Molecular Epidemiology of “Norwalk-like Viruses” in Outbreaks of Gastroenteritis in the United States

Rebecca L. Fankhauser, Jacqueline S. Noel, Stephan S. Monroe, Tamie Ando, and Roger I. Glass

Fecal specimens from 90 outbreaks of nonbacterial gastroenteritis reported to 33 state health departments from January 1996 to June 1997 were examined to determine the importance of and to characterize “Norwalk-like viruses” (NLVs) in these outbreaks. NLVs were detected by reverse transcription–polymerase chain reaction in specimens from 86 (96%) of 90 outbreaks. Outbreaks were most frequent in nursing homes and hospitals (43%), followed by restaurants or events with catered meals (26%); consumption of contaminated food was the most commonly identified mode of transmission (37%). Nucleotide sequence analysis showed great diversity between strains but also provided evidence indicating the emergence of a common, predominant strain. The application of improved molecular techniques to detect NLVs demonstrates that most outbreaks of nonbacterial gastroenteritis in the United States appear to be associated with these viruses and that sequence analysis is a robust tool to help link or differentiate these outbreaks.

“Norwalk-like viruses” (NLVs), also called small round-structured viruses, are a group of genetically diverse, single-stranded RNA viruses belonging to a newly proposed genus in the family Caliciviridae that are recognized as an important cause of outbreaks of acute nonbacterial gastroenteritis [1–3]. Such outbreaks have been reported in a variety of settings, including nursing homes [4–6], hospitals [5, 7, 8], cruise ships [9–12], schools and universities [13, 14], and restaurants and events with catered meals [15–17]. Transmission of the viruses has been documented by contaminated food [8, 15, 18, 19], especially oysters [20–24] and water [25–28], and by person-to-person contact [11, 29, 30].

NLVs can be divided into two distinct genogroups, genogroup I (GI) and genogroup II (GII) [31], each of which can be further divided into 4 and 6 clusters, respectively. GI includes the Norwalk virus, Southampton virus, cruise ship virus, and Desert Shield virus clusters. GII includes the Gwynedd virus, Toronto virus, Lordsdale virus, Snow Mountain agent, and White River and Hawaii virus clusters [32]. Several reports have described the predominant circulation of strains in a particular cluster [5, 33, 34], and Vinje et al. [5] recently reported both a single predominant strain and a shift in predominant clusters over time. The development of sensitive reverse transcription–polymerase chain reaction (RT-PCR) has improved the detection rates for NLVs in outbreaks of gastroenteritis to as high as 91% in a study conducted in The Netherlands [6]. Sequence analysis not only allows for examination of circulating strain types but also has been used to aid in epidemiologic investigations by linking or differentiating outbreaks [35]. Whereas there have been studies examining the role and molecular characteristics of NLVs in outbreaks of nonbacterial gastroenteritis in other countries [5, 6], this analysis has not been done previously in the United States.

In this study, we examined a collection of fecal and emesis specimens from 90 outbreaks of nonbacterial gastroenteritis reported to the Centers for Disease Control and Prevention (CDC) by state and local health departments during an 18-month period between January 1996 and June 1997. For each outbreak, epidemiologic data were recorded, and RT-PCR and nucleotide sequencing studies were performed on the stool and emesis samples. This unique combination of epidemiologic and molecular data has allowed us to examine the role of NLVs in outbreaks of nonbacterial gastroenteritis in the United States, the distribution of strains circulating over time and geographic location, and the patterns of illness with respect to settings, modes of transmission, ages of persons affected, and size of outbreaks. We also assessed several outbreaks in which the molecular data aided the classic epidemiologic investigation by either reenforcing the epidemiologists’ conclusion or by conflicting with it and thereby encouraging the investigator to examine the data for other conclusions.
Materials and Methods

Outbreaks and specimens. Between January 1996 and June 1997, 120 outbreaks of nonbacterial gastroenteritis were reported to the CDC by 33 state and local health departments. Many states (n = 14) reported only a single outbreak, and 1 state, Florida, reported 29 outbreaks. For each outbreak, the epidemiologist was requested to provide information on the setting, date, presumed mode of transmission, number of persons affected and at risk, symptoms, and ages of patients. Oyster-associated outbreaks were classified separately from other foodborne outbreaks to reflect the differences in how each is contaminated; oysters are typically contaminated before harvest and distribution, whereas in many other viral food-related outbreaks, the contamination can be linked to an ill food handler or other on-site contamination. We examined a subset of 90 outbreaks for which (1) epidemiologic data were available, (2) laboratory tests were negative for bacterial and parasitic agents, and (3) at least 4 stool or emesis samples had been collected in a timely fashion (within 48–72 h after onset), stored at 4°C with no preservatives [36], and submitted to our laboratory. Each outbreak was identified by a unique number and 4-letter code. All specimens with adequate testing volume were examined in outbreaks with ≤20 specimens, and 20 specimens were chosen for testing when >20 were submitted.

