Molecular characterisation of the NSP4 gene of group A human rotavirus G2P[4] strains circulating in São Paulo, Brazil, from 1994 and 2006 to 2010

Jéssica Wildgrube Bertol1/+, Maria Clara Duarte Fregolente1,2, Thabata Alessandra Ramos Caruzo1, Márcio José da Silva3, Veridiana Munford4, Marco Aurélio Palazzi Sáfadi5, Maria Lucia Rácz4, Maria Silvia Viccari Gatti1

1Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Genética, Evolução e Bioagentes, Campinas, SP, Brasil 2Universidade da Cidade de São Paulo, São Paulo, SP, Brasil 3Universidade Estadual de Campinas, Centro de Biologia Molecular e Engenharia Genética, Campinas, SP, Brasil 4Universidade de São Paulo, Instituto de Ciências Biomédicas, Departamento de Microbiologia, São Paulo, SP, Brasil 5Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo, SP, Brasil

Group A human rotaviruses (HuRVA) are causative agents of acute gastroenteritis. Six viral structural proteins (VPs) and six nonstructural proteins (NSPs) are produced in RV-infected cells. NSP4 is a diarrhoea-inducing viral enterotoxin and NSP4 gene analysis revealed at least 15 (E1-E15) genotypes. This study analysed the NSP4 genetic diversity of HuRVA G2P[4] strains collected in the state of São Paulo (SP) from 1994 and 2006-2010 using reverse transcription-polymerase chain reaction, sequencing and phylogenetic analysis. Forty (97.6%) G2P[4] strains displayed genotype E2; one strain (2.4%) displayed genotype E1. These results are consistent with the proposed linkage between VP4/VP7 (G2P[4]) and the NSP4 (E2) genotype of HuRVA. NSP4 phylogenetic analysis showed distinct clusters, with grouping of most strains by their genotype and collection year, and most strains from SP were clustered together with strains from other Brazilian states. A deduced amino acid sequence alignment for E2 showed many variations in the C-terminal region, including the VP4-binding domain. Considering the ability of NSP4 to generate host immunity, monitoring NSP4 variations, along with those in the VP4 or VP7 protein, is important for evaluating the circulation and pathogenesis of RV. Finally, the presence of one G2P[4]E1 strain reinforces the idea that new genotype combinations emerge through reassortment and independent segregation.

Key words: human rotavirus - NSP4 genotypes - phylogenetic analysis - VP4/VP7/NSP4 linkage - Brazil

Rotaviruses (RVs) are the most important etiological agents of severe acute gastroenteritis in children younger than five years of age (WHO 2013) and they cause an estimated 1.7 billion episodes of acute diarrhoea, leading to nearly 700,000 deaths worldwide annually (Liu et al. 2012, Walker et al. 2013). In Latin America, RV has been estimated to cause 6,302 deaths and 229,656 hospitalisations annually in 14 countries, including Brazil, in the absence of vaccination (Desai et al. 2012).

RVs belong to the Reoviridae family, Sedoreovirinae subfamily and genus Rotavirus (ICTV 2014). The nonenveloped virion is 100 nm in diameter and a triple-layered capsid surrounds the genome, which is composed of 11 segments of double-stranded RNA (dsRNA). These segments encode six viral structural proteins (VP1-4 and VP6-7) and six nonstructural proteins (NSP1-6) (Estes & Kapikian 2007, Greenberg & Estes 2009).

RVs are classified into eight species (A-H). Most strains that infect humans belong to group A (RVA); however, some group B, C and H strains have been associated with infections in humans (Matthijnssens & Van Ranst 2012, Matthijnssens et al. 2012). Based on the results of reverse transcription-polymerase chain reaction (RT-PCR) and sequence analyses, RVA strains are classified into the G and P genotypes, which represent the two type-specific outer capsid proteins, VP7 (glycoprotein) and VP4 (protease-sensitive protein), respectively. At least 27 G-types and 37 P-types have been described to date (Matthijnssens et al. 2011, Trojanar et al. 2013).

