Transmission Events within Outbreaks of Gastroenteritis Determined through Analysis of Nucleotide Sequences of the P2 Domain of Genogroup II Noroviruses

Jacqueline Xerry,* Chris I. Gallimore, Miren Iturriza-Gómar, David J. Allen, and Jim J. Gray

Enteric Virus Unit, Virus Reference Department, Centre for Infections, Health Protection Agency, Colindale, London, United Kingdom

Received 20 November 2007/New Returned for modification 9 January 2008/Accepted 14 January 2008

Noroviruses (NoVs) are the commonest cause of outbreaks of acute gastroenteritis worldwide (17) and are members of the Caliciviridae family (13). NoV outbreaks are frequently associated with semiclosed or closed institutions such as hospitals and homes for the elderly (11, 15). Outbreaks also occur in other settings, including eating establishments, cruise ships, concert halls (4, 14, 26), and schools (19). Transmission of NoVs is usually person-to-person (18), although food (8, 20), including shellfish and water (9, 21), and environmental or airborne contamination have all been implicated in transmission (11, 24).

The genomic diversity of NoVs includes three genogroups (GI, GII, and GIV), 16 genotypes, and a number of provisional genotypes which have yet to be formally agreed upon. Diversity among NoVs is maintained through the accumulation of point mutations associated with the error-prone nature of RNA replication and genetic recombination involving the exchange of sequences between two related RNA viruses. Several features of RNA viruses (antigenic variation, genotypic diversity, immune escape) have been attributed to the lack of proofreading associated with semiclosed or closed institutions such as hospitals and homes for the elderly (11, 15). Outbreaks also occur in other settings, including eating establishments, cruise ships, concert halls (4, 14, 26), and schools (19). Transmission of NoVs is usually person-to-person (18), although food (8, 20), including shellfish and water (9, 21), and environmental or airborne contamination have all been implicated in transmission (11, 24).

The error-prone nature of template copying by RNA polymerases during virus replication (3) can lead to conserved point mutations, either silent or resulting in amino acid substitutions, but a lack of diversity within genes encoding the RNA-dependent RNA polymerases of the domain of the shell(s) of the capsid within single common genotypes makes it difficult to track outbreaks through comparisons of sequence data derived from these genomic regions (22). The NoV capsid is divided into the S (shell) domain which constitutes the 5' end (amino acids [aa] 1 to 225) and the P (protruding) domain (aa 226 to 530) (28). The S domain contains the more highly conserved regions of the capsid and lends itself to the design of consensus primers which can amplify a wide range of genotypes within a genogroup. The P domain can be further subdivided into two subdomains, P1 and P2. The gene encoding the P2 domain contains the greatest sequence variation and corresponds to the portion of the capsid that is most exposed. Mutations in this region may have a significant effect on virus receptor binding and the host immune response to viral infection.

It is not possible to design a single primer pair which is sufficiently reactive and capable of amplifying all genotypes within a genogroup due to the diversity within the P2 domain. Genotype-specific primers are required in order to amplify this region of NoV genomes of multiple genotypes.

This study was designed to determine the potential of nucleotide sequences, derived from the P2 domain, to identify transmission events. Identification of genotype through nucleic acid sequence analysis of the S domain followed by genotype-specific P2 domain amplification and sequence analysis was able to separate or link patients within and among outbreaks more reliably than analysis of the S domain alone.

MATERIALS AND METHODS

Fecal specimens. A total of 44 fecal specimens were collected in 2006 and 2007 from NoV outbreaks which occurred in wards in two hospitals in the same National Health Service (NHS) trust in London. In addition, 17 fecal specimens from NoV outbreaks occurring throughout the year and in different geographical areas of England and Wales were included for comparison. A total of 23 fecal samples collected from two separate NoV outbreaks on the same cruise ship sailing in the Pacific Ocean in March and June 2007 were also analyzed.

