

RESEARCH ARTICLE

Sequence Analysis of the Capsid Gene during a Genotype II.4 Dominated Norovirus Season in One University Hospital: Identification of Possible Transmission Routes

Barbara Juliane Holzkecht^{1,2*}, Kristina Træholt Franck^{1,3}, Rikke Thoft Nielsen^{2,4}, Blenda Böttiger^{1,5}, Thea Kølsen Fischer¹, Jannik Fonager¹

1 Department of Microbiological Diagnostics and Virology, Statens Serum Institut, Copenhagen, Denmark, **2** Department of Clinical Microbiology, Copenhagen University Hospital Herlev, Herlev, Denmark, **3** Research Unit for Clinical Microbiology, University of Southern Denmark, Odense, Denmark, **4** Department of Clinical Microbiology, Odense University Hospital, Odense, Denmark, **5** Medical Microbiology, Department of Laboratory Medicine Malmö, Lund University, Malmö, Sweden

* barbarah@web.de



OPEN ACCESS

Citation: Holzkecht BJ, Franck KT, Nielsen RT, Böttiger B, Fischer TK, Fonager J (2015) Sequence Analysis of the Capsid Gene during a Genotype II.4 Dominated Norovirus Season in One University Hospital: Identification of Possible Transmission Routes. PLoS ONE 10(1): e0115331. doi:10.1371/journal.pone.0115331

Academic Editor: Cheryl A. Stoddart, University of California, San Francisco, UNITED STATES

Received: August 5, 2014

Accepted: November 21, 2014

Published: January 15, 2015

Copyright: © 2015 Holzkecht et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All nucleotide sequences are available from GenBank. Capsid sequences have accession numbers KJ144938 to KJ145003, and polymerase sequences have accession numbers KJ956701 to KJ956727. A detailed assignment is available in [S1 Table](#).

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Norovirus (NoV) is a leading cause of gastroenteritis and genotype II.4 (GII.4) is responsible for the majority of nosocomial NoV infections. Our objective was to examine whether sequencing of the capsid gene might be a useful tool for the hospital outbreak investigation to define possible transmission routes. All NoV positive samples submitted from one university hospital during the 2007/8 season were selected. Genotyping of selected samples by partial polymerase gene sequencing had shown that the majority belonged to the GII.4 variant Den Haag 2006b and had identical polymerase sequences. Sequences of the capsid gene (1412 nucleotides) were obtained from the first available sample from 55 patients. From six immunocompromised patients with persistent infections a second sample was also included. As a control for a point-source outbreak, five samples from a foodborne outbreak caused by the same GII.4 variant were analyzed. Forty-seven of the inpatients (85%) were infected with the GII.4 variant Den Haag 2006b. Phylogenetic analysis of the Den Haag 2006b sequences identified four distinct outbreaks in different departments and a fifth outbreak with possible inter-department spread. In addition, a more heterogeneous cluster with evidence of repeated introductions from the community, but also possible inter-department spread was observed. In all six patients with paired sequences, evidence for *in vivo* evolution of the virus was found. Capsid gene sequencing showed substantial sequence variation among NoV GII.4 variant Den Haag 2006b strains from one single institution during a nine months' period. This method proved useful to understand the local epidemiology and, when used promptly, has the potential to make infection control measures more targeted.

Introduction

Norovirus (NoV) is a leading cause of acute gastroenteritis worldwide, causing both outbreaks and sporadic cases [1–6]. It is a positive-sense single-stranded non-enveloped RNA-virus belonging to the family *Caliciviridae*. As other RNA-viruses, it is characterized by a high mutation rate, leading to the rapid emergence of genetically diverse strains. At least five genogroups can be distinguished and subdivided into approximately 30 genotypes [1, 7]. The genotype II.4 (GII.4) is responsible for the majority of infections in healthcare settings [3]. Novel GII.4 variants emerge periodically and spread in a pandemic manner [8, 9] and there is growing evidence that this is due to antigenic drift in response to selective pressure from the host population [10–13].

In areas with a temperate climate, NoV infections occur mainly during the winter season. Outbreaks occur frequently in semi-closed or closed environments such as hospitals, schools and cruise ships [3] and can cause major health challenges in health care settings, in particular if effective control measures are not rapidly implemented.

Although symptomatic NoV infection in immunocompetent patients is usually limited to a few days, NoV can cause persistent infections in immunosuppressed patients accompanied by substantial *in vivo* evolution of the virus, leading to the co-existence of quasispecies [14–16].

For the detection and partial sequencing of NoV RNA, primers targeting the polymerase gene [17–19], the capsid gene [20] and the junction of the open reading frame (ORF) 1 and ORF2 [21, 22] have been developed. These methods have been useful in outbreak investigation, especially in determining whether an accumulation of NoV cases is composed of several different genotypes, likely reflecting individual introductions and sometimes called a pseudo outbreak [23]. However, sequencing of longer parts of the capsid gene, especially the hypervariable P2 domain, provides more detailed information and has recently been used to study transmission routes [24–28]. Whole genome sequencing using next generation sequencing technology has also proved feasible and useful in outbreak investigation [29].

Our objective was to investigate the possible transmission routes of GII.4 infections in one hospital during a nine months' period by sequencing of the capsid gene. The aim was to assess if this method can be used to differentiate sustained nosocomial transmission from repeated introductions of different strains from the community and to distinguish between distinct outbreaks in different hospital departments and inter-department spread.

Materials and Methods

Setting and sampling

All stool samples positive for NoV and submitted from of a 770-bed university hospital in the Copenhagen area, Denmark, to the virological department at Statens Serum Institut for diagnosis during November 2007 to July 2008, were identified from the laboratory database. One hundred-and-fifty-three samples from 105 patients had been submitted from seven different departments. Capsid and/or polymerase typing demonstrated that the majority (95% of 60 samples examined) belonged to GII.4.

The first available sample from each patient was included for analysis with a GII.4 specific reverse transcription polymerase chain reaction (RT-PCR). For 29 patients, samples were no longer available and for another 20 patients the GII.4 specific capsid RT-PCR was negative. One sample was excluded since previous genotyping had shown a genotype GI.3. Thus, sequences from a total of 55 patients were analyzed. From six of these patients, additional sequences were obtained from follow-up samples taken at least 14 days after the first sample. In addition, five samples from a foodborne outbreak caused by GII.4 variant Den Haag 2006b, occurring in 2008, were included in the study in order to analyze the capsid gene sequence

variation in a known point-source outbreak. In total, 66 capsid sequences were obtained and analyzed.

For 22 of the inpatients with available capsid sequences and for the five samples from the foodborne outbreak, polymerase gene sequences were available from previous genotyping.

Clinical data

Electronic patient records were reviewed for admission data, onset of gastrointestinal symptoms and presence of immunosuppressive disorders or prescription of immunosuppressive therapy.

Laboratory methods

NoV diagnostics and capsid and/or polymerase typing was performed as previously described [30].

