



P2 domain profiles and shedding dynamics in prospectively monitored norovirus outbreaks

Faizel H.A. Sukhrie^{a,b,*}, Peter Teunis^{c,d}, Harry Vennema^a, Jolanda Bogerman^e, Sebastian van Marm^a, Matthias F.C. Thijs Beersma^b, Marion Koopmans^{a,b}

^a Laboratory for Infectious Diseases and Perinatal Screening, Centre for Infectious Disease Control (RIVM), Bilthoven, The Netherlands

^b Department of Virology, Erasmus Medical Center, Rotterdam, The Netherlands

^c Epidemiology and Surveillance Unit, Centre for Infectious Disease Control (RIVM), Bilthoven, The Netherlands

^d Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, USA

^e Municipal Health Service, Rotterdam, The Netherlands

ARTICLE INFO

Article history:

Received 6 September 2012

Received in revised form 3 December 2012

Accepted 7 December 2012

Keywords:

Norovirus transmission

Shedding

P2 domain

ABSTRACT

Background: Norovirus P2 domain is commonly used to extrapolate transmission within an outbreak (OB) setting. The current definition is that transmission among cases is considered to be proven when no sequence variation is found.

Objectives: Previous studies have shown a high mutation rate and errors during replication of the norovirus genome, therefore the validity of this criterion must be evaluated.

Study design: Sequences of the P2 domain were obtained from patients and health care workers sampled during 4 prospectively GII.4 outbreaks. Fecal samples were tested by RT-PCR for presence of norovirus RNA against a standard control preparation to allow quantification. Estimated time of onset of shedding was derived from shedding kinetics modeled on data from sequential sampling. Thereby P2 sequence variation could be linked to estimated total virus excretion in individual subjects.

Results: In all the outbreaks, P2 domain variation was found that resulted in unique codon changes in some patients. Mutations were found in 14% of initial samples and >50% of follow-up samples taken from patients involved in an outbreak. In three patients, aa mutations was observed in or near sites involved in host or antigen binding.

Conclusions: We concluded that P2 domain variation increases with duration of virus shedding, but was unrelated to total amounts of virus shed. Therefore, we propose that cluster identification based on identical sequences should be relaxed to accommodate minor sequence variation. When using sequence data to support outbreak investigations, sequence diversity should be interpreted in relation to timing of sampling since onset of illness.

© 2012 Elsevier B.V. Open access under the [Elsevier OA license](#).

1. Background

Noroviruses (NoV) are a major cause of gastroenteritis worldwide and are most commonly associated with outbreaks in health care settings.¹ Onward transmission of noroviruses is common when guidelines for outbreak control are not applied rigorously.² For developing effective control measures, a proper understanding

of the transmission patterns during outbreaks is needed, including the role of healthcare workers and asymptomatic shedders. Molecular typing of NoV-positive stool samples can be used to determine links between individual cases. A systematic analysis of genome diversity in a large dataset collected through the Food-borne viruses of Europe network concluded that the optimal target for sequence-based linking of cases was the capsid gene.³ For practical reasons, currently the P2 domain of the NoV is used.^{4–7} This genome region is considered to be the most variable part of the genome since it codes for the protruding domain of the capsid protein, which contains the receptor binding domain and important epitopes targeted by antibodies that inhibit binding.^{8,9} In GII.4 NoV, the P2 domain evolves by accumulation of mutations under selective pressure from host immunity.^{10–12} Accumulation of mutations in this domain was also shown in immunocompromised patients

Abbreviations: NoV, norovirus; HCW, health care workers; OB, outbreaks; EMC, university hospital.

* Corresponding author at: Center for Infectious Disease and Screening, Department of Virology, P.O. Box 1, 3720 BA, A. van Leeuwenhoeklaan 9, Bilthoven, The Netherlands. Tel.: +30 2742909; fax: +30 2744418.

E-mail address: faizel.sukhrie@rivm.nl (F.H.A. Sukhrie).

with prolonged shedding of viruses.¹³ Recently the use of next generation sequencing identified minority variants present during transmission events.¹⁴

2. Objectives

P2 domain sequencing has been used for identifying the transmission pathways and links during outbreaks^{4–7} and identical P2 domain sequences are considered evidence for a cluster. However, with the high mutation rate of norovirus,¹⁵ nucleotide changes may occur within a short time interval, raising the question what would be an appropriate minimum number of nucleotide changes for defining a cluster of cases connected by direct transmission links. This question is relevant because outbreaks may be missed with common cluster detection algorithms that use time and space, or pseudo-outbreaks may occur when many patients are hospitalized during peak season.^{6,7} Therefore, we set out to quantify P2 domain variation, during four prospectively monitored outbreaks in three nursing homes and a tertiary care hospital.¹⁶ We sampled NoV-positive patients and health care workers (HCW), identified through an enhanced outbreak protocol irrespective of their symptom status. Variation in sequence data between and within outbreaks, as well as variation between and within infected subjects was analyzed and correlated with the estimated number of viruses shed by each individual. The results can be used to redefine criteria for linking of cases to outbreaks.