Outbreaks. Outbreaks were defined as including two or more cases of gastroenteritis linked in place and time. A new outbreak was arbitrarily defined as occurring at least 7 days after the last case in a previous outbreak or as occurring in a different patient care unit such as a ward or hospital.
RNA extraction of fecal specimens and reverse transcription assay. Small batches of fecal specimens were prepared as described previously using the guanidinium thiocyanate-silica method (10) or in larger batches using a Roche total nucleic acid extraction kit (Roche Diagnostics Ltd., Burgess Hill, United Kingdom) according to manufacturer’s instructions and a Roche MagNAPure automated extractor (7). Briefly, for the guanidinium thiocyanate-silica procedure, a 200-ul fecal extract in phosphate-buffered saline was added to 1 ml of L6 buffer (Severn Biotech, United Kingdom) and 20 uL of silica, and the silica pellet was then washed with L2 buffer (Severn Biotech, United Kingdom)–70% ethanol–acetone. RNA was eluted in 40 uL of nuclease-free water (Promega, Southampton, United Kingdom), and cDNA was prepared as described previously using PdN6 random hexamers (GE Healthcare, Bucks, United Kingdom) and Moloney murine leukemia virus RTase (Invitrogen, Paisley, United Kingdom) (6).

NoV GII consensus primers. For initial genotyping, GII strains were amplified using a single-round PCR and/or a heminested PCR to increase the amplicon concentration (7, 16). PCR mix and cycling conditions were as described previously (5).

NoV GII genotype-specific P2 domain primers. PCR primers for the amplification of a region of the gene encoding the VP1 and containing the P2 domain of NoV GII genotypes 1 to 8 (12) were designed from alignments of complete Orf2 nucleotide sequence data. The forward primers correspond to nucleotide (nt) positions 5661 to 5682 and the reverse primers to nt positions 6432 to 6451 on the GII-4 strain Lordsdale/1995/UK (GenBank X86557), generating amplicons of between 785 and 818 bp (Table 1). The gene fragment encoding the P2 domain of GII strains corresponds to nt 5880 to 6335 on the GII-4 strain Lordsdale/1995/UK.

Detection and sequencing. P2 domain PCR amplicons were examined by gel electrophoresis in 2% agarose gels (MP agarose; Roche Diagnostics Ltd.), stained with ethidium bromide (0.5 g/ml), and photographed using a Bio-Rad GelDoc system (Bio-Rad, Hemel Hempstead, United Kingdom). DNA was purified using a commercially available spin column PCR purification kit (Geneclean; QbioGene) and was sequenced in both directions using NoV GII genotype P2 domain-specific primers (see Table 1), a CEQ Dye cycle sequencing Quick Start kit according to the instructions of the manufacturer (Beckman Coulter, High Wycombe, United Kingdom), and a Beckman Coulter CEQ8000 capillary sequencer.

Generation of consensus sequences and pairwise alignments of the interprimer regions (P2 GII-genotype F/P2 GII-genotype R) of sequences was performed

