



Population Genomics of GII.4 Noroviruses Reveal Complex Diversification and New Antigenic Sites Involved in the Emergence of Pandemic Strains

Kentaro Tohma,^a Cara J. Lepore,^a Yamei Gao,^a Lauren A. Ford-Siltz,^a  Gabriel I. Parra^a

^aDivision of Viral Products, Food and Drug Administration, Silver Spring, Maryland, USA

ABSTRACT GII.4 noroviruses are a major cause of acute gastroenteritis. Their dominance has been partially explained by the continuous emergence of antigenically distinct variants. To gain insights into the mechanisms of viral emergence and population dynamics of GII.4 noroviruses, we performed large-scale genomics, structural, and mutational analyses of the viral capsid protein (VP1). GII.4 noroviruses exhibited a periodic replacement of predominant variants with accumulation of amino acid substitutions. Genomic analyses revealed (i) a large proportion (87%) of conserved residues; (ii) variable residues that map on the previously determined antigenic sites; and (iii) variable residues that map outside the antigenic sites. Residues in the third pattern category formed motifs on the surface of VP1, which suggested extensions of previously predicted and new uncharacterized antigenic sites. The role of two motifs (C and G) in the antigenic makeup of the GII.4 capsid protein was confirmed with monoclonal antibodies and carbohydrate blocking assays. Amino acid profiles from antigenic sites (A, C, D, E, and G) correlated with the circulation patterns of GII.4 variants, with three of them (A, C, and G) containing residues (352, 357, 368, and 378) linked with the diversifying selective pressure on the emergence of new GII.4 variants. Notably, the emergence of each variant was followed by stochastic diversification with minimal changes that did not progress toward the next variant. This report provides a methodological framework for antigenic characterization of viruses and expands our understanding of the dynamics of GII.4 noroviruses and could facilitate the design of cross-reactive vaccines.

IMPORTANCE Noroviruses are an important cause of viral gastroenteritis around the world. An obstacle delaying the development of norovirus vaccines is inadequate understanding of the role of norovirus diversity in immunity. Using a population genomics approach, we identified new residues on the viral capsid protein (VP1) from GII.4 noroviruses, the predominant genotype, that appear to be involved in the emergence and antigenic topology of GII.4 variants. Careful monitoring of the substitutions in those residues involved in the diversification and emergence of new viruses could help in the early detection of future novel variants with pandemic potential. Therefore, this novel information on the antigenic diversification could facilitate GII.4 norovirus vaccine design.

KEYWORDS antigenic variation, caliciviruses, evolution, gastrointestinal infection, norovirus, phylogenetic analysis

Noroviruses are a leading cause of acute gastroenteritis, affecting all ages worldwide (1). They are second only to rotavirus in this respect, although this is changing in places where rotavirus vaccination is effective (2, 3). It is estimated that norovirus is responsible for approximately 685 million infections and 200,000 deaths worldwide, with a primary public health concern in children, the elderly, and immunocompromised

Citation Tohma K, Lepore CJ, Gao Y, Ford-Siltz LA, Parra GI. 2019. Population genomics of GII.4 noroviruses reveal complex diversification and new antigenic sites involved in the emergence of pandemic strains. *mBio* 10:e02202-19. <https://doi.org/10.1128/mBio.02202-19>.

Invited Editor Ralph S. Baric, University of North Carolina at Chapel Hill

Editor Mary K. Estes, Baylor College of Medicine

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Gabriel I. Parra, gabriel.parra@fda.hhs.gov.

K.T., C.J.L., and G.I.P. contributed equally to this article.

Received 19 August 2019

Accepted 30 August 2019

Published 24 September 2019

individuals (4–6). Norovirus outbreaks usually occur during the winter season in enclosed settings such as schools, hospitals, military facilities, and cruise ships. Because norovirus is highly contagious, outbreaks can be hard to control.

The norovirus genome is a positive-sense, single-stranded RNA molecule that is organized into three open reading frames (ORFs). ORF1 encodes a polyprotein that is cotranslationally cleaved by the viral protease into six nonstructural (NS) proteins required for replication. ORF2 encodes the major capsid protein (VP1) and ORF3 the minor capsid protein (VP2). The norovirus capsid consists of 180 copies of VP1, arranged in $T = 3$ icosahedral symmetry. X-ray crystallography of the VP1 revealed two structural domains: the shell (S) domain and the protruding (P) domain. The S domain is relatively highly conserved and forms the core of the capsid, while the P domain is more variable and extends to the exterior of the capsid protein (7, 8). The P domain interacts with host attachment factors, namely, human histo-blood group antigen (HBGA) carbohydrates, which could facilitate infection. Antibody (Ab)-mediated blocking of the VP1:HBGA interaction correlates with protection against norovirus disease (9, 10). Expression of VP1 results in the self-assembly of virus-like particles (VLPs) that are structurally and antigenically similar to native virions (7, 11, 12). Given the lack of a traditional cell culture system for human norovirus, experimentally developed VLPs have been an important tool to study norovirus immune responses and vaccine design.

Norovirus strains are highly diverse, with at least seven genogroups (GI to GVII) and over 40 genotypes defined based on differences in their VP1 sequences (13). While over 30 different genotypes from GI, GII, and GIV can infect humans, noroviruses from the GII.4 genotype are responsible for at least 70% of infections worldwide (14). Since the mid-1990s, six major norovirus GII.4 pandemics have been recorded worldwide and were associated with the following variants: Grimsby 1995 (or US95_96), Farmington Hills 2002, Hunter 2004, Den Haag 2006b, New Orleans 2009, and Sydney 2012 (15–17). The predominance of GII.4 viruses has been linked to the chronological emergence of variants in the human population, with new variants emerging around the time of the previous declines (15). The emergence of these variants has been correlated with changes on five different variable antigenic sites (namely, sites A to E) that map on the surface of the P domain; thus, new viruses can evade the human immune responses elicited to previously circulating variants (11, 18–20). Using the recently developed cell culture system for human noroviruses (21), two of these sites have been confirmed to be involved in virus neutralization (22). Studies have shown that antibodies that map to these sites can block the interaction of VP1 with carbohydrates from the HBGA; however, the antigenic sites of several monoclonal antibodies (MAbs) with blocking activity, raised against GII.4 viruses, have not been determined (11, 12, 19, 23–25). The evolving nature of GII.4 noroviruses could challenge the development of cross-protective vaccines against noroviruses; therefore, a better understanding of the mechanisms responsible for the antigenic diversification will facilitate vaccine design.

In this study, we adopted a large-scale genomics approach to identify sites that play a role in GII.4 evolution and antigenic diversification. We (re)defined the sites involved in the antigenic make-up of GII.4 pandemic variants, and found that intravariant diversification exhibited a stochastic pattern of evolution. Importantly, we identified four sites (amino acids 352, 357, 368, and 378) implicated with the emergence of predominant GII.4 variants, and that could help in the early detection of the next pandemic variant.

RESULTS

GII.4 intervariant evolution is characterized by the accumulation of substitutions in the P domain. In order to investigate the evolutionary patterns of GII.4 strains, we calculated the genetic differences of 1,601 nearly full-length ($\geq 1,560$ -nucleotide [nt]) VP1 sequences from GII.4 strains collected from 1974 to 2016. The phylogenetic tree of these sequences showed the presence of at least 11 different GII.4 variants emerging since 1995 (Fig. 1a). As shown previously (15, 26), GII.4 strains presented a chronological replacement of variants, with several unassigned intermediate strains

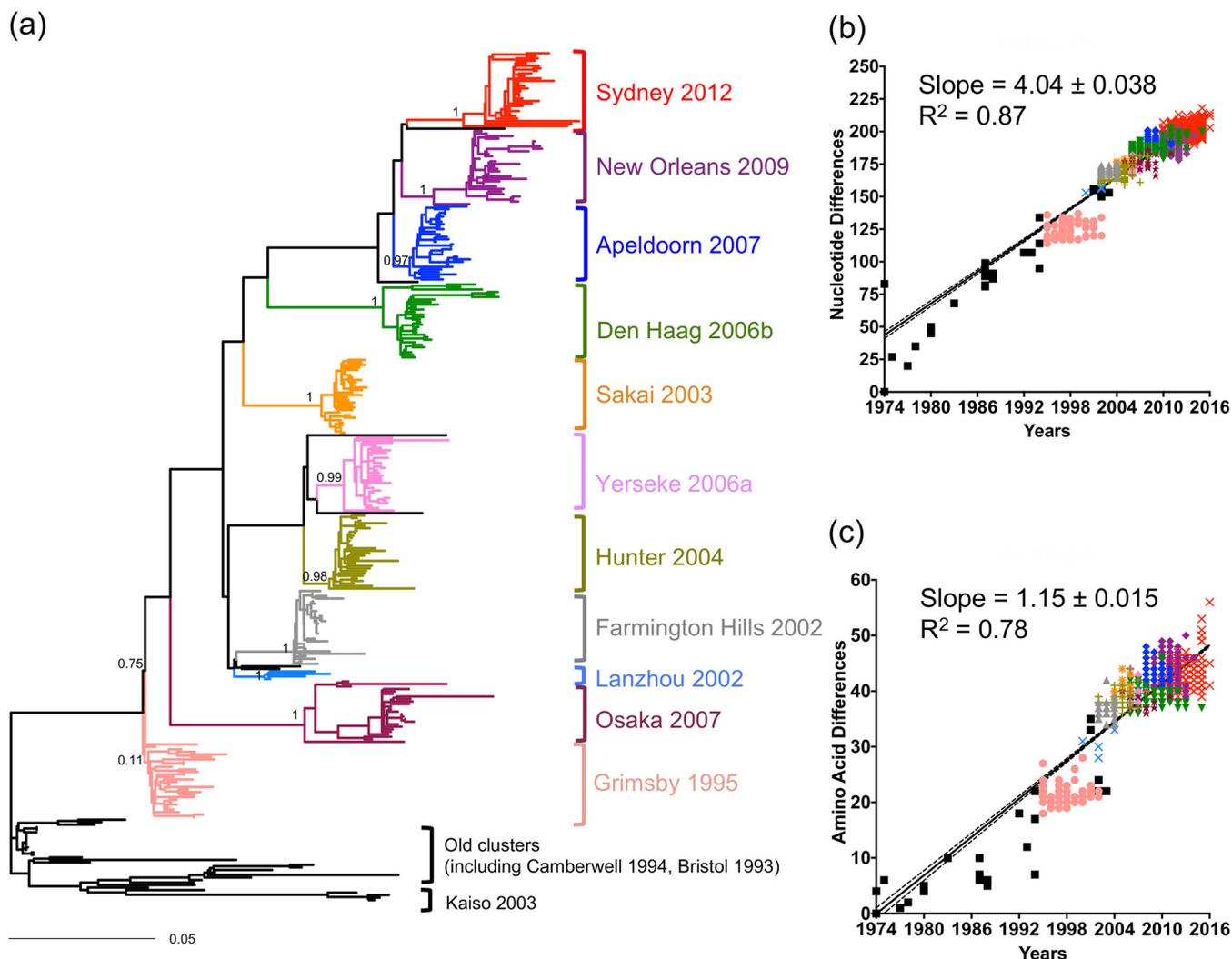
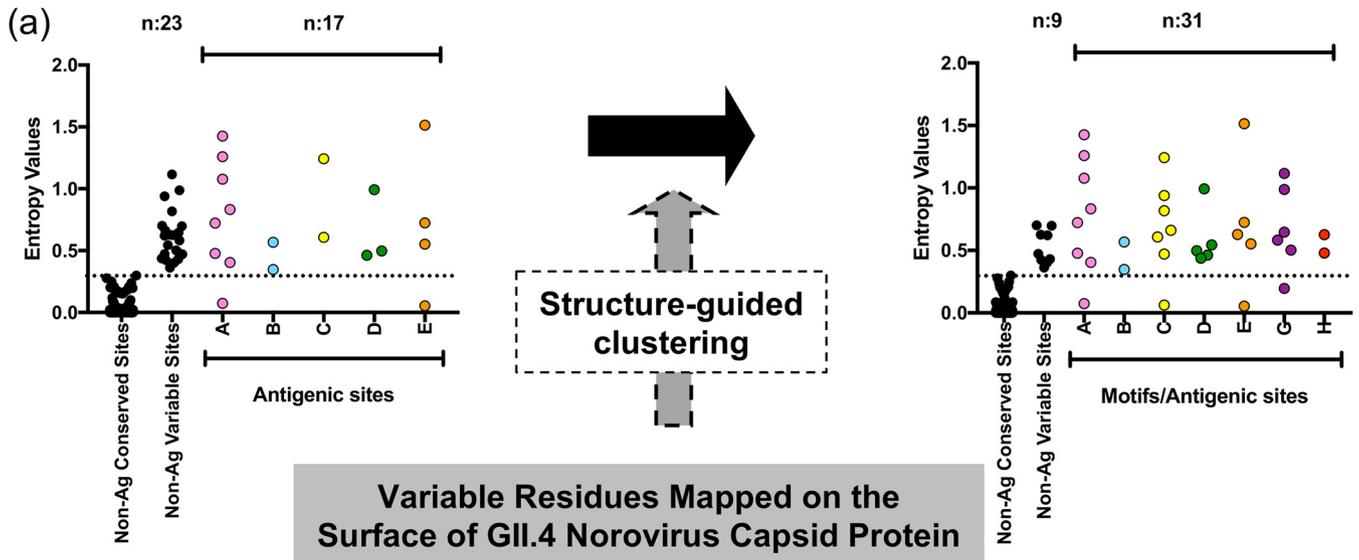


