Application of next-generation sequencing to investigation of norovirus diversity in shellfish collected from two coastal sites in Japan from 2013 to 2014

Imamura, Saiki; Haruna, Mika; Goshima, Tomoko; Kanezashi, Hiromi; Okada, Tsukasa; Akimoto, Keiko

Japanese Journal of Veterinary Research, 64(2): 113-122

2016-05

10.14943/jjvr.64.2.113

http://hdl.handle.net/2115/62217
Application of next-generation sequencing to investigation of norovirus diversity in shellfish collected from two coastal sites in Japan from 2013 to 2014

Saiki Imamura1,*, Mika Haruna1, Tomoko Goshima1, Hiromi Kanezashi1, Tsukasa Okada2 and Keiko Akimoto1

1)Food Safety Consumer Affairs Bureau, Ministry of Agriculture, Forestry and Fisheries, 1-2-1, Kasumigaseki, Chiyoda-ku, Tokyo, Japan, 100-8950
2)Hokkaido System Science Ltd. Co., 1-2-1, Shinkawa Nishi 2, Kita-ku, Sapporo, Hokkaido, Japan, 001-0932

Received for publication, November 5, 2015; accepted, March 14, 2016

Abstract
A better understanding of the role played by shellfish regarding the manner of pathogen contamination, persistence, and selection may help considering epidemiology of noroviruses. Thus, norovirus genotype profiles in shellfish (Crassostrea gigas and Mitilus galloprovincialis) were investigated by using Next-generation sequencing (NGS) technology. In genogroup I (GI), 7 genotypes (abbreviated as GI.2 to GI.7, and GI.9) were detected from C. gigas, whereas 9 genotypes (GI.1 to GI.9) were detected from M. galloprovincialis. The genotype with the highest proportion found in both C. gigas and M. galloprovincialis was GI.4, and the second highest was GI.3. In genogroup II (GII), 17 genotypes (GII.1 to GII.9, GII.11 to GII.17, GII.21 and GII.22) were detected from C. gigas, whereas 16 genotypes (GII.1 to GII.8, GII.11 to GII.17, GII.21 and GII.22) were detected from M. galloprovincialis. The genotype with the highest proportion in both C. gigas and M. galloprovincialis was GII.4, the next highest differed between C. gigas and M. galloprovincialis. To our knowledge, this study may be the first trial to utilize the latest technology in this field, and reveal the diversity of norovirus genotypes present in shellfish.

Key Words: Crassostrea gigas, Mitilus galloprovincialis, Next-generation sequencing, noroviruses

Introduction
Noroviruses are a group of genetically diverse viruses belonging to genus Norovirus, family Caliciviridae, that are recognized as the leading cause of epidemics of gastroenteritis, and have

*Corresponding author: Saiki Imamura; Food Safety Consumer Affairs Bureau, Ministry of Agriculture, Forestry and Fisheries, 1-2-1, Kasumigaseki, Chiyoda-ku, Tokyo, Japan, 100-8950
Phone: +81-3-3502-8111. Fax: +81-3-3597-0329. E-mail: saiki_imamura410@maff.go.jp
doi: 10.14943/jjvr.64.2.113
significant impacts on human health\textsuperscript{7,16}. Norovirus can be divided genetically into at least five genogroups, I to V, which were further segregated into at least more than 30 genotypes\textsuperscript{36}. Because different virus types can differ in the ability to cause epidemics, the host range, incidence, virulence, and stability in the environment\textsuperscript{6,11,17,31,33}, it is important to identify genetically different noroviruses for understanding these epidemiology. Previous studies that used stool specimens from patients with nonbacterial gastroenteritis have demonstrated broad genetic diversity in noroviruses\textsuperscript{6,10,28}. In contrast, information on the epidemiologic characteristics of noroviruses in shellfish is still limited. A better understanding of the role played by shellfish regarding the manner of pathogen contamination, persistence, and selection may help considering epidemiology of noroviruses.

