Genetic linkage of capsid protein-encoding RNA segments in group A equine rotaviruses

Samuel Miño,1† María Barrandeguy,1,2 Viviana Parreño1 and Gabriel I. Parra3‡

1Instituto de Virología, CICVyA, INTA Castelar, Nicolás Repetto y De los Reseros s/n (CP 1816), Hurlingham, Buenos Aires, Argentina
2Escuela de Veterinaria, Universidad del Salvador, Champagnat 1599, Ruta Panamericana km54.5 (B1630AHU), Pilar, Buenos Aires, Argentina
3Departamento de Biología Molecular y Genética, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunciô­n, Asunciô­n, Paraguay

Correspondence
Samuel Miño†
mino.samuel@inta.gob.ar
Gabriel I. Parra‡
gabriel_parra@hotmail.com

Received 1 September 2015
Accepted 8 January 2016

INTRODUCTION

Rotaviruses are major causative agents of severe gastroenteritis in young children and animals (Estes et al., 2001). Equine rotavirus strains were first detected in diarrhoeic foals in England in 1975 (Flewett et al., 1975) and are currently considered a major cause of dehydrating diarrhoea in young foals (Magdesian et al., 2014).

Rotavirus virions are formed by three concentric protein layers that enclose the 11 dsRNA genome segments. Each of those segments encodes one protein, except for segment 11 that encodes two non-structural proteins. The outermost capsid layer is composed of the glycoprotein VP7 and the spike protein VP4. The intermediate layer is composed exclusively of VP6 and the inner layer of the rotavirus virion, also referred to as the core shell, is composed of VP2. Complexes containing the RNA-dependent RNA polymerase (VP1), the capping enzyme (VP3) and the dsRNA segments are located beneath VP2 (Estes & Greenberg, 2013).

The most abundant protein in the virion is VP6. Structural analyses of VP6 showed that is folded into two distinct domains: the B domain (aa 1–150 and 335–397), which corresponds to the base of the molecule, and the H domain (aa 151–334), which is at the top of the molecule. Both domains, B and H, participate in the contacts that stabilize the roughly pear-shaped trimer. The VP6 trimeric structure forms a channel that is involved in transporting the mRNA during the transcriptional process. The B domain of the trimer has different points of interaction with VP2, whilst the H domain interacts with VP7 and VP4 (Charpilienne et al., 2002; Li et al., 2009; Mathieu et al., 2001). The VP7 protein also forms an array of trimers and its assembly locks VP4 into place. Thus, the anchored VP4 protrudes above the surface of the VP7 layer to interact with the host cells during the attachment of the virus. Proteolytic cleavage of the spike protein VP4 into VP5* and VP8* is required for rotavirus infectivity, with the VP8* domain being involved in binding to cellular receptors (Aoki et al., 2009; Chen et al., 2009; Li et al., 2009; Trask et al., 2012).

Interactions between VP6, VP7 and VP4 have been described and the strength of the interaction is important for the fitness...
### VP6 and VP7 genotype distribution

<table>
<thead>
<tr>
<th></th>
<th>G3</th>
<th>G14</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2</td>
<td>1 (3%)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>I6</td>
<td>30 (97%)</td>
<td>–</td>
</tr>
</tbody>
</table>
of the virion (Heiman et al., 2008). The outer capsid assembly appears to be a step-wise process in which VP4 is added to an assembled double-layered particle, made by VP2 and VP6, and then VP7 takes its place and locks VP4 into the triple-layered particle (Li et al., 2009; Trask & Dormitzer, 2006). The VP6:VP7 interaction sites were described as well as the interactions of VP6 and VP7 with VP4 (Settembre et al., 2011).

Antigenic and genetic differences within the VP6 protein have been used to classify rotaviruses into nine groups (A–I, also termed species) (Estes & Greenberg, 2013; Mihalov-Kovács et al., 2015), with group A rotaviruses (RVA) being the most prevalent in humans and animals. Based on differences within RVA, a number of genotypes have been described for the VP4 (P[1]–P[37]), VP6 (I1–I16) and VP7 (G1–G27) gene (Estes & Greenberg, 2013; Matthijnssens et al., 2012; Trojnar et al., 2013).