**TABLE 1. Norovirus genogroup II genotype-specific primers for amplifying a region encompassing the P2 domain**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Primer name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Primer (5’–3’)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Annealing temp (°C)</th>
<th>Amplicon size (bp)</th>
<th>P2 domain size (bp)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>GII-1</td>
<td>P2GI-1 F</td>
<td>GATGATGTNTTTACAGTTTCTT</td>
<td>40</td>
<td>779</td>
<td>445</td>
</tr>
<tr>
<td></td>
<td>P2GI-1 R</td>
<td>CAYTCYTGGGCAAGAACA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-2</td>
<td>P2GI-2 F</td>
<td>GATGACGCTTCTTACAGTTTCTT</td>
<td>45</td>
<td>800</td>
<td>466</td>
</tr>
<tr>
<td></td>
<td>P2GI-2 R</td>
<td>CACTCTTGTCGACGATAGACGCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-3</td>
<td>P2GI-3 F</td>
<td>GATGATTTTTACAGTTTCTT</td>
<td>42</td>
<td>818</td>
<td>484</td>
</tr>
<tr>
<td></td>
<td>P2GI-3 R</td>
<td>CATTCTGGGGGACGAGGCAAAGCCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-4</td>
<td>P2GI-4 F</td>
<td>GANGTGTCCTACAGTTTCTT</td>
<td>45</td>
<td>794</td>
<td>459</td>
</tr>
<tr>
<td></td>
<td>P2GI-4 R</td>
<td>CATTCTGGGGGAGGAGTAGAGCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-5</td>
<td>P2GI-5 F</td>
<td>GACGACTTTTACCGGTCTTCTT</td>
<td>48</td>
<td>794</td>
<td>459</td>
</tr>
<tr>
<td></td>
<td>P2GI-5 R</td>
<td>CACCTCTGAGGCACGACAGACA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-6</td>
<td>P2GI-6 F</td>
<td>GACGACTTTTACCGGTCTTCTT</td>
<td>40</td>
<td>827</td>
<td>492</td>
</tr>
<tr>
<td></td>
<td>P2GI-6 R</td>
<td>CATTCTGGGGGAGGAGTAGAGCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-7</td>
<td>P2GI-7 F</td>
<td>GACGACTTTTACCGGTCTTCTT</td>
<td>45</td>
<td>794</td>
<td>447</td>
</tr>
<tr>
<td></td>
<td>P2GI-7 R</td>
<td>CACCTCTGAGGCACGACAGACA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GII-8</td>
<td>P2GI-8 F</td>
<td>GATGATTTTTTACAGTTTCTT</td>
<td>40</td>
<td>785</td>
<td>450</td>
</tr>
<tr>
<td></td>
<td>P2GI-8 R</td>
<td>CATTCTGGGGGACGAGGCAAAGCCA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> F, forward; R, reverse. The forward primers correspond to nt positions 5661 to 5682 and the reverse primers to nt positions 6432 to 6451 on the GII-4 strain Lordsdale/1995/UK (GenBank X86557).

<sup>b</sup> Y = C or T; R = A or G; N = C or G or T or A.

<sup>c</sup> The P2 domain region of GII strains corresponds to nt 5880 to 6335 on the GII-4 strain Lordsdale/1995/UK.

**FIG. 1. Phylogenetic analysis of the P2 domain of the GII genotypes.** P2 domain sequences from study strains are available from the corresponding author. P2 domain GII-1 to GII-8 primers sets were used (see Table 1).
initially using Genebuilder and Clustal in Bionumerics version 3.5 (Applied Maths, Kortrijk, Belgium). Sequence analysis was performed using the 459-bp region of the P2 domain region.

**RESULTS**

NoV genotyping of outbreaks. A total of 44 NoV strains from 14 outbreaks in the London NHS Trust were identified as GII-4. GII-4 strains from other outbreaks in England and Wales were used for comparative analysis. The NoV responsible for the March 2007 cruise ship outbreak was a GII-4 strain, whereas the outbreak in June 2007 was associated with a GII-5 strain.

Norovirus genogroup II genotyping validation. Fecal samples for NoV genogroup II, genotypes 1 to 8, were all successfully amplified using the genotype-specific primers designed to provide amplicons encompassing the P2 domain (Table 1). Phylogenetic analysis of the P2 domain sequences differentiated between genotypes (Fig. 1).

Diversity within and among hospital GII-4 NoV outbreaks. Between 20 December 2006 and 5 March 2007, a total of eight
presumptive outbreaks were identified in four wards within hospital 1 and six outbreaks in six wards within hospital 2. The nucleotide similarity among strains and within the P2 domain was 100% within each outbreak for which multiple samples were available (Fig. 2).

