Preprints.org

Article

Molecular Evolutionary Analyses of the RNA-Dependent RNA Polymerase (RdRp) Region and VP1 Gene in Human Norovirus Genotypes GII.P6-GII.6 and GII.P7-GII.6

Tomoko Takahashi , Ryusuke Kimura , Tatsuya Shirai , Mitsuru Sada , Toshiyuki Sugai , Kosuke Murakami , Kazuhiko Harada , Kazuto Ito , Yuki Matsushima , Fuminori Mizukoshi , Kaori Okayama , Yuriko Hayashi , Mayumi Kondo , Tsutomu Kageyama , Yoshiyuki Suzuki , Haruyuki Ishii , Akihide Ryo , Kazuhiko Katayama , Kiyotaka Fujita , Hirokazu Kimura [*]

Article

Molecular Evolutionary Analyses of the RNA-Dependent RNA Polymerase (*RdRp*) Region and *VP1* Gene in Human Norovirus Genotypes GII.P6-GII.6 and GII.P7-GII.6

Tomoko Takahashi [1,2], Ryusuke Kimura [3,4], Tatsuya Shirai [3,5], Mitsuru Sada [3,5], Toshiyuki Sugai [6], Kousuke Murakami [7], Kazuhiko Harada [3], Kazuto Ito [3], Yuki Matsushima [8], Fuminori Mizukoshi [9], Kaori Okayama [1], Yuriko Hayashi [1], Mayumi Kondo [10], Tsutomu Kageyama [11], Yoshiyuki Suzuki [12], Haruyuki Ishii [5], Akihide Ryo [13], Kazuhiko Katayama [14], Kiyotaka Fujita [1] and Hirokazu Kimura [1,3,10,*]

[1] Department of Health Science, Graduate School of Health Sciences, Gunma Paz University, Takasaki-shi, Gunma 370-0006, Japan

[2] Iwate Prefectural Research Institute for Environmental Science and Public Health, Morioka-shi, Iwate 020-0857, Japan

[3] Advanced Medical Science Research Center, Gunma Paz University Research Institute, Shibukawa-shi, Gunma 377-0008, Japan

[4] Department of Bacteriology, Graduate School of Medicine, Gunma University, Maebashi-shi, Gunma 371-8514, Japan

[5] Department of Respiratory Medicine, School of Medicine, Kyorin University, Mitaka-shi, Tokyo 181-8611, Japan

[6] Department of Nursing Science, Graduate School of Health Science, Hiroshima University, Hiroshima-shi, Hiroshima 734-8551, Japan

[7] Department of Virology II, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan

[8] Caliciviruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[9] Department of Microbiology, Tochigi Prefectural Institute of Public Health and Environmental Science, Utsunomiya-shi, Tochigi 329-1196, Japan

[10] Department of Clinical Engineering, Faculty of Medical Technology, Gunma Paz University, Takasaki-shi, Gunma 370-0006, Japan

[11] Center for Emergency preparedness and Response, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan

[12] Division of Biological Science, Department of Information and Basic Science, Graduate School of Natural Sciences, Nagoya City University, Nagoya-shi, Aichi 467-8501, Japan

[13] Department of Virology III, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan

[14] Laboratory of Viral Infection Control, Graduate School of Infection Control Sciences, Ōmura Satoshi Memorial Institute, Kitasato University, Minato-ku, Tokyo 108-8641, Japan

* Correspondence: kimhiro@nih.go.jp; Tel.: +81-27-386-5648; Fax: +81-27-386-8594

Abstract: To understand the evolution of the human norovirus GII.P6-GII.6 and GII.P7-GII.6 strains, we analyzed both the *RdRp* region and *VP1* gene in globally collected strains using authentic bioinformatics technologies. A common ancestor of the P6- and P7-type *RdRp* region emerged approximately 50 years ago and a common ancestor of the P6- and P7-type *VP1* gene emerged approximately 110 years ago. Subsequently, the *RdRp* region and *VP1* gene evolved. Moreover, the evolutionary rates were significantly faster for the P6-type *RdRp* region and *VP1* gene than the P7-type *RdRp* region and *VP1* genes. Large genetic divergence was observed in the P7-type *RdRp* region and *VP1* gene compared with the P6-type *RdRp* region and *VP1* gene. The phylodynamics of the *RdRp* region and *VP1* gene fluctuated after the year 2000. Positive selection sites in VP1 proteins were located in the antigenicity-related protruding 2 domain, and these sites overlapped with conformational epitopes. These results suggest that the GII.6 *VP1* gene and VP1 proteins evolved uniquely due to recombination between the P6- and P7-type *RdRp* regions in the HuNoV GII.P6-GII.6 and GII.P7-GII.6 virus strains.

1. Introduction

Human norovirus (HuNoV) is a major causative agent of acute gastroenteritis in humans of all ages [1,2]. Previous epidemiological data suggest that HuNoV may be associated with 30-60% of patients with gastroenteritis [3–5]. Moreover, this agent has caused large outbreaks of food poisoning worldwide [6,7]. However, effective vaccines and antiviral agents are not available at present [7]. Therefore, this agent may be a public health concern [8].

The HuNoV genome is a single-stranded plus-sense RNA with approximately 7.5 kb of nucleotide sequence [9]. The genome contains three open reading frames (ORFs): ORF1, ORF2, and ORF3 [9]. ORF1 encodes six nonstructural proteins designated as nonstructural proteins (NS) 1/2–7 [9]. Of these, the NS7 region encodes the RNA-dependent RNA polymerase (RdRp) protein, while ORF2 and ORF3 encode structural proteins, such as viral protein (VP) 1 and VP2, respectively [7,9]. The VP1 protein acts as an antigen and also shows large antigenic variations [6], although it is not exactly known.

Previous genetic and molecular epidemiological studies have suggested that the HuNoV genome shows large genetic divergence [8]. Currently, HuNoV is classified into two genogroups, genogroup I (GI) and genogroup II (GII) [10]. Furthermore, the GI and GII HuNoVs are classified into many genotypes. Such genotypes are based on the *RdRp* coding region (*RdRp* region) and *VP1* gene sequences [10]. To date, 60 *RdRp* (P-types) and 49 *VP1* genotypes have been confirmed [10]. Moreover, recombination occurs relatively frequently between ORF1 and ORF2 [11,12], resulting in many chimeric viruses acting as recombinants [12]. Based on this evidence, HuNoV can represent both the *RdRp* region genotype and *VP1* genotypes, such as GII.P6 (*RdRp* genotype)-GII.6 (*VP1* genotype) [13]. However, the role of these chimeric viruses remains unclear.

Molecular epidemiological data on HuNoV infections in humans suggest that certain GI and GII genotypes are prevalent [14]. These reports also show that GII HuNoV is more dominant than GI HuNoV [14]. Of these, some GII genotypes corresponding to *VP1* genotypes, such as GII.2, GII.3, GII.4, GII.6, and GII.17, are prevalent types [15]. However, these epidemiological data may not explain the reasons for the HuNoV epidemics.

Recently, authentic bioinformatic technologies have been used in population genetics, including the study of the evolution of various viruses [16]. Indeed, these methods may allow us to estimate the phylogeny, genome population, and antigenicity using three-dimensional antigen structures. Information that reflects viral evolution may contribute to a better response to these questions. To date, we have studied the molecular evolution of chimeric HuNoVs, such as GII.P17-GII.17, GII.P2-GII.2, and GII.P16-GII.2 [17–19]. However, such studies have not been performed on other GII genotypes to better understand GII HuNoV. Therefore, in this study, we performed a comprehensive molecular analysis of globally collected HuNoV GII.P6-GII.6 and GII.P7-GII.6 strains.

2. Materials and Methods

2.1. Strains used in this study

To analyze the molecular evolution of HuNoV GII.6, the complete genome sequences of HuNoV were downloaded from GenBank (last accessed on December 28, 2022). In total, 11,810 strains were collected. They were classified into genotypes using the Norovirus Typing Tool (Ver.2.0), and GII.6 strains were selected [10]. HuNoV GII.6 collected from each local government public health institution was added to the dataset. Among them, strains with an uncertain sequence or an unclear year of collection or area were excluded. Finally, 141 strains belonging to GII.6 remained and were used to analyze the molecular evolution of VP1. Similarly, 141 strains belonging to HuNoV GII.6

were obtained and used to analyze the molecular evolution of *RdRp* region. Details of the strains used in this study are presented in Supplementary Table 1.

2.2. Time-scaled phylogenetic analyses

To evaluate the molecular evolution of the present strains, phylogenetic trees of the HuNoV *RdRp* region (1,530 bp, excluding the stop codon) and the *VP1* gene (1,641–1,650 bp, excluding the stop codon) were constructed using the Bayesian Markov chain Monte Carlo (MCMC) method in the BEAST package (v.2.6.7), as previously described [20,21]. First, the jModelTest2 program was used to determine the suitable substitution models [22]. Second, the path-sampling/stepping-stone sampling marginal likelihood estimation method was used to evaluate the best of the four clock models (strict clock, relaxed clock exponential, relaxed clock log normal, and random local clock) and the two prior tree models (coalescent constant population and coalescent exponential population). Although these were performed independently for *VP1* gene and *RdRp* region analyses, SYM-Γ-I, Relaxed clock exponential, and Coalescent exponential population were selected for the molecular evolutionary analysis of *VP1* gene. However, SYM-Γ, Relaxed clock exponential, and Coalescent exponential population were adopted for the molecular evolutionary analysis of *RdRp*. The lengths of the Bayesian MCMC chains and samples are listed in Table S2. Effective sample sizes (ESS) were calculated using Tracer and the convergence of all parameters was confirmed if the ESS was greater than 200. After a 10% burn-in, phylogenetic trees were generated using TreeAnnotator (v.2.6.7) and rendered using FigTree (v.1.4.0). Molecular evolutionary rates were estimated using suitable models selected for each dataset, as described above. Statistical analyses were performed using the Kruskal-Wallis *t*-test for EZR [23].