Although RVA strains with unusual G/P combinations (such as G3P[3], G5P[7], G6P[1], G8P[1], G10P[1], G10P[11] and G13P[18]) have been detected occasionally in horses (Browning et al., 1991; Ciarlet et al., 2003; Isa et al., 1996; Miño et al., 2013; Taniguchi et al., 1994), G3P[12] and G14P[12] are the most prevalent strains associated with diarrhoea in foals worldwide. Rotavirus strains carrying the G3 genotype have been found in a wide range of host species (Estes & Greenberg, 2013; Martella et al., 2010), but the equine-like G3, G14 and P[12] genotypes have never been found producing outbreaks in any other host species than equines (Bailey et al., 2013; Garaicoechea et al., 2011; Matthijnssens et al., 2015; Weinberg et al., 2013). Genomes of equine RVA G3P[12] and G14P[12] are highly conserved, and only three gene segments have shown different genotypes (i.e. VP7: G3 and G14; VP6: I2 and I6; and NSP4: E2 and E12) in all complete genomes available currently (Ma et al., 2015; Matthijnssens et al., 2012; Nemoto et al., 2015).

Although each of the 11 gene segments can theoretically segregate independently during strain reassortment, a genetic linkage between VP6, VP7, VP4 and NSP4 has been reported for certain RVA strains (Martella et al., 2010; McDonald et al., 2009; Song & Hao, 2009). Thus, viral protein interactions might drive the selection of strains that can replicate and spread better in different species (Iturriza-Góñara et al., 2003; Santos et al., 2003). In this study we characterized the VP4 (VP8*), VP6 and VP7 genes from a large sample collection of equine RVA strains. Using genomic and structural analyses, we described the genetic linkage of variants of VP6 and VP7 carried by the RVA strains that infect equines.

RESULTS

Characterization of Argentinean equine RVA strains

The VP6, VP7 and VP8* sequences from 30 equine RVAs were obtained, and the VP6 and VP7 sequences from another 13 equine strains available in the public databases were retrieved and included in the analyses (Table S1, available in the online Supplementary Material). Phylogenetic analyses of the VP7 gene showed that all Argentinean G14P[12] strains clustered closely together, showing a common geographical origin (Fig. 1a). All Argentinean G3P[12] strains clustered together with other equine G3 strains (Fig. 1b) and apart from the G3 strains from other species, including humans. Phylogenetic analyses of the VP6 gene showed that all G14 equine strains presented the I2 genotype (Fig. 1c), whilst most (34/35) of the G3 equine strains presented the I6 genotype (Fig. 1d). The only exception was the BI strain that was isolated in Japan in 1981 from a diarrhoeic foal and isolated by tissue culture in MA104 cells (Table S1). Thus, these data suggest a genetic linkage between the RNA segments encoding VP7 and VP6 from equine RVA strains (Fig. 1e).

VP6 protein analysis

Comparison between strains bearing the I2 VP6 genotype and the I6 VP6 genotype revealed 38 amino acid differences. All strains carrying the I6 genotype showed the same conserved amino acid sequence. Equine strains carrying the I2 genotype presented a conserved amino acid sequence showing differences in only three residues (L11F, M295I and Q300P). Of note is the presence of an insertion of two amino acid (aa 299 and 300) when I6 and I2 sequences were compared (Table 1).

A modelled VP6 was obtained using I2 and I6 sequences, and fitted to the trimer structure by using the solved VP6 [Protein Data Bank (PDB) IDs: 3N09 and 4V7Q] (Fig. 2a). Most of the changes mapped to the surface of the VP6 protein (Fig. 2c), with three amino acid changes (aa 151, 340 and 341) on the surface of the monomer, but within the intra-trimer contact region (data not shown). Of note is that most of the amino acid changes (21/38) mapped to the top of the H domain of the protein, in or near to the VP6 antigenic sites (Fig. 2b, c). Five VP6: VP7 contact regions were previously defined for the VP6 protein (A, 170–175; B, 240–245; C, 294–300; D, 304–308; and E, 310–315) (Fig. 2b) (Li et al., 2009;
Eight out of the 38 I2–I6 differences mapped to these regions, including the insertion of two amino acids present in I2 strains (Fig. 2c). When analysing the VP6 region that interacts with VP2 (aa 31–39 and 64–76) (Charpilienne et al., 2002; McClain et al., 2010), only two substitutions (I38V and V39I) were found (Fig. 2c). The two-amino-acid insertion (aa 299 and 300) in the I2 sequences when compared with I6 sequences was located in a loop of region D (Fig. 2d).

VP7 protein analyses

Comparison of the VP7 sequences between G3 and G14 equine strains revealed 33 amino acid substitutions (Table 2). In concordance with cross-neutralization assays (Ciarlet et al., 1994), 17 of those substitutions mapped in the VP7 antigenic sites (Table 2). All differences in the antigenic sites as well as three amino acid substitutions (aa 118, 125 and 129) were located in the outer face of the trimer and accessible to antibodies (Fig. 3a, b). Of note is that eight amino acid substitutions (aa 68, 74, 75, 267, 268, 281, 282 and 303) amongst the G3 and G14 strains were located on the surface of the internal face of the VP7 trimer, which is in close contact with the H domain of VP6 (Fig. 3a, b).