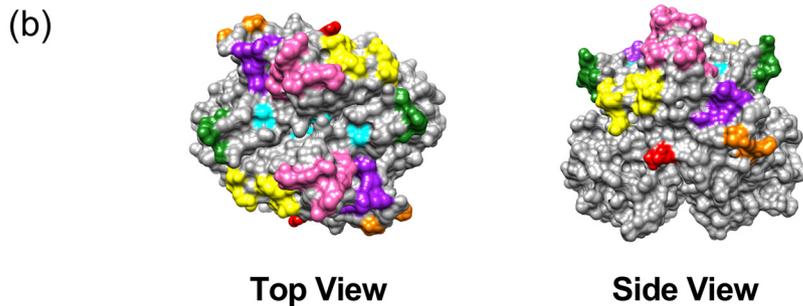
FIG 1 Evolution of the major capsid protein (VP1) from GII.4 noroviruses results in the accumulation of substitutions and the periodic emergence of variants. (a) Maximum likelihood tree of GII.4 noroviruses showing the circulation of different variants over time. Branches are colored based on variant determination results from the online Norovirus Typing Tool (<https://www.rivm.nl/mpf/typingtool/norovirus/>). Black branches in the tree represent sequences that did not cluster into a variant that circulated between 1995 and 2016. A subset of 308 sequences, from a total of 1,601, were used for tree reconstruction as indicated in Materials and Methods. Node support values calculated by the approximate likelihood-ratio test are shown on the major branches. (b and c) Graphical representations of nucleotide (b) and amino acid (c) pairwise differences of each sequence in the data set compared to the earliest strain, a GII.4 strain collected in 1974 (AB303922). A total of 1,601 nearly full-length VP1 GII.4 sequences were included for the pairwise analyses. The black solid line indicates the linear regression line, with dotted lines showing the 95% confidence interval of the best-fit line.

(see Fig. S1 in the supplemental material). Genetic analyses revealed an accumulation of substitutions in both nucleotide (Fig. 1b) and amino acid (Fig. 1c) sequences (coefficients of determination [R^2] of linear regression = 0.87 and 0.78, respectively). This pattern of accumulation of mutations was observed in all the subdomains of VP1 (S, P1, and P2; Fig. S2); however, higher slopes were noted in P2, where the five variable antigenic sites (A to E) are located, thus suggesting their role in the evolution and antigenic diversification of GII.4 noroviruses.

Identification of new antigenic sites of GII.4 noroviruses. To pinpoint the role of each amino acid within the P domain in the evolution of GII.4, we calculated the Shannon entropy to measure the residue diversity at the intervariant and intravariant levels. Because Grimsby-like viruses were the first recorded to cause large outbreaks worldwide, entropy values were calculated with strains (1,572 VP1 sequences) detected from 1995 to 2016. While minimal diversity was detected at the intravariant level (Fig. S3), the intervariant level revealed three substitution patterns: (i) a large number



Variable Residues Mapped on the Surface of GII.4 Norovirus Capsid Protein



Motif/Antigenic site A*	Motif B	Motif/Antigenic site C†	
294 295 296 297 298 368 372 373	333 389	339 340 341 375 376 377 378	
Motif/Antigenic site D*	Motif/Antigenic site E*	Motif/Antigenic site G†	Motif H
393 394 395 396 397	407 411 412 413 414	352 355 356 357 359 364	309 310

*Confirmed as antigenic sites by MABs
 †Confirmed as antigenic sites by MABs in this study

FIG 2 Conservation analyses redefined antigenic sites of the major capsid protein (VP1) from GII.4 noroviruses. (a) Shannon entropy was calculated to quantify amino acid variation for each site in the VP1. Analyses were calculated with strains (1,572 VP1 sequences) detected from 1995 to 2016. Data from the P domain are included here (amino acids 216 to 540). Residues were grouped depending on the degree of variability into the following categories: conserved sites (Shannon entropy value ≤ 0.3), sites mapping on antigenic sites (columns A to E), and variable sites that map outside antigenic sites (left-side dot plot). Based on structural analyses, 14 variable residues that mapped outside antigenic sites were clustered as part of novel motifs (potential antigenic sites) or extension of previously defined antigenic sites (right-side dot plot). (b) Residues forming these novel or expanded motifs/antigenic sites on the surface of the GII.4 major capsid protein are colored accordingly.

(87%, 285/325) of highly conserved residues, which include conserved antigenic site F (27) and likely maintain the structural integrity of the P domain; (ii) a small number (5%, 17/325) of variable residues that map on previously determined antigenic sites A to E (19); and (iii) 23 variable residues that map outside those antigenic sites (Fig. 2a). Among the 23 nonantigenic variable sites comprising the third pattern, 2 residues (residues 228 and 255) were surrounded by conserved residues on the surface of VP1, and 14 residues formed clusters (motifs) on the surface that could represent new antigenic sites or extensions of previously predicted, uncharacterized antigenic sites

(Fig. 2b). One motif comprises residues 339, 340, 341, 375, 376, 377, and 378. Because this motif included two residues (340 and 376) previously described as being among the five major antigenic sites (20), we extended the number of residues forming this antigenic site, site C (Fig. 2b). Although residue 375 represented low variability, it was included as part of the antigenic site as it mapped on the surface of the molecule and might potentially play a role in antibody recognition. Antigenic site D was also extended to include two additional residues (396 and 397), as they clustered with original residues (393, 394, and 395) on the surface (19). The new motif, denominated motif G, included residues 352, 355, 356, 357, 359, and 364; the last motif, denominated motif H, included residues 309 and 310. A summary of the residues corresponding to each of the motifs/antigenic sites is shown in Fig. 2b. Profiling of the temporal frequency of the amino acid sequence patterns (mutational patterns) of previously characterized antigenic site A, expanded antigenic site C, and new antigenic site G (confirmed in this study) indicated that their mutational patterns correlated well with the fluctuation of GII.4 variants in the human population, suggesting a major role in the emergence of variants (Fig. 3; see also Fig. S4). Correlation between the antigenic site mutational patterns and GII.4 variants was assessed using adjusted Rand index values, and antigenic site G was the one presenting the best correlation with the GII.4 variant circulation pattern while showing low sequence variation (Fig. 3b and c). Of note is that the mutational pattern determined for old antigenic site C did not correlate with the circulation of variants in nature (Fig. S5a and b), suggesting that our population-guided antigenic site characterization provided a better resolution of the antigenic profiling of GII.4 noroviruses. Neither previously defined putative epitope (motif) B (19) nor new variable motif H correlated with GII.4 variant circulation (Fig. 3c; see also Fig. S4). Motif B presented differences for Farmington Hills 2002 and Hunter 2004 strains, while motif H showed differences for Apeldoorn 2007, New Orleans 2009, and Sydney 2012 (Fig. S4). Mutations on antigenic sites D and E have been shown to alter both antigenicity and binding to HBGA carbohydrates (19, 24, 28) while showing modest correlation with GII.4 variant emergence (Fig. 3c). Thus, the evolutionary pressure correlating to antigenic sites D and E might be different from that correlating to sites involved only in the antigenic characteristics of the virus. Of note, the mutational pattern from expanded antigenic site D correlates much better than that from the original antigenic site D (Fig. S5a and b), and this improvement of correlation is consistent with the recent addition of residue 396 to this antigenic site (29). In contrast, the mutational pattern determined for expanded antigenic site E showed a lower level of correlation than that determined for the original antigenic site (Fig. S5a and b). This might have been due to the expanded site (inclusion of residue 414) showing variation within the variants Den Haag 2006b and Sydney 2012, making the profiles less consistent with the intervariant diversity. Since our data set represents sampling bias, i.e., over 50% of the sequences in the data set were collected during 2010 to 2016, the Shannon entropy analysis was reconducted using a randomly subsampled data set that included a maximum of 50 sequences per variant (Fig. S6a). This data set sensitively reflected the variation of viruses circulating before 2010 and showed 13 additional variable sites mapping outside the newly defined variable motifs/antigenic sites. Only four of them were located and clustered on the surface of the VP1; two of them (residue 250 and 504) were located together and might represent a site that differentiates early strains (Grimsby 1995 variant) from the others, and the remaining two (residues 300 and 329) mapped close to motif/antigenic site G (Fig. S6b). These two residues could be a part of antigenic site G; however, both represented minor variations over decades, suggesting a subtle impact on variant emergence (Fig. S6c). Residue 504 was recently shown to be a part of the GII.4 cross-protective epitope (30). Thus, despite the variation shown in Fig. S6c, the substitution (P504Q) might have a small impact on the antigenic diversification of GII.4 variants.

