PLOS ONE

RESEARCH ARTICLE

Characterization of intra- and inter-host norovirus P2 genetic variability in linked individuals by amplicon sequencing

Aurora Sabrià[1,2], Rosa M. Pintó[1,2], Albert Bosch[1,2], Josep Quer[3,4], Damir Garcia-Cehic[3,4], Josep Gregori[3,4,5], Angela Dominguez[6,7], Mónica Carol[8], Maria-Rosa Sala-Farré[8], Susana Guix[1,2]*, the Working Group for the Study of Outbreaks of Acute Gastroenteritis in Catalonia[¶]

1 Enteric Virus Laboratory, Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona, Spain, **2** Institute of Nutrition and Food Safety (INSA·UB), University of Barcelona, Barcelona, Spain, **3** Liver Unit, Internal Medicine, Laboratori de Malalties Hepàtiques, Vall d'Hebron Institut Recerca-Hospital Universitari Vall d'Hebron (VHIR-HUVH), Autonomous University of Barcelona, Barcelona, Spain, **4** Centro de Investigación Biomédica en Red (CIBER) de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain, **5** Roche Diagnostics S.L., Sant Cugat del Vallès, Spain, **6** CIBER Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain, **7** Department of Public Health, University of Barcelona, Barcelona, Spain, **8** Public Health Agency of Catalonia, Generalitat de Catalunya, Barcelona, Spain

¶ Membership of the Working Group for the Study of Outbreaks of Acute Gastroenteritis in Catalonia is provided in the Acknowledgments.
* susanaguix@ub.edu

Check for updates

OPEN ACCESS

Citation: Sabrià A, Pintó RM, Bosch A, Quer J, Garcia-Cehic D, Gregori J, et al. (2018) Characterization of intra- and inter-host norovirus P2 genetic variability in linked individuals by amplicon sequencing. PLoS ONE 13(8): e0201850. https://doi.org/10.1371/journal.pone.0201850

Editor: Yury E Khudyakov, Centers for Disease Control and Prevention, UNITED STATES

Received: April 14, 2018

Accepted: July 22, 2018

Published: August 9, 2018

Copyright: © 2018 Sabrià et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All sequence data have been submitted to SRA database (accession number SRP150980).

Funding: This work was supported in part by grants from the Fondo de Investigaciones Sanitarias Instituto de Salud Carlos III, Madrid, Spain [PS09/02516 (AD), PI 16/02005 (AD) and PI16/00337 (JQ)]; https://portalfis.isciii.es], partially by the Spanish Ministry of Economy and Competitiveness (MINECO) [BIO2011-23461

Abstract

Noroviruses are the main cause of epidemics of acute gastroenteritis at a global scale. Although chronically infected immunocompromised individuals are regarded as potential reservoirs for the emergence of new viral variants, viral quasispecies distribution and evolution patterns in acute symptomatic and asymptomatic infections has not been extensively studied. Amplicons of 450 nts from the P2 coding capsid domain were studied using next-generation sequencing (454/GS-Junior) platform. Inter-host diversity between symptomatic and asymptomatic acutely infected individuals linked to the same outbreak as well as their viral intra-host diversity over time were characterized. With an average of 2848 reads per sample and a cutoff frequency of 0.1%, minor variant haplotypes were detected in 5 out of 8 specimens. Transmitted variants could not be confirmed in all infected individuals in one outbreak. The observed initial intra-host viral diversity in asymptomatically infected subjects was higher than in symptomatic ones. Viral quasispecies evolution over time within individuals was host-specific, with an average of 2.8 nt changes per day (0.0062 changes per nucleotide per day) in a given symptomatic case. Nucleotide polymorphisms were detected in 28 out of 450 analyzed nucleotide positions, 32.14% of which were synonymous and 67.86% were non-synonymous. Most observed amino acid changes emerged at or near blockade epitopes A, B, D and E. Our results suggest that acutely infected individuals, even in the absence of symptoms, which go underreported and may enhance transmission, may contribute to norovirus genetic variability and evolution.

(RMP); http://www.mineco.gob.es/], by the Centro para el Desarrollo Tecnológico Industrial-CDTI from MINECO [IDI-20151125 (JQ); https://www.cdti.es/], and by the XRB Biotechnology Reference Network (Generalitat de Catalunya) [(AB); http://www.xrb.cat]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare a commercial affiliation of JG with Roche Diagnostics S.L. which only provided support in the form of JG's salary, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: Competing Interests Statement Dr. Josep Gregori is a member of "Roche Diagnostic, S.L.". This does not alter our adherence to PLoS ONE policies on sharing data and materials.

Introduction

Human Norovirus (HuNoV) is the dominant agent involved in acute gastroenteritis causing one fifth of all cases worldwide, and a major cause of foodborne illness [1]. It is highly infectious [2] and is transmitted by the fecal-oral route, either through contact with infected individuals or through exposure to contaminated food and water, although several other modalities have been described such as via aerosolized viral particles in vomitus [3]. It is responsible for 685 million cases every year affecting both, developed and undeveloped countries. In developed countries, it is estimated that every year HuNoVs cost $60 billion due to health care costs like complications and productivity losses due to absenteeism [4].

The viral RNA genome is about 7500 nt long with three open reading frames (ORFs). The ORF1 encodes for a polyprotein that is processed co- and post-translationally by the viral protease to generate six mature active proteins [5]. ORF2 and ORF3 are translated from subgenomic RNA and encode VP1 and VP2, respectively. VP1 is the major capsid protein and is organized into two domains [6], the shell (S) and the protruding (P) domain. P domain is subdivided into two subdomains P1 and P2. P1 subdomain is more conserved than P2 which is the most exposed region of the capsid and contains the Human Histo-Blood Group Antigen (HBGA) binding function and antigenic determinants [7].

Noroviruses are classified into seven genogroups (GI to GVII) and nearly 40 different genotypes, being genogroup I and II the most relevant for human infection [8]. Genogroup II is the most prevalent one, accounting for around 90% of all infections, and GII.4 has been the most commonly isolated genotype in most parts of the world during the last decade [9,10]. New genetic clusters of GII.4 arise every 2 to 4 years under the pressure of herd immunity [11,12], although the reservoir for such variants is still poorly characterized. In individuals chronically infected with HuNoV (mostly immunocompromised patients) new variants may arise due to the accumulation of mutations during the protracted virus replication. In addition, a lower replication fidelity of GII.4 pandemic strains has also been related to an increased intra-host diversity [12,13,14,15,16].