2.3. Phylogenetic distance analyses

To calculate the phylogenetic distances among the strains, we used MEGA7 software [24]. The best substitution models were estimated using the jModelTest2 program. The phylogenetic distances between the present GII.6 strains were calculated from the pairwise Maximum Likelihood (ML) tree of the ML tree using the Patristic program [25].

2.4. Phylodynamic analyses

To assess the phylodynamics of the GII.6 strains, the effective population sizes of the *RdRp* region and *VP1* gene were calculated using Bayesian skyline plot (BSP) analysis implemented in the BEAST package [20]. Similar to the Bayesian MCMC method, the best substitution and clock models were selected. A Bayesian skyline plot and the 95% highest probability density (HPD) were visualized using Tracer [26].

2.5. Selective pressure analyses

The non-synonymous (*dN*) and synonymous (*dS*) substitution rates at each amino acid site were calculated to identify the selective pressure sites for the *RdRp* region and *VP1* gene using the Datamonkey server (https://www.datamonkey.org/) [27]. Five algorithms, single likelihood ancestor counting (SLAC); fixed-effects likelihood (FEL); internal fixed-effects likelihood (IFEL); the mixed effects model of evolution (MEME) method; and the fast, unconstrained Bayesian approximation (FUBAR) method, were used to identify positively selected sites, and all of them, except FUBAR, were used to detect negatively selected sites. The significance level was set at $p < 0.05$ for SLAC, FEL, IFEL, and MEME. Evidence of selective pressure for FUBAR was supported by a posterior probability > 0.9. In the positive selection analysis, sites common to more than four methods were regarded as positive selection sites, whereas in the negative selection analysis, sites common to more than three methods were considered negative selection sites.

2.6. Construction of the 3D structure of RdRp and VP1 proteins

To compare the VP1 and RdRp protein structures among genotypes, three-dimensional (3D) structural models of VP1 and RdRp proteins were constructed for each genotype using homology modeling. First, 3D structural models of VP1 in representative strains of each genotype (AB039777, LC122916, MH791993, MK956199, and JX989075) were generated using Protein Data Bank (PDB) ID: 6OTF as a template. Then, five models for each *VP1* genotype were generated using Modellar software (version 9.23) [28]. These models were evaluated by Ramachandran plot analysis using WinCoot implemented in the CCP4 package [29] and the best-scoring models were chosen. Finally, the energy of the selected models for each strain was minimized using GROMOS96 implemented in the Swiss PDB viewer (ver4.1.0) [30]. Using a similar procedure, the models of RdRp protein in each representative strain (AB039777, LC122916, LC760173, MK956199, and JX989075) (Table. S1) were constructed using the crystal structure of RdRp (PDB ID:1SH0) as a template.

2.7. Conformational B-cell epitope prediction

PDB files of the crystal structures of the GII.2 VP1 protein (PDB ID:6OTF) and FASTA files of their amino acid sequences were downloaded from PDB (https://pdbj.org/?lang=ja) to use as templates in the homology modeling method. To assess the conformational B-cell epitopes of the constructed VP1 protein models, four methods, DiscoTope 2.0 [31], ElliPro [32], SEMA [33], and SEPPA [34], were used with cutoff values of -3.7, 0.5, 0.76, and 0.064, respectively. Regions with amino acid sequences predicted by three or more of these methods and those contiguous with three or more residues were regarded as conformational epitopes. Furthermore, conformational epitopes were mapped onto the VP1 protein models constructed above.

3. Results

3.1. Time-scaled phylogeny of the RdRp region and VP1 gene in HuNoV GII.P6-GII.6 and GII.P7-GII.6

Time-scale phylogenetic trees were constructed based on the full-length nucleotides of the *RdRp* region and *VP1* gene using the Bayesian MCMC method. First, as shown in Figure 1A, a common ancestor of the P6- and P7-type *RdRp* regions diverged around December 1966 (mean; 95% HPDs, January 1941–March 1984). Subsequently, the P6- or P7-type *RdRp* regions further diverged and formed clusters 1 and 3, respectively. The main divergence times are shown in Figure 1A. The results suggested that a common ancestor of the P6- and P7-type *RdRp* region diverged approximately 50 years ago and evolved.

Next, as shown in Figure 1B, a common ancestor of the GII.6 *VP1* gene diverged around March 1904 (mean; 95% HPDs, September 1823–April 1962). Thereafter, the genes diverged to form four clusters. The main divergence times are shown in Figure 1B. Finally, the GII.6 strains having the P6-type formed only one cluster, while the GII.6 strains having the P7-type formed three independent clusters. Furthermore, this phylogenetic tree estimated that a common ancestor of the P6- and P7-type *VP1* genes diverged around November 1982 (mean; 95% HPDs, September 1970–February 1992). Thus, this time may be estimated as a recombination event between the GII.P6-GII.6 and GII.P7-GII.6 genomes in the present strains.
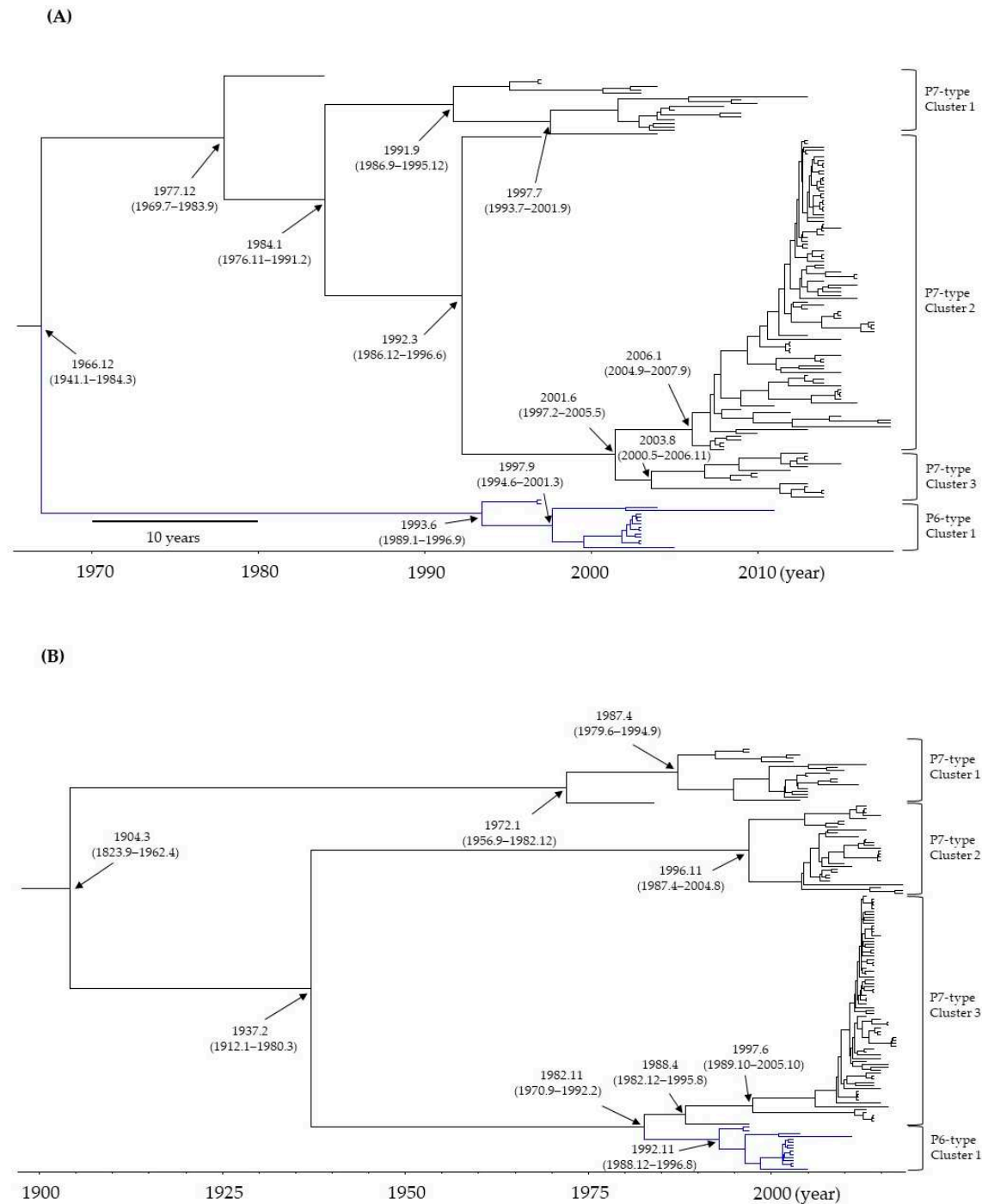
Figure 1. Time-scaled phylogenetic tree of the *RdRp* region and the *VP1* gene in GII.P6-GII.6 (15 strains) and GII.P7-GII.6 (126 strains) of the human norovirus (HuNoV) constructed using the Bayesian MCMC method. The phylogenetic trees of the *RdRp* region and *VP1* gene are shown in 1A and 1B, respectively. The divergence times with 95% highest probability densities (HPDs) are indicated on the phylogenetic tree.

3.2. Evolutionary rates of the RdRp region and VP1 gene in HuNoV GII.P6-GII.6 and GII.P7-GII.6

We also calculated the evolutionary rates using the Bayesian MCMC method. As shown in Table 1, the evolutionary rate was higher for GII.6 VP1 than the RdRp region, including P6- and P7-types (141 strains). The evolutionary rate was higher for the P6-type RdRp region than the P7-type RdRp. The evolutionary rate was higher for the P6-type GII.6 VP1 than the P7-type GII.6 VP1. These results

suggest that the RdRp region and VP1 gene in the present strains evolved independently, and the evolutionary rates were significantly distinct.

