

SHORT REPORT

A large outbreak of acute gastroenteritis in Shippensburg, Pennsylvania, 1972 revisited: evidence for common source exposure to a recombinant GII.Pg/GII.3 norovirus

J. A. JOHNSON, G. I. PARRA†, E. A. LEVENSON AND K. Y. GREEN*

Caliciviruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

*Received 18 October 2016; Final revision 17 January 2017; Accepted 8 February 2017;
first published online 15 March 2017*

SUMMARY

Historical outbreaks can be an important source of information in the understanding of norovirus evolution and epidemiology. Here, we revisit an outbreak of undiagnosed gastroenteritis that occurred in Shippensburg, Pennsylvania in 1972. Nearly 5000 people fell ill over the course of 10 days. Symptoms included diarrhea, vomiting, stomach cramps, and fever, lasting for a median of 24 h. Using current techniques, including next-generation sequencing of full-length viral genomic amplicons, we identified an unusual norovirus recombinant (GII.Pg/GII.3) in nine of 15 available stool samples from the outbreak. This particular recombinant virus has not been reported in recent decades, although GII.3 and GII.Pg genotypes have been detected individually in current epidemic strains. The consensus nucleotide sequences were nearly identical among the four viral genomes analysed, although each strain had three to seven positions in the genome with heterogenous non-synonymous nucleotide subpopulations. Two of these resulting amino acid polymorphisms were conserved in frequency among all four cases, consistent with common source exposure and successful transmission of a mixed viral population. Continued investigation of variant nucleotide populations and recombination events among ancestral norovirus strains such as the Shippensburg virus may provide unique insight into the origin of contemporary strains.

Key words: Norwalk agent and related viruses, noroviruses, caliciviruses, gastroenteritis.

Noroviruses are recognized now as the leading cause of hospitalization due to viral gastroenteritis in the USA [1]. However, immediately after the discovery of the prototype strain, Norwalk virus, in 1972 by Kapikian *et al.* at the National Institutes of Health (NIH) [1], it was unclear whether Norwalk and related viruses were a major cause of epidemic gastroenteritis.

A number of outbreak investigations at the time could not be linked to a specific pathogen due to technical limitations of the early diagnostic assays. Here, we apply a genomics approach to the elucidation of the etiology of one such outbreak.

In November 1972, an outbreak of acute gastroenteritis was reported in the town of Shippensburg, Pennsylvania. Approximately 1500 students at Shippensburg State College (now Shippensburg University of Pennsylvania), and 3500 other residents of the surrounding area complained of illness. A range of signs and symptoms were reported including nausea, vomiting, diarrhea, fever, myalgia and sore throat

* Author for correspondence: Dr K. Y. Green, Caliciviruses Section, Laboratory of Infectious Diseases, NIAID, NIH, 50 South Drive, Bldg. 50, Room 6318, Bethesda, MD 20892, USA. (Email: KGREEN@niaid.nih.gov)

† Current address: Division of Viral Products, CBER, FDA, Silver Spring, MD 20993, USA.

(Supplement Table S1) [1, 2]. Symptoms resolved after a median of 24 h (range <1 h to 21 days), consistent with norovirus criteria developed later by Kaplan *et al.* [2, 3].

An investigation of the outbreak by Centers for Disease Control and Prevention (CDC) officials established a timeline of notable events. The Director of Health Services for the college initially reported the outbreak on 17 November, after an increase in visits to the college dispensary for gastrointestinal illness on the 16th. Cases from the surrounding township and borough were reported on the 17th and on subsequent days. A line break in the northern Shippensburg water system had been reported the previous week and water could not be ruled out as the source of the pathogen. This led to a drinking water ban for the college and surrounding area by Saturday, 18 November. The college suspended classes for the following week until the end of the Thanksgiving vacation. Investigation by the Pennsylvania Department of Environmental Resources of the water supply ruled out bacterial contamination by 22 November, leading to the lifting of the ban. Students returned on 26 November, and no additional cases were recorded at the college or in the town of Shippensburg.