The mutational patterns determined for three motifs (A, C, and G) correlated well with the fluctuation of GII.4 variants (Fig. 3c). Motif/antigenic site A was previously confirmed to be a major antigenic site (23), while the expanded/new motifs (C and G)

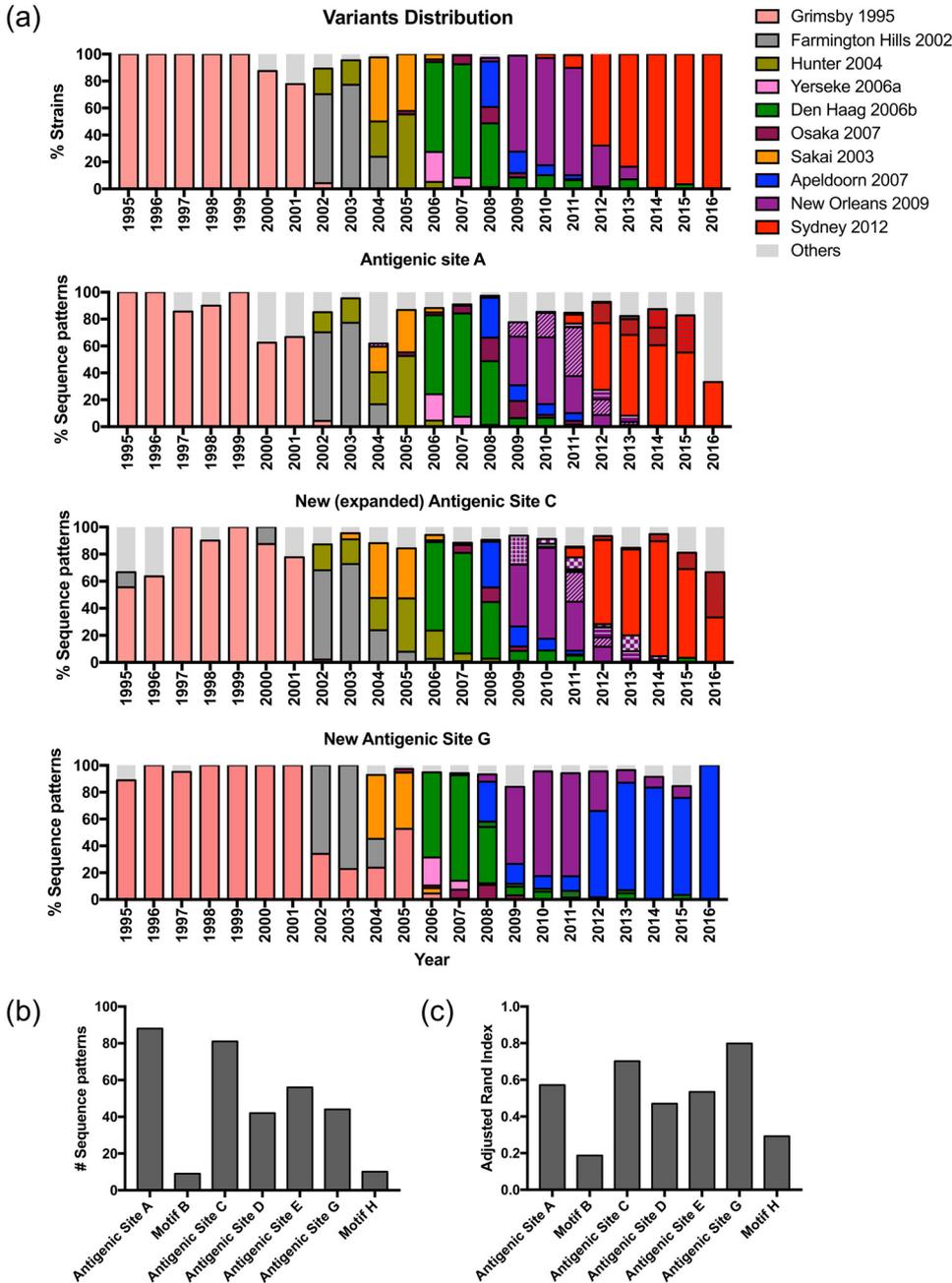


FIG 3 Predominant sequence patterns for proposed antigenic sites correlate with GII.4 variant circulation in the human population. (a) Amino acids from new and expanded motifs/antigenic sites were tabulated using 1,572 sequences of GII.4 norovirus that circulated from 1995 to 2016. GII.4 variant yearly distribution was tabulated using the same sequence database. Colors of the bars for the profiling graphs correspond to the predominant sequence pattern presented in that antigenic site for each GII.4 variant. Colors and variant assignment follow those described for Fig. 1. The patterns of the bars represent minor variations of the sequences in the motifs/antigenic sites. Motifs/antigenic sites A, C, and G appear to follow the pattern of GII.4 variant distribution over time, implying the potential role of these sites in the emergence of new pandemic GII.4 variants. (b and c) The number of sequence patterns of each antigenic site (b) and the correlation between variant classification and the mutational patterns of each motif/antigenic site (c) were calculated. The degree of correlation was assessed by adjusted Rand index values, in which a higher index values indicates a higher level of correlation between variant distribution and the mutational pattern of the motif/antigenic site.

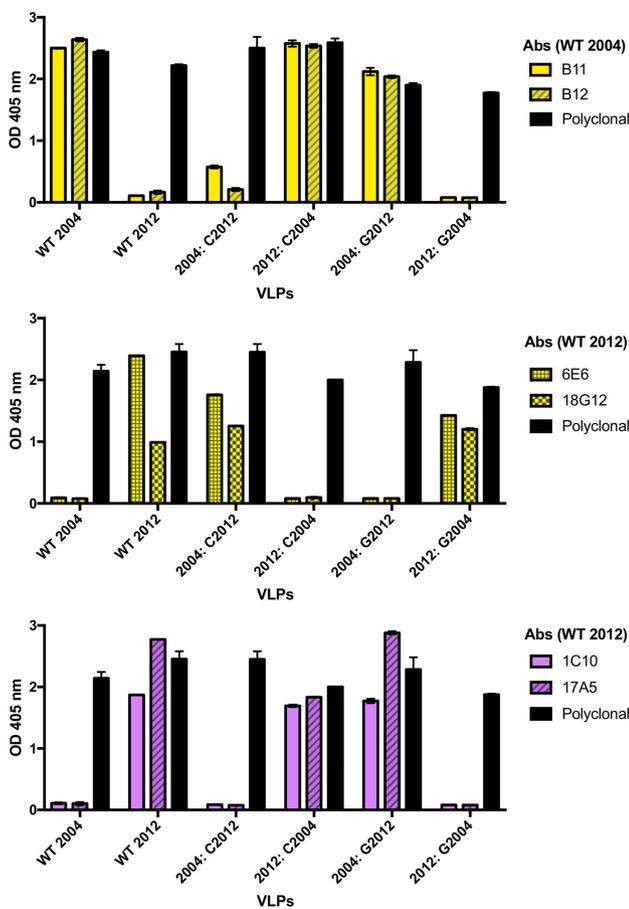
were not yet confirmed experimentally. To confirm the role of these two motifs in the antigenic makeup of the GII.4 capsid, we replaced residues of VP1 from a Farmington Hills 2002 variant (MD2004-3 strain detected in 2004, wild-type [WT] 2004; GenBank accession number [DQ658413](https://www.ncbi.nlm.nih.gov/nuccore/DQ658413)) with those of a Sydney 2012 variant (RockvilleD1 strain

(a)

Antigenic sites		A							C							G						
Residues		294	295	296	297	298	368	372	373	339	340	341	375	376	377	378	352	355	356	357	359	364
WT	WT 2004	A	D	T	H	N	N	N	N	R	G	D	F	E	T	G	S	D	V	H	T	S
VLPs	WT 2012	T	G	S	R	N	E	D	H	R	T	D	F	E	A	N	Y	S	A	D	A	R
	2004: A2012	T	G	S	R	N	E	D	H	R	G	D	F	E	T	G	S	D	V	H	T	S
	2004: C2012	A	D	T	H	N	N	N	N	R	T	D	F	E	A	N	S	D	V	H	T	S
Mutant	2004: G2012	A	D	T	H	N	N	N	N	R	G	D	F	E	T	G	Y	S	A	D	A	R
VLPs	2012: A2004	A	D	T	H	N	N	N	N	R	T	D	F	E	A	N	Y	S	A	D	A	R
	2012: C2004	T	G	S	R	N	E	D	H	R	G	D	F	E	T	G	Y	S	A	D	A	R
	2012: G2004	T	G	S	R	N	E	D	H	R	T	D	F	E	A	N	S	D	V	H	T	S

(b)

ELISA Binding of MAbs



(c)

MAb blocking of WT 2012 VLP

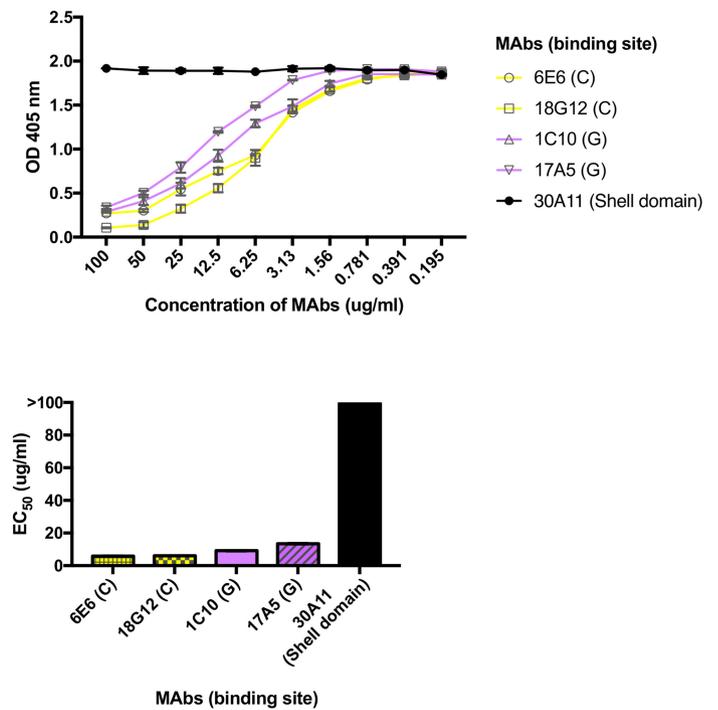


FIG 4 Mutational analyses on the major capsid protein (VP1) of GII.4 noroviruses confirmed newly proposed antigenic sites C and G. (a) Amino acid sequence information corresponding to antigenic sites A, C, and G from the MD2004-3 (wild-type [WT] 2004) strain, the RockvilleD1 (WT 2012) strain, and corresponding mutants. Swapped residues in the mutants are shown in red. (b) Immunoassay performed with virus-like particles (VLPs) and MAbs raised against the Farmington Hills 2002 strain (MD2004-3 [GenBank accession no. [DQ658413](#)]; WT 2004 [11]) (top) and the Sydney 2012 strain (RockvilleD1 [[KY424328](#)]; WT 2012) (middle for antigenic site C and bottom for antigenic site G). Means and standard errors (SE) were calculated from results from duplicate wells. (c) HBGA-blocking assays were performed using MAbs raised against a Sydney strain (WT 2012) and the WT 2012 VLP. The OD curves from each of the MAbs are shown on the top panel, and the corresponding half-maximal blocking values (EC_{50}) ($\mu\text{g/ml}$) are summarized in the bottom panel. Bars represent means and SE. The 30A11 MAb that targets the Shell domain was used as a negative control for blocking assay.

