Application of whole-genome sequencing for norovirus outbreak tracking and surveillance efforts in Orange County, CA

A.J. Silva[1], Z Yang[2], J Wolfe[3], K Hirneisen[4], S Ruelle[4], A Torres[3], D.W Hill[4], M Kulka[2], and R.S. Hellberg[1]

[1] Chapman University, Schmid College of Science and Technology, Food Science Program, Orange, CA;
[2]US Food and Drug Administration, Center for Food Safety and Applied Nutrition, OARSA/Division of Molecular Biology, Laurel, MD;
[3]Orange County Health Care Agency, Public Health Laboratory, Santa Ana, CA;
[4]US Food and Drug Administration, Office of Regulatory Affairs, Office of Regulatory Science, Pacific Southwest Food and Feed Laboratory, Irvine, CA

Abstract

Noroviruses are the leading cause of acute gastroenteritis and foodborne illness in the United States. Traditional Sanger sequencing of short genomic regions (~300-600 bp) is the primary method for differentiation of this pathogen; however, whole-genome sequencing (WGS) offers a valuable approach to further characterize strains of this virus. This study investigated the ability of WGS compared to Sanger sequencing to differentiate norovirus strains and enhance outbreak investigation and surveillance efforts. WGS results for 41 norovirus-positive stool samples from 15 different outbreaks occurring from 2012 to 2019 in Orange County, CA, were analyzed.. All samples were genotyped with both WGS and Sanger sequencing based on the B-C region. WGS generated nearly full-length viral genome sequences (7,029 - 7,768 bp) with 4x to 35,378x coverage. Phylogenetic analysis of WGS data enabled differentiation of genotypically similar strains from separate outbreaks. Single nucleotide variation (SNV) analysis on a subset of strains revealed nucleotide variations (15-79 nt) among isolates from multiple outbreaks of GII.4 Sydney_2015[P31] and GII.17[P17]. Overall, the results demonstrated that coupling norovirus genotype identification with WGS enables enhanced genetic differentiation of strains and provides valuable information for outbreak investigation and surveillance efforts.

Introduction

Norovirus

Human noroviruses (NoV) are the leading cause of acute gastroenteritis and foodborne illness in the United States, with an estimated 19-21 million cases of illnesses every year (Hall et al., 2013b, 2013a). Norovirus spreads through the fecal-oral route and through person-to-person contact with particular concern in closed or semi closed settings, such as hospitals, schools, and food facilities (Parra et al., 2017).

Traditional Identification Methods

Detection of NoV at the species or genogroup level is typically achieved with real-time reverse-transcriptase polymerase chain reaction (RT-qPCR). Genotyping is carried out with Sanger sequencing of up to four different regions (designated A-D). Recently, protocols using a combination of regions B and C have been implemented (Cannon et al., 2017) (Fig 1).

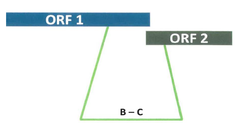

Figure 1: B-C region used for genotyping of NoV, adapted from CDC (Kroneman et al., 2013)

Whole Genome Sequencing

Whole Genome Sequencing (WGS) is a potential solution for enhancement of outbreak investigations through the analysis of the complete norovirus genome. Previous research has identified and differentiated various NoV strains with WGS (Bavellar et al., 2015; Petronella et al., 2018; Yang et al., 2016a, 2016b). However, additional research is needed that includes a wide range of genotypes and compares Sanger and WGS data.

Materials and Methods

Sample Collection

- 41 NoV positive stool samples from 15 foodborne outbreaks in Orange County, CA, were selected for analysis

RNA Extraction/RT-qPCR

- RNA extraction was carried out with the QIAamp viral mini RNA kit (Qiagen)
- Extracts were genogrouped with Quanta qScript One-step RT-qPCR Kit using primers and probes described by Cannon et al. (2017)

Sanger Sequencing

- The B-C Region was sequenced using primers described in Cannon et al. (2017) and an ABI 3500 Genetic Analyzer

Whole Genome Sequencing

- The RNA was treated with Invitrogen Turbo DNA-free kit
- WGS libraries were prepared with a TruSeq mRNA LT sample prep kit
- Library quality was assessed using Agilent High Sensitivity D1000 Screen tape kit
- Library quantity was assessed using Qubit dsDNA HS Assay kit
- No more then 10 samples were pooled and sequenced using the Ilumina MiSeq reagent kit v2
- Data analysis was conducted using the CLC Genomics Workbench
- Raw WGS data submitted to SRA and WGS assembled sequences submitted to GenBank under the ViroTrakr project https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA396739

Table 1. Average of total reads, norovirus reads, % norovirus reads, coverage, and consensus length between genotypes identified in this study

