microorganisms

MDPI

Article

Molecular Evolutionary Analyses of the RNA-Dependent RNA Polymerase (*RdRp*) Region and *VP1* Gene in Sapovirus GI.1 and GI.2

Fuminori Mizukoshi [1,*], Ryusuke Kimura [2,3], Tatsuya Shirai [1], Asumi Hirata-Saito [4], Eri Hiraishi [5], Kosuke Murakami [6], Yen Hai Doan [7], Hiroyuki Tsukagoshi [8], Nobuhiro Saruki [8], Takeshi Tsugawa [9], Kana Kidera [10], Yoshiyuki Suzuki [11], Naomi Sakon [12], Kazuhiko Katayama [10], Tsutomu Kageyama [7], Akihide Ryo [1] and Hirokazu Kimura [3,5,*]

1 Department of Virology III, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan; shirai@niid.go.jp (T.S.); aryo@niid.go.jp (A.R.)
2 Department of Bacteriology, Graduate School of Medicine, Gunma University, Maebashi-shi 371-8511, Japan; m2220015@gunma-u.ac.jp
3 Advanced Medical Science Research Center, Gunma Paz University, Takasaki-shi 370-0006, Japan
4 Department of Microbiology, Tochigi Prefectural Institute of Public Health and Environmental Science, Utsunomiya-shi 329-1196, Japan; saitoa01@pref.tochigi.lg.jp
5 Department of Health Science, Gunma Paz University Graduate School of Health Sciences, Takasaki-shi 370-0006, Japan; er.hiraishi@gmail.com
6 Center for Emergency Preparedness and Response, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan; ko-mura@niid.go.jp
7 Center for Emergency Preparedness and Response, National Institute of Infectious Diseases, Musashimurayama-shi, Tokyo 208-0011, Japan; yendoan@niid.go.jp (Y.H.D.); tkage@niid.go.jp (T.K.)
8 Gunma Prefectural Institute of Public Health and Environmental Sciences, Maebashi-shi 371-0052, Japan; tsuka-hiro@pref.gunma.lg.jp (H.T.); saruki-n@pref.gunma.lg.jp (N.S.)
9 Department of Pediatrics, Sapporo Medical University School of Medicine, Sapporo-shi 060-8543, Japan; tsugawat@sapmed.ac.jp
10 Laboratory of Viral Infection Control, Ōmura Satoshi Memorial Institute, Graduate School of Infection Control Sciences, Kitasato University, 5-9-1, Shirogane, Minato-ku, Tokyo 108-8641, Japan; kidera.kana@st.kitasato-u.ac.jp (K.K.); katayama@lisci.kitasato-u.ac.jp (K.K.)
11 Division of Biological Science, Department of Information and Basic Science, Graduate School of Sciences, Nagoya City University, Nagoya-shi 467-8501, Japan; yossuzuk@nsc.nagoya-cu.ac.jp
12 Department of Microbiology, Osaka Institute of Public Health, Osaka 537-0025, Japan; nsakon@iph.osaka.jp
* Correspondence: mzksh@niid.go.jp (F.M.); h-kimura@paz.ac.jp (H.K.); Tel.: +81-425-61-0771 (F.M.); +81-27-365-3366 (H.K.)

Academic Editor: Agostinho Antunes

Received: 27 December 2024
Revised: 17 January 2025
Accepted: 24 January 2025
Published: 1 February 2025

Citation: Mizukoshi, F.; Kimura, R.; Shirai, T.; Hirata-Saito, A.; Hiraishi, E.; Murakami, K.; Doan, Y.H.; Tsukagoshi, H.; Saruki, N.; Tsugawa, T.; et al. Molecular Evolutionary Analyses of the RNA-Dependent RNA Polymerase (*RdRp*) Region and *VP1* Gene in Sapovirus GI.1 and GI.2. *Microorganisms* **2025**, *13*, 322. https://doi.org/10.3390/microorganisms13020322

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Human sapovirus (HuSaV) is a significant cause of gastroenteritis. This study aims to analyze the evolutionary dynamics of the RNA-dependent RNA polymerase (*RdRp*) and capsid (*VP1*) genes of the HuSaV GI.1 and GI.2 genotypes between 1976 and 2020. Using bioinformatics tools such as the Bayesian phylogenetics software BEAST 2 package (v.2.7.6), we constructed time-scale evolutionary trees based on the gene sequences. Most of the recent common ancestors (MRCAs) of the *RdRp* region and *VP1* gene in the present HuSaV GI.1 diverged around 1930 and 1933, respectively. The trees of the HuSaV GI.1 *RdRp* region and *VP1* gene were divided into two clusters. Further, the MRCAs of the *RdRp* region and *VP1* gene in HuSaV GI.2 diverged in 1960 and 1943, respectively. The evolutionary rates were higher for *VP1* gene in HuSaV GI.1 than that in HuSaV GI.2, furthermore, were higher in GI.1 Cluster B than GI.1 Cluster A. In addition, a steep increase was observed in the time-scaled genome population size of the HuSaV GI.1 Cluster B. These results indicate that the HuSaV GI.1 Cluster B may be evolving more actively than other genotypes. The conformational B-cell epitopes were predicted with a higher probability in *RdRp* for GI.1 and in *VP1* for GI.2, respectively. These results suggest

that the *RdRp* region and *VP1* gene in HuSaV GI.1 and GI.2 evolved uniquely. These findings suggest unique evolutionary patterns in the *RdRp* region and *VP1* gene of HuSaV GI.1 and GI.2, emphasizing the need for a 'One Health' approach to better understand and combat this pathogen.

Keywords: human sapovirus; molecular evolutionary analyses; RNA-dependent RNA polymerase (*RdRp*); capsid (*VP1*)

1. Introduction

Sapovirus (SaV) is a non-enveloped, positive-sense, single-stranded RNA virus that belongs to the Caliciviridae family, which also includes norovirus (NoV) [1]. Various mammals, including humans, pigs, dogs, bats, minks, sea lions, and rats are susceptible to SaV [1–3]. Currently, SaV is classified into nineteen genogroups (GI to GXIX). Four of these genogroups, GI, GII, GIV, and GV, can infect humans and cause clinical disease [1,4]. Moreover, human SaV (HuSaV) is divided into 18 genotypes: GI.1 to GI.7, GII.1 to GII.8, GIV.1, GV.1, and GV.2 [1,5,6]. Molecular epidemiological data on HuSaV infections in humans suggest that two genotypes, GI.1 and GI.2, are frequently detected worldwide [4,7–10].

HuSaV is a causative agent of viral acute gastroenteritis in outbreaks and sporadic cases worldwide [1,4]. The spread of HuSaV usually occurs via the fecal–oral route from contaminated foods and/or via person-to-person transmission. Major clinical symptoms of HuSaV include diarrhea, vomiting nausea, chills, and abdominal cramps. Furthermore, a significant exacerbation of the HuSaV infection occasionally leads to hospitalization [11]. However, at present, effective vaccines and antiviral agents are not available for the control and prevention of HuSaV infections [12]. Therefore, this agent may be a public health concern.

The SaV genome is a positive-sense, single-stranded RNA with an approximately 7.1 to 7.7 kb nucleotide sequence and is composed of two open reading frames (ORF), ORF1 and ORF2 [1]. ORF1 encodes the nonstructural proteins, including RNA-dependent RNA polymerase (RdRp) and the major structural protein, capsid (VP1). ORF2 encodes the minor structural protein, VP2 [1]. The *VP1* gene is more diverse than the *RdRp* gene and is an important region for genotyping [1]. Moreover, similar to NoV, recombinant strains between the *RdRp* gene and *VP1* gene have been reported [1]. Therefore, to understand the characteristics of the SaV genome, it is important to analyze these two major genes, the *RdRp* gene and *VP1* gene, in detail.

It has been recognized that the *RdRp* region and *VP1* gene of NoV have evolved in a co-dependent manner [13,14]. Although most studies have focused on the molecular evolution of NoVs, that of HuSaV is not well known. To gain insight into this theme, we performed a comprehensive molecular analysis of the *RdRp* region and *VP1* gene of the HuSaV GI.1 and GI.2 detected globally between 1976–2020.

2. Materials and Methods

2.1. Strains in This Study

To analyze the molecular evolution of HuSaV GI.1 and GI.2, the complete genome sequences of SaV were downloaded from NCBI Virus [https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/] (last accessed on 2 November 2023) by being searched as the query "Sapovirus, taxid:95341". To begin with, 5750 strains were collected from the database search. These strains were classified into genotypes using the

Norovirus Typing Tool (Ver.2.0) and were confirmed to be defined as SaV GI.1 and GI.2 [15]. Of these, strains with an ambiguous sequence (e.g., N, Y, R, and V) or an unclear year of collection or area were omitted from the dataset. Moreover, sequences with 100% identity were excluded from the dataset. Finally, 113 and 57 strains belonging to HuSaV GI.1 and GI.2 remained and were used to analyze the molecular evolution of the *VP1* gene. On the other hand, 86 and 57 strains belonging to SaV GI.1 and GI.2 were obtained and used to analyze the molecular evolution of the *RdRp* region. Detailed information on the strains used in this study are shown in Supplementary Table S1. Distributions of these genotypes during the investigation periods (1976–2020) are shown in Supplementary Figure S1.

2.2. Maximum Likelihood (ML)

Multiple alignments for these sequences were performed using MAFFT, version 7.520 [16]. The ML phylogenetic analysis was performed using IQ-TREE, version 2.2.2.6 with ultrafast bootstrap test parameters and the SH-like approximate likelihood ratio test [17]. The best-fitting substitution models were automatically selected using ModelFinder within IQ-TREE. FigTree, version 1.4, was used to visualize the resultant trees.

2.3. Time-Scaled Phylogenetic Analyses and Estimation of Evolutionary Rate

To evaluate the molecular evolution of the present strains, phylogenetic trees of the HuSaV *RdRp* region and *VP1* gene were constructed using the Bayesian Markov chain Monte Carlo (MCMC) method in the BEAST package (v.2.7.6) [18]. First, suitable substitution models of each dataset were estimated and determined by the jModelTest2 program [19]. Second, nested sampling was performed to evaluate the best combination from the six clock models (strict clock, random local clock, optimized relaxed clock, relaxed clock exponential, relaxed clock log normal, and fast relaxed clock log normal) and the two prior tree models (coalescent constant population, and coalescent exponential population) [20]. Based on the obtained suitable combinations of these models, listed in Supplementary Table S2, MCMC trees were constructed using the Beast2 software (v.2.7.6). The MCMC chains consisted of 300,000,000 steps, with sampling at every 1000 steps. To confirm convergence, Tracer, version 1.7.2, was used to evaluate effective sample sizes (ESS); values above 200 were accepted. After discarding the first 10% of the MCMC chain as burn-in, the maximum clade credibility tree was produced by TreeAnnotator, version 2.7.6, in the BEAST package. The Bayesian MCMC phylogenetic tree was illustrated by FigTree, version 1.4; 95% of the highest posterior densities (HPDs) of all internal nodes were computed.

