

# Molecular characterization of VP6 and NSP4 genes of unusual G and P Rotavirus group A isolated from children with acute gastroenteritis

Elizabeth-Barbara Tatsi<sup>1</sup>, Charilaos Dellis<sup>1</sup>, Dimitra-Maria Koukou<sup>1</sup>, Filippos Filippatos<sup>1</sup>, Evaggelia-Eirini Vetouli<sup>2</sup>, A. Michos<sup>1</sup>, and Vassiliki Syriopoulou<sup>1</sup>

<sup>1</sup>National and Kapodistrian University of Athens

<sup>2</sup>P & A Kyriakou” Children’s Hospital

August 12, 2023

## Abstract

Group A Rotavirus (RVA), which causes acute gastroenteritis (AGE) in children worldwide, is categorized mainly based on VP7 (genotype G) and VP4 (genotype P) genes. Genotypes that circulate at <1% are considered unusual. Important genes are also VP6 (genotype I) and NSP4 (genotype E). VP6 establishes the group and affects immunogenicity, while NSP4, as enterotoxin, is responsible for the clinical symptoms. Aim of this study was to genotype and molecularly characterize the VP6 and NSP4 genes of unusual RVA. Unusual RVA strains isolated from fecal samples of children [?]16 years with AGE, were genotyped in VP6 and NSP4 genes with Sanger sequencing. Phylogenetics was performed using the MEGA 11 program. In a 15-year period (2007-2021), 54.8% (34/62) of unusual RVA were successfully I and E genotyped. Three different I and E genotypes were identified; I2 (73.5%, 25/34) and E2 (35.3%, 12/34) were the commonest. E3 genotype was detected from 2017 onwards. The uncommon combination of I2-E3 was found in 26.5% (9/34) of the strains and G3-P[9]-I2-E3 was the most frequent G-P-I-E combination (20.6%, 7/34). Statistical analysis showed that children infected with E2 strains had a higher relative frequency of dehydration (50%) compared to those with the E3 genotype ( $p=0.019$ ). Multiple substitutions were detected in both genes, but their functional effect remains unknown. The results of this study highlight the genetic diversity of RVA strains. Continuous surveillance of the RVA based on the whole genome will provide a better knowledge of its evolution.

## Molecular characterization of VP6 and NSP4 genes of unusual G and P Rotavirus group A isolated from children with acute gastroenteritis

Charilaos Dellis<sup>1</sup>, Elizabeth-Barbara Tatsi<sup>1,2</sup>, Dimitra-Maria Koukou<sup>1</sup>, Filippos Filippatos<sup>1</sup>, Evaggelia-Eirini Vetouli<sup>3</sup>, Athanasios Michos<sup>1</sup>, Vasiliki Syriopoulou<sup>1</sup>

1. First Department of Pediatrics, Infectious Diseases and Chemotherapy Research Laboratory, Medical School, National and Kapodistrian University of Athens, “Aghia Sophia” Children’s Hospital, Athens, Greece.
2. University Research Institute of Maternal and Child Health and Precision Medicine, Athens, Greece
3. Department of Microbiology, “P. & A. Kyriakou” Children’s Hospital, Athens, Greece

*Corresponding Author*

**Tatsi Elizabeth-Barbara**, Researcher C, Ph.D., First Department of Pediatrics, Infectious Diseases and Chemotherapy Research Laboratory, Medical School, National and Kapodistrian University of Athens, “Aghia Sophia” Children’s Hospital, 11527, Athens, Greece

Tel: +302107467478

Email: etatsi@med.uoa.gr

## Abstract

Group A Rotavirus (RVA), which causes acute gastroenteritis (AGE) in children worldwide, is categorized mainly based on VP7 (genotype G) and VP4 (genotype P) genes. Genotypes that circulate at <1% are considered unusual. Important genes are also VP6 (genotype I) and NSP4 (genotype E). VP6 establishes the group and affects immunogenicity, while NSP4, as enterotoxin, is responsible for the clinical symptoms. Aim of this study was to genotype and molecularly characterize the VP6 and NSP4 genes of unusual RVA. Unusual RVA strains isolated from fecal samples of children [?]16 years with AGE, were genotyped in VP6 and NSP4 genes with Sanger sequencing. Phylogenetics was performed using the MEGA 11 program. In a 15-year period (2007-2021), 54.8% (34/62) of unusual RVA were successfully I and E genotyped. Three different I and E genotypes were identified; I2 (73.5%, 25/34) and E2 (35.3%, 12/34) were the commonest. E3 genotype was detected from 2017 onwards. The uncommon combination of I2-E3 was found in 26.5% (9/34) of the strains and G3-P[9]-I2-E3 was the most frequent G-P-I-E combination (20.6%, 7/34). Statistical analysis showed that children infected with E2 strains had a higher relative frequency of dehydration (50%) compared to those with the E3 genotype ( $p = 0.019$ ). Multiple substitutions were detected in both genes, but their functional effect remains unknown. The results of this study highlight the genetic diversity of RVA strains. Continuous surveillance of the RVA based on the whole genome will provide a better knowledge of its evolution.

**Keywords:** Rotavirus, acute gastroenteritis, children, NSP4, VP6, genotyping, phylogenetic

## Introduction

Group A Rotavirus (RVA) is one of the most common etiological agents of acute gastroenteritis (AGE) in infants and young children, especially in developing countries. Children with RVA AGE can present severe dehydration that can even lead to death if left untreated. RVA is responsible for more than 100,000 deaths each year worldwide.<sup>1</sup>

RVA is a non-enveloped, icosahedral, double stranded RNA virus (dsRNA) and is a member of the *Reoviridae* family. Its genome consists of 11 linear dsRNA segments which encode six structural viral proteins (VP1-VP4, VP6 and VP7) and six non-structural viral proteins (NSP1-NSP6).<sup>2</sup>

The viral particles consist of a triple layered capsid. The outer capsid consists of the glycoprotein VP7 and the spike protease-sensitive protein VP4. The middle layer consists of VP6 and the core layer comprises of the VP2 which encapsulates genomic RNA and viral replication components.<sup>2</sup> The abundant VP6 protein is commonly used for the detection and classification of rotaviruses. Currently, ten rotavirus species have been identified, A-J, but only A, B, C and H can infect humans including animal-human transmissions.<sup>2-5</sup>