The most common G and P genotype combinations in human RVA (HuRVA) strains are G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8] (Santos & Hoshino 2005, Greenberg & Estes 2009, CDC 2011). Although the prevalence of G2P[4] in Phase III trials with Rotarix® in Brazil was reportedly low, some studies have shown the re-emergence of this genotype from 2005-2008 (Gurgel et al. 2007, Leite et al. 2008, Munford et al. 2009, Gómez et al. 2011). Additionally, the rate of G2P[4] detection has been decreasing since 2009 (Carvalho-Costa et al. 2011, Oliveira et al. 2012, Soares et al. 2012). This genotype fluctuation may be due to a natural phenomenon of re-emergence that occurs approximately every 10 years or may be related to the introduction of the Rotarix® vaccine or both (Gómez et al. 2011).
The nonstructural RV protein NSP4, encoded by RV segment 10, was the first viral enterotoxin to be described (Ball et al. 1996). NSP4 is a multifunctional glycoprotein involved in viral morphogenesis and pathogenesis (Ball et al. 2005) that is important for the assembly of transiently enveloped double-layered particles (DLP) during RV replication (Taylor et al. 1996). NSP4 also promotes an increase in the intracellular calcium concentration, which results in chloride secretion into the intestinal lumen (Ball et al. 1996), the disruption of tight junctions (Tian et al. 1996) and, consequently, aqueous diarrhea (Ball et al. 1996).

The gene that encodes NSP4 has been characterised into six (A-F) (Horie et al. 1997, Ciarlet et al. 2000, Ito et al. 2001, Mori et al. 2001) or 15 genotypes (E1-E15) (Papp et al. 2012). The previously described A, B and C genotypes are typically found in human RVs and correspond to the new E2, E1 and E3 genotypes, respectively (Matthijnssens et al. 2008a). As the diversity among NSP4 genotypes may be important for amino acid (aa) changes that could alter the virulence of HuRVA strains (Rajasekaran et al. 2008), it remains unclear whether NSP4 should be included in RV vaccination strategies because this protein also appears to play roles in immunity and protection (Ball et al. 2005, Araújo et al. 2007a).

Many studies have focused on molecular characterisation of the NSP4 gene of HuRVA collected from different regions of Brazil (Mascarenhas et al. 2006, Araújo et al. 2007b, Tavares et al. 2008, Nozawa et al. 2010, Fumian et al. 2011, Gómez et al. 2011); however, no studies have examined HuRVA strains collected in the state of São Paulo (SP). The aim of this study was to characterise the NSP4 genotype of HuRVA collected from children and adults hospitalised in SP that only one (2.4%) was of the E1 genotype.

Different HuRVA strains detected in Brazilian states were used to construct a phylogenetic tree along with RV reference strains of the E genotypes (Matthijnssens et al. 2011) and strains with major similarities, as assessed by RotaC. The phylogenetic tree produced distinct clusters, grouping most strains by their genotype and collection year (Fig. 1). As expected, E1 strain SP 2186 clustered with E1 strains from different Brazilian states; however, this strain did show 96.5% nucleotide similarity with the strain did show 96.5% nucleotide similarity with the
Most of the strains from SP clustered together with strains from other Brazilian states. Only one branch was solely composed of strains obtained in this study (during the years 2009 and 2010).

Among the 40 deduced aa sequences of the HuRVA E2 NSP4 gene, 16 were aligned to represent the primary observed aa variations (Fig. 2). The alignment results revealed the prevalence of variations at antigenic sites II and III, the VP4-binding domain, the enterotoxin domain and the extracellular matrix-protein binding domain.

Some nucleotide polymorphisms led to aa changes in all of the strains collected in a given year (Table). Only the strains collected in 2006 encoded a different aa at position 87 (E→D). The same type of change was observed at positions 136 (V→M) and 139 (T→I) for strains from 1994. Positions 103 (V→I) and 131 (Y→H) exhibited aa changes in 2007 and 2009, respectively, and these changes were maintained in the strains isolated in later years. Finally, with regard to position 135, the strains from 1994 showed the same aa as the RV reference strain DS-1 (M), whereas the strains from the other years showed the same aa as RV patterns TB-Chen and B1711 (I).

