Identification of a Distinct Common Strain of “Norwalk-like Viruses” Having a Global Distribution.

J. S. Noel, R. L. Fankhauser, T. Ando, S. S. Monroe, and R. I. Glass

“Norwalk-like viruses” (NLVs) are the most common cause of outbreaks of nonbacterial gastroenteritis. During molecular surveillance of NLV strains from 152 outbreaks of gastroenteritis that occurred in the US between August 1993 and July 1997, we identified an NLV strain that predominated during the 1995–1996 season. The “95/96-US” strain caused 60 outbreaks in geographically distant locations within the US and was identified, by sequence comparisons, in an additional 7 countries on 5 continents during the same period. This is the first demonstration linking a single NLV strain globally and suggests that the circulation of these strains might involve patterns of transmission not previously considered. The diagnostic techniques are now available to establish a global network for surveillance of NLV strains that would highlight the importance of NLVs worldwide and allow molecular identification of common strains having a global distribution so as to consider interventions for their control.

“Norwalk-like viruses” (NLVs), members of the family Caliciviridae [1], are a group of single-stranded RNA viruses that are the major cause of outbreaks of acute nonbacterial gastroenteritis [2–4]. Researchers in the 1970s and 1980s demonstrated that NLVs are distributed worldwide [5–9], but methods to detect NLVs at that time were limited to either the direct examination of fecal specimens by electron microscopy (EM), serology using immune-EM, or radio- or enzyme-immunoassays using post-infection human sera. Despite their usefulness in identifying a pathogen in outbreak settings, these methods neither allowed clear differentiation between NLV strains nor permitted tracing the lineage of a single strain.

A review of the epidemiology of outbreaks of foodborne and waterborne viral gastroenteritis in 1993 suggested that the lack of routine laboratory diagnostics had resulted in the underreporting of NLV outbreaks. In the absence of good diagnostic tests, epidemiologic criteria provided the best assessment of the importance of NLV gastroenteritis [10]. With the recent cloning, sequencing, and development of molecular methods of detection and strain characterization [11–23], we now recognize that NLVs are a genetically diverse group of viruses that can be classified phylogenetically into 2 genogroups [22, 24, 25]: genogroup I, which includes Norwalk virus [1, 26, 27], Southampton virus [28], and Desert Shield virus [29]; and genogroup II, which includes Toronto virus [30], Mexico virus [31], Hawaii virus [32], Bristol virus (BV) [24], Lordsdale virus (LV) [33], Camberwell virus (CAV) [34], Snow Mountain agent [35], and Melksham virus [36]. A result of the advances in molecular diagnostics is that partial sequence data are now available for more than 100 NLV strains worldwide [4, 11, 15, 19–21, 23, 24, 34, 37–45]. To date, molecular analysis of NLVs associated with outbreaks has demonstrated both a great diversity of strains in circulation [19] and the presence of a single predominant strain during given time periods [4, 23, 46, 47]. However, in spite of this, few surveillance programs use novel, sensitive molecular methods for the detection of NLVs, and fewer still genetically characterize NLV strains that could potentially link these outbreaks to a common origin.

Since 1993, we have been conducting molecular surveillance of NLV strains causing outbreaks of gastroenteritis within the US that were reported to the Centers for Disease Control and Prevention (CDC). Initially all NLV outbreak strains could be differentiated on the basis of the nucleotide sequence in an 81-base fragment of the RNA polymerase gene [11]. All strains from patients involved in an outbreak showed 100% nucleotide identity [37, 48, 49], whereas no 2 strains from unrelated outbreaks were identical [12, 19]. However, we recently described the molecular analysis of NLV strains from 90 outbreaks where 29/90 (32%) strains from supposedly unrelated outbreaks across the US that occurred between January 1996 and June 1997 showed 100% nucleotide identity in the 81-base region of the RNA polymerase gene. This suggested either the presence of a common strain or our analysis of a fragment too short to permit proper differentiation between outbreak strains [46].
this study, we further characterized these strains that had an identical sequence in the 81-base region and extended our period of surveillance to 4 years, from August 1993 to July 1997, so that we could investigate their appearance and subsequent decline. Strains were sequenced not only in the 81-base region of the RNA polymerase but also in a 277-base region of the capsid gene; for a subset of specimens, the complete open-reading frame 2 (ORF2; 1620 bases) was sequenced to determine their genetic relatedness. Finally, we compared the nucleotide sequences of the NLV strains circulating within the US to those circulating during the same years in other countries to investigate the extent of circulation of this common strain. Our experience in characterizing NLV strains from these outbreaks has allowed us to develop the methods and establish a sequence database to discriminate between circulating strains and consider the transmission of NLVs not only within a single localized outbreak but also nationally and internationally.