Detection and genetic characterization of NLVs. Specimens from all outbreaks were examined by RT-PCR using two primer sets, G-1 and G-2 [37], which amplify a 123-base region of the polymerase gene of GI and GII viruses, respectively. At least 1 RT-PCR-positive specimen from each outbreak was genetically characterized by nucleotide sequencing of both strands of the amplified 123-bp product [37], using an ABI PRISM Dye Terminator Cycle Sequencing Kit with AmpliTaq DNA Polymerase (Perkin-Elmer, Foster City, CA) on an automated sequencer (model 377; Applied Biosystems, Foster City, CA). After removal of primer sequences, the unique 81 bases from each of 86 outbreak strains, 10 reference strains from GenBank, and 15 previously characterized UK strains [37] were analyzed by using the GCG suite of programs [38]. Strain diversity was estimated using the DISTANCES program with the Jukes Cantor method. A phylogram was created using the GROW-TREE program with the neighbor-joining method. The phylogram was used to classify strains into 1 of 10 presumed antigenic clusters [32]. A small subset of strains that could not be amplified in the polymerase region or characterized into an antigenic cluster on the basis of the 81-base nucleotide sequence of the polymerase gene was further examined by amplifying a 322-base region of the capsid gene, using two additional primers, MON381 and MON383 [32]. After classifying strains into clusters based on the phylogram, we examined the strain distribution over time, looked for predominant strains or clusters during the period of investigation, and used the nucleotide sequence information to aid in outbreak investigations by comparing conclusions based on molecular data with those of the classic epidemiologic investigation.

Because the region of sequence examined in this study is small, the sequences have not been submitted to GenBank. The sequences are available upon request. GenBank accession numbers for reference strains used in this analysis are Camberwell virus, U46500; Desert Shield virus, U04469; Hawaii virus, U07611; Lordsdale virus, X86557; Melksham virus, X81879; Mexico virus, U22498; Norwalk virus (NV), M87611; Snow Mountain agent, L23831; Southampton virus, L07418; and Toronto virus, U02030.

Results

Epidemiologic characteristics of outbreaks. We examined the epidemiologic characteristics of 90 outbreaks of nonbacterial gastroenteritis by setting, ages of patients, presumed mode of transmission, and numbers of persons affected and at risk in each outbreak (table 1). These outbreaks occurred in 33 states and in many different settings; nursing homes were the most common (43%), followed by restaurants and events with catered meals (26%). Persons of all ages, from 6 months to 101 years, were affected (average age, 54 years). Outbreak size ranged from small clusters, of 10 people, to epidemics involving >6000 persons, and the numbers of persons at risk ranged from 14 at a family gathering to 55,456 people at a hotel with 3000 guests, in which the outbreak persisted for 34 days. Of the 51 outbreaks

Table 1. Epidemiologic characteristics of 90 outbreaks of gastroenteritis investigated in the United States, January 1996 to June 1997.

<table>
<thead>
<tr>
<th>Setting</th>
<th>No. of outbreaks (%)</th>
<th>Mean age (range)</th>
<th>Mode of transmission</th>
<th>Persons affected</th>
<th>Persons at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursing homes, retirement</td>
<td>39 (43)</td>
<td>81 (23–101)</td>
<td>F 2 PP 8 O 0 W 0 U 9</td>
<td>41 (24–151)</td>
<td>147 (30–780)</td>
</tr>
<tr>
<td>centers, and hospitals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restaurants and events with</td>
<td>23 (26)</td>
<td>37 (9.9–72)</td>
<td>F 13 PP 0 O 0 W 0 U 1</td>
<td>44 (10–6000)</td>
<td>120 (14–12,000)</td>
</tr>
<tr>
<td>catered meals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oyster consumption</td>
<td>5 (6)</td>
<td>43 (24–49)</td>
<td>F 0 PP 0 O 5 W 0 U 0</td>
<td>215 (75–233)</td>
<td>233</td>
</tr>
<tr>
<td>Schools and day care centers</td>
<td>10 (11)</td>
<td>16 (0.5–52)</td>
<td>F 3 PP 0 O 2 W 2 U 2</td>
<td>72 (14–629)</td>
<td>116 (17–1657)</td>
</tr>
<tr>
<td>Vacation settings (including</td>
<td>10 (11)</td>
<td>44 (1–77)</td>
<td>F 1 PP 2 O 1 W 2 U 4</td>
<td>354 (54–642)</td>
<td>1154 (121–55,456)</td>
</tr>
<tr>
<td>cruise ships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other⁵</td>
<td>3 (3)</td>
<td>35 (24–30)</td>
<td>F 0 PP 0 O 0 W 0 U 0</td>
<td>109 (60–157)</td>
<td>960</td>
</tr>
<tr>
<td>Total (%)⁶</td>
<td>90 (100)</td>
<td>54 (0.5–101)</td>
<td>19 (37) F 10 (20) O 5 (10) W 3 (6) U 14 (27)</td>
<td>53 (10–6000)</td>
<td>150 (14–55,456)</td>
</tr>
</tbody>
</table>