Shellfish is a filter feeder that can concentrate more than fourfold in their tissue particles present in the surrounding water, including noroviruses\textsuperscript{4}. Noroviruses present in shellfish is documented in the European Union and some Asian countries, as well as in the United States using conventional sequencing analysis\textsuperscript{1,5,25,30}. However, this conventional approach seems to have limitations to investigate genotype profiles present in shellfish, because shellfish may take up multiple genotypes present in the environment\textsuperscript{8}. In addition, for noroviruses, genogroup-specific differences have been reported with regard to environmental persistence\textsuperscript{9}, sensitivity to removal\textsuperscript{31}, and binding to receptors\textsuperscript{22}. Consequently, the genotypes with low proportion may be hidden by the major genotypes.

Though the expertise and tools required for rapid bioinformatics analyses has been a challenge, to date, NGS technologies have been widely applied to topics as describing the human microbiome\textsuperscript{26}, molecular typing of pathogens\textsuperscript{27,29}, and epidemiological approaches in hospital pathogen tracking\textsuperscript{19}. These suggest that NGS may be useful for drawing more information on the potential association between norovirus genotypes and shellfish.

In order to utilize the NGS technology for investigating norovirus genotype profiles present in shellfish, we conducted two NGS analyses on 1) cDNAs created from a virus suspension derived from \textit{C. gigas} and \textit{M. galloprovincialis} by the Illumina HiSeq, and 2) the capsid N/S region of VP1 gene amplified from the cDNAs by the Illumina MiSeq.

**Materials and Methods**

\textbf{Shellfish and preparation of virus suspension:} During twenty four weeks of September, 2013 to March, 2014, 30 aqua-cultured shellfish (\textit{C. gigas} and \textit{M. galloprovincialis}, respectively) were collected from 2.5 m depth under the surface of the sea, once a week, by two fish producers who were handling the harvest from 2 different coastal sites (A and B), which are the main production areas in Japan. The two fish producers were adopted for the study, so that shellfish supply for the study could be ensured during investigational periods.

For the preparation of viral suspension, fresh shellfish were shucked, and the digestive diverticula were removed by dissection on the day of harvest. Three dissected digestive diverticula were combined as a single sample, so that the combined mass could be over 2.0 ± 0.2 g as described elsewhere\textsuperscript{12}. The combined sample was homogenized with a phosphate-buffered saline solution without magnesium and calcium at 9 times their weight\textsuperscript{18}. The homogenates were incubated with \textit{α}-amylase (Sigma-Aldrich, Tokyo, Japan) at the final concentration of 25 mg/ml for 1 hr at 37°C with shaking at 40 times per min. For the condensation of viral suspension derived from digestive diverticula, the method under consideration by the Japanese Committee for Standardization Virus of Detection in Food was referenced\textsuperscript{13}. Briefly, twenty minutes after centrifuging at 8,000 \textit{×} g, the supernatant was recovered. In order to concentrate the virus,
polyethylene glycol 6000 (Sigma-Aldrich) and sodium chloride (Wako, Tokyo, Japan) was added to the final concentration at 8.0% and 2.1%, respectively. Eighteen hr after incubation at 4°C, the supernatant was centrifuged at 8,000 × g for 20 min. After removing the supernatant, the pellet was re-suspended with 200 μl of SDS-tris-glycine buffer contains 2.5 mM Tris, 19.2 mM glycine, 0.01% SDS, pH 8.3 (BIO RAD, CA USA). The resulting virus suspension was used for RNA extraction.

**RNA extraction, reverse transcription:** Viral RNA was extracted from 200 μl of viral suspension by using High Pure Viral RNA kit (Roche diagnostics, Tokyo, Japan) with DNase I recombinant (Roche diagnostics), and first strand cDNA synthesis was done by using High-Capacity cDNA Reverse Transcription kits (Life technologies, Tokyo, Japan), according to the manufacturer’s instructions.

**Next-generation sequencing analysis of the capsid N/S region of VP1 gene by Illumina MiSeq:** In order to amplify the partial capsid N/S region of noroviruses, total 240 samples each (C. gigas and M. galloprovincialis, respectively) were subjected to the RT-PCR, and primers were prepared as shown in supplemental Table 1. The RT-PCR was conducted by using HotStarTaq Master Mix Kit (QIAGEN), and the PCR protocol included incubation for 3 min at 94°C; subsequently, 40 cycles of 94°C for 60 sec, 50°C for 60 sec, and 72°C for 2 min, and an additional 15 min for elongation at 72°C after the last cycle. This PCR procedure was repeated using inner primers as a nested PCR. Following the purification of the products using AMPure XP beads (Beckman coulter, CA, USA), amplification, and concentration of the nested-PCR products were determined by Agilent 2200 TapeStation System (Agilent Technologies Japan Ltd., Tokyo, Japan). All amplicons were diluted, and pooled to generate a mixture containing an equimolar representation of each sample for 1 plate for sequencing, and then purified using QIAquick PCR Purification Kit (Qiagen, CA, USA). Purified libraries were sequenced by Illumina MiSeq (Illumina K.K., CA, USA) with 300-base paired-end reads to ensure that the average number of reads per sample ranged from 200,000 to 400,000 reads. All kits described above were used according to the manufacturer’s instruction.

