best-fit model of nucleotide substitution on the basis of Akaike Information Criterion (AIC), as implemented in the MEGA 6.0. The uncertainty in the tree was evaluated by bootstrap test with 1000 replicates. To classify each gene segment at the sub-genotype level, lineages and sub-lineages were defined on the basis of the observed clustering pattern with high bootstrap support and pairwise nucleotide identity.

GenBank accession numbers. The nucleotide sequences derived in this study were deposited in the GenBank database under the following accession numbers: KX536652–KX536654 (VP1), KX536655–KX536657 (VP2), KX536658–KX536660 (VP3), KX536661–KX536663 (VP4), KX536664–KX536666 (VP6), KX536667–KX536669 (VP7), KX536670–KX536672 (NSP1), KX536674–KX536676 (NSP2), KX536643–KX536645 (NSP3), KX536664–KX536668 (NSP4) and KX536649–KX536651 (NSP5).

Phylogenetic analysis of NSP4 gene. Full-length NSP4 gene sequences (n=700) of E1–E14 genotypes of RVA strains representing worldwide geographical regions were retrieved from the GenBank database. The geographic regions assigned were Africa (AFR), East Asia (EAs), South Asia (SA), South-east Asia and West Asia (SEAs), South and Central America (SCAm), North America (NAm), Oceania (OCE), East Europe (EEu), South Europe (SEu) and North and West Europe (NWEu). Initially, a distance-based tree was reconstructed, from which down-sampled datasets (n=320) were built by selection of representatives based on year of isolation, country and genotype. The best-fit model of nucleotide substitution was selected on the basis of AIC. The evolutionary model, TN93 (Tamura & Nei, 1993), with gamma distributed rates of variation among sites, was found to be most suitable for the dataset. The rates of nucleotide substitution and the divergence times between NSP4 genotypes (E1–E12) were estimated on the basis of temporal information of 320 full-length NSP4 gene sequences (~751 bp) using the Bayesian Markov Chain Monte Carlo (MCMC) approach available in the BEAST 1.7.1 package (Drummond & Rambaut, 2007). Data were analysed by employing the strict and relaxed (uncorrelated exponential and uncorrelated lognormal) clock (Drummond et al., 2006) models with Bayesian Skyline tree prior. Bayes factors and posterior probability were used to select the best model for the data (Suchard et al., 2001). The Bayes Factor analysis indicated that the uncorrelated exponential clock model fits better than the strict clock or uncorrelated lognormal clock model. The corresponding output files generated by BEAST were utilized for further analysis.

Three independent MCMC chains were each run for 70 million generations, sampling every 1000th generation. The results were combined by using the LogCombiner program available in BEAST with a burn-in of 10% generations. The convergence of the chain was evaluated using Tracer 1.5 (http://tree.bio.ed.ac.uk/software/tracer) (Drummond & Rambaut, 2007). An MCC tree was generated using the Tree Annotator program available in BEAST, and FigTree 1.2.3 (http://tree.bio.ed.ac.uk/software/figtree) was used for 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. The spatial information for the sequences was used to infer the geographic spread patterns of the virus by fitting a standard continuous-time Markov chain (CTMC) model available in BEAST (Lemey et al., 2009).

ACKNOWLEDGEMENTS

The authors thank the Council of Scientific and Industrial Research (CSIR), New Delhi, India for supporting the first author (G.P.) with a Research Fellowship, and Dr D.T. Mourya, Director and Dr A.C. Mishra, Former Director, National Institute of Virology, Pune for their constant support.

REFERENCES


http://jgv.microbiologyresearch.org


Genomic analysis of G9P[4] rotavirus strains from India