NOTE. Data were not available in all categories for all outbreaks. Mode of transmission: F, foodborne; PP, person-to-person; O, oyster-associated; W, waterborne; U, undetermined.

⁵ Data from only 1 outbreak.
⁶ Percentages of each mode of transmission were determined using only outbreaks for which data were available and excluding those in “no data” category.
for which the mode of transmission was reported, foodborne spread was the most common (37%), followed by person-to-person contact (20%), ingestion of contaminated oysters (10%), and consumption of contaminated water (6%); for many outbreaks, the specific mode of transmission was sought but could not be clearly identified. Nursing homes were the setting in which epidemiologists were least able to determine a specific mode of transmission (47%, 9/19 outbreaks). The crude attack rates ranged from 1.2% to 93%, probably because of differences in the thoroughness of the epidemiologic investigation, as some outbreaks were actively investigated by a CDC or state epidemiologist, and some were passively reported to state health departments with less investigation.

Laboratory results. A total of 1084 stool or emesis specimens were received from 90 outbreaks (median, 8.5 specimens per outbreak; range, 4–71). Of these, 901 were suitable for testing by RT-PCR (median, 8 specimens per outbreak; range, 4–38), using both the G-1 and G-2 primer sets. NLVs were detected in at least 1 specimen from 86 outbreaks (96%), and detection rates within outbreaks ranged from 11% (1/9 specimens) in 1 outbreak to 100% in 6 outbreaks. A total of 442 specimens (49%) were positive by RT-PCR, including 424 (96%) that were positive with the G-2 primer set and 18 (4%) with the G-1 primer set.

The RT-PCR product from at least 1 representative strain from each of 86 NLV-positive outbreaks was sequenced, and a rooted dendrogram of genetic distances was constructed (figure 1). The 86 outbreaks were characterized into genetic clusters as defined by Noel et al. [32], rather than probe types, as used previously for polymerase sequences, because not all outbreak specimens were examined by Southern hybridization. Strain clustering in the dendrograms based on the polymerase and capsid regions correlated well with each other with one exception: In the polymerase region, Snow Mountain agent and Melksham virus segregated into separate clusters, whereas in the capsid region, they formed a single cluster. This discrepancy is most likely due to the small region of sequence used for the polymerase analysis, so we have chosen to consider them a single cluster on the basis of the larger capsid sequence. Additionally, the polymerase dendrogram has been simplified by labeling the Norwalk, Southampton, cruise ship, and Desert Shield virus clusters as NV because we detected few of these strains in this study. Using these parameters, we were able to characterize the strains in this study period into 6 clusters: NV, Gwynedd virus, Toronto virus, Hawaii virus, Lordsdale virus, and Snow Mountain agent. No strains representative of the White River cluster were detected during this study period. Overall, the diversity was great between strains, with two exceptions. In the Lordsdale cluster, over the 18-month period of the study, a “common” strain with identical 81-base nucleotide sequence caused 29 (32%) of the 90 outbreaks. This common strain was first identified in April 1995 and, during this study period, was identified from outbreaks in 15 states as geographically disperse as Florida, Alaska, and Hawaii, with no epidemiologic link apparent between the outbreaks it caused. Similarly, in the Toronto cluster, outbreaks were caused by strains with identical sequence. Within other clusters, all strains, except 2 in the Hawaii virus cluster, could be differentiated on the basis of the 81-base nucleotide sequence.