RVAs are further classified based on the outer layer proteins VP7 and VP4 in G and P genotypes, respectively. Although many different G and P types have been identified so far the most common circulating genotypes are G1P[8], G2P[4], G3P[8], G4P[8], G9P[8] and G12P[8].<sup>6,7</sup> Genotyping can also be applied in the whole virus genome, Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, with “x” indicating the number of the corresponding genotype, which represents the genotypes of the genes VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5.<sup>8</sup>

NSP4, except for the full length protein, encodes a toxic-peptide (114-135 amino acids) and both act as enterotoxins that can stimulate  $Ca^{2+}$  release from the endoplasmic reticulum into the cytoplasm.<sup>9,10</sup> NSP4 enterotoxin activity aggravates the symptoms of gastroenteritis and especially diarrhoea and vomiting.<sup>11,12</sup>

Since 2006, several RVA vaccines have been released worldwide. The most widely used are the two-dose monovalent vaccine Rotarix (GlaxoSmithKline Biologicals, Belgium) and the three-dose pentavalent vaccine RotaTeq (Merck, United States), which cover the most common G and P genotypes.<sup>13,14</sup> After their release, notable changes in genotype distribution have been described worldwide, such as the increase in unusual G and P genotypes.<sup>2,15</sup> The aim of this study was the molecular and phylogenetic characterization of VP6 and

NSP4 genes of previously described<sup>16</sup> unusual G and P RVA strains isolated from children [?]16 years with AGE.

1. **Materials and Methods**  
**2.1 Study design** This is a retrospective study involving RVA positive fecal samples with previously described unusual G (G6, G8, G10) and/or P (P[6], P[9], P[10], P[11], P[14]) genotypes collected from children [?]16 years hospitalized with AGE.<sup>16</sup> In the present study, these strains were further genotyped in VP6 (I genotype) and NSP4 (E genotype) genes. Demographic and epidemiological data such as age, gender, residence, and RVA vaccination status were also collected from the children infected with an unusual RVA genotype. Clinical symptoms (diarrhoea, vomiting, fever and dehydration) and laboratory data (measurements of potassium (K<sup>+</sup>), sodium (Na<sup>+</sup>), calcium (Ca<sup>2+</sup>), Chlorine (Cl<sup>-</sup>), C-reactive protein (CRP), urea, creatinine, white blood cells (WBC), polymorphonuclear leukocytes and lymphocytes) were also recorded. The Scientific and Bioethics Committee of “Aghia Sophia” Children’s Hospital approved this study (No 6261).
2. **Reverse transcription and amplification of VP6 and NSP4 genes** Nucleic acid extraction and reverse transcription (RT) were performed as previously described.<sup>16</sup> PCR amplification was performed using GoTaq DNA polymerase (Promega; Madison, WIS, USA) and primers F: 5’-GAC GGV GCR ACT ACA TGG T-3’ and R: 5’-GTC CAA TTC ATN CCT GGT G-3’ for the VP6 gene and F: 5’-GGC TTT TAA AAG TTC TGT TCC GAG-3’ and R: 5’-GTC ACA YTA AGA CCR TTC CTT CCA T-3’ for NSP4 gene<sup>17,18</sup>. The PCR amplification was carried out with an initial denaturation at 94°C for 2 minutes (min), followed by 40 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at 55°C for the VP6 gene and 48°C for the NSP4 gene, extension for 1 min at 72°C and final extension for 10 min at 72°C. The amplification products were analyzed by 2% agarose gel electrophoresis using a 50bp DNA ladder (N3236S; New England Biolabs, Massachusetts, USA) and ethidium bromide staining. The expected size was 379 bp for the VP6 gene, and 749 bp for the NSP4 gene.
3. **Sequencing and phylogenetic analysis**

The I and E genotypes were determined by performing Sanger sequencing with the BigDye Terminator v3.1 cycle sequencer kit on an Applied Biosystems 3500 genetic analyser (Applied Biosystems, Waltham, MA, USA) and using the BLAST bioinformatic tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Phylogenetic evolutionary analysis was performed on VP6 and NSP4 genes using the MEGA 11 software (Molecular Evolutionary Genetics Analysis; [www.megasoftware.net](http://www.megasoftware.net)). Multiple sequence alignment was performed using MUSCLE software (Multiple Sequence Comparison by Log-Expectation). The nucleotide substitution model was selected based on the BIC (Bayesian Information Criterion) scores using MEGA11. The model used in this study was Tamura 3-parameter (T92) using a discrete Gamma distribution (+G) with five rate categories and assuming that a certain fraction of sites is evolutionarily invariable (+I). The evolutionary trees were constructed using the Maximum Likelihood method and bootstrap resampling with 1000 replicates.

### Statistical analysis

Data statistical analysis was carried out using SPSS software (IBM Statistical Package for Social Sciences for Windows, Version 25.0. Armonk, NY: IBM Corp). *p* -value [?] 0.05 was considered statistically significant. A Pearson’s chi-square test ( $\chi^2$  test) was applied to determine the differences for the variables that met the criteria of their application. For variables that did not meet the criteria of the  $\chi^2$  test or both variables had two categories (2x2 double entry matrix), Fisher’s exact test was used.

### Nucleotide sequence accession numbers

The nucleotide sequences of this study were deposited in GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) with accession numbers OM281957-59, OM287400, OM303085, OM303088, OM333185, OM333186, OM323986, OM461377, OM461378, OM972707-OM972710, ON004913, ON009342, ON156796, ON156797, ON185611-18, ON206978, ON206980-83, ON971933, ON971934 for the VP6 gene and OM281953, OM281956, OM283121-26, OM287398, OM287399, OM362404, OM948988-91, ON004914, ON156785-93, ON564370, ON564371, ON971935 for the NSP4 gene.

## Results

### I and E genotyping and genetic linkage with G and P genotypes

From 2007 to 2021, 54.8% (34/62) of the unusual RVA strains were successfully I and E genotyped, and they consisted of 5.9% (2/34) unusual G (G8 and G10), 64.7% (22/34) unusual P (P[6], P[9], P[10] and P[11]) and 29.4% (10/34) unusual G and P (G6P[9], G6P[14] and G8P[14]).