Attack rates were high, ranging from 21.0–33.0 per 100 people in surrounding areas to 38.4–42.3 per 100 people in the township and borough of Shippensburg. The college dormitories were most affected, with an attack rate of 60.4 per 100 people. Approximately 2500 undergraduate students resided in the dormitories, half of the total undergraduate and graduate student population of the college. Records of the initial outbreak and follow-up surveys in the town were documented in a comprehensive outbreak report prepared by the CDC [4]. The number of total cases peaked for the college around 17 November, with more gradual peaks for outlying areas in the following days. The outbreak curve of the percent of the population affected in the university and the community is illustrated in a diagram from the original report in Figure 1a. The sharp epidemic curve for the college is characteristic of a common source epidemic, often a hallmark of exposure to contaminated food or water [2]. Surveys of students and the surrounding area showed an increased attack rate with increasing consumption of water, though not in all areas of the town. No foods were implicated in the follow-up surveys as significantly associated with illness except for consumption of five or more glasses of milk per day. Individuals aged 60 or more reported lower attack

rates than younger age groups, and attack rates were higher for people in Shippensburg than in the surrounding towns.

Stools from 14 students acutely ill with gastroenteritis tested negative for pathogenic bacteria (*Shigella*, *Salmonella*, *Bacillus cereus* and *Klebsiella*). To rule out a viral agent, a set of specimens, including 54 rectal swabs, 16 stools, and 59 paired sera, was sent to Albert Kapikian at NIAID (National Institute of Allergy and Infectious Diseases), NIH for analysis by immune electron microscopy. Although the outbreak exhibited key epidemiological characteristics of a norovirus outbreak [2], no virus particles were detected in stool specimens from the ill individuals. Stool samples were placed into storage at -80°C , and we retrieved 15 stool samples for the current analysis.

Because the original outbreak investigation was conducted prior to current Institutional Review Board (IRB) protocols, the NIH OHSRP (Office of Human Subjects Research and Protection) determined that the stored stool samples were exempt from IRB review for the current study. Stool suspensions were prepared (10% weight:volume) in phosphate-buffered saline. Nucleic acids were extracted with the Mag-MAX™ Total Nucleic Acid Isolation Kit (Applied Biosystems; Foster City, CA) according to manufacturer instructions. Broadly reactive primers 289 hi and 290 hijk, that detect both norovirus and sapovirus in most clinical samples [5], were employed in a diagnostic RT-PCR. Nine of the 15 available samples yielded amplicons that were sequenced and identified as norovirus. Full-length genomes were amplified from positive stool samples as reported previously [6] and four samples yielding amplicons of sufficient quality and concentration (from cases B24, C1, C2, and PB) were analysed by next-generation sequencing (NGS) using the Ion Torrent platform. The precise 5' end sequence of B24 and C2 was confirmed using the 5'/3' RACE system, 2nd generation (Roche Molecular Diagnostics; Pleasanton, CA) according to the manufacturer instructions with the nested gene-specific primers SP1 (5'-GGTTTGTGTACTCCGAGCACC-3') and SP2 (5'-CGTCCCTGTTCTCCCTCTGATT-3'). No differences in the 5' end were noted compared with full-length genomic sequences obtained via NGS. The full viral genome was covered for each sample with median sequencing depths ranging from 3367 to 15 041 reads (Supplemental Table S2). The full-length genomic consensus sequences of the B24 and C2 Shippensburg virus strains (with RACE-verified ends) were submitted to GenBank and assigned Accession numbers KY442319

