# Intra-genotypic Diversity of Archival G4P[8] Human Rotaviruses from Washington, DC

**Sarah M. McDonald**[a,*], **Kristin Davis**[a], **John K. McAllen**[b], **David J. Spiro**[b,§], and **John T. Patton**[a]

[a]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 50 South Drive, Bethesda, MD 20892, USA

[b]The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA

## Abstract

Group A human rotaviruses (RVs) remain the most frequently detected viral agents associated with acute gastroenteritis in infants and young children. Despite their medical importance, relatively few complete genome sequences have been determined for commonly circulating G/P-type strains (i.e., G1P[8], G2P[4], G3P[8], G4P[8], and G9P[8]). In the current study, we sequenced the genomes of 11 G4P[8] isolates from stool specimens that were collected in Washington, DC during the years of 1974-1991. We found that the VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6-encoding genes of all 11 G4P[8] RVs have the genotypes of G4-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1. By constructing phylogenetic trees for each gene, extensive intra-genotypic diversity was revealed among the G4P[8] RVs, and new sub-genotype gene alleles were identified. Several of these alleles are nearly identical to those of G3P[8] isolates previously sequenced from this same Washington, DC collection, strongly suggesting that the RVs underwent gene reassortment. On the other hand, we observed that some G4P[8] RVs exhibit completely different allele-based genome constellations, despite being collected during the same epidemic season; there was no evidence of gene reassortment between these strains. This observation extends our previous findings and supports the notion that stable, genetically-distinct clades of human RVs with the same G/P-type can co-circulate in a community. Interestingly, the sub-genotype gene alleles found in some of the DC RVs share a close evolutionary relationship with genes of more contemporary human strains. Thus, archival human RVs sequenced in this study might represent evolutionary precursors to modern-day strains.

## Keywords

rotavirus; G4P[8]; genetic diversity; gene reassortment; clade

---

[*]Corresponding author current address and affiliation: Virginia Tech Carilion Research Institute and the Department of Pathobiology and Biomedical Sciences, College of Veterinary Medicine, Virginia Tech 2 Riverside Circle, Roanoke, VA 24016, USA; mcdonaldsa@vtc.vt.edu.
[§]Current address and affiliation: Respiratory Diseases Branch, Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, 6610 Rockledge Drive, Room 3217, Bethesda, MD 20817, USA

## 1. Introduction

Group A rotaviruses (RVs) are important pathogens that cause severe, watery diarrhea in infants and young children (Parashar et al., 2009; Parashar et al., 2006). The genome consists of 11 double-stranded (ds) RNA segments and is surrounded by a triple-layered protein capsid (Estes and Kapikian, 2007). In general, each genome segment encodes a single polypeptide, allowing RVs to express 6 structural proteins (VP1-VP4, VP6-7) and 5 or 6 non-structural proteins (NSP1-NSP5, sometimes NSP6) (Estes and Kapikian, 2007). Neutralizing antibodies generated against outer capsid proteins VP7 and VP4 correlate with protection from subsequent RV-induced disease (Offit, 1994). As such, serotypes defined by VP7 (G-serotypes) and VP4 (P-serotypes) antigenicity, as well as the G/P-genotypes based on gene sequences, are a principal method of classifying RV isolates (Estes and Kapikian, 2007). While numerous G/P-type combinations have been detected in nature, strains with specificities of G1P[8], G2P[4], G3P[8], G4P[8], and G9P[8] are responsible for the majority of cases of RV gastroenteritis in humans (Desselberger et al., 2001; Gentsch et al., 2005; Santos and Hoshino, 2005). Recently, a RV classification system was developed that assigns a genotype to each of the 9 internal protein genes (i.e., those encoding VP1-VP3, VP6, NSP1-NSP5/6) in addition to G/P-genotypes (Matthijnssens et al., 2008a; Matthijnssens et al., 2008b). Now, the acronym Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx is used to classify the VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5/6-encoding segments. Using this system, it was shown that the internal protein genes of RVs with G1, G3, G4, G9 and P[8] specificities are usually genotype 1, and are therefore said to belong to the Wa genogroup (Heiman et al., 2008; Matthijnssens et al., 2008a; Matthijnssens et al., 2008b). In contrast, the internal protein genes of G2 and P[4] strains are generally genotype 2, assigning them into the DS-1 genogroup (Heiman et al., 2008; Matthijnssens et al., 2008a; Matthijnssens et al., 2008b). Inter-genogroup reassortant RVs containing both genotype 1 and 2 genes can be generated via reassortment following co-infection with two 'pure' strains (e.g., G1P[8] and G2P[4]) (Ward et al., 1990). Such 'mixed' genotype strains have occasionally been isolated from humans, usually in the context of rare G/P-type combinations (Heiman et al., 2008; Matthijnssens et al., 2008a). However, the lack of sequence information for common human G/P-type strains (i.e., G1P[8], G2P[4], G3P[8], G4P[8], and G9P[8]) limits our understanding of RV genotypic and intra-genotypic diversity, notably for the genes encoding proteins other than VP7 and VP4.

