Genomic characterization and molecular investigation of VP7 epitopes of uncommon G10P[8] group A rotavirus strains detected in Italy in 2009

Giovanni Ianiro,1 Roberto Delogu,2 Lucia Fiore2 and Franco M. Ruggeri1

Correspondence
Franco M. Ruggeri
franco.ruggeri@iss.it

1Department of Veterinary Public Health and Food Safety, Istituto Superiore di Sanità, Rome, Italy
2National Center for Immunobiologicals Research and Evaluation, Istituto Superiore di Sanità, Rome, Italy

Rotavirus strains with the uncommon genotype G10 have been detected sporadically in cases of acute gastroenteritis in humans and are thought to be transmitted zoonotically. During 2009, 10 G10P[8] rotavirus strains were detected in the stools of children hospitalized with acute diarrhoea in several paediatric hospitals in Italy. The phylogenetic analysis of the VP7 gene of the Italian G10P[8] strains analysed revealed nucleotide identities ranging from 94 to 99 %. Molecular characterization of the 11 genomic segments was performed for one of the G10 strains, which displayed a complete genomic constellation for the non-G genes. The analysis of the deduced amino acid sequences of the G10 VP7 epitopes revealed low amino acid identity with common human strains of different G genotype and with the VP7 proteins included in both anti-rotavirus commercial vaccines (Rotarix and RotaTeq). Amongst the common G genotypes, the VP7 amino acid sequence of the G10 strains showed a high similarity with sequences from G9 strains. A hydrophobic cluster analysis (HCA) of the VP7 protein including aa 20–298 was performed for the G10 Italian sequences in comparison with the major human group A rotavirus G genotypes. The HCA analysis confirmed the findings obtained previously by amino acid analysis of the VP7 epitopes, detecting a genotype-specific pattern of hydrophobicity in the hypervariable regions of the major outer capsid protein.

INTRODUCTION

Group A rotaviruses (RVAs) are a leading cause of acute gastroenteritis (AGE) in young children and animals worldwide, and are estimated to cause up to 450 000 deaths amongst children every year, mostly in developing countries (Tate et al., 2012). RVA morbidity in developed countries is also high and nearly every child has been infected at least once by the age of 5. Rotavirus remains a global public health and economic problem (Ogilvie et al., 2012; Parashar et al., 2003), although it is being replaced by norovirus as the most prevalent cause of paediatric AGE in countries that applied universal mass vaccination against RVA disease (Bucardo et al., 2014; Hemming et al., 2013; Koo et al., 2013; Payne et al., 2013). The RVA genome is composed of 11 segments of dsRNA, encoding six structural proteins (VPs) and five or six non-structural proteins (NSPs) (Estes & Cohen, 1989). The segmented nature of their genome favours reassortment between different RVA strains during co-infections, which may result in a progeny with high genomic diversity. In addition, accumulation of point mutations, due to the high error rate of the viral RNA-dependent RNA polymerase, causes further genetic diversity amongst human rotaviruses (Hanada et al., 2004; Parra et al., 2004; Ramig, 1997).

The gene sequences of the two outer capsid proteins VP7 and VP4 are used to classify rotaviruses into G and P genotypes, respectively, on a binary classification base (Matthijnssens et al., 2008). Currently, 27 G genotypes and 37 P genotypes have been described (Matthijnssens et al., 2011; Trojnar et al., 2013), and, despite the high number of possible G/P genotype combinations, ~75 % of RVA infections worldwide are caused by five common human genotypes: G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] (Gentsch et al., 2005). The remaining infections are caused by strains defined as uncommon or rare
genotypes, each occurring in <1 % of cases (Matthijnssens et al., 2011), or untypable in either G or P genes, or are mixed infections.

For a more complete rotavirus characterization, genotypes can be distinguished for all the 11 dsRNA segments, which is of particular value in RVA molecular epidemiology studies. Based on full-genome typing system, most rotaviruses identified can be grouped in two main genomic constellations: constellation 1 (Wa-like RVA strains, GX-P[8]-11-R1-C1-M1-A1-N1-T1-E1-H1) and constellation 2 (DS-1-like RVA strains, GX-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2) with one or more gene genotypes (Matthijnssens et al., 2008).