<table>
<thead>
<tr>
<th>Target</th>
<th>Illumina MiSeq 2500</th>
<th>Illumina HiSeq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Paired-End</td>
<td>Paired-End</td>
</tr>
<tr>
<td>Number of samples</td>
<td>480</td>
<td>267</td>
</tr>
<tr>
<td>Average length of bases obtained/sample</td>
<td>59 Mbp (8 Mbp-228 Mbp)</td>
<td>1,740 Mbp (514 Mbp-6,236 Mbp)</td>
</tr>
<tr>
<td>Average number of reads obtained/sample</td>
<td>195,762 (26,998-756,524)</td>
<td>17,225,214 (5,089,792-61,740,102)</td>
</tr>
</tbody>
</table>

**Next-generation sequencing of cDNAs created from virus suspension by Illumina HiSeq:** First strand cDNAs of 267 samples, from which major capsid protein VP1 was confirmed by the nested RT-PCR, were all used as the template for the library. Double stranded DNAs was prepared by illustra GenomiPhi V2 Kit (GE healthcare Japan, Tokyo, Japan). Following purification of the products using AMPure XP beads (Beckman coulter), amplification, and concentration of...
double-stranded cDNA were determined by Qubit Fluorometer (Invitrogen, CA, USA) with Quant-iT ds DNA BR Assay Kit (Invitrogen). Fifty nanograms of the double-stranded DNA per sample were used for the library preparation by the TruSeq ChIP Sample Prep Kit (Illumina K.K.). Concentration of the library was determined using Agilent 220 TapeStation System. Library preparations more than 4 nM was sequenced by Illumina HiSeq with 100-base paired-end reads to ensure that the average number of reads per sample was approximately 20,000,000 reads. All kits described above were used according to the manufacturer’s instructions.

Read mapping for genotyping of norovirus in shellfish: Genotyping was performed based on the capsid N/S region of norovirus genotype references, and converted to a recent unified proposal for norovirus nomenclature and genotyping. The paired-end reads output from HiSeq were subjected to adaptor-trimming using cutadapt Version 1.1 (http://code.google.com/p/cutadapt/), and quality-trimmed using Trimmomatic Version 0.32 (http://www.usadellab.org/cms/?page=trimmomatic). Resultant reads were mapped to reference sequences using Burrows-Wheeler Aligner Version 0.7.10 (http://bio-bwa.sourceforge.net/), and output alignment data were sorted using SAMtools Version 1.2 (http://www.htslib.org/man/samtools/). Counting of reads mapped to each reference sequence was performed by custom script.

The paired-end reads from MiSeq were also adaptor- and quality-trimmed, and read pairs were assembled into consensus sequences using FastqJoin Version 1.1.2-806 (http://code.google.com/p/ea-utils/wiki/FastqJoin). Consensus sequences were subjected to further analyses as described above.

General information of two Next-generation sequencing was summarized in Table 1. Total 4,350,743,830 reads were obtained from the result of NGS of cDNAs created from virus suspension by Illumina HiSeq. However, the reads mapped with reference sequences were obtained only from 3 samples (2 reads for GI.4, 1 read for GI.5 from same 1 sample: 1 read for GII.3 from 1 sample: 18 reads for GII.4 from 1 sample). Thus, analyses of detection rate and proportion of norovirus genotypes could not be conducted by using the results from NGS of cDNAs created from virus suspension by Illumina HiSeq.

Detection rate and proportion of noroviruses in shellfish: The sample from which the read mapped with reference sequence of noroviruses was confirmed was regarded as a positive sample. Detection rate was calculated that the number of positive samples was divided by the total number of the sample examined.