However, despite it is clear that HuNoV intra-host population diversity is higher in persistently infected immunocompromised individuals, HuNoV quasispecies distribution and evolution patterns in both symptomatic and asymptomatic acute infections has not been extensively studied [12]. Genetic variability is one of the main factors contributing to the high incidence of HuNoV in the human population, and a better knowledge of HuNoV evolution driving forces is of clinical and public health interest.

The aim of this study was to analyze the inter-host HuNoV GII.4 diversity between symptomatic and asymptomatic infected individuals linked to the same outbreak, as well as their viral intra-host diversity over time, using amplicon next-generation sequencing (NGS) approaches.

Material and methods

Sample collection

Fecal samples and epidemiological data (outbreak setting, type of transmission, clinical symptoms and time of sample collection) were collected from food-handlers and health-care workers of one foodborne outbreak that took place in a restaurant (RCC11/10), and one person-to-person transmission outbreak that affected a nursing home (UVEVV51/10) (Table 1). Day 1 was defined as the first day of sample collection and not the first day of disease onset. For symptomatic individuals, first samples were taken during the first 3 days after the onset of symptoms. A serial stool sample was collected several days later from two individuals with

Table 1. Selected samples from food-handlers (FH) and health care workers (HCW) involved in outbreaks under study analyzed by NGS.

Outbreak	Setting	Type of transmission	Date of outbreak	Duration of outbreak (days)	Number of affected individuals/ Attack rate (%)	Average age of affected individuals (years)	Subject	Patient description, symptoms, age (years)	Date of specimen collection	Speci-men	Viral load (genome copies/g stool)	Reads (22587 total)	Number of Haplotypes
RCC 11/10	Restaurant	Foodborne (shellfish, mussels)	18/5/10	4	82/50.3	42.2	A	FH, ill, 43	20/5/10	A-1	$8{,}09 \times 10^{9}$	2700	1
									28/5/10	A-9	$1{,}33 \times 10^{9}$	1795	1
							B	FH, not ill, 36	20/5/10	B-1	$7{,}11 \times 10^{9}$	5038	7
UVEVV 51/10	Nursing home	Person-to-person	10/11/10	12	21/21	75.1	C	HCW, ill, 68	17/11/10	C-1	$2{,}07 \times 10^{8}$	1275	2
							D	HCW, ill, 50	17/11/10	D-1	$1{,}32 \times 10^{11}$	4424	3
									22/11/10	D-6	$3{,}37 \times 10^{10}$	3653	18
							E	FH, not ill, 47	17/11/10	E-1	$6{,}98 \times 10^{7}$	2249	1
							F	FH+HCW, not ill	26/11/10	F-1	$1{,}62 \times 10^{6}$	1453	6

https://doi.org/10.1371/journal.pone.0201850.t001

symptomatic infection. Genotype and virus load for all samples were determined in two previous studies [17,18]. All studied samples (n = 8) were GII.4 New Orleans 2009. The study was approved by the Institutional Review Board (IRB00003099) of the Ethics Committee of the University of Barcelona. Written consent was not required because data were anonymized prior to analysis and gathered for outbreak control and public health purposes.

RT-PCR

Viral RNA was extracted from 150 μl of 10% stool suspension in phosphate-buffered saline (pH 7.4) and was stored at −80˚C. A two-step reverse transcription-PCR (RT-PCR) was employed to amplify a 450-bp region of the HuNoV genome encoding for P2. Expand Reverse Transcriptase (Roche Applied Science, Basel, Switzerland) was used for the RT and high-fidelity *Pwo* DNA Polymerase (Roche Applied Science, Basel, Switzerland) for the PCR. Primers are described in Table 2 and include a template-specific sequence and a universal M13 sequence required in the multiplex Ultra-Deep Pyro Sequencing (UDPS) reaction [19]. RT-PCR products were purified from agarose gel bands using the NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel, Düren, Germany) and quantified using a NanoDrop1000 spectrophotometer (Thermo Scientific, Wilmington, DE).

Ultra-Deep Pyro Sequencing (UDPS)

UDPS was performed using a 454/GS-Junior deep sequencing platform (Roche, Branford, CT, USA). Briefly, after purified RT-PCR products were pooled equimolarly at 10^9 molecules/ul, they were subjected to 15 cycles of re-amplification for the addition of the MIDs, multiplex identifier oligomers of 10bp. Amplification products were purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA, USA), evaluated using the BioAnalyzer DNA 1000 LabChip (Agilent, Santa Clara, CA, USA), and quantified by PicoGreen assay (Invitrogen, Carlsbad, CA, USA). Massive parallel sequencing was performed in a 454/GS-Junior platform (Roche, Branford, CT, USA), using titanium chemistry (GS Junior+ Series, XL+ Kit), which enables sequencing up to 700-nt fragments.

Computations were made as previously described [20,21,22]. Briefly, the fasta file from the GS-Junior was demultiplexed to obtain a fasta file for each sample and strand. Reads not identified by MID and/or primer were discarded. The allowances were two mismatches in the specific sequence, three on the universal M13 sequence and one on the MID sequence with no indels. Sequences not covering the full amplicon or showing more than 2 Ns or 3 gaps were discarded. Sequences not observed on the forward and reverse strands at a minimum abundance of 0.1% were discarded. Different haplotypes, defined as reads covering the full amplicon, were identified and their frequencies computed as the number of observed reads. All sequence data are available from the SRA database (accession number SRP150980).

Table 2. Primers used for the study.

Primer	Sequence (5'-3')	Sense	Position[a]	Genome Region
13uNoV5956	GTTGTAAAACGACGGCCAGTGTCAACATCTGCACCTTCAGA[b]	Forward	5925–5945	P2
13dNoV6447	CACAGGAAACAGCTATGACCTACCCRCTGCATCCGGGCAT[b]	Reverse	6396–6415	P1

[a] GenBank accession number GU445325

[b] Underlined nt correspond to M13 forward and reverse primers, respectively

https://doi.org/10.1371/journal.pone.0201850.t002

Phylogenetic analysis

Phylogenetic analysis was performed using the UPGMA method (distance calculation by number of differences method, pairwise deletion) implemented in the MEGA7 program [23], and results were validated by 1000 bootstrap replicates.

