Diversity of the G3 genes of human rotaviruses in isolates from Spain from 2004 to 2006: cross-species transmission and inter-genotype recombination generates alleles

Jorge Martínez-Laso,1 Angela Román,1 Miriam Rodríguez,1 Isabel Cervera,1 Jacqueline Head,2 Iciar Rodríguez-Avial2 and Juan J. Picazo2

INTRODUCTION

The genus Rotavirus, in the family Reoviridae, is composed of seven different groups (groups A–G), although only groups A, B and C have been identified in humans (Hoshino & Kapikian, 1994). Group A rotavirus (RV) is responsible for most cases of gastroenteritis in humans and its genome consists of 11 double-stranded RNA (dsRNA) segments that encode six structural and six non-structural proteins. Two outer capsid proteins, VP4 and VP7, are used to define serotypes P (pro tease-cleaved protein VP4) and G (glyco-protein VP7) (Estes, 2001). RV serotypes are defined by virus neutralization, but genotypes are studied more often since they can be more easily characterized by sequencing or RT-PCR followed by sequencing of amplicons.

Although 14 G serotypes, 16 G genotypes, 14 P serotypes and 27 P genotypes are known (Fukai et al., 1996; Liprandi et al., 2003; Martella et al., 2003; McNeal et al., 2005; Nakagomi et al., 1990), only 11 G and 15 P types have been identified in humans (Gentsch et al., 2005; Martella et al., 2006a; McNeal et al., 2005; Santos & Hoshino, 2005). Epidemiological studies in many parts of the world have indicated that there are five most common RV G types (G1, G2, G3, G4 and G9) and two P types (P[8] and P[4]) (Hoshino & Kapikian, 2000; Hoshino et al., 2002; Kapikian et al., 2001), and to a lesser extent P[6], P[9] and P[11] (Hoshino et al., 2005; Santos et al., 2001), mainly found in countries of temperate climates. However, in countries of tropical and subtropical climates, other G and P types have been found to be most prevalent (Desselberger et al., 2001; Iturriza-Gómez et al., 2003).

There are a number of reports of RV strains isolated from humans and animals that share genetic and antigenic features of virus strains from heterologous species. In many cases, genetic analysis by hybridization has clearly demonstrated the genetic relatedness of gene segments from RV strains isolated from different species. Together with the observation that some RV strains appear to be transmitted to a different species as a whole genome, these data suggest that interspecies transmission may occur frequently in...
nature (Fujiwara & Nakagomi, 1997; Iizuka et al., 1994; Nakagomi & Nakagomi, 2002; Nakagomi et al., 1990; Palombo, 2002; Matthijssens et al., 2006; Tsugawa & Hoshino, 2008; Nguyen et al., 2007; Parra et al., 2008; De Leener et al., 2004). The factors that promote interspecies transmission of animal RVs to humans or vice versa are poorly understood. Close contact between animals and humans may augment interspecies infections, and genetic reassortment during co-infection with RV strains from different animal species may result in the generation of progeny viruses with novel or atypical genotypes (Palombo, 2002).

RV G3 strains represent the genotype for which the broadest host range has been described. Genotype G3 strains have been detected in several host species including humans, monkeys, rabbits, pigs, birds, cats, dogs, horses, mice, cows and lambs (Gentsch et al., 1996; Khamrin et al., 2006; Lee et al., 2003; Martella et al., 2001, 2003; McNeal et al., 2005). The heterogeneity of G3 RVs, which segregate into various lineages according to species-specific patterns, was first described by Nishikawa et al. (1989). AU-1 was the first G3 genotype isolated and identified as P[9] G3-type human RV, in Akita, Japan, in 1982 (Nakagomi et al., 1987). This genotype has been described as barely present and even absent during the 1990s worldwide and it has re-emerged during recent years (Arista et al., 1990, 1997, 2003; Caprioli et al., 1996; De Grazia et al., 2007; Kebaraetswe et al., 2005; Lai et al., 2005; Reidy et al., 2005; Sanchez-Fauquier et al., 2006), to the extent that it has acquired an important epidemiological role.