3. Study design

We prospectively monitored four GII.4 outbreaks starting from January 2009 until March 2011 in the region of Rotterdam in the University Hospital (EMC) and in nursing homes.¹⁶ Sampling was based on an enhanced outbreak investigation protocol focusing on the identification of possible sources and modes of transmission of NoV.¹⁶ The study protocol included random sampling irrespective of symptoms of all patients and HCW on affected wards with NoV. Patients and their contacts involved in the outbreaks who met the inclusion criteria of the study protocol were tested weekly until a negative test was returned. Each case was confirmed by sequencing region A (genotyping) followed by P2 domain sequencing.^{6,17} The amount of virus shed by each subject was estimated from real time RT-PCR analysis of all fecal samples, using RNA transcribed from a plasmid containing a sequence spanning all commonly used diagnostic targets as a reference template,¹⁸ allowing us to investigate correlation between virus excretion and P2 domain changes over time. Background sequences from the same geographic region were collected from patients outside the studied outbreaks, with NoV infection detected ≤ 2 days after admission.

3.1. Sequence analysis

RNA fragments were reverse transcribed with random hexamers (Invitrogen), yielding cDNA that was amplified by a nested PCR and subsequently sequenced using the ABI-PrismBigDye Terminator v3.0 Ready Reaction Cycle kit. The same primers were used for amplification and sequencing the P2 domain (primers 1st PCR: F: 5' gangatgtcttcacagtctctt 3', R: 5' cattctctgggggagtagaca 3',⁴ Nested primers: F: 5' gtcgccaccacagttgag 3', R: 5' gggagtagacagtcctca 3'). DNA Sequences were entered and assembled in bionumerics 6.6.2 (Applied Maths, Sint-Martens-Latem, Belgium) and evaluated manually for their quality by looking for the number of ambiguities, errors, mismatches and deletions.

Sequences were aligned; genotype and variant assignment was based on the RdRp region¹⁶ using the norovirus typing tool (<http://www.rivm.nl/mpf/norovirus/typingtool>).¹⁹ Full-length P2

domain sequences (600 nucleotide) were then subjected to pairwise analysis (UPGMA) to identify strains linked to the same outbreak, and by advanced cluster analysis (maximum parsimony), to compare diversity within and between outbreaks and robustness of clustering. Sequence diversity within patient and between patients within an outbreak was assessed by comparing the minimum and maximum number of mismatches for each outbreak separately. Translated sequences were reviewed to look for possible directional amino acid mutations.

3.2. Sampling and virus shedding

To study the effect of sampling delay, the time of onset of shedding was estimated by extrapolation from modeled shedding kinetics, based on data from all subjects with follow up samples (Teunis et al., submitted for publication). An RNA standard template was used to translate CT values in fecal samples into an estimated viral load. PCR based estimates of NoV shedding were then used to calculate total numbers of viruses shed by sampled subjects, allowing analysis of sequence variation against viral load, clinical symptoms (symptomatic or asymptomatic), and occupational status (HCW/patients). To characterize the rate of sequence variation all data were pooled and the survivor function for sequence change was calculated, using a Kaplan–Meier estimator²⁰ describing the probability of any nucleotide changes versus time from onset of shedding.

4. Results

4.1. Strain typing and clustering

The four GII.4 outbreaks occurred in 3 nursing homes and 1 university hospital. The ages of the included HCW/patients from the hospital setting ranged from 25 to 77 and 54 to 77 years, respectively. In the nursing homes, ages for the residents were high (72–95 years), while for the HCW this ranged from 20 to 63 years. Details of the outbreaks have been described elsewhere (15). In total 175 HCW and 77 patients consented to enhanced case finding, of which 50 HCW and 47 patients tested positive for NoV infection (Table 1). Capsid gene sequencing was successful in 109 NoV positive stool samples from a total of 252 sampled cases, comprising 44 HCW and 37 patients.¹⁶ OB 4 yielded 48 sampled cases but the data is not published yet. Failed sequences were repeatedly tested but persistently failed to produce sequence information. The success of sequencing was unrelated to the levels of virus shedding (data not shown).