**DISCUSSION**

This study analysed 41 HuRVA strains classified as G2P[4]; the NSP4 sequences were genotyped as E1 or E2, allowing for an investigation of the linkage between VP4, VP7 and NSP4. The deduced aa sequences were determined and variations were observed at different sites of the NSP4 coding sequence.

The results suggested a linkage between VP7, VP4 and NSP4, as 40 (97.6%) G2P[4] HuRVA strains displayed the E2 genotype. In contrast, only one strain (2.4%) displayed the E1 genotype.

A new full genome-based classification system has established 15 NSP4 genotypes, as determined by nucleotide sequencing and phylogenetic analysis. This system establishes NSP4 genotype B (Wa-like) as the E1 genotype and genotype A (DS-1-like) as the E2 genotype, which is the most diverse; the E3 genotype strains include the previous NSP4 genotype C (AU-1-like) (Matthijssens et al. 2008a, b). Intragenotypic diversity has also been reported for the majority of RV strains that infect other species (Papp et al. 2012).

Most HuRVA strains are NSP4 genotype E2 or E1. Genotype E2 appears to be closely related to G2P[4] strains, whereas genotype E1 appears to be most closely related to G1P[8] strains, more distantly related to G3P[8], G4P[8] and G9P[8]/P[6] (Araújo et al. 2007a, Tatte et al. 2010, Fredj et al. 2014) and rarely related to G2P[4] strains (Benati et al. 2010). These results are consistent with those of studies suggesting a linkage among the VP7, VP4 and NSP4 genes. Nonetheless, RV genes may also segregate independently in nature (Maunula & Von Bonsdorff 2002).

NSP4 phylogenetic analysis revealed clustering of the sequences according to genotype and sample collection year and the reference strains for each genotype were also clustered. The phylogenetic tree showed E2 intragenotype variation due to the presence of different branches formed by the strains from the same collection.
year. The same pattern of intragenotypic variation has been observed in E1 strains in Mexico and the authors attributed this finding to the accumulation of single point mutations in the NSP4 gene (González-Ochoa et al. 2013).

The HuRVA strains from different Brazilian states also clustered with the strains from this study. However, strains 16101 09ES/2009, 15900 08MG/2008 and 15860 08RJ/2008 formed a different branch in the phylogenetic tree. Gómez et al. (2011) has reported that these three Southeast Brazilian strains cluster into a monophyletic group due to their common region of origin. Therefore, this group would be expected to cluster together with the Southeast Region strains examined in the present study.

Fig. 2: comparison of the deduced amino acid (aa) sequences of the nonstructural protein 4 of G2P[4]E2 human rotavirus group A (HuRVA) strains detected in this study. The sequences of DS-1, TB-Chen and B1711 with the GenBank accessions AF174305, AY78760 and EF554091, respectively, were used as the reference strains. The dots indicate the identity of the DS-1 strain and the aas that differ from DS-1 are shown for the remaining strains using the single-letter aa code. The extracellular matrix (ECM)-protein binding domain (87-145), viral structural proteins (VP4)-binding domain (112-148), enterotoxin domain (114-135), double-layered particles (DLP)-binding domain (156-175), antigenic site (AS) IV (1-24), AS III (112-133), AS II (136-150) and AS I (152-169) are indicated by overlines.

This finding reinforces the intragenotypic variation exhibited by the NSP4 gene. Finally, the results did not provide evidence of the segregation of strains according to different Brazilian regions and this mixed distribution emphasises that circulation of the HuRVA genotypes in Brazil is homogeneous, despite the wide expanse of this country's territory.