The re-emergence of G3 RV strains (Sanchez-Fauquier et al., 2006; J. J. Picazo and others, unpublished results) and the fact that they can be transmitted from different species to humans mean that a knowledge of their genotype prevalence and interspecies origins has become more relevant, since the newly licensed vaccines RotaTeq and Rotarix (Zomer et al., 2008, 2009) will need to protect against the diversity of strains in circulation.

Post-marketing surveillance is under way at present and relevant data are expected to be published in 6–12 months time. It should be noted that the two vaccines have been developed on the basis of very different rationales: to either elicit an immune response of neutralizing antibodies that is as broad as possible (RotaTeq) or raise an immune response by repeated application of a monotypic vaccine (Rotarix), in the expectation of the emergence of a cross-protective immune response. There is currently uncertainty as to the extent that the different vaccines will drive the emergence of new RV types in humans and whether the vaccine formulae will have to be changed.

During an epidemiological study of RV genotypes in Spain (J. J. Picazo and others, unpublished results), six G3 genotype sequences [the third most frequent G genotype after G9 and G1 in Spain (J. J. Picazo and others, unpublished results)] were obtained and compared with other G genotype sequences of RV strains frequently isolated from humans and other species. Due to the diversity of G3 alleles and the presence of fragments of G1 and G2 in G3 sequences, an evolutionary and phylogenetic analysis (lineages) was performed.

**METHODS**

**Subjects.** During the development of a multicentre epidemiological study of RV genotyping in Spain (using children of Caucasian origin between 0 and 2 years old, staying in the hospital for more than 24 h, with the main selection criteria being severe diarrhoea and testing positive for RV serology) (2006–2008), six G3 genotype samples from Madrid were randomly selected for sequencing. The sequences obtained were compared with G3 sequences of human and different mammalian species and other G type sequences described worldwide (see below).

**RNA extraction.** Double-stranded RNA was extracted from 140 µl 10% diluted stool suspensions with the QiaAmp Viral RNA extraction kit (Qiagen) following the manufacturer’s protocol.

**G genotyping.** Genotyping was performed using a one-step procedure. This method combines RT-PCR and a nested specific multiplex PCR (One-Step RT-PCR kit; Qiagen). The following oligonucleotide primers were used: En9 consensus primer for all G genotypes, and G-specific primers aBT1 (G1 genotype), aCT2 (G2 genotype), aET3 (G3 genotype), aDT4 (G4 genotype) and aFT9 (G9 genotype) (Gouvea et al., 1990).

Briefly, 5 µl extracted RNA was mixed with 5 µl × PCR buffer, 1 µl 10 mM dNTPs, 0.25 µl each primer at 100 µM, 1 µl enzyme mixture (as indicated by the manufacturer) (Omniscript Reverse Transcriptase, Sensiscript Reverse Transcriptase and HotStarTaq DNA Polymerase) and ddH2O to a final volume of 25 µl. The following cycling parameters were used: one RT cycle at 50 °C for 30 min followed by 15 min at 95 °C and a total of 35 cycles at 94 °C for 30 s, 50 °C for 1 min and 72 °C for 45 s. A final extension step of 72 °C for 10 min was performed. The PCR products were analysed by agarose gel electrophoresis and visualized by staining with ethidium bromide. G genotypes were assigned according to the determined electrophoretic migration pattern.

**Sequencing.** The RNA of RV isolates was reamplified with the consensus primers Bg9 and End9 (Gouvea et al., 1990) in order to obtain the whole RNA. The amplified PCR products were purified using the UltraClean PCR Clean-Up kit (MoBio), following the manufacturer’s instructions, and sequenced in both directions using a BigDye terminator cycle sequencing kit (Applied Biosystems) and an ABI PRISM 3700 automated DNA sequencer. To avoid errors in the PCR process, each sample was sequenced using two different PCR amplifications. The Bg9 and End9 RT-PCR primers were used as sequencing primers.

**Statistical analysis.** Generated sequences were analysed with the Seqin program (Applied Biosystems) and contiguous sequences of the forward and reverse reactions from the two different PCRs were assembled.

