

Evolutionary analyses of emerging GII.2[P16] and GII.4 Sydney [P16] noroviruses

Guo-li Zheng,[†] Zheng-xi Zhu,[†] Jia-le Cui,[†] and Jie-mei Yu^{*,†}

College of Life Sciences and Bioengineering, Beijing Jiaotong University, 3 Shangyuan Residency, Hai-Dian District, Beijing 100044, China

[†]These authors contribute to the manuscript equally.

[†]<https://orcid.org/0000-0003-2988-384X>

*Corresponding author: E-mail: jmyu1@bjtu.edu.cn

Abstract

GII.2[P16] and GII.4 Sydney [P16] are currently the two predominant norovirus genotypes. This study sought to clarify their evolutionary patterns by analyzing the major capsid VP1 and RNA-dependent RNA polymerase (RdRp) genes. Sequence diversities were analyzed at both nucleotide and amino acid levels. Selective pressures were evaluated with the Hyphy package in different models. Phylogenetic trees were constructed by the maximum likelihood method from full VP1 sequences, and evolutionary rates were estimated by the Bayesian Markov Chain Monte Carlo approach. The results showed that (1) several groups of tightly linked mutations between the RdRp and VP1 genes were detected in the GII.2[P16] and GII.4[P16] noroviruses, and most of these mutations were synonymous, which may lead to a better viral fitness to the host; (2) although the pattern of having new GII.4 variants every 2–4 years has been broken, both the pre- and the post-2015 Sydney VP1 had comparable evolutionary rates to previously epidemic GII.4 variants, and half of the major antigenic sites on GII.4 Sydney had residue substitutions and several caused obvious changes in the carbohydrate-binding surface that may potentially alter the property of the virus; and (3) GII.4 Sydney variants during 2018–21 showed geographical specificity in East Asia, South Asia, and North America; the antigenic sites of GII.2 are strictly conserved, but the GII.2 VP1 chronologically evolved into nine different sublineages over time, with sublineage IX being the most prevalent one since 2018. This study suggested that both VP1 and RdRp of the GII.2[P16] and GII.4 Sydney [P16] noroviruses exhibited different evolutionary directions. GII.4[P16] is likely to generate potential novel epidemic variants by accumulating mutations in the P2 domain, similar to previously epidemic GII.4 variants, while GII.2[P16] has conserved predicted antigenicity and may evolve by changing the properties of nonstructural proteins, such as polymerase replicational fidelity and efficiency. This study expands the understanding of the evolutionary dynamics of GII.2[P16] and GII.4[P16] noroviruses and may predict the emergence of new variants.

Key words: human norovirus; GII.2[P16]; GII.4[P16]; genetic diversity; selective pressure; evolutionary pattern.

1. Introduction

Human noroviruses are recognized as a major cause of acute gastroenteritis (AGE) in all age groups and are associated with approximately 18 per cent of all AGE cases worldwide (Ahmed et al. 2014; Bartsch et al. 2016). Noroviruses are non-enveloped, positive-sense RNA viruses, with a genome ranging from 7.4 to 7.7 kb in length. The genome consists of three open reading frames (ORFs) (Jiang et al. 1993), of which ORF1 encodes for a large polyprotein that is post-translationally cleaved into six nonstructural proteins, including RNA-dependent RNA polymerase (RdRp), which is critical for viral replication (Arias et al. 2016). ORF2 encodes the major structural protein, VP1 (Hutson et al. 2003; Debbink et al. 2012), while ORF3 encodes the minor structural protein, VP2. Both VP1 and VP2 evolved to modulate RdRp activity (Subba-Reddy, Goodfellow, and Kao 2011; Conley et al. 2019). VP1 consists of an N terminal, a shell (S), and two protruding (P) domains (P1 and P2), with P2 having the most sequence variation

and being critical in immune recognition and receptor binding (Sukhrie et al. 2013).

Noroviruses are a group of genetically diverse viruses of the genus *Norovirus* in the family *Caliciviridae*. Based on amino acid diversity of the complete VP1 and nucleotide (nt) diversity of the RdRp, they can be divided into ten (GI–GX) genogroups with forty-eight confirmed genotypes and sixty confirmed P-types, respectively (Chhabra et al. 2019, 2020). While GI, GII, GIV, GVIII, and GIX genogroups can all infect humans, GII viruses are the most frequently detected. The GII VP1 protein has twenty-seven confirmed genotypes, while RdRp has thirty-seven confirmed P-types (Chhabra et al. 2019).

Approximately 50–70 per cent of human norovirus outbreaks worldwide are caused by GII.4 and its new epidemic variants that emerge every 2–4 years (Lopman et al. 2016). Since the 1990s, six major epidemic GII.4 variants have emerged: Grimsby 1995, Farmington Hills 2002, Hunter 2004, Den Haag 2006, New Orleans

2009, and Sydney 2012 (Siebenga et al. 2009; Parra et al. 2017). These variants have a chronological relationship, which may be driven by selective pressures from herd immunity against older variants. In addition, the VP1 protein of these variants can combine with different polymerases. For example, the first five epidemic GII.4 variants are associated with a GII.P4 polymerase, and the latest Sydney 2012 variant is commonly combined with GII.P31 polymerase before 2016, and GII.P16 polymerase after 2016 (Cannon et al. 2017). The combination of VP1 with different RdRps may affect viral infectivity by changing the replication fidelity and efficiency. The newly recombinant GII.4 Sydney [P16] soon replaced the GII.4 Sydney [P31] as the primary cause of outbreaks in North America and Oceania (Lun et al. 2018; Hasing et al. 2019).

The GII.2 was first identified in the 1970s and accounted for <2 per cent of all genotyped variants worldwide before 2016 (Hoa Tran et al. 2013). GII.2[P16] norovirus was originally detected in outbreaks during 2009–10 in Osaka, Japan (Iritani et al. 2012), before emerging in China in 2016 and becoming predominant in outbreaks in Asian countries (Thongprachum et al. 2017; Ao et al. 2018). In China, 81.2 per cent of all norovirus outbreaks are typed as GII.2[P16] (Jin et al. 2020), while in Japan, this genotype has been associated with the second largest number of pediatric norovirus cases over the past decade (Tohma et al. 2017). Unlike GII.4 noroviruses, the currently circulating GII.2 has exhibited predicted antigenicity like those before 2016. However, the current GII.2[P16] norovirus has unique substitutions in both RdRp and other nonstructural proteins and is associated with a higher viral load in infected individuals (Cheung et al. 2019); thus, it has been speculated that the sudden epidemic linked to GII.2[P16] norovirus is the result of point mutations in nonstructural proteins (Tohma et al. 2017).

Since their simultaneous emergence, as two distinct capsid genotypes, an event linked to the emergence of a new viral polymerase, GII.P16, the GII.2[P16] and GII.4[P16] noroviruses have been extensively investigated (Parra et al. 2017; Tohma et al. 2017, 2021; Ao et al. 2018; Barclay et al. 2019; Li et al. 2021). However, now that these genotypes have been circulating for >5 years, their post-emergence diversification requires a more comprehensive examination. The current study analyzed the large scale of sequences of GII.2 and GII.4 Sydney over time and found that VP1 of GII.4 Sydney had greater diversity and evolved more quickly than GII.2 after acquiring the novel GII.P16 RdRp. Multiple residue substitutions and positively selected sites were found at the major antigenic sites of GII.4 Sydney. The substitutions such as T294A, G295N, R297H/Q, and H373R in epitope A, N412K/S/D and H414L/P in epitope E, and H396P in epitope D had greatly altered the carbohydrate-binding surface, which may potentially change the property of the virus and cause immune escape. The antigenic sites of GII.2 are strictly conserved, although the virus was able to evolve into different clusters and exhibit temporal specificity. This study revealed the different evolutionary patterns of the two major prevalent norovirus genotypes, which could help to predict the emergence of potentially epidemic novel variants of noroviruses.

2. Results

2.1 Sequence dataset

A total of 923 GII.2 VP1 from 1976 to 2020 and 731 GII.4 VP1 from 2015 to 2021, as well as 688 GII.P16 RdRp from 2005 to 2020 with high quality (sequences with no Ns or gaps), were retrieved (updated on 15 September 2021). Of the 731 GII.4 VP1 sequences,

698 (95.5 per cent) were genotyped as Sydney variants. For a better understanding of the evolutionary pattern of GII.4 Sydney [P16], another 638 GII.4 Sydney VP1 sequences before 2015 were downloaded (Supporting information file). Of the sequences that contained both RdRp and VP1 genes, 186 were GII.2[P16] and 113 were GII.4[P16]. The geographical and time information for the sequences is listed in the Supplemental Table.