Uncommon human RVA strains are more often detected in sporadic cases of acute gastroenteritis, but they have also been reported in association with epidemic outbreaks in humans or animals (Martella et al., 2010). Amongst the uncommon RVA reported worldwide, particular attention is given to animal-derived genotypes that may become able to infect the human host, such as RVA strains displaying G6, G9 and G10 genotypes (Libonati et al., 2014; Luchs & Timenetsky, 2014; Matsushima et al., 2012; Mijatovic-Rustempasic et al., 2014; Mukherjee et al., 2012). Some of these are candidates to become major human genotypes, as was previously observed for the G9 genotype (Gentsch et al., 2009; Iturriza-Gomara et al., 2011; Matthijnssens et al., 2010; Santos & Hoshino, 2005), and might represent a challenge for the efficacy of current anti-rotavirus vaccines, posing a need for RVA strain surveillance programs (Banerjee et al., 2006; Iturriza-Gomara et al., 2011).

During the 2009 rotavirus molecular surveillance season in Italy, 10 RVA strains were genotyped as G10P[8] by both reverse transcription-nested PCR (RT-nPCR) and nucleotide sequencing. These G10 strains were collected as sporadic cases in several Italian hospitals, represented the first detection in Italy and were not further detected during the subsequent 5 years of national surveillance. In the last decade, sporadic infections with G10 RVA in either humans or animals (mostly in bovines) have been repeatedly reported worldwide, involving strains that often presented uncommon or animal-derived VP4 P genotypes, such as P[5], P[11], P[14] and P[15] (Badaracco et al., 2013; Cowley et al., 2013; Komoto et al., 2014; Libonati et al., 2014; Paul et al., 2014; Rajendran & Kang, 2014; Steyer et al., 2010). Nonetheless, only limited molecular data are as yet available on human G10 RVA strains that show a complete Wa-like genomic backbone (Esona et al., 2011; Esona et al., 2010; Matsushima et al., 2012).

The present study reports the phylogenetic analysis of the Italian G10 strains and the molecular characterization of their VP7 at the amino acid level. A complete genomic characterization of strain RVA/Human-wt/ITA/RC37/2009/G10P[8] is also presented in order to investigate the molecular epidemiology and origin of this uncommon genotype in more detail.

RESULTS

Detection of G10P[8] rotaviruses

In total, 972 RVA strains were identified in 2009 from children with acute diarrhoea admitted to hospital in Italy (Iturriza-Gomara et al., 2011). As determined by genotyping RT-nPCR and nucleotide sequencing, 31 of the strains (3.2 %) presented an uncommon genotype for at least one of the two capsid protein genes, and 10 of these were characterized to be G10P[8]. The 10 strains were isolated from patients admitted in six different cities throughout Italy, and were identified between January and July 2009, with no apparent epidemiological link amongst cases. Patients presented no difference in clinical severity or specific symptoms with respect to the other subjects investigated.

Phylogenetic analysis of VP7

The phylogenetic analysis of the VP7 (Fig. 1) gene of the Italian G10P[8] RVA strains revealed that they belonged to the G10 genotype, showing nucleotide identities that ranged from 94 to 99 % between them. The phylogenetic tree showed that the Italian strains clustered with either RVA strains detected in Africa, in which the G10 genotype was mostly in combination with a human-like P[8] VP4, or with RVA strains detected in Slovenia, Turkey and Australia, in which G10 was combined with animal-like P[14] or P[15] VP4 genes.

Bayesian analysis of G10 VP7 genes

The G10 RVA strains analysed in Fig. 1 were also subjected to a Bayesian Markov chain Monte Carlo (MCMC) analysis performed in BEAST (Fig. 2), averaging over all plausible trees to account for the phylogenetic error. The phylogenetic tree revealed the same clustering pattern as shown in the nucleotide phylogenetic tree (Fig. 1). The G10 genotype seemed to be generated in Asia in the early 1990s, revealing the separation in two different evolution lines in 1999–2000, reported sporadically in South East Asia, Africa and Europe. The Italian G10 RVA strains belonged to the evolution line also including the African strains since the early 2000s. Within this group, the Italian strains presented varying levels of genetic diversity with global strains reported through the years. Strains CHIER1, PG2 and JES3 were closer to strains detected in Eastern Europe since 2003, whereas strains BERG4 and CHIER2 were related more strictly with the G10 genotype detected in the Ivory Coast in 2005. Although they also shared a possible African origin, strains PG1, BERG14, BERG27, CAMP44 and RC37 were not as strictly correlated with this same group.
Complete genomic characterization of strain RC37

Amongst the Italian G10 RVA strains, the sample containing strain RC37 was still available for further nucleotide sequencing, allowing us to investigate the rest of its genome. After sequencing, a complete genomic constellation 1 (Wa-like, P[8]-I1-C1-M1-A1-N1-T1-E1-H1) was obtained.

The phylogenetic analysis of VP1–4, VP6 and NSP1–5 genes (Fig. 3) of strain RC37 showed high nucleotide identities with common Wa-like human strains detected worldwide, but not with other G10 strains.