Multiple sequence alignment analyses were carried out by using the CLUSTAL W algorithm (Thompson et al., 1994; European Bioinformatics Institute; http://www.ebi.ac.uk/clustalw). Phylogenetic and molecular evolutionary analyses were performed using MEGA version 3.1 and the neighbour-joining method with a bootstrap test (Kumar et al., 2004). A recombination detection program (RDP) version 3.12 (Martin et al., 2005; http://darwin.uvigo.es/rdp/rdp.html) was also used in order to
confirm the phylogenetic analysis and possible recombination patterns between the sequences studied.

**Nucleotide sequences.** The nucleotide sequences obtained in this paper have been deposited in the GenBank database with the following accession numbers: EU159186, EU159187, EU159188, EU159189, EU159190 and EU159191. A total of 160 G sequences were used for comparisons (Supplementary Table S1, available in JGV Online) as follows. G3 genotype: human (75), porcine (10), bovine (seven), lapine (six), rhesus (two), equine (nine), canine (three), caprine (one) and murine (one). G1 genotype: human (five) and porcine (two). G2 genotype: human (three) and porcine (one). G4 genotype: human (14) and porcine (five). G5 genotype: human (one), equine (one) and porcine (five). G8 genotype: human (three). G9 genotype: human (five) and porcine (three). G11 genotype: porcine (two).

**RESULTS AND DISCUSSION**

The complete nucleotide sequences obtained were 1004 nt (the terminal sequences determined by the primers were not considered). These, along with the 978 nucleotide (nt) open reading frames (ORFs) and the deduced amino acid (aa) sequences from the comparison sequences described above (Methods) were compared.

**Comparisons of deduced nucleotide and corresponding amino acid sequences**

**Spanish G3 genotype sequences.** Only two of six G3 VP7 gene RV nucleotide sequences from Spanish individuals presented 100 % nt sequence identity (EU159187 and EU159188). These two nucleotide sequences had the same amino acid sequence as EU159186. Therefore, from the six nucleotide sequences analysed, only four different protein sequences were obtained. This shared amino acid sequence was also found in the sequence AY707794 originating in Thailand. The only characteristic motifs are located at: position 214, with the presence of a serine instead of threonine in sequences EU159190 and EU159191 (present only in the G1 genotype sequences; Supplementary Fig. S1, available in JGV Online); position 218, with asparagine instead of valine in sequence EU159190, which is present in equine G3 sequences and a specific G3 genotype lineage (see below). Similarly, position 253, which is alanine in sequence EU159190, which is present only in one G4 genotype sequence from pig. These changes are mainly due to non-fixed point mutations and may be common with other strains by convergent evolution.

**Identical G3 sequences.** G3 genotype sequences were aligned and compared using the CLUSTAL W program. Comparisons among the rest of the G3 genotype sequences identified a number of identical full cDNA sequences: (i) AY707794, AY707791 and AY707790; (ii) D86270, D86269 and D86271; (iii) DQ873678 and DQ873671; (iv) U04350 and AB081594; and (v) AY740736, AY456382 and NC007468, which are all different accession numbers for the same B4106 sample (Maneekarn et al., 2006).

By comparing sequences with the same nucleotides at the VP7 ORF, the following three groups could be established: (i) L21666 (human; 1043 nt) and AF271089 (canine; 1062 nt); (ii) DQ995489, DQ995490 and DQ873669; and (iii) DQ995488 and DQ873677.

By comparing similarity at the amino acid sequence level only (with a different VP7 ORF nucleotide sequence), the following two groups could be established: (i) D86283 and D86281 and (ii) AY707794 (shares a full sequence with AY707794, AY707791 and AY707790), AY900173, AY870661, DQ904501, DQ995489 (shares a full ORF sequence with DQ995490 and DQ873669), DQ873670, DQ904505, DQ873672, DQ995488 (shares the ORF sequence with DQ873677) and DQ873673.

**G genotype reassignments.** During the process of sequence comparisons, some of the genotype alleles did not prove to be concordant with the genotype assigned (Supplementary Fig. S1 and Fig. 1):

G4-AJ488587. This sequence was registered in GenBank as G4. Comparison of this sequence with other G4 genotypes yielded only 71–74 % nt similarity and 73–78 % aa similarity. In contrast, when the sequence was compared with G3, similarity increased to 80 % (nt) and 89–93 % (aa). Therefore, this sequence belongs to the G3 genotype, although its similarity is low because it is included in a specific G3 genotype lineage (see below).