Results

Relation between viral titer and number of reads

NGS was performed on 450-bp P2 amplicons on 8 selected samples. An average of 2848 reads per sample were obtained (range 1275–5038). Whole genome sequence analyses of other RNA viruses have demonstrated that the depth of coverage improves by increasing viral load [19]. As shown in Fig 1, we observed a moderate correlation ($R^2 = 0.552$; $p = 0.0347$) between viral load and number of reads. In general, samples with highest viral load exhibited higher number of reads, while samples with lower titer yielded less reads. No correlation was found between viral load and number of haplotypes ($R^2 = 0.070$), neither between number of reads and number of haplotypes ($R^2 = 0.143$).

Inter- and intra-host diversity between outbreak-linked symptomatic and asymptomatic cases

Samples from two subjects where analyzed in RCC11/10 foodborne outbreak (subjects A and B). The single haplotype found in subject A was also the most common haplotype occurring in subject B (97.98%), who was infected from the same source but who did not show symptoms (Fig 2A). Of the other 6 minor haplotypes observed in subject B, 4 of them presented an amino acid change at evolving sites: P313L, I330T, T394I and P427L (Table 3 and Fig 3A).

Five samples from four patients were analyzed from the UVEVV51/10 outbreak. The data revealed marked differences between symptomatic and asymptomatic subjects. While symptomatic subjects C and D shared the most abundant haplotype, neither this haplotype nor any of the minor variants detected in these individuals were found in asymptomatic subjects E and F (Figs 2B and 3B). While all 5 haplotypes identified in the first specimen for subjects C and D

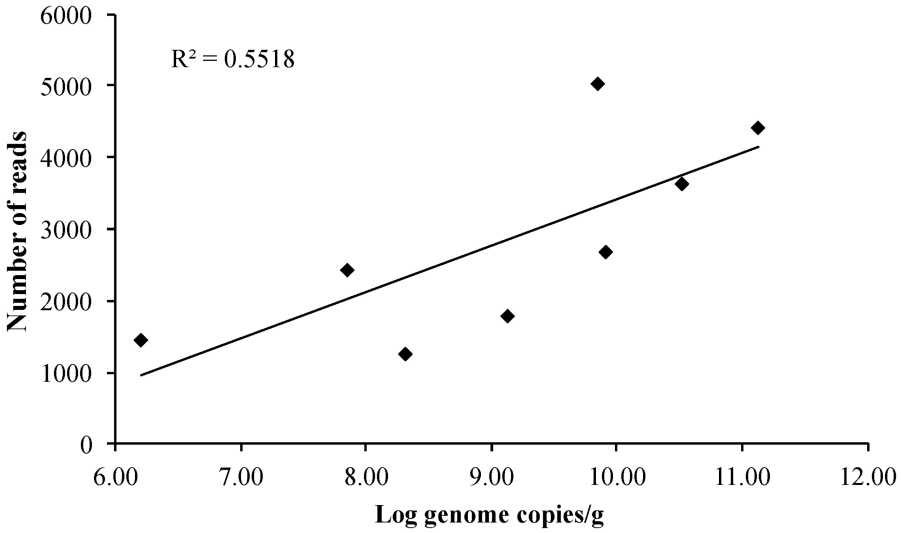

Fig 1. Correlation between viral titer in stool (log genome copies/g) and number of reads.

https://doi.org/10.1371/journal.pone.0201850.g001

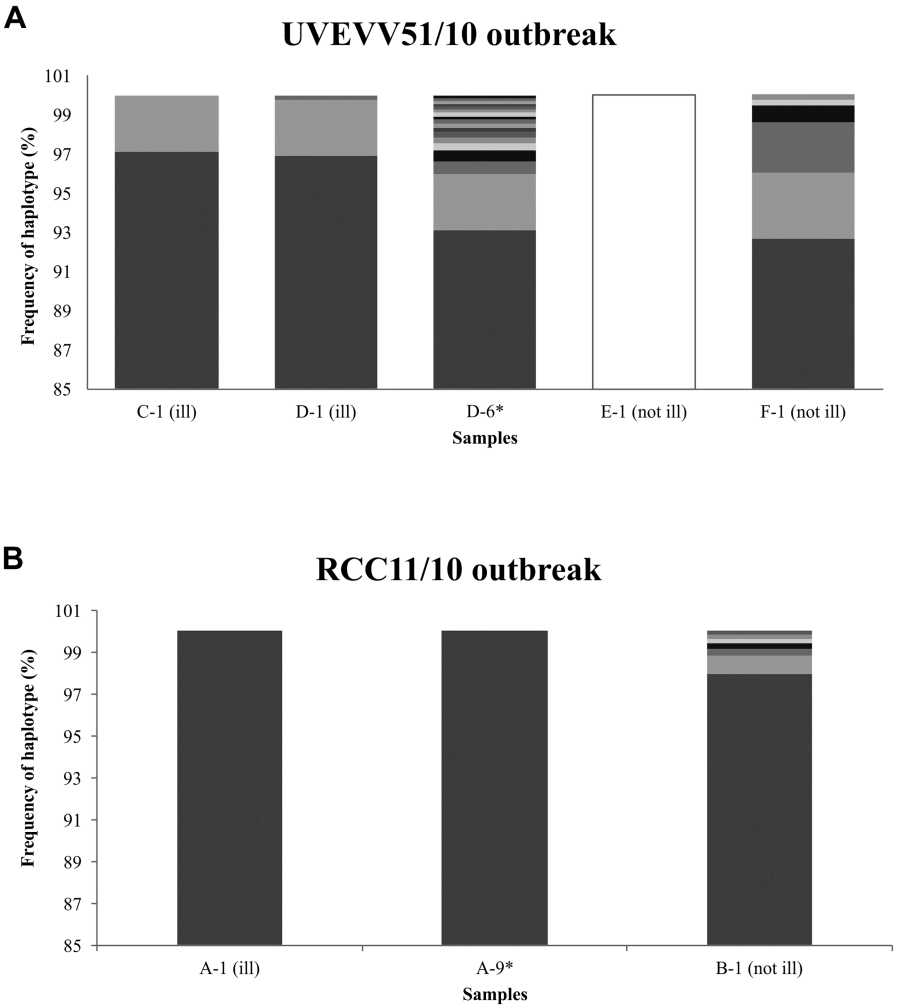

Fig 2. Comparison of the intra-host distribution of HuNoV haplotypes. Each unique haplotype is represented by alternate gray shading. **(A)** Subjects from RCC11/10 outbreak. **(B)** Subjects from UVEVV51/10 outbreak. Asterisk indicate serial sample.