The GII.2 VP1 sequences were combined with eight different P-types, excluding the ones with unknown P-types, the remaining sequences were found to have been combined with eight different P-types (GII.P2, GII.P12, GII.P16, GII.P21, GII.P30, GII.P31, GII.P34, and GII.P35), while the 562 sequences with definite P-types that collected since 2016 were only combined with GII.P2 and GII.P16, and mostly were GII.P16 (95.2 per cent). The GII.4 Sydney VP1 sequences were combined with five different P-types, GII.P4, GII.P12, GII.P13, GII.P16, and GII.P31, except one sequence from the USA in 2014 was combined with GII.P16, while all the remaining ones before 2015 were combined with GII.P31 (83.1 per cent) and GII.P4 (16.2 per cent); the sequences collected after 2015 were most often combined with GII.P16 (52.6 per cent) and GII.P31 (41.9 per cent). The GII.P16 RdRp sequences were associated with nine different capsid types, GII.1, GII.2, GII.3, GII.4, GII.5, GII.12, GII.13, GII.16, and GII.17. Of the 575 sequences collected after 2015, 71.8 and 20.7 per cent were combined with GII.2 and GII.4, respectively (Table 1).

2.2 Genetic diversity of the VP1 and RdRp

To understand the genetic variations, single-nucleotide polymorphism (SNP) callings were performed on the post-2016 GII.2 VP1, the GII.4 Sydney VP1, and post-2016 GII.P16 RdRp genes. A site with a mutation frequency of >1 per cent was defined as an SNP. The following early released sequences were used as references: NC_039476.1 for the post-2016 GII.2 VP1 and GII.P16 RdRp, and JX459908 for the pre-2015 and post-2015 GII.4 Sydney VP1 sequences. The results showed that for the VP1 genes (1,626 bp), the ones from post-2015 GII.4 Sydney had the most SNPs (447 sites), followed by the pre-2015 GII.4 Sydney (341 sites), and the post-2016 GII.2 (271 sites), and the numbers of high-frequency (>50 per cent) SNPs were comparable in the VP1 of post-2016 GII.2, pre-, and post-2015 GII.4, which were 10, 13, and 15, respectively. For the GII.P16 RdRp, there were more SNP sites in the ones associated with GII.2 than those associated with GII.4 (360 versus 271), but much less of the former for its high-frequency (>50 per cent) SNP sites (6 versus 22) (Fig. 1A). All the six high-frequency SNPs on the GII.2-associated GII.P16 RdRp were also high-frequency on the GII.4-associated GII.P16 RdRp, but there were sixteen high-frequency SNPs on the GII.4-associated GII.P16 RdRp that exhibited particularly low occurring frequency or even no mutation in the GII.2-associated GII.P16 RdRp (Fig. 1B).

To clarify the evolutionary directions of the GII.2[P16] and GII.4[P16] noroviruses, we analyzed mutation linkage between the RdRp and VP1 genes by the sequences that simultaneously contained the two genes. It was found that GII.2[P16] had three groups of tightly linked SNPs, involving fifteen sites: 126-387/1226, 165-1212, and 171/367/418/897-126/207/600/675/720/819, while GII.4[P16] also had three groups of linked SNPs, but involving eight sites: 276-438, 279-1252, and 1370-732/753/834. The mutation frequencies of all the linked SNPs were higher than 10 per cent (Table 2). Interestingly, the linked sites on GII.2[P16] were totally different from those on GII.4[P16], and except K457R in the GII.4-associated RdRp and H123Y in the GII.2-associated RdRp, all the other linked sites were synonymous mutations (Table 2).

Table 1. Numbers of combined genotype or P-type of the downloaded sequences.

P-type	GII.2 VP1 (679 of 923 were combined with definite P-type)		GII.4 Sydney VP1 (488 of 1336 were with definite P-type)				GII.P16 (all were with definite genotype)					
	Before 2016		After 2016		Before 2015		After 2015		Before 2016		After 2016	
	n (%)	P-type	n (%)	P-type	n (%)	P-type	n (%)	Genotype	n (%)	Genotype	n (%)	
GII.P2	44 (37.6)	GII.P2	27 (4.8)	GII.P4	1 (0.7)	GII.P4	15 (4.3)	GII.2	54 (47.8)	GII.1	12 (2.1)	
GII.P12	1 (0.9)	GII.P16	535 (95.2)	GII.P16	23 (16.2)	GII.P16	182 (52.6)	GII.3	15 (13.3)	GII.2	413 (71.8)	
GII.P16	56 (47.9)			GII.P31	118 (83.1)	GII.P31	145 (41.9)	GII.4	12 (10.6)	GII.3	11 (1.9)	
GII.P21	2 (1.7)					GII.P12	3 (0.9)	GII.5	1 (0.9)	GII.4	119 (20.7)	
GII.P30	3 (2.7)					GII.P13	1 (0.3)	GII.13	23 (20.4)	GII.12	11 (1.9)	
GII.P31	9 (7.7)							GII.16	6 (5.3)	GII.13	9 (1.6)	
GII.P34	1 (0.9)							GII.17	2 (1.8)			
GII.P35	1 (0.9)											

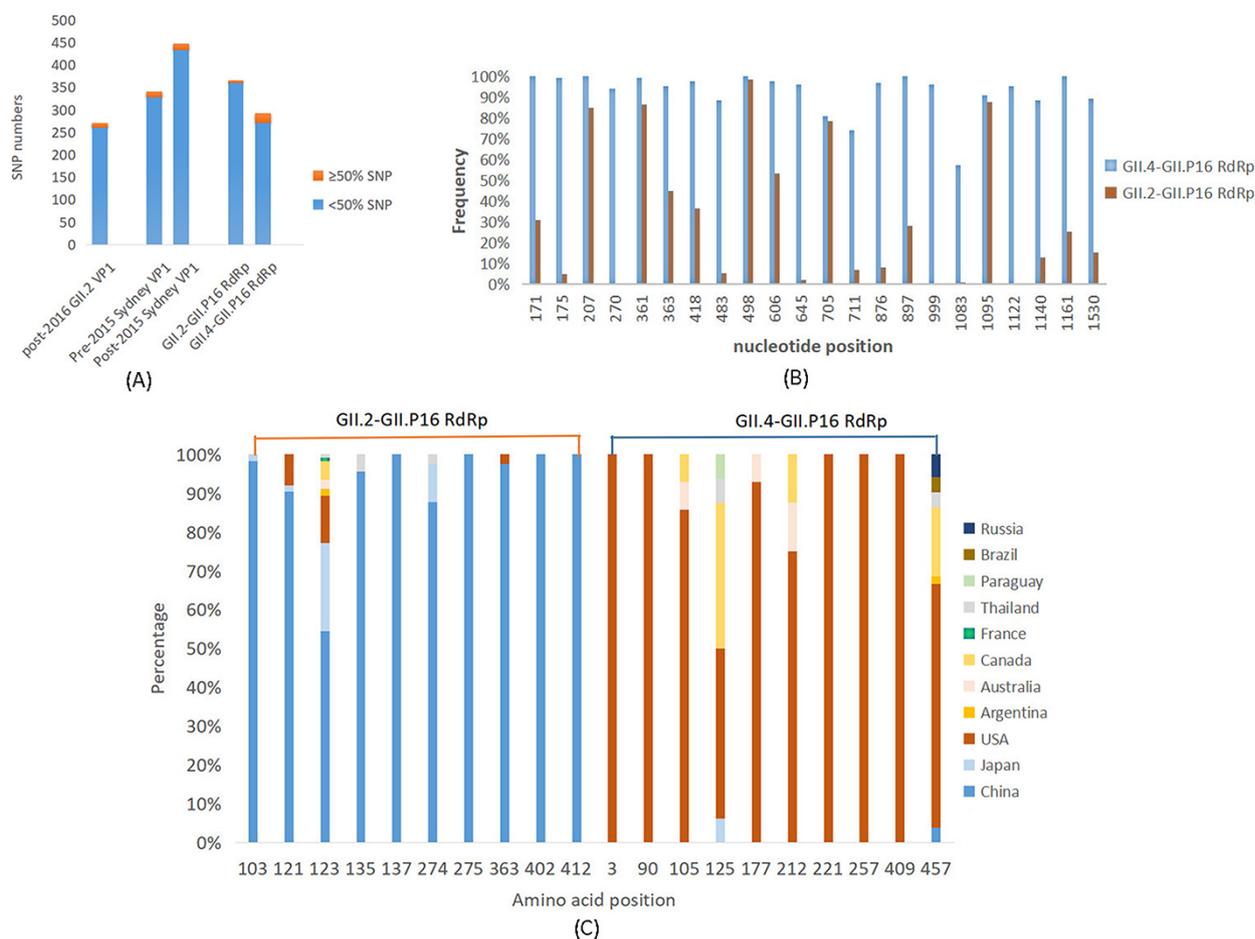


Figure 1. Information for the SNPs and amino acid substitutions. (A) Numbers of SNPs in the VP1 and RdRp genes. GII.4 VP1 was more diversified than GII.2 VP1. GII.P16 RdRp associated with GII.2 had more SNPs than that associated with GII.4 but had less of high-frequency (>50 per cent) SNPs. (B) High-frequency SNP sites on the GII.4-associated GII.P16 RdRp. Six of the twenty-two sites showed very low frequencies or even no mutation in the GII.2-associated GII.P16 RdRp. (C) Geographic distributions of the sequences involved in the changed amino acid sites in the GII.2- and GII.4-associated GII.P16 RdRp. The GII.2-associated RdRp sequences were mostly from East Asia, and the GII.4-associated RdRp sequences were mostly from North America.