G3-J04334. This sequence was registered as G3 but only shares 72–73 % nt and 78–80 % aa sequence with other G3 genotypes. However, compared with the other G genotypes, a 91–92 % nt and 94–95 % aa sequence similarity with the G8 genotype was observed.

G3-AB081593. This sequence was registered as G3 but has only 72–73 % nt and 73–75 % aa sequence similarity. However, compared with the G2 genotype, it shares 97–98 % nt and 98–99 % aa sequence similarity.

G3-AB011973, G3-AB011971, G3-AB011972 and G3-AB011970. These sequences were defined as G3 but have only 72–74 % nt and 70–75 % aa similarity. Compared with other G genotypes, we found complete similarity with G4-AB039032 (nt and aa 100 %), G4-AB039031 (nt 99 %, aa 100 %), G4-AB039033 (nt and aa 100 %) and G4-AB039030 (nt and aa 100 %). In addition, when these sequences were compared with the remaining G4 genotypes, they were found to have 94–97 % nt and 91–95 % aa similarity.

**Generation of polymorphisms: point mutation and multiple recombination events**

As seen in other G genotypes, the most usual mechanism by which polymorphisms are generated in the G3 genotype
J. Martínez-Laso and others

Journal of General Virology 90

Lineage I

IA

Lineage II

IB

IC

Lineage III

Lineage IV

G3 → G8 (RA)

G1 ↔ G3 (GC)

G3 → G4 (RA)

G2 ↔ G3 (GC)

G2 → G3 (RA)

0.05
is by point mutation (Ramig, 1997; Reidy et al., 2005). Of 978 nucleotide sites, 496 were variable: 279 had synonymous substitutions and 217 had non-synonymous substitutions. Supplementary Fig. S1 provides evidence for point mutations as a mechanism of evolution. Inter-genotype recombination events were found for G3 and are described in Supplementary Fig. S1 and Figs 1 and 2. Four G3 sequences were found to be formed by this event, where the receptor sequences were G3 genotypes and the donors were G1 and G2. This evolutionary mechanism has been extensively described for human genes (Martinez-Laso et al., 2004, 2006), probably with variations, and also for RV genes (Cao et al., 2008; Martella et al., 2007; Phan et al., 2007a, b; Parra et al., 2004; Desselberger, 1996; Chnaiderman et al., 1998; Kojima et al., 2000; Patton et al., 2001; Alam et al., 2008; Schnepf et al., 2008).

G3-D86273 and G3-D86277. These sequences were generated by the combination of sequence G3-D86282 as a receptor (lineage I) and a fragment from aa 116 to 231 from the G2 genotype as a donor. In fact, G3-D86273 contains a more evolved G2 fragment that is more divergent than G3-D86277. These data were confirmed in G3-D86273 and G3-D86277 by using RDP, which suggested a possible recombination event in the region of nt 116–231, with statistically significant corrected P-values of $6.9 \times 10^{-17}$ and $6.0 \times 10^{-17}$, respectively (Fig. 2).

G3-D86276 and G3-D86274. In this case, a receptor sequence came mainly from the G3-D86275, D86279, D86282 or D86283 group of G3 genotypes (lineage I) and the donor was a G1 genotype which donates a fragment containing the sequence from aa 116 to 227. These results were also confirmed by RDP analysis with statistically significant corrected P-values of $9.9 \times 10^{-6}$ and $1.2 \times 10^{-19}$, respectively (Fig. 2). Thus, it seems more likely that one sequence comes from the other and not from a different G genotype.

Fig. 1. Phylogenetic tree showing the relationships between G3, G1, G2, G4, G5 and G9 genotypes from several RV species. POR, porcine; CAN, canine; RAB, rabbit; RHE, rhesus monkey; EQU, equine. GC, inter-genotype recombination event; RA, strain(s) reassigned to a different genotype.