https://doi.org/10.1371/journal.pone.0201850.g002

only differed in synonymous changes, the single haplotype found in subject E, and the most abundant haplotype found in subject F, both asymptomatically-infected, differed in 2–3 amino acid positions (Table 4). Genetic diversity was especially notable in asymptomatic subject F, with 4 out of the 6 identified haplotypes containing several amino acid changes (Table 4). These results suggest either multiple sources of infection or a high degree of virus evolution within the selected infected individuals. Compared to GII.4 New Orleans reference strain, all haplotypes identified in subjects C and D differed in 2 common amino acids (P294S and I413T), and all but one also differed at an additional site (R339K) (Table 4). Haplotypes found in subjects E and F were more closely related to GII.4 New Orleans reference strain (Fig 3B), and only the haplotype identified in subject E showed P294S and R339K substitutions. All observed non-synonymous changes had been previously described [14,24,25].

Overall, the intra-host viral diversity in asymptomatically infected subjects (B, E and F) at day 1 was higher than in symptomatic patients (A, C and D). Although differences were not statistically significant, the initial average of haplotypes per subject was of 4.7 in asymptomatic subjects versus 2.0 in symptomatic ones.

Table 3. Amino acid differences between the haplotypes detected in RCC11/10 outbreak with reference to one of the samples within the outbreak. GII.4 reference strains are also provided at the top of the table.

	VP1 Amino acid Position					
	P2					P1
	Epitope A			Epitope D	Epitope E	
	294	313	330	394	413	427
2009 New Orleans (GU445325)	P	P	I	T	I	P
2012 Sydney (JX459908)	T	P	I	T	T	P
A-1 Hpl.1_100%	P	P	I	T	I	P
A-9 Hpl.1_100%						
B-1 Hpl.1_98.0%						
B-1 Hpl.2_0.8%						
B-1 Hpl.3_0.3%						L
B-1 Hpl.4_0.2%						
B-1 Hpl.5_0.2%		L				
B-1 Hpl.6_0.2%			T			
B-1 Hpl.7_0.2%				I		

https://doi.org/10.1371/journal.pone.0201850.t003

Intra-host evolution diversity

While the intra-host viral diversity initially observed in asymptomatically infected subjects (B, E and F) was higher than in symptomatic patients (A, C and D), diversity evolution over time for the 2 individuals with a serial sample (A and D) showed markedly different trends. While in subject A, who was 43 years old, the single initial haplotype remained stable over the 8 days of infection, in subject D, a 50 year-old woman, HuNoV population diversity increased from 3 to 18 haplotypes in only 5 days (Table 1). Of note, this subject was also the one with the highest viral load. The phylogenetic analysis showed that of the 3 initial haplotypes in D-1, the most abundant haplotype slightly decreased from 96.9% to 93.1%, the second haplotype remained at a stable frequency of 2.9% and the third haplotype slightly increased from 0.3% to 0.6% (Fig 3B). During the following 5 days, variation occurred in 14/450 nt positions, which represents an average of 2.8 nt changes per day and 0.0062 changes per nucleotide per day. At the amino acid level, all 3 initial haplotypes detected in subject D were synonymous, and 11 out of the 18 evolved haplotypes contained amino acid changes (61.1%) at previously identified informative sites: D289G, I293T, T314A, F358L, N380D, E399G, S407I, V416A, F433L, and F434L (Table 4).

Quasispecies analysis

Among all sequence variants isolated from the quasispecies from the 4 subjects that showed variation in the studied region (B, C, D, and F), 29 out of 450 analyzed nucleotide positions were variable (6.4%). Of these single nucleotide polymorphisms (SNPs), 34.5% were synonymous and 65.5% were non-synonymous.

Of note, 96.5% of these substitutions were transitions (58.1% of the transitions were substitutions U-to-C, 19.4% were A-to-G), and only 3.5% were transversions (G-to-U) (Table 5). Since HuNoV mutants with a high proportion of U-to-C and A-to-G transitions have been described in clinical samples and double-stranded RNA-specific adenosine deaminases (ADAR)-mediated editing of the viral RNA was suggested as the molecular mechanism responsible for them [26], we analyzed their sequence context. Although the chi-square test did no show significant differences compared to the percentages found for conserved sites, of