Phylogenetic analysis of all P2 domain sequences showed a clear discrimination of the four GII.4 outbreak clusters, but with mixed results for OB 1: here, the outbreak strains segregated into three different clusters: GII.4 2008 (17 cases) and 2 clusters belonging to the GII.4 2006b variant lineage (2 cases each) (Fig. 1). As this suggests that a minority of the patients was from a different, unrelated cluster, detailed molecular analysis was not performed for these strains. Data retrieved from the hospital database showed that one of the samples belongs to a nurse who was involved in patient interviews. The other three subjects were patients who had been admitted into the hospital. From the epidemiological data, it was clear that one patient developed diarrhea after admission, indicating a nosocomial infection.

Comparison of these results against the strain diversity observed in the background dataset (defined as sequences from patients admitted with norovirus infection) showed that these were unique and distinct from the outbreak sequences with few exceptions (4%) (Fig. 1). In OB 4 a unique single case was observed who showed

Table 1
Summary of cases and sequence data.

	Location	NoV cases (HCW/P)	Nr cases with P2 domain sequence (HCW/P)	Nr samples with follow-up P2 domain sequences, HCW/P	Maximum nr nucleotide changes within (patient/OB) ^a
OB 1 (GII.4 2008)	Tertiary care Hosp.	15/7	13/2	3/2	0/22
OB 2 (GII.4 2010)	Nursing home	14/10	15/8	3/13	2/2
OB 3 (GII.4 2010)	Nursing home	12/16	11/15	0/4	1/2
OB 4 (GII.4 2010)	Nursing home	9/14	5/12	0/3	1/15
Total		50/47	44/37	6/22	

^a Including all persons initially considered being part of the outbreak.

at least 15 nucleotide differences compared to the outbreak strain, suggestive for an unrelated introduction.

4.2. Within and between patient sequence diversity

The percentage nucleotide diversity based on P2 domain analysis (600 base pairs) between outbreaks (OBs) including all samples ranged from 6.3 to 7.3% of sequences different at genotype level, and from 0.7 to 1.2% at variant level. Sequence variation within outbreaks was small (0–0.3%) and overlapped with sequence diversity within subjects (0–0.3%). At the time of first sampling 70 (86%) persons of OB 1–4 (both patients and HCW) had an identical sequence, designated the outbreak strain consensus, whereas 6 (38%) were

shedding a virus with a single nucleotide change in the first sample and one person had a sequence with 2 nucleotide changes shortly after onset of shedding. During follow-up, more nucleotide changes were seen; in total 56% of follow-up samples tested yielded one or more mutations.

Fig. 2 shows the timing of sampling in relation to the estimated period of shedding, as estimated from the kinetics of shedding as described elsewhere (Teunis et al., submitted for publication). The sampling delay ranged from 0 to 23 days for symptomatic cases with a median of 8.5 days since the estimated time of infection, and 3–7 days with a median of 5.4 days for asymptomatic cases. When summarizing the rate of sequence change in a Kaplan–Meier plot, sequence changes appeared from 4 days post onset of shedding.