Table 1. Evolutionary rates of the present GII.6 strains.

Region/gene	Evolutionary rates (95% HPDs) (substitutions/site/year)	Compared groups and statistical values
All *RdRp* region (141 strains) P6-type 15 strains; P7-type 126 strains	3.287×10^{-3} (2.489×10^{-3}–4.098×10^{-3})	All *RdRp* region VS
All GII.6 *VP1* gene (141 strains) P6-type 15 strains; P7-type 126 strains	3.345×10^{-3} (2.295×10^{-3}–4.419×10^{-3})	All GII.6 *VP1* gene $p < 0.001$
P6-type *RdRp* region (15 strains)	5.063×10^{-3} (3.525×10^{-3}–6.595×10^{-3})	P6-type *RdRp* region VS P7-type *RdRp* region
P7-type *RdRp* region (126 strains)	3.022×10^{-3} (2.268×10^{-3}–3.775×10^{-3})	$p < 0.001$
GII.P6-GII.6 *VP1* gene (15 strains)	3.725×10^{-3} (1.843×10^{-3}–5.549×10^{-3})	GII.P6-GII.6 *VP1* gene VS GII. P7-GII.6 *VP1* gene
GII.P7-GII.6 *VP1* gene (126 strains)	3.482×10^{-3} (2.419×10^{-3}–4.568×10^{-3})	$p < 0.001$

3.3. Phylogenetic distances among the present strains

To assess the genetic divergence of the RdRp region and VP1 gene in the present strains, we calculated their phylogenetic distances. As shown in Figure 2A and 2D, the phylogenetic distances of the P6- and P7-type RdRp regions and the GII.6 VP1 gene were 0.112 ± 0.098 (mean ± 1 standard deviation [SD]) and 0.317 ± 0.259 (mean ± 1 SD). Statistically, the VP1 gene showed greater genetic divergence than the RdRp region (p < 0.001). Moreover, the genetic divergence was greater for the P7-type RdRp region than the P6-type RdRp region. The detailed statistical data are shown in Table 2.
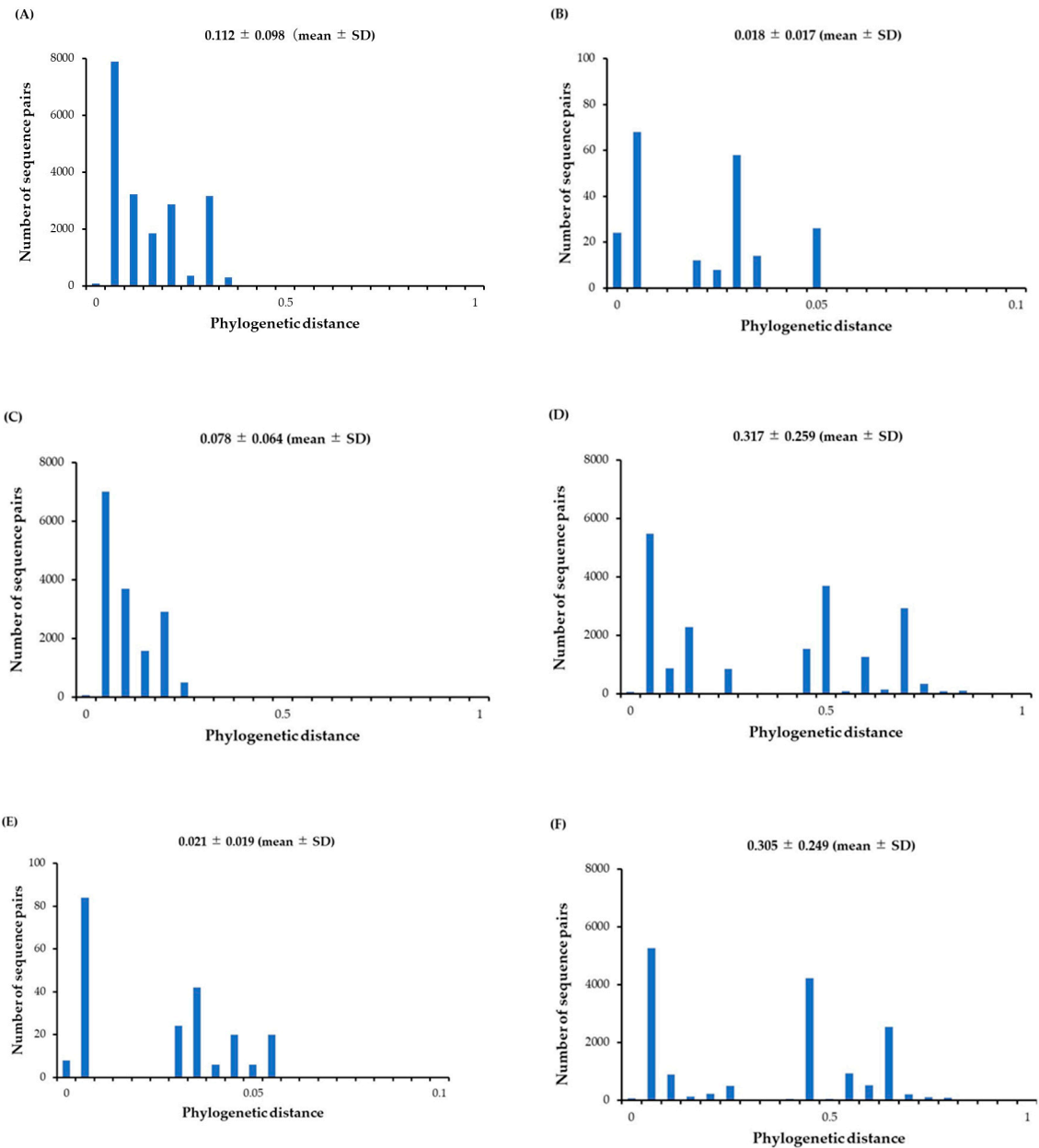
Figure 2. Distribution of phylogenetic distances between HuNoV GII.P6-GII.6 and GII.P7-GII.6 strains. (A) phylogenetic distances of the P6-type and P7-type RdRp regions; (B) P6-type RdRp region alone; (C) P7-type RdRp region alone; (D) GII.P6-GII.6 and GII.P7-GII.6 VP1 gene; (E) P6-type VP1 gene alone; (F) P7-type VP1 gene alone. The y-axis shows the number of sequence pairs for each distance, and the x-axis shows the phylogenetic distances.

Table 2. Phylogenetic distance of the present strains.

Region/gene	Phylogenetic distance (mean ± SD)	Compared groups and statistical values
All *RdRp* region (141 strains) P6-type 15 strains; P7-type 126 strains	0.112 ± 0.098	All *RdRp* region VS
All GII.6 *VP1* gene (141 strains) P6-type 15 strains; P7-type 126 strains	0.317 ± 0.259	ALL GII.6 *VP1* gene $p < 0.001$
P6-type *RdRp* region (15 strains)	0.018 ± 0.017	P6 type *RdRp* region VS
P7-type *RdRp* region (126 strains)	0.078 ± 0.064	P7 type *RdRp* region $p < 0.001$
GII.P6-GII.6 *VP1* gene (15 strains)	0.021 ± 0.019	GII.P6-GII.6 *VP1* gene VS
GII.P7-GII.6 *VP1* gene (126 strains)	0.305 ± 0.249	GII. P7-GII.6 *VP1* gene $p < 0.001$