2.3 Informative amino acid substitutions on the VP1 and RdRp proteins

Amino acid changes were counted on GII.2 VP1, GII.4 Sydney VP1, and GII.P16 RdRp sites if they occurred with >1 per cent frequencies. The GII.2 norovirus became predominant after recombining with the novel GII.P16 RdRp in 2016. Results showed that the 220

pre-2016 GII.2 VP1 sequences had fifty-two amino acid substitutions, while the 703 post-2016 sequences only had twenty-one (Fig. 2A, B). The earliest GII.4 Sydney VP1 has recombined with the novel GII.P16 RdRp since 2015. The 638 pre-2015 and the 698 post-2015 GII.4 Sydney VP1 sequences had thirty-eight and forty-six amino acid substitutions, respectively (Fig. 2C, D). Both the

Table 2. Tightly linked mutations on RdRp and VP1 genes of GII.2- and GII.4-associated GII.P16 norovirus.

	Groups (frequency)	RdRp			VP1		
		nt site	Mutation	aa change	nt site	Mutation	aa change
GII.4 VP1—GII.P16 RdRp	Group 1 (31%)	276	T-A	92 (-)	438	A-G	146 (-)
	Group 2 (27%)	279	C-T	93 (-)	1252	C-T	418 (-)
	Group 3 (43%)	1370	A-G	K457R	732	T-C	243 (-)
					753	G-A	251 (-)
					834	G-T	278 (-)
GII.2 VP1—GII.P16 RdRp	Group 4 (31%)	126	G-A	42 (-)	387	G-A	129 (-)
					1224	T-C	408 (-)
	Group 5 (12%)	165	T-C	55 (-)	1212	A-G	404 (-)
	Group 6 (38%)	171	T-C	57 (-)	126	C-T	42 (-)
		367	C-T	H123Y	207	A-T	69 (-)
		418	C-T	140 (-)	600	T-C	200 (-)
		897	T-C	299 (-)	675	G-A	225 (-)
					720	T-C	240 (-)
					819	A-G	273 (-)

Notes: 'nt', nucleotide. 'aa', amino acid. '-' stands for a synonymous mutation. Sites on the RdRp and VP1 that are in the same row with the same color were linked.

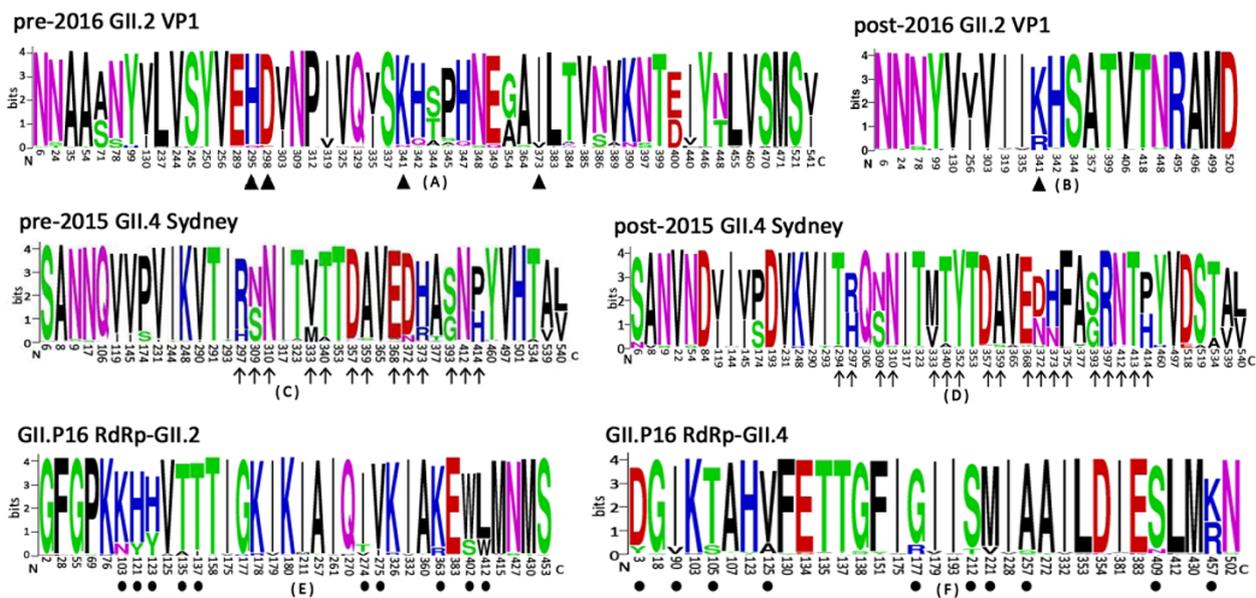


Figure 2. Informative amino acid substitutions on the VP1 and RdRp. (A, B) Pre-2016 and post-2016 GII.2 VP1. (C, D) Pre-2015 and post-2015 GII.4 Sydney VP1. (E, F) GII.P16 RdRp associated with GII.2 and GII.4 Sydney VP1. Multiple changed sites of VP1 were located on major antigenic motifs. The changed sites with >5 per cent frequencies in the GII.2-associated and GII.4-associated RdRp were totally distinct. Antigenic sites on the GII.2 that extrapolated from GII.4 were marked with '▲', the major antigenic sites on GII.4 Sydney were marked with '↑', and amino acid substitutions with frequencies of >5 per cent on the GII.P16 RdRp were marked with '●'.

numbers of changed sites and sites occurring with >5 per cent frequencies on the latest VP1 sequences in GII.4 Sydney were higher than those in GII.2, demonstrating that GII.4 Sydney had a higher level of genetic variation. No amino acid substitutions were found on the sites that involved in binding histo-blood group antigen of the GII.2 or GII.4 Sydney noroviruses. However, for the post-2015 GII.4 Sydney VP1, eighteen amino acid substitutions were located at major antigenic sites, five at antigenic site A (T294A, R297H, E368Q, D372N, and H373N/R), two at antigenic site C (T340A and F375L), two at antigenic site D (S393G and R397Q), three at antigenic site E (N412S/K/D, T413I, and P414H), three at antigenic site G (Y352L, D357N, and A359S), two at motif H (S309N and N310S), and one at motif B (M333V) (Fig. 2F). The pre-2015 sequences had five fewer substitutions (T294A, Y352L, F375L, R397Q, and P414H) than the post-2015 sequences. Until now, no antigenic sites have been described for GII.2 viruses. When extrapolating GII.4 antigenic

sites to GII.2, four substitutions (H295Q, D298E, V373I, and K341R) were found at the antigenic sites of all the GII.2 VP1 sequences, and only one (K341R) was found on the post-2016 VP1.

GII.P16 RdRp has been widely prevalent since 2016 and primarily recombined with GII.2 and GII.4 Sydney VP1. Five unique amino acids in the post-2016 RdRp protein may have altered the viral polymerase kinetics or fidelity and allowed for the predominance of the novel RdRp (Ao et al. 2018). In this study, thirty-four and thirty-three amino acid substitutions were found in the post-2016 GII.P16 RdRp that recombined with GII.2 and GII.4, respectively. Of these changed sites, nine and ten occurred with >5 per cent frequencies in the GII.2-associated and GII.4-associated RdRp, respectively. The changed sites in the GII.2-associated GII.P16 RdRp were totally distinct from those in the GII.4-associated (Fig. 2E, F). For the sequences individually involved in the nineteen changed sites, their geographical distributions revealed that

Table 3. Positively selected sites in the VP1 and RdRp genes by different methods.

Genotype or P-type of the sequences	Methods		
	FUBAR	MEME	SLAC
GII.2 VP1 1971–2020	6, 24, 78, 99	6, 24, 78, 99, 226, 344, 345, 354, 384, 496	6, 78
GII.4 VP1 2011–14	6, 9, 17, 309, 359, 373, 393, 460	6, 8, 9, 17, 74, 78, 232, 236, 309, 311, 352, 359, 368, 373, 393, 419, 420, 421	17, 309, 359, 368, 373, 393
GII.4 VP1 2015–21	6, 8, 9, 372, 373, 393, 460, 534	6, 8, 9, 18, 19, 134, 143, 144, 150, 171, 199, 294, 298, 305, 311, 340, 352, 353, 355, 372, 373, 393, 404, 413, 419, 420, 421, 423, 443, 444, 460, 534	6, 8, 372, 373, 393, 460, 534
GII.2-combined GII.P16 RdRp	412	59, 270, 274, 412 , 430	412
GII.4-combined GII.P16 RdRp	–	137	–
other genotype-combined GII.P16 RdRp	–	138	–

Note: 'MEME', 'FUBAR', and 'SLAC' represent the 'mixed-effects model of evolution', the 'fast unbiased Bayesian approximation', and the 'single-likelihood ancestor counting', respectively. '–' stands for 'none'. Strong positively selected sites are marked in **bold** and those located at major antigenic sites are in *italics* and underlined.

the GII.2-associated RdRps were mostly located in East Asia (China and Japan) and GII.4-associated RdRp in North America (USA and Canada) (Fig. 1C).