To better understand the level of genetic variation among human RVs, we have an ongoing contract with the J. Craig Venter Institute (JCVI) in Rockville, MD to determine the sequences of archival strains in a Washington, DC stool collection. This collection contains 428 G1P[8]-, 27 G2P[4]-, 150 G3P[8]-, and 32 G4P[8]-positive fecal specimens, all of which were isolated from children with RV gastroenteritis at Children's Hospital National Medical Center during the years of 1974-1991 (Table 1). We have previously reported on the analysis of RVs from 51 of the 150 G3P[8]-positive specimens (McDonald et al., 2009). The results of that study showed that the 51 G3P[8] RVs viruses all belong to the Wa genogroup and have identical genotypes (G3-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1); however, we found that that there is significant intra-genotypic variation among their genes. Specifically, by generating phylogenetic trees using the open-reading frame (ORF) sequences, 4 or 5 sub-genotype alleles were identified for each gene. We observed that some G3P[8] RVs maintained completely different allele-based genome constellations, even though they were collected during the same epidemic season. This result suggested that (i) several clades of G3P[8] RVs co-circulated in Washington, DC, and that (ii) minimal gene reassortment occurred among viruses belonging to different clades.

Here, we report on the sequencing and analysis of 11 archival G4P[8] RVs from the same Washington, DC stool collection. To our knowledge, these are the first near-complete

genome sequences available for this medically-important serotype. We found that all 11 of the sequenced G4P[8] RVs have the genotypes G4-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, and that most share sub-genotype gene allele designations with those of the G3P[8] strains. However, for the VP4, VP1, and NSP4 genes, new sub-genotype alleles were identified in the G4P[8] RVs; these alleles are absent in the G3P[8] strains analyzed thus far. Consistent with the previous study, some of the G4P[8] RVs isolated during the same epidemic season exhibit completely different allele-based genome constellations, supporting the idea that several RV clades can circulate concurrently. Interestingly, the genes found in one of the G4P[8] clades, as well as in some of the G3P[8] isolates, are closely related to those of recently-circulating human strains and possibly represent ancestral gene precursors.

## 2. Materials and methods

### 2.1. Sample collection, RNA extraction, and G/P-typing

The methods of sample collection, RNA extraction, and G/P-typing were reported previously (McDonald et al., 2009). Briefly, fecal specimens were collected from children <2 years of age who were hospitalized with diarrhea at Children's Hospital National Medical Center, Washington, DC. Specimens were tested for RV using electron microscopy and enzyme-linked immunosorbent assay (ELISA). RNA was extracted from RV-positive samples using TRIzol (Invitrogen), and samples were classified into G/P-types based on the results of a microtiter plate hybridization-based PCR-ELISA.

### 2.2. RT-PCR and nucleotide sequencing

RT-PCR and nucleotide sequencing was performed at JCVI as described previously (McDonald et al., 2009). Briefly, RT-PCR was performed with 1 ng of RNA using OneStep RT-PCR kits (Qiagen). Reactions were scaled down to 1/5 the recommended volumes, the RNA templates were denatured in 50% DMSO at 95°C for 5 min, and 1.6 units RNase Out (Invitrogen) was added. Following RT-PCR cycling, the reactions were treated with 0.5 units of shrimp alkaline phosphatase and 1 unit of exonuclease I (USB) and then incubated at 37°C for 60 min to degrade remaining dNTPs and digest the single-stranded primers. Enzymes were heat inactivated by incubation at 72° C for 15 min. The RT-PCR products were sequenced with an ABI Prism BigDye v3.1 terminator cycle sequencing kit (Applied Biosystems). The dye terminator was removed using Performa DTR cartridges (Edge Biosystems), and sequences were obtained with a 3730 DNA Analyzer (Applied Biosystems). Raw sequence data was trimmed to remove any primer-derived sequence ,as well as low quality sequence, and gene sequences were assembled using the Elvira and TIGR assemblers (www.jcvi.org/cms/research/software). The gene sequences were then manually edited using CloE (Closure Editor; JCVI) and polymorphisms were re-analyzed by sequencing. The nucleotide sequences deduced for the 11 genome segments of the 11 G4P[8] RVs were deposited into GenBank. Table 2 lists the accession numbers and the full strain name for each G4P[8] isolate according to the guidelines of the RCWG (Matthijnssens et al., 2011). For simplicity, we will refer to viruses by their abbreviated common names (DC2241, DC4996, DC5064, DC5115, DC827, DC1285, DC4613, DC1359, DC1208, DC4608, or DC4320).