3.4. Phylodynamics of GII.P6-GII.6 and GII.P7-GII.6

To assess the phylodynamics of the present GII.P6-GII.6 and GII.P7-GII.6 strains, we calculated time-scaled genome population sizes using the BSP method (Figure 3A-3F). Until approximately 2010, the genome population sizes of both the RdRp region and VP1 gene remained constant. However, significant fluctuations in genome population sizes were observed around 2010–2018.
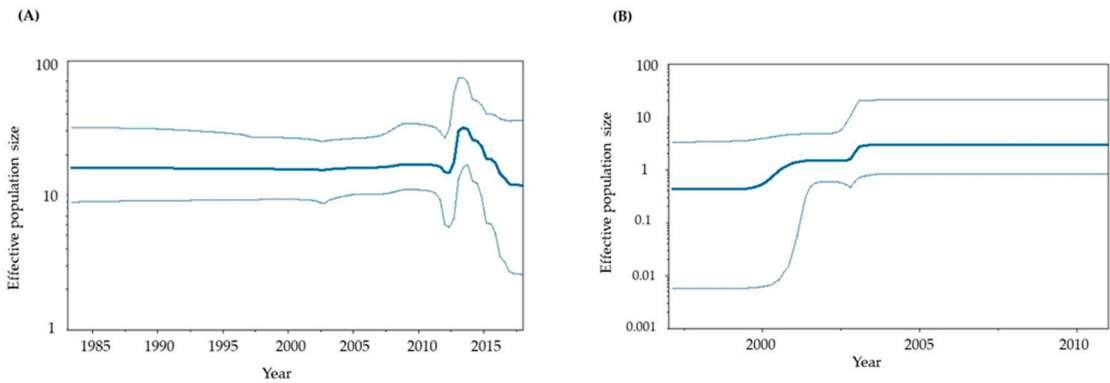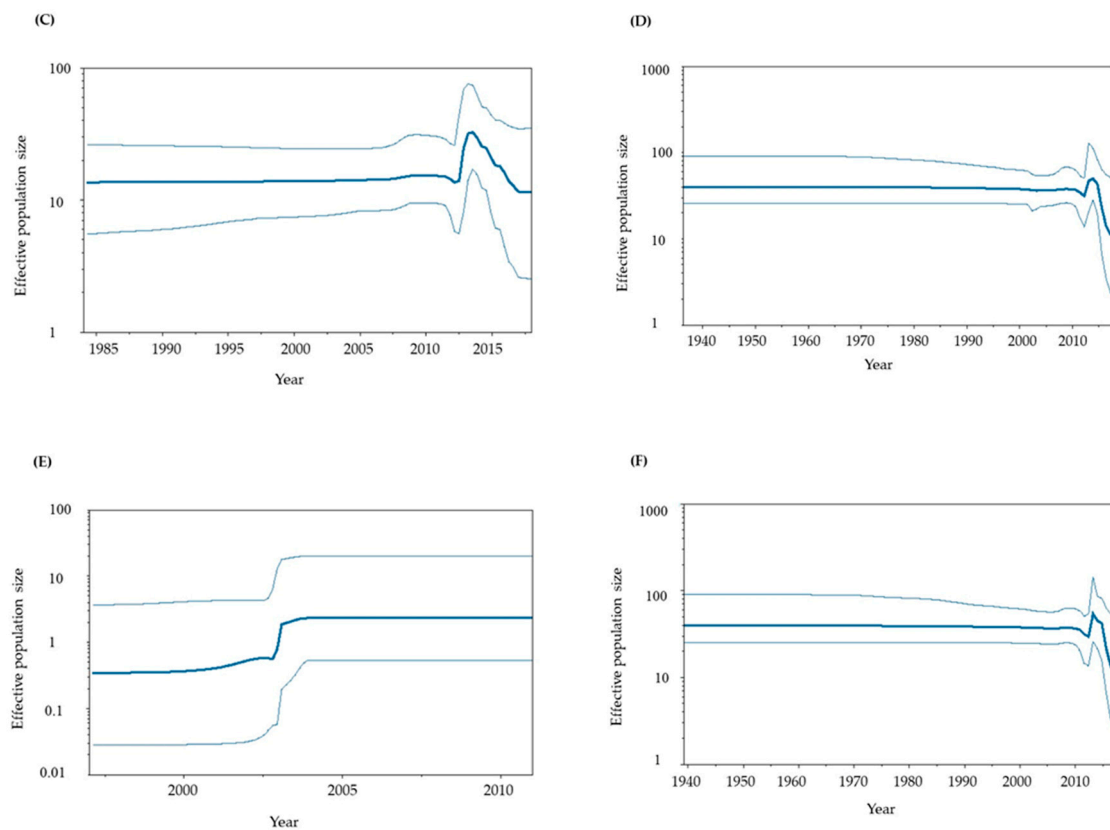
Figure 3. Phylodynamics of the present HuNoV GII.P6-GII.6 and GII.P7-GII.6 strains determined using Bayesian skyline plot analysis. (A) Phylodynamics of the P6-type and P7-type *RdRp* regions; (B) P6-type *RdRp* region; (C) P7-type *RdRp* region; (D) P6- and P7-type *VP1* genes; (E) P6-type *VP1* gene alone; (F) P7-type *VP1* gene alone. The y-axis shows the effective population size for each distance, and the x-axis represents time (years). The thick line in the center shows the median effective population sizes, and the thin lines at the top and bottom indicate the 95% HPDs.

3.5. Positive selection sites in the VP1 protein

We analyzed the positive selection sites in the VP1 protein to estimate selective pressure against the host. First, a few positively selected sites were identified. Of these, only Lys386His was predicted in the P6-type VP1 protein, whereas Pro354Thr, Pro354Ser, Pro354Gln, Asn390Thr, and Asn390Asp were predicted in the P7-type VP1 protein. These sites were located in the protruding 2 (P2) domain of the protein (Table S4). These results suggest that the GII.P7-GII.6 strains may receive stronger selection pressure from the host than the GII.P6-GII.6 strains.

3.6. Negative selection sites in RdRp and VP1 proteins

In general, negative selection sites may prevent the deterioration of protein function. Therefore, we calculated the number of negative-selection sites in these strains. Many negative selection sites were estimated for the P7-type RdRp protein (205 sites) and P7-type VP1 protein (274 sites). However, a small number of negative selection sites were estimated in the P6-type RdRp protein (3 sites) and P6-type VP1 protein (8 sites)(Table 3). Details of the negative selection sites are shown in Supplementary Tables 3A and 3 B.

Table 3. Number of amino acid residues of predicted positive and negative selection sites in HuNoVGII.6.

Region/gene	Number of negative selection sites	Number of positive selection sites	Estimated as positive selective sites
P6 type and P7 type *RdRp* region	258	0	—
P6 type *RdRp* region	3	0	—
P7 type *RdRp* region	205	1	126Lys, Lys126Arg
GII.P6-GII.6 and GII.P7-GII.6 *VP1* gene	298	2	354Pro, 390Asn
GII.P6-GII.6 *VP1* gene	8	1	Lys386His
GII.P7-GII.6 *VP1* gene	274	2	Pro354Thr, Ser and Gln, Asn390Thr and Asp

3.7. 3D mapping relationships between amino acid substitutions and active sites of the RdRp dimer proteins

To better assess the relationships between amino acid substitutions and the active sites of RdRp proteins, we constructed 3D RdRp dimers and mapped them. No amino acid substitutions were found in the RdRp active sites (aa182, 242, 243, 300, 309, 343, and 344) (Figures 4A–4E).
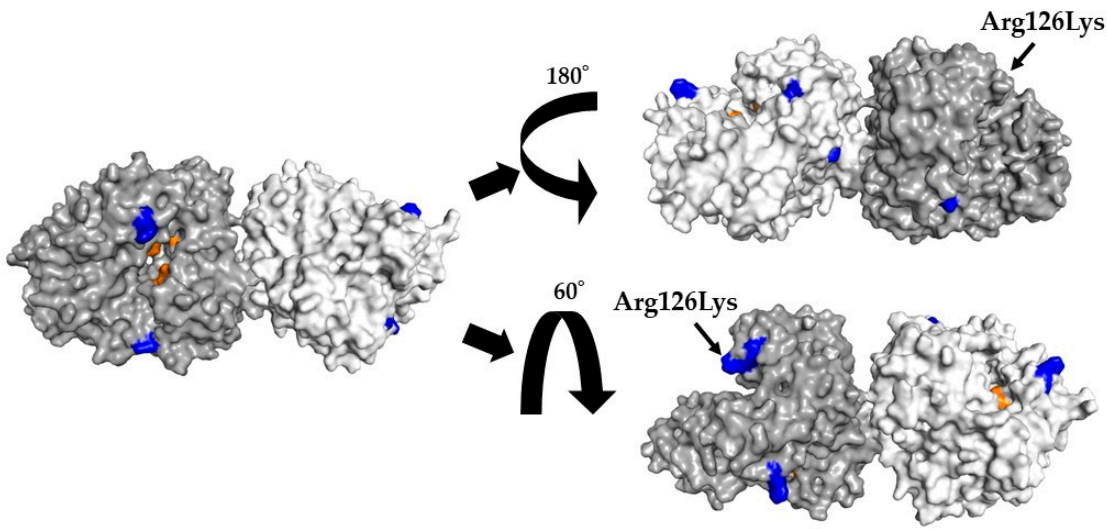
(A) Prototype: P7-type RdRp Protein (AB039777)
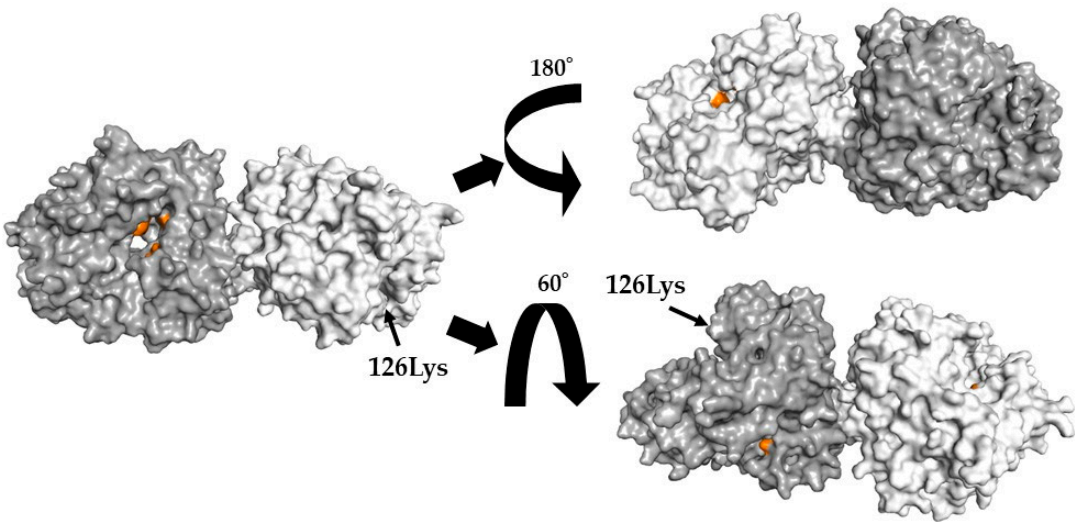
(B) P7-type RdRp Protein (LC122916)
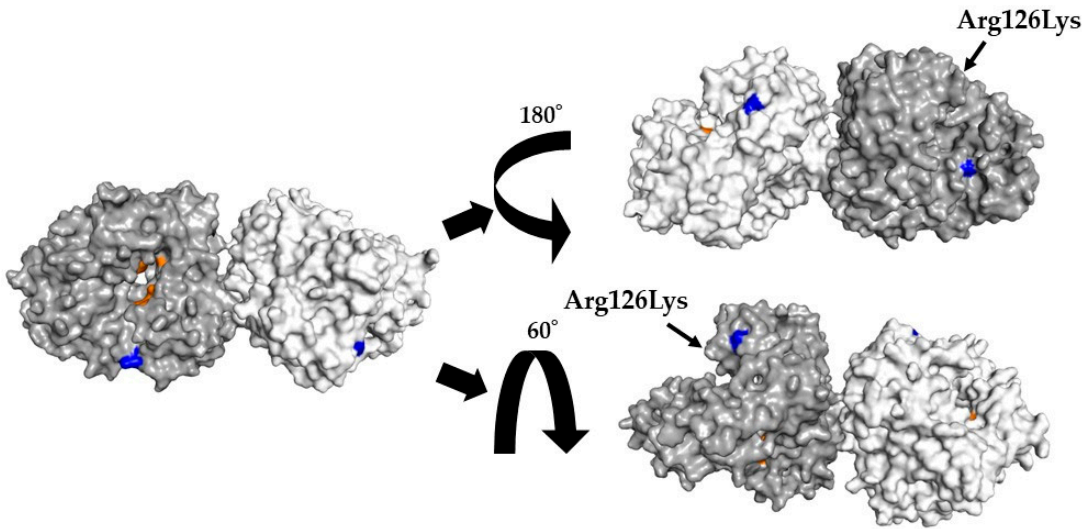
(C) P7-type RdRp Protein (MK956199)