detected in 2012, WT 2012; accession number [KY424328](#)) and vice versa (Fig. 4a) and produced the corresponding VLPs. We tested the mutant and wild-type VLPs against guinea pig hyperimmune sera and mouse MAbs (two uncharacterized MAbs, namely, MAb B11 and MAb B12, raised against the WT 2004 [11] and four MAbs, namely, 1C10,

6E6, 17A5, and 18G12, newly developed against the WT 2012. We found that mutations at new motifs C and G abrogated binding of MAbs B11, B12, 6E6, and 18G12 and of MAbs 1C10 and 17A5, respectively (Fig. 4b). Reconstitution of motif C in the WT 2012 was achieved by reverting those sites to WT 2004 (indicated as “2012: C2004” VLPs in the upper panel of Fig. 4b). Of note, when mutations were introduced at residues 340 and 376, which were regarded as correlating to original antigenic site C, no differences in binding were observed with MAbs B11 and B12 (11). While changes at residues 377 and 378 reduced the levels of binding to those MAbs, an additional mutation at residue 340 was required for complete antigenic site depletion (Fig. S7). Likewise, reconstitution of motifs C and G in the WT 2004 was achieved by reverting those sites to the WT 2012 (indicated as “2004: C2012” and “2004: G2012” VLPs in the middle and bottom panels of Fig. 4b). Blockade potential (a surrogate of norovirus neutralization) was confirmed for all four MAbs using HBGA-blocking assays, except for MAb 30A11 that binds to the S domain of the VP1 (31) and was used as a control (Fig. 4c).

The impact of mutations on antigenic sites C and G with respect to the immune response was evaluated using HBGA-blocking assays. Mutants of antigenic site A were included for comparison. Polyclonal sera against WT 2004 VLP showed high blocking activity against homologous wild-type VLP (low half-maximal effective concentration blocking values [EC₅₀ values]) but reduced blocking against the WT 2012 VLP (high EC₅₀ value) (Fig. 5). Mutations on antigenic sites A, C, and G on those wild-type VLPs altered the blockade potential of polyclonal sera. As shown in previous studies (19, 23), transplanting of antigenic site A between WT 2004 and WT 2012 resulted in a large difference in the levels of blocking activity in experiments using the polyclonal sera against the WT 2004 (Fig. 5a and b, left panels). Mutations on antigenic sites C and G had a minor effect or no effect on the blocking activity of the sera against the WT 2004 (Fig. 5a and b, left panels). Interestingly, the blocking pattern was very different in experiments using the polyclonal sera raised against WT 2012 VLPs. Transplanting of antigenic sites A and G into WT 2004 VLPs (indicated as “2004: A2012” and “2004: G2012” in the right panels of Fig. 5a and b) resulted in blocking activity by sera raised against the WT 2012. Notably, polyclonal sera raised against WT 2012 VLPs did not lose blocking activity against antigenic site A and C mutant VLPs (2012: A 2004 and 2012: C2004) but lost blocking activity against antigenic site G mutant VLP (2012: G2004) (Fig. 5a and b, right panels). In summary, both antigenic sites (A and G) showed distinctive roles as blockade sites; notably, antigenic site G played a role equal to or greater than that played by major antigenic site A in the antigenicity of the WT 2012 strain (Fig. 5), suggesting the potential of both as protective antigenic sites.

Intravariant evolution is driven by stochastic processes. In contrast to the accumulation of nucleotides and amino acids detected at the intervariant level (Fig. 1b and c), there was limited accumulation of nucleotide substitutions (data not shown) and amino acid substitutions (Fig. 6a) within variants (intravariant level). Despite this limited accumulation of substitutions, all of the pandemic variants presented diversity in their sequences (average of 4.8 amino acid substitutions). Interestingly, this diversity was detected at most of the major antigenic sites and variants (Fig. 6b; see also Fig. S8). Thus, while each variant presented a major amino acid combination for each antigenic site, most of the GII.4 variants presented other minor amino acid patterns on those antigenic sites. Moreover, the analysis of intravariant diversification showed that their evolution was stochastic in time and location (Fig. 7), in contrast to the temporally clustered intervariant evolution of GII.4 strains (Fig. 1). While some variants (e.g., Hunter 2004, Den Haag 2006b, New Orleans 2009, and Sydney 2012) presented diversity in their major antigenic sites after 3 to 4 years of circulation, which might suggest that immune pressure acts at the intravariant level (Fig. 6), the numbers of strains (and sequences) are limited and do not represent dominant strains. Two other important observations occurred while analyzing the mutational pattern of each of the antigenic sites at the intravariant level: (i) major differences at the amino acid sequence level were detected in early strains for New Orleans 2009 and Sydney 2012 variants (Fig. 6;

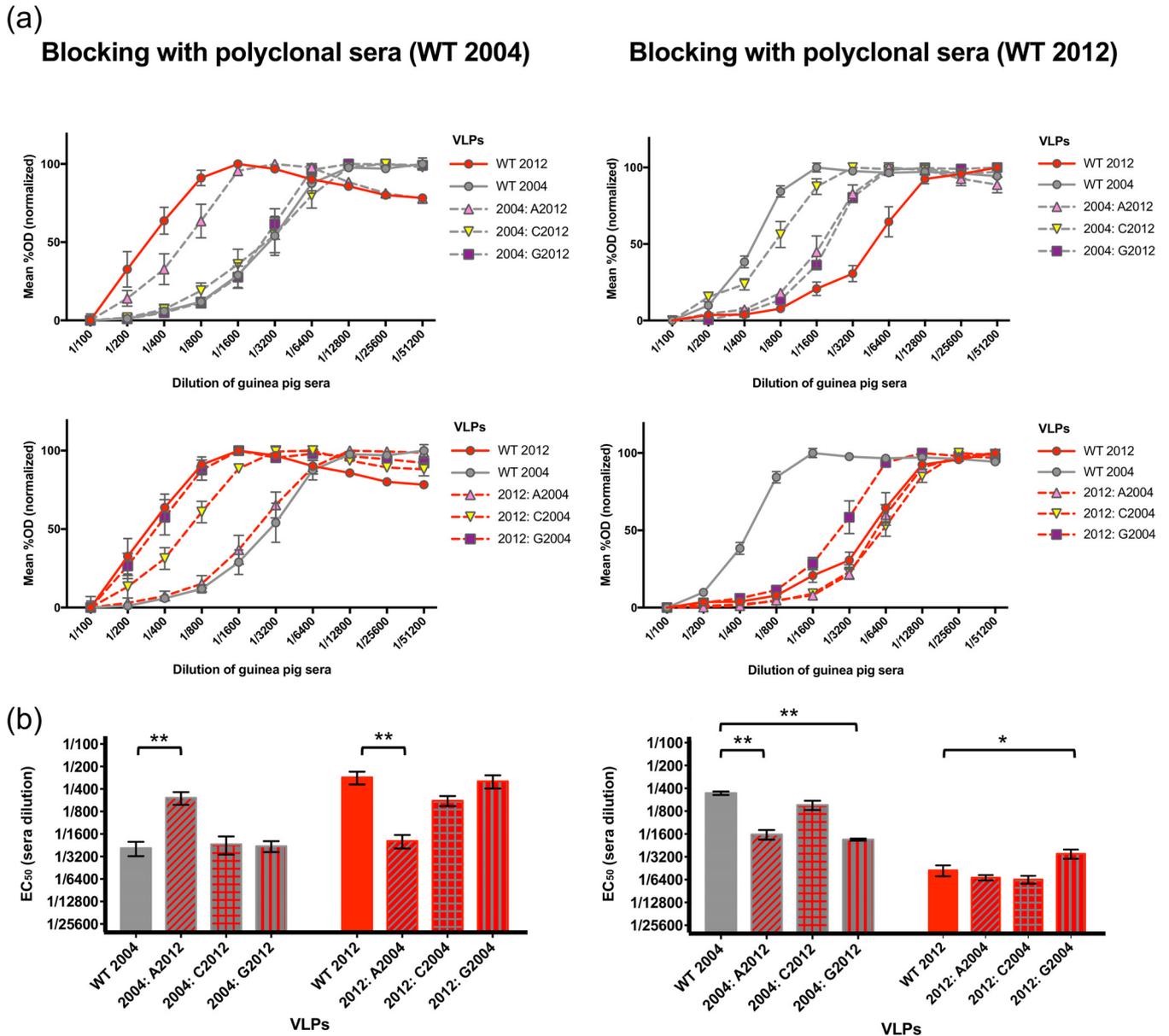


FIG 5 Role of new antigenic sites C and G in overall blocking activity of polyclonal sera. (a) HBGA-blocking assays were performed with hyperimmune (polyclonal) sera raised against a Farmington Hills strain (MD2004-3; WT 2004) and a Sydney strain (RockvilleD1; WT 2012) and wild-type and mutant VLPs. Experiments were performed using sera from guinea pigs for each wild-type strain. The means and standard errors (SE) were calculated for duplicate wells from two guinea pigs. The OD curves from blocking assays were grouped by sera after normalizing the data from each of the VLPs. (b) The half-maximal blocking values (EC₅₀) (serum dilution) corresponding to each of the wild-type strains and the mutant VLPs were calculated from the normalized OD values from the blocking assays. Bars represent means and SE. *, $P < 0.05$; **, $P < 0.01$ (from one-way ANOVA with *post hoc* Dunnett's multiple-comparison test).

see also Fig. S8), and those early strains did not represent the major amino acid combination for any of the four major antigenic sites (A, C, D, and E; Fig. S8); (ii) none of the preceding strains evolved toward (or presented) the amino acid motif seen in the later strains. This, together with the phylogenetic analyses, suggests that each pandemic variant presented a different origin and did not follow a trunk-like linear evolution such as that seen in H3N2 influenza viruses (32).