### 2.3. Phylogenetic analyses

Nucleotide alignments were constructed using MacVector 8.1.2 and ClustalW with the set defaults (open gap penalty of 15.0, extended gap penalty of 6.66, delay divergence of 30%, and transitions weighted). Phylogenetic trees were generated using MacVector 8.1.2 and the neighbor-joining method (1000 bootstrap repetitions and Kimura-2 correction parameter). GenBank accession numbers and genotypes of the previously sequenced RV genes used in the phylogenetic analyses are provided in supplemental materials (Tables S1 and S2).

## 3. Results and discussion

### 3.1. Genotyping

RNA was extracted from all 32 G4P[8]-positive specimens of the Washington, DC stool collection and used as template for RT-PCR and sequencing. Of the 32 samples, 21 yielded partial genome sequences, while 11 contained viral RNA of sufficient quality to obtain near-complete genome sequences. Specifically, the entire ORFs, and in some cases portions of the untranslated regions (UTRs), were determined for each of the 11 genome segments of the 11 G4P[8] RVs. The sequence chromatograms showed strong, single peaks, suggesting that each stool specimen contained a dominant G4P[8] RV (data not shown). Using RotaC, a web-based RV genotyping tool (http://rotac.regatools.be), we confirmed the G/P-genotypes of the 11 G4P[8] RVs and identified the other 9 genes of each isolate as genotype 1 (Maes et al., 2009). Thus, the 11 G4P[8] RVs have the complete genotype descriptor of G4-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1, which is similar to that of the 51 previously sequenced G3P[8] RVs (G3-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1) (McDonald et al., 2009). Therefore, all of the archival RVs sequenced thus far from the Washington, DC stool collection belong to the Wa genogroup.

### 3.2. Intra-genotypic diversity of the G4P[8] RVs and relationships to the G3P[8] RVs

We next constructed neighbor-joining phylogenetic trees for each gene of the 62 sequenced archival DC RVs to examine the intra-genotypic diversity of the G4P[8] strains and their genetic relationships to the G3P[8] strains (Fig. 1). Similar to the previous study, we defined each major phylogenetic grouping of each tree as a sub-genotype gene allele and, for purposes of clarity, assigned it a specific color descriptor (red, orange, green, cyan, navy, or brown) (McDonald et al., 2009). Because there were a limited number of G4P[8] RVs in this study, we were not able to define sub-genotype gene alleles for VP7 (Fig. 1A). For the remaining ten genes (VP1-VP4, VP6, NSP1-NSP5/6), we found that most of the G4P[8] sequences cluster with those of the G3P[8] RVs in the phylogenetic trees, and therefore, they can be given the same color-coded allele designations (Fig. 1B-K). For example, the VP6 genes of G4P[8] isolates DC1285, DC4613, DC1359, DC827, and DC2241 (highlighted in yellow) cluster tightly with the orange VP6 alleles of G3P[8] isolates DC1563, DC1730, DC5751, DC4996, DC5064, and DC5115 (Fig. 1F). However, the VP4, VP1, NSP1, and NSP4 genes of a few G4P[8] RVs do not cluster with those of the G3P[8] RVs (Fig. 1B-C, 1G, and 1J). Instead, they comprise new phylogenetic groupings and sub-genotype alleles, which we designated with the color brown. For VP4, the isolate DC1285 from 1980 is the only virus to show a brown allele (Fig. 1B). For VP1, the brown allele is seen in isolates DC1208 and DC4608 from 1980 and is quite divergent from those of the cyan, red, green, and orange alleles seen for other archival Washington, DC RVs (Fig. 1C). For NSP1 and NSP4, the brown alleles are only detected in viruses isolated in 1977 (DC2241, DC4996, DC5064, and DC5115) (Fig. 1G and 1J). The color-coded allele assignments of the G3P[8] RV genes were consistent with what we had defined previously, with two exceptions (McDonald et al., 2009). First, we found that the NSP2 genes of G3P[8] isolates DC1563 and DC1730 cluster with the red, rather than the orange allele grouping when analyzed along with the G4P[8] RV sequences (McDonald et al., 2009) (Fig. 1H). Also, we found that the NSP3 genes previously described as green and orange alleles share a very close phylogenetic relationship when analyzed in context of the G4P[8] sequences (Fig. 1I). Therefore, the NSP3 genes of isolates DC2262, DC2239, DC130, and DC135 were redefined as orange, rather than green, alleles.

A summary of color-coded allele-based genome constellations of the G3P[8] and G4P[8] RVs sequenced from the Washington, DC stool collection thus far are summarized in Fig. 2. The results show that all of the G4P[8] RVs have genes, and in some cases genomes, quite