(D) P7-type RdRp Protein (LC760173)
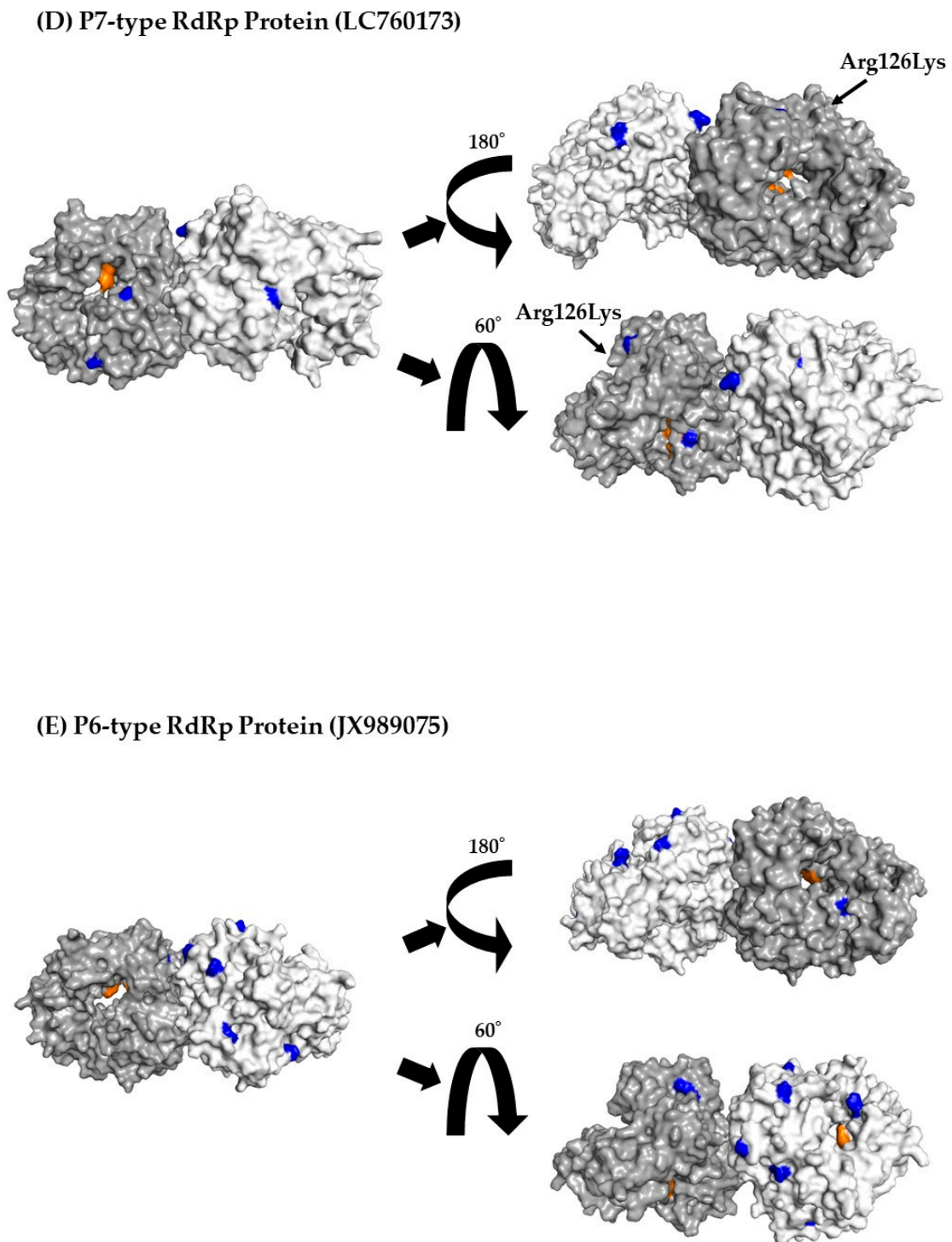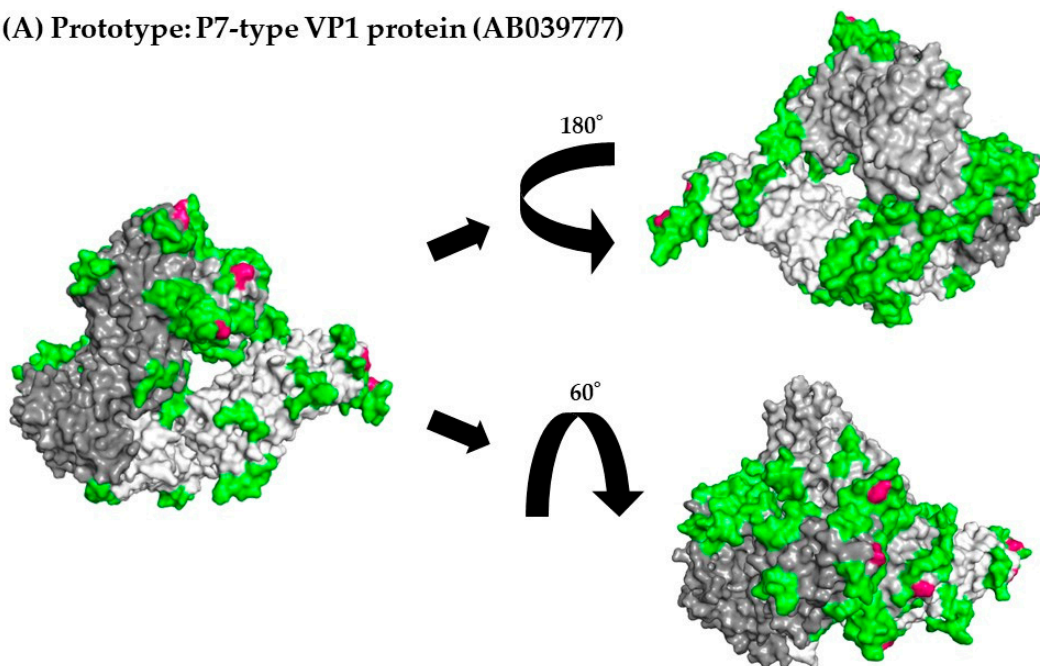

(E) P6-type RdRp Protein (JX989075)

Figure 4. Three-dimensional (3D) RdRp protein (dimer) structure and mapping of amino acid substitutions and active sites. Illustration shows the 3D structures of RdRp protein in the prototype and the most recent strain for each cluster. The strains in each figure are as follows: (A) a P7-type prototype strain (AB039777); (B) a P7-type strain (LC122916) in cluster 1; (C) a P7-type strain (MK956199) in cluster 2; (D) a P7-type strain (LC760173) in cluster 3; (E) a P6-type strain (JX989075) in cluster 1. The chains of the dimer structure are colored dark gray (chain A) and light gray (chain

B). Amino acid substitutions in each variant strain relative to the prototype strain are shown in blue, and the active sites are shown in orange.
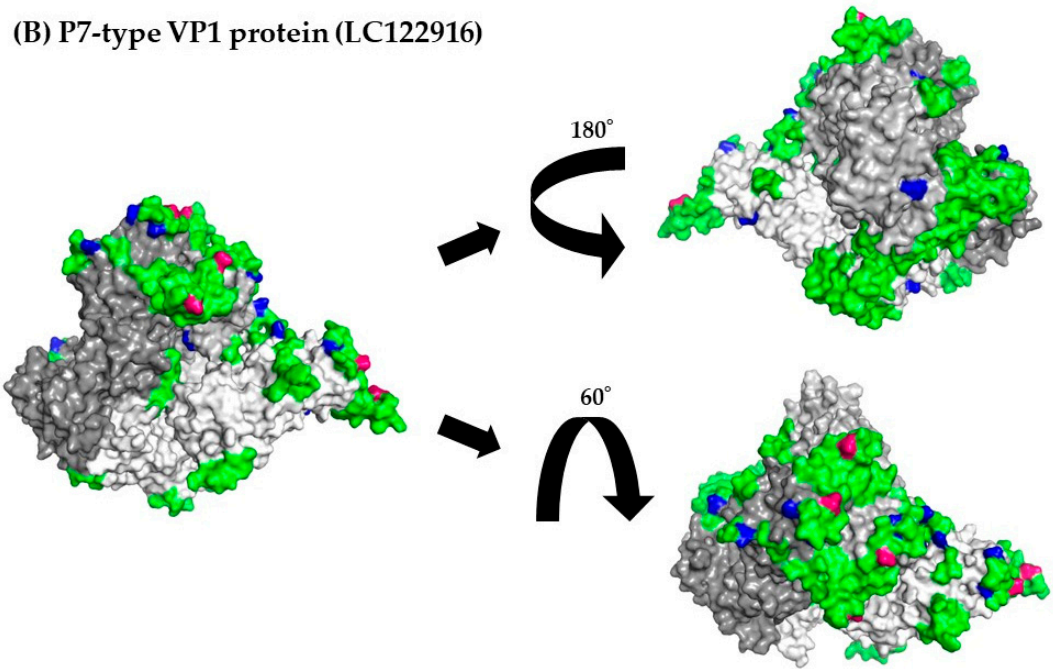
3.8. 3D mapping of the positive selection sites and conformational epitopes in the VP1 trimer proteins