The G2P[4]E2 strains from this study segregated into groups according to the collection year, a finding that has also been observed by Gómez et al. (2011), who analysed G2P[4]E2 strains from different Brazilian states. However, when the strains in the present study were analysed together with the strains from Gómez et al. (2011),
the resulting tree formed mixed branches, regardless of the collection year. This intragenotypic variation according to the sample collection year leads to changes in aa sequence that could also be related to the collection year, thus supporting the cluster formation observed in the phylogenetic tree. In addition, these temporal variations can occur due to point mutations in the NSP4 gene. Most of the variations in NSP4 aas from E2 strains occurred in the carboxy-terminal region, including the VP4-binding domain, and the same result has been observed in studies comparing NSP4 aas from E1 strains (González-Ochoa et al. 2013, Fredj et al. 2014). This finding suggests that NSP4 is involved in viral morphogenesis and that additional studies are necessary to understand how these variations in the NSP4 protein may interfere with the structural conformation of the interaction sites between NSP4 and VP4.

Most mutations were observed in the interspecies variable domain (ISVD) (aa 135-141). Within the ISVD domain, the presence of a glycine at position 140 defines the E2 subgenotype 1. Therefore, all of the strains from 1994 belong to subgenotype 1 and the others are classified as subgenotype 2 (Deepa et al. 2007, Fredj et al. 2014). Additionally, the residues D143, M144 and K146 in the last 16 aas, including the carboxy-terminal methionine, in the VP4- and DLP-binding domains, respectively, are highly conserved among all of the strains (Fredj et al. 2014). Previous studies have demonstrated that this conserved carboxy-terminal region (16-20 aa) is important for DLP-binding activity (O’Brien et al. 2000). With regard to the specific M-I mutation at position 135 in the strains from this study, the mutation most likely occurred between 1994 and 1996 because the DS-1 and TB-Chen strains were collected in 1976 and 1996, respectively.

Although NSP4 is a highly conserved protein, point mutations and genetic drift have led to aa changes over time. Because NSP4 induces a humoral immune response (Johansen et al. 1999), these changes may allow RVs to adapt or to escape host immunity. Therefore, variations in the NSP4 protein, together with variations in the VP4 and VP7 proteins, could be used to monitor RV circulation.

Finally, based on the prevalence of E2 genotypes in G2P[4] HuRVA strains, possible linkage among the G2, P[4] and E2 genotypes was confirmed. The detection of one E1 G2P[4] strain reinforces the idea that new genotype combinations emerge through reassortment and independent segregation.

In conclusion, we have demonstrated NSP4 genetic diversity in HuRVA G2P[4] strains collected in SP over the course of 12 years. The results reinforce the observed NSP4 intragenotype variation, which can likely be attributed to single point mutations and genetic drift events. The predominance of NSP4 aa changes in the VP4-binding domain emphasise the role of NSP4 in viral morphogenesis. Genomic and aa monitoring of circulating RVA strains may aid in identifying the occurrence of variations that might influence host-pathogen relationships. Moreover, the monitoring of genetic diversity among RVA strains is important for predicting the possible reemergence of strains after the introduction of vaccines against RVA.

### REFERENCES

Araújo IT, Assis RMS, Fialho AM, Mascarenhas JDAP, Heinemann MB, Leite JPG 2007a. Brazilian P[8], G1, P[8], G5, P[8], G9 and P[4], G2 rotavirus strains: nucleotide sequence and phylogenetic analysis. *J Med Virol* 79: 995-1001.

Araújo IT, Heinemann MB, Mascarenhas JDP, Assis RMS, Fialho AM, Leite JPG 2007b. Molecular analysis of the NSP4 and VP6 genes of rotavirus strains recovered from hospitalized children in Rio de Janeiro, Brazil. *J Med Microbiol* 56: 854-859.

Ball JM, Mitchell DM, Gibbons TF, Parr RD 2005. Rotavirus NSP4: a multifunctional viral enterotoxin. *Viral Immunol* 18: 27-40.

Ball JM, Tian P, Zeng CQ-Y, Morris AP, Estes MK 1996. Age-dependent diarrhea induced by a rotoviral nonstructural glycoprotein. *Science* 272: 101-104.