In order to calculate proportion of each genotype in a sample, number of reads mapped to each reference was divided by the total reads obtained. Average proportion was described in Figs. 2 and 4.

Statistical analysis: Significant differences for detection rate and proportion of genotype were determined using the Fisher’s exact test and the Mann-Whitney U test, respectively. Significant differences were defined as $p < 0.01$.

Results

Detection of norovirus GI genotypes in shellfish

GI genotypes were confirmed by the NGS on PCR products amplified from the VP1 region (Fig. 1). VP1 region derived from GI was detected from 26 samples of C. gigas at site A, 15 sample of M. galloprovincialis at site A, 62 samples of C. gigas at site B and 57 samples of M. galloprovincialis at site B. The highest detection found in C. gigas was GI.4, and the next highest was GI.3. The other major genotype found in C. gigas was GI.2. As referred to Fig. 1, the results observed in M. galloprovincialis were similar to that in C. gigas. Significant difference observed between shellfish species and the sites were...
Proportion of norovirus GI genotypes in shellfish

Proportion of norovirus GI genotypes was calculated based on the number of the read mapped to each reference (Fig. 2). The highest proportion found in *C. gigas* was GI.4, and the second highest was GI.3. The other major proportion found in *C. gigas* was GI.2. As referred to Fig. 2, the results observed in *M. galloprovincialis* were similar to that in *C. gigas*.

There was no significant difference in the proportion of GI genotypes between *C. gigas* and *M. galloprovincialis*. In addition, differences in proportion of genotypes between sites A and B were not detected.

Detection of norovirus GII genotypes in shellfish

Detection of norovirus GII genotypes was confirmed by NGS of PCR products amplified from the VP1 region (Fig. 3). VP1 region derived from GII was detected from 76 samples of *C. gigas* at site A, 60 samples of *M. galloprovincialis* at site A, 93 samples of *C. gigas* at site B and 91 samples of *M. galloprovincialis* at site B. The highest detection found in *C. gigas* was GII.4, and the second highest was GII.6. The other major genotypes found in *C. gigas* were GII.3 and GII.2.

The highest detection found in *M. galloprovincialis* was GII.4, GII.6, and GII.14, and the second highest was GII.3. The other major genotypes found in *M. galloprovincialis* were GII.2 and GII.13.

Significant differences observed between shellfish species and the sites were indicated in Fig. 3.

Proportion of norovirus GII genotypes in shellfishes

As shown in Fig. 4, the highest proportion found in *C. gigas* was GII.4, and the next highest was GII.14. The other major proportion found in *C. gigas* was GII.3 and GII.6.

Conversely, the highest proportion found in *M. galloprovincialis* was GII.4, and the second highest was GII.3. The other major proportion found in *M. galloprovincialis* was GII.14 and GII.2.

Significant difference observed between
Diversity of noroviruses in shellfish

Fig. 2. Proportion (%) of norovirus GI genotype in shellfish. Proportion represents the average of positive samples. No significant difference of proportion, determined by Mann-Whitney U test (P < 0.01). At site A, proportion of GI.5 to 7 and GI.9 in C. gigas, and that of GI.1, GI.5 GI.7 and GI.9 in M. galloprovincialis, were less than 0.01 %. At site B, proportion of GI.5 and GI.9 in C. gigas, and that of GI.1, GI.5 and GI.7 to 9 in M. galloprovincialis, were less than 0.01%.

Fig. 3. Detection rate (%) of norovirus GII genotype in shellfish. Asterisk (*) demonstrates a significant difference of detection rate in shellfish, determined by Fisher’s exact test (P < 0.01). No read mapped to GII.10 was obtained from the sample prepared from C. gigas at site A. No read mapped to GII.9, GII.10 and GII.15 was obtained from the sample prepared from M. galloprovincialis at site A. No read mapped to GII.10, GII.12 and GII.15 was obtained from the sample prepared from C. gigas at site B. No read mapped to GII.9 and GII.15 was obtained from the sample prepared from M. galloprovincialis at site B.
shellfish species and the sites were indicated in Fig. 4.