Furthermore, to better evaluate the locations of positive selection sites and conformational epitopes on the VP1 protein, we constructed and mapped 3D VP1 trimer proteins. First, as shown in Figure 5A–5E and Table 3, and Section 3.5, positive selection sites for both P6- and P7-type VP1 proteins were located in the P2 domain. Of these, the positive selection sites in the P7-type VP1 proteins (Pro354Thr, Pro354Ser, Pro354Gln, Asn390Thr, and Asn390Asp) overlapped with some conformational epitopes (Table 3 and Table S4), whereas a positive selection site (Lys386His) in P6-type VP1 proteins did not. These results suggest that the positive selection sites in P7-type GII.6 VP1 proteins escaped amino acid mutations.
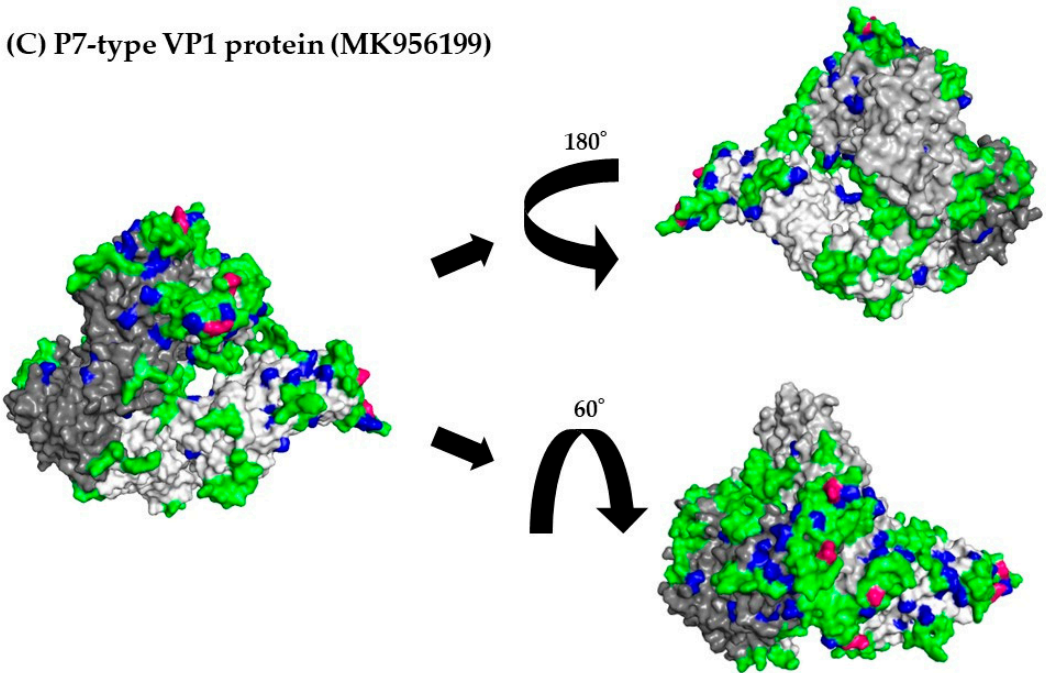

(A) Prototype: P7-type VP1 protein (AB039777)

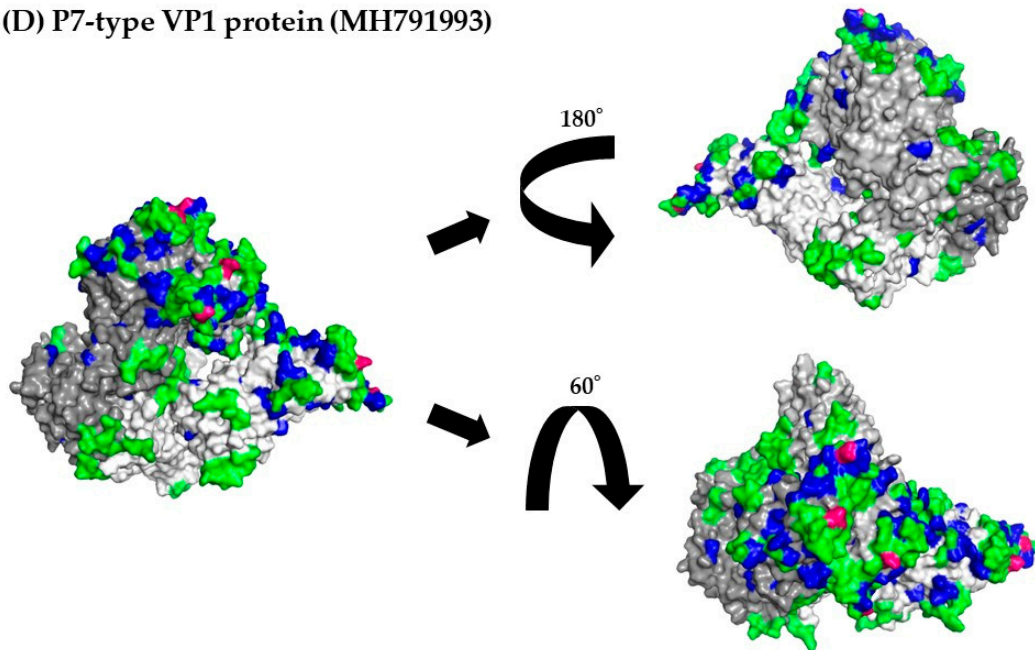**(B) P7-type VP1 protein (LC122916)**
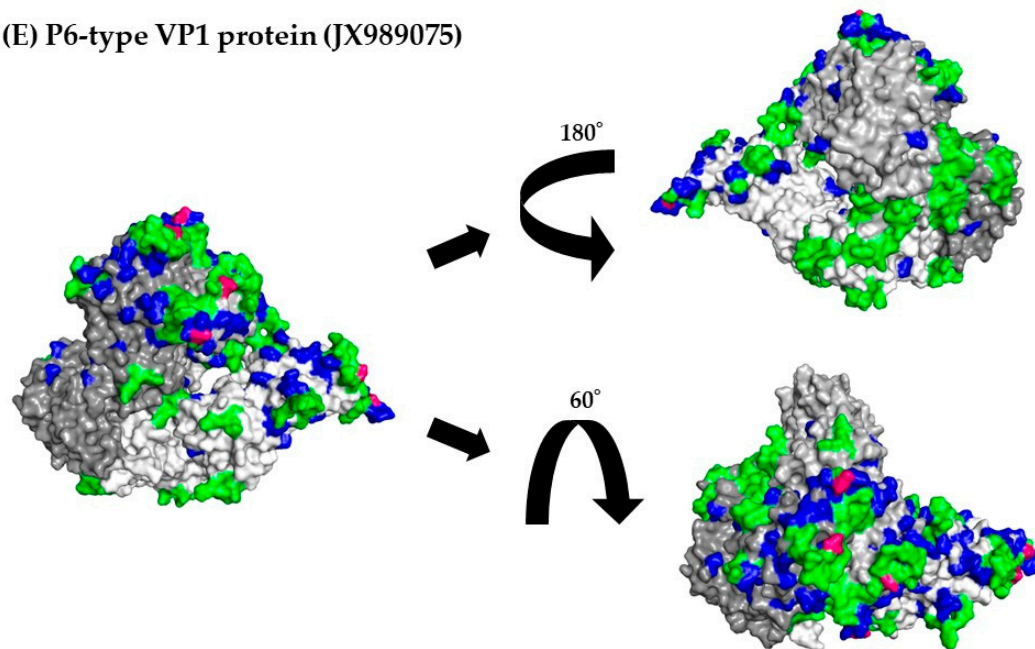

(C) P7-type VP1 protein (MK956199)

(D) P7-type VP1 protein (MH791993)

(E) P6-type VP1 protein (JX989075)

Figure 5. 3D mapping of the positive selection sites and conformational epitopes in the VP1 protein (trimer). Illustration shows the 3D structure of the VP1 protein in the prototype and the most recent strain for each cluster. The strains in each figure are as follows: (A) a P7-type prototype strain (AB039777); (B) a P7-type strain (LC122916) in cluster 1; (C) a P7-type strain (MK956199) in cluster 2; (D) a P7-type strain (MH791993) in cluster 3; (E) a P6-type strain (JX989075) in cluster 1. Chains of the trimeric structures are colored in dark gray (chain A), light gray (chain B), and white (chain C).

Conformational epitopes of each strain are indicated in green. Amino acid substitutions of each strain are indicated in blue. Positive selection sites are colored red. When amino acid substitutions overlapped with conformational epitopes, the amino acid substitutions were given priority and colored blue. The amino acid sequences and details are provided in Supplementary Table S4.

4. Discussion

To better understand the evolution of HuNoV GII.6 strains with different *RdRp* types (P6 and P7), we analyzed both the *RdRp* region and *VP1* gene using various authentic bioinformatics technologies. First, a time-scaled phylogenetic tree showed that a common ancestor of the P6- and P7-type *RdRp* region emerged approximately 50 years ago and uniquely evolved and formed clusters. A common ancestor of P6- and P7-type GII.6 *VP1* gene emerged approximately 110 years ago and formed clusters. The dominant type for both the *RdRp* region and *VP1* gene was P7-type (Figures 1A and 1B). Secondly, the evolutionary rates of both the P6-type *RdRp* region and *VP1* gene were faster than those of the P7-type *RdRp* region and *VP1* gene (Table 1). Next, the phylogenetic distances of the P7-type *RdRp* region and *VP1* gene were wider than those of the P6-type *RdRp* region and *VP1* gene. Furthermore, phylodynamic data showed that the *RdRp* region and *VP1* gene population sizes fluctuated after 2000 (Figure 3A–3F). Some positive selection sites in the VP1 proteins were estimated, and these were located in the antigenicity-related P2 domain. Among these, the positive selection sites in the P7-type VP1 protein overlapped with the conformational epitopes (Figure 5A–D and Table S4). These data imply that the GII.6 *VP1* gene and VP1 protein uniquely evolved because of recombination between the P6- and P7-type *RdRp* regions in the HuNoV GII.P6-GII.6 and GII.P7-GII.6 genomes.