similar to those of the G3P[8] RVs when compared at the allele-level. The orange VP4 allele of the G3P[8] 1979 isolate DC1730 is 99.4% identical to that of the G4P[8] isolate DC4613 collected the following year. Likewise, the green VP1 alleles of the 1976 G3P[8] isolates DC2262, DC2239, and DC130 are ~99.3% identical to those of 1977 G4P[8] isolates DC2241, DC4996, DC5064, and DC5115. Given their genetic similarities, we think it is likely that the G3P[8] and G4P[8] RVs directly reassorted their genes. However, we cannot exclude the possibility that the G3P[8] and G4P[8] RVs independently acquired similar genes by reassorting with strains not yet analyzed, such as the G1P[8] strains. Strikingly, G3P[8] isolates from 1974 and 1979 (DC1563 and DC1730, respectively) show genome constellations indistinguishable at the allele level from the 1980 G4P[8] isolates DC4613 and DC1359 (Fig. 2). Yet, the two other G4P[8] strains from 1980 (DC1208 and DC4608) are very divergent from DC4613 and DC1359, and instead, are much more genetically-similar to the G3P[8] strains (DC1600 and DC792) isolated from the same year (Fig. 2). Moreover, the single 1988 G4P[8] isolate DC4320 is identical at the allele level to the majority of 1991 G3P[8] viruses (Fig. 2). It is quite probable that these particular G4P[8] viruses more similar to G3P[8] strains represent examples of VP7 gene reassortants. Together, the analyses allowed us to identify at least three different clades of G4P[8] RVs that circulated in Washington, DC during the years of collection, all of which have some genes closely related to G3P[8] RVs isolated from the same location. This data is consistent with our previous findings for the G3P[8] strains and clearly show that viruses (i) isolated from the same epidemic season, (ii) collected at the same geographical location, and (iii) having the same G/P-type specificity (i.e., G4P[8]) can contain completely different sub-genotype allele-based genome constellations (McDonald et al., 2009). The results presented here are also in accordance with a previous study, which used electrophoretic analyses of viral genomes to show that co-circulating strains with the same serotype can have distinctly migrating internal protein gene segments (Nakagomi et al, 1988). The biological reason for the existence of clades with preferred allele-based genome constellations remains to be determined. Yet, is interesting to speculate that, in addition to reassortment biases at the genotype level, preferences may also exist for certain gene sets at the sub-genotype level.

### 3.3. Comparison of archival DC RV genes to those of other sequenced strains

Having defined relationships among the sequenced G3P[8] and G4P[8] RVs from the Washington, DC collection, we next sought to determine how similar these archival strains are to other sequenced human and animal RVs. Therefore, we created neighbor-joining phylogenetic trees using the ORF sequences of representative DC RVs and those of various other RVs available in GenBank (Fig. 3). For VP7, the genes of the 11 G4P[8] strains sequenced in this study were compared to those of published human and porcine RVs with known G4 specificities (Fig. 3A). We found that the DC G4P[8] VP7 genes cluster with genes of other human G4 strains within lineage I (Fig. 3A) (Stupka et al., 2009). More specifically, the VP7 genes of the DC strains seem most closely related to those of older, cell culture-adapted human strains (e.g., Hochi, Odelia, and ST3) and to primary human RVs isolated before 2002 (e.g., MW4086/00, RMC100, and Kagawa/90-544) (Fig. 3A) (Kudo et al., 2001; Heiman et al., 2008; Page et al., 2010). Such genetic similarities might be a reflection of the general time period in which the viruses were isolated (during the years of 1977-2002). In support of this notion, the majority of VP7 genes from more recently circulating strains, isolated between 2002-2010, comprise a phylogenetic grouping distinct from that containing the archival DC RV VP7 genes (Fig. 3A). This grouping has been proposed to represent an emerging (i.e., modern) G4 sub-lineage (Stupka et al., 2009). However, the VP7 genes of G4 human RVs isolated in Vietnam in 2003 (strains VN-5, VN-9, VN-21, and VN-565) do not group with the modern sub-lineage and instead cluster with the 1988 G4P[8] isolate DC4320 (Fig. 3A) (Trinh et al., 2010). This result suggests that the VP7 genes of currently circulating G4 strains may show more intra-genotypic diversity

than previously thought. At the amino acid level, the archival DC RV VP7 proteins are nearly identical (92.2 to 100% amino acid identity) to those of other human G4 strains, including those of recently isolated RVs. This observation suggests that, irrespective of nucleotide variation, the VP7 proteins of G4 strains have changed very little over the last 30+ years. We did find that the VP7 protein of the 1978 G4P[8] isolate DC827 has two unique amino acid changes (M63L and A183V) not seen in any other human G4 strain for which a sequence is available in GenBank. Likewise, the VP7 protein of the 1988 isolate DC4320 shows a single unique amino acid substitution (V235I) (Fig. S1). These changes are conservative, and none are located in domains predicted to be bound by neutralizing antibodies; this result suggests that the unique changes do not affect VP7 function or antigenicity (data not shown).