For the GII.4 specific RT-PCR; RNA was extracted from 10% stool suspensions, which had been kept frozen at -80°C, with the MagNAPure LC (Roche Diagnostics, Hvidovre, Denmark). A NoV GII.4-specific nested RT-PCR amplifying 1571 nucleotides (nt) of the capsid gene (ORF2) was performed using primers as previously described [31]. Reverse transcription and the first PCR reaction was carried out in a volume of 25 µl (concentration of dNTP 0.4 mM, primers 0.6 µM, and RNase inhibitor 10 U) using the QIAGEN OneStep RT-PCR Kit (Qiagen Nordic Denmark) and the following cycling conditions: 42°C for 1 h, 94°C for 15 min, 40 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 2 min each and final extension at 72°C for 10 min. The second PCR reaction was carried out in a total volume of 25 µl (concentration of MgCl₂ 4 mM, dNTP 0.4 mM, primers 0.6 µM), using GeneAmp 10X PCR Buffer II, MgCl₂ and AmpliTaq DNA Polymerase (Applied Biosystems, Naerum, Denmark) and cycling conditions: 95°C for 5 min, 35 cycles of 95°C for 30 s, 52°C for 30 s, 72°C for 2 min and final extension at 72°C for 5 min. The PCR product was sequenced with the PCR primers, eventually supplemented with sequencing primers [31] using Big Dye Terminator kit 3.1 (Applied Biosystems, Naerum, Denmark).

Sequence analysis

Sequences were quality assessed and assembled using the BioNumerics version 6.6 software (Applied Maths, Sint-Martens-Latem, Belgium). BioNumerics has an in-built algorithm for sequence quality assessment and automatically discards low quality sequences before assembly. In addition, all sequence chromatograms were inspected manually and finally, all sequences were compared with publically available sequences deposited in GenBank via BLASTN and through manual alignments. For polymerase gene sequences, the full available sequences with a length ranging from 226 to 285 nucleotides (nt) were analyzed. For capsid gene sequences, a 1412 nucleotide (nt) region, corresponding to nt position 110 to 1521 of the reference sequence, was shared by all sequences and was used for further analysis. The norovirus sequence Hu/GII.4/Shellharbour/NSW696T/2006/AUS (GenBank no. EF684915.2) [32] was used as a reference. The P2 region was defined as described by Vega et al. [33], ranging from nt position 820 to 1254 of the reference capsid sequence. For genotype assignment, the online norovirus typing tool was used (<http://www.rivm.nl/mpf/norovirus/typingtool>) [34].

For further analysis, sequence alignment, conceptual translation to amino acid sequences, estimates of evolutionary divergence and phylogenetic analysis were performed in MEGA version 6 [35]. Sequence variation was defined as nucleotide changes among the sequences in this study or the detection of mixed bases (i.e. more than one peak at the same position in the sequence chromatogram in sequences with otherwise good quality). Sites with mixed bases

were not included in the estimates of evolutionary divergence and phylogenetic analysis. For the phylogenetic analysis, the Neighbor Joining (NJ) method (Jukes-Cantor model, bootstrap test with 1,000 replicates, uniform rates among sites, complete deletion of missing data) was used. To evaluate robustness of the phylogenetic analysis, it was repeated by constructing a Maximum Likelihood tree (General Time Reversible model, bootstrap test with 1,000 replicates, uniform rates among sites, complete deletion of missing data, otherwise default settings). Key amino acid positions in the P2 region which have been shown to be involved in receptor binding or to be predicted epitopes [36] were manually identified by alignment with the conceptually translated reference sequence.

Nucleotide sequence accession numbers

All sequences were submitted to GenBank. Capsid sequences have accession numbers KJ144938 to KJ145003, and partial polymerase sequences have accession numbers KJ956701 to KJ956727. A detailed assignment is available as supplemental material (S1 Table).

Statistical analysis and Graphics

The computer program “R: A language and environment for statistical computing”, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria; <http://www.R-project.org/>) was used for statistical analysis. For comparison of categorical data, Fisher’s exact test was used if expected values were <5 , otherwise the Chi-squared test was used. P-values ≤ 0.05 were considered statistically significant.

GraphPad Prism Version 5.04 (GraphPad Software, La Jolla, CA, USA) was used for graphics.

Ethics statement

The study was approved by the Danish Data Protection Agency (Record no. 2012–54–0202). Because of the retrospective design of the study, no written or verbal informed consent could be given by the patients. That being the case, access to patient records has to be approved by the Danish Health and Medicines Authority according to Danish law. This approval was obtained (Record no. 3-3013-292/1/). After retrieving information from electronic patient records data was anonymized by the research team for further analysis.

According to the “Danish Act on Research Ethics Review of Health Research Projects” this study does not require approval by the ethics committees, as it is considered a quality development project. This was confirmed by the Committees on Health Research Ethics for the Capital Region of Denmark in a specific waiver of approval (H-6-2014-FSP-055).

Results

Distribution of GII.4 variants

Capsid sequencing showed that the majority of the inpatients (47 of 55) were infected with the Den Haag 2006b variant and three and five inpatients were infected with the Osaka 2007 variant and the Apeldoorn 2007 variant, respectively. The foodborne outbreak was caused by the Den Haag 2006b variant.

In all 24 samples, where a genotype could be assigned from the polymerase sequence, it was congruent with the capsid genotype. Thus, no recombinants were detected.

Capsid sequence variation between and within variants

Comparing all 66 capsid sequences, nucleotide changes were observed in 238 of 1412 sites (17%). There were no deletions or insertions. The estimated evolutionary divergence between the three GII.4 variants is available as supporting information (S1 Fig).

Between four of the five sequences belonging to the Apeldoorn 2007 variant, no nucleotide changes were found. These samples were from patients in the same department and also temporarily related, while the fifth sequence differed by one nucleotide and was epidemiologically unrelated (sample obtained three months later in another department). Two sequences belonging to the Osaka 2007 variant did not show any nucleotide changes and had been obtained from patients in the same department taken with one day's interval. The third differed by two nucleotides and was obtained two weeks earlier in another department, suggesting that this case was not related to the other two.

Among the 58 Den Haag 2006b sequences, nucleotide changes were detected in 79 sites (6%), 31 of which were located in the P2 region. The number of nucleotide substitutions was higher within the P2 region (31 of 435 nt, 7%) than outside (48 of 977 nt, 5%), but this difference was not statistically significant (Chi-squared test, $p = 0.12$). However, substitutions in the codons of the P2 region were significantly more often located in the first or second position of the codon (45% of nucleotide substitutions in P2, 15% outside P2, Chi-squared test, $p = 0.006$) and transversions (as opposed to transitions) were also significantly more common (32% of nucleotide changes in P2, 4% outside P2, Fischer's Exact test, $p = 0.001$). As both of these observations are associated with non-synonymous nucleotide substitutions, 14 of the 18 amino acid changes occurred in the P2 region (Chi-squared test, $p < 0.001$). The 18 non-synonymous substitutions resulted in 17 amino acid alterations, 5 of which were located in key amino acid positions in the P2 region [36] (marked with asterisks in Table 1).

