JVI Journals.ASM.org

RNA Populations in Immunocompromised Patients as Reservoirs for Novel Norovirus Variants

Everardo Vega,[a] Eric Donaldson,[b] Jeremy Huynh,[b] Leslie Barclay,[a] Ben Lopman,[a] Ralph Baric,[b] Luke F. Chen,[c] Jan Vinjé[a]

Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[a]; Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA[b]; Duke University Hospital, Durham, North Carolina, USA[c]

ABSTRACT

Noroviruses are the leading cause of acute gastroenteritis outbreaks worldwide. The majority of norovirus outbreaks are caused by genogroup II.4 (GII.4). Novel GII.4 strains emerge every 2 to 4 years and replace older variants as the dominant norovirus. Novel variants emerge through a combination of recombination, genetic drift, and selection driven by population immunity, but the exact mechanism of how or where is not known. We detected two previously unknown novel GII.4 variants, termed GII.4 UNK1 and GII.4 UNK2, and a diverse norovirus population in fecal specimens from immunocompromised individuals with diarrhea after they had undergone bone marrow transplantation. We hypothesized that immunocompromised individuals can serve as reservoirs for novel norovirus variants. To test our hypothesis, metagenomic analysis of viral RNA populations was combined with a full-genome bioinformatic analysis of publicly available GII.4 norovirus sequences from 1974 to 2014 to identify converging sites. Variable sites were proportionally more likely to be within two amino acids ($P < 0.05$) of positively selected sites. Further analysis using a hypergeometric distribution indicated that polymorphic site distribution was random and its proximity to positively selected sites was dependent on the size of the norovirus genome and the number of positively selected sites.In conclusion, random mutations may have a positive impact on driving norovirus evolution, and immunocompromised individuals could serve as potential reservoirs for novel GII.4 strains.

IMPORTANCE

Norovirus is the most common cause of viral gastroenteritis in the United States. Every 2 to 3 years novel norovirus variants emerge and replace dominant strains. The continual emergence of novel noroviruses is believed to be caused by a combination of genetic drift, population immunity, and recombination, but exactly how this emergence occurs remains unknown. In this study, we identified two novel GII.4 variants in immunocompromised bone marrow transplant patients. Using metagenomic and bioinformatic analysis, we showed that most genetic polymorphisms in the novel variants occur near 0 to 2 amino acids of positively selected sites, but the distribution of mutations was random; clustering of polymorphisms with positively selected sites was a result of genome size and number of mutations and positively selected sites. This study shows that immunocompromised patients can harbor infectious novel norovirus variants, and although mutations in viruses are random, they can have a positive effect on viral evolution.

Noroviruses are the leading cause of gastroenteritis outbreaks worldwide (1, 2), causing approximately 50% of all such outbreaks. Noroviruses are the primary cause of community and outpatient visits for acute gastrointestinal disease (AGE) (3). Although for most people the disease is relatively mild and self-limiting, morbidity and mortality due to norovirus infection are a concern for young children, immunocompromised individuals, and the elderly (4–6). In the elderly population, noroviruses have been shown to have mortality rates (5, 7) second only to that of *Clostridium difficile* (8).

Norovirus can be classified into seven genogroups (G), of which viruses belonging to GI and GII are responsible for the majority of human disease (9, 10). GII contains 22 recognized genotypes (50), of which GII.4 is the most common, causing 69 to 72% of all norovirus outbreaks in North America (11, 12). The genome of noroviruses is organized into three open reading frames (ORF). ORF1 is translated as a polyprotein that is cleaved by the virus-encoded protease into six nonstructural proteins. ORF2 and ORF3 code for the major capsid protein (VP1) and the minor capsid protein (VP2), respectively. VP1 forms the icosahedral shell that is divided into two well-defined motifs, designated the S (shell) and protruding (P) domains. The P domain is further subdivided into the P1 and P2 subdomains. The surface-exposed P2 subdomain contains the antigenic and histo-blood group antigen (HBGA) binding sites and is characterized by a high mutation frequency (13). The rate of evolution for GII.4 noroviruses has been determined to be 4.3×10^{-3} nucleotide substitutions per site per year in the major capsid protein (14, 15). Studies with other viruses have identified a fine balance between fidelity and viral fitness, where any increase or decrease of fidelity has a negative fitness effect on virus replication and pathogenesis (16–21). For norovirus, aside from the VP1 region, there are few studies identifying positively selected codon sites (14, 16, 22, 23). Thus, the disease burdens associated with norovirus, which vary substantially and temporally, are regulated in part by a poorly understood

Received 2 September 2014 Accepted 23 September 2014

Published ahead of print 1 October 2014

Editor: D. S. Lyles

Address correspondence to Everardo Vega, evega@cdc.gov.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.02494-14.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.02494-14

TABLE 1 Patient information for specimens analyzed in this study[a]

Sample	Variant	Patient age	Patient sex	Norovirus RNA copy no.	Health information	Drug regimen
IC	GII.4 New Orleans	66 yrs	Male	2.7×10^9 RNA copies/ml	No known health complications	None reported
A	GII.4 UNK2	16 yrs	Male	2.4×10^8 RNA copies/ml	Combined immunodeficiency, bone marrow transplant, bilateral lung transplant, and graft-versus-host disease	Tacrolimus, methylprednisolone, infliximab, immunosuppressive drugs
A + 5	GII.4 UNK1, UNK2	17 yrs	Male	3.2×10^6 RNA copies/ml	Combined immunodeficiency, bone marrow transplant, bilateral lung transplant, graft-versus-host disease, and neutropenia	Hydrocortisone, acrolimus, methylprednisolone
B	GII.4 UNK1	16 mo	Female	3.3×10^7 RNA copies/ml	Lysosomal storage disease, bone marrow transplant, graft-versus-host disease	Anti-thymocyte globulin, methylprednisolone, busulfan, cyclophosphamide
C	GII.4 UNK1	22 mo	Female	2.3×10^8 RNA copies/ml	Leukodystrophy, bone marrow transplant, and graft-versus-host disease	Methylprednisolone, tacrolimus, basiliximab, cyclosporine

[a] IC, immunocompetent patient infected with GII.4 New Orleans; UNK1, novel, unknown GII.4 norovirus variant 1; UNK2, novel, unknown GII.4 norovirus variant 2.

process of recombination, continual evolution, and emergence of novel norovirus strains that circulate in human populations.

A growing number of reports have documented chronic norovirus shedding due to an immune disorder, HIV infection, or other underlying conditions (24–30). Recent studies have also shown that noroviruses are the most common pathogen in hospitalized children with an immunodeficiency and that the majority of these children have prolonged virus shedding (31). Previous studies involving immunocompromised individuals have identified a large diversity of norovirus populations in these chronic shedders (24–28), and over time antigenic changes occur (32). Norovirus metagenomic analyses of immunocompromised populations have been attempted, but these involve sequencing site-specific PCR products or using RNA virus population reconstruction software (33–35). Reconstructing the norovirus genome is difficult due to the conserved junction region in the ORF1/ORF2 region and the high error rates associated with reconstruction software (35). A full-genome analysis of norovirus RNA populations and how they relate to norovirus evolution has not been accomplished.

In this study, we determined site-by-site variation in the RNA population (a comparison of individual genomes that compose the mutant spectrum of a sample, also referred to as quasispecies) of four patients infected with norovirus. Two were immunosuppressed using a transplant drug regimen, one patient had combined immunodeficiency disease, and the last patient was a healthy elderly male. We analyzed the norovirus population from 1972 to 2014 using 541 full or near full-length genomes. Full-genome analysis identified NS4 and VP1 as the most positively selected genes in the norovirus genome, indicating an important yet unknown function for the NS4 protein. We tested for convergent sites between the immunocompromised and the circulating populations to determine similar evolutionary trends. We hypothesized that immunocompromised individuals infected with noroviruses are able to serve as reservoirs for novel GII.4 noroviruses by preferentially mutating around norovirus genome positively selected sites. We found that mutations occur disproportionally within 2 and 4 amino acids (aa) of positively selected sites. When genome size, number of selective sites, and available sites were taken into account, the location of mutations was not significant, indicating a random distribution of mutations. These results provide new insight into norovirus evolution.

MATERIALS AND METHODS

Patient information and specimens. A total of five stool specimens from four bone marrow transplant recipients were collected as part of an outbreak investigation (Table 1). Patients exhibited chronic diarrhea and were undergoing immunosuppressive drug therapy. Specimens from immunocompromised patients were collected within 1 day of each other unless otherwise indicated. All specimens tested positive for GII norovirus by real-time reverse transcription-PCR (RT-PCR) and typed as GII.4 by sequence analysis of region D and P2 (36), nucleotides (nt) 6469 to 6622 and nt 5904 to 6338, of Lordsdale virus (GenBank accession number X86557), respectively. The first three specimens were collected in April 2010 from different patients (A, B, and C) with acute gastroenteritis. The fourth specimen (A + 5) was collected from patient A 5 months after the first sample. A GII.4 New Orleans-positive specimen collected from an immunocompetent (IC) patient in a cruise ship outbreak in 2010 was processed concurrently to determine process-induced or naturally occurring polymorphisms (GenBank accession number JN595867).