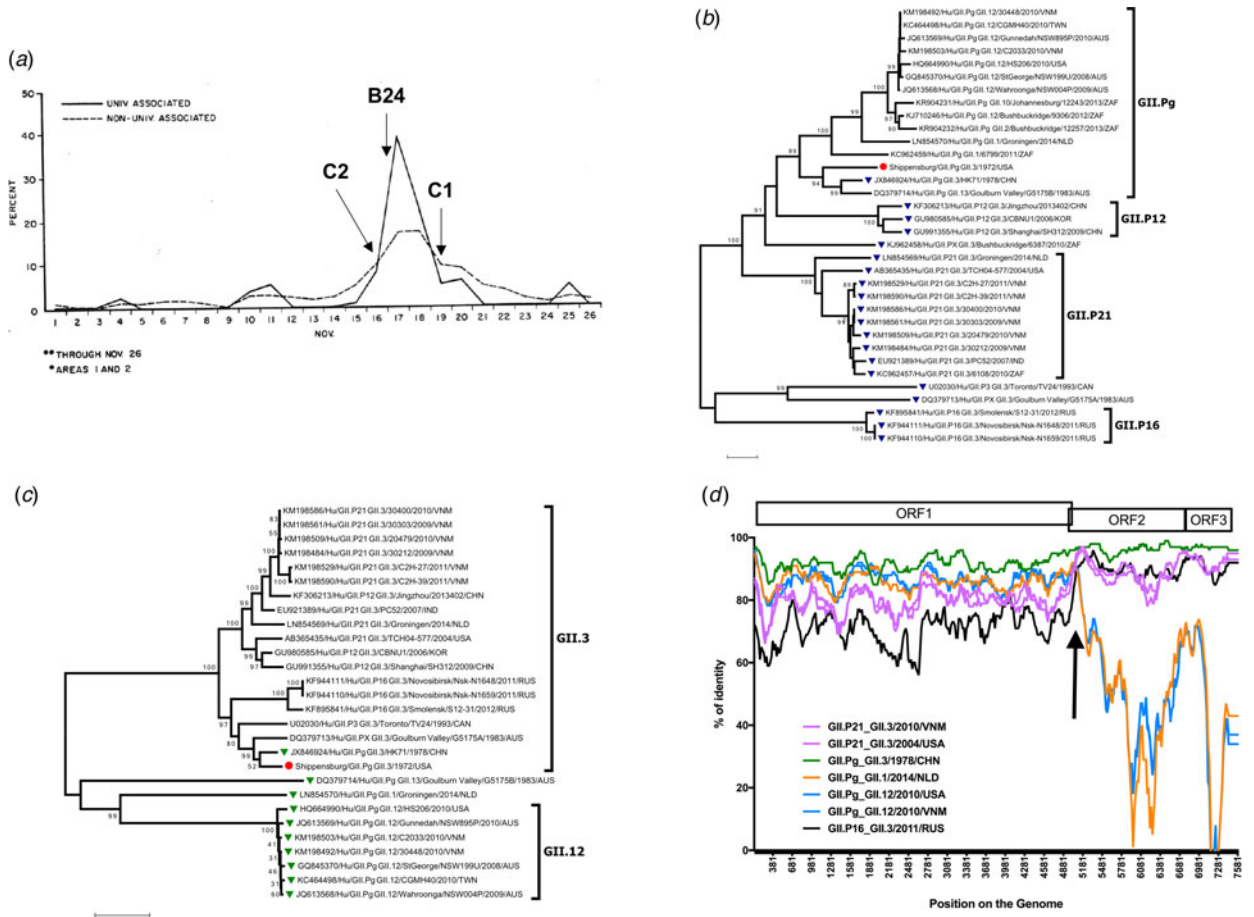


Fig. 1. Features of Shippensburg outbreak with GII.Pg/GII.3 recombinant norovirus. (a) Percent of population affected by university or town association. Figure adapted from initial CDC outbreak report. Percentage of population affected by university association in the college (solid line) and town (dotted line) are displayed. A prominent peak in the university-associated population is seen on 17 November, with a corresponding plateau in the non-university-associated population. Approximately 1500 cases occurred in the college, while an estimated 3500 occurred in the town of Shippensburg and in surrounding areas. Our cases with available onset dates (C2, B24, and C1) have onset indicated by arrows. (b) ORF1-based phylogenetic tree of Shippensburg virus and related strains. Tree was constructed using MEGA version 7.0. Related strains were found via BLAST search, using nucleotides 4242–5110. The Shippensburg B24 strain groups closest with GII.Pg_GII.3/HK71 from 1978. GII.3 viruses are represented by blue triangles. The red circle is the Shippensburg B24 strain. Polymerase genotypes are noted on the right. Bootstrap values ≥ 75 are noted in the tree. (c) VP1-based phylogenetic tree of Shippensburg virus and related strains. Tree was constructed using MEGA version 7.0. Related strains were found via BLAST search, using nucleotides 5091–5788. The Shippensburg B24 strain groups closest with GII.Pg_GII.3/HK71 from 1978. The GII.Pg polymerase strains are noted by green triangles. Shippensburg B24 is denoted by a red circle. Capsid genotypes are noted on the right. Bootstrap values ≥ 75 are noted in the tree. (d) Similarity plot analyses for recombination, Shippensburg strain. A similarity plot was constructed using SimPlot version 3.5.1 with a window size of 200 bases and a step size of 20 bases. Prototype strains of seven GII.Pg and GII.3 combinations were compared with the Shippensburg virus, including: GII.P21/GII.3 (Purple; KM198493, 2010 Vietnam, 30468), GII.P21/GII.3 (Pink; AB365435, 2004 USA, TCH04-577), GII.Pg/GII.3 (Green; JX846924, 1978 Hong Kong, HK71), GII.Pg/GII.1 (Orange; LN854570, 2014 Netherlands, Groningen), GII.Pg/GII.12 (Dark blue; HQ664990, 2010 USA, HS206), GII.Pg/GII.12 (Light blue; KM198503, 2010 Vietnam, C2033), and GII.P16/GII.3 (Black; KF944111, 2011 Russia, Nsk-N1648). GII.Pg viruses (blue and orange lines) have more similarity to the Shippensburg virus in the ORF1, while GII.3 strains (pink and black lines) have greater similarity in ORF2–ORF3. The HK71 virus (GII.Pg/GII.3) has the greatest similarity to Shippensburg for nearly the entire length of the genome. The arrow indicates a possible recombination breakpoint of GII.Pg and GII.3 viruses at the ORF1–ORF2 junction, as shown by the change in percent similarity for the comparison viruses.