Materials and Methods

Specimens/Outbreaks. We examined a subset of 152 outbreaks for which appropriate specimens were available from a national collection of 195 outbreaks of NLV gastroenteritis that occurred in the US during 4 years, August 1993–July 1997, including 4 winter seasons. Specimens and basic epidemiologic information (e.g., setting, ages of patients, location) were submitted to the CDC according to the published guidelines [50] for investigation of NLV outbreaks. The outbreaks occurred in 35 states and in a variety of settings, including healthcare facilities (nursing homes, hospitals, long-term care facilities), n = 67 (44%); events (restaurants, catered meals [e.g., weddings]), n = 47 (31%); institutions (daycare centers, schools, universities, prisons), n = 17 (11%); and cruise ships, n = 17 (11%). No data on setting was available for 4 (3%).

Molecular characterization of NLV outbreak strains. All specimens were characterized by reverse transcription–polymerase chain reaction (RT-PCR) amplification of a 123-bp region of the RNA polymerase gene [11]. In addition, a 322-bp region of the capsid gene from at least 1 specimen from each genogroup II outbreak was amplified and sequenced as previously described [19]. To investigate the possibility of national and international spread of NLVs, the entire capsid encoding region (ORF2) of 11 of the epidemic strains (10 from the US and 1 from Australia) was amplified and sequenced (table 1). These 11 strains had been identified as the cause of outbreaks whose partial sequences, time, and location of isolation suggested national or international transmission.

Briefly, amplification of ORF2 was performed using the Titan One Tube RT-PCR System (Boehringer Mannheim, Indianapolis, IN) and primers Mon396 (5′ 671gtc att atg cac ggt 694 3′) and Mon395 (5′ 580gat gaa gat ggc gtc 599 3′; numbers refer to the sequence of LV), according to the manufacturer’s instructions, and yielded a 1631-bp amplicon. The RT-PCR amplicon was gel purified using a QiAex Gel Purification Kit (Qiagen, Chatsworth, CA). Nucleotide sequencing of both strands was performed using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Foster City, CA) on an automated sequencer (Applied Biosystems model 377; Applied Biosystems, Foster City, CA), using primers Mon395 and Mon396, internal primers Mon381 and Mon383 [19], and sequencing primers Jsn2 (5′ 561taa tgg gga tga tgt ct 670 3′), Jsn16 (5′ 624 aac aag cta cag tga gca 641 3′), Jsn17 (5′ 629 ac aag cca tta ctc agg 630 3′), and s24Fl (5′ 614 act ccc ggt gc 615 3′). Primary sequence data was assembled, and a consensus sequence for each amplicon was generated by using the Sequencher program (GeneCodes). Sequence analysis of the unique 81-, 277-, and 1620-base sequences from the 123-bp RNA polymerase, 322-bp partial capsid, and 1631-bp ORF2 amplicons respectively, was performed using DISTANCES, GROWTREE, and PAUP from the GCG suite of programs [51]. GenBank accession numbers for strains sequenced in ORF2 are given in table 1.

Results

Until early 1995, the nucleotide sequence in the RNA polymerase region of NLV outbreak strains was so diverse that no 2 strains from different outbreaks in the US were identical [19]. However, in April 1995, this changed when strains from multiple dispersed outbreaks appeared to have the same sequence [46]. From April 1995 to July 1997, the sequences of 60 (55%) of 109 strains from apparently unrelated outbreaks were closely related; 39 (65%) of the 60 strains had an identical sequence in the RNA polymerase region, and an additional 21 strains (35%) differed by only 1–4 nucleotides (nt) in this region. These 60 strains, which we defined as the “95/96-US” subset, pre-
Figure 1. GROWTREE phylogram using uncorrected distances of the 81-base nucleotide sequences from the RNA polymerase region of "Norwalk-like virus" outbreak strains. Sequences included are those from 152 outbreaks (designated by outbreak number) that occurred during August 1993-July 1997, 20 previously described UK strains [11], and 13 strains from GenBank. The "95/96-US" subset of strains are indicated. Abbreviations and GenBank accession numbers for strains are: GI, Genogroup I; GII, Genogroup II; DSV, Desert Shield virus, U04469; SOV, Southampton virus, L07418; NV, Norwalk virus, M87611; MV, Mexico virus, U22498; TV, Toronto virus, U02030; HV, Hawaii virus, U07611; SMA, Snow Mountain Agent, L23831; MK, Melksham virus, X81879; LV, Lordsdale virus, X86557; BV, Bristol virus, X76716; CA V, Camberwell virus, U46500; KY89, L23878; OTH25, L23830.
dominated during the 1995–1996 season of this study and were associated with outbreaks that occurred in all settings and 17 US states that submitted outbreak specimens (figure 1).

