



OPEN Genomic diversity and comparative phylogenomic analysis of genus *Norovirus*

Huijeong Doh¹, Changhyeon Lee², Nam Yee Kim³, Yun-Yong Park¹, Eun-jeong Kim¹, Changsun Choi⁴ & Seong-il Eyun¹✉

Noroviruses consist of ten genogroups, five of which (GI, GII, GIV, GVIII, GIX) infect humans. Noroviruses are traditionally classified based on the VP1 (genotype), RdRP (P-type), or dual-typing nomenclature. However, current classifications solely relying on specific proteins may be insufficient to represent the evolutionary history due to their recombination events. Thus, it is challenging to identify the dual-types in environmental or stool samples co-infected with more than two types using the existing system. We performed a comprehensive genomic analysis using ten assembled genomes with 1417 genomes from NCBI. Our study provides a detailed examination of the genomic characteristics of norovirus and the criteria for current genotypes and P-types. The phylogenomic analysis revealed two key findings: (1) GVIII and GIX are nested within GII and (2) strains of GII.11, GII.18, and GII.19 (swine noroviruses) as well as GIV and GVI form host-based clusters, with GIV.2[GVI.P1] strains in particular suggesting the possibility of another instance of zoonotic transmission. We present a comparison of the phylogenetic findings from gene-based and genome-based analyses. Overall, our study represents an initial step towards the phylogenomic analysis of genus *Norovirus*. This is valuable for not only interpreting the evolutionary trajectory among norovirus strains but also developing antiviral targeting strategies.

Keywords Norovirus, Selection pressure, Phylogenomics, Sequence similarity network

Noroviruses are the leading cause of foodborne illness in almost a fifth of all acute gastroenteritis (AGE) cases worldwide¹. Norovirus-associated AGE, characterized by vomiting and dehydrating diarrhea, is highly transmissible and young children and the elderly are especially susceptible². The first major outbreak of human norovirus occurred in 1968 among schoolchildren in Norwalk, Ohio, USA and the causal agent was identified using immune electron microscopy (IEM) to visualize the virus particle in 1972³. In the late 1980s, researchers established the classification of Norwalk virus as a member of the family *Caliciviridae* on the basis of their genome organization⁴.

Noroviruses are non-enveloped positive-sense ssRNA viruses with approximately 7.5 kb genomes⁵. With the exception of the murine norovirus, the genomic structure of noroviruses consists of three open reading frames (ORFs). Of these, ORF1 is translated to a large polyprotein including RNA-dependent RNA polymerase (RdRP) and ORF2 and ORF3 encode the major capsid protein (VP1) and the minor capsid protein (VP2), respectively^{5,6}. Since the 1990s, scientists have conducted more detailed studies on the genes and proteins of noroviruses. In mid-1990, numerous studies were published documenting various attempts to classify noroviruses through various methods—such as IEM, reverse transcription-PCR, and Southern hybridization—based on partial RdRP sequences or complete VP1 amino acid (aa) sequences^{7,8}. In the early stages, researchers classified them using IEM into a minimum of 4 or 6 antigenic types, but these antigenic classification schemes exhibited poor accuracy and reproducibility attributed to the cross-reactivity of antibodies^{9,10}. In research during the 2000s, noroviruses were classified into five genogroups and about 30 genetic clusters based on the VP1 protein sequences^{11–14}. Researchers examined the pairwise distances of strains, clusters, and genotypes using the conserved regions and domains of VP1. However, they observed that the ranges of the three categories overlapped, suggesting that distinguishing norovirus strains based on partial sequences alone may be challenging, leading to inconsistent and confused classification outcomes¹⁴.

¹Department of Life Science, Chung-Ang University, Seoul 06974, Korea. ²Thermo Fisher Scientific Inc, Seoul 06349, Korea. ³Department of Diseases Research, Incheon Metropolitan City Institute of Public Health and Environment, Incheon 22320, Korea. ⁴Department of Food and Nutrition, Chung-Ang University, Gyeonggi 17546, Korea. ✉email: eyun@cau.ac.kr

The conventional genetic classification of noroviruses is based on the aa sequences of the complete VP1 (genotype) or the nucleotide (nt) sequences of the ORF1 RdRP region (P-type)¹⁵. Thus, a dual nomenclature system (genotype + P-type) was introduced for the accurate identification of norovirus strains and is now routinely used in many laboratories worldwide^{15,16}. In 2019, the classification scheme for noroviruses was updated by proposing new genogroups and subtypes based on the 2× standard deviation criteria. In this scheme, noroviruses were divided into ten (GI–GX) genogroups, five of which (GI, GII, GIV, GVIII, and GIX) have the ability to infect humans¹⁷. GI and GII are generally detected in humans, with GII notably accounting for over 85% of norovirus infections^{18,19}. Previous studies have indicated that a majority of norovirus strains causing human infection are GII recombinants, particularly those of the GII.4 variants²⁰.

Recent studies on the genetic characteristics of noroviruses provide evidence supporting the necessity for additional considerations in their phylogenetic classification. Firstly, gene trees cannot fully represent the evolutionary histories due to their incongruence with species trees, especially in the presence of recombination^{21,22}. Recombination of noroviruses has been observed at the ORF2/3 overlap, within ORF2, and at the ORF1/ORF2 junction^{23–26}. The current dual-type system, which relies solely on the partial RdRP and complete VP1 sequences, cannot account for all of recombination events. Also, single-gene analyses often lack sufficient resolution and can sometimes produce conflicting results^{27,28}. Recently, some studies have argued that using multiple genes (or genomic sequences) to reconstruct phylogenies is more important for improved phylogenetic accuracy^{29–31}. Furthermore, VP1 exhibits a high degree of genetic diversity, suggesting its inadequacy as a proper molecular marker. Nevertheless, certain strains previously classified within the GII genogroup were reclassified as GIX and GVIII based on the VP1 classification, despite their high genomic similarity to GII^{17,32–35}. Moreover, GIV strains that infect cats, lions and dogs cluster with GVI strains based on RdRP sequences. Additionally, the similarity in their VP1 protein structure suggests that the genomes of animal-infecting GIV and GVI strains exhibit high similarity, regardless of genogroup^{17,36}.