A previous report regarding the evolutionary analyses of the *RdRp* region of various HuNoV genotypes showed that the P6- and P7-type *RdRp* region diverged from a common ancestor of other *RdRp* genotypes, including P18, P15, and P20 [35]. This report also estimated that the divergence year of the P6- and P7-types of the *RdRp* region was in the 1960s [35]. This may be compatible with the present data (December 1966). Moreover, the topologies of the previous time-scaled evolutionary tree and our tree were similar [35]. Although this and other reports did not show the evolutionary rates of each *RdRp* genotype, the evolutionary rates of various *RdRp* genotypes were estimated as 2.52×10^{-3} s/s/y to 3.12×10^{-3} s/s/y. The present data are also compatible with the data from a previous report [35]. These results suggested that the P6- and P7-type *RdRp* regions are genetically related. Next, the HuNoV *RdRp* region/RdRp protein may have affected the evolution of the *VP1* gene/VP1 protein [18,19]. As shown in Figure 1B, the phylogeny of the *VP1* gene in GII.P6-GII.6 and GII.P7-GII.6 was clearly divided and evolved uniquely. Previous reports have also suggested that recombination between the HuNoV genome ORF1, incorporating the *RdRp* region and ORF2, incorporating *VP1* gene, affects *VP1* gene/VP1 protein evolution [18,19]. For example, during the 2016/17 season, recombination between different lineages of the P16-type *RdRp* region in the GII.P16-GII.2 strains occurred and the recombinant caused large outbreaks of acute gastroenteritis in various countries [36–38]. Moreover, the GII.4 genotype caused a gastroenteritis pandemic between 2006 and 2012 [39,40]. Outbreaks may also be associated with recombination between ORF1 and ORF2 in GII.4 strains [41]. Based on previous and the present results, the prevalence of GII.P7-GII.2 strains was due to the recombination of P6- and P7-type *RdRp* regions. Moreover, the evolutionary rates of the *VP1* gene combined with P6- and P7-type *RdRp* regions were estimated as 5.063×10^{-3} s/s/y and 3.022×10^{-3} s/s/y, respectively. Previous data estimated the mean rates of various GII.2 genotype strains (GII.1 to GII.22) as 3.21×10^{-3} to 4.30×10^{-3} s/s/y [42]. Thus, these values and the present data may be similar [43]. Taken together, these findings provide information on the evolutionary history of these viral strains and suggest that recombination events may have played a pivotal role in their evolution.

Next, we estimated the genetic divergence of the P6- and P7-type *RdRp* regions and P6- and P7-type *VP1* genes in the present strains. First, a larger divergence of P7-type *RdRp* regions and P7-type *VP1* genes was estimated compared to that of the P6-type *VP1* gene. In the present study, the number of P6-type strains was relatively small (15 strains), although statistical analyses were performed.

17

Further studies regarding this may be needed after a greater number of strains with the same genotype are registered.

We also analyzed the phylodynamics of the *RdRp* region and *VP1* gene. The results showed that the genome population size of GII.P6-GII.6 increased around 2000–2003, while the genome population size of GII.P7-GII.6 increased after 2005. A previous molecular epidemiological study suggested that GII.6 had a biphasic prevalence between 2000 and 2005 and 2007 and 2010. Thus, the present phylodynamic data may reflect the prevalence of GII.P6-GII.6 and GII.P7-GII.6.

Moreover, to evaluate the functional and evolutionary characteristics of the P6- and P7-type RdRp proteins, we constructed 3D dimeric RdRp proteins and mapped them with amino acid substitutions (Figure 4). Several amino acid substitutions were also identified. Previous reports have suggested that some amino acid substitutions are associated with replication efficacy [8,41]. For example, the efficacy of HuNoV genome replication is increased by amino acid substitutions (291Thr or 291Val) in various RdRp proteins [43]. However, no substitution in the active sites were found in the P6- and P7-type RdRp proteins.

We also constructed P6- and P7-type 3D trimeric VP1 proteins (Figure 5). Previous reports have shown that the P2 domain may act not only as a host cell-binding site, but also as a major part of the HuNoV antigen [44,45]. Therefore, amino acid substitutions in this domain may be associated with infectivity and antigenicity [44,45]. Moreover, positively selected sites may function as escape mutations in the host [46]. In the present study, some conformational epitopes were identified in both RdRp-type VP1 proteins. Some of these were located in the P2 domain. Positively selected sites were also identified. Moreover, the amino acid positions of the conformational epitopes and positive selections between the P6- and P7-type VP1 proteins were distinct. These results implied that the antigenicity of P6- and P7-type VP1 proteins is distinct, although we did not examine this *in vitro*, and both VP1 proteins may receive selective pressure from host defense systems (i.e., host immunity) [45,47].

Finally, we evaluated negative selection sites for the RdRp and VP1 proteins (Table 3 and Table S3). Many negative selection sites in P7-type RdRp (205 sites) and VP1 proteins (274 sites) were estimated, while P6-type RdRp and VP1 proteins were small. In general, negative selection sites play a role in preventing the deterioration of viral protein function [46]. Thus, the present negative selection data may indicate the maintenance of RdRp and VP1 protein function. Furthermore, a small number of negative selection sites in P6-type RdRp and VP1 proteins were estimated. This may be because of the relatively small number of strains used in this study (15 strains).

5. Conclusions

In this study, to better understand the evolution of the HuNoV GII.P6-GII.6 and GII.P7-GII.6 strains, we performed a detailed analysis of both the *RdRp* region and *VP1* gene in these viruses using various bioinformatics methods. A common ancestor of the P6- and P7-type *RdRp* region emerged approximately 50 years ago and formed clusters. A common ancestor of the P6- and P7-type *VP1* gene emerged approximately 110 years ago. Moreover, both *RdRp* region and *VP1* gene have evolved uniquely. The evolutionary rates of the P6-type *RdRp* region and P6-type *VP1* gene were faster than the evolutionary rates of the P7-type *RdRp* region and *VP1* genes. More genetic divergence was observed in the P7-type *RdRp* region and *VP1* gene than in the P6-type *RdRp* region and *VP1* gene. The phylodynamics of the *RdRp* region and *VP1* fluctuated after 2000. Some positive selection sites in the VP1 proteins were located in the antigenicity-related P2 domain, and these sites in the P7-type VP1 protein overlapped with the conformational epitopes. Taken together, the GII.6 *VP1* and VP1 proteins evolved uniquely due to recombination between the P6- and P7-type *RdRp* regions in the HuNoV GII.P6-GII.6 and GII.P7-GII.6 strains.

S4. Detailed amino acid sequences of the *VP1* gene for representative and prototype strains of each cluster used in this study.

References

1.   Robilotti, E.; Deresinski, S.; Pinsky, B.A. Norovirus. *CLINICAL MICROBIOLOGY REVIEWS* **2015**, *28*, 134-164.
2.   Mans, J. Norovirus Infections and Disease in Lower-MiddleandLow-Income Countries, 1997‒2018. *Viruses* **2019**, *11*, doi:10.3390/v11040341.
3.   Chan-It, W.; Thongprachum, A.; Khamrin, P.; Kobayashi, M.; Okitsu, S.; Mizuguchi, M.; Ushijima, H. Emergence of a new norovirus GII.6 variant in Japan, 2008-2009. *J. Med. Virol.* **2012**, *84*, 1089-1096, doi:10.1002/jmv.23309.
4.   Patel, M.M.; Hall, A.J.; Vinjé, J.; Parashar, U.D. Noroviruses: a comprehensive review. *J. Clin. Virol.* **2009**, *44*, 1-8, doi:10.1016/j.jcv.2008.10.009.
5.   Patel, M.M.; Widdowson, M.A.; Glass, R.I.; Akazawa, K.; Vinjé, J.; Parashar, U.D. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerg. Infect. Dis.* **2008**, *14*, 1224-1231, doi:10.3201/eid1408.071114.
6.   Parrón, I.; Álvarez, J.; Jané, M.; Cornejo Sánchez, T.; Razquin, E.; Guix, S.; Camps, G.; Pérez, C.; Domínguez, À.; Catalonia, W.G.f.t.S.o.O.o.A.G.i. A foodborne norovirus outbreak in a nursing home and spread to staff and their household contacts. *Epidemiol. Infect.* **2019**, *147*, e225, doi:10.1017/S0950268819001146.
7.   Winder, N.; Gohar, S.; Muthana, M. Norovirus: An Overview of Virology and Preventative Measures. *Viruses* **2022**, *14*, doi:10.3390/v14122811.
8.   de Graaf, M.; van Beek, J.; Koopmans, M.P. Human norovirus transmission and evolution in a changing world. *Nat. Rev. Microbiol.* **2016**, *14*, 421-433, doi:10.1038/nrmicro.2016.48.
9.   Jiang, X.; Wang, M.; Wang, K.; Estes, M.K. Sequence and genomic organization of Norwalk virus. *Virology* **1993**, *195*, 51-61, doi:10.1006/viro.1993.1345.
10.   Chhabra, P.; de Graaf, M.; Parra, G.I.; Chan, M.C.; Green, K.; Martella, V.; Wang, Q.; White, P.A.; Katayama, K.; Vennema, H.; et al. Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* **2019**, *100*, 1393-1406, doi:10.1099/jgv.0.001318.
11.   Jiang, X.; Espul, C.; Zhong, W.M.; Cuello, H.; Matson, D.O. Characterization of a novel human calicivirus that may be a naturally occurring recombinant. *Arch. Virol.* **1999**, *144*, 2377-2387, doi:10.1007/s007050050651.
12.   Ludwig-Begall, L.F.; Mauroy, A.; Thiry, E. Norovirus recombinants: recurrent in the field, recalcitrant in the lab - a scoping review of recombination and recombinant types of noroviruses. *J. Gen. Virol.* **2018**, *99*, 970-988, doi:10.1099/jgv.0.001103.
13.   Kroneman, A.; Vega, E.; Vennema, H.; Vinjé, J.; White, P.A.; Hansman, G.; Green, K.; Martella, V.; Katayama, K.; Koopmans, M. Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* **2013**, *158*, 2059-2068, doi:10.1007/s00705-013-1708-5.
14.   van Beek, J.; de Graaf, M.; Al-Hello, H.; Allen, D.J.; Ambert-Balay, K.; Botteldoorn, N.; Brytting, M.; Buesa, J.; Cabrerizo, M.; Chan, M.; et al. Molecular surveillance of norovirus, 2005-16: an epidemiological analysis of data collected from the NoroNet network. *Lancet Infect. Dis.* **2018**, *18*, 545-553, doi:10.1016/S1473-3099(18)30059-8.
15.   Khamrin, P.; Kumthip, K.; Yodmeeklin, A.; Jampanil, N.; Phengma, P.; Yamsakul, P.; Okitsu, S.; Kobayashi, T.; Ushijima, H.; Maneekarn, N. Changing Predominance of Norovirus Recombinant Strains GII.2[P16] to GII.4[P16] and GII.4[P31] in Thailand, 2017 to 2018. *Microbiol. Spectr.* **2022**, *10*, e0044822, doi:10.1128/spectrum.00448-22.