In addition, the molecular evolutionary rates were estimated using suitable models selected for each dataset, as described above. Statistical analyses were performed using the Kruskal–Wallis test for GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA). A significant difference was defined as $p < 0.0001$.

2.4. Bayesian Skyline Plot (BSP) Analyses

To evaluate the phylodynamics of the HuSaV strains and each cluster, the effective population size of the *RdRp* region and *VP1* gene were calculated using BSP analysis with BEAST, v2.7.6 [18]. The best substitution combinations of these models were selected, as described above. The best clock mode,l in combination with the prior tree models (coalescent Bayesian skyline) was selected from the models (strict clock, random local clock, optimized relaxed clock, relaxed clock exponential, relaxed clock log normal, and fast relaxed clock log normal), as described above. The obtained suitable models are listed in Supplementary Table S2. The MCMC chains were run for 300,000,000 steps, with sampling at every 1000 steps. The visualization of the BSPs with 95% of the HPDs were performed using Tracer, v1.7.2 [21].

2.5. Phylogenetic Distance Analyses

The phylogenetic distances (patristic distances) between the HuSaV strains and each cluster were analyzed and calculated from the above-mentioned ML tree using the Patristic program [22]. A violin plot was constructed using Orange DATA MINING, version 3.35 [23]. Statistical analyses were performed using the unpaired t test with Welch's correction for GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA). A significant difference was defined as $p < 0.0001$.

2.6. Similarity Plot Analyses

The nucleotide similarity of each sequence, using SimPlot program 3.5.1, was calculated to clarify the relationships among the aligned nucleotide sequences of the HuSaV *RdRp* region and *VP1* gene [24]. Two strains (GI.1: MN794208/HKG/1977 and GI.2: MN794205/GUF/1978) were used as reference sequences. The similarity was examined using the Kimura two-parameter method, with a window size of 200 nucleotides and a step size of 20 nucleotides.

2.7. Construction of the 3D Structure of RdRp and VP1 Proteins

To compare the RdRp and VP1 protein structures among the subtypes, three-dimensional (3D) structural models of these proteins were predicted using LocalColabFold, version 1.5.3, installed on a local computer [25]. Structural models of the HuSaV RdRp and VP1 proteins were constructed for representative strains from each genotype (GI1: MN794208[1977], LC671561[2020/Cluster A], and OP654153[2020/Cluster B]; and GI2: MN794205[1978], and MN510438[2018]). First, the multiple sequence alignment files (a3m files) were constructed on a local computer, using uniref30(2302) as the uniref database, PDB100(230517) as the template database, and colabfold_envdb(202108) as the environmental sequence database. Second, the following flags were used for the structure prediction: "--amber"; "--templates"; and "--use-gpu-relax". The number of prediction recycles was 30. Of the 5 prediction models created with LocalColabfold for each sequence, the best model was determined by considering parameters such as the predicted local distance difference test (pLDDT), template modeling (TM)-score, and root mean square deviation (RMSD). Lastly, these final models were visualized by UCSF ChimeraX, version 1.7.1 [26].

2.8. Conformational B-Cell Epitope Prediction

The conformational B-cell epitopes of the constructed RdRp and VP1 protein models were evaluated using the following five methods: DiscoTope 3.0 (higher confidence:1.50, recall up to ~30%) [27]; ElliPro (cutoff values of 0.5) [28]; epitope3D [29]; SEPPA 3.0 (cutoff values of 0.089) [30]; and SEMA (cutoff values of 0.76) [31]. Amino acid residues predicted by four or more of these methods were determined as conformational B-cell epitopes. These predicted B-cell epitopes were mapped and colored on each model using UCSF ChimeraX, version 1.7.1 [26].

The number of methods predicted as epitopes were visualized using heat maps. The heat maps were constructed with Orange DATA MINING, version 3.35 [23].

2.9. Selective Pressure Analyses

To assume the selective pressure sites for the *RdRp* region and *VP1* gene, the nonsynonymous (*dN*) and synonymous (*dS*) substitution rates at each amino acid site were estimated using Datamonkey [32]. The following three algorithms were used to identify negatively selected sites: single-likelihood ancestor counting (SLAC) [33]; fixed effects likelihood (FEL) [33]; and fast unconstrained Bayesian approximation (FUBAR) [34]. Next,

in addition to the above three methods (SLAC, FEL, and FUBAR), the mixed effects model of evolution (MEME) [35] method was used to detect positively selected sites. The significance level was set at $p < 0.05$ for SLAC, FEL, and MEME. Evidence of selective pressure for FUBAR was supported by a posterior probability > 0.9.

3. Results

3.1. The Phylogeny of the RdRp Region and VP1 Gene in HuSaV Genogroups

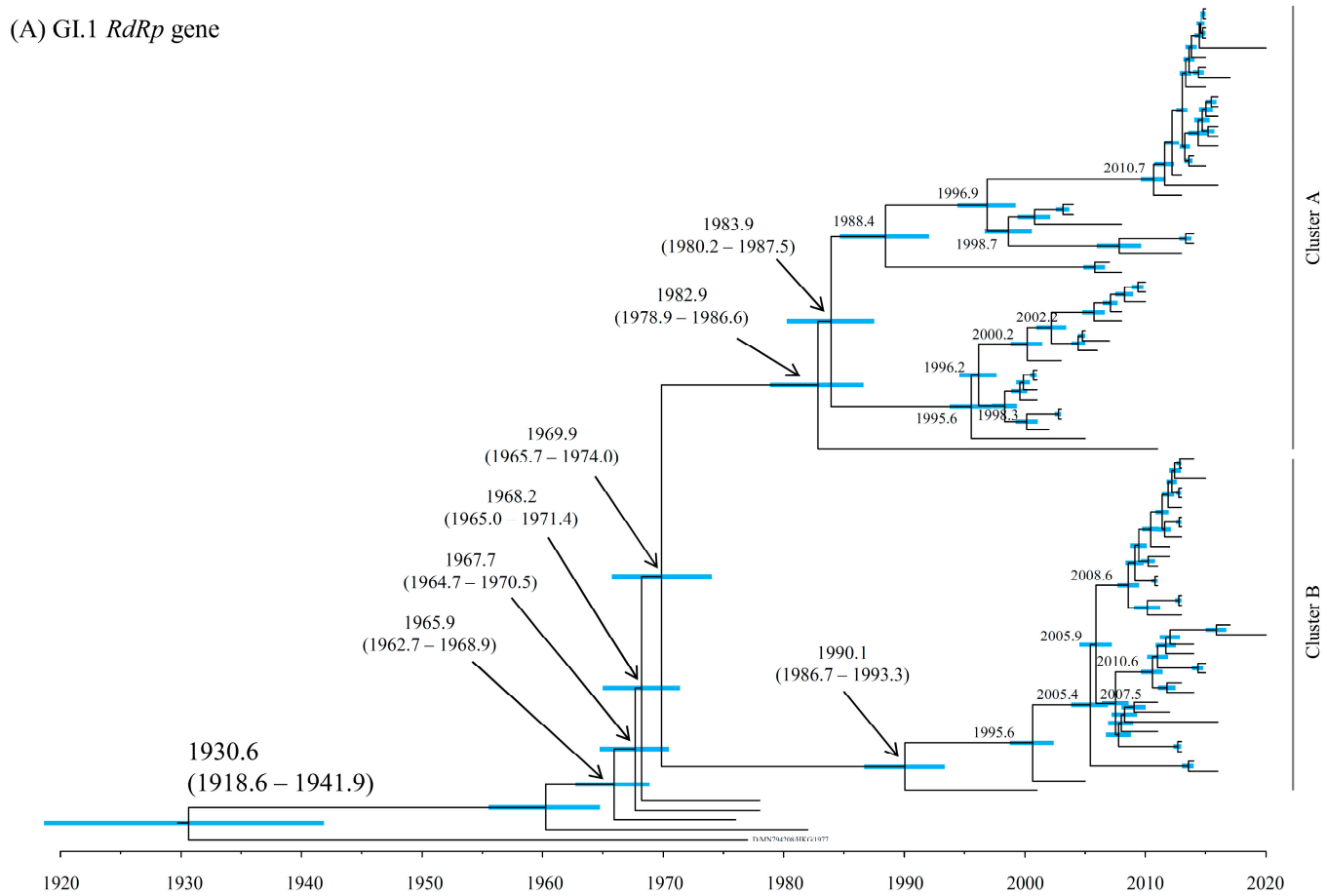
The phylogenetic trees (ML method) of the various HuSaV genogroups (GI.1 and GI.2, plus GI.6, GII, GIV, GV, and porcine SaV GIII) based on the full-length nucleotides of the *RdRp* region and *VP1* gene are shown in Supplementary Figure S2. The trees of *RdRp* and *VP1* showed nearly similar tree shapes, suggesting that these genes may have co-evolved within the SaV genogroups. However, HuSaV GI.6 was located on a different branch, predicting that the recombination between the *RdRp* region and *VP1* gene may have occurred previously. Moreover, the sequences of HuSaV GI.1 and GI.2 were closely related on the *RdRp* region and *VP1* gene, respectively.

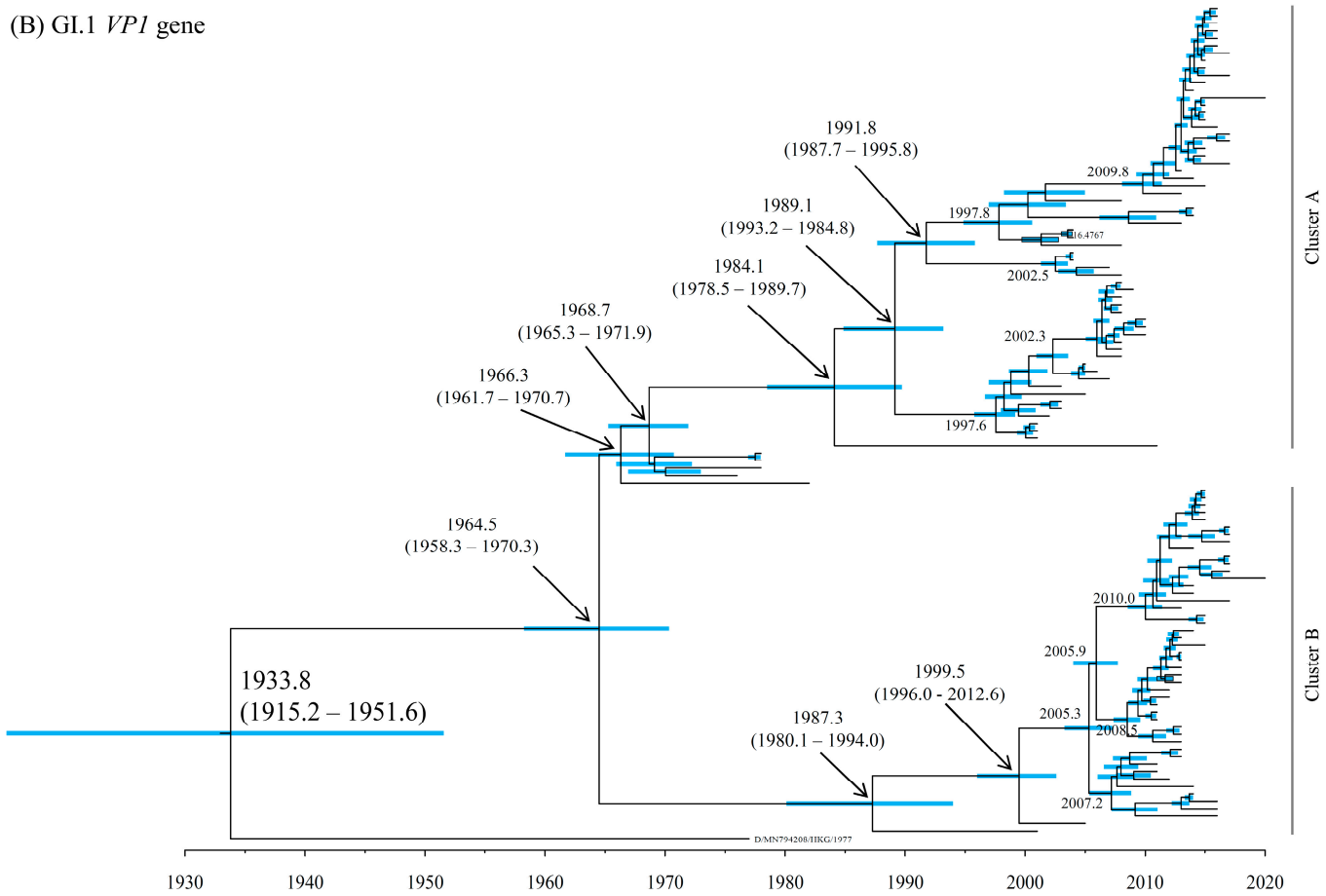
The ML trees of each genotype (GI.1 or GI.2) on the *RdRp*/*VP1* sequences were constructed, respectively (Supplementary Figure S3A–D). The trees of the HuSaV GI.1 *RdRp* region and *VP1* gene were subdivided into two clusters, defined as Cluster A and Cluster B in this study. Distributions of these clusters are shown in Supplementary Figure S1.