Cloning and quantitation of norovirus genomes. The VP1 and partial VP2 regions were amplified and cloned using previously described methods (36). Briefly, RNA was extracted from 10% stool suspensions using an automated KingFisher magnetic particle processor (Thermo Fisher Scientific, Pittsburgh, PA). A 2.3-kbp fragment from VP1 and part of VP2 were amplified using GII norovirus conserved primers Ring2-PCR and panGIIR (37). Ten clones from each specimen, a total of 50, were bidirectionally sequenced and analyzed (GenBank accession numbers KF806495 to KF806534). Ten clones from the immunocompetent individual were identical to the sequence with accession number JN595867. Norovirus RNA copy number was determined by real-time RT-PCR as previously described (38) using a purified and DNase-treated RNA transcript from a GII.7 outbreak (GenBank accession number GU134965).

Deep-sequencing sample preparation. Specimens were processed as previously described (39). Briefly, clarified 10% stool suspensions were prepared in phosphate-buffered saline (PBS). After clarification by low-speed centrifugation, the supernatant was extracted three times with Vertrel XF and pooled supernatants were filtered through a 0.45-μm filter before centrifugation over a 30% sucrose cushion at $237{,}000 \times g$ for 2 h at 4°C. The virus pellet was resuspended in PBS and treated with DNase Turbo, Benzonase, and RNase One for 2 h at 37°C. The RNA from intact

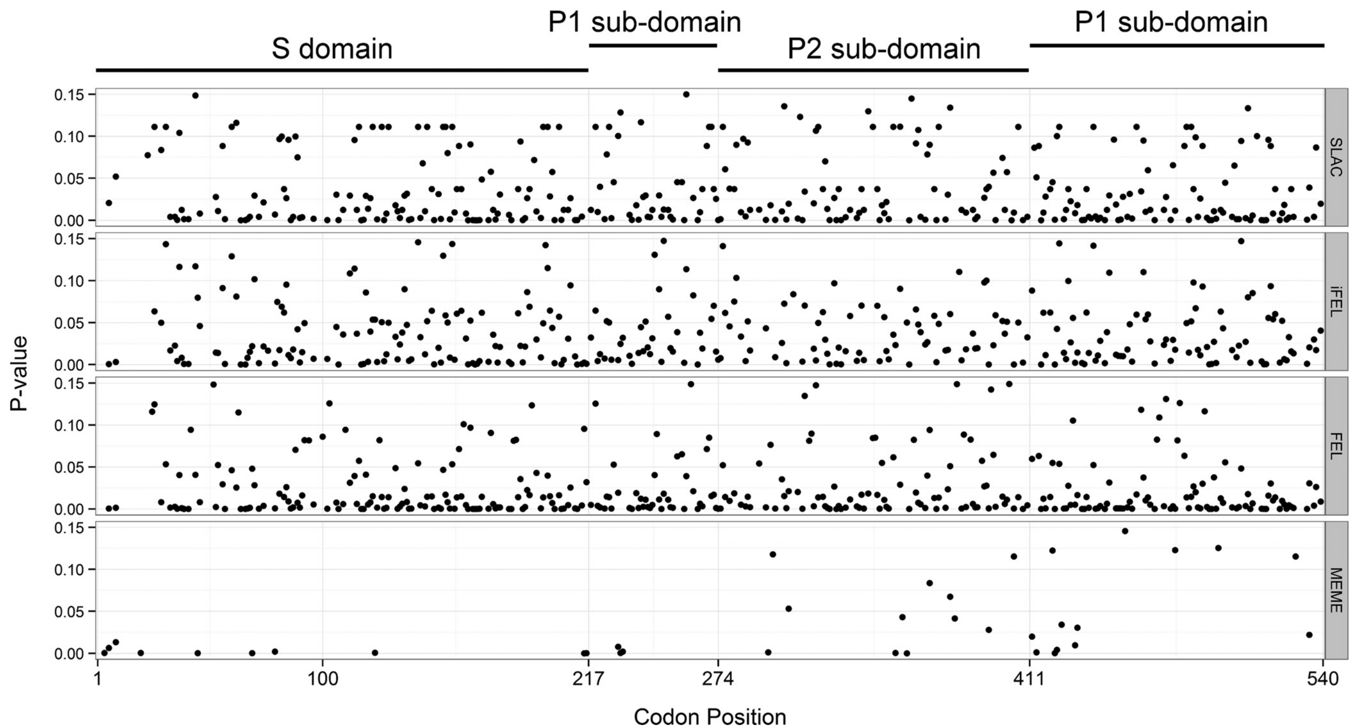

FIG 1 Method for detecting positive selection. Four methods were compared to determine the most appropriate statistical method to determine positive selection. The VP1 gene was parsed from 240 publicly available GII.4 norovirus sequences (S1) and run using the four available methods SLAC (single-likelihood ancestor counting), iFEL (internal fixed-effect likelihood), FEL (fixed-effect likelihood), and MEME (mixed-effect model of evolution) in DataMonkey (http://www.datamonkey.org). The *P* values from the four methods (S2) were graphed for comparison. The S, P1, and P2 domains are labeled.

virions was extracted using the Qiagen viral RNA kit according to the manufacturer's instructions (Valencia, CA).

SISPA reaction. Extracted RNA was reverse transcribed, purified, and amplified as described elsewhere (39). Briefly, RNA was reverse transcribed with Superscript III (Invitrogen, Carlsbad, CA) using random hexamers with a unique barcode. Double-stranded DNA was synthesized by Klenow reaction. The Klenow reaction was used as the template for random amplification using the unique barcode. The sequence-independent single primer amplification (SISPA) reaction mixtures were gel purified using a QiaQuick gel extraction kit (Qiagen). The purified DNA samples were submitted to the University of North Carolina high-throughput sequencing facility for sequencing using Roche 454 Life Science FLX Titanium chemistry (454 Life Sciences of Roche, Branford, CT).

Contig assembly. Sequences derived from Roche 454 sequencing were binned based on the barcode sequence added during the SISPA preparation, and the barcode sequences were trimmed from the reads in each unique bin, representing individual samples. Each sample bin was assembled *de novo* using CLC Genomics Workbench version 5.1 (CLCBio, Aarhus, Denmark) using default settings. Consensus sequences were determined for each contig using the 50% majority rule, and the consensus sequences were used to query the nonredundant nucleotide database using BLAST within the CLC Genomics Workbench interface. This approach was used to determine the nearest sequence neighbor for each consensus sequence, which was subsequently used to conduct a mapping using the closest BLAST match with the reference sequence. Mapping the trimmed sequences to the reference sequence provided a framework for identifying putative gaps in the sequencing information for each sequence, and the *de novo* assembly was used to verify the reference mapping.

Phylogenetic analysis. A total of 10 clones from each specimen for a total of 50 clones were selected and bidirectionally sequenced for phylogenetic analysis. The phylogenetic tree for VP1 of each specimen was built in PhyML 3.0 (http://www.atgc-montpellier.fr/phyml/versions.php) using the general time-reversible model (GTR).

Data processing and statistical analysis. The online webserver Galaxy (http://main.g2.bx.psu.edu/) was used to prepare the Roche 454 data for analysis. In order to differentiate between process-induced mutations and native mutations, only sites with a minimum coverage of 30-fold and minor variants of at least 10% of the total coverage were considered. Using this criterion, the immunocompetent specimen did not have any variable sites, indicating that no significant polymorphisms were present. After excluding minor or spurious mutations, Shannon entropy scores were generated for each nucleotide position using the Miller-Madow correction with a log base 2 using R (http://www.r-project.org/) with the package entropy. Site-by-site variation for each specimen was summarized as "bits," with an absolute maximum entropy or variation of 2 bits per nucleotide position if all four nucleotides were equally likely at any site. Nucleotide entropy scores were grouped by codon triplets, and scores were added together to obtain an entropy score for each codon position with a theoretical maximum of 6 bits for codon; thus, entropy is proportional to the number of mutations for each codon position. To determine positively selected sites in the norovirus population from 1974 to 2014, full- or near-full-genome ($n = 541$) GII.4 noroviruses were parsed and analyzed for each gene using the online webserver DataMonkey (http://www.datamonkey.org). Potential recombination was assayed using SBP (single breakpoint) in HyPhy (http://hyphy.org/w/index.php/Main_Page), and the data were parsed appropriately if a breakpoint was detected. Four methods—SLAC (single likelihood ancestor counting), iFEL (internal fixed-effect likelihood), FEL (fixed-effect likelihood), and MEME (mixed-effect model of evolution)—were run on a subset of VP1 to determine the best model (Fig. 1), based on known antigenic sites and reduction of background noise. The MEME method was determined to be the best model to identify positively selected sites in the norovirus genome (40). All *P* values obtained from DataMonkey MEME analysis were downloaded, and each codon position *P* value was binned into five groups, depending on the *P* value (<0.001, <0.01, <0.05, ≤0.1, or >0.1) and graphed along with the Roche 454 sequencing data in R using the ggplot2

TABLE 2 Norovirus deep-sequencing polymorphisms and MEME sites by gene

Gene	MEME analysis result		No. of variable sites determined by deep sequencing			
	No. of selected sites	Proportion of selected sites by gene size	Specimen A	Specimen B	Specimen C	Unique sites
N terminus (NS1-2)	18	0.05	21	7	3	28
VP1	31	0.06	54	22	4	75
P20 (NS4)	11	0.06	30	9	0	38
NTPase (NS3)	15	0.04	28	7	1	36
VPg (NS5)	2	0.02	14	1	0	15
Protease (NS6)	2	0.01	11	2	0	12
Polymerase (NS7)	12	0.02	58	9	5	68
VP2	22	0.08	19	10	2	29
Total	113	NA[a]	235	67	15	301

[a] NA, not available.

package. All polymorphisms which met the criteria were graphed, as previously discussed. To determine clustering of variable and selective sites, the amino acid absolute distance between all norovirus selective sites ($n = 113$) and variable sites ($n = 321$) was determined for each gene. The nearest selective and variable sites, by amino acid distance, were collected and binned. A proportion test was run on each bin, using R proportion test, to determine significantly different proportions in each bin. To determine significant clustering in bins, a hypergeometric test to account for nonreplacement sampling was run using R.

