GENETIC DIVERSITY OF NOROVIRUS AMONG CHILDREN WITH GASTROENTERITIS IN SÃO PAULO STATE, BRAZIL

JULIANA GALERA CASTILHO¹, VERIDIANA MUNFORD¹, HUGO REIS RESQUE¹, ULYSSES FAGUNDES-NETO², JAN VINJÉ³ †, MARIA LÚCIA RÁCZ¹*

Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 05508-900, Brazil¹; UNIFESP-Escola Paulista de Medicina, Brazil²; Departament of Environmental Sciences and Engineering, School of Public Health, University of North Carolina, Chapel Hill, NC³; Instituto Pasteur, São Paulo, 01311-000⁴.

Running title: Genetic Diversity of Norovirus in Children in Brazil

Corresponding author: Mailing adress: Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Av. Prof. Lineu Prestes, 1374, São Paulo, SP, 05508-900, Brazil. Phone: 55-11-3091-7292. Fax: 55-11-30917354.

Email: mlracz@usp.br

†Present address: Centers for Disease Control and Prevention, Atlanta, GA
Abstract
Norovirus (NoV) is one of the most common causes of acute gastroenteritis in children and adults. To study the prevalence and genetic variability of NoV in children with acute gastroenteritis in São Paulo State, Brazil, we examined 234 stool samples from children with and without gastroenteritis during a 5-year period (1995-1999). NoV RNA was detected by RT-PCR and confirmed by DNA sequence analysis. We employed two different oligonucleotide primer sets targeting the 3’-end of the RNA polymerase gene (region B) as well a partial capsid region at the 3’-end of the VP1 gene (region D). In all, 78 (33.3%) of the samples tested positive for NoV and in region B, of the 66 strains sequenced, four (6.1%) belonged to GI and 52 (78.7%) to GII and in five (7.6%) samples a mixture of GI and GII genotypes was found. Phylogenetic analysis showed that the majority (40/66, 60.6%) of the strains belonged to genotype GII.4. The nucleotide sequence identity of three strains was lower than 77.9% when compared to a region B reference sequence database but showed 85.3 to 88.8% identity with GII.2 Melksham strain in region D, indicating the circulation of a possible recombinant NoV strain. One sample (GII.3) was sequenced only in region D. In conclusion, we have a total of 67 sequenced strains and this is the first report that describes the predominance of GII.4 NoV strains in children visiting the ambulatory of different hospitals in São Paulo State, Brazil and that mixtures of different strains can be found in individual samples including some possible new recombinant strains.
Introduction

Acute gastroenteritis remains a major public health problem worldwide, with more than 700 million estimated cases occurring annually in children less than 5 years of age. The mortality associated with gastroenteritis has been estimated to be 3.5 to 5 million per year (14, 32). Many different pathogens have been found in fecal samples of children with gastroenteritis, including parasites, bacteria and viruses (12). Among the viruses, rotavirus group A, enteric adenovirus, astrovirus and human caliciviruses (norovirus and sapovirus) are well established as etiologic agents of acute gastroenteritis (1) with norovirus (NoV) as the single most common cause of outbreaks of acute gastroenteritis in all age groups (5, 13). The frequency of NoVs in sporadic cases of gastroenteritis is not yet well defined (1).

NoVs, members of the family *Caliciviridae*, are non-enveloped viruses, 27-35 nm in diameter, that possess a positive-sense RNA genome of 7.5-7.7 kb. The genome encodes three open reading frames (ORF) including ORF1 coding for a large polyprotein that after translation is cleaved into nonstructural proteins including RNA-dependent RNA polymerase (POL), helicase and protease peptides. ORF2 encodes a major capsid protein (VP1) and ORF3 encodes for a minor capsid protein (VP2) (1, 13).

Human NoVs can be divided into at least two distinct genogroups (GI and GII) which based on the N-terminal region (region C) of VP1 can be further subdivided into at least 31 different genetic clusters or genotypes (18). When POL sequences are used to classify NoV into genotypes nucleotide similarities of more than 85% for GI strains or 90% for GII strains belonging to the same genetic cluster have been proposed (37). For NoV diagnostics two different genomic regions are most often used. Traditionally,
primers targeting the POL region of the genome (region A) are being used for NoV
detection (36, 39) whereas the Centers for Disease Control developed an assay targeting
the 3’-end of ORF1 (region B; 6). For genotyping of NoV strains two different regions of
the capsid gene have been employed (region C; 21, 25, 36) (region D; 37) that show
similar clustering of strains compared to the complete VP1 protein.