3.2. Time-Scaled Phylogeny

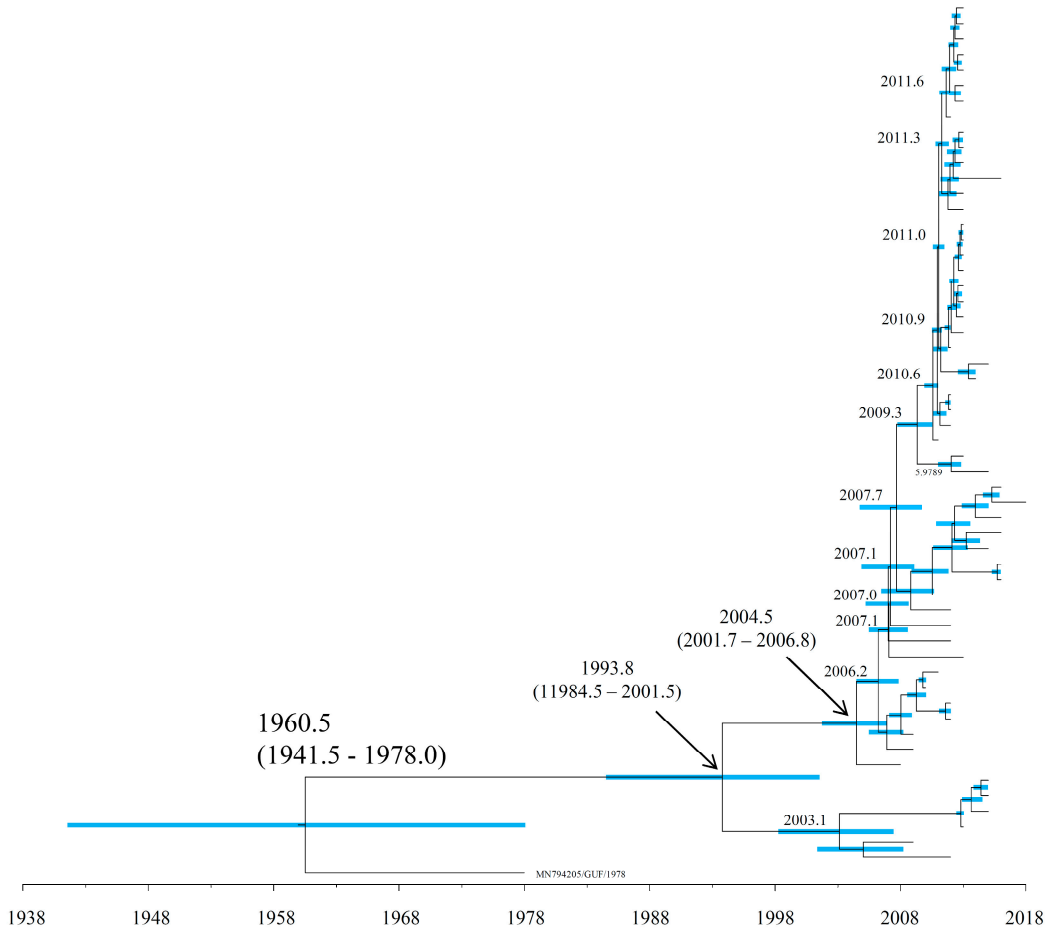
To estimate the evolution of the *RdRp* region and *VP1* gene in HuSaV GI.1 and GI.2, time-scale phylogenetic trees were constructed based on these sequences (Figure 1A–D). Most recent common ancestors (MRCAs) of the *RdRp* region and *VP1* gene in HuSaV GI.1 diverged around 1930.6 (mean; 95% HPDs, 1918.6–1941.9) and 1933.8 (mean; 95% HPDs, 1915.2–1951.6), respectively (Figure 1A,B). Subsequently, these genes of HuSaV GI.1 were further divided into two clusters, Cluster A and Cluster B. The common ancestors of the *RdRp* region of Cluster A and Cluster B diverged around 1982.9 and 1990.1, respectively. On the other hand, the common ancestors of the *VP1* gene of Cluster A and Cluster B diverged around 1984.1 and 1987.3, respectively. Further, the MRCAs of the *RdRp* region and *VP1* gene in HuSaV GI.2 diverged in 1960.5 (mean; 95% HPDs, 1941.5–1978.0) and 1943.6 (mean; 95% HPDs, 1927.9–1958.1), respectively (Figure 1C,D). The common ancestors of the current HuSaV GI.2 strains detected in 2008–2018 diverged in 1993.8 (*RdRp* region) and 1984.8 (*VP1* gene), respectively.

Next, to estimate the ancestors of various genogroups of SaV, time-scale phylogenetic trees were constructed based on the sequences of HuSaV GI.1 and GI.2, together with HuSaV GI.6, GII, GIV, GV, and porcine SaV GIII (Supplementary Figure S4). As a result, the *RdRp* region and *VP1* gene of the HuSaV genogroups (GI, GII, GIV, and GV) diverged from those of the ancestor porcine SaV GIII in 546.6 BC (mean; 95% HPDs, 1269.2 BC–108.4 BC) and 929.7 BC (mean; 95% HPDs, 1668.5 BC–242.8 BC), respectively. The *RdRp* region of HuSaV GI.1 and GI.2 diverged at 1686.6 (mean; 95% HPDs, 1630.0–1737.6). On the other hand, the common ancestors of the HuSaV GI.1 and GI.2 *VP1* genes divided in 1738.2 (mean; 95% HPDs, 1696.1–1778.1). However, because the phylogenetic trees of the various genogroups were constructed with the alignment data of the different nucleotide lengths, this may reflect on the accuracy of the results and this bias may limit the phylogenetic trees of various genogroups.


(A) GI.1 *RdRp* gene
1983.9
(1980.2 – 1987.5)
1982.9
(1978.9 – 1986.6)
1969.9
(1965.7 – 1974.0)
1968.2
(1965.0 – 1971.4)
1967.7
(1964.7 – 1970.5)
1965.9
(1962.7 – 1968.9)
1930.6
(1918.6 – 1941.9)
1990.1
(1986.7 – 1993.3)
1988.4
1996.9
1998.7
2010.7
2002.2
2000.2
1996.2
1995.6
1998.3
2008.6
2005.9
2010.6
2005.4
2007.5
1995.6
Cluster A
Cluster B
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020


(B) GI.1 *VP1* gene
1991.8
(1987.7 – 1995.8)
1989.1
(1993.2 – 1984.8)
1984.1
(1978.5 – 1989.7)
1968.7
(1965.3 – 1971.9)
1966.3
(1961.7 – 1970.7)
1964.5
(1958.3 – 1970.3)
1933.8
(1915.2 – 1951.6)
1999.5
(1996.0 - 2012.6)
1987.3
(1980.1 – 1994.0)
2009.8
1997.8
2002.5
2002.3
1997.6
2010.0
2005.9
2005.3
2008.5
2007.2
Cluster A
Cluster B
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020