Three different I and E genotypes were identified: I1 (7/34, 20.6%), I2 (25/34, 73.5%), I3 (2/34, 5.9%) and E1 (5/34, 14.7%), E2 (12/34, 35.3%), E3 (11/34, 32.4%). The most common genotypes were I2 and E2. The E3 genotype was first detected in samples in 2017 and was the second most common E genotype (**Figure 1**). E3 was detected in strains with P[9] genotype combined with G3 (n=8), G4 (n=1), G6 (n=1), and G9 (n=1) genotypes (**Table 1**).

Six (6/34, 17.6%) samples were not successfully genotyped in NSP4 gene and they were characterized as EUD (unidentified E genotype). These RVA strains were the following: G2-P[6]-I2-EUD (n=1), G6-P[9]-I2-EUD (n=2), G8-P[8]-I1-EUD (n=1), G8-P[14]-I2-EUD (n=1), and G12-P[6]-I1-EUD (n=1) (**Table 1**).

The most frequent combinations of G-P-I-E were G3-P[9]-I2-E3 and G8-P[14]-I2-E2 accounting for 20.6% (7/34) and 11.8% (4/34) of the samples, respectively (**Table 1**).

### Association of I and E genotypes with patient characteristics

Statistical analysis of demographic, clinical, and laboratory data from children depending on RVA I genotype showed no significant correlation. The corresponding analysis with E genotypes showed that children infected with E2 RVA strains had a higher relative frequency of dehydration (6/12, 50%) compared to those with the E3 genotype (0/9, 0%) ( $p = 0.019$ ).

### Molecular characterization and phylogenetic analysis of VP6

Molecular characterization was performed in a 378bp fragment of VP6 gene which encodes the protein amino acids (aa) 243-368. This VP6 sequence was compared to reference strains from the Wa (K02086.1), DS-1 (DQ870507.1), and AU-1 (DQ490538.1) constellations depending on I genotype to detect substitutions. This comparison showed four homozygous missense substitutions [V252I (n=7/7), I281V (n=3/7), A287T (n=7/7), L291S (n=7/7)] in strains carrying the I1 genotype, four [V281I (n=19/25), S303A (n=25/25), M342L (n=1/25), V349I (n=1/25)] in strains carrying the I2 genotype and one (V330I) in both strains carrying the I3 genotype. However, none of these substitutions was *novel* after comparison of VP6 sequences with the 100 most similar strains using BLAST.

In the sequenced fragment of VP6 gene, a part of the antigenic region III (aa 208-274) was included. Genetic analysis revealed the existence of three already known substitutions; the homozygous Y248F carried by all I1, Wa and Rotarix strains, homozygous V252I carried by all I1 strains, and the homozygous I253V carried by I3 (n=1) and AU-1 strains.

Phylogenetic analysis of the VP6 gene in 34 unusual RVA strains revealed three distinct groups corresponding to I1, I2, and I3 genotypes with 100% reliability for I1 and I3 groups and 88% reliability for I2 group. Among the unusual RVA strains carrying the I2 genotype, two distinct clades (I2-A and I2-B) were identified with 94% and 80% reliability, respectively (**Figure 2**). The division of these clades is based on the 9 synonymous substitutions (L265L/c.793T>C, N266N/c.798T>C, Y273Y/c.819T>C, T287T/c.861T>A, L294L/c.882A>G, V304V/c.912G>A, L324L/c.972A>G/T, A344A/c.1032T>A, T347T/c.1041G>A). Clade I2-B is also divided into two subclades (I2-B1 and I2-B2). This separation also occurred due to the substitutions of the missense V281I/c.841G>A (carried by 5 strains) and four synonymous (N25N/c.75T>C, A275A/c.825A>T, T323T/c.969G>A, L324L/c.972T>G) substitutions.

### Molecular characterization and phylogenetic analysis of NSP4

Molecular characterization was performed on the whole NSP4 gene comparing the sequences of this study to Wa (AF093199.1), DS-1 (EF672582.1), and AU-1 (D89873.1) reference strains. Through this comparison,

13 homozygous missense substitutions were found in strains carrying the E1 genotype, 21 homozygous and two heterozygous missense substitutions in strains carrying the E2 genotype and 16 homozygous and one heterozygous missense substitution in strains carrying the E3 genotype (**Figure 3**). Most of these substitutions (n=23) were located in the VP4 binding region (aa 112-148).

The NSP4 gene sequences were compared with the 100 most similar strains using BLAST, and eight possibly *novel* substitutions were identified. These *novel* substitutions were the D140N in one E1 strain, the L25I (n=1), T78A (n=1) and D140N (n=2) in four E2 strains and the D19G (n=1), I24V (n=1), V102I (n=1), K141R (n=2) and T155M (n=2) in six E3 strains. Four of these substitutions were located within significant domains of NSP4. Specifically, D19G and T78A were in the conserved hydrophobic domains 1 and 3 (H1, H3), respectively, and the D140N and K141R were located in VP4 binding domain.

In the toxic peptide region, three already known homozygous substitutions were detected. The H131Y was found in 1/5 (20.0%) E1 strain, in 4/12 E2 strains (33.3%) and in 1/11 (9.1%) E3 strain. The M133V was found in 7/11 (63.6%) E3 strains and the M135V was detected in 1/12 (8.3%) E2 strain (**Figure 3**).

Phylogenetic analysis of the NSP4 gene in 28 unusual RVA strains revealed three distinct groups corresponding to E1, E2, and E3 genotypes with 100% reliability. Among the unusual RVA strains carrying the E2 genotype, three distinct clades (E2-A, E2-B and E2-C) were identified (**Figure 4**). The division of the E2-A clade from E2-B and E2-C is based on four synonymous substitutions (L21L/c.63A>G, I56I/c.168A>T, L116L/c.346C>T, V124V/c.372A>T). The E2-A clade differentiated from the E2-B clade due to one missense (A45T/c.133G>A) and additional five synonymous (N18N/c.54T>C, Q109Q/c.327A>G, L110/c.330A>G, I130I/c.390C>T, S138S/c.414G>A) substitutions and from the E2-C clade due to one missense (G140D/c.419G>A) and another five synonymous (P34P/c.102C>T, E125E/c.375G>A, I130I/c.390A>T, P168P/c.504G>A) substitutions. The E2-B clade differed and separated from the E2-C clade due to two missense (A45T/c.133G>A, G140D/c.419G>A) and nine synonymous substitutions.