Analysis of the VP7 hypervariable regions of G10 RVA strains

The deduced amino acid sequences of the G10 VP7 belonging to the Italian strains investigated were compared with those belonging to the human major RVA genotypes, in the VP7 hypervariable regions A, B, C, E and F (Fig. 4) (Aoki et al., 2009, 2011). The comparison investigated a total of 52 aa residues and showed the highest similarity between G10 amino acid sequences and genotypes G9 (39/52 aa) and G3 (38/52 aa). Conversely, the G10 sequences revealed the highest divergence with VP7 sequences corresponding to either G2 (25/52 aa) or G12 (26/52 aa) RVA genotypes.
The VP7 amino acid sequences of the Rotarix (G1) and RotaTeq (G1–4) anti-rotavirus vaccines were also included in the analysis, showing differences between the G10 and the vaccine strains in all the regions investigated, particularly in hypervariable regions A and C (Fig. 4).

The amino acid comparison (Fig. 4) revealed that the hypervariable region A deduced from the Italian G10 RVA strains showed the highest amino acid identity with that belonging to the G3 genotype (4 aa substitutions). In all other regions analysed (B, C, E and F), the best match of the G10 deduced amino acid sequence was with the G9 genotype, which altogether presented only 7 aa substitutions. The analysis of the G10 VP7 hypervariable regions B and C showed the highest amino acid identities with the G6 and G9 genotypes, and with G3 and G9 RVA VP7 genotypes, respectively, identifying 2 or 3 aa substitutions.

High amino acid identity was shown in the hypervariable region F of VP7 between G10 strains and genotypes G3, G6, and G9.
**Fig. 3.** Phylogenetic trees based on the partial ORF of genes coding for: VP1 (nt 74–813), VP2 (nt 205–802), VP3 (nt 50–632), VP4 (trypsin-cleavage fragment VP8*) (nt 28–875), VP6 (nt 238–1283), NSP1 (nt 66–841), NSP2 (nt 47–1000), NSP3 (nt 38–1035), NSP4 (nt 48–730) and NSP5 (nt 18–615). Italian G10P[8] strain RC37/2009 is marked with a filled circle in each tree. Trees were reconstructed with the maximum-likelihood method (Tamura-3 parameter) and bootstrapped with 1000 repetitions; bootstrap values <70 are not shown. Bar, 0.02 nucleotide substitutions per site.
G9 and G4, but not G1 and G2 genotypes. The bovine-derived genotypes G6 and G8 showed a moderate amino acid identity with G10, revealing a low substitution rate particularly in regions B and F.

**Hydrophobic cluster analysis (HCA)**

The deduced amino acid sequences of the Italian G10 strain BERG4 and RVA strains belonging to the major human genotypes were used for a more detailed study encompassing aa 20–298. The comparison was conducted using HCA to investigate the impact of amino acid substitutions in the hydrophobic pattern of the VP7 protein by RVA G genotype (Fig. 5, Table 1). The VP7 hypervariable regions analysed in Fig. 4 are identified within vertical black frames with the corresponding letters.

The analysis of region A revealed a unique HCA pattern for the G10 genotype, when compared with those of the G1–4 (Fig. 5a), G6, G8, G9 and G12 genotypes (Fig. 5b).

The amino acid similarity value between the G10 VP7 and other RVA genotypes closely paralleled the calculated HCA similarity scores, with their ratios being between 0.929 and 0.981(Table 1).

**DISCUSSION**

This study reports the detection and the molecular analysis of G10P[8] RVA strains detected in Italy in 2009. These viruses were not detected in Italy by the national rotavirus surveillance network RotaNet-Italy during the period 2006–2009 (Iturriza-Gómar et al., 2011) and were not further identified in the following 5 years of surveillance (unpublished). The spread of these viruses cannot be ascribed to an epidemic outbreak, as the 10 cases identified had no epidemiological link or possible common source of infection and cases were dispersed through the entire country. In addition, the phylogenetic analysis indicated that the 10 strains were not identical, suggesting independent spread and evolution in different places.

Eventually, none of the G10 strain was successful in establishing an epidemic spreading in the population of the six cities in either 2009 or in the following years.

The phylodynamic tree suggests that the G10 genotype was likely generated in Asia in the early 1990s, separating in two different evolution lines in 1999–2000. The Italian G10 strains apparently belong to the G10 line evolving in Africa since the early 2000s, within which part of them were strictly correlated with strains detected in Eastern Europe since 2003, whilst others were closer to the G10 genotype that originated in 2005 in the Ivory Coast.