(C) GI.2 *RdRp* gene



(D) GI.2 *VP1* gene

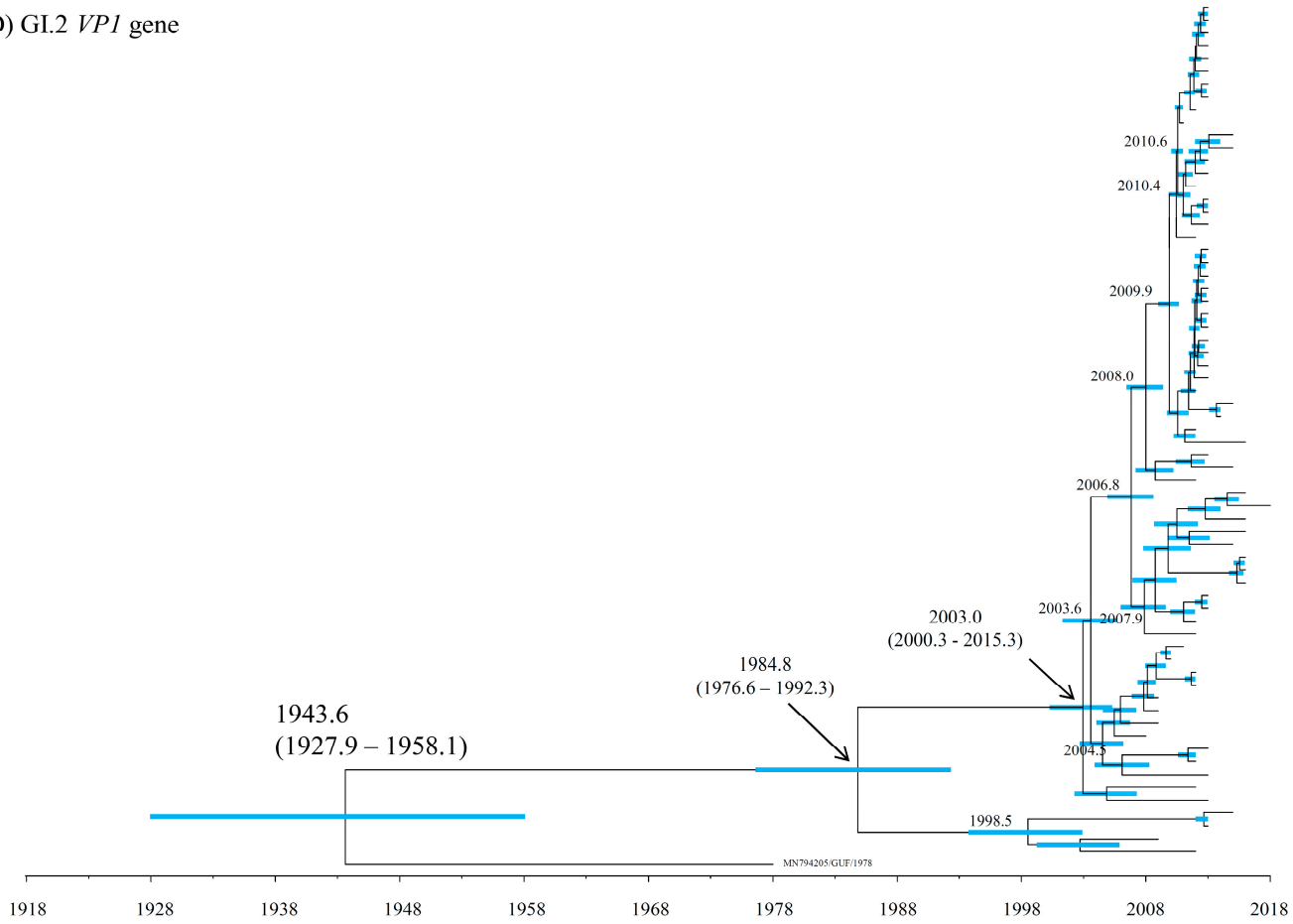


Figure 1. Time-scaled phylogenetic tree of the HuSaV GI.1 constructed by the Bayesian MCMC method: *RdRp* region (**A**); and *VP1* gene (**B**). Time-scaled phylogenetic tree of the HuSaV GI.2 constructed by the Bayesian MCMC method: *RdRp* region (**C**); and *VP1* gene (**D**). The horizontal axis represents time (years). Blue bars indicate the 95% HPD for a branched year.

3.3. Evolutionary Rates

The evolutionary rates of the *RdRp* region and *VP1* gene in HuSaV GI.1 and GI.2 are shown in Figure 2 and Supplementary Table S3. In each *RdRp* region and *VP1* gene, there were significant differences in all combinations of genotypes and clusters.

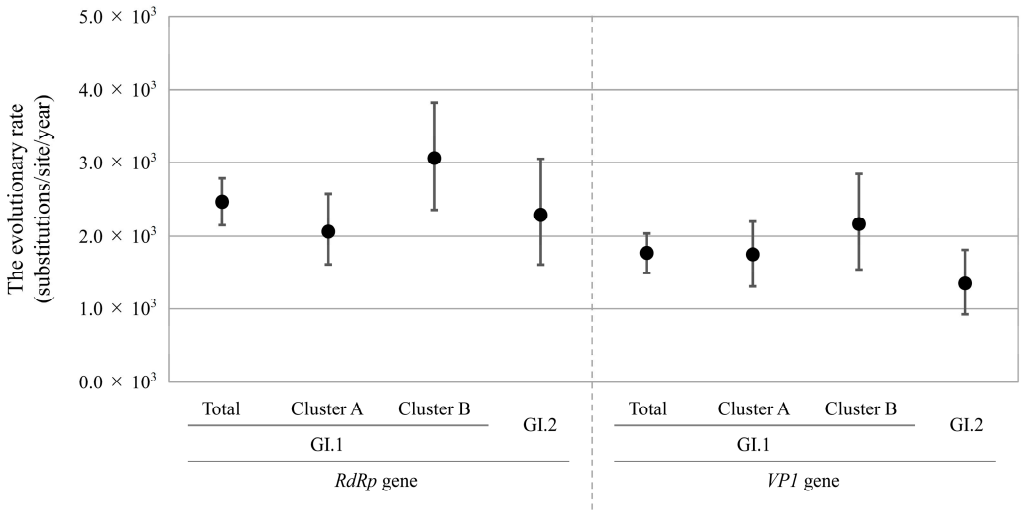


Figure 2. Evolutionary rates of each sequence in the HuSaV of *RdRp* region and *VP1* gene. The y-axis represents the evolutionary rate (substitutions/site/year), and the x-axis indicates each genotype and cluster. The black circles indicate the mean and the bars indicate the intervals of 95% HPD. The detailed evolutionary rates are shown in Supplementary Table S3.

The evolutionary rate of the *RdRp* region was higher for the total HuSaV GI.1 (2.461 × 10^{-3} substitutions/site/year) than the total HuSaV GI.2 (2.289 × 10^{-3} substitutions/site/year). Similarly, the mean evolutionary rate of the *VP1* gene in the total HuSaV GI.1 strains (1.763 × 10^{-3} substitutions/site/year) was higher than that in the HuSaV GI.2 (1.350 × 10^{-3} substitutions/site/year).

Comparing HuSaV GI.1 Cluster A and Cluster B, the evolutionary rates of the *RdRp* region and *VP1* gene were higher in Cluster B (3.059 and 2.168 × 10^{-3} substitutions/site/year, respectively) than in Cluster A (2.065 × 10^{-3} and 1.745 × 10^{-3} substitutions/site/year, respectively).

These results indicate that, comparing the evolutionary rates among the genotypes and clusters, that of GI.1 was faster than that of GI.2, and that of GI.1 Cluster B was the most rapid.

3.4. Phylodynamic Using the BSP Method

The phylodynamics of HuSaV GI.1 and GI.2 in the *RdRp* region and *VP1* gene were estimated and the effective population sizes (EPS) are shown in Figure 3. Until approximately the 2000s, the genome population sizes of both the *RdRp* region and *VP1* gene remained constant. However, significant fluctuations in the EPS were observed around 2010.

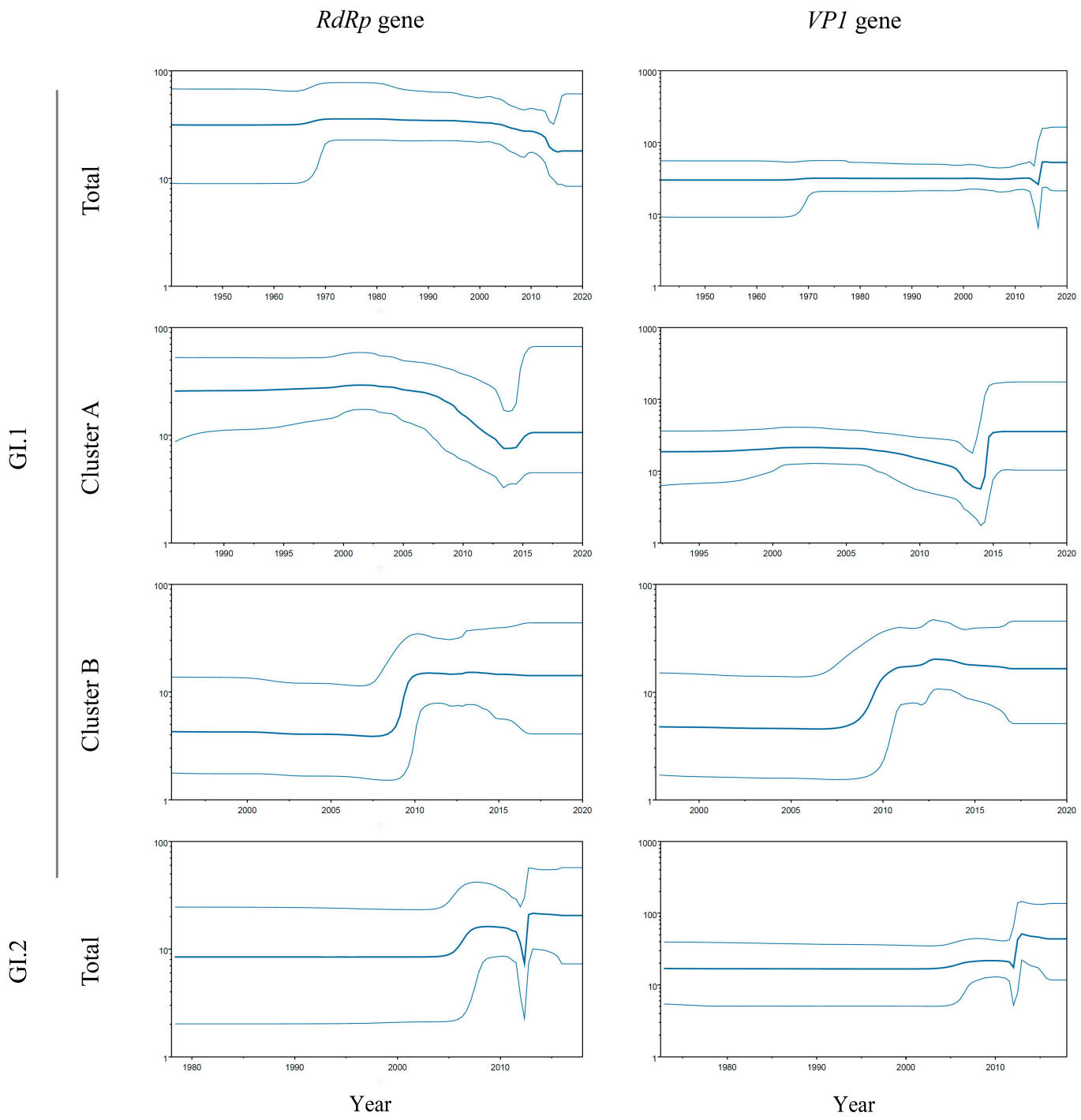


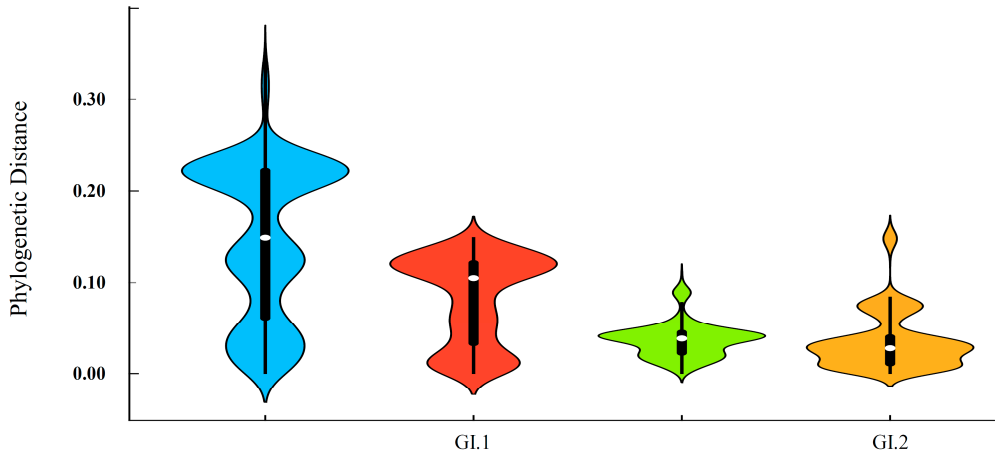
Figure 3. Phylodynamics of the *RdRp* region and *VP1* genes in HuSaV GI.1 and GI.2 determined using BSP analysis. The y-axis represents the effective population size on the logarithmic scale and the x-axis indicates the time in years. The thick line (blue) shows the median value over time. The intervals with the HPDs (95%) are shown by thin lines (blue).