Differences in the evolutionary patterns of the intervariants and intravariants were also confirmed by Bayesian Markov chain Monte Carlo (MCMC) analysis. Substitution rates of seven variants with >50 sequences were calculated, and the results are summarized in Table 1. The rate of (overall) intervariant GII.4 strains was reported previously elsewhere (15). Intravariant substitution rates ranged from 1.57×10^{-3} to

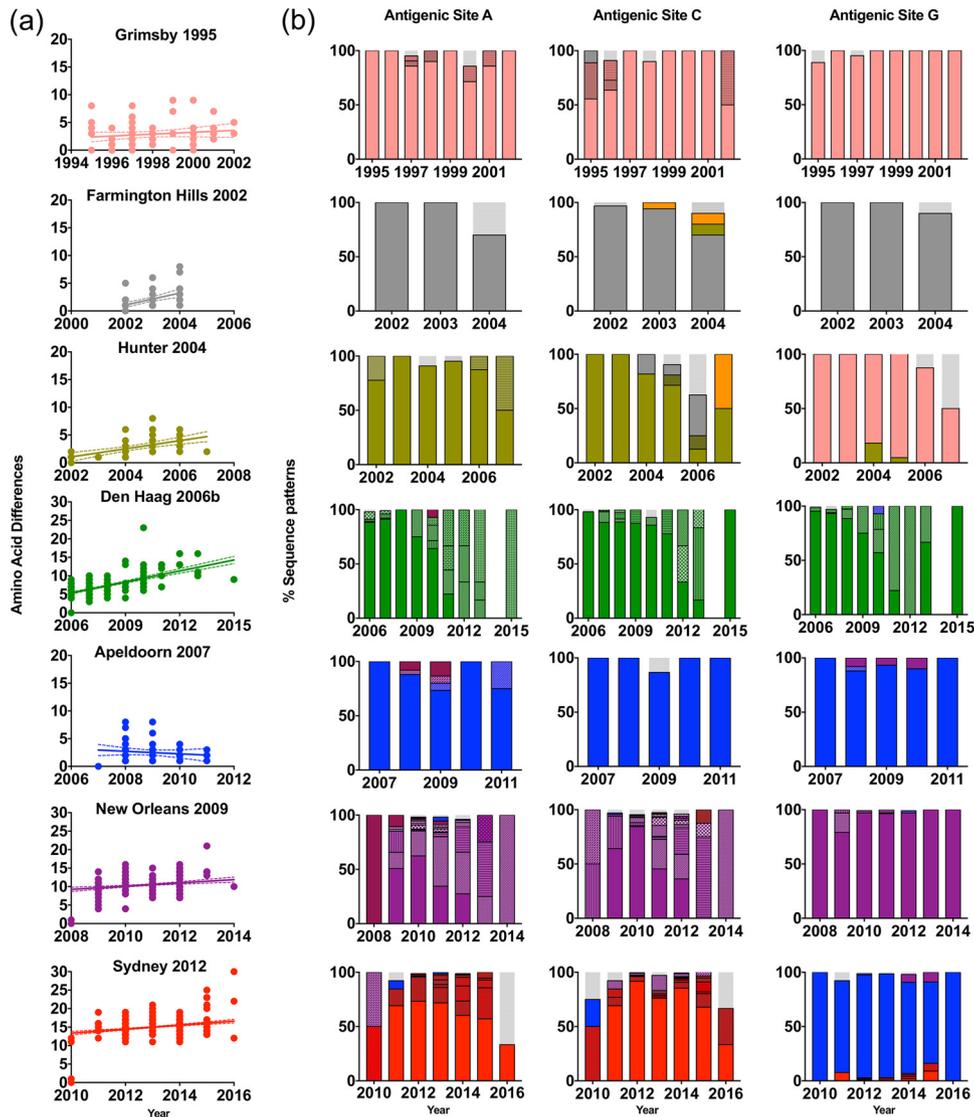
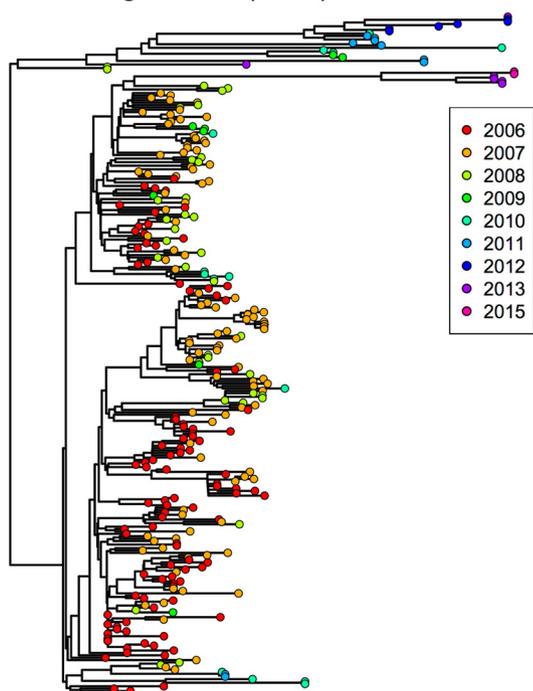


FIG 6 Intravariant diversity of GII.4 noroviruses reveals minimal accumulation of mutations. (a) Amino acid pairwise differences among viruses from each of the pandemic GII.4 variants compared to the earliest strain of each given variant. (b) The intravariant mutational pattern of each antigenic site for each pandemic GII.4 variant was calculated as described for Fig. 3.

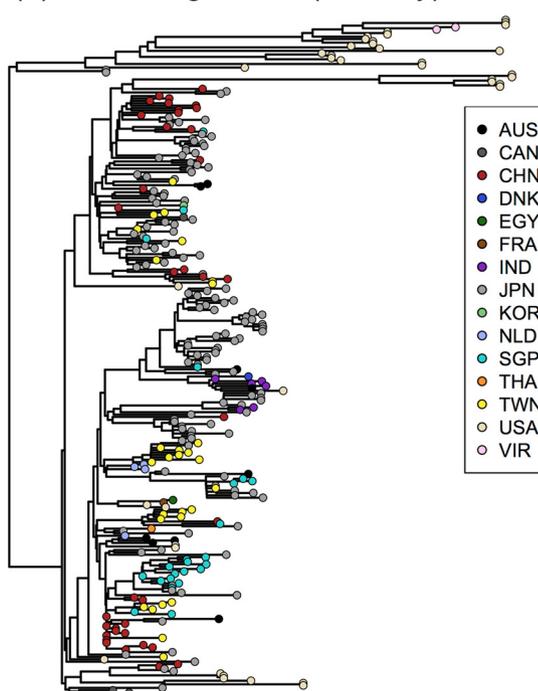
4.64×10^{-3} substitutions/site/year, all of them lower than the intervariant (overall) substitution rate (5.4×10^{-3} substitutions/site/year) (15). Among the variants, the Sydney 2012 variants had the higher substitution rate (Table 1).

Diversifying pressure drives emergence of pandemic GII.4 noroviruses. Because different factors seemed to drive the evolution of the intervariants (overall) and intravariants, we performed selection analysis using the *mixed-effect model of evolution* (MEME) method and looked for evidence of site-by-site episodic diversifying pressure on the VP1 along the branches of its evolutionary tree. To analyze the overall evolutionary process, we randomly subsampled sequences from the original data set that included a maximum of 30 strains per variant. During the overall evolution, we found 9 positively selected sites ($P < 0.05$ and empirical Bayes factor > 100) on the P2 subdomain on the surface (codon sites 327, 335, 352, 355, 357, 366, 368, 375, and 378) distributed on 21 branches of the phylogenetic tree (Fig. 8a). Branches connecting discrete GII.4 variants presented sites undergoing episodic diversification (codon sites 352, 355, 357, 368, and 378) that mapped on the antigenic sites. The mutational pattern

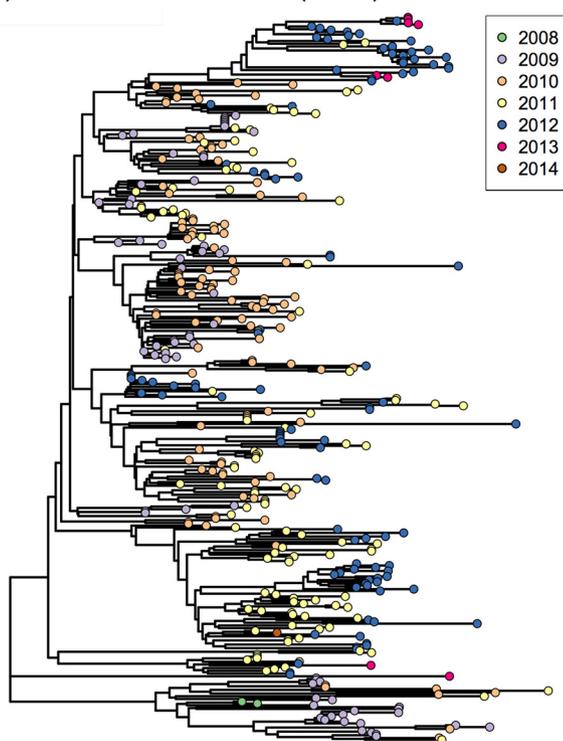
(a) Den Haag 2006b (Year)



(b) Den Haag 2006b (Country)



(c) New Orleans 2009 (Year)



(d) New Orleans 2009 (Country)

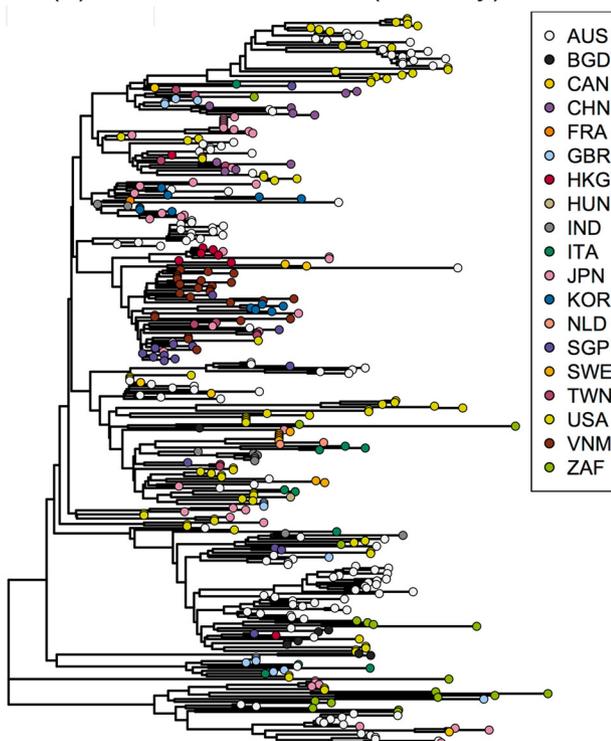


FIG 7 Intravariant diversification of GII.4 noroviruses is governed by stochastic events. Maximum likelihood trees of two major GII.4 variants, Den Haag 2006b (a and b) and New Orleans 2009 (c and d), show the collection year (a and c) or country (b and d) for each of their strains. Phylogenetic clustering of the strains did not present any pattern, indicating randomness of the intravariant evolution in time and space.