To more precisely determine similarities and differences between strains, we sequenced a 277-base region of the capsid gene of at least 1 specimen from each genogroup II outbreak in our collection. Amplified strains from 107/133 genogroup II outbreaks, including all 95/96-US strains, could be differentiated into genetic clusters as previously described (figure 2A) [19], and strains representing every genogroup II cluster were detected during the period of our study. Specimens from 26 genogroup II outbreaks could not be amplified in the capsid region using the primer pair Mon 381/Mon 383. The 95/96-US subset of strains exhibited 95.7%–100% nt similarity and 100% deduced amino acid (aa) identity in the 277-base partial capsid amplicon, with the exception of 2 (345-2 and 355, which differed by a single aa), and formed a unique subset within the LV cluster of NLV strains. This suggested that they represented multiple lineages that originated from a recent common ancestor most closely related to LV (figure 2B).

To determine the validity of the partial capsid amplicon analysis and address the question of local, interstate, and international circulation of NLV strains, the complete ORF2 (1620 bp capsid gene) of 11 of the 95/96-US strains, 10 selected as being representative of the NLV strains in circulation within 6 US states, together with 1 from Australia (table 1, figure 2B), were sequenced. The partial capsid amplicon analysis of the outbreak strains correlated with the results from the complete ORF2 analysis, with the 95/96-US strains forming a distinct subset within the LV cluster (figure 3). The 95/96-US strains were clearly related to (4–47 nt, 0–6 aa differences), “yet differentiable from,” other LV cluster strains in ORF2 (105–140 nt, 18–20 aa differences to either LV, BV, or CAV, the 3 most closely related strains for which complete ORF2 sequences are available in GenBank) or other genogroup II NLV strains, for example, Hawaii virus (567–575 nt, 186–188 aa differences). Furthermore, the ORF2 analysis provided definitive information on the circulation of the 95/96-US NLV strains: strains from 2 outbreaks (364, 379) that occurred in Arizona 53 days apart and strains from 2 outbreaks (345 and 373) that occurred in South Carolina 74 days apart differed by 2 and 8 nt, respectively, and 0 aa in ORF2, suggesting that the same strain persisted in the same community and resulted in repeated outbreaks. Strains from 4 outbreaks that occurred in South Carolina (345, 373), Florida (384), and Louisiana (416) during 1996 differed by 8–32 nt and 0 aa, suggesting that the same strain was transported between communities and resulted in outbreaks caused by the same strain in different states. The strain that caused an outbreak in Australia in August 1995 (004) differed by 13–24 nt and 0 aa from the strains that caused outbreaks in Arizona, South Carolina, Florida, and Louisiana (345, 364, 373, 379, 384, and 416).

This observation prompted us to consider the potential global circulation of the 95/96-US strains previously considered to be prevalent only in the US. On the basis of our phylogenetic analysis of NLV strains in the RNA polymerase, partial, and entire capsid regions, indicating maintenance of lineages among the 3 regions, we compared the sequences of the 95/96-US strains, which now included 1 representative strain from Australia, with the NLV sequences available in GenBank and identified at least 8 additional strains closely related to the 95/96-US subset from 7 other countries (figure 4). Based on these data, the first published sequence of a strain belonging to the 95/96-US subset was from 2 hospital-based outbreaks in Lymington and Southampton in the UK in February 1995. Subsequently, it was detected in the US in April 1995, then in Brazil, Canada, Australia, and the Netherlands in late 1995; Australia, the Netherlands and China in 1996; and in Germany before October 1997, suggesting that the strains forming the 95/96-US subset were also circulating globally during this period.