Although point mutations have likely driven the evolution of the strains investigated in Italy in 2009, based on the analysis and characterization of all 11 genome segments of strain RC37, a possible role of gene reassortment in its origin is also suggested by the identification of a complete Wa-like human backbone (genotype 1) in addition to the animal-derived VP7 genotype G10.

However, the possible occurrence of additional reassortment events in the evolution of strain RC37 may not be ruled out, as the individual genes of this strain show
Fig. 5. HCA of VP7 deduced amino acid sequences of different human RVA genotypes (aa 20–298). The software developed for HCA creates a two-dimensional plot of the protein, based on its amino acid sequence. The plot is presented on a cylinder with 3.6 aa per turn shown as a classical α-helix, and is separated along its axis and unrolled. (a) Comparison between G10 and G1–4. (b) Comparison between G10 and G6, 8, 9 and 12. Residues are coloured following the strength of hydrophobicity: strong VILF (green), medium WMY (green) and mimetic AC (black). Clusters of hydrophobic residues are identified by sets of green letters within a black contour. Hydrophilic residues are highlighted by colours: DENQ (red) and HKR (blue). Special amino acids are marked as follows: P (red stars), G (black diamonds), T (white squares) and S (black empty squares with dot). The vertical black frames correspond to the VP7 hypervariable regions also shown in Fig. 4.
closer similarity with different human Wa-like strains reported in different countries.

Considering the antigenic sites recognized on VP7, the G10 strains shared a similar amino acid sequence with G3 genotype RVA in hypervariable region A, and with G9 viruses in the remaining VP7 regions B, C, E and F, where only 7 aa substitutions were observed on a total of 37 residues considered.

It is unclear whether the G10 VP7 genotype might possess the characteristics to become a major human genotype, but it is interesting to note that this gene was integrated in strains showing an otherwise complete genomic constellation 1, which represents the best RVA genomic backbone fitting with the human host (Matthijnssens & Van Ranst, 2012).

In the absence of a crystallized form of a G10 VP7, the fine three-dimensional structure of this protein remains to be disclosed, but HCA has proven helpful to understand the spatial relationship of amino acids for proteins and enzymes (Callebaut et al., 1997; Gaboriaud et al., 1987). Here, the HCA was performed to obtain more detailed information on the major RVA outer capsid protein VP7, combining results of the hydrophobic pattern and amino acid similarity/difference analyses.

The hydrophobic cluster analysis of the G10 VP7 protein fragments revealed genotype-specific differences compared with cogent fragments of other G genotypes and were closely correlated with the amino acid similarity scores (Table 1). Amongst others, both VP7 HCA analysis and amino acid sequence comparison revealed variations with the VP7 of other G genotypes also at aa 99 and 208, which have been indicated to be within or very close to major neutralization epitopes of both human and animal rotaviruses (Mackow et al., 1988; Taniguchi et al., 1988).

The Italian G10P[8] RVA strains identified in 2009 were apparently unable to spread further and persist in the human population in the same or other areas of Italy during the following 5 years of rotavirus surveillance. Nonetheless, this paper indicates that uncommon rotaviruses such as G10 may sporadically emerge across a country in the absence of rotavirus mass vaccination, and advance into different evolution lines by mutation and possibly by reassortment with other local human strains.

Continuous monitoring programmes of clinical RVA gastroenteritis, together with surveillance of adults, animals and environment, may help identify possible emerging viral strains of possible public health interest.

### METHODS

#### Cases and sample collection. 
Stool specimens were collected in 2009 from 10 children admitted with acute gastroenteritis to public hospitals of six different cities throughout Italy, in the framework of the molecular surveillance activities of RotaNet-Italia (Iturriza-Gómez et al., 2011).

RNA was extracted from 140 µl 10 % faecal suspensions in H2O using a Viral RNeasy Mini kit (Qiagen), according to the manufacturer’s instructions. Final elution from the silica membrane spin column was performed in 60 µl RNase-free water and RNA was stored at −80 °C.

#### RVA genome segments amplification and sequencing. 
RVA G and P genotyping for all samples was accomplished by RT-nPCR as described previously (Gentsch et al., 1992; Iturriza-Gómez et al., 2004). Genotype assignment was first performed based on the molecular size of amplified DNA according to EuroRotaNet shared protocols (http://www.eurorota.net/), and was confirmed by nucleotide sequencing (see below) and the RotaC genotyping tool (Maes et al., 2009). Nucleotide sequencing of the genes amplified was performed at the Macrogen, with the same primers used as for PCR, applying the BigDye chemistry.

RT-PCR and sequencing were performed for each of the 11 RNA segments using specific primers (Esona et al., 2009; Matthijnssens et al., 2008).