For VP4, the genes of the G4P[8] strains sequenced in this study were compared phylogenetically to those of previously-published human RVs with P[8] specificities, including 5 archival G3P[8] DC RVs (Fig. 3B). Like what was seen for VP7, the VP4 genes of archival DC RVs generally cluster with those of older, cell culture-adapted P[8] human strains (e.g., Wa, D, KU, IAL28, and YO), which may reflect the general time period of isolation (during the years of 1977-2002) (Fig. 3B) (Heiman et al, 2008). Nonetheless, we did observe that the orange VP4 alleles found in some archival DC RVs from 1974, 1979, and 1980 (strains DC1563, DC1359, and DC4613) grouped more closely with VP4 genes of P[8] human strains isolated after 2002. This modern lineage includes VP4 proteins the 2006 Indian strain, 0613158, the 2008-2009 Belgium strains, BE00029 and BE00036, the 2008 German strain, GER126-08, and the 2009 United States strains, 2008747332 and 2008747332 (Arora and Chitambar, 2011; Pietsch et al., 2011). At the amino acid level, the orange allele-encoded VP4 is also very similar to proteins of the modern lineage strains (Fig. S2). In particular, the VP4 proteins of 1980 G4P[8] strains DC1359 and DC4613 share five amino acid changes (S144G, N/D195G, I580V, L604V, and K617N) with VP4 proteins of modern lineage strains when compared the proteins of older, pre-2002 RVs. Amino acids at positions 144 and 195 are located in putative antigenic domain 8-1 of VP8*, the distal cleavage fragment of trypsin-activated VP4 (Fig. S3) (Dormitzer et al., 2002; Dormitzer et al., 2004). It is possible that these changes provided a selective advantage to RVs with the orange-allele encoded VP4 proteins, thus allowing their emergence in the human population.

For VP1-VP3, VP6, and NSP1-NSP5/6, we found that most of the archival DC RV genes cluster phylogenetically with the genotype 1 genes of human strains and are separate from the genotype 1 genes of porcine RVs (strains OSU, Gottfried, YM, A131, and A253) and the porcine-like bovine RVs (strains KJ25-1 and KJ44) (Fig. 3C-K). However, we did notice that the newly-designated brown VP1 allele of DC G4P[8] isolates DC1208 and DC4608 comprises a phylogenetic grouping along with porcine RV genotype 1 VP1 genes (Fig. 3C). In fact, the VP1 gene of DC1208 is 93.2% identical at the nucleotide level to the VP1 gene of porcine strain YM, but only ~86.4% identical to the VP1 genes of the Washington, DC RVs (Almanza et al., 1994). A few other alleles of the archival DC strains also showed a close evolutionary relationship with genes of porcine RVs: the cyan and navy VP2 alleles, the cyan NSP1 allele, and the navy NSP3 allele (Fig. 3D, 3G, and 3I). It is not clear whether this phylogenetic pattern is indicative of direct reassortment among human and porcine strains or simply reflects a shared ancestral origin (Matthijnssens et al., 2008a).

Interestingly, when compared phylogenetically to other human strains, the orange alleles of many internal protein genes are closely related to genes of primary human RVs isolated in many different parts of the world during the years of 2002-2010 (e.g., 6361, 061060, BE00036, BE00029, Dhaka16-03, AM06-I, 2008747332, 2008747336, B3458, Matlab36-02, Dhaka12-03, Matlab13-03, GER172-08, GER126-08, B4633-03, and Dhaka25-02) (Fig. 3C-K) (Rahman et al., 2007; Matthijnssens et al., 2008a; Arora and

Chitambar, 2011; Matthijnssens et al., 2010; Pietsch et al., 2011). This pattern is similar to what we observed for the orange VP4 alleles (Fig. 3B). In contrast, the red, navy, cyan, green, and brown alleles of the archival DC RVs seem more similar to older, cell culture-adapted RVs and to primary human strains isolated before 2002 (Fig. 3C-K). For the NSP1 and NSP4 trees, the phylogenetic pattern is more complicated; there is no clear grouping of orange alleles with genes of contemporary human RVs (Fig. 3G and 3J). Specifically, while the orange NSP1 alleles cluster with genes of the 2008 United States strains 2008747332 and 2008747336, they are more distantly related to genes of other recent isolates (Fig. 3G). The contemporary Belgium and German strains, BE00039 and GER126-08, respectively, instead group with the green and navy NSP1 alleles (Fig. 3G) (Pietsch et al., 2011). For NSP4, the genes of many modern isolates comprise a phylogenetic grouping with the orange alleles of G4P[8] isolates DC1359, DC1285, and DC4613 (Fig. 3J). However, the NSP4 gene of G3P[8] isolate DC1563, which is also classified as orange, does not cluster with those of the G4P[8] isolates, and the branching separation shows a weak bootstrap value (<70%) (Fig. 3J). Nonetheless, it is clear that the sub-genotype alleles designated as orange in this study are genetically similar to genes of recently circulating human strains.