To further investigate trends of epidemic strains over time, we examined the quarterly distribution of outbreak strains spanning the 18-month study period (figure 2). This analysis was done using a denominator of 92 strains, rather than 90 outbreaks, because multiple strains of virus were identified in 2 outbreaks. The Lordsdale virus cluster of strains was the predominant cluster identified from the outbreaks examined (51%, 47/92), while strains belonging to the NV cluster were implicated in only 5 (5%) of the outbreaks and were not detected for almost a year between June 1996 and April 1997. During each quarter, strains representing at least 3 clusters were detected, and between April and June 1997, at least 1 strain from each of the 6 clusters was detected. During the first quarter of 1996, 83% of outbreaks were caused by strains belonging to the Lordsdale virus cluster, and of these, 65% could be attributed to the common strain. After this peak, the number of outbreaks caused by the common strain steadily fell until the final quarter of the study, when it was detected only once. The number of outbreaks attributed to the Lordsdale virus cluster in general followed a similar pattern as the common strain, peaking during the first quarter of 1996 and steadily declining until the final quarter of 1996, when it was equaled in number by strains of the Toronto virus cluster. During the first 6 months of 1997, no cluster of strains predominated.

Of note, within the limited 18-month period of surveillance, outbreaks appeared to have a winter-spring seasonality, with the number of outbreaks being the greatest in the first quarter of 1996 and 1997. However, outbreaks occurred throughout the year, and September 1996 was the only month in which no outbreaks were reported.

Epidemiologic and laboratory examination of outbreaks. For several outbreaks, results of the sequence analysis provided strong information that either confirmed or disputed the conclusion of the field investigation. We describe four such examples (figure 1).

Outbreak A. Epidemiologists investigating an outbreak (403dade) of gastroenteritis in an elementary school in Miami, Florida, in October 1996, involving 629 students and teachers, also investigated a concurrent outbreak in a nearby day care center (408kind) involving at least 6 children. The spread was linked to ill schoolchildren who had siblings attending the day care center. Although epidemiologists linked the 2 outbreaks, sequence analysis of the implicated NLVs indicated that the strains were unrelated and the outbreaks were independent.

Outbreak B. A cruise ship, which sailed out of Alaska, had
3 outbreaks during a 6-month period in June (387west), October (401west2), and November (405west3) of 1996. Investigators were uncertain as to whether these outbreaks were due to onboard transmission of the virus or independent introductions of new strains. Molecular analysis of the polymerase region identified 2 distinct strains, with the strains from the first and third voyages being identical. Further analysis using the larger capsid region differentiated these strains, confirming that they represented independent introductions.

Outbreak C. Epidemiologists investigated outbreaks from three successive voyages of a cruise ship, which carried ~1150 passengers and 380 crew members and sailed out of Miami, Florida, during March and April 1997. It was suggested by investigators that a common source of contamination was being harbored aboard the ship. Specimens were available from the first (435roya) and third (442roy2) cruises, and sequence analysis showed identical strains of NLVs from both voyages, which supported the epidemiologists’ conclusion that the source of contamination had remained onboard between cruises.

Outbreak D. A small cruise ship had an outbreak of gastroenteritis during a single voyage. The investigating official concluded that this outbreak was caused by a common source of exposure. Sequencing analysis identified two distinct strains (426natt and 426natt2) from the passengers, indicating either multiple sources of contamination or a single source with multiple strains.

Discussion

Our understanding of the etiology of outbreaks of nonbacterial gastroenteritis and the epidemiology of NLVs has increased in parallel with the development of novel, more sensitive, detection methods. A decade ago, when electron microscopy and serologic tests were the only diagnostic methods available, NLVs could be identified in only 20% of these outbreaks, which left open the search for many new etiologic agents [39]. Kaplan and colleagues [40, 41] and Kuritsky et al. [42], examining serologic test results along with clinical and epidemiologic features of patients, suggested that NLVs probably accounted for 45%–50% of these outbreaks in the United States. More recently, Vinje et al. [6], using RT-PCR as the diagnostic tool, attributed 91% of nonbacterial gastroenteritis outbreaks in The Netherlands to NLVs, similar to our own findings using the same methods. These observations underscore our conclusion that NLVs are the most important agents of nonbacterial epidemic gastroenteritis in the United States, responsible for 96% of such outbreaks.