Hypergeometric distribution. The hypergeometric distribution is similar to a binomial distribution but accounts for sampling in a nonrenewed population. To accurately determine clustering at each binned site, once each site is selected it is not reincorporated into the available population or is nonrenewed. The hypergeometric distribution for at least X number of counts is determined as follows:

$$p(X > x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{N}{k}}$$

where m is the number of all possible sites within a bin, n is nonbinnable sites, N is the number of all sites, k is the total number of all variable sites, and x is the number of counts in the bin.

Norovirus simulation. Three norovirus simulations were run to determine the effects of genome size, random proximity of polymorphisms to selective sites, and number of mutations on norovirus RNA populations. All simulations were run and results graphed in R. The hypergeometric test was run using the same parameters obtained from the next-generation sequence data set. The size of the genome or available sites varied by 5%, 10%, and 40% of the norovirus genome using a uniform distribution; all the data were combined to achieve a uniform sample size to obtain the full range of P scores. To obtain the randomization of polymorphic and selective sites, a uniform distribution was used to sample the location of the sites. The number of mutations was obtained from the next-generation data and selective sites obtained from DataMonkey. The randomization was repeated 1,000 times, and data from each iteration were collected and graphed. In order to determine the association of the number of mutations and proximity to positively selected sites, the model as previously described was run, but instead of randomizing the location of sites, the number of variable sites was randomized from zero to all possible sites. The simulation was repeated 1,000 times. Data were collated as previously described by three possible results—0 to 2, 3 to 4, and >4 amino acids from the nearest selected site—and graphed.

Nucleotide sequence accession numbers. All norovirus sequences in this study are available in GenBank. The UNK1 and UNK2 GII.4 virus sequences from the three patients are available with the following accession numbers: KF806495 to KF806534. The GII.4 New Orleans 2009 norovirus from the immunocompetent patient has accession number JN595867, and the RNA transcript was based on sequence data for accession number GU134965.

RESULTS

Norovirus viral loads ranged from 10^6 to 10^9 genomic copies per ml. The specimen from the immunocompetent patient had the highest titer, 2.7×10^9 genome copies per ml. Patient A, the only patient with immunodeficiency disease, had the greatest number of polymorphisms detected (Tables 1 and 2). The other two bone marrow transplant patients, B and C, with immunosuppression by drug therapy had norovirus genome copy titers that were 10-fold lower than and equivalent to, respectively, that of patient A. There was no association between norovirus genome counts and number of polymorphisms detected.

Identification of viral RNA populations by Sanger sequencing. Three patients—patients A, B, and C—tested positive for two novel GII.4 variants, UNK1 and UNK2, distinct from each other and not belonging to any known GII.4 virus based on patristic distance (Fig. 2) or phylogenetic analysis (Fig. 3) (50). Both variants UNK1 and UNK2 contained deletions in the P2 domain of the capsid compared to modern noroviruses since GII.4 Farmington Hills 2002. One deletion occurred in aa 391, adjacent to the external HBGA binding site 2, and the other at aa 298, at a protruding site on the capsid surface (not shown due to 100% deletion in all clones). None of the deletions were present in the GII.4 norovirus variants circulating in 2010. The deletion at aa 298 had not been previously identified in any GII.4 variant known to date. The average mutation frequencies of VP1, based on a VP1 of 1,620 nt, obtained from each specimen were 6.6×10^{-2}, 3.0×10^{-2}, and 6.2×10^{-4} substitution per site in the GII.4 major capsid protein for specimens A, B, and C, respectively. Specimen A + 5, collected 5 months after the first specimen, contained a mixed infection: the GII.4 variants UNK1 and UNK2 were present in patient A, similar to the virus in patient C. After separating the two variants, the mutation frequencies were 1.8×10^{-3} and 6.2×10^{-4} substitution per site for each variant in the patient. The mutation frequency for the immunocompetent specimen could not be determined because it did not contain any nucleotide substitutions among the 10 clones recovered.

Seven of the 50 clones had a premature stop codon at the end of the S domain. Six out the seven mutants with this deletion were from the same GII.4 variant, GII.4 UNK2, and were from the same person,

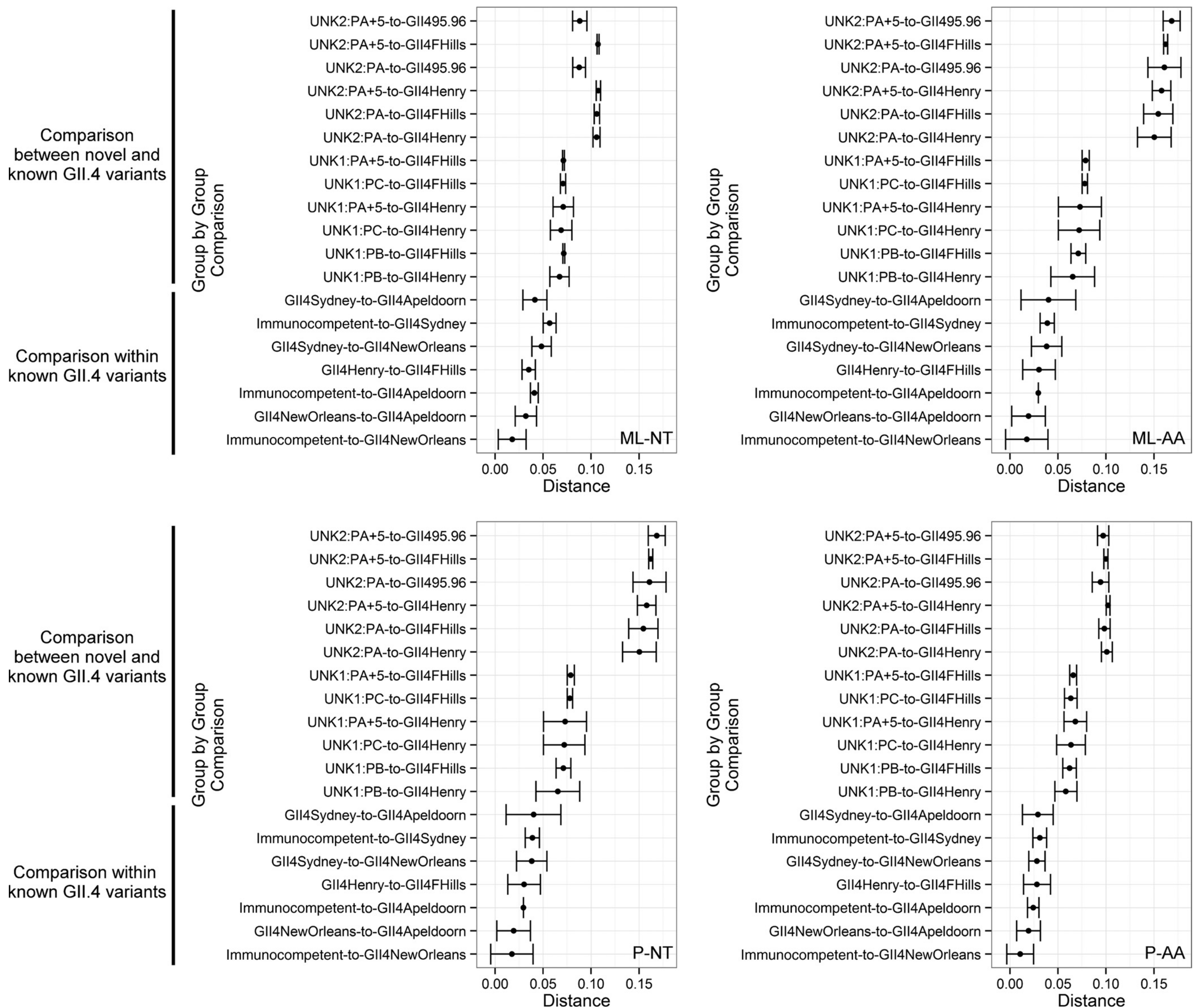

FIG 2 Patristic and P-distance comparison of novel GII.4 variants. The nucleotide maximum likelihood tree (ML-NT), amino acid maximum likelihood tree (ML-AA), nucleotide P-distance (P-NT), and amino acid P-distance (P-AA) between GII.4 variants were graphed. The error bars represent ±2 SDs for each group comparison. UNK1 and UNK2 are described in this study. The other variants and their GenBank accession numbers are GII495.96 (United States, 1995-1996), AY741811; GII4Henry (Henry, 2001), EU310927; GII4FHills (Farmington Hills, 2002), AY502023; GII4Apeldoorn (Apeldoorn, 2008), HQ009513; GII4NewOrleans (New Orleans, 2009), GU445325; and GII4Sydney (Sydney, 2012), JX459908. PA, PB, and PC refer to patients A, B, and C, respectively.

with three of the deletion mutants present in the cloning data 5 months afterwards (Fig. 3). The deletion mutants were phylogenetically distinct from the other clones from the same person (Fig. 3). The deletions in the S domain led to premature stop codons with the predicted protein encompassing most of the S domain. Most of the deletion mutants in patient A were nearly identical to the deletion mutants found 5 months after the first sample (Fig. 3).