Epidemiologic data submitted with outbreak specimens allowed us to examine some of the characteristics of NLV infections within the US. The emergence of strains belonging to the 95/96-US subset in 1995 coincided with a 145% increase in the number of NLV outbreaks reported to the CDC over the previous year. Although the increase may represent a reporting artifact, the proportion of outbreaks attributed to the 95/96-US strains increased significantly from 0% (0/16) during the first year of our study (August 1993–July 1994) to 12% (2/17) during the 1994–1995 season, peaking at 69% (38/55) during the 1995–1996 season and declining to 31% (20/64) during the 1996–1997 season of our study (figure 5A). While the number of NLV outbreaks reported to the CDC continues to increase, strains belonging to the 95/96-US subset accounted for only 26% (19/72) during the 1997–1998 season (data not shown). The 95/96-US strains appeared to be more prevalent among the elderly and, following their emergence in April 1995, were responsible for 66% (38/58) of the outbreaks in healthcare fa-

---

Figure 2. A, 1 of 100 equally parsimonious trees generated by PAUP with a 5:1 weighting of transversions to transitions using the 277-base nucleotide sequences from the partial capsid region of all “Norwalk-like virus” outbreak strains. Clusters of strains are labeled as described in [19], GV, Gwynedd virus cluster and WR, White River cluster, named for the location where the first outbreak in these clusters occurred. Sequences included are those from 107 strains amplified from the 152 outbreaks that occurred during August 1993–July 1997, 8 UK strains [11], 4 previously described genogroup I outbreak strains (184, 277, 292, 316) [19], and 13 GenBank strains. B, enlargement of Lordsdale virus (LV)–cluster strains showing relationship between the 95/96-US subset and other LV cluster strains. The outbreak numbers and US states in which the outbreaks occurred are given for each strain. Arrows indicate strains for which the complete open-reading frame 2 sequence was determined. Abbreviations and GenBank accession numbers for strains are as for figure 1. AU, Australia; UK, United Kingdom.
Figure 3. Distance matrix showing actual number of amino acid (aa) or nucleotide (nt) substitutions in open-reading frame 2 of selected “Norwalk-like virus” (NLV) outbreak strains. Dark shading highlights local circulation of NLV strains. Light shading highlights interstate circulation of NLV strains. Abbreviations for strains are as for figure 1 and table 1. AU, Australia; UK, United Kingdom.

Discussion

Through our molecular epidemiologic surveillance of NLV strains causing outbreaks in the US over 4 years, we first identified in April 1995 what we called the 95/96-US subset of strains. These strains showed identical deduced aa sequence in both the polymerase and partial capsid regions and for a group, the complete ORF2, and rapidly became the predominant strains detected in outbreak specimens in the US during the 1995–1996 season. In the past, identification of a single distinct strain with no nt sequence differences has often been used to link patients to a common-source exposure, such as those who consumed contaminated water [37], oysters [12, 52–54], or food [48]. The detection of the 95/96-US subset of strains provided us with the invaluable opportunity to investigate the dynamics of NLVs causing outbreaks nationally within the US and, by sequence comparisons, to propose for the first time the global spread of these closely related strains between 7 countries on 5 continents. Identification of what is essentially a single strain responsible for national and international outbreaks challenges us to trace and understand how this strain was transmitted. For example, we must explain how the virus spread to geographically distant locations in the US and abroad with no obvious common source of exposure and what mode(s) of transmission or selective factor(s) allowed the sudden emergence and rapid global spread of this strain.

Well-defined methods to discriminate between strains form the basis of many international surveillance networks to track enteric and respiratory pathogens, including *Vibrio cholerae* O1, *Escherichia coli* O157:H7, Salmonella, influenza, and polioviruses, whose circulation can have global public health implications. The current study extends this work for the first time to viral enteric pathogens and has allowed us to identify the global presence of a single strain of NLV. The study further demonstrates that the diagnostic tools are now available to...
implement an international molecular-based surveillance system for NLVs in a manner similar to that used to trace Salmonella. For example, Enter-Net (formerly Salm-Net) [55, 56] has been used to identify international outbreaks of salmonellosis [57–59] by comparing isolates typed using a standardized protocol against an international database of human Salmonella isolates.