In particular, the EPS of the *RdRp* region and *VP1* gene in HuSaV GI.1 Cluster B rapidly increased before 2010 and then remained constant. On the other hand, the EPS of the *VP1* gene in Cluster A slightly increased around 2010 but that in the *RdRp* gradually decreased after around 2015. These results suggest that HuSaV GI.1 Cluster B has become more adapted to humans.

3.5. Phylogenetic Distances

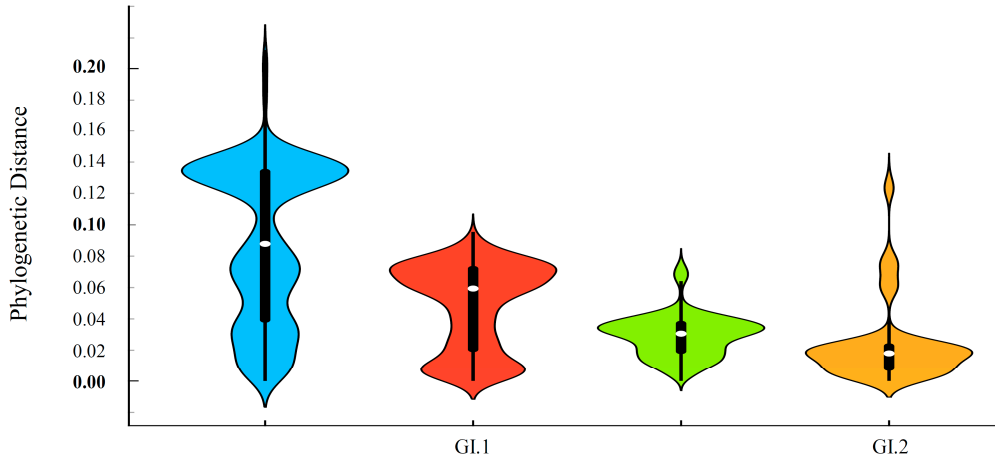
The genetic divergence of the *RdRp* region and *VP1* gene was assessed by their phylogenetic distances and distribution (Figure 4A,B). In both the *RdRp* region and *VP1* gene, there were significant differences in all combinations of genotypes and clusters.

(A) *RdRp* gene



		Total	cluster A	cluster B	GI.2
Strain No.		113	60	47	68
Period		1976 - 2022 (47 years)	2001 - 2020 (20 years)	2001 - 2020 (20 years)	1978 - 2018 (41 years)
Phylogenetic-Distance	Mean	0.147	0.082	0.037	0.035
	(Std. Deviation)	(± 0.084)	(± 0.048)	(± 0.019)	(± 0.032)
	Median	0.149	0.105	0.038	0.028
	(Interquartile range)	(0.061-0.222)	(0.033-0.122)	(0.022-0.045)	(0.011-0.040)

(B) *VP1* gene



		Total	cluster A	cluster B	GI.2
Strain No.		113	60	47	68
Period		1976 - 2022 (47 years)	2001 - 2020 (20 years)	2001 - 2020 (20 years)	1978 - 2018 (41 years)
Phylogenetic-Distance	Mean	0.088	0.048	0.029	0.023
	(Std. Deviation)	(± 0.050)	(± 0.027)	(± 0.014)	(± 0.026)
	Median	0.088	0.059	0.031	0.018
	(Interquartile range)	(0.039-0.134)	(0.021-0.072)	(0.019-0.037)	(0.008-0.022)

Figure 4. Phylogenetic distances of the *RdRp* region (**A**); and *VP1* gene (**B**) in HuSaV GI.1 and GI.2, illustrated by violin plots. The width of the violin plot represents the kernel density, indicating the distribution shape of the data. Total HuSaV GI.1, GI.1 Cluster A, GI.1 Cluster B, and HuSaV GI.2 are illustrated in light blue, red, green, and gold. The central box plot (black) and white dots represent the interquartile range and the median, respectively. The whiskers from the box plots represent

the data intervals. The detailed statistical data are shown below the violin plots. There were significant differences in all combinations of genotypes and clusters (unpaired *t* test; $p < 0.0001$).

The phylogenetic distance values of the *RdRp* region in HuSaV GI.1 Cluster A, Cluster B, and in the GI.2 strains were 0.082 ± 0.048, 0.037 ± 0.019, and 0.035± 0.032 (mean ± SD). Moreover, the phylogenetic distance values of the *VP1* gene in HuSaV GI.1 Cluster A, Cluster B, and in the GI.2 strains were 0.048 ± 0.027, 0.029 ± 0.014, and 0.023 ± 0.026 (mean ± standard deviation, SD). In both the *RdRp* region and *VP1* gene, the branches of the HuSaV GI.1 Cluster A were longer than those of the others ($p < 0.0001$).

These results indicate that the genetic divergence of GI.2 was more highly conservative than that of the other, and that GI.1 Cluster A has a statistically greater genetic divergence than GI.1 Cluster B.

3.6. Similarity Plot Analysis

To understand the genetic diversity of intra-genotypes, a similarity plot analysis based on the nucleotide sequences of the *RdRp* region and *VP1* gene was conducted and is shown in Figure 5 and in Supplementary Figure S5. The results show that the S domain (N-terminal region) and P2 domain (central variable region) in the *VP1* gene of the HuSaV GI.1 total strain, Cluster A, and Cluster B, are characterized by lower similarities (approximately 85% and 82%, respectively), indicating that high divergence sites are localized in the *VP1* gene. On the other hand, the similarity of the entire *VP1* gene in HuSaV GI.2 strains was slightly low (80–90%).

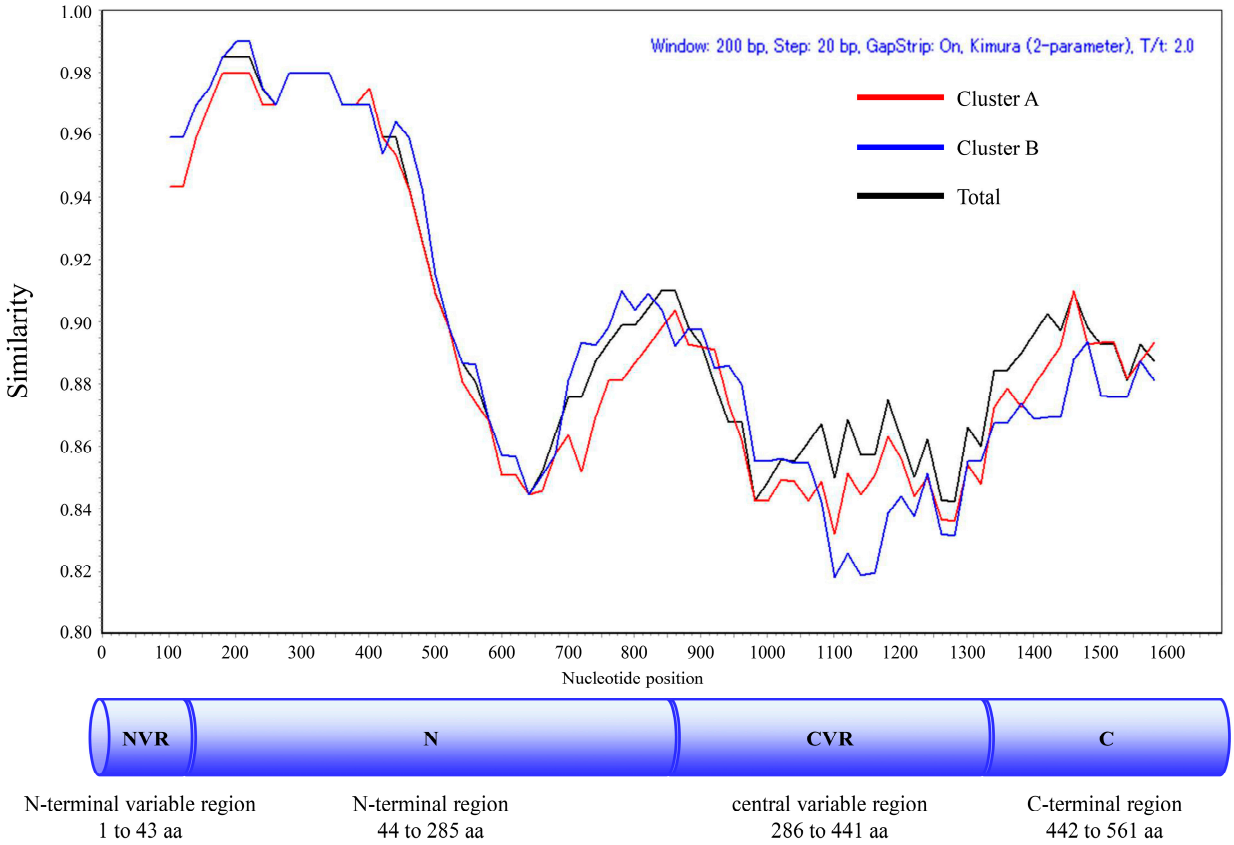


Figure 5. Similarity plot analysis of the *VP1* gene in HuSaV GI.1. Nucleotide similarity to the prototype strain (GenBank accession no. MN794208, collected in 1977) was calculated using SimPlot analysis. Total HuSaV GI.1, GI.1 Cluster A, and GI.1 Cluster B are illustrated in black, red, and blue. Nucleotide position numbers correspond to the *VP1* gene in the prototype strain. The positions of each subunit are shown below the graph [1].

3.7. 3D Mapping of the Conformational Epitopes in the RdRp Dimer and VP1 Trimer Proteins

The predicted conformational epitopes mapped on the RdRp and VP1 protein structure models of the recent strains and ancient representative strains of each genotype are shown in Figures 6A and 7A. To assess the reliability of the predicted sites, the number of methods predicted as epitopes are visualized in a heat map (Figures 6B and 7B). Details of the sites predicted as the conformational epitopes are presented in Supplementary Tables S4 and S5.