### 3.4. Conclusions

The results of this study reveal that extensive intra-genotypic diversity existed among 11 G4P[8] and 51 G3P[8] archival RVs from Washington, DC. These different G/P-type strains were found to share sub-genotype gene allele designations, suggesting that they might have undergone reassortment. Still, we found some G4P[8] strains that were isolated during the same epidemic season, but exhibit completely different allele-based genome constellations. This observation is similar to what we have previously reported for the G3P[8] RVs and supports the notion that stable, genetically-distinct clades of human RVs with the same G/P-type can co-circulate without reassorting genes. Interestingly, the genes designated as orange alleles in this study share a close evolutionary relationship with genes of more contemporary human strains. Based on this data, we hypothesize that human RVs with predominantly orange allele-like genes emerged in the population over the last few decades. If so, the archival DC RVs sequenced in this and our previous study might represent the oldest known ancestors of modern-day strains. However, we acknowledge that this study only analyzes a small subset of RVs relative to the vast number that infect children everyday. Future, large-scale RV genomics studies are certainly required to fully elucidate the evolutionary relationship among archival and contemporary strains.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

Almanza L, Arias C, Lopez S. Amino acid sequence of the porcine rotavirus YM VP1 protein. Res. Virol. 1994; 145(5):313–317. [PubMed: 7839009]

Arora R, Chitambar SD. Full genomic analysis of Indian G1P[8] rotavirus strains. Infect. Gen. Evol. 2011; 11:504–511.

Desselberger U, Iturriza-Gómara M, Gray JJ. Rotavirus epidemiology and surveillance. Novartis Found. Symp. 2001; 238:125–47. [PubMed: 11444024]

Dormitzer PR, Sun ZY, Wagner G, Harrison SC. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO. 2002; 21(5):885–897.

Dormitzer PR, Nason EB, Prasad BV, Harrison SC. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. Nature. 2004; 430(7003):1053–1058. [PubMed: 15329727]

Estes, MK.; Kapikian, AZ. Rotaviruses. In: Knipe, DM.; Howley, PM., editors. Fields Virology. 5th Edition. Lippincott Williams and Wilkens; Philadelphia: 2007. p. 1917-1974.

Gentsch JR, Laird AR, Bielfelt B, Griffin DD, Banyai K, Ramachandran M, Jain V, Cunliffe NA, Nakagomi O, Kirkwood CD, Fischer TK, Parashar UD, Bresee JS, Jiang B, Glass RI. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. J. Infect. Dis. 2005; 192(1):S146–159. [PubMed: 16088798]

Heiman EM, McDonald SM, Barro M, Taraporewala ZF, Bar-Magen T, Patton JT. Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. J. Virol. 2008; 82(22):11106–11116. [PubMed: 18786998]

Kudo S, Zhou Y, Cao X, Yamanishi S, Nakata S, Ushijima H. Molecular characterization in the VP7, VP4 and NSP4 genes of human rotavirus serotype 4 (G4) isolated in Japan and Kenya. Microbiol. Immunol. 2001; 45(2):167–171. [PubMed: 11293484]

Maes P, Matthijnssens J, Rahman M, Van Ranst M. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. BMC Microbiol. 2009; 9:238. [PubMed: 19930627]

Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van Ranst M. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. J. Virol. 2008a; 82(7):3204–3219. [PubMed: 18216098]

Matthijnssens J, Ciarlet M, Rahman M, Attoui H, Banyai K, Estes MK, Gentsch JR, Iturriza-Gomara M, Kirkwood CD, Martella V, Mertens PP, Nakagomi O, Patton JT, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K, Desselberger U, Van Ranst M. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. Arch. Virol. 2008b; 153(8):1621–1629. [PubMed: 18604469]

Matthijnssens J, Rahman M, Ciarlet M, Zeller M, Heylen E, Nakagomi T, Uchida R, Hassan Z, Azim T, Nakagomi O, Van Ranst M. Reassortment of human rotavirus gene segments into G11 rotavirus strains. Emerging. Infect. Dis. 2010; 16(4):625–630. [PubMed: 20350376]

Matthijnssens J, Ciarlet M, McDonald S, Attoui H, Bányai K, Brister R, Buesa J, Esona M, Estes M, Gentsch J, Iturriza-Gómara M, Johne R, Kirkwood C, Martella V, Mertens P, Nakagomi O, Parreño V, Rahman M, Ruggeri F, Saif L, Santos N, Steyer A, Taniguchi K, Patton J, Desselberger U, Van Ranst M. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Arch. Virol. 2011 in press.