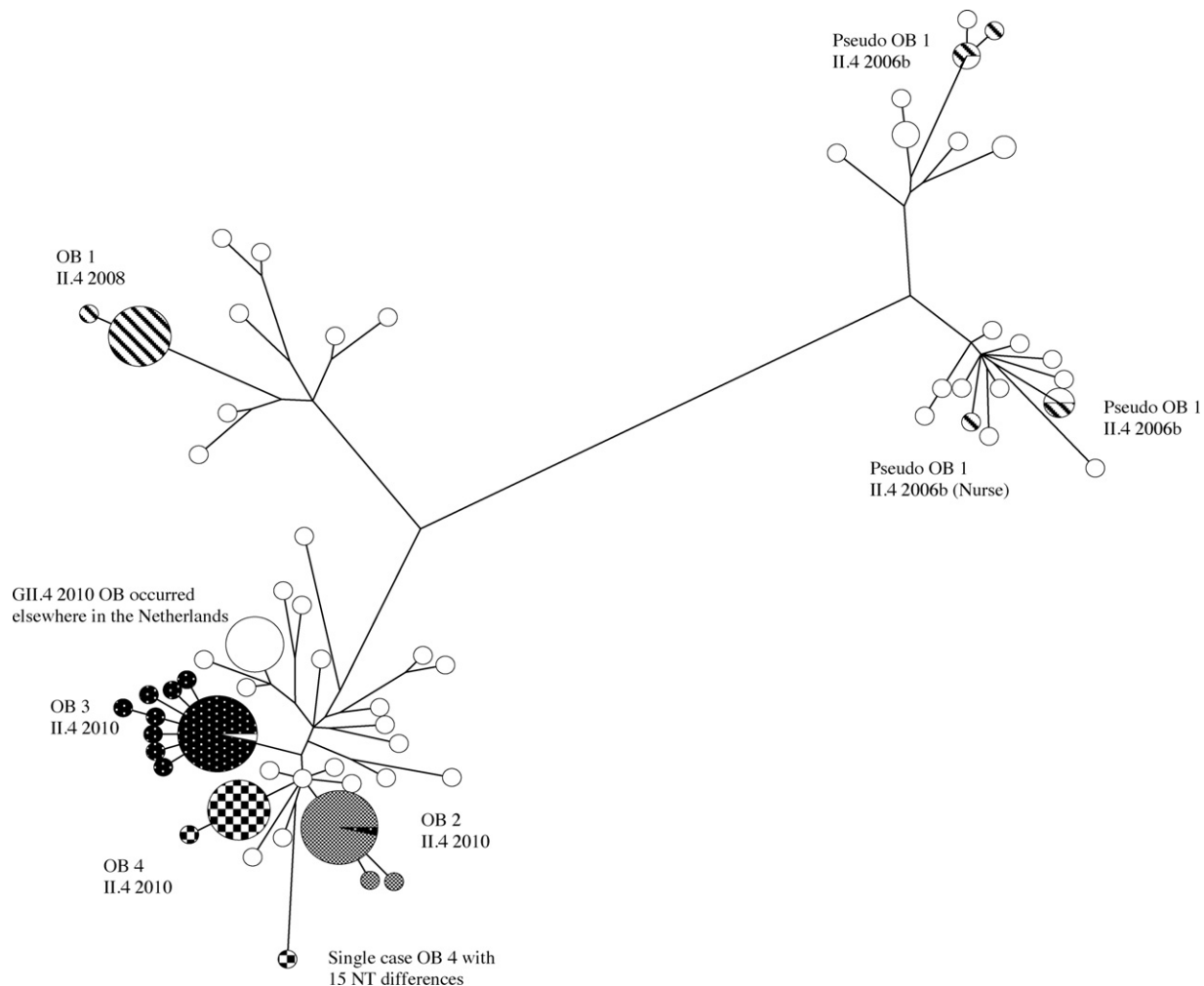


Fig. 1. Cluster analysis (maximum parsimony) showing diversity between and within healthcare outbreaks against background diversity in community cases with NoV (white circles). Size of circle reflects number of sequences.