Fig. 2. RDP statistical analyses of G3/G2 (a) and G3/G1 (b) inter-genotype recombination events. The most probable points of recombination for these events are located surrounding nt 116–237 and 241, respectively. Grey boxes, portion of sequence with a recombinant origin.
Phylogeny

Comparisons of the translated region (sequence reassignments and alleles generated by multiple recombination were excluded) of the nucleotide and amino acid sequences of the human G3 genotype revealed several degrees of similarity between them (98–79 % for nt sequences and 99–89 % for aa sequences; Supplementary Table S2, available in JGV Online). These results clearly showed that different types of evolution may have taken place. For this reason, G3 genotypes from other species (pig, horse, goat, sheep, rabbit, mouse and rhesus monkey) were included in the study.

Taking into account the nucleotide and amino acid sequence similarity (Supplementary Table S2), phylogenetic relationships between them (Fig. 1) and the parameters already established to define the lineages in other genotypes (i.e. G9; reviewed by Hoshino et al., 2005), four clearly different lineages were established. These lineages are found in particular host species, namely pigs, dogs, rhesus monkeys and rabbits. No correlation was found between human RV G3 sequences and those of horse (Supplementary Table S2), mouse, sheep and goat (data not shown).

Lineage I

Sixty-five G3 sequences were included in lineage I: D86264–D86272, D86275, D86278–D86284, AY707789–AY707791, AY707793, AY707794, AY753650, AY900173, AY870661, DQ995488–DQ995490, DQ873669–DQ873673, DQ873676–DQ873679, DQ904498–DQ904506, AY165009, DQ674933, DQ674934, AF3260957, AF260958, EF088831, EF088832, U04350, AB081594, AF386915, AF450293, AJ311738, AJ311739 and the Spanish sequences EU159186–EU159191. These sequences are clearly associated with the G3 sequence of porcine RV, with a nucleotide similarity of 84–89 % and an amino acid similarity of 90–96 % and are not associated with those of other host species (dog, rhesus monkey, rabbit and horse) (Supplementary Table S2). In contrast, this similarity diminished when the comparison was made with the other lineages: 80–81 % (nt), 91–93 % (aa) with lineage II; 78–81 % (nt), 89–92 % (aa) with lineage III; and 78–80 % (nt) and 89–92 % (aa) with lineage IV. Moreover, within lineage I, three different sublineages could be defined based on sequence similarities (Supplementary Fig. S1) and phylogenetic relationships (Fig. 1).

Sublineage IA includes the majority of the G3 sequences from lineage I (58) with the exception of D86264, AJ311738 and AJ311739. The sequences belonging to this sublineage have a similarity of 94–99 % (nt) and 93–99 % (aa) between them and 90–91 % (nt) and 92–96 % (aa) with respect to sublineage IB and 87–90 % (nt) and 90–92 % (aa) with respect to sublineage IC (Supplementary Table S2).

Sublineage IB includes only sequences AJ311738 and AJ311739. The sequence similarity between them is 99 % (nt) and 98 % (aa) and the sequences show 90–91 % (nt) and 92–93 % (aa) similarity with the G3 sequences of sublineage IC.

Sublineage IC is only formed by sequence D86264 and it is the most divergent of the three sublineages considered.

Sequence AY753650 is related to lineage I based on the phylogenetic tree (Fig. 1) and particularly related to sublineage IA; however, it has low percentage similarity values.

Lineage II

Sequence AY707792 (CMH222; Khamrin et al., 2006) has already been described in a rhesus monkey and is also shared by a canine species (Khamrin et al., 2006). Its sequence similarity is 87 % (nt) and 97 % (aa) with AF271090 (RV52; canine) and 87–88 % (nt) and 97 % (aa) with AF295303 (rhesus monkey) and M21650 (RRV; rhesus monkey). When compared with others of the same species (AF271089 and U97199; lineage III, see below), the similarity of the canine sequence is 82 % (nt) and 94–95 % (aa). The similarity between these two canine sequences (AF271089 and U97199) is 98 % (nt and aa). These results show that the RV described in the three canine species is different and that the G3 RV in AF271090 probably comes from rhesus monkeys or humans with RV sequences of G3 genotype lineage II (related to rhesus monkey). The sequence similarity with lineage III is 85–86 % (nt) and 95 % (aa), and with lineage IV is 85 % (nt) and 95 % (aa) (Supplementary Table S2). These results were also confirmed by phylogenetic analysis (Fig. 1).