Along with the aforementioned obstacles, another challenge in norovirus research is the extremely low levels of norovirus concentrations in environmental or stool samples³⁷. Hence, it is essential to precisely detect the norovirus types within samples using minimal analytical methods. Complicating matters further, there are numerous cases of co-infection with more than two types and the recombination in the environmental sources (e.g., oysters)^{38–41}. Accurately identifying their dual-types with the existing system is challenging, emphasizing the need for complementary genomic databases as well as RdRP and VP1 sequences.

In this study, we evaluate the genetic diversity of norovirus genomes to clarify their genomic characteristics and the criteria of existing classification. Thereafter, we reconstruct phylogenetic trees to compare the evolutionary relationships derived from gene-based and genome-based analyses.

Materials & methods

Data mining & identification

A total of 1417 norovirus genome sequences were downloaded from the National Center for Biotechnology Information (NCBI) database. Ten genogroups, including 50 genotypes and 71 P-types, were represented in the dataset (Table S1). From these, we extracted the nucleotide and peptide sequences of the ORFs using the Entrez retrieval system based on the accession numbers. The dual types from NCBI were updated through the Norovirus Typing Tool (ver. 2.0) (<https://www.rivm.nl/mpf/typ-ingtool/norovirus>) and phylogenetic analyses based on the RdRP and VP1 sequences.

Similarity plot

The ORF protein sequences of genogroups GI and GII were respectively aligned using MAFFT with the L-INS-I algorithm (ver. 7.505). The results were concatenated in the order of the ORFs for each genogroup. The percent similarities of sequences in the concatenated alignments were calculated using a Python script based on the sum-of-pairs scoring function with a sliding window of 5 aa and a step size of 1 aa⁴² and similarity plots were visualized in R (ver. 4.2.0) using the *ggplot2* package.

Evolutionary selection pressure

To examine positive selection acting on the norovirus GI and GII genogroups, we subsampled 100 sequences from established databases and used site models in codeml as implemented in the PAML software package (ver. 4.10.5)⁴³. We carried out likelihood ratio tests (LRTs) comparing a null model and an alternative model: M0 (one ratio) vs. M3 (discrete), M1a (nearly neutral) vs. M2a (positive selection), and M7 (beta) vs. M8 (beta&ω). Positively selected amino acid sites were identified based on Bayes empirical Bayes posterior probabilities. All PAML analyses were carried out using the F3×4 codon frequency model. The level of significance (*P*) for the LRTs was estimated using a χ^2 distribution with the corresponding degrees of freedom. The test statistic is calculated as twice the difference of the log-likelihood between the models ($2\Delta\ln L = 2[\ln L_1 - \ln L_0]$) where L_1 and L_0 are the likelihoods of the alternative and null models, respectively).

Global pairwise alignment

To assess the distance between genogroups, we used the LAGAN (Limited Area Global Alignment of Nucleotides) tool, which is an efficient and reliable pairwise aligner that is suitable for genomic comparisons of distantly related organisms⁴⁴. Global pairwise alignments produced by LAGAN were visualized with mVISTA. We compared the ten genogroups using GenBank genome sequences—accession numbers MT031988, JQ622197, JX145650, KC894731, KC792553, MW662289, OL757872, AB985418, MN473468, and KJ790198—as references for GI–GX, respectively.

To get the distance matrices of genotypes and P-types, we aligned the RdRP and VP1 nucleotide sequences of the various types using the G-INS-I algorithm in MAFFT (ver. 7.505). We generated distance matrices based on the alignments, including gaps, on the UGENE platform (ver. 48.1).

Phylogenetic analyses

We constructed phylogenetic trees of noroviruses using two methods: alignment-based and alignment-free. For the alignment-based tree, a multiple sequence alignment of the 1417 downloaded sequences and our 10 assembled results was generated using MAFFT with the L-INS-I algorithm (ver. 7.505). To determine the best-fit substitution models, the ModelFinder in IQ-TREE (ver. 1.6.12) was used. The phylogenetic trees were reconstructed by the maximum likelihood (ML) and Bayesian inference methods. The ML method was performed using RAxML-NG (ver. 1.1.0) with the GTR + F + I + G4 nucleotide substitution model and Bayesian phylogenetic inference was performed using the MrBayes package (ver. 3.2.7a) with the same model. The Markov chain Monte Carlo search was run for 10^6 generations with a sampling frequency of 5×10^2 using three heated and one cold chain. A method for the alignment-free tree is described in Supplementary Materials.

Results

Genomic diversity

In this study, a comprehensive analysis was conducted on 1427 norovirus genomes from the NCBI database and human stool samples. All genogroups were represented in these genomes, although some genotypes, such as GIII.3, GIV.NA1, GNA1.1, and GNA2.1, were not included due to the absence of their genome data (Table S1). For a more accurate analysis, we verified or revised the dual type of certain strains.

We assessed the genomic diversity of the GI and GII, which are the genogroups most commonly infecting humans. Considering the degeneracy in the third base of codons, we used the protein sequences of the three ORFs. Consistent with previous research findings^{45,46}, our genome database confirmed that the RdRP region (located at the 3' end of ORF1) is the most conserved region, while VP1 and VP2 exhibit greater variability (Figure 1). In ORF1, the N-terminal region displayed variability within the GI and GII genogroups and sequence conservation increased toward the C-terminus of the polyprotein. Amino acid positions 700 to 900, corresponding to the p22 protein, showed significantly lower similarity. Previous analyses of p22, one of the most variable genomic