Table 1. Amino acid substitutions in the 58 translated GII.4 Den Haag 2006b sequences.

Reference position ^a	Amino acid substitution ^b
144	Ile / Met
244	Ile / Thr
245	Pro / Ser
297*	Arg / His
305	Ser / Ala
317	Ile / Thr
323	Thr / Ala
338	Thr / Ser
340*	Gly / Ala
355*	Ser / Gly
357	Pro / Asp
372*	Glu / Asp
376	Glu / Val
378	His / Leu
398	Asn / Ser
407*	Ser / Asn
501	His / Tyr

^a In relation to reference capsid sequence (GenBank Accession Nr. EF684915.2). Key amino acid positions [36] are marked with asterisks.

^b More common / less common.

doi:10.1371/journal.pone.0115331.t001

To exclude the possibility of RT-PCR introduced nucleotide changes or sequencing mistakes, cDNA synthesis, PCR and sequencing was repeated for 5 samples with unique substitutions not shared with any of the other sequences. This repeated testing showed 100% reproducibility for a total of ten nucleotide substitutions.

In addition to the above described nucleotide changes, a total of 47 mixed bases were detected in 15 sequences at 39 different positions. These mixed bases indicate quasispecies, as a mixed signal would be expected, if more than one viral strain has been co-PCR amplified and co-sequenced. The 15 sequences were from samples of 12 different patients, ten of which were receiving immunosuppressive therapy (one cardiology patient receiving high dose prednisolone, one nephrology patient receiving immunosuppressive therapy after renal transplantation and eight hematology patients receiving chemotherapy).

Phylogenetic analysis and molecular epidemiology

The first capsid sequence from each of the 55 inpatients, the five samples from the foodborne outbreak and the reference sequence were included in the phylogenetic analysis.

[Fig. 1](#) shows the Neighbor Joining subtree of the Den Haag 2006b capsid sequences. There were 23 distinct sequence patterns which formed seven clusters and four singletons. Clusters 2 to 7 were supported by bootstrap values >80 . Cluster 1 was only supported by a bootstrap value of 36, but as it was comprised of five distinct sequence patterns only differing by four nucleotide changes it was considered a cluster in this study. When the phylogenetic analysis was repeated with the Maximum Likelihood method, the separation in these seven clusters and four singletons did not change ([S2 Fig.](#)). The estimated evolutionary divergence between the different clusters and singletons is available as supporting information ([S1 Fig.](#)).

Cluster 7 represented the foodborne point source outbreak and these five sequences were identical. Four of the six hospital clusters were locally confined to one department. The three smallest clusters (3, 5 and 6) were restricted to one department each and to a period of less than one month and the sequences differed by no more than one nucleotide substitution. Thus, in these clusters the molecular epidemiology confirmed the epidemiological relation and added the information, that each of these clusters was due to a new introduction of the virus into the hospital environment. Cluster 4 included 13 samples from the hematology department over a period of 78 days. Six distinct sequence patterns differing at 12 sites (0.8%) were observed. A more detailed description of the sequence variation in this cluster, taking into account mixed bases and follow up samples from four of the patients, is given in [Fig. 2](#) and described more detailed in the next section.

The clusters 1 and 2 included sequences from samples from different departments. Cluster 2 was composed of sequences from eight samples obtained from the nephrology department and one from the hematology department. The earliest of the nephrology samples was taken one day after admission in a patient undergoing chronic hemodialysis. The ninth sample was from the hematology department and was taken at the patient's admission two weeks earlier. Thus, this cluster could have been caused by either two separate introductions or inter-department spread of the virus. Cluster 1 was epidemiologically more heterogeneous, with sequences of samples obtained from four different departments. The first sample (Neph-12/18/2007) was taken in the nephrology department in December 2007. About two months later, four samples belonging to the same cluster were taken in the cardiology department within three weeks; one of these samples differed by one nucleotide from the other three. Another two samples in this cluster were taken in the internal medicine department 1 within three days after the first sample from the cardiology department (IntMed1-01/29/2008, IntMed1-01/31/2008). One of these had an identical sequence compared to the three cardiology sequences and the

