Journal of Clinical Virology 72 (2015) 122–125

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus

Herjan H.J. Bavelaar[a,*], Janette Rahamat-Langendoen[a], Hubert G.M. Niesters[b], Jan Zoll[a], Willem J.G. Melchers[a]

[a] Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, The Netherlands
[b] Department of Medical Microbiology, Division of Clinical Virology, The University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form
30 September 2015
Accepted 3 October 2015

Keywords:
Norovirus
Next generation sequencing
Outbreak management

ABSTRACT

Background: Norovirus is a major cause of gastroenteritis, causing yearly epidemics and hospital outbreaks resulting in a high burden on health care. Detection and characterization of norovirus directly from clinical samples could provide a powerful tool in infection control and norovirus epidemiology.
Objectives: To determine whether next-generation sequencing directly on fecal samples can accurately detect and characterize norovirus.
Study design: Whole genome sequencing was performed on fecal samples from 10 patients with gastroenteritis. Norovirus infection had previously been confirmed by RT-PCR. Genotyping was performed using phylogenetic analysis.
Results: From all clinical samples sufficient amounts of RNA were retrieved to perform whole-transcriptome sequencing for the detection of RNA-viruses. Complete genomic norovirus sequences were obtained from all clinical samples, permitting accurate genotyping by phylogenetic analysis. In addition, a complete coxsackie B1 virus genome was isolated.
Conclusion: Detailed information on viral content can be obtained from fecal samples in a single-step approach, supporting clinical and epidemiological purposes. Next-generation sequencing performed directly on clinical samples can become a powerful tool in patient care and infection control.

1. Background

Norovirus is a highly infectious virus with the potential of causing severe gastroenteritis in adults and children. With its unusually low infective dose [1], norovirus is particularly known for causing hospital outbreaks, necessitating elaborate infection control measures or even ward closure [9].

Norovirus, part of the family of *Caliciviridae*, is non-enveloped and contains a positive-sense single-strand RNA genome. Norovirus genome consists of 3 open-reading frames (ORF1-3), of which ORF-1 (containing RNA dependent RNA polymerase) and ORF-2 (encoding VP-1, the major capsid protein) are being used for typing of norovirus into genogroups and genotypes. Norovirus can be classified into 5 different genogroups (GI-GV), of which three (GI, GII, GIV) are known to cause disease in humans. Genogroups can be further divided into genotypes, with genotype GII.4 being the predominant genotype responsible for causing approximately 60% of global epidemics [21]. Recombination is common in norovirus [4]. Especially genotype GII.4 is known to be highly recombinant, with new strains of this genotype rapidly replacing old strains approximately every 2–3 years [5,15]. Surveillance of these new strains is important for understanding of norovirus evolution and global epidemiology.

Advances in molecular diagnostic techniques allow rapid identification of infected patients. Besides, characterization of norovirus by sequence analysis provides useful information for understanding transmission during hospital outbreaks. Rapid characterization of norovirus immediately after detection has shown to be essential in outbreak management. [20] In recent years, next generation sequencing (NGS) techniques are increasingly being used for outbreak monitoring, metagenomic studies and virus discovery [7,12,17,25,26]. Incorporating deep sequencing techniques into routine virus diagnostics on clinical samples broadens the range of viruses that can be detected and at the same time provides additional information on the characterization of the detected viruses.

* Corresponding author at: Radboud University Medical Center Nijmegen, Department of Medical Microbiology. P.O. Box 9101, 6500HB Nijmegen, The Netherlands.
E-mail address: Herjan.Bavelaar@radboudumc.nl (H.H.J. Bavelaar).

http://dx.doi.org/10.1016/j.jcv.2015.10.003
1386-6532/

Therefore, NGS is a promising, powerful tool in infection control and understanding norovirus epidemiology.

Until now, reports on the use of NGS in norovirus detection and characterization focused on the most prevalent genotypes, most notably genotype GII.4 [7,25]. A prerequisite for incorporating NGS into routine diagnostics is the ability of these techniques to detect and characterize novel recombinant norovirus genotypes.

1.1. Objectives

In this report we describe the detection of different genotypes of norovirus directly from clinical samples using multiplexed whole genome sequencing.

2. Study design

2.1. Sample selection

Eight norovirus positive fecal samples were included that had previously been characterized using the Noronet typing tool [11]. Different genotypes were selected based on reported prevalence. The eight most prevalent genotypes (as reported in the Noronet surveillance reports) were included: GI.P4, GI.Pb, GII.P4 Den_Haag_2006b, GII.P4 New_Orleans 2009, GII.Pe Sydney, GII.P7, GII.P21 and GII.Pg [3].