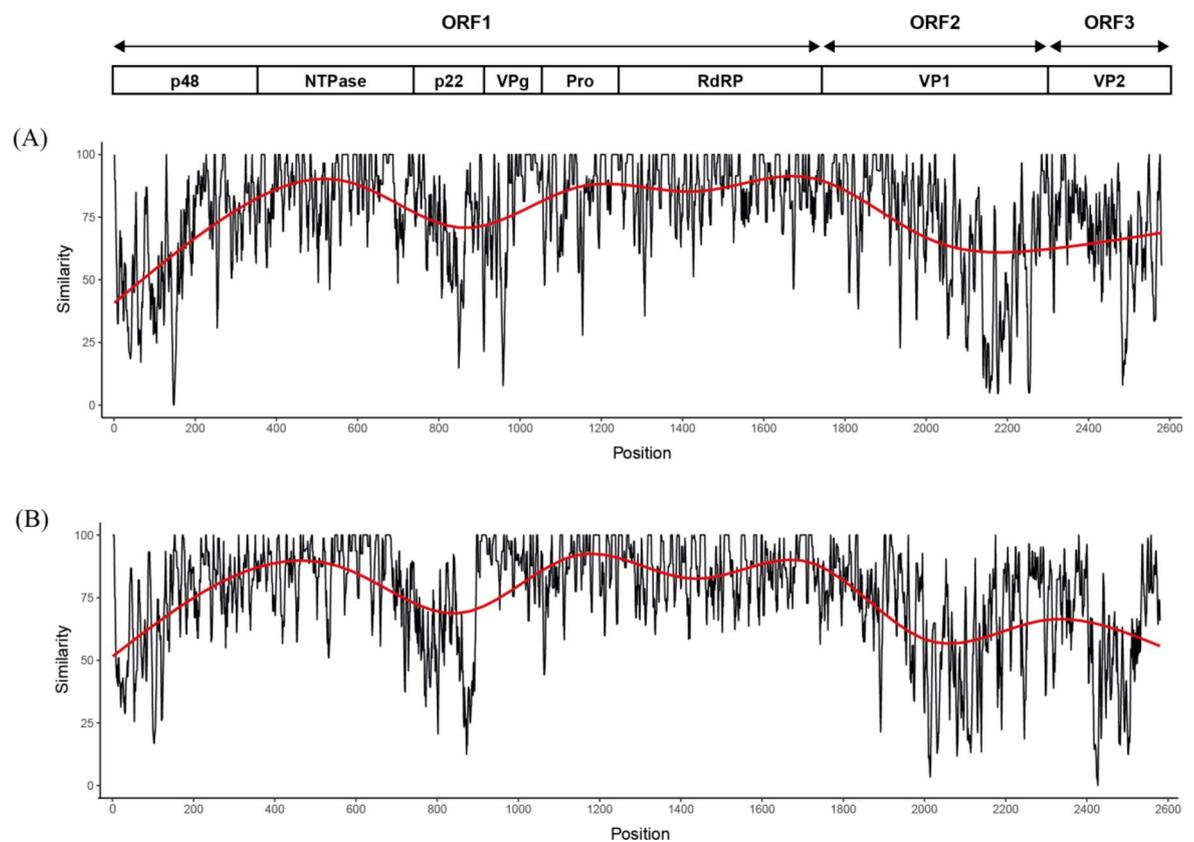


Figure 1. Similarity plots of norovirus GI genogroup and GII genogroup genomes. The plots are based on the concatenated sequences from 94 and 1161 complete genomes in GI and GII, respectively. Analyses were performed using the sum of pairs scoring with a sliding window of 5 amino acids (aa) and a step size of 1 aa. The plot depicts the percent similarity (Y-axis) of aa positions (X-axis). (A) Similarity plots of GI genomes. (B) Similarity plots of GII genomes. A schematic representation of the human norovirus open reading frames (ORFs) and the encoded proteins are shown above the graphs.

regions, revealed that it plays a role in Golgi disassembly and the antagonism of Golgi-dependent cellular protein secretion, which were observed during norovirus replication^{47,48}. We thus concluded that the conservation of the ORF1 polyprotein is not limited to the RdRP but extends across the majority of the sequence.

Despite the 5' end of ORF1 showing a similarity trendline of less than 50%, the initial five amino acids remained highly conserved (Figure S1). The sequence logo analysis showed that both ends of ORF1 have conserved nucleotide sequences in all genogroups. Upon translating the conserved nucleotides from the 5' and 3' ends of ORF1 into protein sequences, we observed an intriguing pattern. Most genogroups associated with strains infecting humans show identical deduced protein sequences at both ends. Since the sequence logos of the GVII, GVIII, and GX genogroups were constructed with one or two sequences due to their limited availability in the current genomic database, further research will be needed.

Selection pressure

To conduct a phylogenetic analysis, it is essential to identify genomic regions that contain sufficient phylogenetic signals. Thus, we measured the selective pressure for the three ORFs of both genogroup GI and GII. We carried out likelihood ratio tests (LRTs) comparing null and alternative codon substitution models. Across all ORFs, M3 was selected over M0 in the first comparison, indicating that the GI and GII genogroups have variable ω values among sites (Table 1). Following that, the null hypothesis M1a was consistently chosen over M2a, and the test was concluded. Consequently, no predicted positive selection sites were identified, but we confirmed that both GI and GII exhibit their lowest ω (d_N/d_S) ratios in ORF1. This suggests lower selection pressure on ORF1, signifying its phylogenetic significance compared to other ORFs. The capsid proteins of most viruses undergo rapid evolution to evade host immune detection, reach different host organs, and trigger pathological effects, ultimately promoting efficient transmission to new hosts. Our results also demonstrate that capsid proteins, encoded by ORF2 and ORF3, experience a high degree of selection pressure. Even though the major capsid protein, VP1 interacts directly with the entry receptors and antibodies of its host, VP2 showed a higher ω ratio than VP1. Although higher evolutionary rates in VP2 have been previously documented, the functional drivers behind the observed variability remain unclear^{49,50}. When comparing GI and GII, each ORF of GII exhibited a higher selection pressure value than its counterpart in GI.