and KY442320. The onset of illness date was available for cases B24 (17 November), C1 (19 November), and C2 (16 November).

The Shippensburg virus genome was 7527 nucleotides in length and organized into three open reading frames (ORFs) spanning nucleotides 5–5110 (ORF1, encoding the non-structural polyprotein), 5091–6737 (ORF2, encoding the VP1 capsid protein), and 6737–7501 (ORF3, encoding the VP2 protein). A predicted cleavage map for the polyprotein was similar to that of other genogroup II noroviruses (Supplement Fig. S1A). The norovirus genotyping tool [7] assigned the polymerase region as GII.Pg and the capsid as GII.3, making these the earliest documented representatives of these genotypes described thus far. The consensus nucleotide sequence was identical among the four genomes analysed with two exceptions. One non-synonymous mutation was detected in one patient's sample (case C2) at nucleotide 5292 (in ORF2), where a C to G substitution led to an alanine instead of a proline at amino acid residue 68 in the VP1 Shell (S) domain. Additionally, a synonymous A to G substitution at position 6713 of the ORF2 in one sample (case C1) left a glycine at residue 539 in the VP1 Protruding (P) domain unchanged.

We examined the underlying variant nucleotide populations below the predominant consensus sequence determined by NGS in the four genomes. Previous investigations have shown successful transmission of variants with frequencies as low as 0.01% of reads [8], supporting mixed viral population transmission as an important source of norovirus diversity. Variant nucleotide and amino acid residues for our study were defined as those where at least 10% of the total sequencing reads differed from the predominant consensus sequence at that position. We detected underlying nucleotide variation within the noroviruses of this outbreak, with between 126 and 221 sites (1.67–2.94% of the genome) meeting this definition (data not shown). This did not translate into equivalently high amino acid variation, with each sample exhibiting between 3 and 7 variable amino acids (0.12–0.28% of all amino acids) per genome below the predominant consensus sequence (Supplement Fig. S1B). Most of the polymorphic residues in our samples occurred in the N-terminal protein (NS1-2) region of ORF1, with one variable amino acid population detected in the VP2 encoded in ORF3 of one sample virus (case C2). The VP1 showed little to no amino acid variation among the four genomes. The VP1 amino acid substitution at residue 68 noted in the consensus sequence of case C2 was not associated with a mixed population of nucleotides at residue