Limited studies have investigated the prevalence and epidemiology of NoV in
Brazil and no studies have been reported on the NoV strain diversity in children with
acute gastroenteritis in São Paulo State, Brazil. The knowledge of genomic diversity of
NoV in Brazil is limited to two reports, one in Ceará (27) and Rio de Janeiro (11).

Genetic diversity of RNA viruses can be generated by recombination. For this to
occur, different genotypes of NoVs need to coinfect the same cell. Different topologies of
the same strain when analyzing different parts of the genome (e.g., RNA polymerase
gene and capsid gene) are an indication of recombination. Jiang et al., 1999 has shown
that strain Arg320 is a recombinant strain of natural origin. Natural recombination has
also been suggested for strains Snow Mountain (16) and Rotterdam (37).

Due to the tremendous genetic diversity of NoVs, molecular characterization of
different genotypes is essential to better understand the epidemiology of strains
associated with pediatric gastroenteritis. Therefore, the objective of this study was to
identify and genetically characterize NoV in stools samples from children in São Paulo
State, Brazil.

MATERIAL AND METHODS

Specimen collection. A total of 234 fecal specimens were collected from children
(<3 years of age), in the State of São Paulo between August 1995 and November 1999;
94 samples were from people with diarrheal symptoms, 45 samples were from children with persistent diarrhea, 55 from controls and 40 samples with no clinical data available. Samples were collected at the Ambulatories of Hospital São Paulo, Hospital Municipal de Jundiaí, Hospital Darci Vargas and Gastro-Pediatric Unit of UNIFESP (University Federal of São Paulo). After collection, samples were shipped frozen on dry ice to the Virology laboratory and stored at –20°C. All samples had been previously tested for rotavirus group A, enteric adenovirus (30) and astrovirus (28).

**Nucleic acid extraction and RT-PCR.** Fecal suspensions (10%, wt/vol) were prepared in phosphate buffered saline (0.01M, [pH 7.4]). Viral RNA was extracted using TRIzol™ (Invitrogen) according to the manufacturers’ instructions and stored at -20°C. Reverse transcription-polymerase chain reaction (RT-PCR) specific for NoV was performed using two different assays, targeting regions B (3’-end of ORF1) and region D (3’-end of ORF2) of the viral genome. All samples were first screened by the region B primer set (6) and some genogroups II positive samples were further tested and characterized by the region D primer set (38). The oligonucleotide primer sequences are listed in Table 1. All assays were performed with appropriate positive and negative controls (ultra-pure water) and four separate laboratory rooms (extraction, master mix preparation; adding viral RNA, post PCR) were used to avoid cross-contamination.

**Nucleotide sequencing and phylogenetic analysis.** The 213 bp (region B) and 253 bp (GGII region D) RT-PCR products were purified using either the Concert Rapid PCR Purification System kit (Gibco BRL™) or the QIAquick PCR purification kit (Qiagen™). Sequencing reactions were carried out in both directions using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction I kit (Applied
Biosystems™). The resulting product was precipitated using the DyeEx™ 2.0 Spin kit (Qiagen) and the nucleotide sequence was determined using an automated sequencer (Applied Biosystems, ABI Prism 3100 DNA).

Consensus sequences were obtained using SeqManII (Lasergene, DNASTar Inc.) software. Multiple sequence alignments and phylogenetic analyses were performed by the Clustal V method and MegAlign (Lasergene, DNASTar Inc.) software, respectively.

**Reference strains.** The following reference strains (Genbank accession numbers between brackets) were included in the analysis: GI.1, Norwalk (M87661); GI.2, Southampton (L07418); GI.3, Desert Shield (U04469); GI.3b, Arkansas (AF414405); GI.4, Louisiana (AF414402); GI.5, Florida 1995 (AF414406); GI.6, Hesse (AF093797); GII.1, Hawaii (U07611); GII.1b, 314USA (AF414420); GII.2, Melksham (X81879); GII.3, Toronto (U02030); GII.4, Common Florida (AF080549); GII.4, Bristol (X76716); GII.5, Hillingdon (AJ277607); GII.5, Vermont (AF414423); GII.6, Seacroft (AJ277620); GII.6, Florida 1993 (AF414407); GII.7, Pennsylvannia (AF414409); GII.8, Idaho (AY054299); GII.8, Amsterdam (AF195848); GII.9, VA97207 (AY038599.1); Arg320, (AF190817); GII, Swine NoV Sw43/1997/JP, (AB074892); GIII, Jena (AJ011099); GIV.1 Cruise Ship, (AF414427).