Pairwise distances of norovirus types

We examined the sequence similarity at the whole genome level to figure out the probable genetic relationships within norovirus genogroups. A global pairwise alignment was performed based on genomic sequences of all ten genogroups. The alignments of GII with GVIII, GII with GIX, GVIII with GIX, and GIV with GVI revealed high degrees of similarity across their genomes, particularly in ORF1 and the ORF1/ORF2 junction, when compared to the other comparisons (Figures 2 and S2). To further characterize genome similarities, we counted the base pairs in conserved regions between genogroups (Figure 2). The GI and GII genogroups, which predominantly infect humans, shared the least conserved regions among the comparisons. The GV genogroup, the murine norovirus, distinctively possesses ORF4, which encodes virulence factor 1 (VF1), a mitochondria-localized protein that acts as an innate immune antagonist and contributes to viral adaptation during ongoing murine norovirus infection^{51,52}. In the figures, GV generally had low similarity with all other genogroups. Most notably, while the whole genome size is about 7.5 kb, genogroups GII, GVIII, and GIX shared conserved regions exceeding half of the genome size by a significant margin, as did groups GIV and GVI.

To clarify the sorting criteria among subtypes, including P-types and genotypes, we measured the pairwise distances of RdRP and VP1 nucleotide sequences of all types present in our dataset (Figure 3 and Tables S2 and S3). All sequences used in the subtype analysis were complete except for the GII.P38 RdRP sequence. In the P-type distance matrix (Figure 3A and Table S2), the minimum and maximum identity values were 55% and 95%, respectively. Intra-genogroup identities were 71–91% in GI, 71–92% in GII, 95% in GIII, 79% in GIV, 67% in GV, and 80% in GVI. The results indicated that the inter-genogroup identity range for P-types is 55–70%, and intra-genogroup identity exceeds 70%. Notably, intra-genogroup identity within GV, between GV.P1 and GV.P2, is relatively low. In the genotype distance matrix (Figure 3B and Table S3), where the percent identity ranges from a minimum of 47% to a maximum of 87%, the values were largely lower than those for the P-types. Intra-genogroup identities of VP1 were 67–75% in GI, 65–87% in GII, 73% in GIII, 68–74% in GIV, 68% in GV, and 65% in GVI. It could be inferred that the inter-genogroup identity is less than 65%. Ironically, GIX.1 showed 65% identity with some GII genotypes, equivalent to the intra-genogroup identity of GII and GVI, while it also had values greater than 62% with all GII types. Among the alignments, identity scores of 80% or higher were only evident in genotypes GII.22–GII.27, GII.NA1, and GII.NA2, which were identified recently.

Phylogenomic analysis

Since the d_N/d_S ratio of ORF1 implied their phylogenetic significance, we reconstructed two norovirus phylogenies based respectively on this region and genomic sequences using the ML method. The phylogenomic analysis, including the downloaded dataset and assembled genomes, was performed based on the complete or partial genome nucleotide sequences. This tree's topology was identical to that of the ORF1-based tree, indicating that the phylogenetic relationships of most genogroups were well-supported by the genomic sequences (Figures S3A and S4A). Consistent with the pairwise distances, the trees showed that the GV genogroup had distant phylogenetic relationships with all other genogroups and that there was a notable genetic distance between groups GI and GII.

However, in this genomic based tree, GVIII and GIX—two genogroups (formerly GII) that had been reclassified through a highly variable VP1-based analysis¹⁷—were found to be part of the same clade as GII (Figures 4 and 5B). This result, along with the pairwise distance analysis, strongly indicates a high degree of genetic similarity among the genomes of the GII, GVIII, and GIX genogroups, as well as an ability to effectively

	Model	d_N/d_S	Estimates of parameters	lnL	$2\Delta\ln L$	p-value	Models compared
GI							
ORF1	M0	0.02362	$\omega = 0.02362$	-61193.22009	3088.85006	9.4877	M0 versus M3
	M3	0.02730	$p_0 = 0.68919, p_1 = 0.25076 (p_2 = 0.06005)$ $\omega_0 = 0.00188, \omega_1 = 0.04223, \omega_2 = 0.25705$	-59648.79506			
	M1a	0.05440	$p_0 = 0.96035 (p_1 = 0.03965)$ $\omega_0 = 0.01536, \omega_1 = 1$	-60602.65793	4E-05	5.9915	M1a versus M2a
	M2a	0.05440	$p_0 = 0.96035, p_1 = 0.03965 (p_2 = 0.00000)$ $\omega_0 = 0.01536, \omega_1 = 1, \omega_2 = 7.67257$	-60602.65795			
ORF2	M0	0.02541	$\omega = 0.02541$	-21797.45591	2094.41732	9.4877	M0 versus M3
	M3	0.02500	$p_0 = 0.54586, p_1 = 0.31114 (p_2 = 0.14300)$ $\omega_0 = 0.00053, \omega_1 = 0.02181, \omega_2 = 0.12501$	-20750.24725			
	M1a	0.09780	$p_0 = 0.91942 (p_1 = 0.08058)$ $\omega_0 = 0.01872, \omega_1 = 1$	-21504.19462	0.00030	5.9915	M1a versus M2a
	M2a	0.09780	$p_0 = 0.91942, p_1 = 0.04448 (p_2 = 0.03610)$ $\omega_0 = 0.01872, \omega_1 = 1, \omega_2 = 1$	-21504.19447			
ORF3	M0	0.02922	$\omega = 0.02922$	-9224.85787	920.21614	9.4877	M0 versus M3
	M3	0.02730	$p_0 = 0.68919, p_1 = 0.25076 (p_2 = 0.06005)$ $\omega_0 = 0.00188, \omega_1 = 0.04223, \omega_2 = 0.25705$	-8764.74980			
	M1a	0.12550	$p_0 = 0.89315 (p_1 = 0.10685)$ $\omega_0 = 0.02084, \omega_1 = 1$	-9079.34467	0.00032	5.9915	M1a versus M2a
	M2a	0.12550	$p_0 = 0.89315, p_1 = 0.04742 (p_2 = 0.05943)$ $\omega_0 = 0.02084, \omega_1 = 1, \omega_2 = 1$	-9079.34484			
GII							
ORF1	M0	0.02854	$\omega = 0.02854$	-85976.53396	5015.70878	9.4877	M0 versus M3
	M3	0.03090	$p_0 = 0.69068, p_1 = 0.24094 (p_2 = 0.06838)$ $\omega_0 = 0.00294, \omega_1 = 0.05084, \omega_2 = 0.24296$	-83468.67957			
	M1a	0.05780	$p_0 = 0.96228 (p_1 = 0.03772)$ $\omega_0 = 0.02084, \omega_1 = 1$	-85212.00749	0	5.9915	M1a versus M2a
	M2a	0.05780	$p_0 = 0.96228, p_1 = 0.03772 (p_2 = 0.00000)$ $\omega_0 = 0.02084, \omega_1 = 1, \omega_2 = 2.88909$	-85212.00749			
ORF2	M0	0.06469	$\omega = 0.06469$	-54006.90074	6199.48555	9.4877	M0 versus M3
	M3	0.07090	$p_0 = 0.50059, p_1 = 0.35924 (p_2 = 0.14017)$ $\omega_0 = 0.00281, \omega_1 = 0.06943, \omega_2 = 0.31769$	-50907.15796			
	M1a	0.17250	$p_0 = 0.87051 (p_1 = 0.12949)$ $\omega_0 = 0.04946, \omega_1 = 1$	-52599.29607	0.00013	5.9915	M1a versus M2a
	M2a	0.17250	$p_0 = 0.87051, p_1 = 0.08360 (p_2 = 0.04589)$ $\omega_0 = 0.04946, \omega_1 = 1, \omega_2 = 1$	-52599.29614			
ORF3	M0	0.10407	$\omega = 0.10407$	-18677.05316	2227.55627	9.4877	M0 versus M3
	M3	0.11480	$p_0 = 0.48691, p_1 = 0.26937 (p_2 = 0.24372)$ $\omega_0 = 0.00884, \omega_1 = 0.09880, \omega_2 = 0.34426$	-17563.27503			
	M1a	0.32360	$p_0 = 0.72288 (p_1 = 0.27712)$ $\omega_0 = 0.06426, \omega_1 = 1$	-18103.28879	0.000326	5.9915	M1a versus M2a
	M2a	0.32360	$p_0 = 0.72288, p_1 = 0.19246 (p_2 = 0.08466)$ $\omega_0 = 0.06426, \omega_1 = 1, \omega_2 = 1$	-18103.28879			