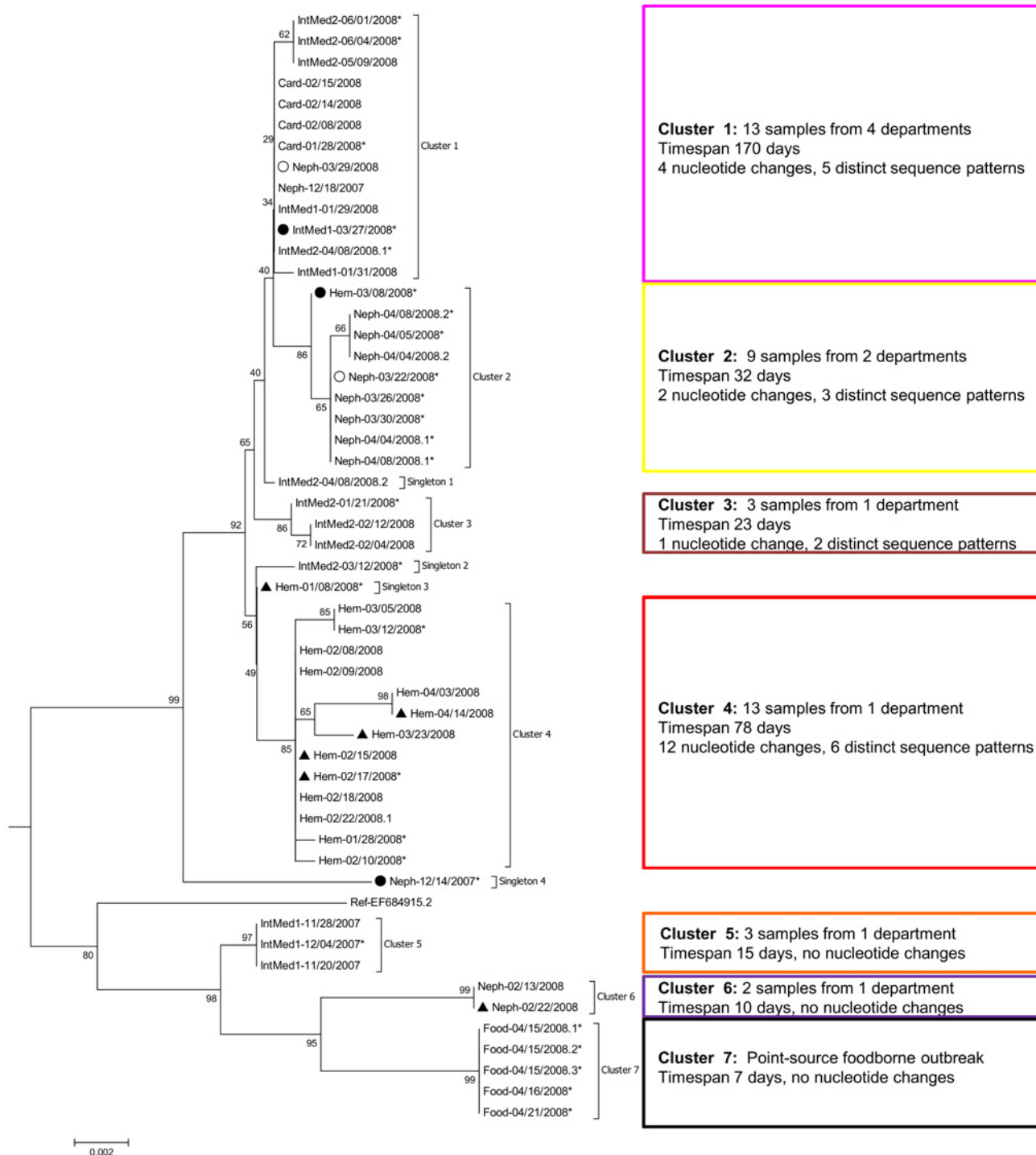


Figure 1. Phylogenetic tree of 53 NoV genotype II.4 variant Den Haag 2006b capsid sequences. All capsid sequences (first sample per patient) were aligned and a Neighbor Joining tree was constructed and rooted on the branch between the Den Haag 2006b variant and the Osaka 2007 and Apeldoorn 2007 sequences. The subtree of the 2006b variant is displayed here. Sites with mixed bases were ignored for phylogenetic analysis. Sequence names represent hospital department (Hem: hematology, Neph: nephrology, IntMed1/2: internal medicine 1/2, Card: cardiology, Food: foodborne outbreak) and date of sampling (mm/dd/yyyy), eventually followed by a serial number to ensure uniqueness. Sequences from patients with symptom onset within 48 hours after hospital admission are marked with a circle (filled circle: no previous contact to hospital, empty circle: frequent contact to hospital due to hemodialysis). Sequences from the first samples from six patients with persistent diarrhea are marked with a filled triangle. Sequences from samples with an available polymerase gene sequence from previous genotyping (n = 27) are marked with asterisks and phylogenetic analysis of this subgroup is shown in Fig. 4. On the right hand side of the tree epidemiological information is given as well as numbers of nucleotide changes within the cluster and the number of distinct sequence patterns included in the cluster.

doi:10.1371/journal.pone.0115331.g001

Sequence name	Patient No.	Nucleotide position																												
		274	432	436	498	525	615	694	731	831	874	883	884	890	950	1012	1037	1105	1110	1114	1116	1173	1177	1178	1193	1194	1242	1344	1419	1501
Hem-01/28/2008		T	A	G	T	C	G	T	C	A	C	G	G	G	T	A	G	A	T	G	C	T	A	G	A	T	T	T	C	C
Hem-02/08/2008		T
Hem-02/09/2008		T
Hem-02/10/2008		T	G
Hem-02/15/2008	1.1	T
Hem-02/17/2008	2.1	T	A/G	.	.	C/G
Hem-02/18/2008		T
Hem-02/22/2008.1		T
Hem-03/05/2008		.	.	.	C/T	.	.	T	G	G	C	.	.
Hem-03/06/2008	2.2	T	A/G	.	.	.	A/G
Hem-03/12/2008		T	G	G	C	.	.
Hem-03/23/2008	3.1	C	T	.	A/C	.	.	A	C	A/T	.	.	.
Hem-04/03/2008		.	G	T	A	.	T	.	.	.	C	T
Hem-04/14/2008	4.1	.	G	T	A	.	T	.	.	.	C	T
Hem-04/30/2008	4.2	.	G	A/G	C/T	.	A/G	C/T	T	.	.	A/G	A/G	A/G	.	A/T	C/G	.	C/T	A/G	.	C/T	C/T	T	
Hem-05/02/2008	1.2	T	C	A/G
Hem-05/07/2008	3.2	C	T	.	A/C	.	.	A/G	A/G	.	.	C	A/T

Figure 2. Sequence variation in cluster 4 of the GII.4 Den Haag 2006b variant. Seventeen sequences from cluster 4 are shown, including follow up samples from four patients with chronic NoV infection. Sequences are sorted chronologically and nucleotide positions with mixed bases and/or nucleotide changes are shown. Sequence names indicate department (Hem: hematology), sample date (mm/dd/yyyy) and eventual serial number. For patients with chronic NoV infections, a patient number is given followed by “.1” for first sample and “.2” for second sample (marked with grey background). The nucleotide position is given in relation to the reference capsid sequence from GenBank Accession Nr. EF684915.2. Period indicates the same nucleotide as in the first sample (Hem-01/28/2008).

doi:10.1371/journal.pone.0115331.g002

other one differed by one nucleotide. This indicates that transmission between these two departments has occurred. Six weeks after the last sample was taken in the cardiology department, a sample taken from a patient in the internal medicine department 1 (IntMed1-03/27/2008) revealed a sequence differing by one nucleotide. As this sample was taken within 48 hours after admission in a patient without previous hospital contact, this most likely represents a new introduction of the virus into the hospital environment. A sequence obtained from a sample taken two days later from a chronic hemodialysis patient within 48 hours after admission (Neph-03/29/2008) and a third sequence taken two weeks later from a patient in the internal medicine department 2 (IntMed2-04/08/2008.1) were identical. Inter-department spread could have occurred here. Again one month later, three samples from the internal medicine department 2 yielded identical sequences differing from the last mentioned ones by one nucleotide substitution. This could either represent a new introduction or an ongoing outbreak in that department. Taken together, in cluster one, most likely several new introductions have taken place, but inter-department spread is also likely to have occurred (at least between the cardiology department and the internal medicine department 1).

One sample with a sporadic sequence (“singleton 1”) was taken on the same day and in the same department as a sample from cluster 1 (IntMed2-04/08/2008.1). The sequences from these two samples differed only by one nucleotide substitution (represented by the short horizontal branch length), making nosocomial transmission very likely and indicating that a more reasonable classification of this sequence would be in cluster 1.

The temporal distribution of the different genotypes, variants and clusters in each department is shown in Fig. 3. This figure illustrates consecutive outbreaks with distinct clusters of the GII.4 Den Haag 2006b variant in the nephrology department and both internal medicine departments. It also demonstrates the occurrence of sequences belonging to cluster 1 during a prolonged time period in four different departments. The temporarily related cases in the cardiology department and the internal medicine department 1 (week 5–7/2008) likely represent inter-department spread. During week 13–15/2008, three samples belonging to cluster 1 were taken in three different departments (nephrology, internal medicine 1 and 2), also suggesting inter-department spread.