of these sites correlated well with GII.4 variant emergence, with higher adjusted Rand index values than any of the antigenic sites (Fig. 8b and c). This suggests that residues on the antigenic sites experienced episodic diversifying pressure during intervariant evolution. Analyzing each variant only, most of the diversifying pressure was found to

virus to evade the immunity acquired to previously circulating GII.4 variants, a process similar to that seen with H3N2 influenza viruses (26, 28, 32). Since the mid-1990s, over 10 different variants have been reported, with 6 of them associated with large outbreaks worldwide. The overall evolutionary pattern of GII.4 viruses presents a strong linear accumulation of amino acid substitutions during intervariant diversification (15), with most substitutions occurring in the P2 subdomain (28). Antigenic differences among variants have been largely attributed to highly variable residues that map on the surface of the P domain, leading to the identification of five (A to E) motifs that are part of GII.4-specific antigenic sites (18–20). While the binding site was characterized for different GII.4-specific MAbs, the same studies reported numerous GII.4-specific MAbs whose binding sites have not been determined (11, 12, 19, 23–25). Applying a population genomics approach, we found new (or expanded) motifs on the surface of the capsid that presented mutational patterns that correlated with the circulation of GII.4 variants. The role of antigenicity of two of those motifs (antigenic sites C and G) was confirmed with (previously) uncharacterized HBGA-blockade MAbs (B11 and B12), newly developed HBGA-blockade MAbs (1C10, 6E6, 17A5, and 18G12), and polyclonal sera from guinea pigs immunized with wild-type VLPs. Antigenic site G presented the strongest correlation with the emergence and circulation of new GII.4 variants, while being more highly conserved than other large antigenic sites (e.g., site A, C, or E). This indicates that newly discovered antigenic site G plays a pivotal role in transforming the GII.4 noroviruses into new variants with pandemic potential. Antigenic site C is close to previously defined antigenic site A, but substitutions on antigenic site A did not affect the binding of those MAbs (11). Notably, competition analyses among different MAbs showed that MAbs B11 and B12 partially blocked interactions with MAbs mapping to antigenic site A (11), suggesting that this new motif is part of an antigenic site that involves two or more epitopes (at least antigenic sites A and C). Recently, Koromyslova and colleagues showed that the footprint of an antibody that neutralized human noroviruses mapped to residues from antigenic sites C and D (22). This demonstrates that the same epitope could be shared by the different antigenic sites and that their interaction could result in differences in the evolutionary patterns presented. In addition, differences in HBGA-blocking ability between antigenic sites (A, C, and G) and between variants (Farmington Hills 2002 and Sydney 2012) further highlight the complexity of the antigenic topology of GII.4 noroviruses.

Large-scale analysis reduces biases in determining the role of individual amino acid substitutions in the emergence of the new GII.4 variants. Recently, it was suggested that mutations in the capsid protein of Sydney strains circulating in 2015 resulted in viruses that were antigenically different from those circulating in 2012 (34). Our large-scale intravariant analyses show that (i) the strain selected as Sydney 2012 for the antigenic study (34) was not representative of the predominant virus in terms of the sequences of the antigenic sites and that (ii) strains with antigenic site sequences similar to those regarded as “GII.4 2015” by Lindesmith et al. (34) have been circulating since 2010 in the human population as part of the overall population of the Sydney 2012 variant (see Fig. S8 in the supplemental material). While there is no indication of the replacement of antigenically distinct predominant strains within the Sydney variants, since 2015, the Sydney 2012 variants have been detected with a different RNA-dependent RNA polymerase (GII.P16), which might have influenced the transmissibility and spread of this virus (35, 36). Similarly, the NERK motif, which was suggested to occlude conserved GII.4 antigenic sites and affect the antibody blockade potency (37, 38), was shown to be highly conserved among GII.4 variants. Previous studies pointed out a mutation in this motif in a Sydney 2012 variant; however, our large-scale analysis showed that the New Orleans variant was the only one showing—at the population level—any mutation at this motif (Fig. S10). Thus, the profiling of mutational patterns that we implemented, which included the use of a large number of sequences, could provide a better understanding of the role of individual mutations in the circulation and predominance of the pandemic GII.4 variants. Immunological analyses that include multiple different viruses from each of the pandemic variants and in-depth population analyses are

needed to better delineate the meaning of minor mutations with respect to the antigenic differences among the variants.

Improvements in the understanding of viral dynamics and of the correlation between antigenic and genetic changes in influenza viruses have facilitated the selection of virus strains to be included in upcoming seasonal vaccines (39–41). To better understand the viral dynamics of GII.4 noroviruses, we performed selection analyses that included all variants reported for over 4 decades. Episodic diversifying (positive) selection was observed in five residues (352, 355, 357, 368, and 378) during intervariant evolution (Fig. 7a); these residues are part of antigenic sites A, C, and G. Three of these residues (352, 357, and 378) showed positive selection on major branches of the GII.4 tree, indicating a role in the emergence of new pandemic noroviruses. In these three sites, strains that emerged and predominated from 1995 to 2006 (Grimsby 1995, Farmington Hills 2002, Hunter 2004, and Yerseke 2006a) presented the motif SHG; the strain Den Haag 2006b that emerged and predominated from 2006 to 2009 presented the motif YPH; the strains that have predominated since 2009 (Apeldoorn 2007, New Orleans 2009, and Sydney 2012) presented the motif YDN. Notably, while these three residues have been understudied pertaining to the evolution of GII.4 noroviruses (26, 28, 42), they seem to play a major role in the antigenic topology and emergence of new GII.4 variants. Of note, residue 368 presented episodic diversification that distinguished the Sydney 2012 strains from Apeldoorn 2007 and New Orleans 2009. Changes on this residue were shown to be involved in the antigenic diversification of the Sydney 2012 strain (43), and our mutational analyses indicated a pivotal role of sites A and G in the antigenic differences, thus supporting our *in silico* observation. Initial studies suggested that most of the positively selected sites for GII.4 noroviruses were located in the S domain (28, 42, 44); however, our analyses, with an expanded number of sequences and variants, showed that the emergence of new variants stemmed from the positive selection of residues mapping on the surface of the P domain. Pinpointing of the changes required for emergence of a new antigenic variant of seasonal influenza virus or emergence of a pandemic virus is the “holy grail” for controlling viruses undergoing constant change. While prediction of viral emergence requires a holistic approach that includes studies of the virus, the environment, and the host (45–47), careful monitoring of the substitutions in residues involved in the diversification and emergence of new GII.4 viruses could help in the early detection of future novel variants with pandemic potential.

The GII.4 intervariant diversification, likely driven by the immune status of the population, is correlated with the accumulation of amino acid substitutions in the major capsid protein. In contrast, our analyses suggest that diversification at the intravariant level is much restricted, with amino acid substitutions occurring without indications of diversifying selection. Minimal substitutions of the amino acids that mapped on the major antigenic sites were observed over the predominance of each variant, and most seemed to follow a stochastic process. The latter could represent a result of multiple pressures exerted on the virus, including but not limited to individual virus-host interactions and dispersion. For two variants (New Orleans 2009 and Sydney 2012), a larger, more comprehensive data set was available, and prepandemic strains have been reported previously (48–51). Interestingly, although these prepandemic strains cluster within their respective variants, they present multiple differences (mostly mapping at the antigenic sites) from the virus population that later established worldwide dominance. Together, these results suggest that emergence of GII.4 variants could occur in the following sequence: (i) a prepandemic stage characterized by acquisition of mutations that facilitate viral emergence and episodic diversification (exemplified here by residues 352, 357, 368, and 378); (ii) a short period (1 to 2 years) of adaptability (with different antigenic motifs) that precedes the pandemic phase; and, finally, (iii) the pandemic phase, where the virus is dominant and explores only a narrow space of sequence diversity. A similar pattern has been observed for H3N2 influenza viruses (52) and rotaviruses (53), in which viruses that circulate at low levels are able to predominate in the following season without major changes in their genetic background. This

order of events might have occurred during the recent emergence and predominance of GII.17 noroviruses in many Asian countries. During 2013 to 2014, a “new” GII.17 norovirus (variant C) was detected circulating in different countries (Japan, Hong Kong, and China), but in the next epidemic season, this variant was ultimately replaced by variant D, which spread worldwide and predominated from 2014 to 2016 (15, 54–56). In the same context, we found six GII.4 strains that did not cluster with any GII.4 variant (Fig. S1) and that might represent strains that did not adapt well to the human population. These strains presented different sequence patterns in their antigenic sites and therefore might constitute strains in the prepandemic stage that explored the sequence space but failed to thrive sufficiently to reach the next evolutionary level.

Our findings suggest two different mechanisms behind the evolutionary dynamics of the major capsid protein from GII.4 noroviruses: (i) a steady pandemic phase governed by stochastic processes, preceded by (ii) substitutions that arise from positive selection. This report also provides a methodological framework that could facilitate the characterization of the variable antigenic sites that play a relevant role in the emergence of new viruses in the human population. Studies that include large and long-time-scale data sets of full-length genomes would help to determine the factors involved in norovirus predominance and persistence in the human population.

MATERIALS AND METHODS

Data mining and sequence analyses. A total of 1,601 full-length (1,623-nt) and nearly full-length ($\geq 1,560$ nt) VP1 sequences of the GII.4 genotype, from which sequences from immunocompromised patients and environmental samples were removed, were downloaded from GenBank (accessed July 2017) (data available upon request). The sequences spanned the 42 years from 1974 to 2016. Sequences were aligned using ClustalW as implemented in MEGA v7 (57) and were visually inspected to confirm proper alignment. The nomenclature used for the GII.4 variants was adopted as previously indicated (15), and variant data sets were parsed following such information. Sequence analyses were performed using the total data set except where indicated. Entropy analysis and profiling of mutational patterns included only strains (1,572 VP1 sequences) collected from 1995 to 2016, as Grimsby-like viruses were the first recorded to cause large outbreaks worldwide (33), and included the following 11 variants in approximate order of emergence: Grimsby 1995 (or US95_96), Farmington Hills 2002, Lanzhou 2002, Sakai 2003 (or Asia 2003), Hunter 2004, Yerseke 2006a, Den Haag 2006b, Osaka 2007, Apeldoorn 2007, New Orleans 2009, and Sydney 2012. Intravariant and intervariant Shannon entropy values were calculated using the Shannon Entropy-One tool as implemented in Los Alamos National Laboratory (www.hiv.lanl.gov). Entropy values for each position were plotted in GraphPad Prism v7. The structural model of the GII.4 norovirus P domain dimer (Protein Data Bank [PDB] accession number 2OBS) was rendered using UCSF Chimera (version 1.11.2) (58). Profiling of mutational patterns of motifs/antigenic sites was performed using R 3.4.2 (59). Amino acid sequences of each motif/antigenic site were profiled by year, and the corresponding mutational patterns were plotted in a composite bar graph that showed the number of strains with each pattern as a fraction (percentage) of a whole (total number of strains). The correlation of the mutational patterns and the variant distribution was assessed using adjusted Rand index values, known as a clustering analysis method, which evaluated the degree of the matches between mutational patterns and variant classification. Adjusted Rand index values were calculated using R and the *mclust* package (60). To account for the sampling bias associated with the original data set, we repeated the entropy and structural analyses using a randomly subsampled data set that included a maximum of 50 strains/variant ($n = 474$) from the original GII.4 data set (1,572 sequences).