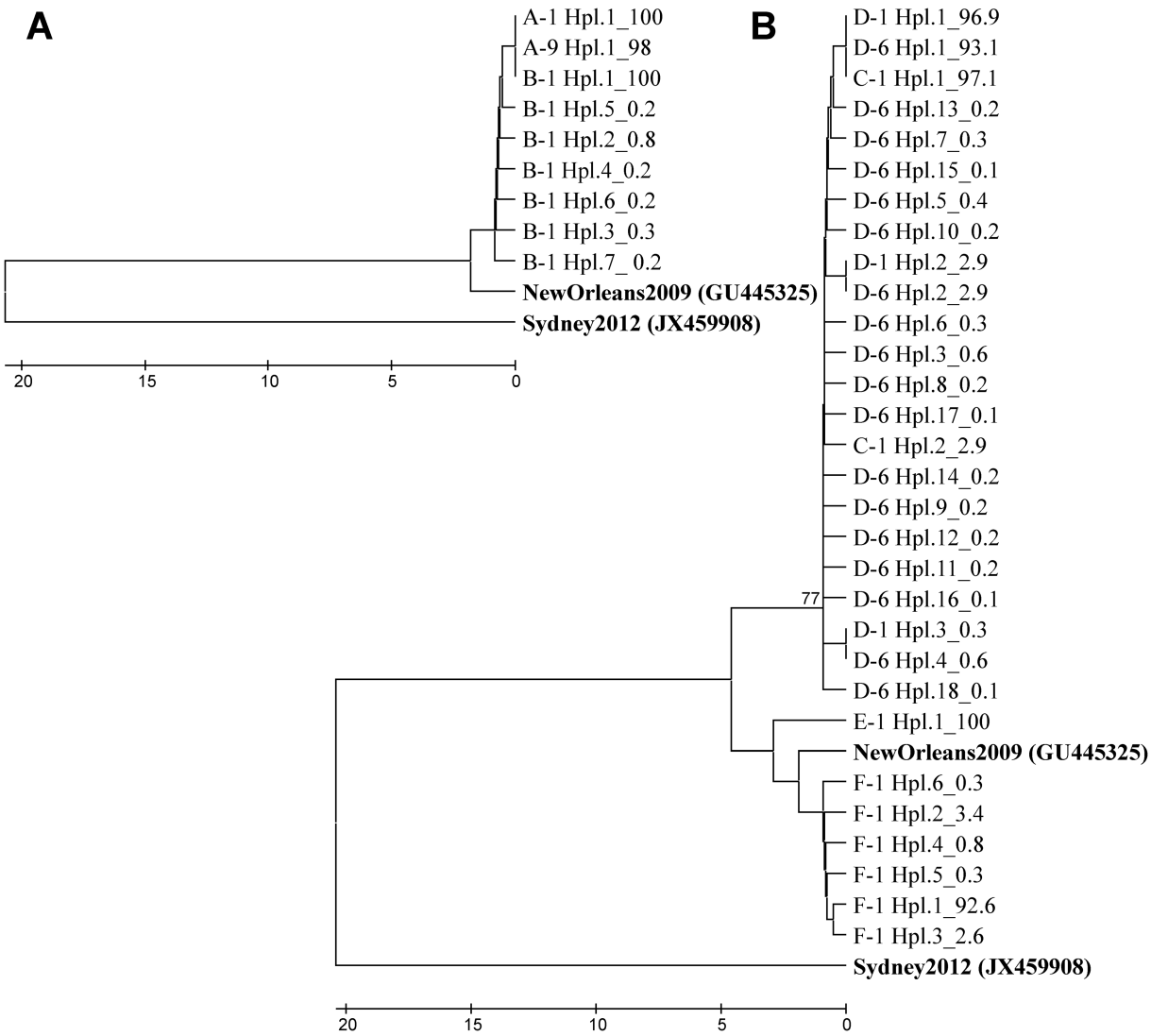

Fig 3. Phylogenetic analysis of haplotypes from RCC11/10 (A) and UVEVV51/10 (B) outbreak based on 450 bp of the P2 region. The nucleotide dendrogram was inferred using the UPGMA method with distance calculation by number of differences method using the MEGA7 software. A bootstrap of 1000 replicates was performed, and values above 75 are shown. Bold indicates reference strains. Haplotypes names are composed of sample name, number of haplotype.

https://doi.org/10.1371/journal.pone.0201850.g003

all U-to-C transitions, the 3' neighbor nucleotide was U in 40% of cases, A in 26.7%, G in 20% and C in 13.3%, which tend to the sequence context compatible with ADAR-mediated editing of the viral RNA.

At the amino acid level, the majority of the observed changes emerged at or near epitopes A, D and E, while some others were located at internal regions of the P dimer (Tables 3 and 4, and Fig 4) [14,24,25,27]. No changes were observed in HBGA binding pocket sites. Identified substitutions matched previously identified changes between different GII.4 variants (VP1 residues 289, 293, 294, 313, 314, 330, 339, 356, 358, 380, 394, 399, 404, 407, 413, 416, 426, 427, 433, 434, and 436). In two different individuals, mutations were detected at specific amino acids directly involved in epitope D (T394I) and epitope E (S407I) among the minor variants, suggesting that they may affect antibody recognition and HBGA binding.

Table 4. Amino acid differences between the haplotypes detected in UVEVV51/10 outbreak with reference to one of the samples within the outbreak. GII.4 reference strains are also provided at the top of the table.

	VP1 Amino acid Position																
	P2												P1				
			Ep. A								Ep. E						
	289	293	294	314	339	356	358	380	399	404	407	413	416	426	433	434	436
2009 New Orleans (GU445325)	D	I	P	T	R	A	F	N	E	V	S	I	V	F	F	F	S
2012 Sydney (JX459908)	D	I	T	T	R	A	F	N	E	V	S	T	V	F	F	F	S
C-1 Hpl.1_97.1%	D	I	S	T	K	A	F	N	E	V	S	T	V	F	F	F	S
C-1 Hpl.2_2.9%																	
D-1 Hpl.1_96.9%																	
D-1 Hpl.2_2.9%																	
D-1 Hpl.3_0.3%																	
D-6 Hpl.1_93.1%																	
D-6 Hpl.2_2.9%																	
D-6 Hpl.3_0.6%							L										
D-6 Hpl.4_0.6%																	
D-6 Hpl.5_0.4%				A													
D-6 Hpl.6_0.3%																	
D-6 Hpl.7_0.3%																	
D-6 Hpl.8_0.2%		T															
D-6 Hpl.9_0.2%								D									
D-6 Hpl.10_0.2%																	
D-6 Hpl.11_0.2%															L		
D-6 Hpl.12_0.2%																L	
D-6 Hpl.13_0.2%													A				
D-6 Hpl.14_0.2%									G								
D-6 Hpl.15_0.1%																	
D-6 Hpl.16_0.1%											I						
D-6 Hpl.17_0.1%					R												
D-6 Hpl.18_0.1%	G																
E-1 Hpl.1_100%												I					
F-1 Hpl.1_92.6%			P		R							I					
F-1 Hpl.2_3.4%			P		R							I					
F-1_Hpl.3_2.6%			P		R	V						I					
F-1_Hpl.4_0.8%			P		R							I		S			
F-1_Hpl.5_0.3%			P		R					M		I					
F-1_Hpl.6_0.3%			P		R							I					P

https://doi.org/10.1371/journal.pone.0201850.t004

Discussion

In this work we used amplicon-NGS analysis to investigate the viral population of selected individuals linked to the same outbreak, as well as to analyze the intra-host HuNoV evolution dynamics over time. Since all investigated subjects suffered an acute infection, viral populations observed were highly homogeneous as expected. However, studied individuals who were asymptomatically infected had a higher haplotype diversity, suggesting that they may also act

Table 5. Characterization of mutant spectrum found in fecal specimens.