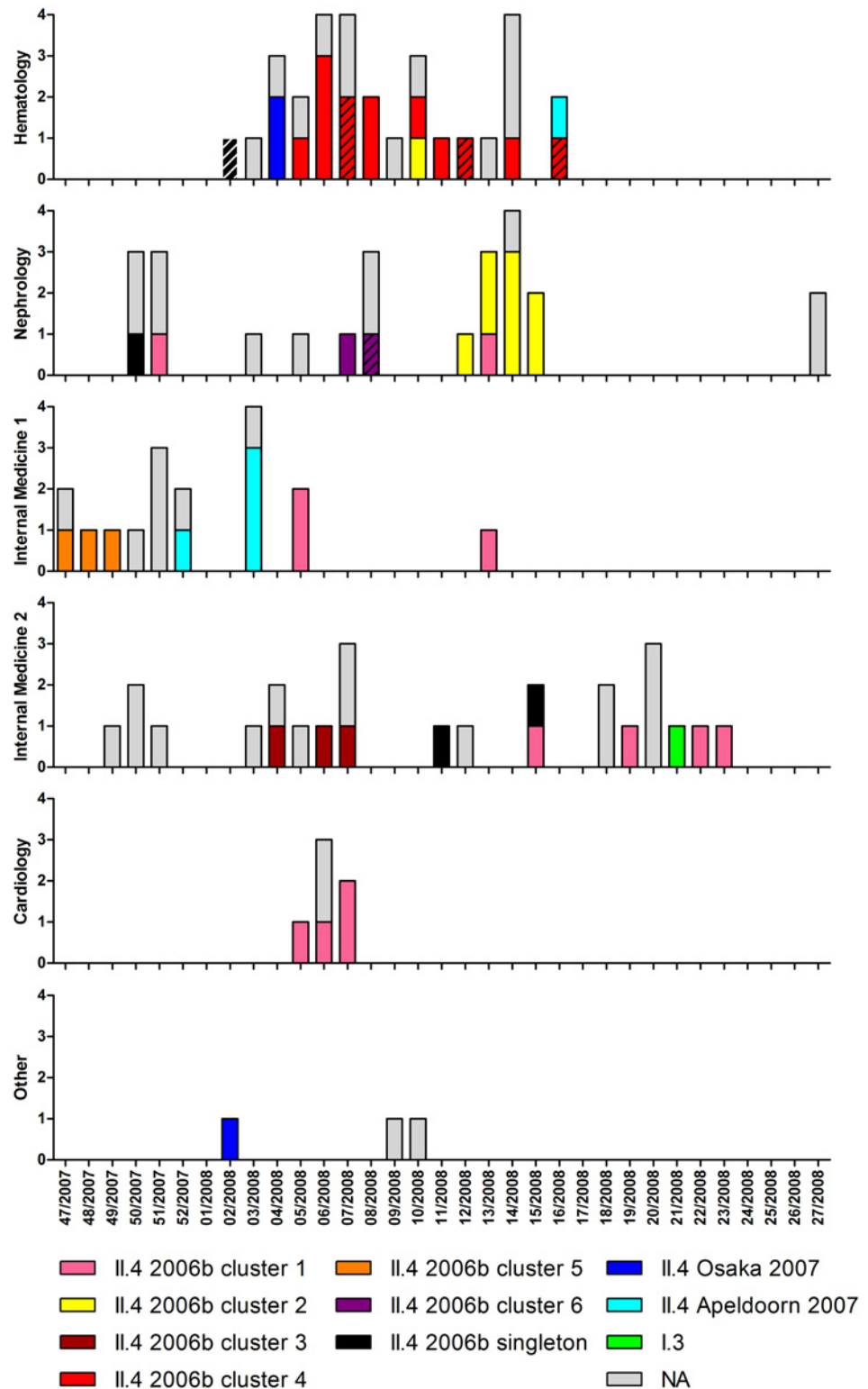


Figure 3. Local and temporal distribution of genotypes, variants and GII.4 2006b clusters. The first sample per inpatient is shown (n = 105). Units on vertical axes represent numbers of patients. NA (n = 49): no sequence information available (n = 29: sample not stored, n = 20: GII.4 specific PCR negative). The sequences from the first samples from six patients with persistent diarrhea are marked with diagonal shading.

doi:10.1371/journal.pone.0115331.g003

strain. However, sequence variation (in 2–13 positions, median 3) was detected between all paired samples and mixed bases were observed in all later samples. In total, mixed bases indicating NoV quasispecies were detected at 37 sites in the paired samples. A significantly higher frequency of mixed bases were observed in the second samples (32 in six 1412 nt sequences) than in the first samples (five mixed bases in six 1412 nt sequences, Chi-squared test, $p < 0.001$).

The nucleotide changes and mixed bases in the sequences from the four patients belonging to cluster 4, together with the remaining sequences of that cluster, are shown in [Fig. 2](#). Sustained nucleotide changes, occurring in more than one patient and suggesting person-to-person transmission after *in vivo* evolution, were found in nine positions.

Discussion

We used sequencing of a large part of the capsid gene to investigate NoV infections in a university hospital during a nine months' period in the winter season 2007/2008. Routine polymerase genotyping had shown that the majority of strains belonged to the GII.4 Den Haag 2006b variant and that most of the sequences were identical ([Fig. 4](#)). For further outbreak investigation, a method with a higher resolution was needed and substantially longer sequences of the variable capsid gene were therefore examined.

The phylogenetic analysis of the 1412 nt long capsid sequence showed that the accumulation of NoV infections in the hospital was due to several outbreaks mostly confined to one department each. In two clusters (GII.4 Den Haag 2006b clusters 1 and 2), possible inter-department spread was observed. The heterogeneous cluster 1 contained sequences from four different departments over a time span of nearly half a year. One of the later samples in this cluster was taken from a patient without earlier hospital contact within 48 hours after admission, thus indicating a new introduction of the virus from the community into the hospital. However, inter-department spread could have occurred in temporarily related samples from this cluster. An example of this could be the simultaneously occurring norovirus infections in the cardiology department and the internal medicine department 1 in week 5–7/2008. In other temporary and/or locally related cases with sequences differing by no more than one nucleotide the transmission route was unclear. Identical sequences belonging to cluster 1 were for example found in three patients on three different wards in week 13–15/2008. Both nosocomial transmission and repeated introduction from the community could explain these infections. The same is true for the first two samples from cluster 2 which were from patients from two different departments.

In the nephrology department and both internal medicine departments, consecutive outbreaks with distinct clusters of the GII.4 variant Den Haag 2006b were detected, indicating repeated introductions of different strains from the community. As opposed to this, a prolonged outbreak over 12 weeks with the same cluster occurred in the hematology department.