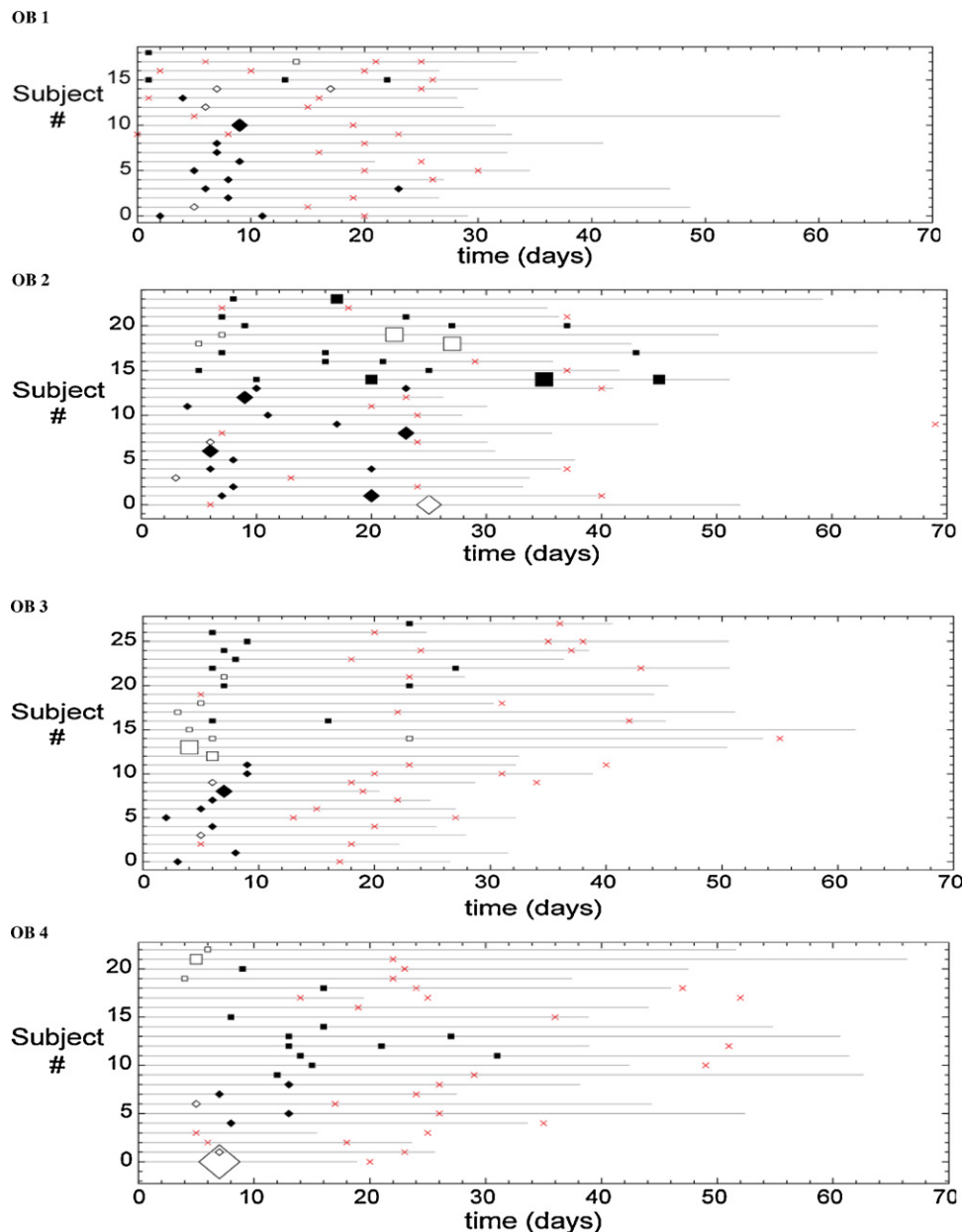


Fig. 2. Time course of virus shedding (modeled) and sequence changes for individual subjects in 4 NoV GI.4 outbreaks. Squares indicate patients, diamonds health care workers; filled symbols indicate symptomatic cases, empty symbols asymptomatic subjects. Size of the symbols indicates sequence change: smallest 0 bp change, medium size 1 bp change, large symbols 2 bp change. One subject in OB 4 (bottom) has a sequence with 15 bp different (extra large square) from the consensus sequence for that outbreak. \times indicates when a sample was taken that was either negative, or for which a sequence could not be established. Horizontal lines (gray) indicate (estimated) duration of shedding, from estimated onset ($t=0$) to the time when the CT increases above 40.

The rate of increase in probability of a sequence change indicates that for samples collected at 3 weeks post onset of shedding there is a 10% probability of (1–2) nucleotide changes (Fig. 3).

When the shedding data were combined with the sequence data no significant association could be seen between shedding and virus excretion or sequence variation, although patients seemed to excrete slightly higher numbers of viruses than HCWs, for a longer period (Fig. 4).

4.3. Codon changes

In total, 11 nucleotide changes resulted in codon changes (6 in HCW, 5 in patients). Codon changes were observed in OB's 2 and 3 only. In these outbreaks, almost all nucleotide sequence changes

(8 out of 10 and 3 out of 3, respectively) were codon changes. Six of the amino acid changes were located at positions in the P2 domain that have been identified as informative sites, because they were one of the marker mutations for global variants of GI.4 (S255G, S310R, T340A, Y352H, S393N, K248R).^{11,12,21,22} Amino acid changes at positions A256T and N373S were seen sporadically in the past, as illustrated in Fig. 5 (Genbank accession numbers). The remaining amino acid changes were unique, located at the following sites; D312N, D312E, R411K. One of the mutations was located in the histo-bloodgroup antigen-binding site (position 393), one in a position that was resulted in an additional RGD motif on the GI.4 2002 strains, and one mutation was near epitope A. No pattern of amino acid changes was observed to verify transmission between subjects (Fig. 5).

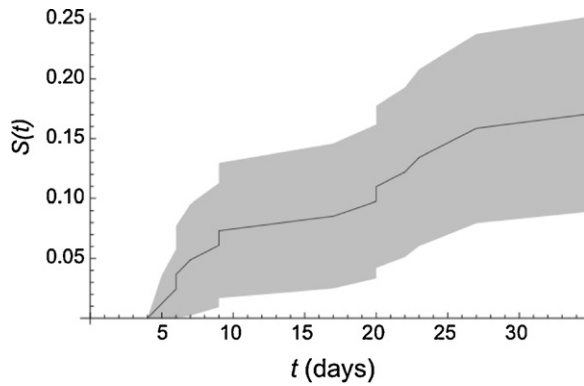


Fig. 3. Sequence changes Y -as ($S/100\%$) over time after onset of shedding X -as (days), shown by nonparametric (Kaplan–Meier) estimate of the probability of sequence change calculated for the complete set of infected subjects from all four outbreaks (mean curve and 95% range).

4.4. Previously reported outbreaks

Background information from the hospital and the nursing homes with OB 1, OB 2, OB 3 and OB 4 indicated that no outbreaks had been reported on the same department shortly before these outbreaks were identified.