In the cloning data set, most nonsynonymous mutations occurred in the P2 subdomain, followed by the P1 and S domains. Because mutations in VP1 were clustered by domain, each domain was analyzed separately. The nonsynonymous/synonymous (dn/ds) ratios for the P1 and S domains were less than 1, indicating overall negative selection. The P2 domain for A, B, and C had dn/ds ratios of 1.7, 1.25, and 1.15, respectively, indicating the presence of positive selection pressure in all specimens.

Identification of norovirus polymorphisms in the specimen population and comparison to circulating viruses. To further investigate the RNA virus population (quasispecies of each specimen), specimens from patients A, B, and C and the IC patient were analyzed by deep sequencing using Roche 454 sequencing. The A + 5 sample was not processed further because of the presence of a mixed infection. Genetic polymorphisms from the deep-sequencing data were summarized as bit scores as previously described. Bit scores for each gene were then overlaid on the *P* values obtained from the MEME analysis to identify similar sites between noroviruses from the immunocompromised patients and circulating noroviruses. In the GII.4 circulating population, a total of 113 positively selected codon sites were identified by MEME analysis with a *P* value score of ≤0.1 (Table 2). All scores from the MEME analysis that were ≤0.1 were selected and binned accord-

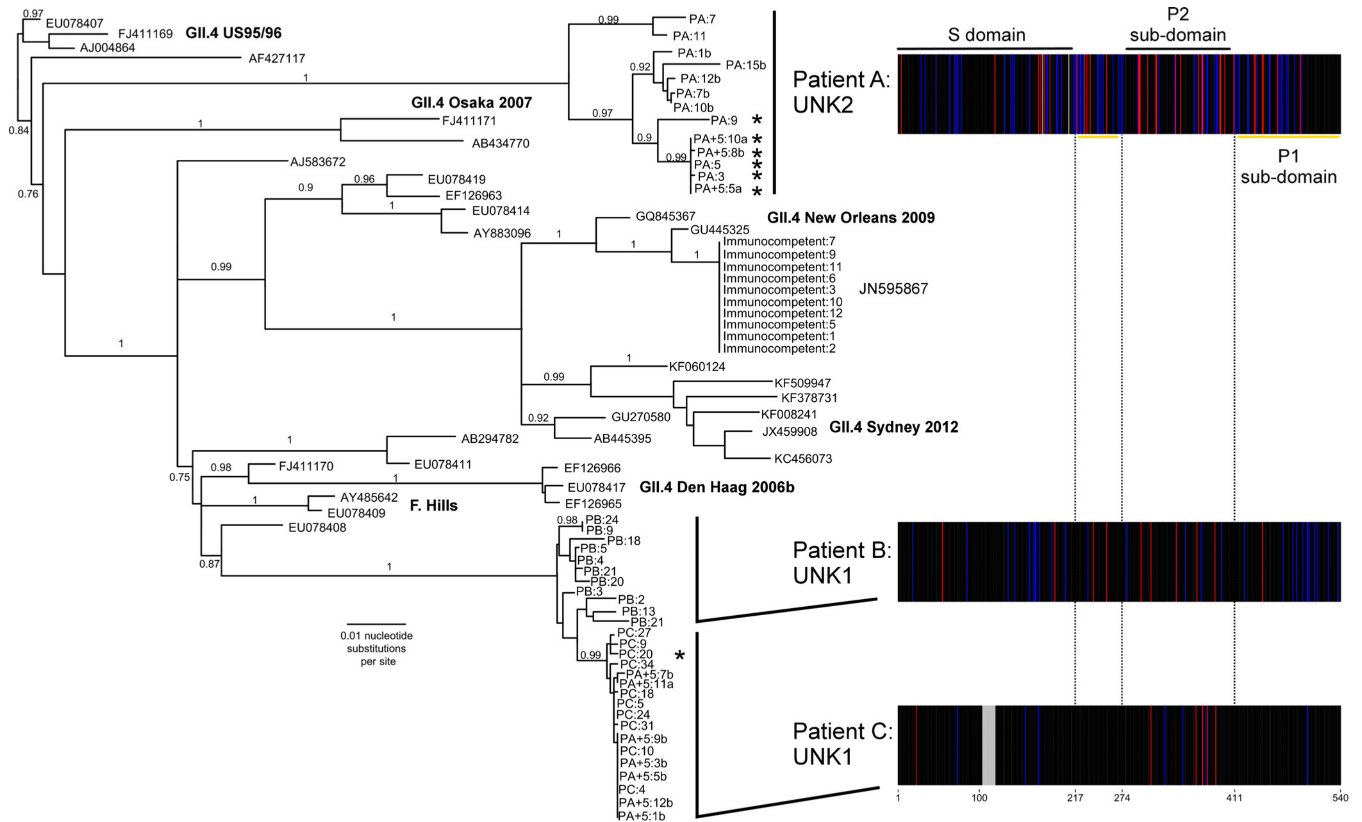

FIG 3 Sanger sequencing analysis. Fifty clones, 10 per specimen, were selected and bidirectionally sequenced. The major capsid protein for all sequences along with known GII.4 variants was run in PhyML. The sequences of the clones were clustered phylogenetically, and all nucleotide sequences were translated. All synonymous (blue), nonsynonymous (red), and deleted (gray) codons were summarized graphically. A subset (*) were found to have premature stop codons. The amino acid positions are noted on the bottom of the graphic. The S domain and P1 and P2 subdomains are labeled accordingly. GII.4 Farmington Hills 2002 is labeled as F. Hills.

ing to their P values (<0.001, <0.01, <0.05, ≤ 0.1, or >0.1). Within this data set, the VP1 gene had the highest number of positively selected sites, 31, whereas the NS6 and NS5 were the most conserved, with two positively selected sites each (Table 2). Adjusting for gene size, VP2 contained the largest number of positively selected sites, with 8% of the gene under positive selection. The VP1 and NS4 were equal, at 6% each. The NS7 contained one of the lowest proportions of selected sites, at 2%.

Regarding the deep-sequencing data, the GII.4-positive specimen from the IC patient contained no mutations after filtering data for low coverage or potentially spurious mutations, but due to low coverage, only VP1 was analyzed thoroughly (Fig. 4). A

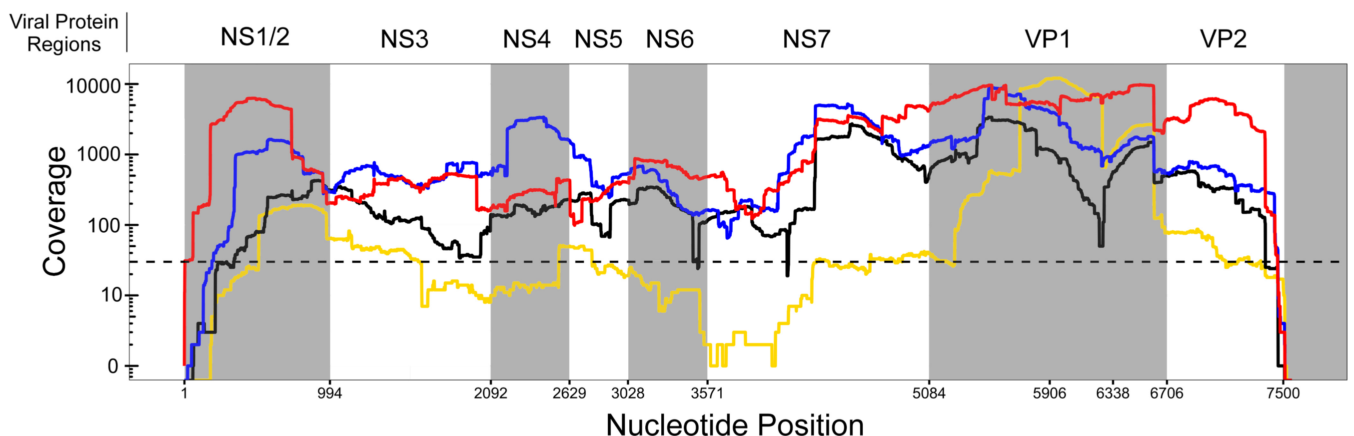

FIG 4 Genome sequence coverage of four GII.4 noroviruses. Four specimens were analyzed by next-generation sequencing. The total number of reads for specimens A (black), B (blue), and C (red) and the immunocompetent specimen (gold) are shown. The horizontal line is the cutoff below which data were not included for analysis. Viral protein regions are labeled accordingly.