Sample	Mutations[a]/ Nucleotides Sequenced						Nucleotide Mutation Frequency[c]		Amino Acid Mutation Frequency[c]		S_N[d]
		Ts[b]	Tv[b]	Nsyn[b]	Syn[b]	Indel[b] or Stop	Minimum	Maximum	Minimum	Maximum	
B-1	102/2267100	6	0	4	2	0	2.65×10^{-6}	4.50×10^{-5}	5.29×10^{-5}	6.22×10^{-5}	0.02
C-1	37/573750	1	0	0	1	0	$1{,}74\times10^{-6}$	$6{,}45\times10^{-5}$	0	0	0.02
D-1	139/1990800	2	0	0	2	0	$1{,}00\times10^{-6}$	$6{,}98\times10^{-5}$	0	0	0.02
D-6	252/1643850	16	1	11	6	0	$1{,}03\times10^{-5}$	$1{,}53\times10^{-4}$	2.01×10^{-5}	1.70×10^{-4}	0.05
F-1	107/653850	5	0	1	4	0	$7{,}65\times10^{-6}$	$1{,}64\times10^{-4}$	1.84×10^{-5}	2.66×10^{-4}	0.05

[a] Mutations are those that vary relative to the corresponding consensus sequence. Nucleotides sequenced are the total number of nucleotides sequenced.

[b] The mutations are classified into: transitions (Ts), transversions (Tv), nonsynonymous (Nsyn), synonymous (Syn) and insertion/deletion (Indel or Stop).

[c] The minimum nucleotide mutation frequency is the number of different mutations found divided by the total number of nucleotides sequenced. The maximum nucleotide mutation frequency is the total number of mutations found divided by the total number of nucleotides sequenced. The maximum amino acid mutation frequency is the total number of nonsynonymous mutations divided by the number of amino acids encoded in the sequence analyzed. Mutation frequencies are expressed as substitutions per nucleotide or amino acid substitutions per amino acid.

[d] The Shannon entropy was calculated as $S_N = [\Sigma_i (p_i \times \ln p_i)]/\ln N$, where p_i is the frequency of each sequence and N is the total number of sequences.

https://doi.org/10.1371/journal.pone.0201850.t005

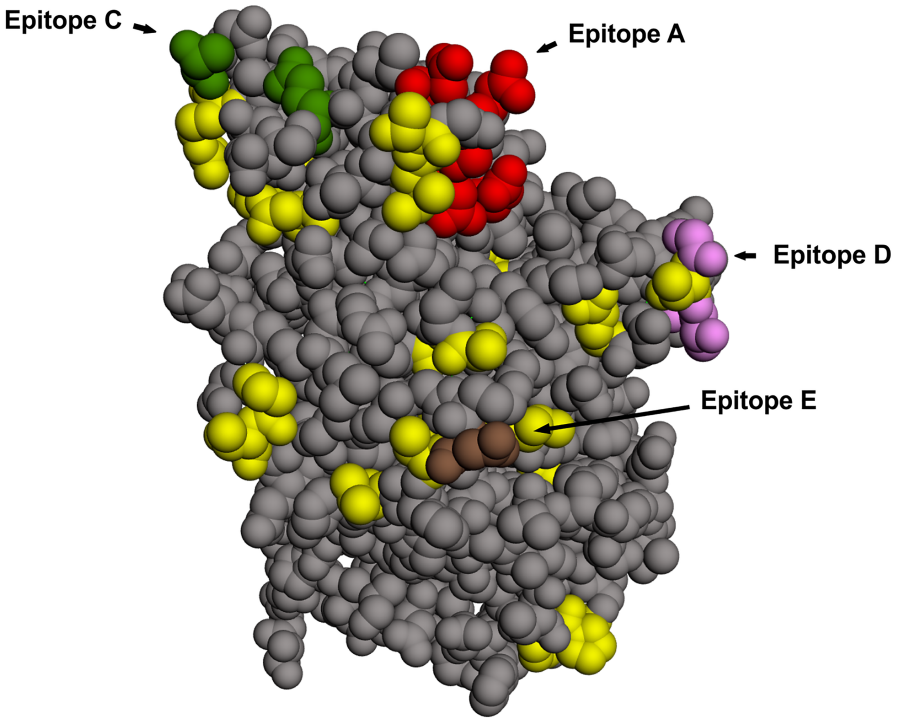

Fig 4. Amino acid variations in P domain. Amino acid changes observed in minor variants (yellow) were highlighted on the GII.4 P domain (side view), using the GII.4 2012 P2 domain structure (PDB id 4OOS; available at https://www.rcsb.org/structure/4OOS). The reference structure shows epitopes A (red), C (green), D (pink) and E (brown). The Jmol software was used to localize the amino acid residues (Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/).

https://doi.org/10.1371/journal.pone.0201850.g004

as a reservoir of novel genetic variants. Asymptomatically infected individuals are usually underreported and may transmit the virus in the absence of maximized hygiene habits.

Of note, the pattern of evolution of HuNoV variability over time differed in the 2 analyzed symptomatic subjects. While the single initial haplotype was maintained in one of them (subject A) during at least 8 days, diversity in subject D, which also was initially higher, increased from 3 to 18 haplotypes in only 5 days, revealing an average of 0.0062 nucleotide changes per nucleotide per day. Although still remarkable, this variation rate is significantly lower than what was described for immunocompromised patients chronically infected with HuNoV, which ranged between 0.03 to 0.37 nucleotide changes per day considering the full genome consensus sequence variation over time [16]. In addition to unknown individual factors, such as genetic differences or the history of previous infections by different strains of each subject, the degree of diversity in the transmitted/founding population that initiated the infection in both individuals may have also been different. Although our analysis did not indicate a correlation between haplotype diversity and viral load within a given sample, the initial viral load may be important in determining the evolution pattern. Individual D, who had the highest initial viral load at day 1 showed the highest diversity over time.