Outbreaks in hospital 1 occurred in December 2006 and January, February, and March 2007, whereas outbreaks in hospital 2 were confined to January 2007. Sequencing and analysis of the gene fragment encoding the P2 region of the GII-4 strains identified three genetic clusters (Fig. 3). Cluster 1 was found in three wards in hospital 1 in December 2006, cluster 2 in five wards in hospital 2 in January 2007, and cluster 3 in three wards in hospital 1 in January, February, and March 2007 and in one ward in hospital 2 in January 2007.

Outbreaks associated with GII-4 strains occurring during the same period, but from different geographical regions of England, were 89.2% to 99.4% similar to cluster 1, 88.1% to 97.8% to cluster 2, and 89.2% to 98.2% to cluster 3 at the nucleotide level. Two strains from the northeast region of England collected in December 2006 and January 2007 shared 100% similarity (Fig. 2). Although these strains were collected in the same region, they were from outbreaks occurring in different hospitals. Analysis of 18 strains, representative of all outbreaks, showed less than 100% similarity at the nucleotide level among strains from different geographical regions and at different times of the year (Fig. 4).

Diversity within cruise ship outbreaks. Fecal specimens S1/2007 to S16/2007 and S1-1/2007 to S1-16/2007 were from separate NoV outbreaks on the same cruise ship sailing in the Pacific Ocean in March and June 2007, respectively. All strains in the March outbreak were GII-4. P2 domain sequences showed seven different genetic clusters with one or more samples in each cluster. Cluster 1 was represented by S-8, cluster 2 by S-4, cluster 3 by S-9, cluster 4 by S-5, cluster 5 by S-2 and

FIG. 3. GII-4 NoV outbreaks reported between December 2006 and March 2007 in two neighboring hospitals within the same NHS trust. Three clusters are shown as well as the ward designations and outbreak numbers related to strains shown in Fig. 2.

FIG. 4. Similarity among NoV outbreaks reported between December 2006 and September 2007 from different regions of England and Wales. NE, northeast; NW, northwest; SW, southwest. P2 domain GII-4 primer sets were used (see Table 1).
S-3, cluster 6 by S-1, S-6, and S-7, and cluster 7 by S-10, S-12, and S-14 to S-16, with 86 to 96% homology between clusters. However, strains in the June outbreak were GII-5 and sequences derived from the gene encoding the P2 domain showed 100% homology among the strains (Fig. 5).

DISCUSSION

Noroviruses are spread by contact with an infected individual, by ingestion of contaminated food or water, or by contact with contaminated environmental surfaces. Outbreaks may result from multiple introductions, and onward transmission is frequent, resulting in high attack rates. Methods that can determine transmission events accurately can help to identify common or significant transmission routes and focus interventions on reducing onward transmission, shortening the duration of the outbreak and reducing the number of individuals infected.

The ubiquitous nature of noroviruses and their association with symptomatic and asymptomatic infection (1) and the global distribution of a predominant strain, GII-4, make it difficult to follow transmission events. Some success in plotting transmission events has been reported, and onward transmission is frequent, resulting in high attack rates. Methods that can determine transmission events accurately can help to identify common or significant transmission routes and focus interventions on reducing onward transmission, shortening the duration of the outbreak and reducing the number of individuals infected.

The hypervariable region of the genome encoding P2 was chosen for analysis, as this was the least conserved, whereas whole-capsid sequencing is likely to dilute differences among strains through the inclusion of conserved sequences. Also, the practicality of sequencing up to 827 bp compared with 3,255 bp allows this method to be used for the routine surveillance of transmission events. Strains with identities of 100% were regarded as having been derived from a common source, whereas strains with one or more mutations in the region sequenced were regarded as representing unrelated transmission events. Only one sample was obtained from each patient, and the short duration of NoV illness would make it unlikely that more than one sample would be submitted.