Benati FJ, Maranhão AG, Lima RS, da Silva RC, Santos N 2010. Multiple-gene characterization of rotavirus strains: evidence of genetic linkage among the VP7-, VP4-, VP6- and NSP4-encoding genes. *J Med Virol* 82: 1797-1802.

Carvalho-Costa FA, Volotão EM, de Assis RMS, Fialho AM, de Andrade JSR, Rocha LN, Tort LFL, da Silva MFM, Gómez MM, de Souza PM, Leite JPG 2011. Laboratory-based rotavirus surveillance during the introduction of a vaccination program, Brazil, 2005-2009. *Pediatr Infect Dis J* 30: S35-S41.

CDC - Centers for Disease Control and Prevention 2011. Rotavirus surveillance - worldwide, 2009. *MMWR Morb Mortal Wkly Rep* 60: 514-516.

Ciarlet M, Liprandi F, Conner ME, Estes MK 2000. Species specificity and interspecies relatedness of NSP4 genetic groups by comparative NSP4 sequence analyses of animal rotaviruses. *Arch Virol* 145: 371-383.

Deepa R, Rao CD, Suguna K 2007. Structure of the extended diarrhoea-inducing domain of rotavirus enterotoxigenic protein NSP4. *Arch Virol* 152: 847-859.

Desai R, Parashar UD, Lopman B, Oliveira LH, Clark AD, Sanderson CFB, Tate JE, Matus CR, Andrus JK, Patel MM 2012. Potential intussusception risk versus health benefits from rotavirus vaccination in Latin America. *Clin Infect Dis* 54: 1397-1405.
Sáfadi MAP, Berezin EN, Munford V, Almeida FI, de Moraes IC, Pinheiro CF, Racz ML 2010. Hospital-based surveillance to evaluate the impact of rotavirus vaccination in São Paulo, Brazil. *Pediatr Infect Dis J* 29: 1019-1022.

Santos N, Hoshino Y 2005. Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev Med Virol* 15: 29-56.

Soares LS, Lobo PS, Mascarenhas JDP, Neri DL, Guerra SFS, de Oliveira ASL, Maestri RP, Oliveira DS, de Menezes EMFC, Linhares AC 2012. Identification of lineage III of G12 rotavirus strains in diarrheic children in the northern Region of Brazil between 2008 and 2010. *Arch Virol* 157: 135-139.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.

Tatte VS, Rawal KN, Chitambar SD 2010. Sequence and phylogenetic analysis of the VP6 and NSP4 genes of human rotavirus strains: evidence of discordance in their genetic linkage. *Infect Genet Evol* 10: 940-949.

Tavares TM, de Brito WMED, Fiaccadori FS, Parente JA, da Costa PSS, Giugliano LG, Andreasi MSA, Soares CMA, Cardoso DDP 2008. Molecular characterization of VP6-encoding gene of group A human rotavirus samples from Central-West Region of Brazil. *J Med Virol* 80: 2034-2039.

Taylor JA, O’Brien JA, Yeager M 1996. The cytoplasmic tail of NSP4, the endoplasmic reticulum-localized non-structural glycoprotein of rotavirus, contains distinct virus binding and coiled coil domains. *EMBO J* 15: 4469-4476.

Tian P, Ball JM, Zeng CQ-Y, Estes MK 1996. The rotavirus non-structural glycoprotein NSP4 possesses membrane destabilization activity. *J Virol* 70: 6973-6981.

Trojnar E, Sachsenröder J, Twardziok S, Reetz J, Otto PH, Johne R 2013. Identification of an avian group a rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses. *J Gen Virol* 94: 136-142.

Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, O’Brien KL, Campbell H, Black RE 2013. Global burden of childhood pneumonia and diarrhea. *Lancet* 381: 1405-1416.

WHO - World Health Organization 2013. Rotavirus vaccines. WHO position paper - January 2013. *Wkly Epidemiol Rec* 88: 49-64.