Sequencing of the capsid gene has previously been shown to be helpful for investigating outbreaks. Comparable to our study, which found a presumed larger hospital outbreak to consist of several distinct smaller outbreaks, Xerry et al. [25] could differentiate two distinct simultaneous outbreaks in different departments of the same hospital and Sukhrie et al. [24] found that one of four epidemiological outbreaks involved several unrelated strains. However, in the study of Xerry et al., P2 sequences also detected links between presumably unrelated outbreaks in different departments and even two different hospitals [25]. Similarly, in another study by Sukhrie and coworkers only three out of 14 molecular clusters found by sequencing of the P2 region had been identified through epidemiological investigations, even though the conservative approach of defining a molecular cluster by 100% sequence identity was chosen [28].

The frequency of nucleotide changes found in our study (0 to 12 substitutions per cluster in 1412 nt sequences) is in concordance with previous studies investigating nosocomial NoV GII.4 outbreaks: Dingle et al. showed a diversity of six nucleotide changes in the entire NoV genome in a 1-week outbreak with presumed person-to-person transmission, whereas the strains of an outbreak with a presumed point source only differed by one substitution [26]. Sukhrie et al. found a similar sequence variation of two to 22 nucleotide changes in the P2 domain within prospectively monitored nosocomial outbreaks, where patients and health care workers were sampled once weekly for one to two months [24]. However, in contrast to these and our findings, Xerry et al. described 100% identity in P2 regions of strains in different nosocomial clusters, the largest of which involving four wards in two different hospitals over a seven weeks' period where 24 samples were examined [25].

A possible epidemiological explanation for outbreaks consisting of identical strains could be a point source such as contaminated food (as seen in the foodborne outbreak in our study) or continuous transmission of the same strain from a contaminated environment, the extent of which is still debated [37]. In our study, no nucleotide changes were found within the sequences in four clusters representing outbreaks confined to one distinct department within a one month's period (Den Haag 2006b clusters 5 and 6, the Osaka 2007 cluster and the Apeldoorn 2007 cluster). In two clusters (Den Haag 2006b clusters 2 and 3) closely related sequence patterns (one to two nucleotide changes per 1412 nt) were found temporarily related in the same department. Our study does not permit a definite conclusion whether these outbreaks with identical or nearly identical sequences were due to environmental transmission or shorter person-to-person transmissions between individuals with acute infections without evidence of *in vivo* evolution.

In immunocompromised patients, the *in vivo* evolution of the virus leading to the development of quasispecies has been shown [14–16]. In accordance to this, we found the highest sequence diversity in the prolonged outbreak in the hematology department, where the majority of patients was immunosuppressed either due to their underlying disease or induced by chemotherapy. This outbreak (Den Haag 2006b cluster 4) included four of the six patients, where sequences from paired samples were available. The high frequency of mixed bases in these patients' second samples shows that viral quasispecies had evolved and there was indication for subsequent person-to-person transmission (Fig. 2). The frequencies of sequence variation in paired samples from immunocompromised patients with chronic NoV infection in our study are in accordance with previous studies [16, 26].

Repeated analysis of the data by generation of a Maximum Likelihood tree allocated the sequences to the same clusters (S2 Fig.), demonstrating the method's robustness despite of the low bootstrap value of cluster 1. However, one of the Den Haag 2006b singleton sequences ("singleton 1") was obtained from a sample taken on the same day and in the same department as a sample belonging to the heterogeneous cluster 1. As these two sequences only differ by one nucleotide substitution, nosocomial transmission is very likely. This underlines that manual inspection of the sequences, eventually supplemented with epidemiological information, is helpful in interpreting phylogenetic analysis in areas of uncertainty.

This study was designed to investigate sequence variation as an indicator for NoV evolution to study transmission routes. However, further characterization of the nucleotide changes showed interesting findings. Transversions (compared to transitions) and nucleotide changes in the first two codon positions were significantly more common within the P2 region than outside the P2 region, as were non-synonymous substitutions. These findings are in accordance with previous studies and indicate positive selection in this surface-exposed region, which is consistent with its predicted key role for antigenic variation and receptor binding [10, 13, 16, 38]. Five amino acid substitutions were found in key positions of the Den Haag 2006b variant

which shows that potentially functional changes in virus—host interacting sites can occur even in a locally and temporarily confined setting.

A drawback of this study is that the available sequences only represent a part of the actual burden of NoV infections. Usually, most samples are taken in the beginning of an outbreak and subsequent cases are diagnosed based on clinical findings. In addition, in our study a substantial part of the samples were not available or sequences could not be obtained. However, as the phylogenetic analysis shows a clear clustering according to department and time of sampling, we feel confident that this bias has not influenced the major findings and conclusions of the study.

This study also lacks information about the genetic variability of strains occurring in the community during the study period. Especially for the interpretation of the heterogeneous Den Haag 2006b cluster 1, this information would have been valuable.

Interestingly, it was difficult to find a clear consensus in the literature about the exact demarcation of the P2 region in NoV GII.4 strains. Several slightly different regions, ranging in length from 125 to 152 amino acids, have been reported [13, 25, 26, 33]. In this study, the definition from a recent article referring to a Den Haag 2006b variant [33] was used. However, the above described characteristics of nucleotide substitutions were still applicable, when using the other cited definitions (data not shown).

In conclusion, sequencing of the capsid gene proved useful for the investigation of a NoV season dominated by the GII.4 variant Den Haag 2006b in one university hospital. Polymerase sequences which were available from genotyping performed continuously during the outbreak were identical. Phylogenetic analysis of the capsid sequences, however, showed that repeated introductions were common, leading to smaller outbreaks, which were most often confined to a single department. However, a few cases of possible inter-department transmission were also detected. If promptly available, these results would have been of value for outbreak management. In this setting, one should have focused on correct isolation measures upon patient admission to prevent subsequent intra-department spread instead of regarding the outbreak as a pan-institutional outbreak. The sustained outbreak on the hematology ward was shown to be due to repeated person-to-person transmission indicating that outbreak management should have focused on isolation measures of chronically infected patients.

Supporting Information

S1 Table. Assignment of accession numbers and sequence names. The accession numbers and sequence names of the 66 capsid sequences and 27 polymerase sequences included in this study are given.

(PDF)

S1 Fig. Estimates of evolutionary divergence over sequence pairs between GII.4 variants (a.) and GII.4 Den Haag 2006b clusters (b.). The number of base substitutions per site from averaging over all sequence pairs are shown. Analyses were conducted using the Jukes-Cantor model.

(TIF)

S2 Fig. Maximum Likelihood (ML) tree of 53 NoV genotype II.4 variant Den Haag 2006b sequences. A ML tree of all capsid sequences (first sample per patient) was constructed and rooted on the branch between the Den Haag 2006b variant and the Osaka 2007 and Apeldoorn 2007 sequences. The subtree of the 2006b variant is displayed here. Sites with mixed bases were ignored for phylogenetic analysis.

Sequence names represent hospital department (Hem: hematology, Neph: nephrology, IntMed1/2: internal medicine 1/2, Card: cardiology, Food: foodborne outbreak) and date of sampling (mm/dd/yyyy), eventually followed by a serial number to ensure uniqueness.