Table 1. Selection pressures (d_N/d_S) and statistical test values for the three ORFs in genogroups GI and GII. For the tests, we subsampled 100 sequences from established databases and used site models in codeml as implemented in the PAML software package (ver. 4.10.5). Using likelihood ratio tests (LRTs), null models and alternative models were sequentially compared: M0 (one ratio) vs. M3 (discrete), M1a (nearly neutral) vs. M2a (positive selection), and M7 (beta) vs. M8 (beta& ω). All PAML analyses were carried out using the F3X4 codon frequency model. * ω , d_N/d_S ratio; p_n , proportion of codons in each class; $p < 0.05$ was considered significant.

distinguish between GII.4 variants and their recombinants (Figure 5A). Moreover, GII dual types with swine as hosts were conclusively categorized alongside strains that infect humans (Figure 5B and Table 2). We also reconstructed a tree solely for the GII clustering, which is the predominant genogroup associated with human diseases (Figures S3B and S4B). In the tree, strains can be divided into three major groups, named GII.A, GII.B, and GII.C. The GII.A clade, encompassing strains with P-types P4, P12, P16, P21, and P31, included prominent variants like GII.4 and GII.17, which collectively account for a significant proportion of infections. Strains with P-types P6, P7, and P8 were classified within the GII.B clade, while types recently reported to be in GII.C clustered together. Variant GVIII.1 [GII.P28] was affiliated with GII.A, and variant GIX.1 [GII.P15] was grouped within GII.B.

Furthermore, there was a mixing of branches between GIV and GVI based on their host specificity. Upon confirming their hosts, the GIV strains that infect animals were grouped together within the GVI genogroup,

	GI	GII	GIII	GIV	GV	GVI	GVII	GVIII	GIX	GX
GI	–	1338	2073	356	426	415	284	703	1075	351
GII		–	593	3238	616	2880	1824	5467	4848	1510
GIII			–	780	153	552	830	633	638	217
GIV				–	617	4200	1879	2710	3162	1786
GV					–	439	287	594	666	651
GVI						–	2261	2298	2735	2687
GVII							–	1764	1576	1731
GVIII								–	4131	1237
GIX									–	1165
GX										–

Figure 2. Alignment plots and total base pairs of representative genome sequences of the ten norovirus genogroups. In the plots, regions with over 70% identity in a 150 bp sliding window are marked in blue. The analysis used GenBank genome sequences—accession numbers MT031988, JQ622197, JX145650, KC894731, KC792553, MW662289, OL757872, AB985418, MN473468, and KJ790198—as references for genogroups GI–GX, respectively.