5. Discussion

Molecular analysis can help define transmission pathways during outbreaks, particularly when combined with metadata.^{5,6,17} For NoV the proposed molecular marker is the P2 domain, since this domain is considered hypervariable thus providing sufficient resolution for use of sequence data for linking of cases (17). According to our results, each outbreak has a unique consensus sequence based on P2 domain homology. Phylogenetic analysis of P2 domain sequence data can therefore unravel pseudo-outbreaks and specifically can serve to exclude strains that do not belong to the outbreak (Fig. 1). It can also provide information regarding the extent of the outbreak, for instance concurrent sampling from the population extraneous of the outbreak can provide additional linked cases based on the P2 domain homology.

More detailed analysis of the domain revealed that in addition to diversity of the P2 domain between outbreaks there was minor variation within each outbreak, and among follow-up samples from individual patients and HCW. Most of the samples were collected between 4 and 7 days of post onset illness when the transmission was most intense; however, the majority of the NT changes occurred in a later stage (Fig. 2).

In the outbreaks studied here, the variation within the P2 domain does not exceed 2 nucleotide changes and therefore a maximum number of nucleotide changes of 0.33%/600 bp is a conservative threshold to suspect a new introduction. The time interval to the first sequence change can be relevant for future outbreak investigations, with changes observed as early as 4 days following infection.

The position of each nucleotide change was unique and was only found in single cases with one exception. This suggests that minor sequence changes during NoV outbreaks are a random phenomenon in otherwise healthy individuals. However, an indication of immune driven selection is the finding that the majority of mutations in follow-up samples were codon changes, with two remarkable mutations: mutations in position 393 have been associated with alterations in histo-bloodgroup antigen binding patterns of GII.4 strains,²³ and therefore such strains potentially could target a different segment of the population. One mutation was at a position adjacent to epitope A (position 373) and therefore potentially influencing antibody binding. Amino acid changes at position 340 have been found in almost all GII.4 variant transitions, and therefore may be significant as well.^{12,21,22} While speculative, this suggests that new variants may arise quite rapidly in patients during the course of a single infection. Still, an antibody response has to be mounted after an infection since we do not know the exposure history of the individuals involved. Although all of them were adults and therefore most likely had experienced multiple norovirus infections prior to the present one, given the high incidence of these infections in the population.²⁴

Obviously, more information is needed to define transition to a new (epidemic) variant. In our study, we found no evidence of onward transmission of the viruses with potentially informative mutations. Without changes, affecting transmissibility such variants would most likely not emerge as major strains, given the omnipresence of competing strains circulating in the general

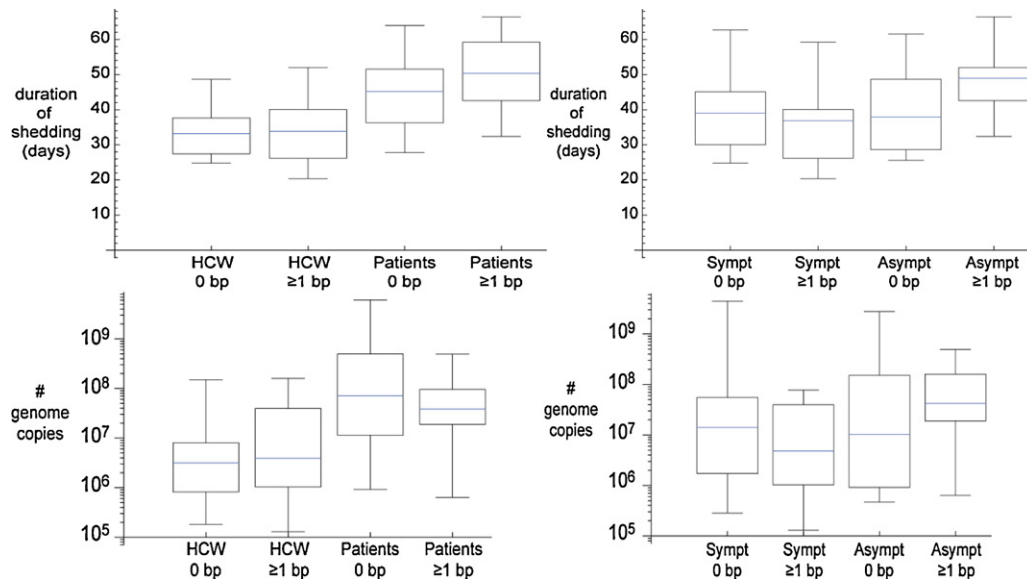


Fig. 4. Box plots (arithmetic mean, quartile box and 95% whiskers) of the duration of shedding for HCW and patients without and with sequence changes, and for symptomatic and asymptomatic cases without and with sequence changes. Also shown (bottom row): same for the estimated total numbers of viruses shed.