**Discussion**

NGS analysis of about 300 bp encoding the capsid N/S region amplified from the VP1 region seems to be a useful tool to survey norovirus genotypes in shellfish, and revealed the diversity of norovirus genotypes in the analyzed shellfish. However, we recognized the method had limitations for genotyping, because the sequence of ORF1/ORF2 junction region, which have many recombination of the genome, and produce a diversity of noroviruses\(^3\), should be taken into account for the classification\(^16\). For this reason, the use of direct NGS on cDNAs prepared from concentrated virus suspension was thought to be advantageous to investigate the classification of noroviruses in the shellfish. However, it was not successful due to the existence of the sequences by the transcripts derived from shellfish and other bacterial species. In order to conclude the major contaminant(s), *de novo* assembly can be useful to identify the cause(s). If contaminated RNA from shellfish and other bacterial species were ribosomal RNA molecules (rRNA), the technique to deplete rRNA appeared to be useful to utilize. Further studies were needed to develop and optimize the method to use NGS for classification of noroviruses in shellfish.

The results obtained in this study demonstrated the genotype profile of noroviruses in shellfish, and a certain genotypes present in shellfish significantly differed between *C. gigas* and *M. galloprovincialis*. These findings suggest that the risk for gastroenteritis from shellfish differed depending on the species. Strain-dependent norovirus bioaccumulation in oysters was well studied by Maalouf *et al.*\(^22\). They obtained data on binding ability of the norovirus...
ligand in the oyster with norovirus GI.1, GII.3, and GII.4, and demonstrated a difference of bioaccumulation efficacy by the ligand. The genotype profiles of noroviruses found in shellfish may be correlated to the ligand expression in the shellfish. However, further studies are needed to refer the association between proportion of noroviruses and bioaccumulation.

In addition, multiple genotypes were detected from a single sample by using the NGS analysis in the current study. With sequence analysis instead of NGS, multiple norovirus strains are frequently identified in shellfish. In earlier studies, both GI and GII noroviruses were detected in shellfish in China [15]. GI.3 and GII.12 were also detected in shellfish in the other study [21]. However, the current study used a pool of 3 digestive diverticula as a single sample. Thus, it was not clear that the multiple genotype detection was caused by pooling or multiple contaminations. Modification of the method should be taken into consideration in the future.

Seasonal variation in the detection rate and proportion of norovirus genotypes in shellfish were not demonstrated in the current study, since seasonality covers various kinds of factors. Wang and Deng (2015) suggested that oyster norovirus outbreaks generally occur when extremely low gage height occurs in combination with low water temperature, low salinity, strong offshore wind, and heavy antecedent rainfall [34], implying complexity of analyzing interaction of seasonality with noroviruses in shellfish. Thus, it appears to be hard to discuss seasonality behind the prevalence of norovirus in shellfish from the current results at the present moment. However, in many countries, norovirus outbreak occur more frequently during winter [20], and shellfish contains less norovirus in the summer months [24]. Though it was difficult to discuss the relationship between prevalence of noroviruses in the shellfish and outbreaks in humans due to lack of the data of human outbreak in the region, frequent detection of GII in shellfish may reflect the true virus circulation from the cases with infectious gastroenteritis. In fact, MHLW information reported that GII genotypes were detected from 30 out of 35 cases of human gastroenteritis in 2013 and 2014 [23].

Recently, norovirus genotype profiles have been used to discriminate origins of foodborne outbreaks. A study on norovirus outbreaks in the United States demonstrated that, of the outbreaks with a known transmission route, 16% were food-borne, and that GI.3, GI.6, GI.7, GII.6, and GII.12 were the norovirus genotypes most often associated with food-borne outbreaks [32]. Bruggink et al. reported that GI.3 was significantly higher in children than in adults [2]. Furthermore, patients in long-term care facilities and people aged over 65 years were reported to be higher risk factors for GII.4 infections than those in other settings [32]. The investigational method used in the current study may contribute to investigate epidemiologic trends of the noroviruses.

To our knowledge, this study may be the first trial to utilize the latest technology in this field, and reveal the diversity of norovirus genotypes present in shellfish. Though the method still has some limitations, it may be widely applied in the field of public health.

Acknowledgement

This work was performed as part of the surveillance/monitoring program of microbiological hazards in the Ministry of Agriculture, Forestry and Fisheries, Tokyo, Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.14943/jjvr.64.2.113
References


Diversity of noroviruses in shellfish