Additionally, fecal samples from 2 patients known to have gastro-enteritis caused by norovirus were included. Norovirus was confirmed by real-time PCR targeting the RDRP/Capsid junction of norovirus genotypes GI and GII. [24]. The genotype of these specimens was unknown. Specimens were stored at −80 °C.

2.2. Nucleic acid isolation

RNA was isolated directly from fecal samples suspended in 200 μl PBS. Suspensions were centrifuged for 10 min. at 10,000 × g in order to remove cellular debris. The supernatant was treated with RNase-One (Promega, Madison (WI), USA) and Turbo DNase-One (Life Technologies, USA) according to the manufacturer's protocols. Total nucleic acid was purified using Qiagen DNeasy Blood kit according to the manufacturer's protocol. Nucleic acids were eluted with 100 μl elution buffer. Eluted RNA was used for the pre-amplification of viral genomes as previously described [26].

2.3. Ribosomal RNA depletion and reverse transcription reaction

Virus identification by NGS requires enrichment of viral particles. Presence of human and bacterial cells will cause an overrepresentation of non-viral reads, especially rRNA. Therefore, human and bacterial rRNA from the purified nucleic acids fractions was depleted using the Ribo-Zero Gold rRNA removal kit epidemiology (Illumina, San Diego CA, USA) according to the manufacturer's protocol. rRNA-depleted RNA was purified with the RNeasy minikit (Qiagen, Valencia CA, USA). RNA was eluted in 15 μl elution buffer. For the reverse transcription reaction, 13 μl RNA was mixed with 8 μl lysis-buffer supplied with the Qiagen Repli-G WTA single cell kit and incubated for 5 min. at 24 °C, 3 min at 95 °C and cooled on ice. Reverse transcription was performed on 10 μl RNA using the Qiagen Repli-G WTA single cell kit according the manufacturer's protocol.

2.4. Amplification and sequencing of the DNA and cDNA samples

The cDNA fractions obtained after rRNA-depletion and reverse transcription were used for whole-transcriptome amplification (WTA) for the detection of RNA viruses, using the Qiagen Repli-G WTA single cell kit. Approximately 100 ng DNA per sample was used for sequencing on an IonTorrent PGM system with an Ion 318 sequencing chip (Life Technologies, Waltham (MA), USA).

2.5. Assembly of virus genomes

Partial or full virus genomes were assembled from selected reads covering the most significant reference sequences found during the virus identification analysis. Assembly and determination of sequence coverage was done with Bowtie2 using the reference sequence as template and the default setting "local-sensitive" switched on [13]. Sequence depth was calculated using Samtools available at the Galaxy web server [16,14]. Sequences are submitted to GenBank (accession numbers norovirus: LN854563 to LN854572; coxsackievirus: LN854562).

2.6. Phylogenetic analysis

The full genomic norovirus sequences from clinical samples were aligned with a subset of full genomes obtained from GenBank. Alignment and phylogenetic analysis was performed with CLC Genomic Workbench software package (Qiagen, Aarhus, Denmark). For norovirus genotype determination, an alignment was made with different norovirus strains submitted to GenBank (accession numbers upon request). A phylogenetic tree was constructed used a K-mer based Neighbour Joining method [8], using K-mer length of 15 and a Jukes-Cantor nucleotide distance matrix. Phylogenetic tree constructed also included bootstrap analysis with 100 replicates.

3. Results

3.1. Sequencing of clinical samples

From all samples, sufficient amounts of RNA were obtained to perform whole-transcriptome sequencing for the detection of RNA viruses. All reads were used as input for the virus identification (Table 1). With an average read length of 150 bp, the settings used for Bowtie2-alignment allows a sequence variation of approximately 26%.

3.2. Virus-classification of reads

The distribution of reads over human and viral sequences and the results of virus classification of reads are shown in Table 1. Complete genomic norovirus sequences were obtained from all samples. In addition, a complete coxsackievirus B1 genome could be assembled from reads obtained from sample 1. The mean sequence coverage of the viral genomes corresponded with the percentage of virus associated sequencing reads (Table 1).

3.3. Phylogenetic analysis

Norovirus genotypes could be determined from all clinical samples by phylogenetic analysis (Fig. 1, Table 1). For samples 1–8 genotyping results were identical to the results based on norovirus sequence analysis using the typing tool of Noronet. Norovirus detected in sample 9 and 10 was characterized as norovirus G II.4. Both viruses appeared to be identical.

4. Discussion

Here we present a proof of principle study using next-generation sequencing techniques for detection and characterization of different genotypes of norovirus in fecal samples. Direct multiplexed whole genome sequencing confirmed the results obtained by real time PCR with subsequent sequence analysis of the ORF1/ORF2 region of the norovirus genome.