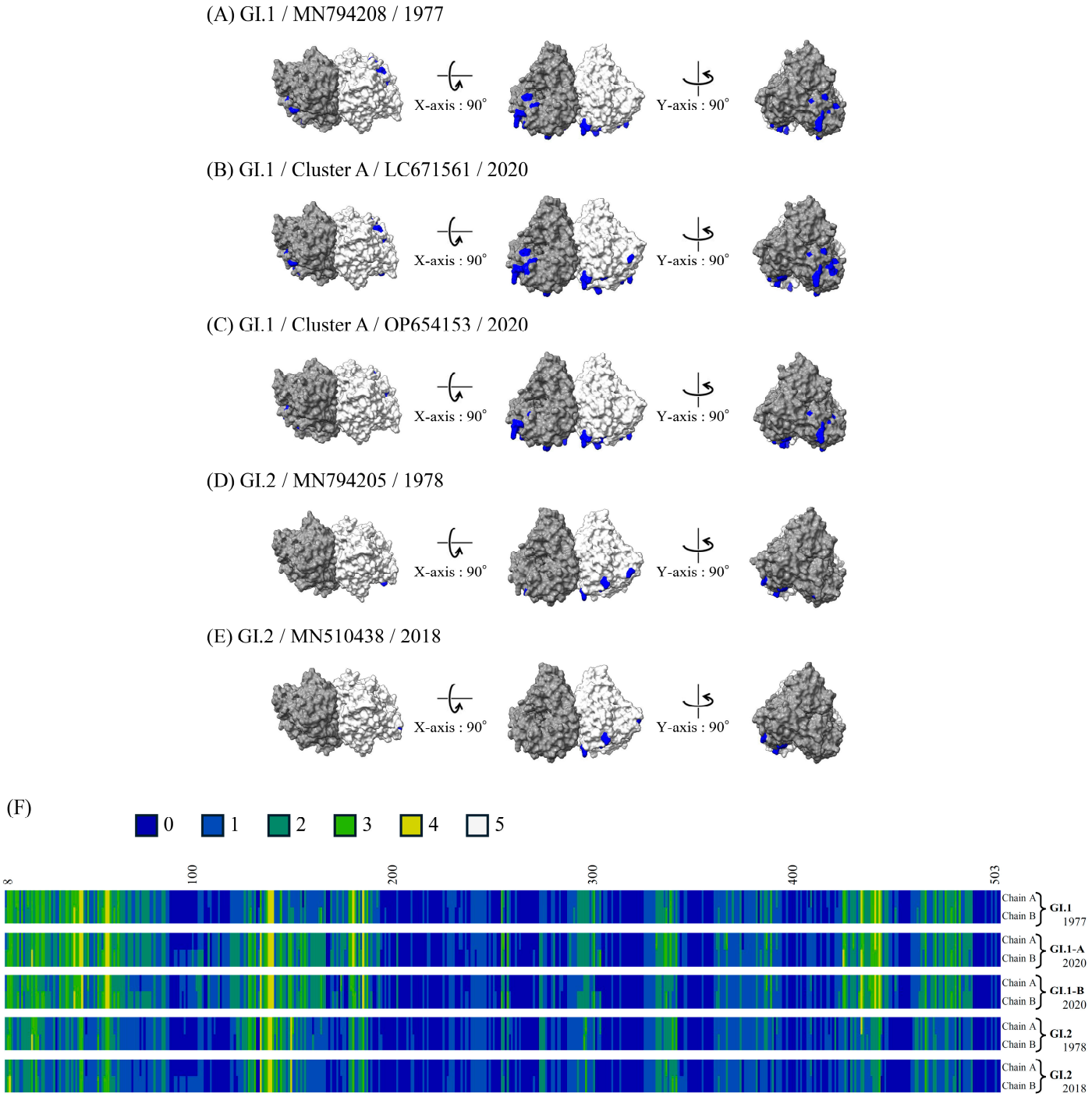


Figure 6. Structural models of the RdRp dimer proteins and mapping of predicted conformational epitopes. Chains A and B are colored by gray and light gray, respectively. The predicted conformational epitopes are shown in blue. The strains are as follows: (**A**) GI.1 prototype strain collected in 1977 (MN794208); (**B**) GI.1 strain belonging to Cluster A collected in 2020 (LC671561); (**C**) GI.1 strain belonging to Cluster B collected in 2020 (OP654153); (**D**) GI.2 prototype strain collected in 1978

(MN794205); (**E**) GI.2 strain collected in 2018 (MN510438); and (**F**) the sites and the number of methods predicted as epitopes and visualized in heat maps. Amino acid residues predicted as conformational B-cell epitopes by 0, 1, 2, 3, 4, and 5 of the five methods are shown in dark blue, blue, dark green, green, yellow, and white.

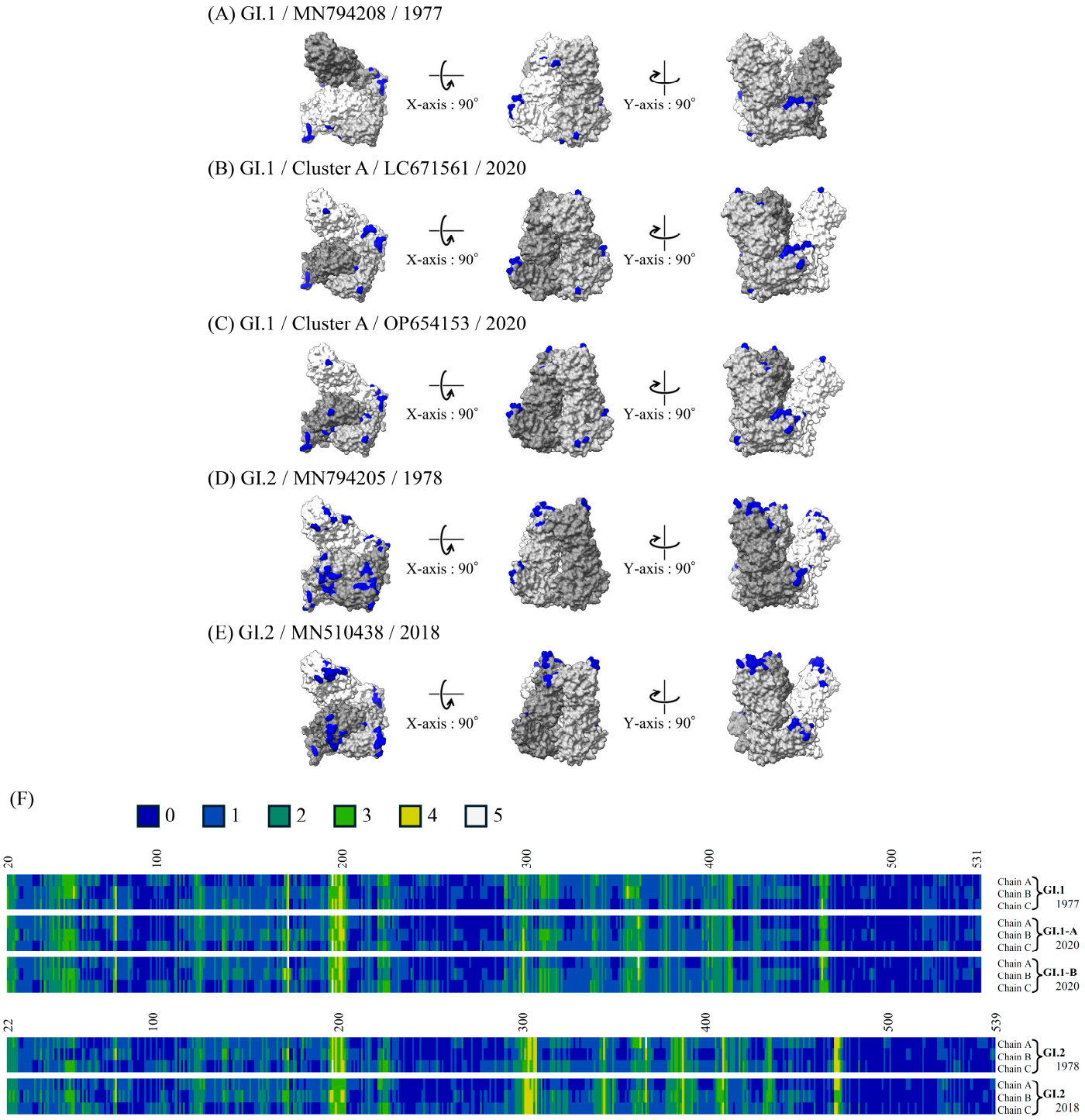


Figure 7. Structural models of the VP1 trimer proteins and mapping of predicted conformational epitopes. Chains A, B, and C are colored in dark gray, light gray, and white, respectively. The predicted conformational epitopes are shown in blue. The strains are as follows: (**A**) GI.1 prototype strain collected in 1977 (MN794208); (**B**) GI.1 strain belonging to Cluster A collected in 2020 (LC671561); (**C**) GI.1 strain belonging to Cluster B collected in 2020 (OP654153); (**D**) GI.2 prototype strain collected in 1978 (MN794205); (**E**) GI.2 strain collected in 2018 (MN510438); and (**F**) the sites and the number of methods predicted as epitopes and visualized in heat maps. Amino acid residues

predicted as conformational B-cell epitopes by 0, 1, 2, 3, 4, and 5 of the five methods are shown in dark blue, blue, dark green, green, yellow, and white.

More sites were predicted as epitopes in the RdRp protein of the GI.1 strains than in that of the GI.2 strains. Most of these predicted epitopes were located on the sides of the RdRp protein dimer in the GI.1 strains (residues aa42–59, 181,186, and 434). On the other hand, more sites predicted as epitopes (residues aa304–310, 347, 370, 390, 412, and 473–475) were found in the VP1 trimer protein of GI.2 than that of GI.1. Many of the predicted epitopes are located in the protruding position (mainly, P2 domain) of the trimer.

3.8. Negative and Positive Selection Sites

To determine selective pressure against the host, positive and negative selection sites in the HuSaV RdRp and VP1 proteins were analyzed and are shown in Supplementary Table S6.

The number of negative selection sites was calculated with three methods (FUBAR, FEL, and SLAC). For the VP1 protein in the GI.1 total strain, Cluster A Cluster B, and GI.2, the predicted negative selection sites were 214, 71, 48, and 37, respectively. On the other hand, the negative selection sites of the RdRp protein were 256, 65, 37, and 45, respectively.

First, the episodic positive selection that occurs on only a subset of evolutionary branches were analyzed using one method (MEME). For the VP1 protein in the GI.1 total strain, Cluster A, Cluster B, and GI.2, the predicted positive selection sites were 4, 1, 1, and 5, respectively. On the other hand, the positive selection sites of the RdRp protein were 2, 1, 0, and 1, respectively. Second, the pervasive positive selection across all the lineages in a phylogenetic tree was detected using three methods (FUBAR, FEL, and SLAC). One residue and three residues in the RdRp protein of the GI.1 total strain were calculated by FEL and FUBAR, respectively. One method (FUBAR) predicted four sites of positive selection in the VP1 protein of the GI.2. However, no positive selection sites were detected in all datasets when calculated by SLAC. Overall, a residue (aa510) in the RdRp protein of the GI.1 total strain, and two residues (aa13, 421) in the VP1 protein of the GI.2 strain were identified as common to episodic and pervasive positive selections. Therefore, these sites are strongly suggested to be under positive selection. Nevertheless, none of these positively selected sites matched the predicted epitope sites.

4. Discussion

The purpose of this study was to investigate the molecular evolution of the *RdRp* region and *VP1* gene in HuSaV using various bioinformatics technologies. In particular, this study focused on two HuSaV genotypes, GI.1 and GI.2, which have been the predominant genotypes worldwide [7,9–11]. These genotypes are closely related genetically and phylogenetically [36,37]. It is extremely important to compare and comprehensively analyze the *RdRp* region and *VP1* gene of these genotypes. Moreover, the HuSaV GI.1 strains were divided into two clusters and the differences between these clusters were examined in detail.