Unusual strains carrying the E1 and E3 genotypes were also divided into 3 (E1-A, E1-B, E1-C) and 2 (E3-A, E3-B) distinct clades, respectively (**Figure 4**). Separation between the E1-A and E1-B strains occurred due to three missense (I141V/c.421A>G, T145S/c.433A>T, I169S/c.505\_506AT>TC) and 22 synonymous substitutions. The E1-C clade differed from both E1-A and E1-B clades due to two missense (I76V/c.226A>G, S161N/c.482G>A) and three synonymous (K3K/c.9G>A, L82L/c.244\_246TTG>CTA, P138P/c.414A>G) substitutions. Furthermore, the E1-C clade differed from the E1-A clade in three missense (V141T/c.421\_422GT>AC, S145T/c.433T>A, S169I/c.505\_506TC>AT) and 21 synonymous substitutions and from the E1-B clade in one missense (I141T/c.422T>C) and seven synonymous substitutions. The division among the E3 cluster appeared due to six missense (I51V/c.151A>G, R59K/c.176G>A, R141K/c.422G>A, F148I/c.442T>A, R151K/c.452G>A, Q152H/c.456A>C) and 25 synonymous substitutions (**Figure 4**).

## Discussion

There are limited studies that investigate the molecular characterization of VP6 and NSP4 genes in human RVA strains worldwide as the interest has mainly focused on G and P distribution. This 15-year study focusses on the genotyping and molecular characterization of the VP6 and NSP4 genes of unusual G and P RVA strains isolated from children hospitalized with AGE.

Genotyping revealed three different I (I1, I2, I3) and E (E1, E2, E3) genotypes in unusual RVA strains, I2 and E2 being the most common. According to the Rotavirus Classification Working Group, 32 I and E genotypes are known so far, with I1, I2 and E1, E2 being the most commonly detected genotypes among humans.<sup>9,21-23</sup> I1-E1 are strongly associated with G1/G3/G4/G5/G9-P[8] and follow the Wa-like genotype constellation, I2-E2 are associated with G2-P[4] typical of the DS-1 like genotype constellation and I3-E3 are associated with G3-P[9] typical of the AU-1 like constellation.<sup>22,24</sup>

Similarly, in a 10-year study (1996-2006) conducted in Brazil, which included both common and unusual

strains, they found the same three I and E genotypes. In their study the most prevalent I and E genotypes were I1 (82.7%) and E1 (81.5%), respectively. However, among strains with an unusual G (G6, G8, G10) and/or P (P[6], P[9], P[10], P[11], P[14]) genotype, I2 and E2 were the most prevalent I and E genotype (n=7/13),<sup>24</sup> as in the present study. In other epidemiological studies such as a 4-year study conducted in the Democratic Republic of Congo, although the number of unusual G and/or P strains recorded was substantial, only two I (I1, I2) and E (E1, E2) genotypes were recorded.<sup>25</sup>

In this study, I3 and E3 were detected in 2019 and 2017 onwards, respectively. Strains carrying the E3 genotype showed a significant increase between 2019-2021, during the COVID-19 pandemic period and they were mostly found in combination with G3-P[9]-I2. G3-P[9]-I2-E3 was the most prevalent G-P-I-E genotype combination throughout this study. The increase in E3 was observed in the same period with the increase of P[9] strains in Greece, as recorded by Tatsi et al.<sup>16</sup> The first record of the G3-P[9]-I2-E3 genotype in humans was in 2012 in Korea, where it was isolated from a 9 year old female.<sup>26</sup> However, a similar strain (G3-P[9]-I2-R2-C2-M2-A3-N2-T3-E3-H3) was recently identified in 2021 in Thailand, and was originated from a feline with diarrhoea.<sup>27</sup>

The rare combination of G3-P[9]-I2-E3 that was detected in this study is possibly derived from a reassortment event. Reassortment is common among RVs and is a crucial mechanism for the evolution of the virus. Molecular characterization of multiple RVA genes is important, as it may contribute to detect strains that do not fit into any of the major constellations (Wa, DS-1 and AU-1) and are probably products of reassortment events. Furthermore, this finding supports that VP6 and NSP4 can segregate independently, contradicting a study in 2003 that reported a genetic linkage among these two proteins in common, unusual and reassortant human strains<sup>22</sup>. Similarly to our observation, many studies reported such reassortment events at VP6 and NSP4, but at a lower rate. In an 11-year study (1996-2006) in Brazil, the I1-E2 unusual I-E genotype combination was found in 1.2% of circulating strains<sup>24</sup>. The combinations I2-E1 and I1-E2 were detected in 15.4% of RVA strains in India during 1990-2000 and in 6.5% in Iran during 2021-2022.<sup>28,29</sup>

Multiple amino acid substitutions were detected in both VP6 and NSP4 genes. While many of these variants own key positions in the proteins, their functional impact remains unknown. VP6 protein is crucial as it is used in molecular and serological diagnostic tests for RVA due to its high conservation.<sup>30,31</sup> The genetic analysis in this study showed that I2 was more conserved compared to I1 and I3, since only 16% of the I2 strains carried substitutions. This finding is in concordance with a similar study in South Africa, in which only I1 and I2 genotypes were described and I2 was found more conserved than I1 as it carried only two substitutions.<sup>31</sup>

VP6 also contains four major antigenic regions.<sup>32</sup> Nyaga et al. described many substitutions in I1 antigenic region III, three of which (Y248F, V252I and I253V) were also presented in our samples, but their functional effect is unknown.<sup>31</sup> Changes at the antigenic regions should be closely monitored since it could potentially affect the efficiency of the rotavirus detection methods and the future development of a VP6-based vaccine as it also induces the development of neutralization antibodies like the capsid proteins VP7 and VP4.<sup>33</sup>

NSP4 is an essential protein for virus morphogenesis and pathogenesis. In the present study, nine possibly *novel* substitutions were found in the NSP4 gene. Most substitutions were detected in VP4-binding domain which also contains the toxic peptide and the interspecies variable domain (ISVD). According to other studies characterizing the nucleotide sequence of the NSP4 gene, the ISVD region shows great heterogeneity and the amino acid vary according to genotype.<sup>18,34-37</sup> Limited functional studies exist and therefore the effects of these variants on the functionality and immunogenicity of the corresponding protein remain unknown.