While the seasonal prevalence of NLV strains have been previously described [47, 60], the methods used to detect these strains have often been limited in their ability to differentiate between strains. Although increased awareness in NLV outbreaks may have partly contributed to the peak in NLV outbreaks during the 1995–1996 season, the proportion of outbreaks caused by the 95/96-US strains compared with previous and later years indicates that this was a unique event and that the associated epidemiological features are more likely a feature of the strains themselves rather than an increase in reporting. Vinje et al. described the epidemic spread of closely related strains of NLVs in the Netherlands during 1995 and 1996 [4, 23]. Our own findings correlated with their results and, using...
Figure 5. “Norwalk-like virus” (NLV) outbreaks in the US by quarters, reported to the CDC August 1993 through July 1997. As surveillance began in August 1993, the first reporting period is defined as the third quarter of 1993 and includes the months August–October. Subsequent quarters follow sequentially. A, molecular characterization of the NLV outbreak strains. The number above each bar indicates the percentage of outbreaks caused by the 95/96-US strains for that quarter. B, the distribution by setting of the 95/96-US outbreak strains and “other strains” which includes HV, Hawaii virus; TV, Toronto virus; NV, Norwalk virus; GV, Gwynedd virus; WR, White River virus; LV, Lordsdale virus; and SMA, Snow Mountain Agent cluster strains [19]. Percentages of outbreaks caused by the 95/96-US subset for each setting following the emergence of the 95/96-US strain are given.
our database, we were able to identify that the strains forming the NET96 cluster, responsible for 88% (53/60) of the outbreaks (the majority in nursing homes) during their surveillance period, were identical to the 95/96-US strains (figure 4). Similarly, Wright et al. describe a significant increase in “Genogroup 2B strains” of which strains 22–36 (table 1, [45]) detected in Australia during 1995–1996 cluster with the 95/96-US subset of strains (figure 4). Furthermore, although no sequence information is available, the appearance of the 95/96-US, NET96, and Australian strains coincided with an equally large increase in NLV-associated outbreaks reported in the UK [61, 62], where in 1995 and 1996 the number of reported NLV outbreaks \( (n = 367 \text{ and } n = 314 \text{ respectively}) \) was more than twice the number of outbreaks \( (n = 154) \) reported for 1994 [63]. Molecular characterization of 3 strains detected for the first time and in the same year in 3 hospital-based outbreaks in Southampton were identified as belonging to the 95/96-US subset (figure 4).

This study is predicated on the genetic principle that strains with the same sequence have a clonal origin and that, with continued passage, they accumulate more and more sequence changes, first as silent changes in third base positions, and then as coding differences as well. We do not yet understand the biological interactions that regulate the rate of introduction of these genetic changes, and elucidation of the rate of accumulation of these sequence changes would be of great help in understanding the dynamics in the spread of closely related strains. A further limitation is that our study relies on a passive reporting system and most likely only reflects a small subset of NLV outbreaks. More complete or active surveillance of NLV strains from both outbreaks and sporadic hospital admissions might help us understand the diversity of strains in circulation, their mode of transmission, and common features that might link outbreaks to a common source.

The results of our molecular epidemiologic surveillance have raised many important public health questions that need to be further addressed. Person-to-person, foodborne, or aerosol transmission of NLVs have been suggested, which could explain the sudden emergence and global spread of an epidemic strain of NLV. For example, infections of international travelers such as cruise ship, air or tour group passengers or importation of fecally contaminated fruits and vegetables or oysters that are consumed uncooked could permit international dissemination of a highly infectious agent [37, 64]. Similarly, our results challenge us to explain why the 95/96-US strain predominated during 1995 and 1996 and has been replaced the following season by a greater variety of NLV strains. Although we previously focused our investigations on the spread of virus within the outbreak situation, our current data suggest that we now have local, national, and global considerations to address when investigating outbreaks. The implementation of an international surveillance network with laboratories providing sequence data from a common genetic region would be helpful to highlight the importance of NLVs worldwide and to demonstrate the emergence or global circulation of particular strains. Moreover, it would permit the rapid identification of international outbreaks of NLV-related gastroenteritis by allowing us to link outbreaks never previously considered to be associated. This could enable the identification of their mode(s) of transmission, such as a common imported food, so that strategies for prevention could be considered.

Acknowledgments

We gratefully acknowledge the contributions provided by Maria do Carmen Timenetsky, Dr. Gary Grohmann, and Dr. Zhao-Yin Fang and the editorial assistance of John O’Connor.

References

Global identification of a common NLV strain


25. Liu BL, Clarke IN, Caul EO, Lambden PR. Human enteric caliciviruses have a unique genome structure and are distinct from the Norwalk-like viruses. Arch Virol 1995; 140:1345–56.


54. ENTER-Net (online). National Surveillance Network for the enteric infections - Salmonella and VTEC O157; Available at http://www.phls.co.uk/ _private/enter-net.htm


58. Threlfall EJ, Hampton MD, Ward LR, Rowe B. Application of pulsed-field