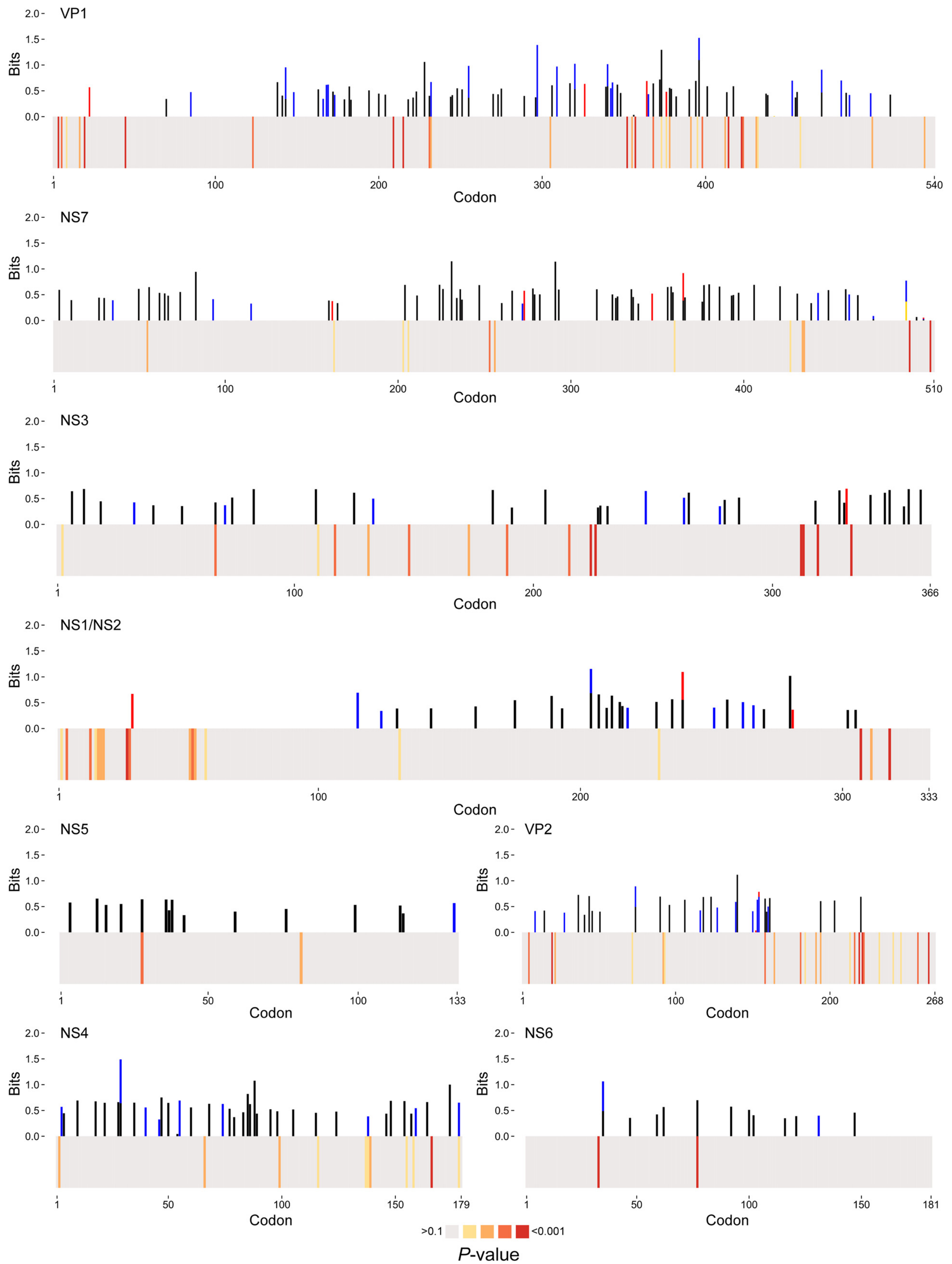

FIG 5 Deep-sequencing analysis of norovirus genome. Entropy bit scores at each codon position for each gene are graphed onto the binned *P* value results for each codon position from the MEME analysis. All specimens' genetic polymorphisms, if present, are graphed. Specimens A (black), B (blue), and C (red) are included. No data from the immunocompetent specimen were available for analysis. The *P* values were binned according to score: <0.001, <0.01, <0.05, ≤0.1, or >0.1. Each virus gene is graphed separately and labeled accordingly. Position summaries for counts, statistical analysis, and results are included as Data Sets S3 to S6 in the supplemental material. Genome alignments can be provided upon request.

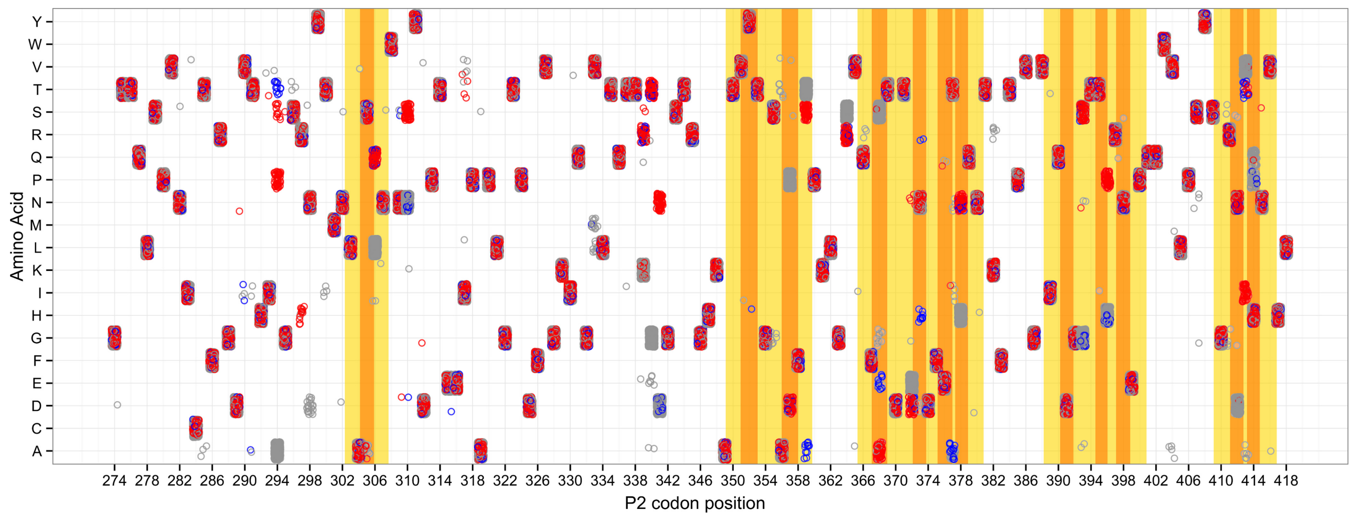

FIG 6 Norovirus P2 subdomain polymorphisms for GII.4 noroviruses, 2006 to 2012 major variants. Norovirus GII.4 polymorphisms for 431 noroviruses from GII.4 2006b (Minerva [gray]), GII.4 2009 (New Orleans), and GII.4 2012 (Sydney) variants were identified. Residues within two codons (yellow) of positively selected sites (orange) were identified.

total of 317 sites throughout the norovirus genome were available for analysis after discarding minimal or spurious mutations. Of these sites, 301 were unique among the three specimens (Table 2 and Fig. 5). Specimen A contained the highest number of mutations (235), followed by specimen B (67), and specimen C (15) (Table 2). The NS5, NS3, and NS6 genes contained the lowest proportions of mutations by gene size, 0.01 to 0.04 (Table 2).

Distribution and magnitude of mutations by gene. To determine if mutations were randomly distributed across the norovirus genome, we analyzed the number of mutations per gene using the hypergeometric test. Only the NS4 gene had a statistically significant ($P < 0.001$) number of mutations (Table 2). The VP1 ($P = 0.22$), NS7 ($P = 0.37$), and NS5 ($P = 0.75$) genes did not have significant numbers of mutations, whereas the NS1-2 ($P = 0.03$), NS3 ($P = 0.024$), and NS6 ($P = 0.002$) genes had significantly fewer mutations. To further investigate VP1, we separated VP1 based on the three recognized domains, S, P1, and P2. P2 had a significantly greater number of mutations detected ($P = 0.013$), the number of mutations for P1 was not significant ($P = 0.502$), and the S domain had significantly fewer mutations ($P < 0.001$).

The VP1 gene had four (1.53, 1.39, 1.29, and 1.16 bits) of the five (NS4: 1.49 bits) greatest entropy sites on the norovirus genome corresponding to codon sites 396, 297, 373, and 228, respectively (Fig. 5). The VP2 had one codon site, 140, with an entropy score over 1 and a bit score of 1.12. The majority of nonstructural genes, NS1-2, NS4, NS6, and NS7, had at least one site with ≥1 bit of entropy. Gene NS1-2 codons with scores of ≥1 bits were at positions 204, 239, and 280, NS4 at positions 29, 88, and 174, NS6 at position 35, and NS7 at codons 231, 291, and 365.

Clustering analysis between MEME and entropy sites. To obtain an unbiased estimation of clustering between variable sites from the norovirus RNA virus population of each patient and the positively selected sites from the MEME analysis, the difference between all sites was calculated. Based on previous GII.4 variants, residue changes occur within 0 to 2 amino acids of positively selected sites (Fig. 6). The minimal difference or nearest site between any two sites (variable to positively selected and positively selected to variable) from each gene was binned into a series of amino acid distance groups (Fig. 7). There was a predominance of sites between both variable to positively selected and positively selected to variable within two amino acids of each other (Fig. 7). The number of variable to selective and selective to variable sites decreased as the distance between the sites increased. There were significantly more positively selected, $n = 56$, sites within 0 to 2 amino acids of polymorphic sites ($P < 0.001$). Additionally, there were significantly more variable sites within four amino acids (0 to 2 aa, $n = 66$, and 3 or 4 aa, $n = 32$) of positively selected MEME sites ($P < 0.001$) for both groups. All other groups were not significant ($P \geq 0.65$). A breakdown by gene indicated that NS3, NS4, VP1, and VP2 had a variable to selective site distribution similar to that of the complete norovirus genome (Fig. 7).