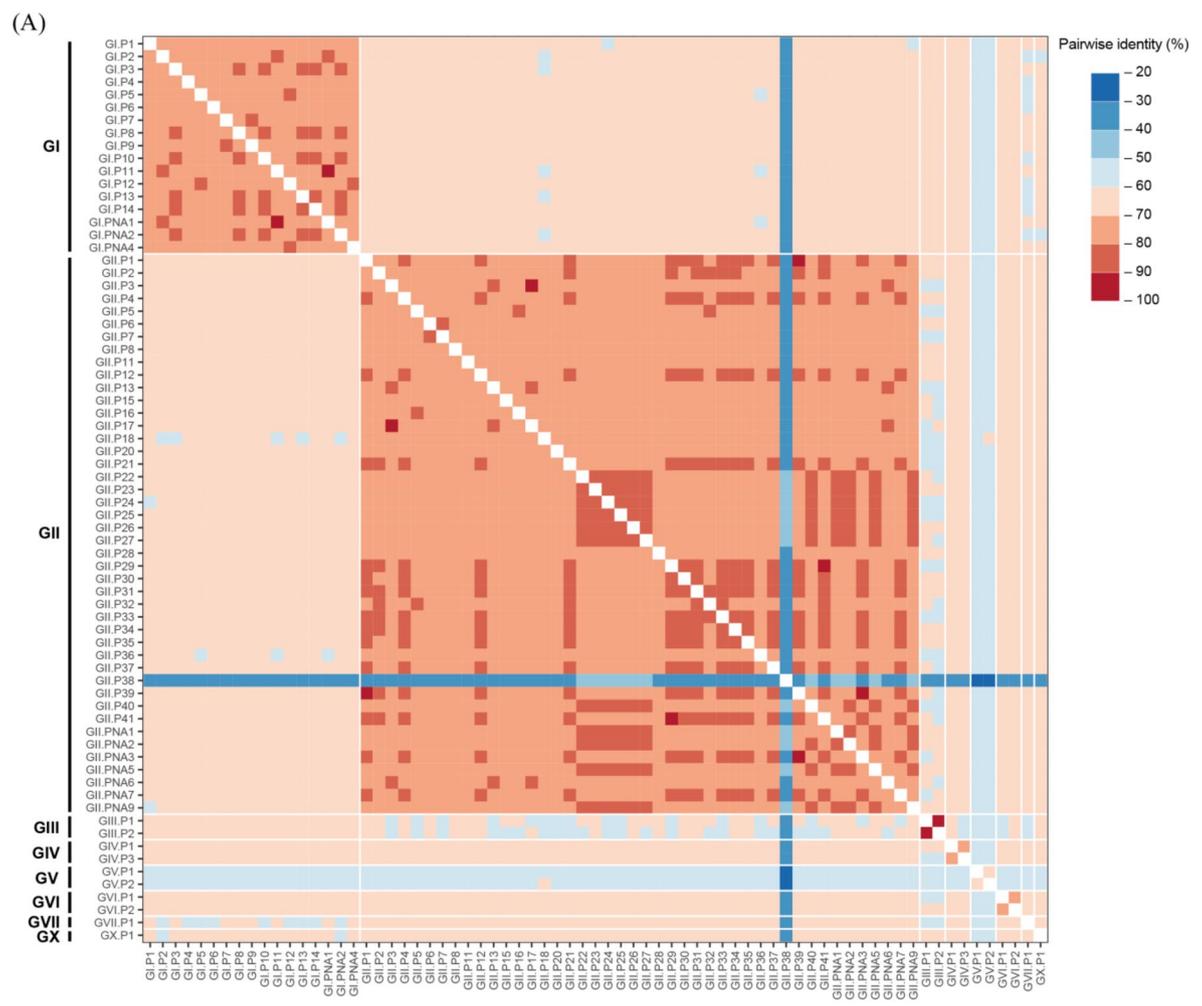


Figure 3. Pairwise distance matrices of the P-types and genotypes of strains from the ten genogroups. Vertical and horizontal lines separate the types into ten genogroups. Percent sequence identity is indicated by the color-coded boxes. (A) The pairwise alignments between the RdRP sequences of 71 P-types are plotted. Only the GII.P38 sequence is partial. (B) The alignments from the VP1 sequences of 50 genotypes are represented.

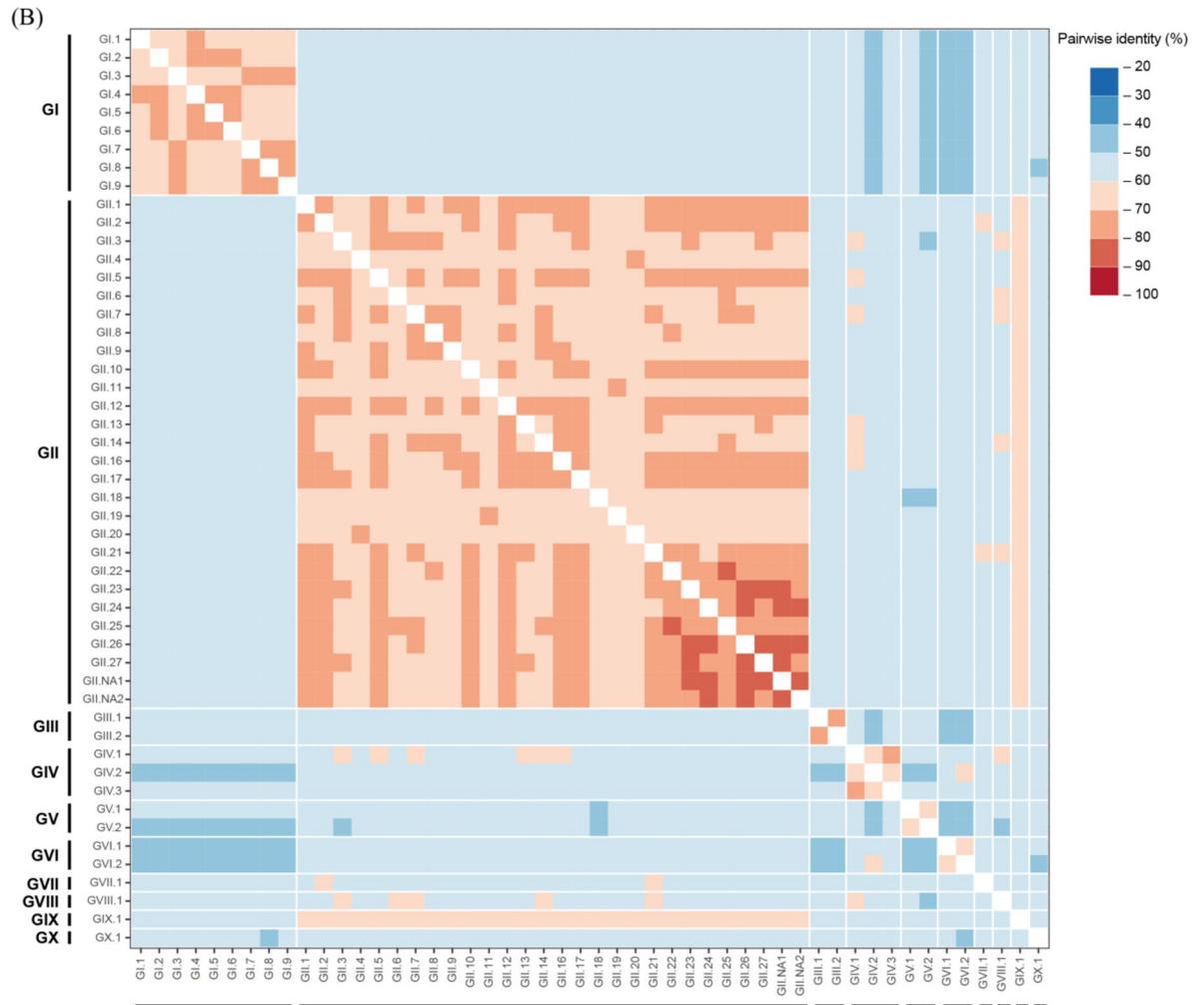


Figure 3. (continued)

which specifically targeting only carnivores and human noroviruses GIV.1 [GIV.P1] and GIV.3 [GIV.P3] were clustered into same clade (Figure 5C and Table 2).