#### Software analysis and phylogenesis. 
Sequences generated were assembled to obtain a single consensus sequence. The resulting sequencing files were analysed and corrected with ChromasPro2.23 (Technelysium). Contig assembly was performed with SeqMan II (DNASTAR). Nucleotide and amino acid sequence similarity searches were performed using the BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignments and phylogenetic tree construction were performed with MEGA5 (Tamura et al., 2011), applying the maximum-likelihood and using substitution models for each tree, as suggested by MEGA5 ModelTest.

Bayesian phylogenetic reconstructions were performed using MCMC analysis implemented in BEAST (Drummond & Rambaut, 2007) molecular clock models to estimate rooted, time-measured phylogenetic trees with a coalescent prior. The topology of the G10 strains were analysed using an HKY nucleotide substitution model with a G1-distributed rate variation, a log-normal relaxed clock model (Drummond et al., 2006) and a flexible Bayesian skyline tree prior (Drummond et al., 2005). Four independent MCMC analyses were

<table>
<thead>
<tr>
<th>Strain</th>
<th>G genotype</th>
<th>Amino acid identity (%)</th>
<th>HCA similarity score (%)</th>
<th>Amino acid identity/HCA similarity score ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wa</td>
<td>G1</td>
<td>75.9</td>
<td>77.5</td>
<td>0.979</td>
</tr>
<tr>
<td>DS-1</td>
<td>G2</td>
<td>73.1</td>
<td>78.7</td>
<td>0.929</td>
</tr>
<tr>
<td>AU-1</td>
<td>G3</td>
<td>82.7</td>
<td>84.8</td>
<td>0.975</td>
</tr>
<tr>
<td>Hochi</td>
<td>G4</td>
<td>75.2</td>
<td>76.8</td>
<td>0.979</td>
</tr>
<tr>
<td>CEC06</td>
<td>G6</td>
<td>82.0</td>
<td>85.1</td>
<td>0.964</td>
</tr>
<tr>
<td>DRC88</td>
<td>G8</td>
<td>80.2</td>
<td>82.8</td>
<td>0.969</td>
</tr>
<tr>
<td>MMC38</td>
<td>G9</td>
<td>83.4</td>
<td>85.0</td>
<td>0.981</td>
</tr>
<tr>
<td>VU080939</td>
<td>G12</td>
<td>78.0</td>
<td>81.3</td>
<td>0.959</td>
</tr>
</tbody>
</table>

*Genotypes G5, G7, G11 and other rare G genotypes infecting mainly animals were excluded.

†Per cent values are calculated versus the amino acid sequence of G10 strain BERG4.
run for 100 million generations and diagnosed using Tracer (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility trees were annotated using TreeAnnotator and visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

**HCA of VP7.** HCA was performed on the VP7 protein, based on the deduced amino acid sequences of RVA strains belonging to the major human genotypes (aa 20–298). For this analysis, the G10 genotype was represented by the consensus amino acid sequence generated from the Italian G10 RVAs involved in this study; the comparison was conducted using the same reference sequences for each genotype included in Fig. 4.

The HCA was performed using the online tool at Mobyle@RPBS (http://bioserv.rpbs.univ-paris-diderot.fr/services/HCA/) (Callebaut et al., 1997; Gaboriaud et al., 1987).

The software developed for HCA creates a two-dimensional plot of the protein, based on the amino acid sequence. The plot is presented on a cylinder with 3.6 aa per turn shown as a classical α-helix, and is separated along its axis and unrolled. The HCA plot was defined by encircling the hydrophobic residues, marking the proline and glycine residues as presenting loops, and the cysteine as involving disulphide bonds.

HCA similarity scores (Table 1) were obtained by the following formula: HCA similarity score (%) = (2CR × 100)/(RC1 + RC2), where RC1 and RC2 are the numbers of hydrophobic residues in protein 1 and 2, respectively, and CR is the number of hydrophobic residues conserved between the compared sequences (Xiang et al., 2014).

**ACKNOWLEDGEMENTS**

We acknowledge the RotaNet-Italy collaborators for detection of rotavirus strains. The full list of members is present at: http://www.iss.it/criv/index.php?lang=it&id=363&tipo=9. This study was supported by grants from the Ministry of Health, Italy (Italy/USA ‘Investigating the evolution of zoonotic norovirus and rotavirus strains from swine’, 2011) and by EuroRotaNet (http://www.eurorota.net).

**REFERENCES**


Iturriza-Gómez, M., Dallman, T., Bányai, K., Böttiger, B., Buesa, J., Diedrich, S., Fiore, L., Johansen, K., Koopmans, M. & other