5292, suggesting that a non-synonymous mutation occurred *de novo*. Two amino acid positions exhibited significant subpopulation variation within all four cases due to nucleotide heterogeneity. Amino acid 315 in the NS1-2 region was a mixture of threonine and proline, with proline as the predominant sequence in all cases and position 1293 in the NS7 region had both proline and glutamine, with glutamine predominant. These data are consistent with exposure and infection with the same mixed viral population, likely due to the suspected common source nature of this outbreak.

GII.3 viruses were detected in a variety of settings later in the 1970s [9, 10], and a GII.Pg/GII.3 virus (HK71) similar to the Shippensburg virus was detected in Hong Kong in 1978 [9]. Our BLAST searches did not identify recently occurring strains with this particular polymerase and capsid combination. A phylogenetic analysis was performed comparing the individual Shippensburg GII.Pg polymerase (nucleotides 4242–5110; Fig. 1b) and capsid region (nucleotides 5091–5788, S domain of VP1; Fig. 1c) with related sequences in GenBank. Viruses with the GII.Pg polymerase have been detected in outbreaks in Victoria, Australia from at least 1983 to the present [11]. This polymerase has also been found in combination with numerous other GII capsid genotypes, including GII.1, GII.12, GII.10, GII.2, and GII.13, suggesting that GII.Pg can readily recombine (Fig. 1b). Interestingly, a GII.3 capsid similar to that of the Shippensburg virus has been paired with a variety of polymerase genotypes in recent years, predominantly GII.P21 and GII.P12 (Fig. 1c). In both trees, the Shippensburg virus clustered with HK71, separate from more contemporary strains. Norovirus strains used for phylogenetic analysis can be found in Supplemental Table S3.

A similarity plot was constructed using SimPlot version 3.5.1 [12], comparing the Shippensburg strain to the historical HK71 strain as well as modern GII.3 and GII.Pg viruses (Fig. 1d). Evidence was found for a cross-over event at the ORF1–ORF2 junction, which suggests that the Shippensburg virus was likely a recombinant virus, at least when compared with current strains. Recombination is considered an important mechanism in norovirus evolution, particularly in GII.3 viruses [13]. However, due to the age of these samples, it is also possible that GII.Pg and GII.3 were an ancestral prototype that frequently paired together. This possibility is supported by the circulation of other GII.Pg/GII.3 viruses during this decade [9]. Like most noroviruses, the GII.3 noroviruses exhibit a complex pattern of evolution [10, 13], including both recombination and gradual drift, which requires further study.

We cannot confirm the source of the outbreak, but evidence recorded in the initial outbreak report supported a common source epidemic involving the Shippensburg water supply. Norovirus has been shown to be infectious in low doses [14], so it is plausible that infectious particles could enter a compromised water line for the college and town, leading to a sharp widespread outbreak. A number of waterborne outbreaks have been linked to improper sanitation [15]. The similarity of the viral variant populations among the infected individuals further supported the likelihood of exposure to a common source.