16.   Pybus, O.G.; Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **2009**, *10*, 540-550, doi:10.1038/nrg2583.

17.   Matsushima, Y.; Mizukoshi, F.; Sakon, N.; Doan, Y.H.; Ueki, Y.; Ogawa, Y.; Motoya, T.; Tsukagoshi, H.; Nakamura, N.; Shigemoto, N.; et al. Evolutionary Analysis of the VP1 and RNA-Dependent RNA Polymerase Regions of Human Norovirus GII.P17-GII.17 in 2013–2017. *Front. Microbiol.* **2019**, *10*, 2189, doi:10.3389/fmicb.2019.02189.

18.   Mizukoshi, F.; Nagasawa, K.; Doan, Y.H.; Haga, K.; Yoshizumi, S.; Ueki, Y.; Shinohara, M.; Ishikawa, M.; Sakon, N.; Shigemoto, N.; et al. Molecular Evolution of the RNA-Dependent RNA Polymerase and Capsid Genes of Human Norovirus Genotype GII.2 in Japan during 2004-2015. *Front. Microbiol.* **2017**, *8*, 705, doi:10.3389/fmicb.2017.00705.

19.   Nagasawa, K.; Matsushima, Y.; Motoya, T.; Mizukoshi, F.; Ueki, Y.; Sakon, N.; Murakami, K.; Shimizu, T.; Okabe, N.; Nagata, N.; et al. Phylogeny and Immunoreactivity of Norovirus GII.P16-GII.2, Japan, Winter 2016-17. *Emerg. Infect. Dis.* **2018**, *24*, 144-148, doi:10.3201/eid2401.170284.

20.   Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS COMPUTATIONAL BIOLOGY* **2014**, *10*, e1003537-e1003537.

21.   Saito, M.; Tsukagoshi, H.; Sada, M.; Sunagawa, S.; Shirai, T.; Okayama, K.; Sugai, T.; Tsugawa, T.; Hayashi, Y.; Ryo, A.; et al. Detailed Evolutionary Analyses of the F Gene in the Respiratory Syncytial Virus Subgroup A. *Viruses* **2021**, *13*, doi:10.3390/v13122525.

22.   Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods.* **2012**, *9*, 772, doi:10.1038/nmeth.2109.

23.   Kanda Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 48 452–458. 10.1038/bmt.2012.244

24.   Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870-1874, doi:10.1093/molbev/msw054.

25.   Fourment M., Gibbs M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol. Biol.* 6:1. 10.1186/1471-2148-6-1

26.   Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901-904, doi:10.1093/sysbio/syy032.

27.   Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **2018**, *35*, 773-777, doi:10.1093/molbev/msx335.

28.   Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1-15, doi:10.1007/978-1-4939-0366-5_1.

29.   Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. Acta. *Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486-501, doi:10.1107/S0907444910007493.

30.   Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714-2723, doi:10.1002/elps.1150181505.

31.   Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput. Biol.* **2012**, *8*, e1002829, doi:10.1371/journal.pcbi.1002829.

32.   Ponomarenko, J.; Bui, H.H.; Li, W.; Fusseder, N.; Bourne, P.E.; Sette, A.; Peters, B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* **2008**, *9*, 514, doi:10.1186/1471-2105-9-514.

33.   Shashkova, T.I.; Umerenkov, D.; Salnikov, M.; Strashnov, P.V.; Konstantinova, A.V.; Lebed, I.; Shcherbinin, D.N.; Asatryan, M.N.; Kardymon, O.L.; Ivanisenko, N.V. SEMA: Antigen B-cell conformational epitope prediction using deep transfer learning. *Front. Immunol.* **2022**, *13*, 960985, doi:10.3389/fimmu.2022.960985.

34.   Zhou, C.; Chen, Z.; Zhang, L.; Yan, D.; Mao, T.; Tang, K.; Qiu, T.; Cao, Z. SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic. Acids. Res.* **2019**, *47*, W388-W394, doi:10.1093/nar/gkz413.

35.   Ozaki, K.; Matsushima, Y.; Nagasawa, K.; Motoya, T.; Ryo, A.; Kuroda, M.; Katayama, K.; Kimura, H. Molecular Evolutionary Analyses of the RNA-Dependent RNA Polymerase Region in Norovirus Genogroup II. *Front. Microbiol.* **2018**, *9*, 3070, doi:10.3389/fmicb.2018.03070.

36.   Bidalot, M.; Théry, L.; Kaplon, J.; De Rougemont, A.; Ambert-Balay, K. Emergence of new recombinant noroviruses GII.p16-GII.4 and GII.p16-GII.2, France, winter 2016 to 2017. *Euro. Surveill.* **2017**, *22*, doi:10.2807/1560-7917.ES.2017.22.15.30508.

37.   Ao, Y.; Wang, J.; Ling, H.; He, Y.; Dong, X.; Wang, X.; Peng, J.; Zhang, H.; Jin, M.; Duan, Z. Norovirus GII.P16/GII.2-Associated Gastroenteritis, China, 2016. *Emerg. Infect. Dis.* **2017**, *23*, 1172-1175, doi:10.3201/eid2307.170034.

38.   Liu, L.T.; Kuo, T.Y.; Wu, C.Y.; Liao, W.T.; Hall, A.J.; Wu, F.T. Recombinant GII.P16-GII.2 Norovirus, Taiwan, 2016. *Emerg. Infect. Dis.* **2017**, *23*, 1180-1183, doi:10.3201/eid2307.170212.

39. Siebenga, J.J.; Vennema, H.; Zheng, D.P.; Vinjé, J.; Lee, B.E.; Pang, X.L.; Ho, E.C.; Lim, W.; Choudekar, A.; Broor, S.; et al. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. *J. Infect. Dis.* **2009**, *200*, 802-812, doi:10.1086/605127.

40. Vinjé, J. Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* **2015**, *53*, 373-381, doi:10.1128/JCM.01535-14.

41. Eden, J.S.; Tanaka, M.M.; Boni, M.F.; Rawlinson, W.D.; White, P.A. Recombination within the pandemic norovirus GII.4 lineage. *J. Virol.* **2013**, *87*, 6270-6282, doi:10.1128/JVI.03464-12.

42. Kobayashi, M.; Matsushima, Y.; Motoya, T.; Sakon, N.; Shigemoto, N.; Okamoto-Nakagawa, R.; Nishimura, K.; Yamashita, Y.; Kuroda, M.; Saruki, N.; et al. Molecular evolution of the capsid gene in human norovirus genogroup II. *Sci. Rep.* **2016**, *6*, 29400, doi:10.1038/srep29400.

43. Bull, R.A.; Eden, J.S.; Rawlinson, W.D.; White, P.A. Rapid evolution of pandemic noroviruses of the GII.4 lineage. *PLoS Pathog.* **2010**, *6*, e1000831, doi:10.1371/journal.ppat.1000831.

44. Hutson, A.M.; Atmar, R.L.; Marcus, D.M.; Estes, M.K. Norwalk virus-like particle hemagglutination by binding to h histo-blood group antigens. *J. Virol.* **2003**, *77*, 405-415, doi:10.1128/jvi.77.1.405-415.2003.

45. Schroten, H.; Hanisch, F.G.; Hansman, G.S. Human Norovirus Interactions with Histo-Blood Group Antigens and Human Milk Oligosaccharides. *J. Virol.* **2016**, *90*, 5855-5859, doi:10.1128/JVI.00317-16.

46. Holmes, E.C. Virus evolution. In *Fields Virology*, 6th ed.; Knipe, D.M., Howley, P.M., Cohen, J.I., Griffin, D.E., Lamb, R.A., Martin, M.A., Racaniello, V.D., Roizman, B., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013; Volume 1, pp. 286–313.

47. Lindesmith, L.C.; Donaldson, E.; Leon, J.; Moe, C.L.; Frelinger, J.A.; Johnston, R.E.; Weber, D.J.; Baric, R.S. Heterotypic humoral and cellular immune responses following Norwalk virus infection. *J. Virol.* **2010**, *84*, 1800-1815, doi:10.1128/JVI.02179-09.