Similar to NoV, SaV can undergo recombination between the non-structural protein-encoding region (including the *RdRp* region) and the *VP1* gene, resulting in new chimera viruses [1,38]. In fact, the occurrence of recombination in SaVs has been reported to date [39–41]. Recombination, as well as evolution, is essential for viral survival. As previously reported, the molecular evolution of NoV may be influenced by the recombination between *RdRp* and *VP1* [14]. In this study, the phylogenetic trees (both ML and time-scaled trees) of the *RdRp* region and *VP1* gene did not find recombination events in HuSaVs GI.1 and GI.2. The *RdRp* region and *VP1* gene may have co-evolved in HuSaV GI.1 and GI.2, respectively.

The evolutionary rates of the *VP1* gene in HuSaV GI.1 and GI.2 were 1.763 and 1.350 × 10^{-3} substitutions/site/year, respectively. The rates for these HuSaV genotypes are likely lower than that of various NoV genotypes such as GII.2 (2.987 × 10^{-3} substitutions/site/year) [14], Gll.3 (4.82 × 10^{-3} substitutions/site/year) [42], GII.4 (7.68 × 10^{-3} substitutions/site/year) [43], GII.6 (3.345 × 10^{-3} substitutions/site/year) [13], and GII.P17-GII.17 (2.7 × 10^{-3} substitutions/site/year) [44]. Therefore, it is possible that the genes of HuSaVs have evolved more slowly than those of the NoVs. To evaluate the evolutionary rates of viral genes, it may be possible to predict the emergence of new variants that are better able to evade host immunity [43]. Thus, NoVs frequently causes global pandemics because the high evolutionary rates permit the NoVs to mutate for immune evasion [45]. On the other hand, SaV, which mutates less frequently than NoV, may not have escaped human herd immunity to date. Consequently, continuous evolutionary analyses of these genes are needed in the future.

As previously reported for NoV, the evolution of the *VP1* gene affects the *RdRp* region [14,46]. In this study, the time-scaled phylogenetic trees suggested that the *RdRp* region and *VP1* gene of HuSaV GI.1 divided into two clusters, Cluster A and Cluster B, around 1969 and 1964, respectively. Moreover, the different evolutionary rates between the two HuSaV GI.1 clusters were estimated. The evolutionary rates of the *VP1* gene were higher in Cluster B (2.168 × 10^{-3} substitutions/site/year) than in Cluster A (1.745 × 10^{-3} substitutions/site/year). Similarly, the rate in the *RdRp* region of Cluster B (3.037 × 10^{-3} substitutions/site/year) was faster than that of Cluster A (3.037 × 10^{-3} substitutions/site/year). The rapid evolutionary rate of Cluster B may be related to the increase in population genome size. These results suggest that Cluster A and B may have evolved independently. Because no recombination between the *RdRp* region and *VP1* gene was observed within SaV GI in this study, the evolutionary differences between clusters are not due to the recombination, as previously reported [14]. Further work is needed to clarify the differences in the evolutionary characteristics of the genotypes and clusters.

The phylogenetic distance analyses estimated that the *RdRp* region and *VP1* gene of HuSaV GI.1 Cluster A were more genetically wide than those of Cluster B (Figure 4). These results suggest that the *RdRp* region and *VP1* gene in Cluster A have high diversity. The possible reason for this diversity may be that HuSaV GI.1 Cluster A has been detected for a long period of time (2001–2020) (Supplementary Figure S1) and has branched into two sub-clusters topologically in the phylogenetic trees (Figure 1, Supplementary Figure S3). On the other hand, the *RdRp* region and *VP1* gene of HuSaV GI.1 Cluster B may be highly conserved, because Cluster B was detected over a short period of time (2011–2017) (Supplementary Figure S1). Nevertheless, a high evolutionary rate and an increase in the effective population sizes were observed in HuSaV Cluster B. Thus, it is possible that the gene of HuSaV Cluster B may be rapidly evolving.

The HuSaV strains diverged from the porcine SaV GIII around 900 BC (Supplementary Figure S4). The domestication of pigs is believed to have started around 9000 to 10,000 years ago [47]. It has been suggested that SaV may have been transmitted from pigs to humans. On the other hand, a previous report suggested that NoV GIII (bovine type) was divided from NoV GII (human type) around 854 CE [48]. Moreover, Group A rotavirus is a zoonotic pathogen that infects a variety of mammalian animals, particularly domestic animals such as cattle and pigs [49]. Domestic animals that come into close contact with humans increase the risk of the cross-species transmission of pathogens, which may potentially cause pandemics [50]. Some genotypes of SaV can infect various animals; thus, SaV is also considered a zoonosis [51]. Since it is possible that SaV from animals may infect humans, it is necessary to study the relationship between genotypes that infect other animal species and HuSaV in the future. Such an approach would contribute to the development of One Health surveillance.

The RdRp diner protein of HuSaV GI.1 had more sites that were strongly predicted as epitopes than that of GI.2. In contrast, the VP1 protein of HuSaV GI.2 had more sites that were strongly predicted as epitopes than GI.1. In particular, the apex of the protrusion of the VP1 protein was strongly predicted as epitopes. Because it has been reported that the VP1 protein of SaV is a target of host immunity [52], it is possible that GI.1 has evolved through escape mutations. Therefore, in the future, a variant that evades host immunity may emerge from genotype GI.1. In particular, Cluster A, which has a rapid evolutionary rate and an increasing genome population size among SaV GI, should be noticeable.

In a recent report, Ji et al. also analyzed the whole genome, focusing on HuSaV GI.1 [53]. They constructed a phylogenetic tree using the whole genome of GI.1 and classified it into two lineages: Lineage I and Lineage II, which correspond to Clusters B and A in this study. In contrast, we analyzed the *RdRp* region and *VP1* gene of HuSaV GI.1 and GI.2, separately. It was found that the two genes have evolved with different characteristics, providing important insights into understanding the function and evolution of these viral proteins. For molecular evolutionary research in viral genes, it is important to analyze individual genes as well as the whole genome sequence.

Similar to NoV, it is important to study the *RdRp* and *VP1* of SaV. However, there is no clear classification definition for the *RdRp* genotype of SaV. Therefore, there may be limitations to the molecular evolution studies and molecular epidemiological surveys focusing on the SaV *RdRp* gene and RdRp protein. As with NoV, an international definition of the *RdRp* classification in SaV is eagerly awaited.

5. Conclusions

In conclusion, the most recent common ancestors of the *VP1* genes in HuSaV GI.1 and GI.2 emerged around the 1930s. The evolutionary rate of strains belonging to HuSaV GI.1 Cluster B was significantly higher than that of strains belonging to HuSaV GI.1 Cluster A and GI.2. Moreover, the effective population size of the HuSaV GI.1 Cluster B rapidly increased before 2010. These results suggest that the HuSaV GI.1 Cluster B may be evolving more actively than other genotypes. Furthermore, conformational B-cell epitopes were more strongly predicted in the VP1 protein of GI.2 than in that of GI.1. Together, these findings may contribute to a better understanding of HuSaV virology and molecular evolution.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Timeline distribution of sequences by collected year in this study; Figure S2: Phylogenetic trees of the *RdRp* and *VP1* sequences of the various SaV genogroups constructed by the ML method; Figure S3: Phylogenetic trees of *RdRp* regions and *VP1* gene of the HuSaV GI.1, and *RdRp* region and *VP1* gene of the HuSaV GI.2 constructed by the ML method; Figure S4: Time-scaled phylogenetic trees of *RdRp* region and *VP1* gene of t the various SaV genogroups constructed by the Bayesian MCMC method; Figure S5: Similarity plot analysis of the *RdRp* region of the HuSaV GI.1, and *RdRp* region and *VP1* genes of the HuSaV GI.2; Table S1: List of HuSaV strains used in this study; Table S2: Parameters used in the Bayesian Markov chain Monte Carlo (MCMC) analyses; Table S3: Evolutionary rates (Substitutions/Site/Year) of the *RdRp* region and *VP1* gene in HuSaV GI.1 and GI.2; Table S4: Putative conformational epitopes for the RdRp proteins of HuSaV; Table S5: Putative conformational epitopes for the VP1 proteins of HuSaV; Table S6: Number of amino acid residues of predicted positive and negative selection sites in HuSaV.

Author Contributions: Conceptualization, F.M. and H.K.; methodology, F.M., R.K., T.S., A.H.-S., E.H., K.M., Y.H.D., H.T. and H.K.; investigation, F.M. and H.K.; formal analysis, F.M.; visualization, F.M.; writing (original draft) was performed by F.M.; writing (review and editing) was performed by K.K. (Kana Kidera), K.K. (Kazuhiko Katayama) and H.K.; supervision, N.S. (Nobuhiro Saruki),

T.T., Y.S., N.S. (Naomi Sakon), K.K. (Kazuhiko Katayama), T.K., A.R. and H.K.; funding acquisition, T.K. and H.K.; All authors participated in the investigation, and interpretation of data, and provided important intellectual contributions to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the Japan Agency for Medical Research and Development (AMED) (grant number: JP24fk0108667).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest in association with the present study.