The effect of other factors related to the host on the degree of intra-host haplotype diversity could be especially observed in foodborne outbreak RCC11/10, which had been caused by a common source (contaminated shellfish). In this outbreak, a greater level of variability was found in the subject who did not develop symptoms. The relationship between the immune response raised to clear infection and the occurrence of symptoms is still poorly understood, and although a larger number of asymptomatic infected individuals should be tested to confirm this idea, our observations suggest that people infected by HuNoV who do not suffer symptoms may act as a reservoir for viral antigenic variants. While a properly functioning immune response may induce a higher fixation rate of mutations [28], it would be interesting to study whether symptomatology and/or virus severity may also affect it.

The NGS analysis of samples belonging to UVEVV51/10 outbreak occurring at a nursing home allowed us to confirm earlier results reported by Sanger sequencing [18]. While subjects C and D shared some of the haplotypes, sequences detected in subjects E and F were markedly different, suggesting more than one introduction of HuNoV in the nursing home. However, since transmission has been shown to be an important source of diversity and minor variants at frequencies as low as $<0.01\%$ have been shown to be efficiently transmitted from a donor to a recipient [12], viral transmission between subjects C or D and E and F cannot be completely ruled out. Quality filters performed during our NGS analysis excluded haplotypes at frequencies lower than 0.1%. Indeed, the type of transmission event (foodborne, fecal-oral, vomit-oral, etc) may also affect viral population diversity. Workers at nursing homes may be frequently exposed to episodes of vomiting and diarrhea and minor variants may disseminate more easily [29]. Indeed, the total number of different haplotypes found in the nursing home outbreak was higher than in the foodborne outbreak.

It is described that HuNoV GII.4 is able to generate new antigenic variants that have the potential to escape herd immunity by accumulation of mutations at major blockade epitopes [30,31,32,33]. In our study, numerous SNPs were found in sequence variants from both outbreaks, and 21 of them produced amino acid changes at sites previously determined to be functionally significant. Some of these changes occurred at or near described epitopes A, B, D, and E, while none of them affected any of the described receptor binding pocket sites.

In summary, NGS analysis is useful to characterize the quasispecies diversity and evolution over time, and may be a powerful tool to understand and confirm viral transmission between different hosts. Although evolution of viral population within an infected host may be subject-specific, it may also occur in asymptomatically infected individuals.

Acknowledgments

We are indebted to the Working Group for the Study of Outbreaks of Acute Gastroenteritis in Catalonia, and we are grateful for the collaboration of the physicians reporting the outbreaks and the technicians of the Epidemiological Surveillance Units of the Department of Health of the Generalitat of Catalonia and the Public Health Agency of Barcelona. The full membership list of the Working Group for the Study of Outbreaks of Acute Gastroenteritis in Catalonia are: M. Alsedà, J. Álvarez, C. Arias, A. Artigues, I. Barrabeig, P. J Balanyà, R. Bartolomé, A. Bosch, A. Domínguez, N. Camps, M. Carol, M. Company, T. Cornejo, G. Ferrús, N. Follia, P. Godoy, S. Guix, A. Martínez, S.Minguell, I. Parrón, R.M. Pintó, A. Recasens, A. Rovira, M. R. Sala-Farré, M. de Simon, R. Torra, N. Torner, J. Torres (Public Health Agency of Catalonia, Barcelona, Spain), D. Ferrer, A. Moreno, M. Sanz (Public Health Agency of Barcelona, Barcelona, Spain), and S. Broner (CIBER Epidemiology and Public Health, Instituto de Salud Carlos III, Madrid, Spain).

Author Contributions

Conceptualization: Aurora Sabrià, Rosa M. Pintó, Albert Bosch, Josep Quer, Josep Gregori, Susana Guix.

Data curation: Josep Gregori.

Formal analysis: Josep Gregori, Susana Guix.

Funding acquisition: Rosa M. Pintó, Albert Bosch, Josep Quer, Angela Dominguez, Susana Guix.

Investigation: Aurora Sabrià, Damir Garcia-Cehic, Josep Gregori, Susana Guix.

Methodology: Damir Garcia-Cehic, Mónica Carol, Maria-Rosa Sala-Farré, Susana Guix.

Resources: Angela Dominguez.

Software: Josep Gregori.

Supervision: Rosa M. Pintó, Albert Bosch, Josep Quer.

Writing – original draft: Aurora Sabrià, Susana Guix.

Writing – review & editing: Rosa M. Pintó, Albert Bosch, Josep Quer, Angela Dominguez.

References

1. Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, Parashar UD, et al. (2014) Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. The Lancet Infectious Diseases 14: 725–730. https://doi.org/10.1016/S1473-3099(14)70767-4 PMID: 24981041
2. Teunis PF, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, et al. (2008) Norwalk virus: how infectious is it? J Med Virol 80: 1468–1476. https://doi.org/10.1002/jmv.21237 PMID: 18551613
3. Robilotti E, Deresinski S, Pinsky BA (2015) Norovirus. Clin Microbiol Rev 28: 134–164. https://doi.org/10.1128/CMR.00075-14 PMID: 25567225
4. Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY (2016) Global Economic Burden of Norovirus Gastroenteritis. PLOS ONE 11: e0151219. https://doi.org/10.1371/journal.pone.0151219 PMID: 27115736
5. Thorne LG, Goodfellow IG (2013) Norovirus gene expression and replication. J Gen Virol.
6. Prasad BV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK (1999) X-ray crystallographic structure of the Norwalk virus capsid. Science 286: 287–290. PMID: 10514371
7. Debbink K, Donaldson EF, Lindesmith LC, Baric RS (2012) Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. J Virol 86: 1214–1226. https://doi.org/10.1128/JVI.06189-11 PMID: 22090110

8. Vinje J (2015) Advances in laboratory methods for detection and typing of norovirus. J Clin Microbiol 53: 373–381. https://doi.org/10.1128/JCM.01535-14 PMID: 24989606

9. Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinje J (2014) Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. J Clin Microbiol 52: 147–155. https://doi.org/10.1128/JCM.02680-13 PMID: 24172151

10. van Beek J, de Graaf M, Al-Hello H, Allen DJ, Ambert-Balay K, Botteldoorn N, et al. (2018) Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network. The Lancet Infectious Diseases.