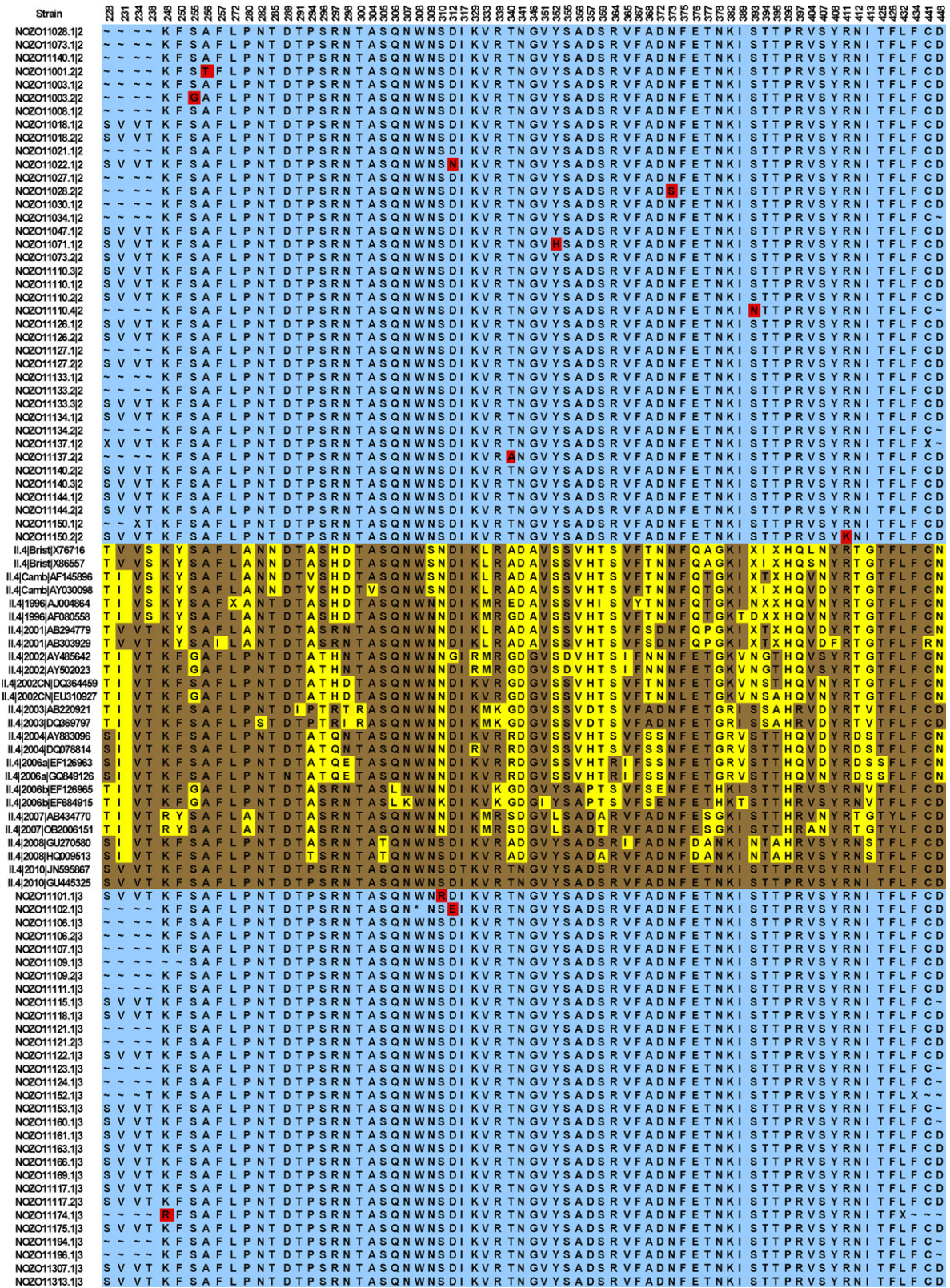


Fig. 5. Amino acid (AA) changes (informative sites) in P2 domain sequences of GGII.4 outbreak strains collected between 2009 and 2010. The informative sites throughout the protein are listed from left to right, AA numbering is indicated at the top, and names are given on the left. From top to bottom, blue color indicates identical amino acids, and overlapping AA of the background (Genbank accession numbers) are illustrated in brown and distinct AA with yellow. The red colors indicate the locations of insertions of the GGII.4 strains during the OB's. Sequences in the middle area (brown) are reference GII4 strains. Symbol (-) indicates failure of partial sequence.

population. This may be different when such infections occur outside the norovirus winter season, and a question is whether chronic norovirus infection in immune-compromised individuals could serve as a reservoir for new variants.¹³

Finally, the sequence variations observed in the present study demonstrate the need to reconsider the guidelines^{4–7} for identifying clusters: the currently used cut-off of 100% identical P2 domain sequence should be relaxed to allow minor variations, thereby potentially increasing the attribution of cases in health care settings. Conversely, since the variation in the P2 domain is limited within the same cluster, it is often difficult to trace transmission events using only sequence data, particularly for defining transmission between individual infected subjects. We have established that the resolution is insufficient to conclusively identify links between individuals within outbreaks (who infected whom): for such purposes, more enhanced sequencing or sequencing of a larger part of the genome is required by considering the presence of minority variants/quasi species.