Sequences from patients with symptom onset within 48 hours after hospital admission are marked with a circle (filled circle: no previous contact to hospital, empty circle: frequent contact to hospital due to hemodialysis). Sequences from the first samples from six patients with persistent diarrhea are marked with a filled triangle. Sequences from samples with an available polymerase gene sequence from routine genotyping ($n = 27$) are marked with asterisks. (TIF)

Acknowledgments

We thank Jesper Rønn for his technical assistance.

Meeting presentation: Preliminary data of this study has partly been presented as a poster (P1815) at the 23rd European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) in Berlin, Germany, 27.-30.04.2013.

Author Contributions

Conceived and designed the experiments: BJH KTF BB TKF JF. Performed the experiments: BJH KTF RTN JF. Analyzed the data: BJH JF. Contributed reagents/materials/analysis tools: BB TKF JF. Wrote the paper: BJH KTF RTN BB TKF JF.

References

1. Glass RI, Parashar UD, Estes MK (2009) Norovirus gastroenteritis. *N Engl J Med* 361: 1776–1785. doi: [10.1056/NEJMra0804575](https://doi.org/10.1056/NEJMra0804575)
2. Patel MM, Widdowson MA, Glass RI, Akazawa K, Vinje J, et al. (2008) Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg Infect Dis* 14: 1224–1231. doi: [10.3201/eid1408.071114](https://doi.org/10.3201/eid1408.071114)
3. Hall AJ, Vinje J, Lopman B, Park GW, Yen C, et al. (2011) Updated Norovirus Outbreak Management and Disease Prevention Guidelines. *Morbidity and Mortality Weekly Report* 60: 1–15.
4. Payne DC, Vinje J, Szilagyi PG, Edwards KM, Staat MA, et al. (2013) Norovirus and medically attended gastroenteritis in U.S. children. *N Engl J Med* 368: 1121–1130. doi: [10.1056/NEJMsa1206589](https://doi.org/10.1056/NEJMsa1206589)
5. Barclay L, Park GW, Vega E, Hall A, Parashar U, et al. (2014) Infection control for norovirus. *Clin Microbiol Infect* 20:731–740. doi: [10.1111/1469-0691.12674](https://doi.org/10.1111/1469-0691.12674) PMID: [24813073](https://pubmed.ncbi.nlm.nih.gov/24813073/)
6. Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, et al. (2014) Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect Dis* 14: 725–730. doi: [10.1016/S1473-3099\(14\)70767-4](https://doi.org/10.1016/S1473-3099(14)70767-4) PMID: [24981041](https://pubmed.ncbi.nlm.nih.gov/24981041/)
7. Kroneman A, Vega E, Vennema H, Vinje J, White PA, et al. (2013) Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol* 158: 2059–2068. doi: [10.1007/s00705-013-1708-5](https://doi.org/10.1007/s00705-013-1708-5) PMID: [23615870](https://pubmed.ncbi.nlm.nih.gov/23615870/)
8. Siebenga JJ, Vennema H, Zheng DP, Vinje J, Lee BE, et al. (2009) Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J Infect Dis* 200: 802–812. doi: [10.1086/605127](https://doi.org/10.1086/605127) PMID: [19627248](https://pubmed.ncbi.nlm.nih.gov/19627248/)
9. van Beek J, Ambert-Balay K, Botteldoorn N, Eden J, Fonager J, et al. (2013) Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro Surveill* 18: 8–9. PMID: [23305715](https://pubmed.ncbi.nlm.nih.gov/23305715/)
10. Lindesmith LC, Donaldson EF, LoBue AD, Cannon JL, Zheng DP, et al. (2008) Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med* 5: e31. Available: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0050031>. Accessed 30 June 2014. doi: [10.1371/journal.pmed.0050031](https://doi.org/10.1371/journal.pmed.0050031) PMID: [18271619](https://pubmed.ncbi.nlm.nih.gov/18271619/)
11. Bull RA, White PA (2011) Mechanisms of GII.4 norovirus evolution. *Trends Microbiol* 19: 233–240. doi: [10.1016/j.tim.2011.01.002](https://doi.org/10.1016/j.tim.2011.01.002) PMID: [21310617](https://pubmed.ncbi.nlm.nih.gov/21310617/)