VP6: VP7 interaction sites

In addition to the five regions of VP6 shown to be in contact with VP7 (Fig. 2b), three specific VP6: VP7 interaction sites have been fine-mapped (Chen et al., 2009; Li et al., 2009; Mathieu et al., 2001). Accordingly, these residues presented conservative substitutions between G3 and G14 equine strains (Tables 1 and 2), with the exception of the G14 strain FI-23 that presented four amino acid differences (V237I, E238N, D267N and A282E) compared with the other G14 strains (data not shown). Although these eight specific residues showed conservative substitutions, there were another 11 (V151T, F199I, A217V, I225L, I253V, L254F, I261V, M295I, Q312N, D329E and V332I) conservative substitutions between G3 and G14 strains that mapped to the surface of the top of the VP6 H domain, and six (F118L, A125T, V129I, V260I, D267N and I307V) that mapped to the inner surface of the VP7. Therefore, we explored whether these substitutions play a role in the overall hydrophobicity or the electrostatic potential of the modelled structures. We noticed that the electrostatic potential of the VP6 structures of I2 and I6 mostly differed at the top of the VP6 H domain, and six (F118L, A125T, V129I, V260I, D267N and I307V) that mapped to the inner surface of the VP7. Therefore, we explored whether these substitutions play a role in the overall hydrophobicity or the electrostatic potential of the modelled structures. We noticed that the electrostatic potential of the VP6 structures of I2 and I6 mostly differed at the top of the VP6 trimer, where VP6 interacted with the VP7 protein (Fig. 4a), and the modelled VP7 proteins from G3 and G14 strains mostly differed in the inner face of the trimer, which was in close contact with the H domain of VP6 (Fig. 4b).

DISCUSSION

Although rotaviruses have the potential of creating new viruses by reassortment of their segmented genomes, large-scale analyses of natural strains have shown that
Fig. 2. Structural analyses of equine VP6. (a) Representation of VP6:VP7 heterohexamer. The VP7 trimer is shown in yellow and the VP6 trimer is shown in grey. (b) Ribbon diagram of the VP6 subunit with the VP6:VP7 interaction sites shown in pink (A–E) and the VP6:VP2 interaction sites shown in green. (c) Representation of the VP6 trimer with the I2/I6 differences shown in red. (d) Ribbon diagram of superimposed modelled VP6 I2/I6 proteins (grey, I2; light blue, I6). The two-amino-acid insertion (P299 and P300) is shown in red. Inset: zoom of the loop containing the two-amino-acid insertion.
rotaviruses tend to present constellations of genes that remain stable over time (Heiman et al., 2008; Matthijnssens et al., 2012; McDonald et al., 2009). Human RVA can be divided into three major genome constellations (GI–GIII), being the Wa strain as the prototype for GI, DS-1 for GII and AU-1 for GIII. This strong association of genes (genome constellations) has been explained by specific interactions of the viral proteins (Benati et al., 2010; Heiman et al., 2008; Iturriza-Gomara et al., 2003; McDonald et al., 2009). Thus, genes from one specific genome constellation appear to co-segregate, restricting reassortment amongst different genotypes. Equine RVAs have also been shown to present a high level of conservation in their genome constellations, with the VP7, VP6 and NSP4 gene segments being distinguished into two different genotypes, i.e. G3/G14, I2/I6 and E2/E12, respectively (Ghosh et al., 2013; Matthijnssens et al., 2012; Nemoto et al., 2015). Importantly, only one P type (P[12]) has been found in RVAs infecting horses, which indicates a high restriction at the species level.

It was shown for human and animal RVAs that the VP6 protein plays a major role in restricting the reassortment of genes because it forms the intermediate capsid layer, and interacts with both the outer capsid (VP4 and VP7) and the inner capsid (VP2) (Heiman et al., 2008). All the equine RVA strains sequenced to date present the same genotype for the VP2 gene (C2) (Ghosh et al., 2013; Matthijnssens et al., 2012; Nemoto et al., 2015). As only two amino acid substitutions (I38V and V39I) were found in the VP6 : VP2 interacting region, which seem to compensate for each other (Ile for Val and vice versa), no major differences can be expected with the VP2 interaction, and therefore the same VP2 protein could be shared by the equine strains showing G3P[12] and G14P[12] genotypes. However, our analyses support a strong and specific association of the VP6 : VP7 proteins from the equine RVA strains. A strong correlation of genotypes G3 and G14 segregating with I6 and I2, respectively, was detected, suggesting that both proteins which functionally interact with one another co-evolve in equine RVA. Inspection of the amino acid sequence and structural analyses using modelled VP6 proteins showed that many differences between each of the two VP6 genotypes mapped in the top domain of VP6 (H domain), with 21 substitutions mapping at or near defined VP6 antigenic sites. These amino acid substitutions could result in antigenic differences between I6 and I2 genotypes, and have important implications in RVA diagnostics when using VP6 as target antigen (Miño et al., 2015), as well as in defining structural restrictions whilst interacting with the VP7 protein. In this regard, the VP6 : VP7 contact region has been defined (Chen et al., 2009; Li et al., 2009; Mathieu et al., 2001), and a co-evolution of the residues involved in this interaction amongst the G3-I6 and G14-I2 equine strains was detected (Tables 1 and 2). Although the two-residue insertion in one of the loops of the VP6 protein from G14P[12] strains could be involved
in defining the specificity of VP6 : VP7 interacting sites, the current data suggest that other specific substitutions might also be involved in the functional interaction and co-evolution of these two proteins. Further experiments, including isolation and characterization of reassortant viruses rescued from equine G3 and G14 mixed infections, or experimental information on the structure of these equine proteins will be needed to confirm these observations.