11. Siebenga JJ, Vennema H, Zheng DP, Vinje J, Lee BE, Pang XL, 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. https://doi.org/10.1086/605127 PMID: 19627248

12. Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA (2012) Contribution of intra- and interhost dynamics to norovirus evolution. J Virol 86: 3219–3229. https://doi.org/10.1128/JVI.06712-11 PMID: 22205753

13. Bull RA, White PA (2011) Mechanisms of GII.4 norovirus evolution. Trends Microbiol 19: 233–240. https://doi.org/10.1016/j.tim.2011.01.002 PMID: 21310617

14. Debbink K, Lindesmith LC, Ferris MT, Swanstrom J, Beltramello M, Corti D, et al. (2014) Within-host evolution results in antigenically distinct GII.4 noroviruses. J Virol 88: 7244–7255. https://doi.org/10.1128/JVI.00203-14 PMID: 24648459

15. Hoffmann D, Hutzenthaler M, Seebach J, Panning M, Umgelter A, Menzel H, et al. (2012) Norovirus GII.4 and GII.7 capsid sequences undergo positive selection in chronically infected patients. Infect Genet Evol 12: 461–466. https://doi.org/10.1016/j.meegid.2012.01.020 PMID: 22310302

16. van Beek J, de Graaf M, Smits S, Schapendonk CME, Verjans G, Vennema H, et al. (2017) Whole-Genome Next-Generation Sequencing to Study Within-Host Evolution of Norovirus (NoV) Among Immunocompromised Patients With Chronic NoV Infection. J Infect Dis 216: 1513–1524. https://doi.org/10.1093/infdis/jix520 PMID: 29029115

17. Sabria A, Pinto RM, Bosch A, Bartolome R, Cornejo T, Torner N, et al. (2014) Molecular and clinical epidemiology of norovirus outbreaks in Spain during the emergence of GII.4 2012 variant. J Clin Virol 60: 96–104. https://doi.org/10.1016/j.jcv.2014.03.013 PMID: 24746342

18. Sabria A, Pinto RM, Bosch A, Bartolome R, Cornejo T, Torner N, et al. (2016) Norovirus shedding among food and healthcare workers exposed to the virus in outbreak settings. J Clin Virol 82: 119–125. https://doi.org/10.1016/j.jcv.2016.07.012 PMID: 27479175

19. Nasheri N, Petronella N, Ronholm J, Bidawid S, Corneau N (2017) Characterization of the Genomic Diversity of Norovirus in Linked Patients Using a Metagenomic Deep Sequencing Approach. Front Microbiol 8: 73. https://doi.org/10.3389/fmicb.2017.00073 PMID: 28197136

20. Quer J, Gregori J, Rodriguez-Frias F, Buti M, Madejon A, Perez-del-Pulgar S, et al. (2015) High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. J Clin Microbiol 53: 219–226. https://doi.org/10.1128/JCM.02093-14 PMID: 25378574

21. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, et al. (2013) Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. PLoS One 8: e83361. https://doi.org/10.1371/journal.pone.0083361 PMID: 24391758

22. Gregori J, Salicru M, Domingo E, Sanchez A, Esteban JI, Rodriguez-Frias F, et al. (2014) Inference with viral quasispecies diversity indices: clonal and NGS approaches. Bioinformatics 30: 1104–1111. https://doi.org/10.1093/bioinformatics/btt768 PMID: 24389655

23. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33: 1870–1874. https://doi.org/10.1093/molbev/msw054 PMID: 27004904

24. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. (2007) Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. J Virol 81: 9932–9941. https://doi.org/10.1128/JVI.00674-07 PMID: 17609280

25. Lindesmith LC, Costantini V, Swanstrom J, Debbink K, Donaldson EF, Vinje J, et al. (2013) Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. J Virol 87: 2803–2813. https://doi.org/10.1128/JVI.03106-12 PMID: 23269783

26. Cuevas JM, Combe M, Torres-Puente M, Garijo R, Guix S, Buesa J, et al. (2016) Human norovirus hyper-mutation revealed by ultra-deep sequencing. Infect Genet Evol 41: 233–239. https://doi.org/10.1016/j.meegid.2016.04.017 PMID: 27094861

27. Carmona-Vicente N, Vila-Vicent S, Allen D, Gozalbo-Rovira R, Iturriza-Gomara M, Buesa J, et al. (2016) Characterization of a Novel Conformational GII.4 Norovirus Epitope: Implications for Norovirus-Host Interactions. J Virol 90: 7703–7714. https://doi.org/10.1128/JVI.01023-16 PMID: 27307569

28. Siebenga JJ, Beersma MF, Vennema H, van Biezen P, Hartwig NJ, Koopmans M (2008) High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J Infect Dis 198: 994–1001. https://doi.org/10.1086/591627 PMID: 18774885

29. Petrignani M, van Beek J, Borsboom G, Richardus JH, Koopmans M (2015) Norovirus introduction routes into nursing homes and risk factors for spread: a systematic review and meta-analysis of observational studies. J Hosp Infect 89: 163–178. https://doi.org/10.1016/j.jhin.2014.11.015 PMID: 25601744

30. Bull RA, Eden JS, Rawlinson WD, White PA (2010) Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog 6: e1000831. https://doi.org/10.1371/journal.ppat.1000831 PMID: 20360972

31. Lindesmith LC, Donaldson EF, Baric RS (2011) Norovirus GII.4 strain antigenic variation. J Virol 85: 231–242. https://doi.org/10.1128/JVI.01364-10 PMID: 20980508

32. Lindesmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, et al. (2012) Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. PLoS Pathog 8: e1002705. https://doi.org/10.1371/journal.ppat.1002705 PMID: 22615565

33. Debbink K, Lindesmith LC, Donaldson EF, Costantini V, Beltramello M, Corti D, et al. (2013) Emergence of new pandemic GII.4 Sydney norovirus strain correlates with escape from herd immunity. J Infect Dis 208: 1877–1887. https://doi.org/10.1093/infdis/jit370 PMID: 23908476