Diversifying selection analysis. Maximum likelihood (ML) phylogenetic trees of VP1-encoding nucleotide sequences for all variants (intervariant analyses) and for each variant (intravariant analyses) were constructed using PhyML (61). The best substitution models were selected based on the lowest corrected Akaike information criterion (AICc) value for each data set using jModelTest v2 (62, 63). Analyses of larger data sets (i.e., the Sydney 2012 and New Orleans 2009 variants) tended to favor a generalized time-reversible (GTR) substitution model, while analyses of variants with fewer reported sequences favored a Tamura-Nei 93 (TN93) model. Diversifying selection analyses of the VP1-encoding sequence through its ML phylogenetic tree were performed by using MEME methods (64, 65). We aimed to detect codon sites subjected to positive selection (i.e., with more nonsynonymous substitutions than synonymous substitutions) during their evolution and focused on the sites at or near the major antigenic sites of GII.4. Significant positive selection was indicated by *P* values of < 0.05 . The branches that were subjected to diversifying positive selection were explored by using empirical Bayes factors of > 100 in MEME. To reduce the data size for the intervariant analyses of positive selection, we randomly subsampled a maximum of 30 strains/variant ($n = 308$) from the original GII.4 data set (1,601 sequences) as indicated previously.

Bayesian analyses of nucleotide substitution rates. Using the VP1 sequences and data corresponding to the respective collection years from each strain, temporal phylogenetic analysis was performed using Bayesian Markov chain Monte Carlo (MCMC) methodology in BEAST v1.8.3 (66). The best substitution models were selected based on the lowest corrected Akaike information criterion (AICc)

value as mentioned above. The clock models (strict or relaxed lognormal clock) and tree priors (constant population size, exponential growth, or skyline) were tested, and the best models were selected based on the model selection procedure using AIC through MCMC. The MCMC runs were performed until all the parameters reached convergence. MCMC runs were analyzed using Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>). The initial 10% of the logs from the MCMC run was removed before summarizing the mean and the 95% highest posterior density interval of the substitution rates.

Site-directed mutagenesis and VLP production. The VP1-encoding sequences from a GII.4 Farmington Hills 2002 (MD2004-3) strain and a Sydney 2012 (RockvilleD1) strain were ligated into pFastBac1 vectors using Sall and NotI restriction sites (11, 67). Site-directed mutagenesis of pFastBac-MD2004-3 and pFastBac-RockvilleD1 was performed using mutation-specific forward and reverse primers, followed by purification with illustra MicroSpin G-50 columns (GE Healthcare, Buckinghamshire, United Kingdom). Parental DNA was digested using the DpnI enzyme (New England Biolabs, MA, USA). VLPs presenting multiple mutations were developed by cloning a chemically synthesized P domain into a pFastBac1 plasmid containing the S domain by the use of PspXI and NotI restriction sites (11). Each pFastBac construct was transformed via electroporation into ElectroMAX DH10B cells (Thermo Fisher Scientific, CA, USA) and grown on LB plates with ampicillin overnight at 37°C. Selected colonies were used to extract plasmid DNA (QIAprep Spin Miniprep kit; Qiagen, Hilden, Germany). Introductions of mutations were confirmed by Sanger sequencing. VLPs were produced using a Bac-to-Bac baculovirus expression system (Invitrogen, CA, USA) and purified through a cesium chloride gradient as previously described (67). Expression of VP1 protein was confirmed by Western blotting, and VLP integrity was confirmed by electron microscopy.

Immunoassays. Mutants and wild-type norovirus VLPs were analyzed for reactivity to monoclonal antibodies by enzyme-linked immunosorbent assay (ELISA) as described previously (11). The B11 and B12 MAbs were obtained from mice immunized with GII.4 Farmington Hills 2002 variant (MD2004-3 strain) VLP (11) and were generously provided by Kim Y. Green (National Institutes of Health, USA). The GII.4 Sydney 2012 variant-specific MAbs (1C10, 6E6, 17A5, and 18G12) were developed from mice immunized with RockvilleD1 strain VLP (GenScript, NJ, USA). HBGA-blocking assays were performed using mutant and wild-type VLPs, HBGA molecules derived from human saliva, the Sydney 2012 variant-specific MAbs, and polyclonal antibodies. The polyclonal antibodies were obtained from guinea pigs and mice immunized with GII.4 Farmington Hills 2002 (MD2004-3) and Sydney 2012 (RockvilleD1 strain) VLPs (11, 67). Human saliva was collected from a healthy adult volunteer. The saliva sample was boiled at 100°C for 10 min immediately after collection and centrifuged at 13,000 rpm for 5 min. The saliva supernatant was collected and used for HBGA-binding and HBGA-blocking assays. Briefly, serial dilutions of mouse MAb or guinea pig polyclonal sera were mixed with wild-type or mutant VLPs and incubated on the saliva-coated plate for 1 h at 37°C. Plates were washed four times to remove unbound (i.e., blocked by mouse MAbs or guinea pig polyclonal sera) VLPs. Pooled sera from guinea pigs or mice immunized with Farmington Hills 2002 VLP and Sydney 2012 VLP were used to detect the VLPs attached to the plate. Goat anti-guinea pig or anti-mouse IgG conjugated with horseradish peroxidase and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) substrate (SeraCare, MA, USA) was used to develop the blue-green color on the VLP-attached plate. The degree of blocking was evaluated using optical density (OD) at 405 nm and EC_{50} calculated for the serum dilutions. The EC_{50} was calculated from the normalized OD curve using GraphPad Prism v7. One-way analysis of variance (ANOVA) and Dunnett's multiple-comparison test were conducted to analyze the differences in EC_{50} values between wild-type and mutant VLPs using GraphPad Prism v7. Differences in EC_{50} values corresponding to P values of <0.05 were considered statistically significant.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.02202-19>.

FIG S1, TIF file, 1.1 MB.

FIG S2, TIF file, 2.1 MB.

FIG S3, TIF file, 2.5 MB.

FIG S4, TIF file, 2.2 MB.

FIG S5, TIF file, 2.1 MB.

FIG S6, TIF file, 2.7 MB.

FIG S7, TIF file, 0.5 MB.

FIG S8, TIF file, 2.6 MB.

FIG S9, TIF file, 0.7 MB.

FIG S10, TIF file, 1.1 MB.

ACKNOWLEDGMENTS

We thank Steve Rubin for the critical reading of the manuscript.

Financial support for this work was provided by Food and Drug Administration intramural funds (Program Number Z01 BK 04012-01 LHV to G.I.P.). K.T. and C.J.L. are recipients of a CBER/FDA-sponsored Oak Ridge Institute for Science and Education (ORISE) fellowship.