McDonald SM, Matthijnssens J, McAllen JK, Hine E, Overton L, Wang S, Lemey P, Zeller M, Van Ranst M, Spiro DJ, Patton JT. Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. PLoS Pathogens. 2009; 5(10):e1000634. [PubMed: 19851457]

Nakagomi T, Akanti K, Ikegami N, Katshushima N, Nakagomi A. Occurrence of changes in human rotavirus serotypes with concurrent changes in genomic RNA electropherotypes. J. Clin. Microbiol. 1988; 26(12):2586–2592. [PubMed: 2852673]

Offit P. Immunologic determinants of protection against rotavirus disease. Curr. Top. Microbiol. Immunol. 1994; 185:229–254. [PubMed: 8050280]

Page N, Esona M, Seheri M, Nyangao J, Bos P, Mwenda J, Steele D. Characterization of genotype G8 strains from Malawi, Kenya, and South Africa. J. Med. Virol. 2010; 82(12):2073–2081. [PubMed: 20981795]

Parashar UD, Burton A, Lanata C, Boschi-Pinto C, Shibuya K, Steele D, Birmingham M, Glass RI. Global mortality associated with rotavirus disease among children in 2004. J. Infect. Dis. 2009; 200(1):S9–S15. [PubMed: 19817620]

Parashar UD, Gibson CJ, Bresse JS, Glass RI. Rotavirus and severe childhood diarrhea. Emerg. Infect. Dis. 2006; 12(2):304–306. [PubMed: 16494759]

Pietsch C, Schuster V, Liebert UG. A hospital-based study on inter- and intra-genotypic diversity of human rotavirus VP4 and VP7 gene segments, Germany. J. Clin. Virol. 2011; 50:136–141. [PubMed: 21074491]

Rahman M, Matthijnssens J, Yang X, Delbeke T, Arijs I, Taniguchi K, Iturriza-Gomara M, Iftekharuddin N, Azim T, Van Ranst M. Evolutionary history and global spread of the emerging G12 human rotaviruses. J. Virol. 2007; 81(5):2382–2390. [PubMed: 17166908]

Santos N, Hoshino Y. Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. Rev. Med. Virol. 2005; 15(1): 29–56. [PubMed: 15484186]

Stupka JA, Carvalho P, Amarilla AA, Massana M, Parra GI. National Rotavirus Surveillance in Argentina: high incidence of G9P[8] strains and detection of G4P[6] strains with porcine characteristics. Infect. Genet. Evol. 2009; 9(6):1225–1231. [PubMed: 19631763]

Trinh Q, Pham N, Nguyen T, Phan T, Yan H, Hoang P, Khamrin P, Maneekarn N, Li Y, Okitsu S, Mizuguchi M, Ushijima H. Sequence analysis of the VP7 gene of human rotaviruses G2 and G4 isolated in Japan, China, Thailand, and Vietnam during 2001-2003. J. Med. Virol. 2010; 82(5): 878–885. [PubMed: 20336736]

Ward RL, Nakagomi O, Knowlton DR, McNeal MM, Nakagomi T, Clemens JD, Sack DA, Schiff GM. Evidence for natural reassortants of human rotaviruses belonging to different genogroups. J. Virol. 1990; 64(7):3219–3225. [PubMed: 1693701]

**Highlights**

- Group A rotaviruses (RVs) are important pathogens that cause severe diarrhea in infants and young children.

- In the current study, we sequenced the genomes of 11 G4P[8] isolates from clinical stool specimens that were collected in Washington, DC during 1974-1991.

- By constructing phylogenetic trees for each gene, we revealed extensive intra-genotypic diversity among the G4P[8] RVs and identified sub-genotype alleles.

- We observed that some G4P[8] RVs exhibit completely different allele-based genome constellations, despite being collected during the same epidemic season, supporting the notion that stable, genetically-distinct clades can co-circulate.