The outbreak in Shippensburg, Pennsylvania was caused by a rarely seen polymerase and capsid genotype combination that may no longer circulate. Attack rates were high in the small community, characteristic of modern norovirus outbreaks in similarly confined environments such as cruise ships, hospitals, and day care centers. The epidemiological impact of genomic recombination among norovirus strains is poorly understood, but a dual genotyping system was developed recently to track evidence of such recombination in epidemic strains [7]. Our analysis is consistent with the occurrence of multiple norovirus recombination events over the past several decades, with individual polymerase and capsid genes from the Shippensburg virus now appearing in various combinations in contemporary strains. Ancestral viruses such as the Shippensburg virus provide not only a retrospective history of recombinants and their potential parental donors, they contribute baseline sequences for evolutionary time clocks that inform selective pressures on the viral genome. As norovirus vaccines move forward, it will be important to understand the mechanisms responsible for the emergence of new strains so that vaccine antigens and immunization strategies can be optimally designed.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268817000498>

ACKNOWLEDGEMENTS

This work was funded by the Division of Intramural Research, NIAID, NIH. The authors appreciate the assistance of Andrew J. Oler of the Bioinformatics and Computational Biosciences Branch, NIAID, NIH, Bethesda, MD, for his expertise in data processing of next-generation sequences. They acknowledge the excellent outbreak records and CDC outbreak report

(compiled by Michael H. Merson, David Rimland, Richard J. Haber, Stanley M. Martin, Robert A. Pollard, William H. Barker, and Eugene J. Gangarosa) as well as William A. Nickles, Director of Health Services for Shippensburg State College, and their response teams that made the current analysis possible. They also thank Albert Z. Kapikian for preservation of these samples and records over time.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Green K.** Caliciviridae: the noroviruses. In: Knipe DMPMH, ed. *Fields Virology*, 6 edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2013, pp. 582–608.
2. **Kaplan JE, et al.** Epidemiology of Norwalk gastroenteritis and the role of Norwalk virus in outbreaks of acute nonbacterial gastroenteritis. *Annals of Internal Medicine* 1982; **96**(6 Pt 1): 756–761.
3. **Kaplan JE, et al.** The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis. *American Journal of Public Health* 1982; **72**(12): 1329–1332.
4. **Bacterial Diseases Branch BoE.** Outbreak of Gastrointestinal Illness in Shippensburg, Pennsylvania. In: *Bacterial Diseases Branch, Bureau of Epidemiology. Outbreak of Gastrointestinal Illness in Shippensburg, Pennsylvania*. Atlanta, GA: Public Health Service-Center for Disease Control, 1973; EPI-73-45-2.
5. **Jiang X, et al.** Design and evaluation of a primer pair that detects both Norwalk- and Sapporo-like caliciviruses by RT-PCR. *Journal of Virological Methods* 1999; **83**(1–2): 145–154.
6. **Parra GI, et al.** Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLOS Pathogens* 2017; **13**(1): e1006136.
7. **Kroneman A, et al.** An automated genotyping tool for enteroviruses and noroviruses. *Journal of Clinical Virology* 2011; **51**(2): 121–125.
8. **Bull RA, et al.** Contribution of intra- and interhost dynamics to norovirus evolution. *Journal of Virology* 2012; **86**(6): 3219–3229.
9. **Rackoff LA, et al.** Epidemiology and evolution of rotaviruses and noroviruses from an archival WHO Global Study in Children (1976–79) with implications for vaccine design. *PLOS One* 2013; **8**(3): e59394.
10. **Boon D, et al.** Comparative evolution of GII.3 and GII.4 norovirus over a 31-year period. *Journal of Virology* 2011; **85**(17): 8656–8666.
11. **Bruggink LD, Dunbar NL, Marshall JA.** The emergence of GII.Pg norovirus in gastroenteritis outbreaks in Victoria, Australia. *Journal of Medical Virology* 2016; **88**(9): 1521–1528.
12. **Lole KS, et al.** Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in

- India, with evidence of intersubtype recombination. *Journal of Virology* 1999; **73**(1): 152–160.
13. **Mahar JE, et al.** The importance of intergenic recombination in norovirus GII.3 evolution. *Journal of Virology* 2013; **87**(7): 3687–3698.
 14. **Teunis PF, et al.** Norwalk virus: how infectious is it? *Journal of Medical Virology* 2008; **80**(8): 1468–1476.
 15. **Bitler EJ, et al.** Norovirus outbreaks: a systematic review of commonly implicated transmission routes and vehicles. *Epidemiology and Infection* 2013; **141**(8): 1563–1571.