**Nucleotide sequence accession numbers.** The nucleotide sequence data of the RNA polymerase gene and capsid gene have been submitted to GenBank and assigned accession numbers DQ386915 thru DQ386995.
RESULTS

RT-PCR with region B primers. NoV RNA was detected in 78 (33.3%) of 234 fecal samples. NoV positive results were: 34/94 (36.2%) in acute diarrhea samples, 12/45 in persistent diarrhea samples (26.7%); 20/55 (36.4%) control samples and 12/40 (30.0%) in samples with no information.

Of these 78 NoV positive samples, 3 (1.8%), 5 (2.1%) and 26 (11.1%) also tested positive for rotavirus, adenovirus and astrovirus, respectively (data not shown).

RT-PCR products of 66 NoV strains were sequenced from 61 samples. From the remaining 17 samples we did not have enough fecal material for the complete sequencing reaction and therefore, strains could not be genotyped although some small fragments of the genome were sequenced and were confirmed as NoV.

We found 58 GII strains, 52 in samples with a single strain and 6 in samples with mixtures of two different strains. Eight GI strains were found, 4 in samples with single strains and 4 in mixtures.

In 52 samples with single GII strains, 40 (76.9%) could be typed as GII.4, four (7.7%) had sequences similar to the recombinant Arg320 strain (21) and 5 (9.6%) strains that could not be differentiated in region B (GII.6/7/8 strains), as the nucleotide sequence of this region does not discriminates between these three genotypes (Table 2; Figure 1).

Three strains -ICB1521, ICB1915 and ICB2230- (5.8%) had less than 77.9% nucleotide sequence similarity when compared to the known GII reference strains indicating they may present a new genetic cluster (Figure 1). Four single GI strains were detected, one GI.3b and three GI.4.
The sequence results of five region B RT-PCR products indicated the presence of mixtures of two different genotypes in each sample. After analyzing the GI and GII region B sequence reactions separately, we found that two samples contained both GII.4 and GI.3b mixtures, two samples contained GII.4 and GI.4 strains and one sample contained a GII.4 and a ICB2230-like GII strain (Figure 2).

**RT-PCR with region D primers.** Six samples that could not be assigned to a defined genotype, three samples showing GII sequences in region B with identity with the recombinant strain Arg320, one sample not sequenced in region B, and 5 GII.4 strains were analyzed by region D RT-PCR. Of the three unresolved (GII6/7/8) region B strains, one strain (ICB1963) showed 91.8% nucleotide identity with the GII.6 reference strain (Seacroft) and 2 samples showed high sequence identity with the GII.3 reference strain (Toronto) in region D (Table 2; Figure 3). Three strains (ICB1529, ICB1912, ICB1200) showing region B identity to the recombinant Arg320 strain clustered with GII3 prototype strain Toronto in region D.

The three (ICB1521, ICB1915 and ICB2230) unresolved region B sequences that showed <77.9% nucleotide sequence identity compared to the region B sequence database at CDC showed 85.3 to 88.8% nucleotide sequence identity and 92.9 to 100.0% amino acid sequence identity with GII.2 prototype strain Melksham in region D (Figure 3). One sample (ICB1443), for which no region B sequence was obtained, was identified as GII.3.

**DISCUSSION**

RT-PCR has become a routine diagnostic method for the detection of NoV RNA in clinical laboratories (1). Application of RT-PCR based methods to screen stool
specimens has not only shown that the overwhelming majority of outbreaks of acute
gastroenteritis are attributable to NoVs but also that these viruses are the cause of
numerous cases of sporadic gastroenteritis (2, 3, 4, 7-9, 15, 20, 22, 26, 27, 33, 35).