REFERENCES

- Ahmed SM, Lopman BA, Levy K. 2013. A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS One* 8:e75922. <https://doi.org/10.1371/journal.pone.0075922>.
- Payne DC, Vinje 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. <https://doi.org/10.1056/NEJMsa1206589>.
- Koo HL, Neill FH, Estes MK, Munoz FM, Cameron A, DuPont HL, Atmar RL. 2013. Noroviruses: the most common pediatric viral enteric pathogen at a large university hospital after introduction of rotavirus vaccination. *J Pediatric Infect Dis Soc* 2:57–60. <https://doi.org/10.1093/jpids/pis070>.
- Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleeschauwer B, Dopfer D, Fazil A, Fischer-Walker CL, Hald T, Hall AJ, Keddy KH, Lake RJ, Lanata CF, Torgerson PR, Havelaar AH, Angulo FJ. 2015. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Med* 12:e1001921. <https://doi.org/10.1371/journal.pmed.1001921>.
- Patel MM, Widdowson MA, Glass RI, Akazawa K, Vinje J, Parashar UD. 2008. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg Infect Dis* 14:1224–1231. <https://doi.org/10.3201/eid1408.071114>.
- Pires SM, Fischer-Walker CL, Lanata CF, Devleeschauwer B, Hall AJ, Kirk MD, Duarte AS, Black RE, Angulo FJ. 2015. Aetiology-specific estimates of the global and regional incidence and mortality of diarrhoeal diseases commonly transmitted through food. *PLoS One* 10:e0142927. <https://doi.org/10.1371/journal.pone.0142927>.
- Green KY. 2013. Caliciviridae: the noroviruses, p 582–608. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), *Fields virology*, 6th ed. Lippincott, Williams & Wilkins, Philadelphia, PA.
- 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. <https://doi.org/10.1126/science.286.5438.287>.
- Reeck A, Kavanagh O, Estes MK, Opekun AR, Gilger MA, Graham DY, Atmar RL. 2010. Serological correlate of protection against norovirus-induced gastroenteritis. *J Infect Dis* 202:1212–1218. <https://doi.org/10.1086/656364>.
- Czako R, Atmar RL, Opekun AR, Gilger MA, Graham DY, Estes MK. 2012. Serum hemagglutination inhibition activity correlates with protection from gastroenteritis in persons infected with Norwalk virus. *Clin Vaccine Immunol* 19:284–287. <https://doi.org/10.1128/CLVI.05592-11>.
- Parra GI, Abente EJ, Sandoval-Jaime C, Sosnovtsev SV, Bok K, Green KY. 2012. Multiple antigenic sites are involved in blocking the interaction of GII.4 norovirus capsid with ABH histo-blood group antigens. *J Virol* 86:7414–7426. <https://doi.org/10.1128/JVI.06729-11>.
- Alvarado G, Ettayebi K, Atmar RL, Bombardi RG, Kose N, Estes MK, Crowe JE, Jr. 2018. Human monoclonal antibodies that neutralize pandemic GII.4 noroviruses. *Gastroenterology* 155:1898. <https://doi.org/10.1053/j.gastro.2018.08.039>.
- 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>.
- Hoa Tran TN, Trainor E, Nakagomi T, Cunliffe NA, Nakagomi O. 2013. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII.4 variants. *J Clin Virol* 56:185–193. <https://doi.org/10.1016/j.jcv.2012.11.011>.
- Parra GI, Squires RB, Karangwa CK, Johnson JA, Lepore CJ, Sosnovtsev SV, Green KY. 2017. Static and evolving norovirus genotypes: implications for epidemiology and immunity. *PLoS Pathog* 13:e1006136. <https://doi.org/10.1371/journal.ppat.1006136>.
- Siebenga JJ, Vennema H, Zheng DP, Vinje J, Lee BE, Pang XL, Ho EC, Lim W, Choudekar A, Broor S, Halperin T, Rasool NB, Hewitt J, Greening GE, Jin M, Duan ZJ, Lucero Y, O’Ryan M, Hoehne M, Schreier E, Ratcliff RM, White PA, Iritani N, Reuter G, Koopmans M. 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>.
- van Beek J, Ambert-Balay K, Botteldoorn N, Eden JS, Fonager J, Hewitt J, Iritani N, Kroneman A, Vennema H, Vinje J, White PA, Koopmans M, NoroNet. 2013. Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro Surveill* 18:8–9.
- Debbink K, Lindesmith LC, Donaldson EF, Baric RS. 2012. Norovirus immunity and the great escape. *PLoS Pathog* 8:e1002921. <https://doi.org/10.1371/journal.ppat.1002921>.
- 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. <https://doi.org/10.1371/journal.ppat.1002705>.
- Garaicoechea L, Aguilar A, Parra GI, Bok M, Sosnovtsev SV, Canziani G, Green KY, Bok K, Parreno V. 2015. Llama nanoantibodies with therapeutic potential against human norovirus diarrhea. *PLoS One* 10:e0133665. <https://doi.org/10.1371/journal.pone.0133665>.
- Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, Neill FH, Blutt SE, Zeng XL, Qu L, Kou B, Opekun AR, Burrin D, Graham DY, Ramani S, Atmar RL, Estes MK. 2016. Replication of human noroviruses in stem cell-derived human enteroids. *Science* 353:1387–1393. <https://doi.org/10.1126/science.aaf5211>.
- Koromyislova AD, Morozov VA, Hefele L, Hansman GS. 2018. Human norovirus neutralized by a monoclonal antibody targeting the HBGA pocket. *J Virol* 93:e02174-18. <https://doi.org/10.1128/JVI.02174-18>.
- 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>.
- Lindesmith LC, Debbink K, Swanstrom J, Vinje J, Costantini V, Baric RS, Donaldson EF. 2012. Monoclonal antibody-based antigenic mapping of norovirus GII.4–2002. *J Virol* 86:873–883. <https://doi.org/10.1128/JVI.06200-11>.
- Lindesmith LC, Costantini V, Swanstrom J, Debbink K, Donaldson EF, Vinje J, Baric RS. 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>.
- Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, Koopmans M. 2007. Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J Virol* 81:9932–9941. <https://doi.org/10.1128/JVI.00674-07>.
- Lindesmith LC, Mallory ML, Debbink K, Donaldson EF, Brewer-Jensen PD, Swann EW, Sheahan TP, Graham RL, Beltramello M, Corti D, Lanzavecchia A, Baric RS. 2018. Conformational occlusion of blockade antibody epitopes, a novel mechanism of GII.4 human norovirus immune evasion. *mSphere* 3:e00518-17. <https://doi.org/10.1128/mSphere.00518-17>.
- 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. <https://doi.org/10.1371/journal.pmed.0050031>.
- Lindesmith LC, Brewer-Jensen PD, Mallory ML, Yount B, Collins MH, Debbink K, Graham RL, Baric RS. 2018. Human norovirus epitope D plasticity allows escape from antibody immunity without loss of capacity for binding cellular ligands. *J Virol* 93:e01813-18. <https://doi.org/10.1128/JVI.01813-18>.
- Lindesmith LC, McDaniel JR, Changela A, Verardi R, Kerr SA, Costantini V, Brewer-Jensen PD, Mallory ML, Voss WN, Boutz DR, Blazeck JJ, Ippolito GC, Vinje J, Kwong PD, Georgiou G, Baric RS. 2019. Sera antibody repertoire analyses reveal mechanisms of broad and pandemic strain neutralizing responses after human norovirus vaccination. *Immunity* 50:1530–1541.e8. <https://doi.org/10.1016/j.immuni.2019.05.007>.
- Parra GI, Azure J, Fischer R, Bok K, Sandoval-Jaime C, Sosnovtsev SV, Sander P, Green KY. 2013. Identification of a broadly cross-reactive epitope in the inner shell of the norovirus capsid. *PLoS One* 8:e67592. <https://doi.org/10.1371/journal.pone.0067592>.
- Koelle K, Cobey S, Grenfell B, Pascual M. 2006. Epochal evolution shapes the phylodynamics of interpancemic influenza A (H3N2) in humans. *Science* 314:1898–1903. <https://doi.org/10.1126/science.1132745>.
- Vinje J, Altena SA, Koopmans MP. 1997. The incidence and genetic variability of the small round-structured viruses in outbreaks of gastroenteritis in The Netherlands. *J Infect Dis* 176:1374–1378. <https://doi.org/10.1086/517325>.
- Lindesmith LC, Brewer-Jensen PD, Mallory ML, Debbink K, Swann EW, Vinje J, Baric RS. 2018. Antigenic characterization of a novel recombinant

- GII.P16-GII.4 Sydney norovirus strain with minor sequence variation leading to antibody escape. *J Infect Dis* 217:1145–1152. <https://doi.org/10.1093/infdis/jix651>.
35. Tohma K, Lepore CJ, Ford-Siltz LA, Parra GI. 2017. Phylogenetic analyses suggest that factors other than the capsid protein play a role in the epidemic potential of GII.2 norovirus. *mSphere* 2:e00187-17. <https://doi.org/10.1128/mSphereDirect.00187-17>.
 36. Cannon JL, Barclay L, Collins NR, Wikswo ME, Castro CJ, Magana LC, Gregoricus N, Marine RL, Chhabra P, Vinje J. 2017. Genetic and epidemiologic trends of norovirus outbreaks in the United States from 2013 to 2016 demonstrated emergence of novel GII.4 recombinant viruses. *J Clin Microbiol* 55:2208–2221. <https://doi.org/10.1128/JCM.00455-17>.
 37. Lindesmith LC, Donaldson EF, Beltramello M, Pintus S, Corti D, Swanson 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. <https://doi.org/10.1128/JVI.01192-14>.
 38. Tamminen K, Malm M, Vesikari T, Blazevic V. 2019. Immunological cross-reactivity of an ancestral and the most recent pandemic norovirus GII.4 variant. *Viruses* 11:91. <https://doi.org/10.3390/v11020091>.
 39. Neher RA, Bedford T, Daniels RS, Russell CA, Shraiman BI. 2016. Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. *Proc Natl Acad Sci U S A* 113:E1701–E1709. <https://doi.org/10.1073/pnas.1525578113>.
 40. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, Grenfell BT, Lässig M, McCauley JW. 2018. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol* 26:102–118. <https://doi.org/10.1016/j.tim.2017.09.004>.
 41. Kligen TR, Reimering S, Guzman CA, McHardy AC. 2018. In silico vaccine strain prediction for human influenza viruses. *Trends Microbiol* 26:119–131. <https://doi.org/10.1016/j.tim.2017.09.001>.
 42. Siebenga JJ, Lemey P, Kosakovsky Pond SL, Rambaut A, Vennema H, Koopmans M. 2010. Phylogenetic reconstruction reveals norovirus GII.4 epidemic expansions and their molecular determinants. *PLoS Pathog* 6:e1000884. <https://doi.org/10.1371/journal.ppat.1000884>.
 43. Debbink K, Lindesmith LC, Donaldson EF, Costantini V, Beltramello M, Corti D, Swanson J, Lanzavecchia A, Vinje J, Baric RS. 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>.
 44. 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. <https://doi.org/10.1128/JVI.00864-09>.
 45. Holmes EC, Rambaut A, Andersen KG. 2018. Pandemics: spend on surveillance, not prediction. *Nature* 558:180–182. <https://doi.org/10.1038/d41586-018-05373-w>.
 46. Holmes EC. 2013. What can we predict about viral evolution and emergence? *Curr Opin Virol* 3:180–184. <https://doi.org/10.1016/j.coviro.2012.12.003>.
 47. Lipsitch M, Barclay W, Raman R, Russell CJ, Belsler JA, Cobey S, Kasson PM, Lloyd-Smith JO, Maurer-Stroh S, Riley S, Beauchemin CA, Bedford T, Friedrich TC, Handel A, Herfst S, Murcia PR, Roche B, Wilke CO, Russell CA. 2016. Viral factors in influenza pandemic risk assessment. *Elife* 5:e18491. <https://doi.org/10.7554/eLife.18491>.
 48. Giammanco GM, De Grazia S, Terio V, Lanave G, Catella C, Bonura F, Saporito L, Medici MC, Tummolo F, Calderaro A, Banyai K, Hanson G, Martella V. 2014. Analysis of early strains of the norovirus pandemic variant GII.4 Sydney 2012 identifies mutations in adaptive sites of the capsid protein. *Virology* 450–451:355–358. <https://doi.org/10.1016/j.virol.2013.12.007>.
 49. 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. <https://doi.org/10.1016/j.virol.2013.12.005>.
 50. 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. <https://doi.org/10.1128/JCM.00663-13>.
 51. Eden JS, Bull RA, Tu E, McIver CJ, Lyon MJ, Marshall JA, Smith DW, Musto J, Rawlinson WD, White PA. 2010. Norovirus GII.4 variant 2006b caused epidemics of acute gastroenteritis in Australia during 2007 and 2008. *J Clin Virol* 49:265–271. <https://doi.org/10.1016/j.jcv.2010.09.001>.
 52. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. 2006. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol Direct* 1:34. <https://doi.org/10.1186/1745-6150-1-34>.
 53. Parra GI. 2009. Seasonal shifts of group A rotavirus strains as a possible mechanism of persistence in the human population. *J Med Virol* 81:568–571. <https://doi.org/10.1002/jmv.21423>.
 54. Chan MC, Lee N, Hung TN, Kwok K, Cheung K, Tin EK, Lai RW, Nelson EA, Leung TF, Chan PK. 2015. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat Commun* 6:10061. <https://doi.org/10.1038/ncomms10061>.
 55. Lu J, Fang L, Zheng H, Lao J, Yang F, Sun L, Xiao J, Lin J, Song T, Ni T, Raghwanji J, Ke C, Faria NR, Bowden TA, Pybus OG, Li H. 2016. The evolution and transmission of epidemic GII.17 noroviruses. *J Infect Dis* 214:556–564. <https://doi.org/10.1093/infdis/jiw208>.
 56. Matsushima Y, Ishikawa M, Shimizu T, Komane A, Kasuo S, Shinohara M, Nagasawa K, Kimura H, Ryo A, Okabe N, Haga K, Doan YH, Katayama K, Shimizu H. 2015. Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro Surveill* 20:21173. <https://doi.org/10.2807/1560-7917.ES2015.20.26.21173>.
 57. 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>.
 58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>.
 59. R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
 60. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J* 8:289–317. <https://doi.org/10.32614/RJ-2016-021>.
 61. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
 62. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. <https://doi.org/10.1038/nmeth.2109>.
 63. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704. <https://doi.org/10.1080/10635150390235520>.
 64. 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. <https://doi.org/10.1371/journal.pgen.1002764>.
 65. Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457. <https://doi.org/10.1093/bioinformatics/btq429>.
 66. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973. <https://doi.org/10.1093/molbev/mss075>.
 67. Parra GI, Green KY. 2014. Sequential gastroenteritis episodes caused by 2 norovirus genotypes. *Emerg Infect Dis* 20:1016–1018. <https://doi.org/10.3201/eid2006.131627>.