2.4 Positive selection pressure on the VP1 and RdRp proteins

All the sequences of the three proteins, GII.2 VP1, GII.4 Sydney VP1, and GII.P16 RdRp, have been undertaken through selective pressure analysis. To reduce the high possibility of occasional events, the sites supported by at least two methods were considered as candidates under positive selection, while those supported by all three methods were defined as strongly selected sites (P-value of <0.1 in mixed-effects model of evolution (MEME) and single-likelihood ancestor counting (SLAC); posterior probabilities of >0.9 in fast unbiased Bayesian approximation (FUBAR)). Results suggested that two sites (6 and 78) on the GII.2 VP1, six sites (17, 309, 359, 368, 373, and 393) on the pre-2015 GII.4 Sydney VP1, and seven sites (6, 8, 372, 373, 393, 460, and 534) on the post-2015 GII.4 Sydney VP1 were under strong positive selection. The two positively selected sites on the GII.2 VP1 were located at the S domain, while four of the six and three of the seven positively selected sites on the pre- and post-2015 GII.4 VP1 were located at major antigenic sites: 359 on antigenic G, 368, 372, and 373 on antigenic A, and 393 on antigenic D. Only one positively selected site was found on the GII.P16 RdRp combined with GII.2, while no positively selected sites were found on the RdRp recombined with GII.4 and other genotypes (Table 3).

2.5 Phylogenetic analysis of the GII.2 and GII.4 Sydney noroviruses

The maximum likelihood trees for the GII.2 and GII.4 Sydney noroviruses were performed based on the complete VP1 nucleotide sequences. Results suggested that the circulating GII.2 norovirus sequences formed different transmission subclades with an obvious time specificity; sequences circulating before 2000 were at the bottom of the tree, those circulating between 2001–15 were at the middle, and those after 2016 were at the top. The post-2016 GII.2 viruses spread rapidly and exhibited a wide range of genetic diversity, which can be further divided into nine sublineages (I–IX). The sequences from 2016 and 2017 exhibited the greatest diversity and covered all nine sublineages, while

most of the sequences from 2018 to 2020 clustered in sublineage IX (Fig. 3A). No geographical specificity was observed in GII.2 norovirus.

The Sydney 2012 norovirus was the most prevalent GII.4 variant identified since 2012 with sequences that evolved into three lineages (I–III). Lineage I, formed by eleven sequences from the USA during 2012–5, was the least prevalent, and lineages II and III were widely detected during 2012–20, with lineage III being the most prevalent (Fig. 3B). No obvious time specificity was observed in lineages II and III, but the geographical distributions were distinct. The sequences in lineage II were primarily from North America (33.0 per cent), followed by East Asia (20.4 per cent) and Europe (14.7 per cent), Oceania (12.5 per cent), South Asia (9.3 per cent), Africa (6.3 per cent), and South America (3.8 per cent), while the sequences in lineage III were mostly from East Asia (>67.9 per cent), followed by North America (9.4 per cent), Europe (8.0 per cent), Oceania (7.5 per cent), South America (5.9 per cent), and South Asia (0.8 per cent), with no sequences from Africa (Fig. 3C). Notably, the GII.4 Sydney sequences that were collected during 2018–21 were mainly distributed in three sublineages: two in lineage II (sublineages II-1 and II-2) and one in lineage III (sublineage III-1). Importantly, these 2018–21 sequences showed some geographical specificity with the sequences from East Asia being primarily distributed in sublineage III-1, those from North America and South Asia being primarily distributed in sublineage II-1, and those from Africa being distributed in sublineage II-2 (Fig. 3D).

2.6 Evolutionary rate comparisons

The nucleotide substitution rates for the VP1 of GII.2 and GII.4 Sydney from different periods, and GII.P16 RdRp associated with GII.2 and GII.4 were, respectively, estimated. The estimated mean evolutionary rate of the post-2016 GII.2 VP1 was 2.44×10^{-3} nt substitutions/site/year (95 per cent highest posterior density (HPD) interval: $2.49\text{--}3.09 \times 10^{-3}$), which was comparable to the pre-2016 GII.2 VP1 (2.77×10^{-3} nt substitutions/site/year, 95 per cent HPD interval: $2.04\text{--}2.87 \times 10^{-3}$), and the mean rate of the post-2015 GII.4 Sydney VP1 (5.18×10^{-3} nt substitutions/site/year, 95 per cent HPD interval: $4.46\text{--}5.88 \times 10^{-3}$) was comparable to the pre-2015 GII.4 Sydney VP1 (4.56×10^{-3} nt substitutions/site/year, 95 per cent HPD interval: $3.67\text{--}6.98 \times 10^{-3}$). Moreover, the mean rate of the GII.P16 RdRp that associated with GII.2 VP1 (2.64×10^{-3} nt substitutions/site/year, 95 per cent HPD interval: $2.16\text{--}3.11 \times 10^{-3}$) was

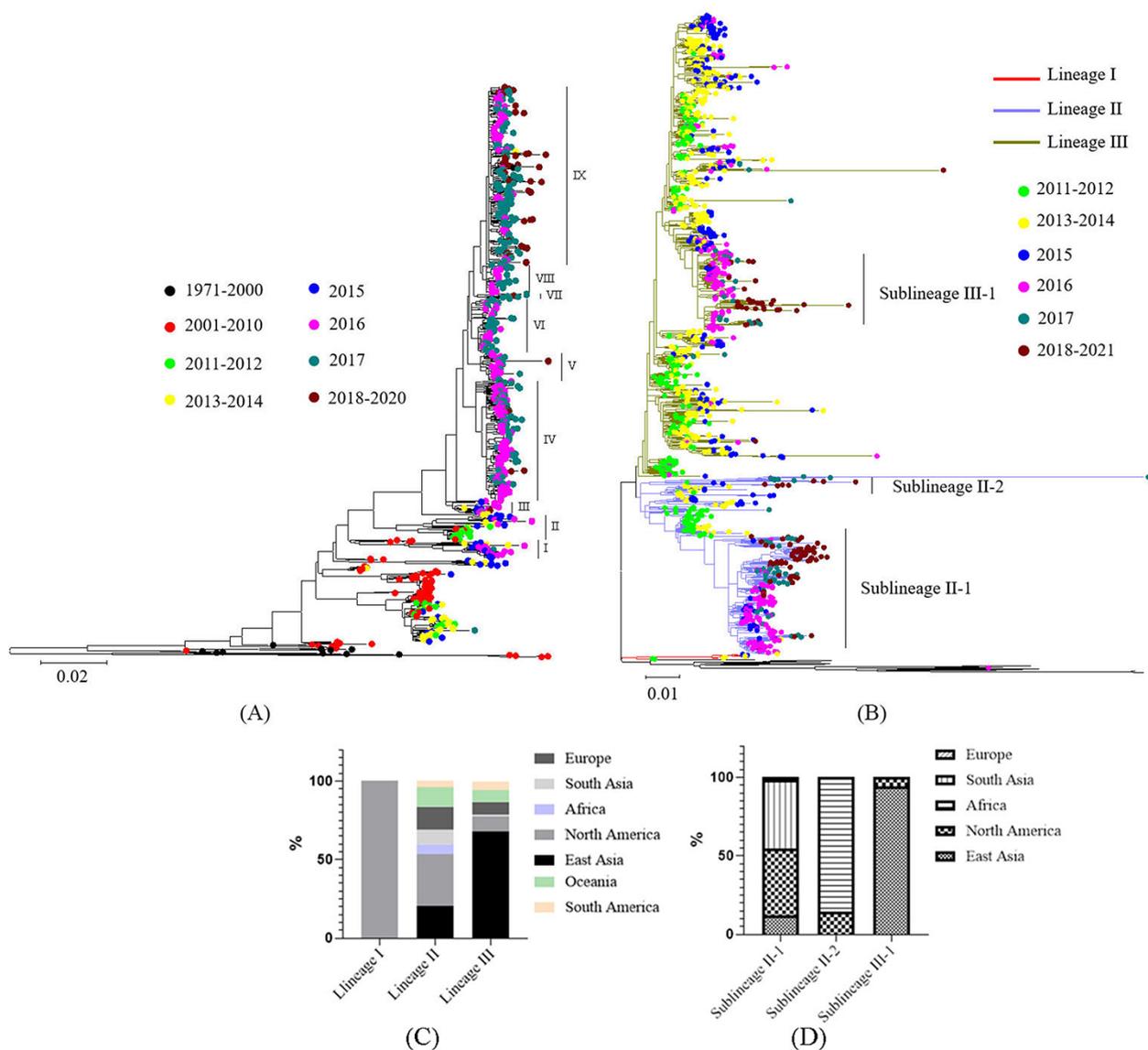


Figure 3. Phylogenetic analysis using VP1 nucleotide sequences of GII.2 and GII.4 Sydney by the maximum likelihood method and spatial distributions of the cluster/subclusters in the GII.4 Sydney norovirus. Sequences of the early GII.2 VP1 in the 1970s and the previously reported epidemic GII.4 norovirus strains (New Orleans, DenHaag, Hunter, Farmington, Grimsby, and Osaka) were used as their roots, respectively. (A) GII.2 VP1. The sequences from 2018–20 evolved into nine determined sublineages. (B) GII.4 Sydney VP1. The sequences evolved into three lineages (I–III). (C) Geographical distributions of the clusters in the GII.4 Sydney VP1. Lineage I only contained eleven sequences from North America, lineage III contained most sequences from South Asia, and Lineage II from North America. (D) Geographical distributions of the latest GII.4 Sydney VP1. The sequences from East Asia mainly distributed in sublineage III-1, while the ones from North America and South Asia mostly distributed in sublineage II-1.

comparable to that associated with GII.4 Sydney VP1 (3.29×10^{-3} nt substitutions/site/year, 95 per cent HPD interval: $2.59\text{--}3.89 \times 10^{-3}$). The mean rate of VP1 was higher for the post-2015 GII.4 Sydney than the post-2016 GII.2 (5.18×10^{-3} nt substitutions/site/year versus 2.77×10^{-3} nt substitutions/site/year), and the 95 per cent HPD intervals were 4.46×10^{-3} to 5.88×10^{-3} and 2.49×10^{-3} to 3.09×10^{-3} , respectively (Fig. 4).