12. Bull RA, Eden JS, Rawlinson WD, White PA (2010) Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog* 6: e1000831. Available: <http://www.plospathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1000831>. Accessed 30 June 2014. doi: [10.1371/journal.ppat.1000831](https://doi.org/10.1371/journal.ppat.1000831) PMID: [20360972](https://pubmed.ncbi.nlm.nih.gov/20360972/)
13. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van de Veer B, et al. (2007) Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J Virol* 81: 9932–9941. doi: [10.1128/JVI.00674-07](https://doi.org/10.1128/JVI.00674-07) PMID: [17609280](https://pubmed.ncbi.nlm.nih.gov/17609280/)
14. Bok K, Green KY (2012) Norovirus gastroenteritis in immunocompromised patients. *N Engl J Med* 367: 2126–2132. doi: [10.1056/NEJMra1207742](https://doi.org/10.1056/NEJMra1207742) PMID: [23190223](https://pubmed.ncbi.nlm.nih.gov/23190223/)
15. Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, et al. (2012) Contribution of intra- and interhost dynamics to norovirus evolution. *J Virol* 86: 3219–3229. doi: [10.1128/JVI.06712-11](https://doi.org/10.1128/JVI.06712-11) PMID: [22205753](https://pubmed.ncbi.nlm.nih.gov/22205753/)
16. Nilsson M, Hedlund KO, Thorhagen M, Larson G, Johansen K, et al. (2003) Evolution of human calicivirus RNA *in vivo*: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J Virol* 77: 13117–13124. doi: [10.1128/JVI.77.24.13117-13124.2003](https://doi.org/10.1128/JVI.77.24.13117-13124.2003) PMID: [14645568](https://pubmed.ncbi.nlm.nih.gov/14645568/)
17. Schreier E, Döring F, Künkel U (2000) Molecular epidemiology of outbreaks of gastroenteritis associated with small round structured viruses in Germany in 1997/98. *Arch Virol* 145: 443–453. doi: [10.1007/s007050050038](https://doi.org/10.1007/s007050050038) PMID: [10795514](https://pubmed.ncbi.nlm.nih.gov/10795514/)
18. Vennema H, de Bruin E, Koopmans M (2002) Rational optimization of generic primers used for Norwalk-like virus detection by reverse transcriptase polymerase chain reaction. *J Clin Virol* 25: 233–235. doi: [10.1016/S1386-6532\(02\)00126-9](https://doi.org/10.1016/S1386-6532(02)00126-9) PMID: [12367660](https://pubmed.ncbi.nlm.nih.gov/12367660/)
19. Hoebe CJ, Vennema H, de Roda Husman AM, van Duynhoven YT (2004) Norovirus outbreak among primary schoolchildren who had played in a recreational water fountain. *J Infect Dis* 189: 699–705. doi: [10.1086/381534](https://doi.org/10.1086/381534) PMID: [14767824](https://pubmed.ncbi.nlm.nih.gov/14767824/)
20. Kojima S, Kageyama T, Fukushi S, Hoshino FB, Shinohara M, et al. (2002) Genogroup-specific PCR primers for detection of Norwalk-like viruses. *J Virol Methods* 100: 107–114. doi: [10.1016/S0166-0934\(01\)00404-9](https://doi.org/10.1016/S0166-0934(01)00404-9) PMID: [11742657](https://pubmed.ncbi.nlm.nih.gov/11742657/)
21. Kageyama T, Kojima S, Shinohara M, Uchida K, Fukushi S, et al. (2003) Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *J Clin Microbiol* 41: 1548–1557. doi: [10.1128/JCM.41.4.1548-1557.2003](https://doi.org/10.1128/JCM.41.4.1548-1557.2003) PMID: [12682144](https://pubmed.ncbi.nlm.nih.gov/12682144/)
22. Gallimore CI, Cheesbrough JS, Lamden K, Bingham C, Gray JJ (2005) Multiple norovirus genotypes characterised from an oyster-associated outbreak of gastroenteritis. *Int J Food Microbiol* 103: 323–330. doi: [10.1016/j.ijfoodmicro.2005.02.003](https://doi.org/10.1016/j.ijfoodmicro.2005.02.003) PMID: [15967530](https://pubmed.ncbi.nlm.nih.gov/15967530/)
23. Rahamat-Langendoen JC, Lokate M, Schölvincq EH, Friedrich AW, Niesters HG (2013) Rapid detection of a norovirus pseudo-outbreak by using real-time sequence based information. *J Clin Virol* 58: 245–248. doi: [10.1016/j.jcv.2013.06.034](https://doi.org/10.1016/j.jcv.2013.06.034) PMID: [23880160](https://pubmed.ncbi.nlm.nih.gov/23880160/)
24. Sukhrie FH, Teunis P, Vennema H, Bogerman J, van Marm S, et al. (2013) P2 domain profiles and shedding dynamics in prospectively monitored norovirus outbreaks. *J Clin Virol* 56: 286–292. doi: [10.1016/j.jcv.2012.12.006](https://doi.org/10.1016/j.jcv.2012.12.006) PMID: [23294532](https://pubmed.ncbi.nlm.nih.gov/23294532/)
25. Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ (2008) Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses. *J Clin Microbiol* 46: 947–953. doi: [10.1128/JCM.02240-07](https://doi.org/10.1128/JCM.02240-07) PMID: [18216210](https://pubmed.ncbi.nlm.nih.gov/18216210/)
26. Dingle KE, Norovirus Infection Control in Oxfordshire Communities Hospitals (2004) Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. *J Clin Microbiol* 42: 3950–3957. doi: [10.1128/JCM.42.9.3950-3957.2004](https://doi.org/10.1128/JCM.42.9.3950-3957.2004) PMID: [15364974](https://pubmed.ncbi.nlm.nih.gov/15364974/)
27. Xerry J, Gallimore CI, Iturriza-Gomara M, Gray JJ (2009) Tracking the transmission routes of genogroup II noroviruses in suspected food-borne or environmental outbreaks of gastroenteritis through sequence analysis of the P2 domain. *J Med Virol* 81: 1298–1304. doi: [10.1002/jmv.21517](https://doi.org/10.1002/jmv.21517) PMID: [19475614](https://pubmed.ncbi.nlm.nih.gov/19475614/)
28. Sukhrie FH, Beersma MF, Wong A, van de Veer B, Vennema H, et al. (2011) Using molecular epidemiology to trace transmission of nosocomial norovirus infection. *J Clin Microbiol* 49: 602–606. doi: [10.1128/JCM.01443-10](https://doi.org/10.1128/JCM.01443-10) PMID: [21159934](https://pubmed.ncbi.nlm.nih.gov/21159934/)
29. Kundu S, Lockwood J, Depledge DP, Chaudhry Y, Aston A, et al. (2013) Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. *Clin Infect Dis* 57: 407–414. doi: [10.1093/cid/cit287](https://doi.org/10.1093/cid/cit287) PMID: [23645848](https://pubmed.ncbi.nlm.nih.gov/23645848/)
30. Franck KT, Fonager J, Ersbøll AK, Böttiger B (2014) Norovirus epidemiology in community and health care settings and association with patient age, Denmark. *Emerg Infect Dis* 20: 1123–1131. doi: [10.3201/eid2007.130781](https://doi.org/10.3201/eid2007.130781) PMID: [24960024](https://pubmed.ncbi.nlm.nih.gov/24960024/)

31. Fonager J, Hindbaek LS, Fischer TK (2013) Rapid emergence and antigenic diversification of the norovirus 2012 Sydney variant in Denmark, October to December, 2012. *Euro Surveill* 18: pii:20413. Available: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20413>. Accessed 30 June 2014. PMID: [23470017](#)
32. Tu ET, Bull RA, Greening GE, Hewitt J, Lyon MJ, et al. (2008) Epidemics of gastroenteritis during 2006 were associated with the spread of norovirus GII.4 variants 2006a and 2006b. *Clin Infect Dis* 46: 413–420. doi: [10.1086/525259](https://doi.org/10.1086/525259) PMID: [18177226](#)
33. Vega E, Barclay L, Gregoricus N, Williams K, Lee D, et al. (2011) Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerg Infect Dis* 17: 1389–1395. PMID: [21801614](#)
34. Kroneman A, Vennema H, Deforche K, Avoort H, Penaranda S, et al. (2011) An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 51: 121–125. doi: [10.1016/j.jcv.2011.03.006](https://doi.org/10.1016/j.jcv.2011.03.006) PMID: [21514213](#)
35. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](#)
36. Zakikhany K, Allen DJ, Brown D, Iturriza-Gomara M (2012) Molecular evolution of GII-4 Norovirus strains. *PLoS One* 7: e41625. Available: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0041625>. Accessed 30 June 2014. doi: [10.1371/journal.pone.0041625](https://doi.org/10.1371/journal.pone.0041625) PMID: [22844506](#)
37. Lopman B, Gastanaduy P, Park GW, Hall AJ, Parashar UD, et al. (2012) Environmental transmission of norovirus gastroenteritis. *Curr Opin Virol* 2: 96–102. doi: [10.1016/j.coviro.2011.11.005](https://doi.org/10.1016/j.coviro.2011.11.005) PMID: [22440972](#)
38. Debbink K, Lindesmith LC, Donaldson EF, Baric RS (2012) Norovirus immunity and the great escape. *PLoS Pathog* 8: e1002921. Available: <http://www.plospathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1002921>. Accessed 30 June 2014. doi: [10.1371/journal.ppat.1002921](https://doi.org/10.1371/journal.ppat.1002921) PMID: [23093932](#)