Funding

This work was supported by the ZonMw, the Netherlands (grant number 125010002).

Competing interests

None declared.

Ethical approval

Ethical approval was obtained.

Acknowledgment

The authors thank all the patients and healthcare workers for their contribution and participation in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jcv.2012.12.006>.

References

- Lopman BA, Gallimore C, Gray JJ, Vipond IB, Andrews N, Sarangi J, et al. Linking healthcare associated norovirus outbreaks: a molecular epidemiologic method for investigating transmission. *BMC Infect Dis* 2006;**6**:108.
- Gould D. Management and prevention of norovirus outbreaks in hospitals. *Nurs Stand* 2008;**23**:51–6, quiz 58–60.
- Verhoef L, Williams KP, Kroneman A, Sobral B, van Pelt W, Koopmans M. Selection of a phylogenetically informative region of the norovirus genome for outbreak linkage. *Virus Genes* 2012;**44**(1):8–18.
- Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ. Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses. *J Clin Microbiol* 2008;**46**:947–53.
- Xerry J, Gallimore CI, Iturriza-Gomara M, Gray JJ. 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* 2009;**81**:1298–304.
- Sukhrie FH, Siebenga JJ, Beersma MF, Koopmans M. Chronic shedders as reservoir for nosocomial transmission of norovirus. *J Clin Microbiol* 2010;**48**:4303–5.
- Sukhrie FH, Beersma MF, Wong A, van der Veer B, Vennema H, Bogerman J, et al. Using molecular epidemiology to trace transmission of nosocomial norovirus infection. *J Clin Microbiol* 2011;**49**:602–6.
- Bu W, Mamedova A, Tan M, Xia M, Jiang X, Hegde RS. Structural basis for the receptor binding specificity of Norwalk virus. *J Virol* 2008;**82**:5340–7.
- Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, et al. Structural basis for the recognition of blood group trisaccharides by norovirus. *J Virol* 2007;**81**:5949–57.
- Allen DJ, Noad R, Samuel D, Gray JJ, Roy P, Iturriza-Gomara M. Characterisation of a GII-4 norovirus variant-specific surface-exposed site involved in antibody binding. *Virol J* 2009;**6**:150.
- Bok K, Abente EJ, Realpe-Quintero M, Mitra T, Sosnovtsev SV, Kapikian AZ, et al. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. *J Virol* 2009;**83**:11890–901.
- Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J Virol* 2007;**81**:9932–41.
- Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. *J Infect Dis* 2008;**198**:994–1001.
- Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of intra- and inter-host dynamics to Norovirus evolution. *J Virol* 2012;**86**(6):3219–29.
- Marshall JA, Bruggink LD. The dynamics of norovirus outbreak epidemics: recent insights. *Int J Environ Res Public Health* 2011;**8**:1141–9.
- Sukhrie FH, Teunis P, Vennema H, Copra C, Thijs Beersma MF, Bogerman J, et al. Nosocomial transmission of norovirus is mainly caused by symptomatic cases. *Clin Infect Dis* 2012;**54**(7):931–7.
- Sukhrie FH, Beersma MF, Wong A, van der Veer B, Vennema H, Bogerman J, et al. Using molecular epidemiology to trace transmission events of nosocomial norovirus infection. *J Clin Microbiol* 2011;**49**(2):602–6.
- Stals A, Baert L, Botteldoorn N, Werbrouck H, Herman L, Uyttendaele M, et al. Multiplex real-time RT-PCR for simultaneous detection of GI/GII noroviruses and murine norovirus 1. *J Virol Methods* 2009;**161**:247–53.
- Kroneman A, Vennema H, Deforche K, v d Avoort H, Penaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 2011;**51**:121–5.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;**53**:457–81.
- Lindsmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog* 2012;**8**:e1002705.
- Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-Gomara M. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. *PLoS ONE* 2008;**3**:e1485.
- Debbink K, Donaldson EF, Lindsmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *J Virol* 2012;**86**:1214–26.
- de Wit MA, Koopmans MP, Kortbeek LM, Wannet WJ, Vinje J, van Leusden F, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *Am J Epidemiol* 2001;**154**:666–74.