This method allowed the identification of outbreaks linked within a hospital but involving different wards and between hospitals within the same NHS Trust, where patients were likely to be transferred between NHS care units. Also, it allowed outbreaks to be better defined in terms of duration and

FIG. 5. Dendrogram of P2 domain nucleotide sequences and similarity matrix of two NoV outbreaks on the same cruise ship sailing in the Pacific Ocean. The outbreaks in 2007 were S in March (a) and S1 in June (b). P2 domain GII-4 and GII-5 primer sets were used (see Table 1).
location. Previously, outbreak definitions have suggested that a new outbreak has occurred when there is a break of 7 days between the last case in the first outbreak and the first case of the second outbreak, but this takes no account of asymptomatic spread among immune staff members or patients. With no clear history of the movement of ill patients or staff members between wards or hospitals, outbreaks are usually defined as occurring within a ward or care unit. Based on this study, analysis would suggest that in hospital 1, only two outbreaks occurred, one in December 2006, lasting for 16 days and involving three wards, that was initially thought to represent four separate outbreaks, and a second, of longer duration, initially recorded as five outbreaks in four wards, lasting from 18 January to 6 March 2007. The long duration of this outbreak may have been linked to the constant introduction of susceptible individuals into this environment. Hospital 2 suffered two outbreaks in January 2007, the first initially recorded as five separate outbreaks in five wards and lasting 7 days and the second outbreak as caused by a GII-4 strain associated with an outbreak occurring at the same time in a sister hospital within the same NHS Trust. Although two outbreaks occurred simultaneously in hospital 2, sequence analysis allowed discrimination between the GII-4 strains causing these outbreaks and clearly identified two separate introductions.

GII-4 strains from outbreaks occurring at the same time but in different geographical regions of England were clearly different from those isolated in the London hospitals, and although the majority of the strains were different from each other, two strains from two different hospitals located in the same geographical region were identical. This may suggest movement of infected individuals or staff members between the two hospitals. The analysis of nucleotide similarity among representative strains of all outbreaks confirmed that among unrelated outbreaks, similarity was less than 100%.

Noroviruses can be introduced onto cruise ships by symptomatic or asymptomatic embarking passengers, by infected crew members, by contaminated food stuffs, or, during the cruise, by passengers returning from daily excursions. The numerous sources of infectious virus often result in outbreaks associated with more than one introduction and possibly occurring associated by more than one route. Different interventions may be used to prevent the introduction of NoVs by these routes, but such interventions can only be successful when the routes are accurately identified. Again, as with the hospital outbreaks, the defining of an outbreak may suggest a point source or person-to-person transmission of an outbreak strain but cannot account for multiple introductions of closely related strains.

Two NoV outbreaks occurred on the same cruise ship but at different times and during different cruises. As both were associated with different NoV genotypes, they were easily identified as separate introductions. However, analysis of the P2 regions of strains from the outbreak that occurred during the first cruise in which a GII-4 outbreak was identified revealed the introduction of seven different strains of GII-4. This may not be an unusual event, as GII-4 was the predominant global strain at the time. Gathering passengers from many locations will allow the introduction of many virus variants (23), some excreted from asymptotically infected individuals (1). Alternatively, foodstuffs, contaminated with feces or raw sewage, are often associated with outbreaks involving multiple strains of a pathogen, but multiple pathogens are also often found in this context (5, 8). Only GII-4 NoV strains were associated with this outbreak.

The second outbreak on the cruise ship was associated with a GII-5 strain, and sequences of the gene encoding the P2 domain were analyzed using nine samples collected during this outbreak. All samples showed 100% identity at the nucleotide level, indicating a point source outbreak followed by person-to-person spread.

The analysis of genes encoding the P2 domain provides a powerful tool for the tracking of outbreaks. Interventions are often hampered by an inability to understand transmission events, which may be complicated by multiple introductions of closely related viruses into semiclosed communities. This method will provide insights into food- and water-borne transmission as well as transmission through contact with contaminated environmental surfaces.

REFERENCES

17. Kaplan, J. E., G. W. Gary, R. C. Baron, N. Singh, L. B. Schonberger, R.