2.7 Tertiary structure dynamics of norovirus VP1

As listed above, multiple amino acid substitutions were located at the major antigenic sites of GII.4 Sydney VP1. Previous studies suggested that simultaneous mutations at predominant antigenic sites A (294–298, 368, and 372–373), G (352, 355–357, 359, and 364), D (393–397), and E (407, 411, and 412–414) are needed for major antigenic property shifts of GII.4 norovirus, especially for sites 352,

355, 357, 368, and 378 that are involved in the emergence of new GII.4 variants (Kendra et al. 2021). Substitutions were observed on sites 294, 295, 297, 368, 372, and 373 in antigenic site A, 352 and 359 in antigenic site G, 393 and 396 in antigenic site D, and 412 and 414 in antigenic site E, and distinct combinations were found in different profiles. Specifically, when compared with the wild-type GII.4 Sydney [P16] from the USA in 2015 (KY94750), 510 out of the 698 post-2015 GII.4 Sydney [P16] sequences (73.1 per cent) exhibited amino acid substitutions in the major epitopes, of which seventeen (3.3 per cent) had substitutions in all the four epitopes, 227 (44.5 per cent) had substitutions in three of the four epitopes, and 107 (21.0 per cent) had substitutions in two of the four epitopes.

To evaluate the impact of the substitutions on the carbohydrate-binding interface of GII.4 Sydney [P16], the amino

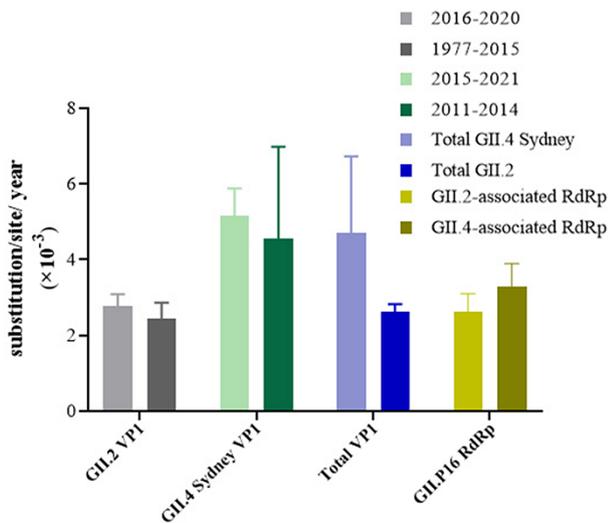


Figure 4. Evolutionary rate comparison of VP1 and RdRp genes between different groups. The mean rate of VP1 for GII.4 Sydney was higher than that for GII.2, and the rates of the pre-2016 GII.2 VP1, pre-2015 GII.4 Sydney VP1, and GII.2-associated GII.P16 RdRp were, respectively, comparable to those for the post-2016 GII.2 VP1, post-2015 GII.4 Sydney VP1, and GII.4-associated GII.P16 RdRp.

acids located at the major blockade epitopes (A, G, D, and E) and changes involving three or four major epitopes were analyzed in homology models. The structure of a wild-type GII.4[P16] Sydney VP1 sequence from the USA in 2015 (KY94750) was used as a reference (Fig. 5A). Some sequences, such as those from the USA in 2018 (MT028542) and those from Canada in 2019 (MW661256), had amino acid substitutions occurring simultaneously at the four major epitopes (A, G, D, and E), while other sequences, such as those from Botswana in 2017 (MW661261) and China in 2019 (MW661256), had amino acid changes that occurred in three of the four major epitopes. It was shown that the substitutions of T294A, G295N, R297H/Q, and H373R in epitope A, N412K/S/D and H414L/P in epitope E, as well as H396P in epitope D greatly altered the structure of the binding surface, while E368Q, D372N, and H373N in epitope A, S393G in epitope D, and Y352L and A359S in epitope G exhibited marginal effects (Fig. 5B–G). In summary, simultaneous amino acid changes in the four major epitopes, especially A, D, and E, resulted in obvious changes in carbohydrate-binding surface that may potentially alter the antigenic property and binding ability of the virus.

The structure of the norovirus RdRp resembles a partially closed right hand with finger, palm, and thumb subdomains, and seven organized motifs, A–G, are involved in RNA synthesis (Fig. 6). These motifs were highly conserved in different P-types of RdRp. However, site 163 in motif F of GII.P2 and GII.P16, and site 337 in motif C of GII.P16 were different from other P-types. Notably, 29.8 per cent of GII.P16 RdRps that combined with GII.2 VP1 had the same residue as GII.P2 at site 121 of motif G (Fig. 6). The A–G motifs on RdRp interact with the template, the nascent RNA, and the nucleoside triphosphates for RNA synthesis and, thus, play an important role in norovirus replication (Gorbalenya et al. 2002; Deval et al. 2017). The three combining residue differences in the conserved motifs may affect the fidelity or other properties of GII.P16 RdRp.

3. Discussion

Worldwide GII.4 noroviruses have been responsible for most of the norovirus-associated AGE for nearly three decades because of the chronologically sequential emergence of novel variants every

2–4 years. The evolutionary pattern of GII.4 viruses involves the accumulation of amino acid substitutions in the P2 subdomain that lead to antigenic differences (Lindsmith et al. 2008; Tohma et al. 2019). Sydney 2012 was the predominant variant of GII.4 over the past decade. GII.4 Sydney [P31] was the primary variant before 2016, and GII.4 Sydney [P16] emerged and became as predominant strain in 2016. Norovirus RdRp can shape viral evolution by changing viral fidelity and replication rate (Smerina et al. 2019). The GII.P16 RdRp with five unique amino acid substitutions after 2016 exhibited a better fitness than other P-types. This RdRp is also associated with GII.2 viruses and has rapidly become predominant in outbreaks in Asia (Ao et al. 2017; Thongprachum et al. 2017; Jin et al. 2020). Recombination and genetic drift are key mechanisms in the evolution and diversity of noroviruses. Previous studies have shown minor variations in non-GII.4 VP1 over the past few decades (Siebenga et al. 2009; Parra et al. 2017). While both GII.2[P16] and GII.4[P16] noroviruses have been circulating for >5 years, their evolutionary patterns have not been comprehensively characterized. Thus, this study systematically analyzed the genetic diversity, amino acid substitution, selective pressure, phylogeny, and spatial changes in the major epitopes of the viruses.

For the VP1, it was more diversified in GII.4 Sydney than in GII.2 as expected. For the GII.P16 RdRp, there were more SNPs in the GII.2-associated RdRp than those in the GII.4-associated RdRp, whereas the ones with high frequencies were less. Since the frequencies of those high-frequency SNPs in the GII.4-associated RdRp were not the same, and their corresponding sites in the GII.2-associated RdRp exhibited great differences (some were high frequency, some were low frequency, and some even have no mutation), we may infer that the high-frequency SNPs were not caused by the founder effect. Further analysis revealed that there were groups of tightly linked mutations between VP1 and RdRp genes both in the GII.2[P16] and GII.4[P16] noroviruses, and the involved sites had high mutation frequencies and were totally distinct between the two viruses, indicating that different evolutionary directions may exhibit. It is worth mentioning that the high-frequency SNPs and the tightly linked mutations were mostly synonymous mutations. Although these kinds of mutations do not alter the amino acid, previous studies reported that they can change the structure or function of an mRNA and were predicted to affect translational speed (Hunt et al. 2014; Kristofich et al. 2018), we inferred here that these synonymous mutations may lead to a better fitness of the noroviruses.