- Surprisingly, the genes found in one of the clades are closely related to those of recently-circulating human strains and might represent evolutionary precursors.
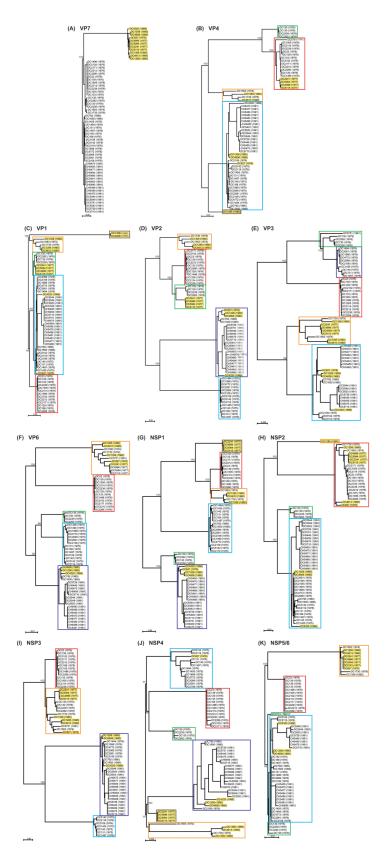
**Fig. 1.**
Genetic relationships among the genes of G3P[8] and G4P[8] RVs from an archival Washington, DC stool collection. The neighbor-joining phylogenetic trees were generated for each gene using the ORF nucleotide sequences. The trees are mid-group rooted for purposes of clarity, horizontal branch lengths are drawn to scale (nucleotide substitutions per base), and bootstrap values are shown as percentages for key nodes. Strain names are listed followed parenthetically by the year of specimen collection. Colored boxes indicate the genotype 1 allele designation (red, orange, green, cyan, navy, or brown). The archival G4P[8] RVs sequenced in this study are highlighted in yellow.
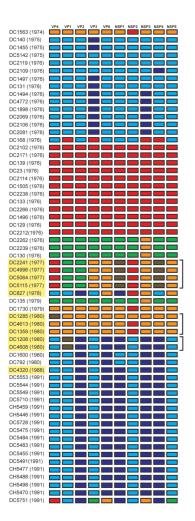
**Fig. 2.**
Allele-based genome constellations of the G3P[8] and G4P[8] RVs. The schematic
illustrates the color-coding of each gene based on the phylogenetic trees shown in Fig. 1.
The strain name and year of isolation are listed to the left of the corresponding genome
constellation (excluding the VP7 gene), and the protein encoded by each gene is listed at the
top. The G4P[8] RVs sequenced in this study are highlighted in yellow, and those of the
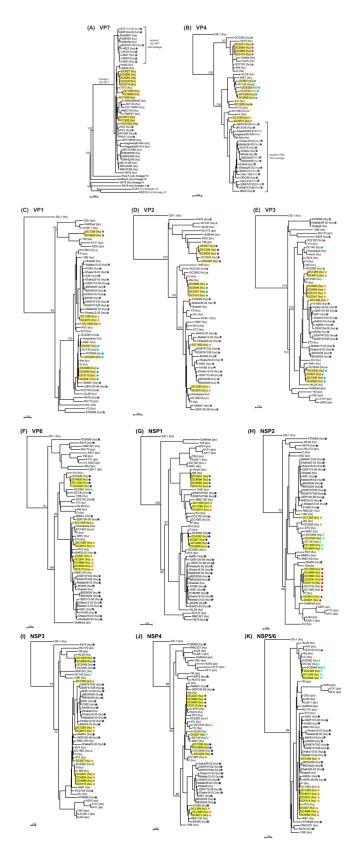same clade are indicated with brackets.

**Fig. 3.**
Genetic relationships of DC RV genes to those of other human and animal strains. The neighbor-joining phylogenetic trees were generated for each gene using the ORF nucleotide sequences. Horizontal branch lengths are drawn to scale (nucleotide substitutions per base), and bootstrap values are shown as percentages for key nodes. Strain names are listed followed parenthetically by host species: human (hu), porcine (po), and bovine (bo). The VP7 tree (A) is mid-group rooted and the VP1-VP4, VP6, and NSP1-NSP5/6 trees (B-K) are outgroup rooted to strain DS-1. Colored dots indicate the allele designation of the DC RV gene according to Fig.1. The archival G4P[8] RVs sequenced in this study are highlighted in yellow. Asterisks (*) indicate human RVs isolated during the years of 2002-2010.

**Table 1**

Distribution of RV G/P-types from an archival Washington, DC stool collection

| Year | # Samples for each G/P-type | | | |
|---|---|---|---|---|
| | G1P[8] | G2P[4] | G3P[8] | G4P[8] |
| 1974 | 20 | 0 | 1 | 1 |
| 1975 | 31 | 1 | 8 | 0 |
| 1976 | 8 | 2 | 49 | 2 |
| 1977 | 20 | 6 | 4 | 8 |
| 1978 | 55 | 1 | 11 | 1 |
| 1979 | 44 | 8 | 6 | 1 |
| 1980 | 37 | 3 | 10 | 18 |
| 1981 | 5 | 0 | 0 | 0 |
| 1987 | 6 | 0 | 0 | 0 |
| 1988 | 112 | 0 | 7 | 1 |
| 1989 | 60 | 0 | 8 | 0 |
| 1991 | 30 | 6 | 46 | 0 |
| **Total** | **428** | **27** | **150** | **32** |

**Table 2**

Full strain names and accession numbers for the G4P[8] RVs sequenced in this study

| Full Strain Name[*] | GenBank Accession #s |
|---|---|
| RVA/human/USA/1977/**DC2241**/G4P[8] | HM773876-HM773886 |
| RVA/human/USA/1977/**DC4996**/G4P[8] | HM773920-HM773930 |
| RVA/human/USA/1977/**DC5064**/G4P[8] | HM773931-HM773941 |
| RVA/human/USA/1977/**DC5115**/G4P[8] | HM773942-HM773952 |
| RVA/human/USA/1978/**DC827**/G4P[8] | HM773953-HM773963 |
| RVA/human/USA/1980/**DC1285**/G4P[8] | FJ947164-FJ947174 |
| RVA/human/USA/1980/**DC4613**/G4P[8] | HM773909-HM773919 |
| RVA/human/USA/1980/**DC1359**/G4P[8] | HM773865-HM773875 |
| RVA/human/USA/1980/**DC1208**/G4P[8] | HM773854-HM773864 |
| RVA/human/USA/1980/**DC4608**/G4P[8] | HM773898-HM773908 |
| RVA/human/USA/1988/**DC4320**/G4P[8] | HM773887-HM773897 |

[*] common name is shown in bold print