Viral protein interactions play a major role in the co-evolution of genes in segmented viruses (Dugan et al., 2008; Graham et al., 1987; Heiman et al., 2008; Lubeck et al., 1979; Nibert et al., 1996; Trask et al., 2012) which restricts the generation of viable viral particles and the ability to infect different species. Reassortants of human and bovine RVA strains experimentally rescued from mixed infections in different cell lines have shown that both genetic determinants and host factors contribute to the outcome of reassortment events (Graham et al., 1987). Our work provides strong evidence for the genetic linkage between the VP6 and VP7 genes from equine RVAs, which could be related to protein–protein interactions that define the fitness of the rotavirus virions, restricting the reassortment of strains whilst infecting the same host.

**METHODS**

**Samples and sequences.** A total of 30 stool samples from young foals (aged 1 day to 10 months old) with acute diarrhoea and positive for RVA were used in this study. Samples were collected in Argentina between 2009 and 2011. In addition, the sequences of three older RVA strains isolated from horses in Argentina in 1993, 2006 and 2008 were used. RNA extraction, reverse transcription PCR and gene segment sequencing (VP6 nt 24–1224, VP7 nt 49–1029 and VP8* nt 10–861) were carried out as described previously (Miño et al., 2015) (Table S1).

**Phylogenetic analyses.** Datasets were constructed with 43 equine RVA strains (Table S1). Multiple sequence alignments were performed using CLUSTAL_X 2 (Larkin et al., 2007) and manually edited in BioEdit (Hall, 1999). Phylogenetic and molecular evolutionary analyses were conducted using MEGA6 (Tamura et al., 2011). Genetic distances were calculated using the Tamura three-parameter + G (gamma distribution) model at the nucleotide level and the phylogenetic trees were reconstructed using the maximum-likelihood

---

**Fig. 3.** Representation of the modelled equine VP7 trimer: (a) side view and (b) top view. The monomers of VP7 are shown in yellow, golden and light yellow. The G3/G14 differences are shown in red.
algorithm and SPR as the heuristic method with gamma distribution (G=5), and 1000 bootstrap replicates as statistical support.

**Structural modelling and analyses.** The molecular modelling was carried out using the I-TASSER online server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010), without customizing any template model. The VP6 and VP7 amino acid sequences of E30/1993/G3P[12] and E4040/2008/G14P[12] strains were used for structural modelling. The C-score, which is a confidence score provided by I-TASSER and ranges from −5 to 2 (where higher values signify higher confidence), was 2.00 for the two VP6 models and 0.57 for the two VP7 models. Note that modelling was based on experimentally solved structures from RVA strains (bovine UK strain and rhesus RRV strain) that present high similarity (≥90%) with the

---

**Fig. 4.** Representation of the modelled equine (a) VP6 and (b) VP7 trimers showing the electrostatic potential of the surfaces. Positive charges are indicated by blue and negative charges are indicated by red. Scale, kcal mol$^{-1}$. 

---

http://jgv.microbiologyresearch.org 919
proteins from the equine strains used in this study (Table S2). VP6 and VP7 trimers were reconstructed using the atomic model of the RRV strain (G3P[3]; PDB ID: 4V7Q) and the UCSF Chimera package (Petterson et al., 2004). Electrostatic potentials of the surface were calculated using the Coulombic Surface Coloring option with default settings as implemented in UCSF Chimera. Electric charges for each atom were assigned using AMBER ff14SB. Molecular visualization and graphing were done using UCSF Chimera.

ACKNOWLEDGEMENTS

This work was supported by CVT INTA, Haras Agreement, Argentina. V. P. is a member of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

REFERENCES