Our results show that NoVs are commonly found in sporadic cases of
gastroenteritis as well as in controls in children in São Paulo State, Brazil. The overall
frequency (33.3%) of NoV found in our study is substantial higher compared to other
studies which report a prevalence in the range of 6-19% (3, 4, 7, 8, 15, 20, 22, 27).
Circulation of more pathogenic strains, different requirements for hospital admission, and
the very low income of the population under study could perhaps explain the high
frequency of NoVs. In addition, the presence of NoVs as opportunistic agent in infections
with other agents are other conditions that may need further study. Another striking
finding is the high proportion (36.4%) of NoV positives in children without diarrheal
symptoms which might be due to the fact that NoV RNA can be detected up to 3 weeks
after the onset of illness. Also, children could have been infected with different genotypes
of norovirus in the last 6 to 12 months, which offers immunological protection against
disease but not infection. Asymptomatic infection is common in children under 5 years of
age.

The predominance of NoVs GII strains detected in this study is in agreement with
previous reports of a higher prevalence of GII strains over GI in outbreaks as well as
studies on sporadic gastroenteritis (3, 4, 8, 9, 11, 20, 23, 24, 25, 26, 33, 35). The reason
for this is unknown, although differences in biological properties, such as virulence,
routes of transmission or stability of the virus in the environment, are possible
explanations (4).
Genotyping of the NoV strains detected in our study showed that 40/67 (59.7%) of the strains belonged to GII.4 genotype. Viruses of this genotype caused 60 outbreaks in geographically distant locations within the USA and were identified, by sequence comparisons, in additional 7 countries on 5 continents during a same period (25). Strains of this genotype were detected in the USA in April 1995, then in Brazil, Canada, Australia, and the Netherlands in late 1995, Australia, the Netherlands and China in 1996 and in Germany in 1997, suggesting that the strain was circulating globally during this time-period (25). Recent studies on sporadic cases of pediatric acute gastroenteritis have also shown the predominance of the GII.4 genotype in several countries (2, 4, 10, 15, 20, 22). In Ceará (Northeastern Region), Brazil, strains similar to the GII.4 Lordsdale strain have been reported as early as 1990 (27). Gallimore et al. (11) examined outbreaks of acute gastroenteritis in a children’s day care facility in Rio de Janeiro (Southeast Region), from 1996 to 1998, and showed that the majority of strains belonged to GII.4. Our study confirms previous findings in our country that GII.4 are the most prevalent NoV strains in this time period.

Because first generation NoV RT-PCR assays targeted a small region of the POL gene, many reports on NoV genotyping have been based upon sequences amplified from this region (38). However, because VP1 is responsible for the differences in antigenicity, the complete VP1 protein has been proposed as the gold standard for classification of NoVs into genogroups and genotypes (37; 40). Amplification of a small region of the capsid gene (region C or region D) has been shown to accurately classify NoVs comparable to the complete VP1 (38). Therefore, testing fecal specimens for NoV by conventional RT-PCR using region B primers (6) followed by genotyping of strains by
region C or region D seems to be an adequate strategy for sensitive detection and reliable
characterization of NoV strains (38). We detected four strains that had POL and capsid
sequences similar to the Arg320 recombinant strain (17) Similar strains have also been
identified in Argentina in 1995 (17) and were also identified in children with diarrhea in
Argentina in the same time period our samples were collected (23).

Region B primers used in this work do not discriminate between GII.6, GII.7 and
GII.8 genotypes because the genomic region targeted by these primers is conserved (6).
Three of the five samples identified as GII.6/7/8 in region could be classified as GII.6 (1
strain) and GII.3 (2 strains). Differences in tree topology when different regions of the
genome are analyzed may be an indication of a recombinant strain with most likely the
conserved ORF1-ORF2 junction region as the cross-over site (37). The fact that we found
several samples that contained genetically different strains illustrates that the conditions
for recombination between strains exist in the studied population.

In region B, three strains showed less than 77.9% identity with representative
strains of all known GI and GII strains, however, region D sequences of these strains
showed nucleotide sequence identities from 85.3-88.8% and amino acid identities from
92.9-100% with GII.2 Melksham strain. These results discard the possibility of a new
genotype, but reinforce the possibility of a new recombinant, containing the GII.2
genotype capsid protein and a novel polymerase sequence of a unknown parent strain.

Mixed infections with different genotypes of NoVs were previously described in
outbreaks associated with contaminated oysters or water (34), but to our knowledge,
mixed infections have not been previously described in sporadic cases of NoV
gastroenteritis. In five of our samples, two different GI and GII genotypes were found.
using all four region B primers in separate sequencing reactions. All genotypes that were
detected in the mixtures were also found as individual strains in other fecal samples in
this study.