Of interest are the substitutions in amino acid 131 in the region of the toxic peptide, in which the majority of the strains of this study carried the H131 and E2 strains mainly carried Y131. Ball et al. conducted functional study for this amino acid on infant mice and they found that substitutions in amino acid 131 has an effect on the enterotoxin properties of NSP4.<sup>38</sup> Specifically, they reported that the Y131K substitutions resulted in the absence of diarrhoea. Studies from Brazil between 1990-2000 and 1987-2003 have reported that Y131 was detected only in E2 strains, while E1 strains had H131, and there was no data regarding E3

strains.<sup>39,40</sup> Srivastava et al. showed that patients infected with a strain carrying Y131 experienced more severe diarrhoea.<sup>34</sup> Even though the severity of symptoms was not evaluated in the present study, statistical analysis showed that children infected with an unusual strain carrying the E2 genotype had a higher chance to exhibit dehydration, which may indicate more severe diarrhoea. This result may also be related to the fact that Y131 was detected more in E2 strains.

Limitations of the present study included partial sequencing of the VP6 gene and moderate detection rates of both VP6 and NSP4 genes in RVA-positive fecal samples. However, similar detection rates have also been reported in other studies, possibly due to poor sample storage conditions or the presence of RNases resulting in fragmentation of the viral RNA genome, presence of PCR inhibitors or inability of primers to hybridize.<sup>9,41</sup> Another limitation of our study was that the analysis was based only in four genes (VP7, VP4, VP6 and NSP4) and not in the complete genotype constellation, which would provide more information about the genetic evolution of the strains.

This is the first study of VP6 and NSP4 epidemiology and molecular characterization of unusual RVA strains in Greece, in which the unusual I3 and E3 genotypes, the reassortant I2-E3 human strains and many substitutions in significant domains of VP6 and NSP4 genes were detected. Furthermore, a significant clinical association between dehydration and E2 genotype was described.

In this study, the genotype distribution of the VP6 and NSP4 gene in unusual rotavirus strains was described. The association between RVA genotype and the severity of the symptoms needs to be further investigated. Continuous surveillance of the distribution of RVA genotypes based on the whole genome, the molecular characterization and their association with epidemiological and clinical data is important for the better knowledge of the virus' evolution, the disease prognosis and upgrading RVA vaccines.

## Declarations

**Conflict of interest:** All authors declare that they have no competing interests regarding the present study.

**Ethical approval:** The study protocol was approved by the scientific and bioethics committee of “Aghia Sophia” Children’s Hospital (No. 6261) and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

**Data availability:** All relevant data are within the paper.

**Author contributions** E.B.T., D.M.K., and V.S. contributed to the conception of the study. C.D., E.B.T., D.M.K., F.F. and E.E.V. collected samples and demographic data. C.D., D.M.K., and E.B.T carried out the experiments. C.D. and E.B.T. analyzed the data. C.D. performed the phylogenetic analysis and wrote the initial manuscript. A.M., V.S. and E.B.T. supervised the study. All authors reviewed and approved the final manuscript.

## References

1. Troeger C, Khalil IA, Rao PC, et al. Rotavirus Vaccination and the Global Burden of Rotavirus Diarrhea among Children Younger Than 5 Years. *JAMA Pediatr* . 2018;172(10):958-965. doi:10.1001/jamapediatrics.2018.1960
2. Crawford SE, Ramani S, Tate JE, et al. Rotavirus infection. *Nat Rev Dis Prim* . 2017;3:17083. doi:10.1038/nrdp.2017.83
3. Rojas M, Dias HG, Gonçalves JLS, et al. Genetic diversity and zoonotic potential of rotavirus A strains in the southern Andean highlands, Peru. *Transbound Emerg Dis* . 2019;66(4):1718-1726. doi:10.1111/TBED.13207
4. Shoja Z, Jalilvand S, Latifi T, Farzin R. Rotavirus VP6: involvement in immunogenicity, adjuvant activity, and use as a vector for heterologous peptides, drug delivery, and production of nano-biomaterials. *Arch Virol* . 2022;167(4):1013-1023. doi:10.1007/s00705-022-05407-9