Discussion

Noroviruses are regarded as rapidly evolving viruses with a large host range and present an extensive diversity driven by the accumulation of point mutations and recombination. Presently, their classification is determined by VP1 (genogroups and genotypes) and RdRP (P-types)^{15,16}. The number of genogroups has been expanded to ten (GI–GX), with some genotypes having been recently updated¹⁷. Research focusing on VP1 is essential for the prevention and treatment of norovirus infections. However, due to the rapid evolution of this protein and recombination events at the three regions (ORF1/2 and ORF2/3 junction, and within ORF2), gene-based analysis may inadequately reflect phylogenomic history of the genus, as exemplified by GVIII and GIX. Since gene trees do not always align with the species tree topology, it is essential to incorporate genome sequence analysis to comprehend the evolutionary history of a species^{53–56}. Moreover, since environmental samples can be co-infected with more than two types, relying solely on RdRP and VP1 typing is inadequate for accurately identifying norovirus strains within them. Therefore, in this study, we have detailed the criteria for genotypes and P-types and established a comparison of the phylogenetic relationships between gene-based and whole-genome-based analysis to achieve a more precise evolutionary lineage of the genus *Norovirus*.

According to prior research, the hypervariable VP2 region may interact with its VP1 interaction domain, and VP2 could function in the stability of norovirus particles or in regulating the maturation of antigen-presenting cells and protective immunity induction in a virus-strain-specific manner^{57–59}. Moreover, VP2 seems to undergo covariation with VP1 in the GII, GIV, and GVI genogroups^{36,49,60}. Our genomic diversity analyses also indicated the conservation pattern of norovirus genomes and the variability and high ω (d_N/d_S) ratios in the two capsid proteins, supporting their coevolution. Furthermore, it was observed that ORF1 carries a significant phylogenetic signal, playing a crucial role in the evolutionary trajectory of noroviruses. We also measured the criteria for current subtypes and observed some genotypes exhibit overlapped range of intra-genogroup and