Multiple amino acid substitutions have been detected both in the GII.2 and GII.4 Sydney VP1. Although there were no specific antigenic sites determined for GII.2, it has been proved that its major antigenic region is located at the P2 domain, and several antigenic sites that extrapolated from GII.4 affect monoclonal antibody binding (Mallory et al. 2020). Only one site with amino acid changes on the post-2016 GII.2 VP1 was found located at the major antigenic sites that were extrapolated from GII.4. While fifty-two amino acid changes were detected on the pre-2016 GII.2 VP1, only four were located at the extrapolated major antigenic sites, with three, H295Q, D298E, and V373I, occurring in low frequency and not concurrent. Since minor changes in sequence are not enough for a meaningful antigenic variation on GII.4 norovirus (Kendra et al. 2021), it is likely that amino acid changes on the pre-2016 GII.2 VP1 sequences had little effect on the predicted antigenicity of the virus, and may instead be caused by the low-level accumulation of random mutations over nearly 50 years of prevalence. Consistent with a previous study showing that GII.2 VP1 lacks the plasticity of GII.4 viruses (Parra et al. 2017), this

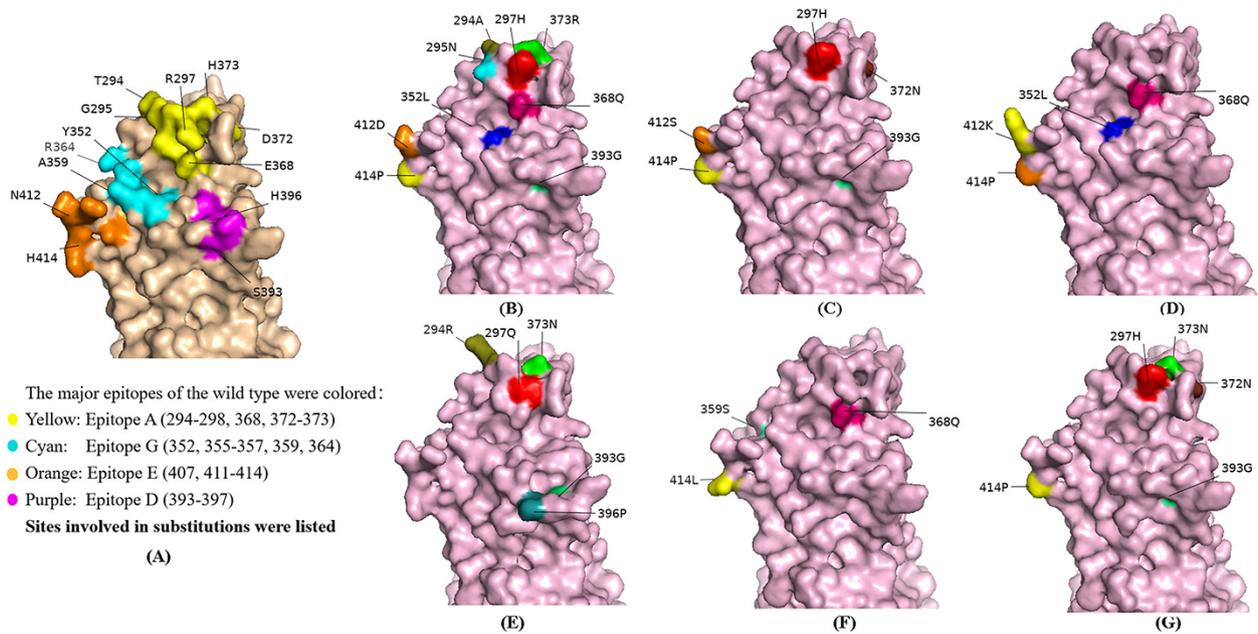


Figure 5. Spatial structures of the major epitopes on GII.4 Sydney norovirus. (A) Wild-type GII.4 Sydney VP1 (from the USA in 2015); (B–G) GII.4 Sydney norovirus mutants. Substitutions of T294A, G295N, R297H/Q, and H373R in epitope A, N412K/S/D and H414L/P in epitope E, and H396P in epitope D had greatly altered the binding surface, while E368Q, D372N, and H373N in epitope A, S393G in epitope D, and Y352L and A359S in epitope G exhibited marginal effects. The regions with amino acid change were marked in color.

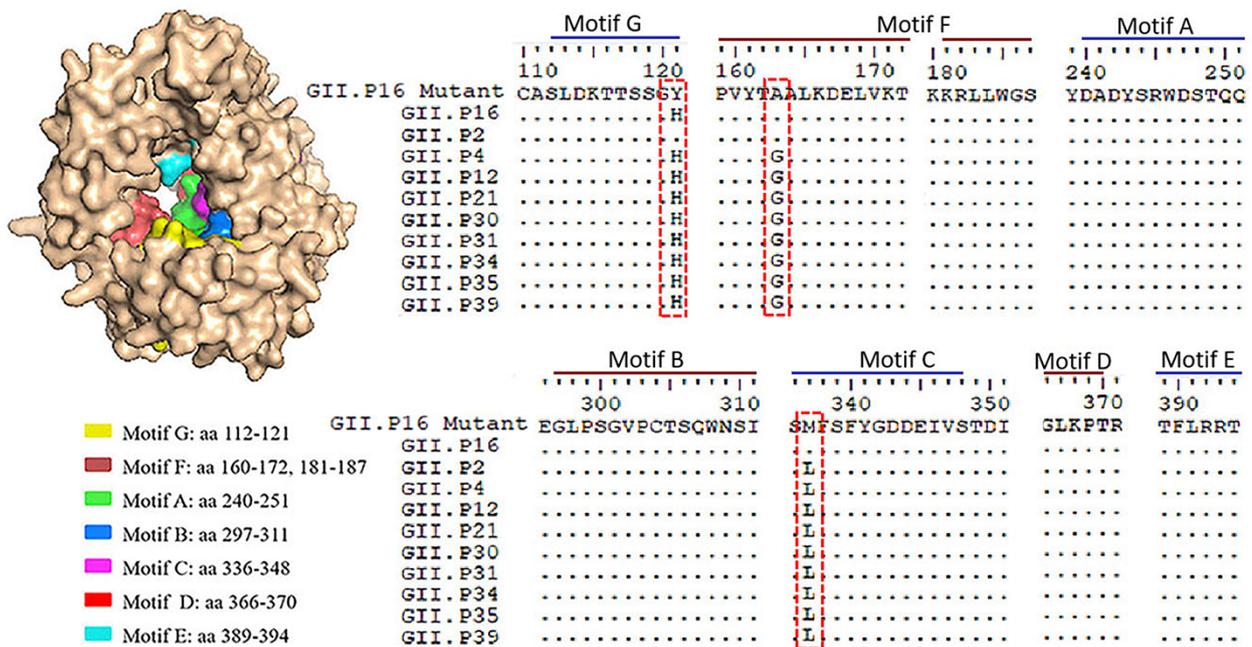


Figure 6. Seven organized motifs on the GII.P16 RdRp. Motifs A, B, D, and E were strictly conserved in different P-types of RdRp. Site 163 in F motif and 337 in C were different in some P-types. Numbers of GII.P16 RdRps from 2016–9 and GII.P2 had the same amino acid at site 121 in motif G.

study found that the antigenic motifs on GII.2 VP1 were conserved. The GII.4 Sydney VP1 occurring before and after 2015 exhibited a high level of genetic diversity and distinct from the GII.2 VP1, and half of the amino acids located at major antigenic sites exhibited amino acid substitutions. As continuous selective pressures from host immunity lead to accumulated amino acid substitutions on the GII.4 VP1, mutations on antigenic sites could alter viral antigenicity and lead to the emergence of a novel variant (Lindesmith et al. 2013; Tohma et al. 2019).

Norovirus RdRp plays a central role in viral genome replication, and the RdRp region is shown to evolve as rapidly as other regions of the norovirus genome, including VP1 (Tohma et al. 2021). In this study, multiple amino acid substitutions were detected in the GII.P16 RdRp, and interestingly, the substitutions that occurred with >5 per cent frequencies in the GII.P16 RdRp that associated with GII.2 and GII.4 were completely distinct. Given that VP1 can enhance RdRp activity (Subba-Reddy, Goodfellow, and Kao 2011), it was assumed that RdRp, or at least GII.P16 RdRp, may exhibit

distinct evolutionary patterns by combining with different VP1. However, due to the low-occurring frequencies and no conserved differences were found in the corresponding VP1, we cannot rule out the possibility that these mutations were caused by the founder effect. But at least, the differences revealed that recombination only occurred once, and there was limited recombination among the two co-circulating viruses. Since GII.2 is antigenically stable, and the GII.P16 RdRp that associated with GII.2 had one site under positive selection and three amino acid differences in the conserved motifs, it was speculated that GII.2[P16] noroviruses may generate novel epidemic variants by changing the properties of RdRp, such as fidelity and efficiency. This is distinct from prior GII.4 noroviruses that have substitutions at antigenic sites in the P2 region and facilitate escape from herd immunity (Kendra et al. 2021).