5. Lestari FB, Vongpunsawad S, Wanlapakorn N, Poovorawan Y. Rotavirus infection in children in Southeast Asia 2008-2018: disease burden, genotype distribution, seasonality, and vaccination. *J Biomed Sci* . 2020;27(1):66. doi:10.1186/S12929-020-00649-8
6. Manouana GP, Niendorf S, Tomazatos A, et al. Molecular surveillance and genetic divergence of rotavirus A antigenic epitopes in Gabonese children with acute gastroenteritis. *EBioMedicine* . 2021;73:103648. doi:10.1016/J.EBIOM.2021.103648
7. Bergman H, Henschke N, Hungerford D, et al. Vaccines for preventing rotavirus diarrhoea: vaccines in use. *Cochrane Database Syst Rev* . 2021;11(11):CD008521. doi:10.1002/14651858.CD008521.PUB6
8. Matthijssens J, Ciarlet M, Rahman M, et al. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. *Arch Virol* . 2008;153(8):1621-1629. doi:10.1007/S00705-008-0155-1
9. Ben Hadj Fredj M, Zeller M, Fodha I, et al. Molecular characterization of the NSP4 gene of human group A rotavirus strains circulating in Tunisia from 2006 to 2008. *Infect Genet Evol* . 2012;12(5):997-1004. doi:10.1016/J.MEEGID.2012.02.011
10. Caddy S, Papa G, Borodavka A, Desselberger U. Rotavirus research: 2014-2020. *Virus Res* . 2021;304:198499. doi:10.1016/J.VIRUSRES.2021.198499
11. Lorrot M, Vasseur M. How do the rotavirus NSP4 and bacterial enterotoxins lead differently to diarrhea? *Virol J* . 2007;4:31. doi:10.1186/1743-422X-4-31
12. Greenberg HB, Estes MK. Rotaviruses: from pathogenesis to vaccination. *Gastroenterology* . 2009;136(6):1939-1951. doi:10.1053/J.GASTRO.2009.02.076
13. Matthijssens J, Joëlsson DB, Warakomski DJ, et al. Molecular and biological characterization of the 5 human-bovine rotavirus (WC3)-based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq. *Virology* . 2010;403(2):111-127. doi:10.1016/J.VIROL.2010.04.004
14. Zeller M, Rahman M, Heylen E, et al. Rotavirus incidence and genotype distribution before and after national rotavirus vaccine introduction in Belgium. *Vaccine* . 2010;28(47):7507-7513. doi:10.1016/J.VACCINE.2010.09.004
15. Koukou DM, Michos A, Chatzichristou P, et al. Rotavirus epidemiology and genotype distribution in hospitalised children, Greece, 2008 to 2020: A prospective multicentre study. *Eurosurveillance* . 2022;27(47):2101133. doi:10.2807/1560-7917.ES.2022.27.47.2101133
16. Tatsi E-B, Koukou D-M, Dellis C, et al. Epidemiological study of unusual rotavirus strains and molecular characterization of emerging P[14] strains isolated from children with acute gastroenteritis during a 15-year period. *Arch Virol* . 2023;168(5):149. doi:10.1007/S00705-023-05769-8
17. DOCUMENTS AND METHODS – European Rotavirus Network. Accessed April 28, 2023. <https://www.eurorotanet.com/project-information/documents-and-methods/>
18. Khalkhali P, Khavandegar A, Mozhgani SH, et al. Genotyping and sequence characterization of the NSP4 gene of human group A rotavirus strains in Northern Iran. *J Med Virol* . 2021;93(8):4824-4830. doi:10.1002/JMV.26998
19. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* . 1992;9(4):678-687. doi:10.1093/OXFORDJOURNALS.MOLBEV.A040752
20. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* . 2021;38(7):3022-3027. doi:10.1093/MOLBEV/MSAB120
21. Rotavirus Classification Working Group: RCWG – Laboratory of Viral Metagenomics. Accessed August 1, 2022. <https://rega.kuleuven.be/cev/viralmetagenomics/virus-classification/rcwg>

22. Iturriza-Gòmara M, Anderton E, Kang G, et al. Evidence for Genetic Linkage between the Gene Segments Encoding NSP4 and VP6 Proteins in Common and Reassortant Human Rotavirus Strains. *J Clin Microbiol* . 2003;41(8):3566-3573. doi:10.1128/JCM.41.8.3566-3573.2003
23. Araújo IT, Heinemann MB, Mascarenhas JDAP, Santos Assis RM, Fialho AM, Leite JPG. Molecular analysis of the NSP4 and VP6 genes of rotavirus strains recovered from hospitalized children in Rio de Janeiro, Brazil. *J Med Microbiol* . 2007;56(Pt 6):854-859. doi:10.1099/JMM.0.46787-0
24. Benati FJ, Maranhão AG, Lima RS, Da Silva RC, Santos N. Multiple-gene characterization of rotavirus strains: evidence of genetic linkage among the VP7-, VP4-, VP6-, and NSP4-encoding genes. *J Med Virol* . 2010;82(10):1797-1802. doi:10.1002/JMV.21816
25. Heylen E, Likele BB, Zeller M, et al. Rotavirus Surveillance in Kisangani, the Democratic Republic of the Congo, Reveals a High Number of Unusual Genotypes and Gene Segments of Animal Origin in Non-Vaccinated Symptomatic Children. *PLoS One* . 2014;9(6):e100953. doi:10.1371/JOURNAL.PONE.0100953
26. Jeong S, Than VT, Lim I, Kim W. Whole-genome analysis of a rare human Korean G3P rotavirus strain suggests a complex evolutionary origin potentially involving reassortment events between feline and bovine rotaviruses. *PLoS One* . 2014;9(5):e97127. doi:10.1371/JOURNAL.PONE.0097127
27. Lestari FB, Chandranoi K, Chuchaona W, Vongpunsawad S, Poovorawan Y. A G3P[9] rotavirus strain with an unusual genome constellation in a diarrheic cat in Thailand. *Arch Virol* . 2023;168(1):24. doi:10.1007/S00705-022-05641-1
28. Tatte VS, Chitambar SD. Evidence of discordant genetic linkage in the VP4, VP6, VP7 and NSP4 encoding genes of rotavirus strains from adolescent and adult patients with acute gastroenteritis. *Infect Genet Evol* . 2012;12(8):1630-1634. doi:10.1016/J.MEEGID.2012.07.005
29. Kachooei A, Tava Koli A, Minaeian S, et al. Molecular characterization of rotavirus infections in children less than 5 years of age with acute gastroenteritis in Tehran, Iran, 2021–2022: Emergence of uncommon G9P[4] and G9P[8] rotavirus strains. *J Med Virol* . 2023;95(2):e28529. doi:10.1002/JMV.28529
30. Ward RL, McNeal MM. VP6: A candidate rotavirus vaccine. *J Infect Dis* . 2010;202 Suppl:S101-7. doi:10.1086/653556
31. Nyaga MM, Esona MD, Jere KC, Peenze I, Seheri ML, Jeffrey Mphahlele M. Genetic diversity of rotavirus genome segment 6 (encoding VP6) in Pretoria, South Africa. *Springerplus* . 2014;3:179. doi:10.1186/2193-1801-3-179
32. Buragohain M, Cherian SS, Prabhakar G, Chitambar SD. VP6 capsid protein of chicken rotavirus strain CH2: Sequence, Phylogeny and In Silico antigenic analyses. *Virus Res* . 2008;137(2):173-178. doi:10.1016/J.VIRUSRES.2008.07.008
33. Sadiq A, Bostan N, Jadoon K, Aziz A. Effect of rotavirus genetic diversity on vaccine impact. *Rev Med Virol* . 2022;32(1):e2259. doi:10.1002/RMV.2259
34. Srivastava S, Jain A, Khan DN, Prakash S, Singh M, Awasthi S. Molecular characterization of the rotavirus enterotoxin NSP4 gene of strains causing diarrhoea in children aged 0-5 years in northern India. *J Appl Pharm Sci* . 2015;5(11):43-49. doi:10.7324/JAPS.2015.501107
35. Teimoori A, Nejati M, Ebrahimi S, Makvandi M, Zandi M, Azaran A. Analysis of NSP4 Gene and Its Association with Genotyping of Rotavirus Group A in Stool Samples. *Iran Biomed J* . 2018;22(1):42-49. doi:10.22034/IBJ.22.1.42
36. Ben Hadj Fredj M, Ben Hamida-Rebaï M, Zeller M, et al. Sequence and structural analyses of NSP4 proteins from human group A rotavirus strains detected in Tunisia. *Pathol Biol (Paris)* . 2014;62(3):146-151. doi:10.1016/J.PATBIO.2013.10.006