References

1. Oka, T.; Wang, Q.; Katayama, K.; Saif, L.J. Comprehensive review of human sapoviruses. *Clin. Microbiol. Rev.* **2015**, *28*, 32–53. https://doi.org/10.1128/CMR.00011-14.
2. Scheuer, K.A.; Oka, T.; Hoet, A.E.; Gebreyes, W.A.; Molla, B.Z.; Saif, L.J.; Wang, Q. Prevalence of porcine noroviruses, molecular characterization of emerging porcine sapoviruses from finisher swine in the United States, and unified classification scheme for sapoviruses. *J. Clin. Microbiol.* **2013**, *51*, 2344–2353. https://doi.org/10.1128/JCM.00865-13.
3. Oka, T.; Lu, Z.; Phan, T.; Delwart, E.L.; Saif, L.J.; Wang, Q. Genetic Characterization and Classification of Human and Animal Sapoviruses. *PLoS ONE* **2016**, *11*, e0156373. https://doi.org/10.1371/journal.pone.0156373.
4. Becker-Dreps, S.; Gonzalez, F.; Bucardo, F. Sapovirus: An emerging cause of childhood diarrhea. *Curr. Opin. Infect. Dis.* **2020**, *33*, 388–397. https://doi.org/10.1097/QCO.0000000000000671.
5. Liu, X.; Jahuira, H.; Gilman, R.H.; Alva, A.; Cabrera, L.; Okamoto, M.; Xu, H.; Windle, H.J.; Kelleher, D.; Varela, M.; et al. Etiological Role and Repeated Infections of Sapovirus among Children Aged Less than 2 Years in a Cohort Study in a Peri-urban Community of Peru. *J. Clin. Microbiol.* **2016**, *54*, 1598–1604. https://doi.org/10.1128/JCM.03133-15.
6. Kagning Tsinda, E.; Malasao, R.; Furuse, Y.; Gilman, R.H.; Liu, X.; Apaza, S.; Espetia, S.; Cama, V.; Oshitani, H.; Saito, M. Complete Coding Genome Sequences of Uncommon GII.8 Sapovirus Strains Identified in Diarrhea Samples Collected from Peruvian Children. *Genome Announc.* **2017**, *5*, e01137-17. https://doi.org/10.1128/genomeA.01137-17.
7. Kumthip, K.; Khamrin, P.; Maneekarn, N. Molecular epidemiology and genotype distributions of noroviruses and sapoviruses in Thailand 2000-2016: A review. *J. Med. Virol.* **2018**, *90*, 617–624. https://doi.org/10.1002/jmv.25019.
8. Mann, P.; Pietsch, C.; Liebert, U.G. Genetic Diversity of Sapoviruses among Inpatients in Germany, 2008–2018. *Viruses* **2019**, *11*, 726. https://doi.org/10.3390/v11080726.
9. Zhuo, R.; Ding, X.; Freedman, S.B.; Lee, B.E.; Ali, S.; Luong, J.; Xie, J.; Chui, L.; Wu, Y.; Pang, X.; et al. Molecular Epidemiology of Human Sapovirus among Children with Acute Gastroenteritis in Western Canada. *J. Clin. Microbiol.* **2021**, *59*, e0098621. https://doi.org/10.1128/JCM.00986-21.
10. Doan, Y.H.; Yamashita, Y.; Shinomiya, H.; Motoya, T.; Sakon, N.; Suzuki, R.; Shimizu, H.; Shigemoto, N.; Harada, S.; Yahiro, S.; et al. Distribution of Human Sapovirus Strain Genotypes over the Last Four Decades in Japan: A Global Perspective. *Jpn J. Infect. Dis.* **2023**, *76*, 255–258. https://doi.org/10.7883/yoken.JJID.2022.704.
11. Magwalivha, M.; Kabue, J.P.; Traore, A.N.; Potgieter, N. Prevalence of Human Sapovirus in Low and Middle Income Countries. *Adv. Virol.* **2018**, *2018*, 5986549. https://doi.org/10.1155/2018/5986549.
12. Wang, X.; Li, Y.; Deloria-Knoll, M.; Madhi, S.A.; Cohen, C.; Arguelles, V.L.; Basnet, S.; Bassat, Q.; Brooks, W.A.; Echavarria, M.; et al. Global burden of acute lower respiratory infection associated with human parainfluenza virus in children younger than 5 years for 2018: A systematic review and meta-analysis. *Lancet Glob. Health* **2021**, *9*, e1077–e1087. https://doi.org/10.1016/S2214-109X(21)00218-7.
13. Takahashi, T.; Kimura, R.; Shirai, T.; Sada, M.; Sugai, T.; Murakami, K.; Harada, K.; Ito, K.; Matsushima, Y.; Mizukoshi, F.; et al. 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. *Viruses* **2023**, *15*, 1497. https://doi.org/10.3390/v15071497.

14. 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. https://doi.org/10.3389/fmicb.2017.00705.
15. Chhabra, P.; Graaf, M.; Parra, G.I.; Chan, M.C.; Green, K.; Martella, V.; Wang, Q.; White, P.A.; Katayama, K.; Vennema, H.; et al. Corrigendum: Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* **2020**, *101*, 893. https://doi.org/10.1099/jgv.0.001475.
16. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. https://doi.org/10.1093/molbev/mst010.
17. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
18. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchene, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kuhnert, D.; De Maio, N.; et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650. https://doi.org/10.1371/journal.pcbi.1006650.
19. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. https://doi.org/10.1038/nmeth.2109.
20. Russel, P.M.; Brewer, B.J.; Klaere, S.; Bouckaert, R.R. Model Selection and Parameter Inference in Phylogenetics Using Nested Sampling. *Syst. Biol.* **2019**, *68*, 219–233. https://doi.org/10.1093/sysbio/syy050.
21. 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. https://doi.org/10.1093/sysbio/syy032.
22. Fourment, M.; Gibbs, M.J. PATRISTIC: A program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol. Biol.* **2006**, *6*, 1. https://doi.org/10.1186/1471-2148-6-1.
23. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
24. Lole, K.S.; Bollinger, R.C.; Paranjape, R.S.; Gadkari, D.; Kulkarni, S.S.; Novak, N.G.; Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **1999**, *73*, 152–160. https://doi.org/10.1128/JVI.73.1.152-160.1999.
25. Mirdita, M.; Schutze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.
26. Meng, E.C.; Goddard, T.D.; Pettersen, E.F.; Couch, G.S.; Pearson, Z.J.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **2023**, *32*, e4792. https://doi.org/10.1002/pro.4792.
27. Hoie, M.H.; Gade, F.S.; Johansen, J.M.; Wurtzen, C.; Winther, O.; Nielsen, M.; Marcatili, P. DiscoTope-3.0: Improved B-cell epitope prediction using inverse folding latent representations. *Front. Immunol.* **2024**, *15*, 1322712. https://doi.org/10.3389/fimmu.2024.1322712.
28. 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 Bioinform.* **2008**, *9*, 514. https://doi.org/10.1186/1471-2105-9-514.
29. da Silva, B.M.; Myung, Y.; Ascher, D.B.; Pires, D.E.V. epitope3D: A machine learning method for conformational B-cell epitope prediction. *Brief. Bioinform.* **2022**, *23*, bbab423. https://doi.org/10.1093/bib/bbab423.
30. 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. https://doi.org/10.1093/nar/gkz413.
31. 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. https://doi.org/10.3389/fimmu.2022.960985.
32. 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. https://doi.org/10.1093/molbev/msx335.
33. Kosakovsky Pond, S.L.; Frost, S.D. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. https://doi.org/10.1093/molbev/msi105.
34. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. https://doi.org/10.1093/molbev/mst030.

35. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **2012**, *8*, e1002764. https://doi.org/10.1371/journal.pgen.1002764.
36. Wu, F.T.; Oka, T.; Kuo, T.Y.; Doan, Y.H.; Tzu-Chi Liu, L. Sapoviruses detected from acute gastroenteritis outbreaks and hospitalized children in Taiwan. *J. Formos. Med. Assoc.* **2021**, *120*, 1591–1601. https://doi.org/10.1016/j.jfma.2020.11.020.
37. Luo, X.; Deng, J.K.; Mu, X.P.; Yu, N.; Che, X. Detection and characterization of human astrovirus and sapovirus in outpatients with acute gastroenteritis in Guangzhou, China. *BMC Gastroenterol.* **2021**, *21*, 455. https://doi.org/10.1186/s12876-021-02044-5.
38. Desselberger, U. Caliciviridae Other Than Noroviruses. *Viruses* **2019**, *11*, 286. https://doi.org/10.3390/v11030286.
39. Katayama, K.; Miyoshi, T.; Uchino, K.; Oka, T.; Tanaka, T.; Takeda, N.; Hansman, G.S. Novel recombinant sapovirus. *Emerg. Infect. Dis.* **2004**, *10*, 1874–1876. https://doi.org/10.3201/eid1010.040395.
40. Wang, Q.H.; Han, M.G.; Funk, J.A.; Bowman, G.; Janies, D.A.; Saif, L.J. Genetic diversity and recombination of porcine sapoviruses. *J. Clin. Microbiol.* **2005**, *43*, 5963–5972. https://doi.org/10.1128/JCM.43.12.5963-5972.2005.
41. Tohma, K.; Kulka, M.; Coughlan, S.; Green, K.Y.; Parra, G.I. Genomic Analyses of Human Sapoviruses Detected over a 40-Year Period Reveal Disparate Patterns of Evolution among Genotypes and Genome Regions. *Viruses* **2020**, *12*, 516. https://doi.org/10.3390/v12050516.
42. Saito, M.; Tsukagoshi, H.; Ishigaki, H.; Aso, J.; Ishii, H.; Okayama, K.; Ryo, A.; Ishioka, T.; Kuroda, M.; Saruki, N.; et al. Molecular evolution of the capsid (VP1) region in human norovirus genogroup II genotype 3. *Heliyon* **2020**, *6*, e03835. https://doi.org/10.1016/j.heliyon.2020.e03835.
43. Motoya, T.; Nagasawa, K.; Matsushima, Y.; Nagata, N.; Ryo, A.; Sekizuka, T.; Yamashita, A.; Kuroda, M.; Morita, Y.; Suzuki, Y.; et al. Molecular Evolution of the VP1 Gene in Human Norovirus GII.4 Variants in 1974-2015. *Front. Microbiol.* **2017**, *8*, 2399. https://doi.org/10.3389/fmicb.2017.02399.
44. 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. https://doi.org/10.3389/fmicb.2019.02189.
45. Winder, N.; Gohar, S.; Muthana, M. Norovirus: An Overview of Virology and Preventative Measures. *Viruses* **2022**, *14*, 2811. https://doi.org/10.3390/v14122811.
46. Nagasawa, K.; Matsushima, Y.; Motoya, T.; Mizukoshi, F.; Ueki, Y.; Sakon, N.; Murakami, K.; Shimizu, T.; Okabe, N.; Nagata, N.; et al. Genetic Analysis of Human Norovirus Strains in Japan in 2016–2017. *Front. Microbiol.* **2018**, *9*, 1. https://doi.org/10.3389/fmicb.2018.00001.
47. Groenen, M.A. A decade of pig genome sequencing: A window on pig domestication and evolution. *Genet. Sel. Evol.* **2016**, *48*, 23. https://doi.org/10.1186/s12711-016-0204-2.
48. 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. https://doi.org/10.1038/srep29400.
49. Luchs, A.; Timenetsky Mdo, C. Group A rotavirus gastroenteritis: Post-vaccine era, genotypes and zoonotic transmission. *Einstein* **2016**, *14*, 278–287. https://doi.org/10.1590/S1679-45082016RB3582.
50. Tomori, O.; Oluwayelu, D.O. Domestic Animals as Potential Reservoirs of Zoonotic Viral Diseases. *Annu. Rev. Anim. Biosci.* **2023**, *11*, 33–55. https://doi.org/10.1146/annurev-animal-062922-060125.
51. Bank-Wolf, B.R.; Konig, M.; Thiel, H.J. Zoonotic aspects of infections with noroviruses and sapoviruses. *Vet. Microbiol.* **2010**, *140*, 204–212. https://doi.org/10.1016/j.vetmic.2009.08.021.
52. Amin, M.R.; Siddiqui, M.S.; Ahmed, D.; Ahmed, F.; Hossain, A. B- and T-cell epitope mapping of human sapovirus capsid protein: An immunomics approach. *Int. J. Bioinform. Res. Appl.* **2011**, *7*, 287–298. https://doi.org/10.1504/IJBRA.2011.041739.
53. Ji, X.; Guo, C.; Dai, Y.; Chen, L.; Chen, Y.; Wang, S.; Sun, Y. Genomic Characterization and Molecular Evolution of Sapovirus in Children under 5 Years of Age. *Viruses* **2024**, *16*, 146. https://doi.org/10.3390/v16010146.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.