Positive selection of individual codons usually reflects changes that provide an immunological fitness advantage. Previous analyses did not reveal episodic positive selections in the GII.2 VP1 (Nagasawa et al. 2018; Li et al. 2021); however, two sites, N6S and N78S, located at the N terminal and the S domain, respectively, were found to be under strong positive selection in this study. It was inferred that the S domain may bind to RdRp and enhance species-specific RdRp activity; thus, positive selection in the S domain may optimize the stability of the replication complex (Subba-Reddy et al. 2012, 2017). No positive selections were detected in the major antigenic motifs of GII.2 VP1, suggesting that its variation is generated by natural evolution, not by selective pressures from the host immune system. Consistent with results from previous studies that most positively selected residues mapped to the P2 subdomain (Tohma et al. 2019, 2021), several positively selected sites (four in the pre-2015 and two in the post-2015) were located at major antigenic sites of GII.4 Sydney VP1 and may be associated with changes in GII.4 Sydney norovirus antigenicity.

While no antigenic diversification of GII.2 VP1 was observed over time, phylogenetic analysis linked the prevalence of GII.2 viruses to the chronological emergence of new variants in the human population. The post-2016 GII.2 VP1 has evolved into nine independent sublineages. The most prevalent sublineage since 2018 was IX. Sublineages at the early stage (I–III) were no longer detected after 2017. GII.4 Sydney VP1 represented three different clusters, with cluster I being the least popular, only prevalent on a small scale in the USA between 2012 and 2015, while clusters II and III spread in parallel worldwide and lacked obvious temporal specificity. Clusters II and III correspond to the two larger clades of the Sydney lineage reported previously (Hernandez et al. 2020), suggesting that despite co-circulating, GII.4[P31] and GII.4[P16] viruses rarely recombine, further confirming results from a recent study (Tohma et al. 2021). Interestingly, the GII.4 Sydney VP1 from 2018–21 has geographical specificity in East Asia, South Asia, and North America, suggesting that the genetic background of these populations or other unknown factors may affect the evolutionary direction of the virus. However, it is possible that these differences are attributed to surveillance reporting biases. A previous study revealed that the evolutionary rate was 5.40×10^{-3} nt substitutions/site/year in GII.4 VP1 and 2.99×10^{-3} nt substitutions/site/year in GII.2 VP1 (Parra et al. 2017). In the current study, the VP1 evolutionary rates of the GII.4 Sydney and GII.2 VP1 genes were, respectively, estimated to be comparable to those previously reported (Parra et al. 2017; Li et al. 2021). The rates were similar in the pre-2016 and post-2016 GII.2 sequences, exhibiting overlap with 95 per cent HPD intervals. While a previous study showed that the evolutionary rates of GII.4 VP1 differed between

variants (Motoya et al. 2017), the current study found that the pre- and post-2015 Sydney VP1 had comparable evolutionary rates. Results also showed that after recombining with the novel GII.P16 RdRp, the post-2015 GII.4 VP1 had a higher evolutionary rate than the post-2016 GII.2 VP1, which is consistent with well-established findings that GII.4 evolves more quickly than other genotypes. However, the evolutionary rates of the GII.P16 RdRp that associated with GII.2 and GII.4 were comparable, indicating that different VP1 probably had marginal effects on them.

Studies have revealed that the novel emergent GII.4 variants can be monitored and detected. Pre-epidemic variants usually circulate at low levels and are associated with limited outbreaks prior to spreading globally. For example, pre-epidemic New Orleans 2009 was first detected in 2008, and Sydney 2012 was identified in 2010 (Eden et al. 2014; White 2014). Small changes to the GII.4 VP1 may not be enough to antigenically separate these viruses from the same variant, and amino acid changes at major antigenic sites, especially synchronous changes in major epitopes A, D, G, and E, are necessary for major shifts in antigenic features (Tohma et al. 2019; Kendra et al. 2021). Substitutions of T294A, G295N, R297H/Q, and H373R in epitope A, N412K/S/D and H414L/P in epitope E, and H396P in epitope D showed substantial changes in the corresponding carbohydrate-binding surface. Of note, seventeen sequences with amino acid substitutions in all the four antigenic sites, A, D, G, and E, and 334 sequences with amino acid substitutions at two or three antigenic sites deserve careful monitoring as they may have epidemic potential. Although E368Q in antigenic site A and Y352L in G seemed to have little effect on the spatial structure, as they were of the five key residues (352, 355, 357, 368, and 378) involved in the evolution and emergence of GII.4 norovirus new variants (Tohma et al. 2019), special attention should be paid to them. Combined with the fact that multiple amino acid substitutions occur in major antigenic sites and the evolutionary rates of both the pre- and post-2015 GII.4 VP1 were comparable to the evolutionary rate of GII.4 VP1 that was reported previously higher than other 'static' genotypes (Parra et al. 2017), it is inferred that the evolutionary pattern of GII.4[P16] was thought to be similar to prior epidemic GII.4 variants, accumulating mutations in the P2 domain that create a new prevalent variant and changing host susceptibility. In fact, a recent study reported a novel GII.4[P16] variant with multiple mutations on the antigenic sites, GII.4 Hong Kong, has the potential to become the next pandemic variant (Chan et al. 2021). Continuous monitoring and further functional verification are needed.

In conclusion, our finding suggested that although genetic diversities were observed both in the GII.2[P16] and GII.4[P16] noroviruses, the evolutionary patterns of the two viruses are different. The GII.P16 RdRps that associated with GII.2 and GII.4 Sydney VP1 exhibit different evolutionary patterns, and GII.4[P16] may generate potential novel epidemic variants through the accumulation of mutations in the P2 domain, while GII.2[P16] may evolve by changing its properties of replication, such as fidelity and efficiency. It is important to monitor variants with amino acid substitutions in the major antigenic sites of VP1 and the conserved motifs of polymerase in GII.2[P16] and GII.4[P16] noroviruses.

4. Materials and methods

4.1 Sequences retrieval

All the available nearly complete (>95 per cent) GII.2 VP1 sequences through the history and GII.4 VP1 sequences since 2016, as well as all the complete GII.P16 RdRp sequences with specific sampling date and geographic distribution, were downloaded

from the NCBI GenBank Database (<http://www.ncbi.nlm.nih.gov/genbank>). Low-quality sequences with Ns or gaps were removed. In addition, the complete VP1 of GII.4 Sydney before 2016 and several other common P-types of RdRp were also downloaded from the NCBI. The sequences were aligned by MAFFT 7 (online version, <https://mafft.cbrc.jp/alignment/software/>) and then adjusted manually by MEGA 7 software. All the analyzed sequences were annotated by GenBank number, genotype, sampling time, and location.

4.2 Genetic diversity calculation and amino acid logo analysis

SNP calling and mutation linkage analysis were completed by R (4.1.2). The relative frequencies of amino acid occurrence (bits) of the proteins were visualized using sequence logos online version (<http://weblogo.berkeley.edu/logo.cgi>). The informative sites on the major antigenic sites on the GII.4 VP1 can also be visualized in the amino acid logos.

4.3 Selective pressure analysis

Diversities of the genes were quantified as the mean genetic distances that were calculated for the pairs of nucleotide sequences using MEGA software. The ratio of the number of nonsynonymous substitutions per nonsynonymous site (dN) and synonymous substitutions per synonymous site (dS) was calculated. The dN/dS ratio is an indicator of the strength of positive (>1) or negative (<1) or neutral (=1) selection pressure on specific proteins. Moreover, statistically supported positively selected sites were localized through the Hyphy package. The MEME, FUBAR, and SLAC were applied for the analysis, and the results from the three models were merged.

4.4 Temporal-scaled distribution

The aligned nucleotide sequences for the GII.2 VP1 and GII.4 Sydney 2012 VP1 genes were used for phylogeny analysis. The phylogenetic trees were generated by MEGA 7.0.26 software, using the maximum likelihood method. Since the rates of transitions are two times higher than those of transversions for both the GII.2 VP1 and GII.4 Sydney 2012 VP1 genes, Kimura two-parameter substitution is the best model. Sequences of the early GII.2 VP1 in the 1970s and the previously reported epidemic GII.4 norovirus strains (New Orleans, Den Haag, Hunter, Farmington, Grimsby, and Osaka) were used as roots for the GII.2 and GII.4 VP1 trees, respectively.

4.5 Evolutionary rate estimation

Using BEAST software (version 2.6.3), the Bayesian Markov Chain Monte Carlo (MCMC) approach was implemented to precisely estimate the substitution rates of GII.2 and GII.4 Sydney VP1 and GII.P16 RdRp genes. The best fit evolutionary model was estimated by the Iqtree package (version 1.6.12). The running results were presented and analyzed by Tracer v.1.7.1. The effective sample size values for all the estimated parameters in the MCMC were >200. Statistical uncertainty of the data parameter values was reflected by the 95 per cent HPD values.

4.6 Construction and analysis of tertiary structures

By using Phyre 2 based on the template of NV GII.4 strain 60UU (c60uuB), the VP1 models were constructed for the wild type of

GII.4 Sydney 2012 (KY947550.1, USA, 2015) and for the different mutants with amino acid substitutions located at the major antigenic sites. The amino acids in the immunodominant motifs of the wild type and the mutants were plotted onto P-domain homology models to illustrate tertiary structure changes caused by sequence variations.