37. Bertol JW, Fregolente MCD, Caruzo TAR, et al. 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. *Mem Inst Oswaldo Cruz* . 2015;110(6):786-792. doi:10.1590/0074-02760150199
38. Ball JM, Tian P, Zeng CQY, Morris AP, Estes MK. Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. *Science* . 1996;272(5258):101-104. doi:10.1126/SCIENCE.272.5258.101
39. Mascarenhas JDAP, Linhares AC, Gabbay YB, et al. Molecular characterization of VP4 and NSP4 genes from rotavirus strains infecting neonates and young children in Belém, Brazil. *Virus Res* . 2007;126(1-2):149-158. doi:10.1016/J.VIRUSRES.2007.02.010
40. Tavares TDM, De Brito WMED, Fiaccadori FS, et al. Molecular characterization of the NSP4 gene of human group A rotavirus samples from the West Central region of Brazil. *Mem Inst Oswaldo Cruz* . 2008;103(3):288-294. doi:10.1590/S0074-02762008000300011
41. Tavares TDM, De Brito WMED, Fiaccadori FS, et al. Molecular characterization of VP6-encoding gene of group A human rotavirus samples from Central West region of Brazil. *J Med Virol* . 2008;80(11):2034-2039. doi:10.1002/jmv.21306

#### Figure legends

**Figure 1.** I-E genotype annual distribution of 34 unusual Rotavirus group A strains isolated from children aged [?]16 years hospitalized with acute gastroenteritis during 2007-2021.

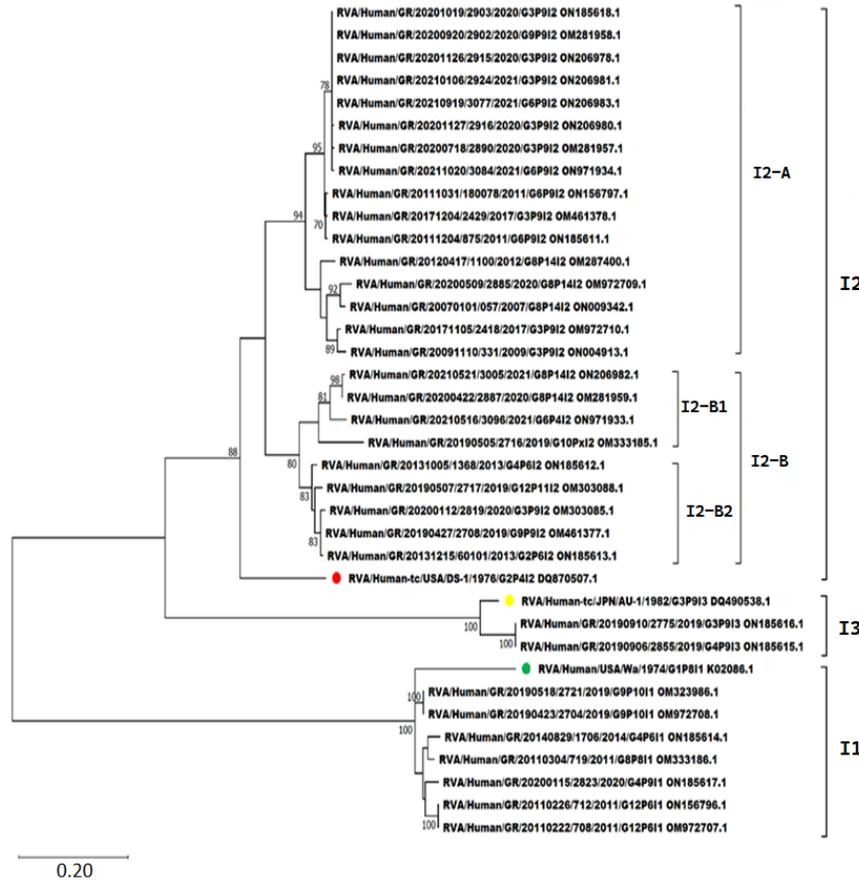
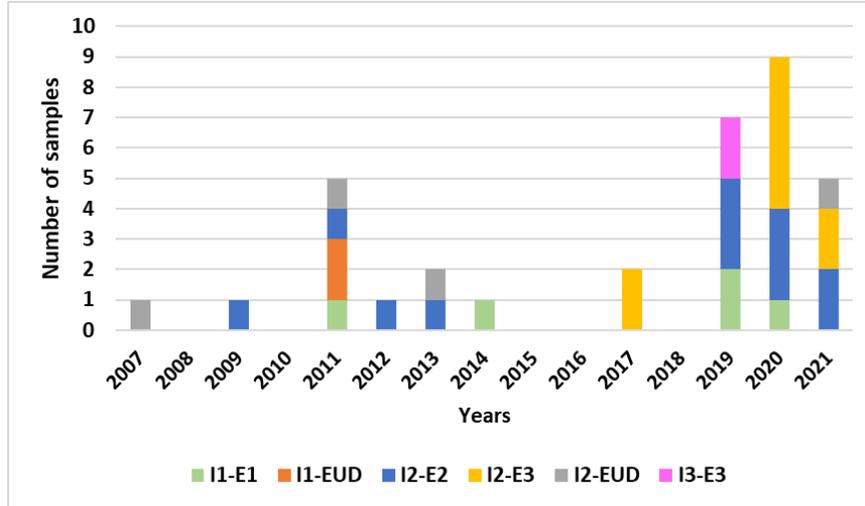
Abbreviation: EUD= unidentified E genotype.