Lineage III

This lineage comprises sequences AJ488587 (previously considered to be G4) and L21666 (HCR3; Khamrin et al., 2006), which share a similarity of 93 % (nt) and 98 % (aa) and which are grouped with the AF271089 (RV198/95; Khamrin et al., 2006) and U97199 (K9; Khamrin et al., 2006) G3 canine sequences. Their sequence similarities with humans are 91–96 % (nt) and 96–98 % (aa), and there is 100 % aa similarity between human L21666 and canine AF271089. The similarity with lineage IV sequences is 87 % (nt) and 93–94 % (aa) (Supplementary Table S2). This evolutionary group has been described elsewhere (Khamrin et al., 2006; Maneekarn et al., 2006) and has presented a different antigenic pattern, which has led to it being described as a G3 ‘subtype’ (Di Stefano et al., 2005).

Lineage IV

This lineage comprises the human G3 sequence AY740736, which is the same as AY456382, and NC007468 (B4106), and is grouped with G3 rabbit sequences (AF528201–AF528204). The sequence similarity of the human and rabbit sequences is 93–95 % (nt) and 96–97 % (aa) (Supplementary Table S1) (Khamrin et al., 2006; Maneekarn et al., 2006).

These lineages are clearly demonstrated based on the sequence similarity (Supplementary Table S2) and can be confirmed by the direct analysis of the sequences (Supplementary Fig. S1) and phylogenetic analysis (Fig. 1).
However, sequence DQ674932, described as G3 (Theamboonlers et al., 2008), could not be included in any of the lineages. In fact, when its sequence was compared with other G3 genotypes, a similarity of only 71–78 % (nt) and 56–82 % (aa) was found. The comparison with other G genotypes yielded only a 53–76 % (nt) and 81–88 % (aa) similarity. Furthermore, the phylogenetic tree (Fig. 1) groups this sequence with lineages II, III and IV but with a low bootstrap value (34 %; Fig. 1). These results allow the proposal that this sequence constitutes a different evolutionary stage of the G3 genotype or that it comes from a species that has not yet been tested. The data in this paper suggest that it may be related to G8 (Fig. 1).

On the other hand, taking into account the three antigenic regions in VP7, named A (aa 86–101), B (aa 142–152) and C (aa 208–221) (Kapikian et al., 2001), that are relevant for serotype antigenic variation, no important amino acid variations have been found in the G3 genotype sequences described.

The increased detection of RV strains bearing an unusual combination of cross-species phenotypes of human and animal RV has been well documented (El-Attar et al., 2001; Palombo et al., 2000; Ramig, 1997). This observation supports the hypothesis that interspecies transmission of RVs, including to and from humans, might take place in nature (Cook et al., 2004; Gouvea et al., 1990; Nakagomi & Nakagomi, 1993; Palombo et al., 2000). Interspecies transmission could be the result of infection with an animal RV (Nakagomi & Nakagomi, 2002) or of genetic reassortment between human and animal RV strains during co-infection of the same cell (Das et al., 1993; Dunn et al., 1993; Urasawa et al., 1992; Varghese et al., 2004, Steyer et al., 2008; Martella et al., 2006b).

These findings strongly suggest multiple interspecies transmission events of RV strains among different species and humans in nature and provide convincing evidence that the evolution of human RVs is closely linked to the evolution of animal RVs.

Monitoring prevalent genotypes using molecular methods such as sequencing will enable us to detect new strains or to establish different phylogenetic pathways of evolution for each genotype. Furthermore, it is clear that epidemiological studies, vaccine design and studies associating clinical features and infection would benefit from this kind of analysis.

ACKNOWLEDGEMENTS

This work was supported in part by the Ministerio de Sanidad (PI05/0338, CP04/00026), Instituto de Salud Carlos III (MPI1180/06) and Fundación para el Estudio de la Infección. We also thank the Fundación para la Investigación Biomédica del Hospital Clínico San Carlos.

REFERENCES