In order to obtain a more accurate estimate of significance, a hypergeometric test was run over the entire norovirus genome, taking into account mutations by distance, number of positive selected sites, genome size, and total number of polymorphic sites. Although there were proportionally more sites within 0 to 2 amino acids of each other, the counts were not significant when accounting for all variables ($P = 0.836$), indicating that clustering was due to random chance and was not preferential. To further investigate the discrepancy between the proportion of sites within 0 to 2 amino acids of selective sites and the results of the hypergeometric test, data were simulated to identify the effect of genome size where all other variables remained constant. Genome size data simulation indicated that the size of the norovirus genome would have to be 30% smaller (corresponding to ≈1,800 aa) in order for the number of mutations within 0 to 2 amino acids of selective sites to be significant (Fig. 8A). To verify that a high proportion of mutations would randomly fall within 0 to 2 amino acids of selective sites given the data obtained from this study, the location of both selective sites and genetic polymorphisms were randomized and binned. The results of the randomized site simulation were similar to the RNA virus population data from the immu-

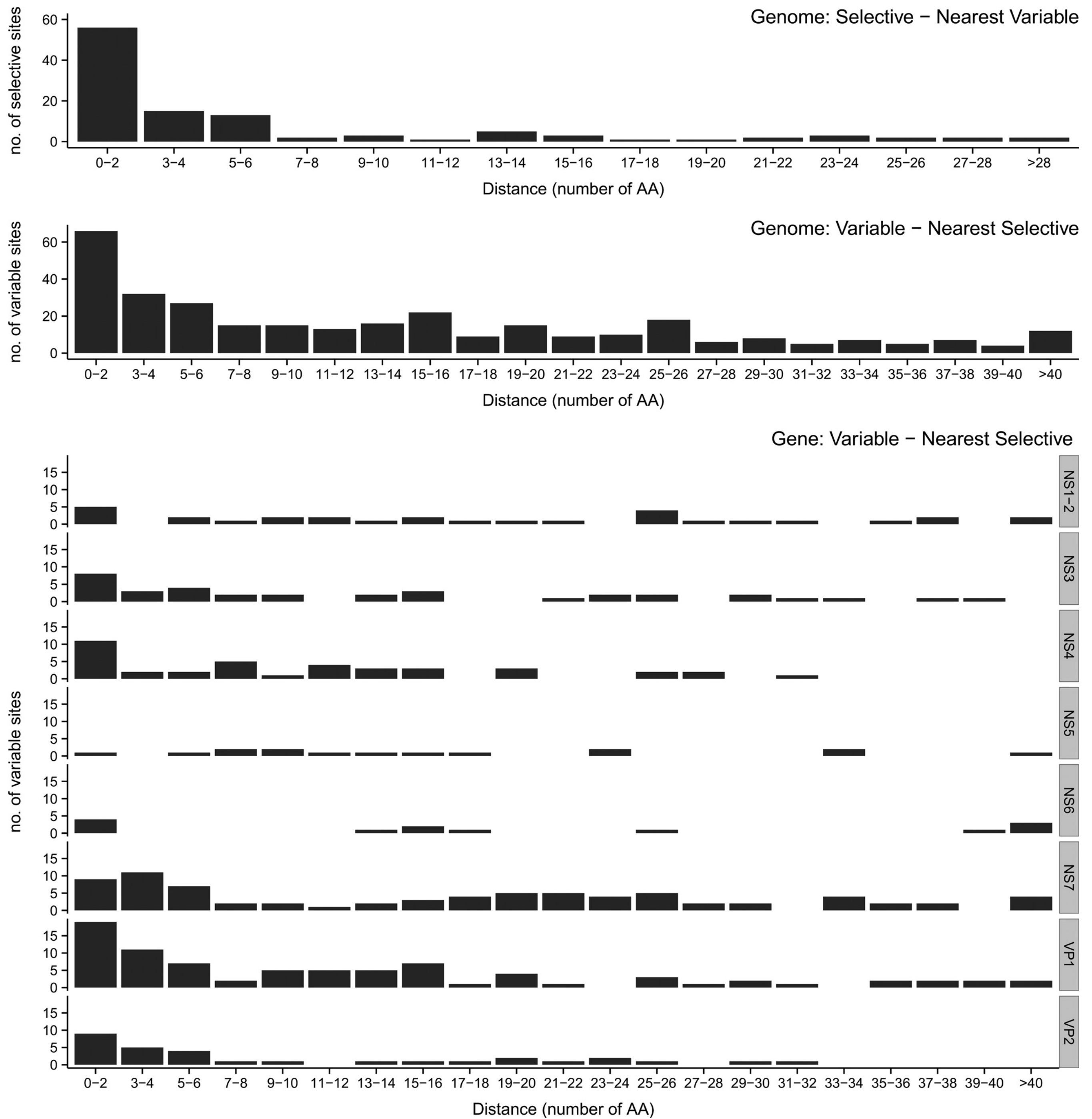

FIG 7 Cluster analysis of variable and positively selected sites. Distance counts for positively selected to nearest genetic polymorphic site (selective to nearest variable) in the norovirus genome was graphed and labeled accordingly. Distance counts for polymorphic to nearest positively selected (variable to nearest selective) sites in the norovirus genome and for each gene were graphed and labeled accordingly.

nocompromised people (Fig. 7 and 8B). The proportions of polymorphic sites within 0 to 2 amino acids of selective sites were nearly identical, 0.206 and 0.202 for the immunocompromised patients and simulated data, respectively. To better understand why the number of polymorphic sites seemed to randomly cluster around selective sites, we ran an additional simulation, varying only the number of mutations. The results show that as mutations increase, the residues around the selective sites become rapidly saturated with genetic polymorphisms, and increasing the number of mutations does not provide an increase in positively selective site changes (Fig. 8C). Mutations occurring >4 amino acids away from positively selected sites will have little effect and cause neutral or negative effects on virus viability.

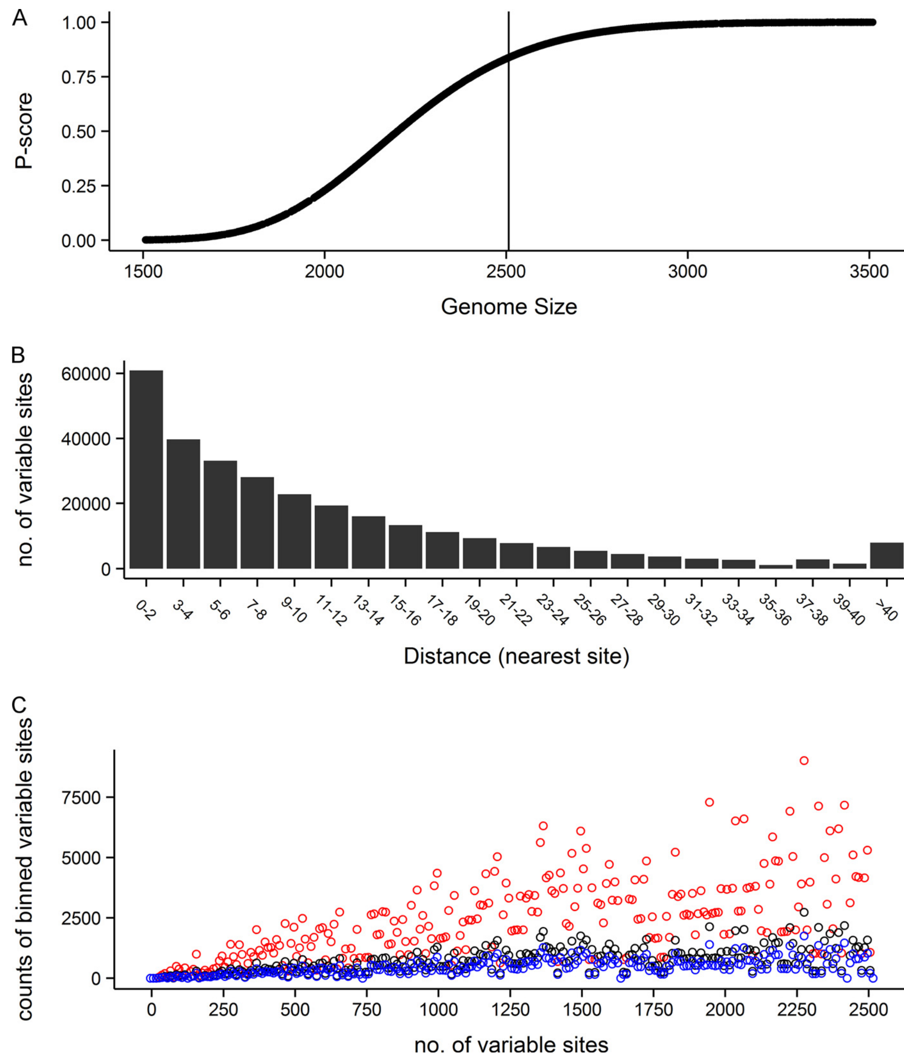

FIG 8 Norovirus data simulation. Data obtained from deep sequencing of immunocompromised patients were used to simulate the effects of several variables. (A) The effect of genome size on hypergeometric scores of the norovirus data were simulated, ranging from 1,500 to 3,500 amino acids. The norovirus genome size is indicated by the vertical line. (B) The random clustering between genetic polymorphisms and positively selected sites was investigated. The distance from the positively selected to nearest variable site was simulated 1,000 times, each randomly distributed and binned according to distance. (C) The number of mutations and its effect on random localization near positively selected site are shown. The number of variable sites was varied from zero to all sites, and the distance between a genetic polymorphism and the nearest positively selected site was obtained and binned based on three possible outcomes: amino acid distance of 0 to 2 (black), 3 or 4 (blue), and >4 (red).