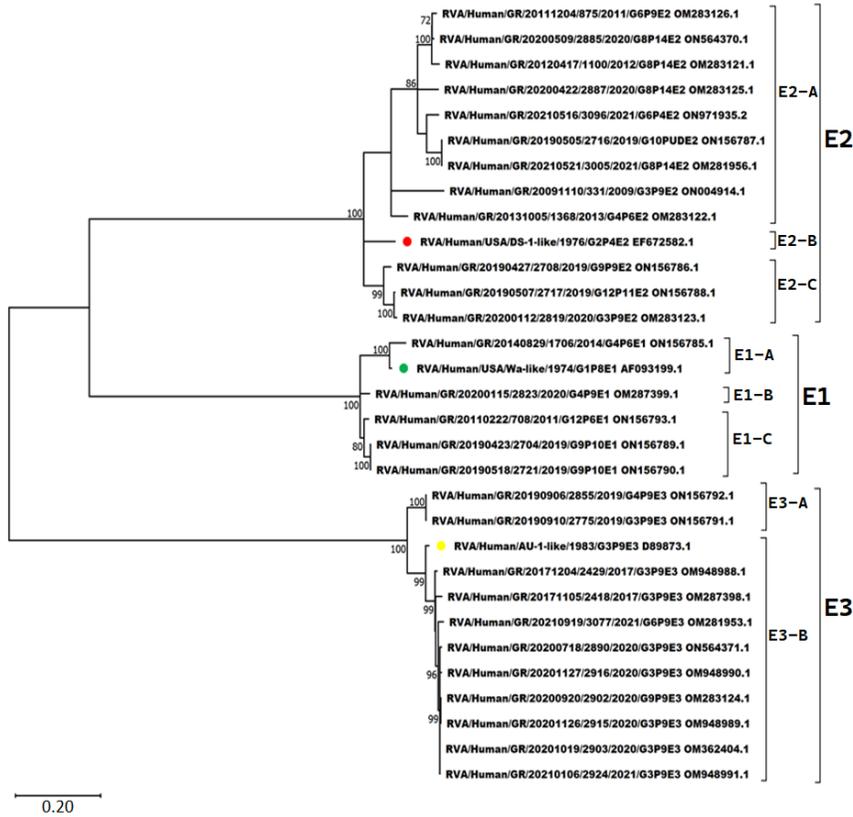
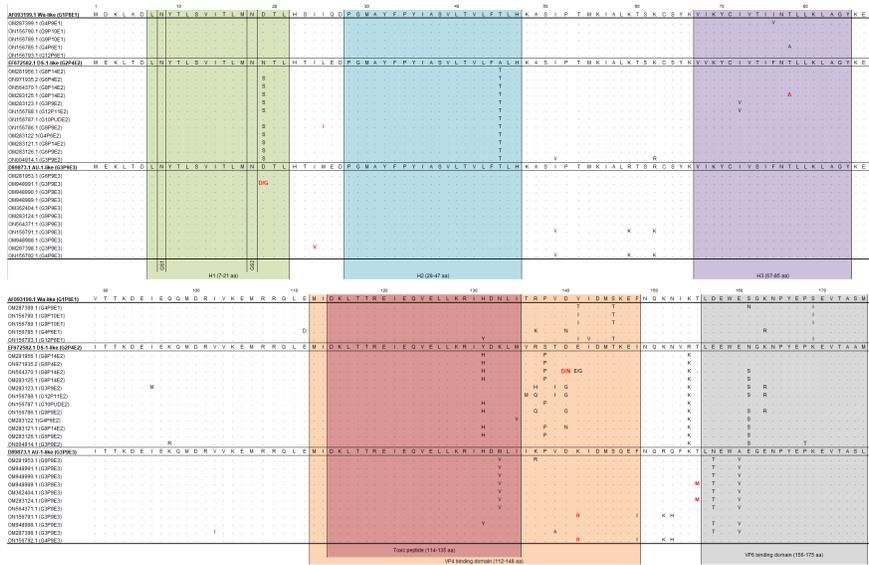
**Figure 2.** Phylogenetic tree of VP6 gene of unusual Rotavirus group A strains (n=34) circulating in Greece between 2007-2021. Reference strains are indicated by a colored circle. The tree was constructed using the Maximum Likelihood method and Tamura 3-parameter model.<sup>19</sup> A discrete Gamma distribution was used to model evolutionary rate differences among sites. Bootstrap values (1000 replicates) above 70% are shown. The scale bar indicates the branch length for 5% nucleotide differences. Evolutionary analysis was conducted with MEGA 11 software.<sup>20</sup>

**Figure 3** . Multiple sequence alignment of the NSP4 proteins (n=28) of unusual human Rotavirus group A strains. E1, E2 and E3 strains were compared to Wa (AF093199.1), DS-1 (EF672582.1), and AU-1 (D89873.1), respectively.

Abbreviations: H1= Hydrophobic domain 1, H2= Hydrophobic domain 2, H3= Hydrophobic domain 3, GS1= Glycosylation site 1 and GS2= Glycosylation site 2.

**Figure 4.** Phylogenetic tree of NSP4 gene of unusual RVA strains circulating in Greece between 2007-2021. Reference strains are indicated by a colored circle. The tree was constructed using the Maximum Likelihood method and Tamura 3-parameter model.<sup>19</sup> A discrete Gamma distribution was used to model evolutionary rate differences among sites. Bootstrap values (1000 replicates) above 70% are shown. The scale bar indicates the branch length for 5% nucleotide differences. Evolutionary analysis was conducted with MEGA 11 software.<sup>20</sup>





## Hosted file

2023-8-11 RVA RARE table.docx available at <https://authorea.com/users/651884/articles/659637-molecular-characterization-of-vp6-and-nsp4-genes-of-unusual-g-and-p-rotavirus-group-a-isolated-from-children-with-acute-gastroenteritis>