In conclusion, we found a high frequency of NoV strains in stool samples
collected from children in São Paulo State, Brazil with and without acute gastroenteritis.
GII.4 strains was the predominant strain detected and different potential recombinant
strains as well as mixed infections that are required to generate such strains.

ACKNOWLEDGMENTS

This work was supported by FAPESP grant 2001/07298-8. J.G. Castilho and M.L.
Rácz have CNPq scholarships. We also thank Dr. Paulo Brandão for critical review of the
manuscript.
REFERENCES


### TABLE 1. Oligonucleotide primers used in this study

<table>
<thead>
<tr>
<th>Region</th>
<th>Name</th>
<th>Sequence (5’- 3’)</th>
<th>Sense</th>
<th>Positiona</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>MON 431</td>
<td>tgg acl agR ggl ccY aaY</td>
<td>+</td>
<td>5093 - 5113</td>
</tr>
<tr>
<td>B</td>
<td>MON 432</td>
<td>tgg acl cgY ggl ccY aaY</td>
<td>+</td>
<td>5093 - 5113</td>
</tr>
<tr>
<td>B</td>
<td>MON 433</td>
<td>gaa Yct cat cca Yct gaa</td>
<td>-</td>
<td>5284 - 5305</td>
</tr>
<tr>
<td>B</td>
<td>MON 434</td>
<td>gaa Scg cat cca Rcg gaa</td>
<td>-</td>
<td>5284 - 5305</td>
</tr>
<tr>
<td>D</td>
<td>CapD3</td>
<td>tgY ctY Itl ccH caR caa</td>
<td>+</td>
<td>6432 - 6453</td>
</tr>
<tr>
<td>D</td>
<td>CapC</td>
<td>cct tYc caK Wtc cca Ygg</td>
<td>-</td>
<td>6666 - 6684</td>
</tr>
</tbody>
</table>

International Union of Biochemistry ambiguity codes: I = Inosine; H = (A/C/T); Y = pYrimidine (C/T); R = puRine (A/G); S = Strong (C/G); K = Keto (G-T); W = Weak (A-T)

a Equivalent position within the NV genomic sequence (M87661)
TABLE 2. Genotype classification by sequencing of regions B (polymerase) and region D (capsid) of 62 strains found in children from São Paulo, Brazil.

<table>
<thead>
<tr>
<th>Region B</th>
<th>GII.2</th>
<th>GII.3</th>
<th>GII.4</th>
<th>GII.6</th>
<th>Not done</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI.3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GI.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>GII.4</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>GII.6/7/8</td>
<td>2</td>
<td>1</td>
<td></td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><em>GII.rec</em></td>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Undetermined</td>
<td>3</td>
<td>1</td>
<td></td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>52</td>
<td>67</td>
</tr>
</tbody>
</table>

*GII.rec = similar to Arg320 strain*
FIG. 1. Phylogenetic tree of region B nucleotide sequences from 56 norovirus samples identified by number detected in children in São Paulo, SP, Brazil and of norovirus GI and GII reference strains. The length of branches represents the distance between sequence pairs and the dendrogram was constructed by the Clustal V method, using the MegAlign program (LaserGene/DNAStar). The units at the bottom of the tree indicate the number of substitution events.
FIG. 2. Genetic relatedness of the region B nucleotide sequences from 5 samples (ICB1242, ICB1969, ICB1987, ICB1346, ICB1436) containing a mixture (c1 and c2) of norovirus genotypes and norovirus GI and GII reference strains. The length of branches represents the distance between sequence pairs and the dendogram was constructed by the Clustal V method, using the MegAlign program (LaserGene/DNAStar). The units at the bottom of the tree indicate the number of substitution events.
FIG. 3. Phylogenetic tree of region D nucleotide sequences from 15 norovirus strains detected in this study (ICB1200, ICB1207, ICB1244, ICB1443, ICB1521, ICB1529, ICB1912, ICB1913, ICB1915, ICB1963, ICB1977, ICB1983, ICB2230, ICB2784, and ICB2785) GI and GII norovirus reference strains. The length of branches represents the distance between sequence pairs and the dendogram was constructed by the Clustal V method, using the MegAlign program (LaserGene/DNAStar). The units at the bottom of the tree indicate the number of substitution events.