DISCUSSION

GII.4 is the most common genotype associated with norovirus gastroenteritis outbreaks. The GII.4 viruses found in immunocompromised patients described here not only are novel but also are more closely related to GII.4 variants which currently do not circulate. The presence of these variants indicates that there may be a much larger pool of GII.4 noroviruses circulating in human populations than is captured by routine outbreak surveillance. These patients may serve as reservoirs of novel norovirus variants based on the random nature of mutations and the patients' lack of a fully functioning immune system. These results also show that the sequence space available for evolution of norovirus is strongly suppressed by the immune system. Only patient A, with primary immunodeficiency disease, had large numbers of polymorphisms. In addition, the high degree of genetic variability found in the NS4 gene, the only gene which had a significantly enriched number of genetic polymorphisms, indicates an important yet unknown role for this gene in norovirus pathogenesis.

The identification of these viruses in bone marrow transplant patients made it possible to assess the genetic diversity of an active infection *in vivo*. All three patients were undergoing immunosuppression to various degrees with drug regimens aimed at preventing graft rejection. The immunosuppressant drugs were broadly acting, such as corticosteroids, IL-2 competitors, and tumor necrosis factor alpha (TNF-α) binders, but there were also specific drugs which depleted T cells (Table 1). Using deep sequencing along with bioinformatic analysis, we were able combine data from an ongoing infection and analyze them in terms of norovirus evolution. Because human noroviruses are nonculturable, many of the norovirus gene products and their functions are unknown. Many of the variable sites in the RNA virus population of the three patients were within 0 to 2 amino acids of

the 113 positively selected sites. This result would argue that RNA virus populations in immunocompromised patients are mutating in a manner similar to that of circulating noroviruses. More thorough analysis argues, however, that this is not true and that mutations are occurring randomly. These divergent results from the same data set highlight norovirus evolutionary benefits and limitations.

We showed that the high number of mutations within 0 to 2 amino acids of positively selective sites is a result of the balance between genome size and the number of positively selected sites. The positively selective sites encompass 535 residues within 0 to 2 amino acids of the 113 identified in this study. Norovirus has ≈2,507 total residues in the entire genome, which means that 21% of the norovirus genome, or greater than 1 of 5 mutations, would have a positive effect on norovirus in the occurrence of a random mutation. Previous studies have shown that immunocompromised patients infected with norovirus shed antigenically distinct viruses over time (32). Taken together, this study and that of Debbink et al. (32) show that immunocompromised people can shed antigenically distinct viruses and that antigenic changes occur randomly. If either the number of selected sites or genome size is different, it would change the balance between positive and negative selection, with potentially detrimental effects for the virus. The number of mutations *in vivo* is balanced against the negative effect of mutations. Although we may not have identified all possible mutations, data simulation shows that an increase in the number of mutations may not be tolerated due to the increased risk of negative selection. To our knowledge, this is the first time this has been determined for norovirus.

The location of polymorphic sites was determined to be random across the genome, and the number of mutations per gene was not significant except in the case of the NS4 gene. The NS4 gene, P18, P20, and P22 in GV, GII, and GI noroviruses, respectively, has been shown to be the most variable gene of the nonstructural proteins (22, 41). Interestingly, the function of NS4 has not been identified, and it is unclear why this protein is highly variable. Two published studies, one involving human Norwalk virus (GI.1) and the other with murine noroviruses (MNV), indicate that NS4 proteins may not have the same function (42, 43). The NS4 in GI.1 has an endoplasmic reticulum (ER) export signal and is involved in Golgi disassembly, whereas MNV has no effect on the Golgi and lacks the ER export signal (42, 43). The significance of this difference is currently unknown. The presence of a membrane-associated domain (42) suggests that NS4 is involved in the replication complex and may associate with VP1. The clustering of mutations within this gene makes the NS4 a potentially important target for future research.

The VP1 gene is the best-studied norovirus gene. VP1 contains the antigenic determinants for norovirus and evolves in response to host population immunity. We did not find any increase in the number of polymorphic sites in VP1 or an increase in number of sites within 0 to 2 amino acids of positively selected sites. Nevertheless, the magnitudes of mutations are noteworthy. Of the five largest entropy bit scores (1.16 to 1.53), four occur in VP1, indicating that there is high turnover at the same sites. Sites 297, 373, and 396 are all known to affect antigenic or HBGA binding (44–46). Site 228 is not located in an antigenic or HBGA binding site but is predicted to be buried in the hinge region. Buried and noncontiguous residues have been previously determined to have conformational effects on the norovirus VLP and consequently antibody binding (47). Additionally, Lindesmith et al. (47) also showed through mutagenesis and surrogate neutralization assays that not only are the surface epitopes of importance but so are the buried sites which may alter the conformational surface and change the neutralization capacity of antibodies. Further studies are needed to characterize the hinge region as a possible site affecting immune evasion.

We showed that norovirus-infected immunocompromised patients can be potential reservoirs for novel norovirus strains by the presence of two novel GII.4 variant viruses, one of which infected another immunocompromised person. Our comparison of an immunodeficient norovirus RNA population with positively selective sites allowed us to further determine that mutations in such populations are random, and only the NS4 gene has a significant number of mutations. Although most mutations occurred within 0 to 2 amino acids of selective sites, this was a function of genome size and the number of selective sites and not specific clustering. Because immunocompromised patients are often hospitalized for long periods, there is a significant risk of direct or indirect contact with other susceptible patients, and it is imperative that strict hygiene practices be used to prevent further spread of these viruses (48). While a norovirus vaccine may not be effective for these patients, the potential use of antiviral compounds which are under development may assist in treating and preventing norovirus infections in the immunocompromised population (49).

ACKNOWLEDGMENTS

This work was supported by the Intramural Food Safety Program at the Centers for Disease Control and Prevention and in part by a grant from the National Institutes of Health to R.B. (AI 056351), a Gillings Innovation Laboratory award from the UNC Gillings School of Global Public Health to E.D., and a University of North Carolina research grant.

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. **Glass RI, Parashar UD, Estes MK.** 2009. Norovirus gastroenteritis. N. Engl. J. Med. **361:**1776–1785. http://dx.doi.org/10.1056/NEJMra0804575.
2. **Patel MM, Hall AJ, Vinjé J, Parashar UD.** 2009. Noroviruses: a comprehensive review. J. Clin. Virol. **44:**1–8. http://dx.doi.org/10.1016/j.jcv.2008.10.009.
3. **Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, Gray JJ, Letley LH, Rait G, Tompkins DS, O'Brien SJ, IID2 Study Executive Committee.** 2012. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut **61:**69–77. http://dx.doi.org/10.1136/gut.2011.238386.
4. **Payne DC, Vinjé J, Szilagyi PG, Edwards KM, Staat MA, Weinberg GA, Hall CB, Chappell J, Bernstein DI, Curns AT, Wikswo M, Shirley SH, Hall AJ, Lopman B, Parashar UD.** 2013. Norovirus and medically attended gastroenteritis in U.S. children. N. Engl. J. Med. **368:**1121–1130. http://dx.doi.org/10.1056/NEJMsa1206589.
5. **Trivedi TK, DeSalvo T, Lee L, Palumbo A, Moll M, Curns A, Hall AJ, Patel M, Parashar UD, Lopman BA.** 2012. Hospitalizations and mortality associated with norovirus outbreaks in nursing homes, 2009–2010. JAMA **308:**1668–1675. http://dx.doi.org/10.1001/jama.2012.14023.
6. **Bok K, Green KY.** 2012. Norovirus gastroenteritis in immunocompromised patients. N. Engl. J. Med. **367:**2126–2132. http://dx.doi.org/10.1056/NEJMra1207742.
7. **van Asten L, van den Wijngaard C, van Pelt W, van de Kassteele J, Meijer A, van der Hoek W, Kretzschmar M, Koopmans M.** 2012.