Outbreak reference #	Genotype	Average Total Reads	Average Norovirus Reads	Average % Norovirus Reads[a]	Average Genome-wide Depth of Coverage[b]	Average Consensus Length (bp)
3, 15	G1.3 (n=6)	4,044,887	64,592	3.87	800x	7,689
12	GII.2 (n=3)	2,037,426	3,275	0.18	41x	7,499
8	GII.3 (n=3)	1,915,296	461,808	30.16	5,788x	7,515
1, 5, 10, 11, 14	GII.4 (n=17)	3,356,795	315,512	14.32	4,012x	7,949
13	GII.5 (n=2)	3,736,694	732.243	0.246	9,489x	7,539
4	GII.6 (n=2)	2,304,608	3,791	0.28	47x	7,306
2	GII.13 (n=1)	754,649	473,411	62.73	6,020x	7,438
6, 7, 9	GII.17 (n=7)	5,884,514	185,607	3.51	2,278x	7,467

[a]Percent Norovirus reads were calculated as follows: (Average norovirus reads/Average total reads) *100
[b]Coverage was calculated as follows: total mapped reads/average length of the reference sequence

Significance

The objective of this study was to investigate the ability of WGS as compared to Sanger sequencing to differentiate NoV strains from 15 NoV outbreaks in Orange County, CA from 2012-2019. This study will provide a scientific basis for regulations enhancing outbreak investigation and surveillance efforts.

Results and Discussion

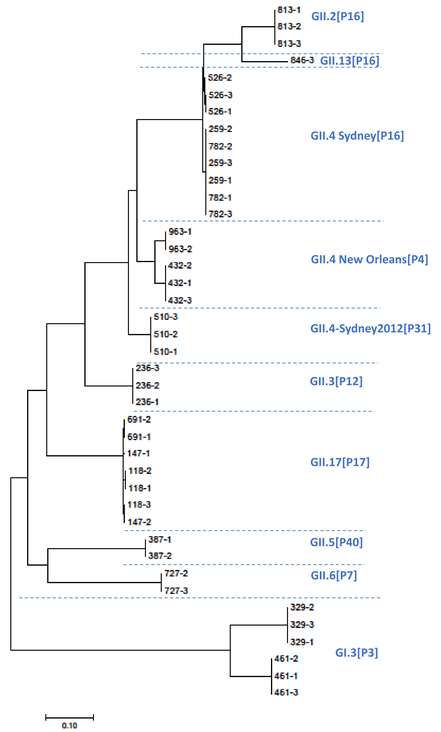

Figure 3: Phylogenetic tree of the norovirus strains analyzed in this study with WGS. The tree was constructed with whole genomic sequences of 41 viral samples in MEGA7 using the Neighbor-Joining method.

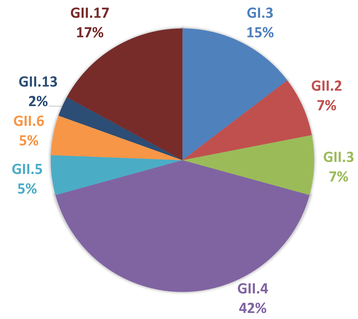

Figure 2: Diversity of 41 NoV isolates identified from 2012-2019 in Orange County, CA. Genotypes include GI.3, GII.2, GII.3, GII.4, GII.5, GII.6, GII.13, and GII.17

Conclusion

- All the samples in this study were 100% identical for genotype identification with both WGS and Sanger sequencing.
- WGS identified genome-wide nucleotide variations (15-79 nt) of multiple outbreaks of GII.4 Sydney_2015[P31] and GII.17[P17].
- While Sanger sequencing remains an important method for NoV genotyping, WGS provides additional nucleotide sequences important for discrimination of virus strains present within and between outbreaks.
- WGS can be used in a complementary manner with Sanger sequencing for NoV surveillance and outbreak investigations.

References

Bavelaar, H.H.J., Rahamat-Langendoen, J., Niesters, H.G.M., Zoll, J., Melchers, W.J.G., 2015.. Journal of Clinical Virology 72, 122–125.
Hall, A.J., Lopman, B.A., Payne, D.C., Patel, M.M., Gastanaduy, P.A., Vinje, J., Parashar, U.D., 2013a. Emerging Infectious Diseases 19, 1198–1205.
Hall, A.J., Mary E. Wikswo, Karunya Manikonda, Virginia A. Roberts, Jonathan S. Yoder, L. Hannah Gould, 2013b. Emerging Infectious Diseases 19, 1305–1309.
Cannon, J.L., Barclay, L., Collins, N.R., Wikswo, M.E., Castro, C.J., Magana, L.C., Gregoricus, N., Marine, R.L., Chhabra, P., Vinje, J., 2017. Journal of Clinical Microbiology 55, 2208–2221.
Kroneman, A., Vega, E., Vennema, H., Vinjé, J., White, P.A., Hansman, G., Green, K., Martella, 2013. Proposal for a unified norovirus nomenclature and genotyping. Arch Virol 158, 2059–2068.
Parra GI, Squires RB, Karangwa CK, Johnson JA, Lepore CJ, Sosnovtsev SV, Green KY. 2017. Plos Pathogens 13(1).
Petronella N, Ronholm J, Suresh M, Harlow J, Mykytczuk O, Corneau N, Bidawid S, Nasheri N. 2018. Bmc Infectious Diseases 18.
Yang ZH, Leonard SR, Mammel MK, Elkins CA, Kulka M. 2016a. Journal of Virological Methods 236:221-30.
Yang ZH, Vinje J, Elkins CA, Kulka M. 2016b. Microbiology Resource Announcements 4(5).