Data availability

Data are available in the supplementary material.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Funding

This work was supported by the Fundamental Research Funds for the Central Universities (2020RC016).

Conflict of interest: None declared.

References

- Ahmed, S. M. et al. (2014) 'Global Prevalence of Norovirus in Cases of Gastroenteritis: A Systematic Review and Meta-analysis', *The Lancet Infectious Diseases*, 14: 725–30.
- Ao, Y. et al. (2018) 'Genetic Analysis of Reemerging GII.P16-GII.2 Noroviruses in 2016–2017 in China', *The Journal of Infectious Diseases*, 218: 133–43.
- et al. (2017) 'Norovirus GII.P16/GII.2-Associated Gastroenteritis, China, 2016', *Emerging Infectious Diseases*, 23: 1172–5.
- Arias, A. et al. (2016) 'Norovirus Polymerase Fidelity Contributes to Viral Transmission in Vivo', *mSphere*, 1(5): e00279–16.
- Barclay, L. et al. (2019) 'Emerging Novel GII.P16 Noroviruses Associated with Multiple Capsid Genotypes', *Viruses*, 11: 535.
- Bartsch, S. M. et al. (2016) 'Global Economic Burden of Norovirus Gastroenteritis', *PLoS One*, 11: e0151219.
- Cannon, J. L. et al. (2017) 'Genetic and Epidemiologic Trends of Norovirus Outbreaks in the United States from 2013 to 2016 Demonstrated Emergence of Novel GII.4 Recombinant Viruses', *Journal of Clinical Microbiology*, 55: 2208–21.
- Chan, M. C. et al. Noropatrol for. (2021) 'Detection of Norovirus Variant GII.4 Hong Kong in Asia and Europe, 2017–2019', *Emerging Infectious Diseases*, 27: 289–93.
- Cheung, S. K. C. et al. (2019) 'Higher Viral Load of Emerging Norovirus GII.P16-GII.2 Than Pandemic GII.4 And Epidemic GII.17, Hong Kong, China', *Emerging Infectious Diseases*, 25: 119–22.
- Chhabra, P. et al. (2019) 'Updated Classification of Norovirus Genogroups and Genotypes', *Journal of General Virology*, 100: 1393–406.
- et al. (2020) 'Corrigendum: Updated Classification of Norovirus Genogroups and Genotypes', *Journal of General Virology*, 101: 893.
- Conley, M. J. et al. (2019) 'Calicivirus VP2 Forms a Portal-like Assembly following Receptor Engagement', *Nature*, 565: 377–81.
- Debbink, K. et al. (2012) 'Genetic Mapping of a Highly Variable Norovirus GII.4 Blockade Epitope: Potential Role in Escape from Human Herd Immunity', *Journal of Virology*, 86: 1214–26.
- Deval, J. et al. (2017) 'Structure(s), Function(s), and Inhibition of the RNA-Dependent RNA Polymerase of Noroviruses', *Virus Research*, 234: 21–33.
- Eden, J. S. et al. (2014) 'The Emergence and Evolution of the Novel Epidemic Norovirus GII.4 Variant Sydney 2012', *Virology*, 450–451: 106–13.

- Gorbalenya, A. E. et al. (2002) 'The Palm Subdomain-based Active Site Is Internally Permuted in Viral RNA-Dependent RNA Polymerases of an Ancient Lineage', *Journal of Molecular Biology*, 324: 47–62.
- Hasing, M. E. et al. (2019) 'Changes in Norovirus Genotype Diversity in Gastroenteritis Outbreaks in Alberta, Canada: 2012–2018', *BMC Infectious Diseases*, 19: 177.
- Hernandez, J. M. et al. (2020) 'Evolutionary and Molecular Analysis of Complete Genome Sequences of Norovirus from Brazil: Emerging Recombinant Strain GII.P16/GII.4', *Frontiers in Microbiology*, 11: 1870.
- Hoa Tran, T. N. et al. (2013) 'Molecular Epidemiology of Noroviruses Associated with Acute Sporadic Gastroenteritis in Children: Global Distribution of Genogroups, Genotypes and GII.4 Variants', *Journal of Clinical Virology*, 56: 185–93.
- Hunt, R. C. et al. (2014) 'Exposing Synonymous Mutations', *Trends in Genetics*, 30: 308–21.
- Hutson, A. M. et al. (2003) 'Norwalk Virus-like Particle Hemagglutination by Binding to H Histo-blood Group Antigens', *Journal of Virology*, 77: 405–15.
- Iritani, N. et al. (2012) 'Increase of GII.2 Norovirus Infections during the 2009–2010 Season in Osaka City, Japan', *Journal of Medical Virology*, 84: 517–25.
- Jiang, X. et al. (1993) 'Sequence and Genomic Organization of Norwalk Virus', *Virology*, 195: 51–61.
- Jin, M. et al. (2020) 'Norovirus Outbreak Surveillance, China, 2016–2018', *Emerging Infectious Diseases*, 26: 437–45.
- Kendra, J. A. et al. (2021) 'Antigenic Cartography Reveals Complexities of Genetic Determinants that Lead to Antigenic Differences among Pandemic GII.4 Noroviruses', *Proceedings of the National Academy of Sciences*, 118: e2015874118.
- Kristofich, J. et al. (2018) 'Synonymous Mutations Make Dramatic Contributions to Fitness When Growth Is Limited by a Weak-link Enzyme', *PLoS Genetics*, 14: e1007615.
- Li, X. et al. (2021) 'Molecular Evolution of Human Norovirus GII.2 Clusters', *Frontiers in Microbiology*, 12: 655567.
- Lindesmith, L. C. et al. (2013) 'Emergence of a Norovirus GII.4 Strain Correlates with Changes in Evolving Blockade Epitopes', *Journal of Virology*, 87: 2803–13.
- et al. (2008) 'Mechanisms of GII.4 Norovirus Persistence in Human Populations', *PLoS Medicine*, 5: e31.
- Lopman, B. A. et al. (2016) 'The Vast and Varied Global Burden of Norovirus: Prospects for Prevention and Control', *PLoS Medicine*, 13: e1001999.
- Lun, J. H. et al. (2018) 'Recombinant GII.P16/GII.4 Sydney 2012 Was the Dominant Norovirus Identified in Australia and New Zealand in 2017', *Viruses*, 10: 548.
- Mallory, M. L. et al. (2020) 'Bile Facilitates Human Norovirus Interactions with Diverse Histoblood Group Antigens, Compensating for Capsid Microvariation Observed in 2016–2017 GII.2 Strains', *Viruses*, 12: 989.
- Motoya, T. et al. (2017) 'Molecular Evolution of the VP1 Gene in Human Norovirus GII.4 Variants in 1974–2015', *Frontiers in Microbiology*, 8: 2399.
- Nagasawa, K. et al. (2018) 'Genetic Analysis of Human Norovirus Strains in Japan in 2016–2017', *Frontiers in Microbiology*, 9: 1.
- Parra, G. I. et al. (2017) 'Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity', *PLoS Pathogens*, 13: e1006136.
- Siebenga, J. J. et al. (2009) 'Norovirus Illness Is a Global Problem: Emergence and Spread of Norovirus GII.4 Variants, 2001–2007', *The Journal of Infectious Diseases*, 200: 802–12.
- Smertina, E. et al. (2019) 'Calicivirus RNA-Dependent RNA Polymerases: Evolution, Structure, Protein Dynamics, and Function', *Frontiers in Microbiology*, 10: 1280.
- Subba-Reddy, C. V., Goodfellow, I., and Kao, C. C. (2011) 'VPg-Primed RNA Synthesis of Norovirus RNA-Dependent RNA Polymerases by Using a Novel Cell-based Assay', *Journal of Virology*, 85: 13027–37.
- Subba-Reddy, C. V. et al. (2012) 'Norovirus RNA Synthesis Is Modulated by an Interaction between the Viral RNA-Dependent RNA Polymerase and the Major Capsid Protein, VP1', *Journal of Virology*, 86: 10138–49.
- et al. (2017) 'Retraction for Subba-Reddy Et Al., "Norovirus RNA Synthesis Is Modulated by an Interaction between the Viral RNA-Dependent RNA Polymerase and the Major Capsid Protein, VP1"', *Journal of Virology*, 91: e01708–17.
- Sukhrie, F. H. et al. (2013) 'P2 Domain Profiles and Shedding Dynamics in Prospectively Monitored Norovirus Outbreaks', *Journal of Clinical Virology*, 56: 286–92.
- Thongprachum, A. et al. (2017) 'Emergence of Norovirus GII.2 and Its Novel Recombination during the Gastroenteritis Outbreak in Japanese Children in Mid-2016', *Infection, Genetics and Evolution*, 51: 86–8.
- Tohma, K. et al. (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.
- et al. (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.
- et al. (2021) 'Genome-Wide Analyses of Human Noroviruses Provide Insights on Evolutionary Dynamics and Evidence of Coexisting Viral Populations Evolving under Recombination Constraints', *PLoS Pathogens*, 17: e1009744.
- White, P. A. (2014) 'Evolution of Norovirus', *Clinical Microbiology and Infection*, 20: 741–5.