In this study, we applied RT-PCR to the routine screening of fecal and emesis specimens from outbreaks of gastroenteritis. Our ability to detect a virus in 96% of these outbreaks demonstrates the success of this approach and indicates that this method could be used for routine diagnosis of outbreaks in the field. Nonetheless, since only 49% of the specimens examined from these outbreaks were positive, further attention needs to be directed to increasing the sensitivity of the assay by addressing issues such as inhibitors of RT-PCR before this method can be applied to screening individual specimens from sporadic or individual cases of disease. The addition of sequence analysis of RT-PCR products allowed us to monitor the variety of strains and the emergence and disappearance of individual strains over time. The fact that so many genetically different variants are present suggests that the majority of outbreaks were generally unrelated, independent events. By contrast, we observed in some outbreaks the emergence of a common, predominant strain with no obvious epidemiologic link between the outbreaks attributed to it, which occurred in different settings, via unrelated exposures, in distinct age groups, and in distant areas of the country. The sudden emergence and spread of a single strain raises important public health questions about the mode of transmission that permitted the rapid radiation of a single virus. This observation challenges us to explain how it spread to become so predominant and then gradually disappeared.

The usefulness of the molecular data was severely limited by the quality of epidemiologic information available. Our epidemiologic database could identify some descriptive features of public health interest and importance; for example, nursing homes were the most common setting for outbreaks, followed by restaurants or events with catered meals, while transmission was most common via contaminated food or water and person-to-person contact, with great strain diversity between outbreaks. At the same time, more detailed epidemiologic infor-

Figure 1. Phylogram of 86 outbreak strains, 10 reference strains from GenBank, and 15 UK strains [37] based on 81 bases of RNA polymerase gene created using DISTANCES program with Jukes Cantor algorithm, followed by GROWTREE analysis. Eighty-six outbreak sequences reflect 82 outbreaks with single sequences and 2 outbreaks with 2 sequences. For simplification, 4 GI clusters described by Noel et al. [32] were considered as single cluster in this analysis. Snow Mountain agent (SMA) and Melksham virus (MK) appear to form separate clusters on basis of polymerase phylogram, but we have chosen to consider them single cluster based on larger capsid region [32]. Insets show outbreak strains that had identical sequence. Arrows indicate outbreaks discussed individually in text, and each letter corresponds to example with same letter. Although 2 strains for example B were identical in polymerase region, further sequence analysis in capsid region distinguished 2 outbreaks, and only 1 sequence is shown in this figure (405west3) because first outbreak in this example did not qualify for analysis in this study. GenBank accession numbers for reference strains used in this analysis: Camberwell virus (CAV), U46500; Desert Shield virus (DSV), U04469; Hawaii virus (HV), U07611; Lordsdale virus (LV), X86557; MK, X81879; Mexico virus (MV), U22498; Norwalk virus (NV), M87611; SMA, L23831; Southampton virus (SOV), L07418; and Toronto virus (TV), U02030. GV, Gwynedd virus.
mation could aid in determining in a more timely fashion the links between outbreaks or clusters in which a common strain was identified. Our surveillance now relies upon investigations and voluntary reporting by state and local health departments to CDC. The knowledge that these strains might be clustered should encourage more intense and timely future investigations to identify important links between these outbreaks. Prospective determination of unique viruses in multiple outbreaks would allow for rapid efforts to trace the spread of these strains and try to link them to a common vehicle or exposure.

Future studies should focus on furthering our understanding of the epidemiology of the viruses in both epidemic and endemic settings. The availability of new, sensitive detection methods of NLVs not only allows for the detection of virus in more outbreaks but also improves our ability to link outbreaks through the combined efforts of classic epidemiology and molecular analysis. The advances in molecular techniques challenge epidemiologists to identify links between outbreaks caused by the same virus in geographically distinct locations and to explore the modes of transmission that allow the virus to spread rapidly to all areas of the country. Resolution of key epidemiologic questions regarding spread of the virus in outbreaks may lead to new prevention measures to interrupt transmission. While most outbreaks of gastroenteritis can be attributed to NLVs, very little is known about the role of the viruses in sporadic cases. Examination of stool specimens for bacterial pathogens from both hospital inpatients and outpatients has found that no etiologic agent can be determined in 91% of the cases [43]. A future challenge will be to assess the role of NLVs in these cases and to determine if the sporadic cases are linked to outbreaks, but this will require the development of simpler, more sensitive diagnostic methods that can be used in a clinical setting. Improved surveillance of gastroenteritis outbreaks, along with studies examining the role of NLVs in sporadic cases, will
allow for comparisons between mode of transmission, seasonality, and strain types of epidemic and endemic NLV infections. This combination of information will provide a more thorough understanding of NLVs and the illnesses they cause.

Acknowledgments

We thank the epidemiologists and laboratory staff at the state and local health departments for the reporting and investigation of these outbreaks and the collection of specimens. We thank David Kim and Umesh Parashar for their assistance in collecting epidemiologic information and investigating reported outbreaks and John O’Connor for editorial assistance.

References