Mortality attributable to 9 common infections: significant effect of influenza A, respiratory syncytial virus, influenza B, norovirus, and parainfluenza in elderly persons. J. Infect. Dis. **206:**628–639. http://dx.doi.org/10.1093/infdis/jis415.
8. **Hall AJ, Curns AT, McDonald LC, Parashar UD, Lopman BA.** 2012. The roles of Clostridium difficile and norovirus among gastroenteritis-associated deaths in the United States, 1999–2007. Clin. Infect. Dis. **55:** 216–223. http://dx.doi.org/10.1093/cid/cis386.
9. **Green K.** 2013. Caliciviridae: the noroviruses, p 583–609. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields virology, 6th ed, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.
10. **Vinjé J.** 2 July 2014. Advances in laboratory methods for detection and typing of norovirus. J. Clin. Microbiol. http://dx.doi.org/10.1128/JCM.01535-14.
11. **Hasing ME, Lee BE, Preiksaitis JK, Tellier R, Honish L, Senthilselvan A, Pang XL.** 2013. Emergence of a new norovirus GII.4 variant and changes in the historical biennial pattern of norovirus outbreak activity in Alberta, Canada, from 2008 to 2013. J. Clin. Microbiol. **51:**2204–2211. http://dx.doi.org/10.1128/JCM.00663-13.
12. **Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinjé J.** 2014. Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. J. Clin. Microbiol. **52:**147–155. http://dx.doi.org/10.1128/JCM.02680-13.
13. **Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, Baric RS.** 2008. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med. **5:**e31. http://dx.doi.org/10.1371/journal.pmed.0050031.
14. **Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M.** 2010. Phylodynamic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. PLoS Pathog. **6:**e1000884. http://dx.doi.org/10.1371/journal.ppat.1000884.
15. **Bok K, Abente EJ, Realpe-Quintero M, Mitra T, Sosnovtsev SV, Kapikian AZ, Green KY.** 2009. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. J. Virol. **83:**11890–11901. http://dx.doi.org/10.1128/JVI.00864-09.
16. **Bull RA, Eden JS, Rawlinson WD, White PA.** 2010. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog. **6:**e1000831. http://dx.doi.org/10.1371/journal.ppat.1000831.
17. **Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R.** 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature **439:**344–348. http://dx.doi.org/10.1038/nature04388.
18. **Graham RL, Becker MM, Eckerle LD, Bolles M, Denison MR, Baric RS.** 2012. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. Nat. Med. **18:** 1820–1826. http://dx.doi.org/10.1038/nm.2972.
19. **Gnädig NF, Beaucourt S, Campagnola G, Bordería AV, Sanz-Ramos M, Gong P, Blanc H, Peersen OB, Vignuzzi M.** 2012. Coxsackievirus B3 mutator strains are attenuated in vivo. Proc. Natl. Acad. Sci. U. S. A. **109:**2294–2303.
20. **Levi LI, Gnädig NF, Beaucourt S, McPherson MJ, Baron B, Arnold JJ, Vignuzzi M.** 2010. Fidelity variants of RNA dependent RNA polymerases uncover an indirect, mutagenic activity of amiloride compounds. PLoS Pathog. **6:**e1001163. http://dx.doi.org/10.1371/journal.ppat.1001163.
21. **Pfeiffer JK, Kirkegaard K.** 2005. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. **1:**e11. http://dx.doi.org/10.1371/journal.ppat.0010011.
22. **Motomura K, Yokoyama M, Ode H, Nakamura H, Mori H, Kanda T, Oka T, Katayama K, Noda M, Tanaka T, Takeda N, Sato H, Norovirus Surveillance Group of Japan.** 2010. Divergent evolution of norovirus GII/4 by genome recombination from May 2006 to February 2009 in Japan. J. Virol. **84:**8085–8097. http://dx.doi.org/10.1128/JVI.02125-09.
23. **Eden JS, Hewitt J, Lim KL, Boni MF, Merif J, Greening G, Ratcliff RM, Holmes EC, Tanaka MM, Rawlinson WD, White PA.** 2014. The emergence and evolution of the novel epidemic norovirus GII.4 variant Sydney 2012. Virology **450-451:**106–113. http://dx.doi.org/10.1016/j.virol.2013.12.005.
24. **Nilsson M, Hedlund KO, Thorhagen M, Larson G, Johansen K, Ekspong A, Svensson L.** 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. http://dx.doi.org/10.1128/JVI.77.24.13117-13124.2003.
25. **Carlsson B, Lindberg AM, Rodriguez-Díaz J, Hedlund KO, Persson B, Svensson L.** 2009. Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection. J. Gen. Virol. **90:**432–441. http://dx.doi.org/10.1099/vir.0.005082-0.
26. **Hoffmann D, Hutzenthaler M, Seebach J, Panning M, Umgelter A, Menzel H, Protzer U, Metzler D.** 2012. Norovirus GII.4 and GII.7 capsid sequences undergo positive selection in chronically infected patients. Infect. Genet. Evol. **12:**461–466. http://dx.doi.org/10.1016/j.meegid.2012.01.020.
27. **Capizzi T, Makari-Judson G, Steingart R, Mertens WC.** 2011. Chronic diarrhea associated with persistent norovirus excretion in patients with chronic lymphocytic leukemia: report of two cases. BMC Infect. Dis. **11:** 131. http://dx.doi.org/10.1186/1471-2334-11-131.
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. http://dx.doi.org/10.1086/591627.
29. **Wingfield T, Gallimore CI, Xerry J, Gray JJ, Klapper P, Guiver M, Blanchard TJ.** 2010. Chronic norovirus infection in an HIV-positive patient with persistent diarrhoea: a novel cause. J. Clin. Virol. **49:**219–222. http://dx.doi.org/10.1016/j.jcv.2010.07.025.
30. **Robles JD, Cheuk DK, Ha SY, Chiang AK, Chan GC.** 2012. Norovirus infection in pediatric hematopoietic stem cell transplantation recipients: incidence, risk factors, and outcome. Biol. Blood Marrow Transplant. **18:** 1883–1889. http://dx.doi.org/10.1016/j.bbmt.2012.07.005.
31. **Frange P, Touzot F, Debré M, Héritier S, Leruez-Ville M, Cros G, Rouzioux C, Blanche S, Fischer A, Avettand-Fenoël V.** 2012. Prevalence and clinical impact of norovirus fecal shedding in children with inherited immune deficiencies. J. Infect. Dis. **206:**1269–1274. http://dx.doi.org/10.1093/infdis/jis498.
32. **Debbink K, Lindesmith LC, Ferris MT, Swanstrom J, Beltramello M, Corti D, Lanzavecchia A, Baric RS.** 2014. Within-host evolution results in antigenically distinct GII.4 noroviruses. J. Virol. **88:**7244–7255. http://dx.doi.org/10.1128/JVI.00203-14.
33. **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. http://dx.doi.org/10.1128/JVI.06712-11.
34. **Chan MC, Lee N, Ho WS, Law CO, Lau TC, Tsui SK, Sung JJ.** 2012. Covariation of major and minor viral capsid proteins in norovirus genogroup II genotype 4 strains. J. Virol. **86:**1227–1232. http://dx.doi.org/10.1128/JVI.00228-11.
35. **Prosperi MC, Yin L, Nolan DJ, Lowe AD, Goodenow MM, Salemi M.** 2013. Empirical validation of viral quasispecies assembly algorithms: state-of-the-art and challenges. Sci. Rep. **3:**2837. http://dx.doi.org/10.1038/srep02837.
36. **Vega E, Barclay L, Gregoricus N, Williams K, Lee D, Vinjé J.** 2011. Novel surveillance network for norovirus gastroenteritis outbreaks, United States. Emerg. Infect. Dis. **17:**1389–1395. http://dx.doi.org/10.3201/eid1708.101837.
37. **Vega E, Vinjé J.** 2011. Novel GII.12 norovirus strain, United States, 2009–2010. Emerg. Infect. Dis. **17:**1516–1518. http://dx.doi.org/10.3201/eid1708.110025.
38. **Trujillo AA, McCaustland KA, Zheng DP, Hadley LA, Vaughn G, Adams SM, Ando T, Glass RI, Monroe SS.** 2006. Use of TaqMan real-time reverse transcription-PCR for rapid detection, quantification, and typing of norovirus. J. Clin. Microbiol. **44:**1405–1412. http://dx.doi.org/10.1128/JCM.44.4.1405-1412.2006.
39. **Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB.** 2010. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. J. Virol. **84:**13004–13018. http://dx.doi.org/10.1128/JVI.01255-10.
40. **Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL.** 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. **8:**e1002764. http://dx.doi.org/10.1371/journal.pgen.1002764.
41. **Motomura K, Oka T, Yokoyama M, Nakamura H, Mori H, Ode H, Hansman GS, Katayama K, Kanda T, Tanaka T, Takeda N, Sato H, Norovirus Surveillance Group of Japan.** 2008. Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. J. Virol. **82:**11247–11262. http://dx.doi.org/10.1128/JVI.00897-08.
42. **Sharp TM, Guix S, Katayama K, Crawford SE, Estes MK.** 2010. Inhibition of cellular protein secretion by norwalk virus nonstructural protein p22 requires a mimic of an endoplasmic reticulum export signal. PLoS One **5:**e13130. http://dx.doi.org/10.1371/journal.pone.0013130.
43. **Hyde JL, Mackenzie JM.** 2010. Subcellular localization of the MNV-1

ORF1 proteins and their potential roles in the formation of the MNV-1 replication complex. Virology **406:**138–148. http://dx.doi.org/10.1016/j.virol.2010.06.047.

44. **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. http://dx.doi.org/10.1128/JVI.06189-11.
45. **Lindesmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, Lanzavecchia A, Baric RS.** 2012. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. PLoS Pathog. **8:**e1002705. http://dx.doi.org/10.1371/journal.ppat.1002705.
46. **Lindesmith LC, Costantini V, Swanstrom J, Debbink K, Donaldson EF, Vinjé J, Baric RS.** 2013. Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. J. Virol. **87:**2803–2813. http://dx.doi.org/10.1128/JVI.03106-12.
47. **Lindesmith LC, Donaldson EF, Beltramello M, Pintus S, Corti D, Swanstrom J, Debbink K, Jones TA, Lanzavecchia A, Baric RS.** 2014. Particle conformation regulates antibody access to a conserved GII.4 norovirus blockade epitope. J. Virol. **88:**8826–8842. http://dx.doi.org/10.1128/JVI.01192-14.
48. **Barclay L, Park GW, Vega E, Hall A, Parashar U, Vinjé J, Lopman B.** 2014. Infection control for norovirus. Clin. Microbiol. Infect. **20:**731–740. http://dx.doi.org/10.1111/1469-0691.12674.
49. **Costantini VP, Whitaker T, Barclay L, Lee D, McBrayer TR, Schinazi RF, Vinjé J.** 2012. Antiviral activity of nucleoside analogues against norovirus. Antivir. Ther. **17:**981–991. http://dx.doi.org/10.3851/IMP2229.
50. **Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, Green K, Martella V, Katayama K, Koopmans M.** 2013. Proposal for a unified norovirus nomenclature and genotyping. Arch. Virol. **158:**2059–2068. http://dx.doi.org/10.1007/s00705-013-1708-5.