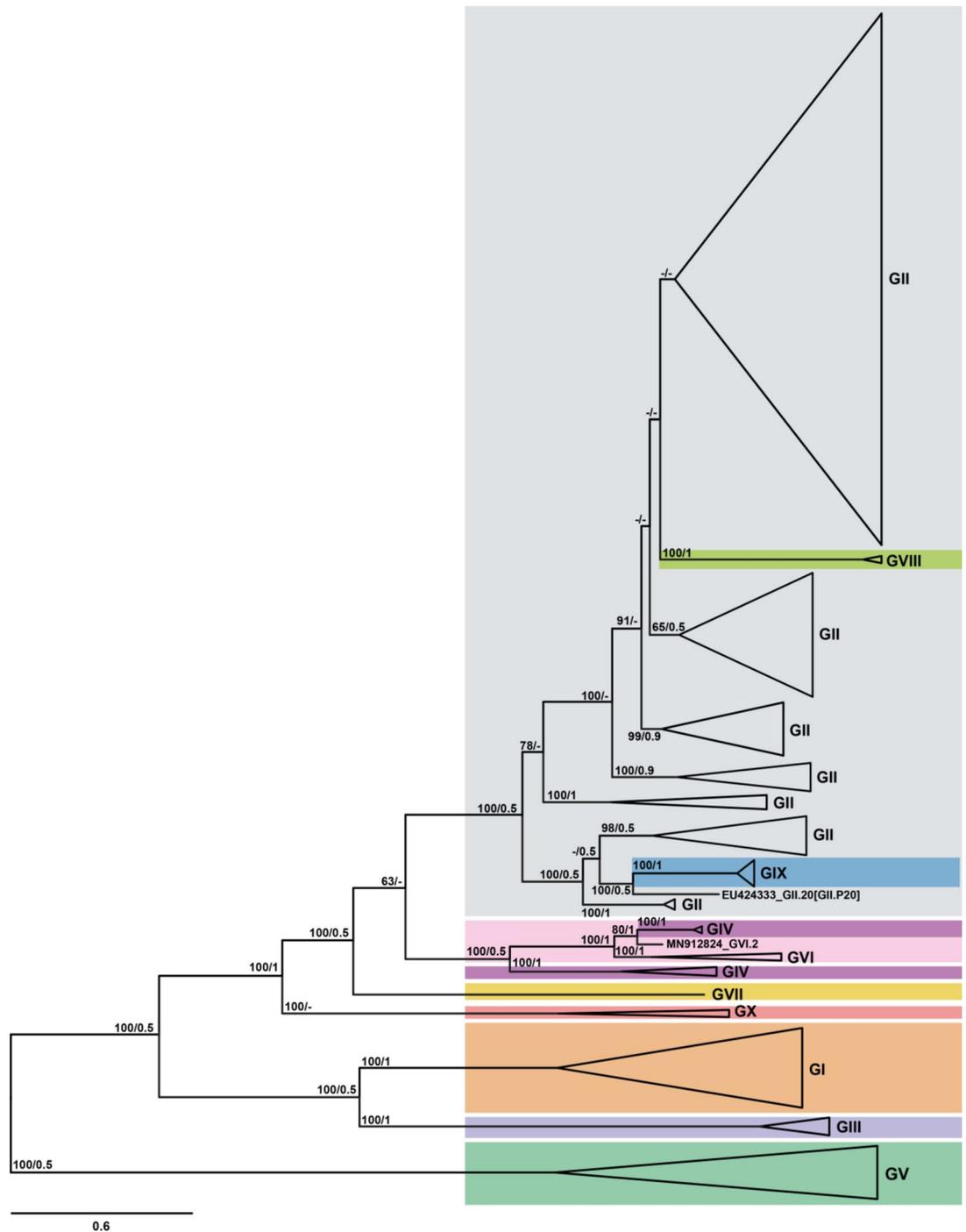


Figure 4. Midpoint-rooted phylogenomic trees of norovirus ten genogroups. The phylogenetic trees were reconstructed by the maximum likelihood and Bayesian inference methods based on the 1417 downloaded sequences and our 10 assembled results. Branches were collapsed by genogroup. Bootstrap value and Bayesian Inference posterior are depicted on branches, and dash (-) indicates with PPBI < 50% or ML < 60%.

inter-genogroup similarity. Consequently, we inferred that the gene-based classification could not present the phylogenetic relationships of genus *Norovirus*.

Since the mid-1990s, norovirus GII.4 variants have been responsible for 62 to 80% of norovirus outbreaks globally and contributed to at least six pandemics of acute gastroenteritis⁶¹. Additionally, intragenotype recombination within GII.4 has the potential to give rise to new GII.4 variants, further hastening the occurrence of pandemics^{62,63}. Our phylogenomic tree can distinguish each dual type and even intragenotype recombinant strains of GII.4. This feature also enables the accurate type prediction of norovirus strains, even with short reads from environmental or stool samples. Additionally, the whole-genome-based tree showed that the GIV, GVI, GVIII, and GIX strains segregate independently of their corresponding capsid genogroups. GVIII and

(A)

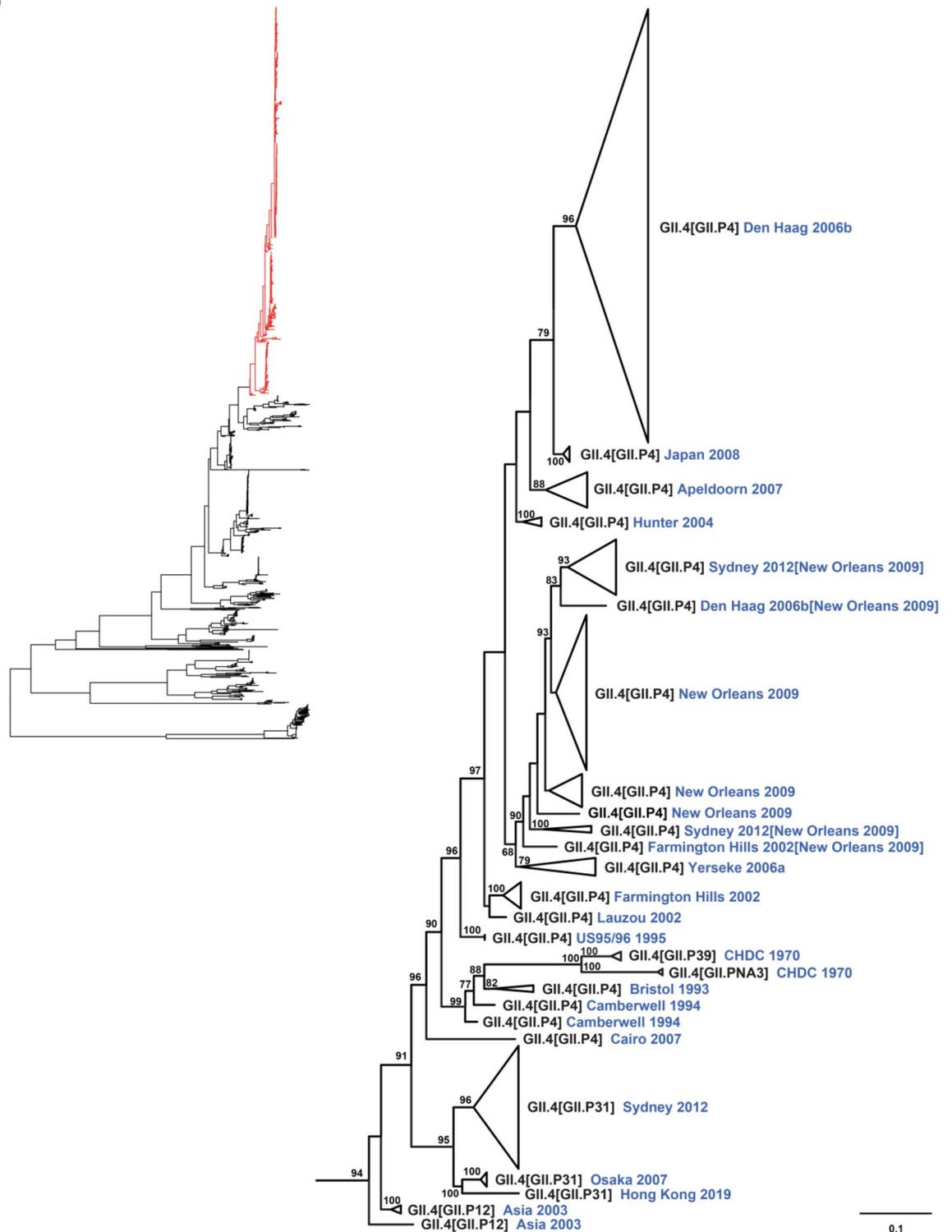


Figure 5. Midpoint-rooted phylogenetic trees of norovirus strains. The phylogenetic trees were reconstructed using ML method based on the 1417 downloaded sequences and our 10 assembled results. Branches were collapsed by dual-types and bootstrap values above 60% are depicted on branches. (A) Phylogenetic tree of GII strains. (B) Phylogenetic tree of rest of GII strains, GVIII, and GIX strains. (C) Phylogenetic tree of norovirus strains except for GII, GVIII, and GIX.

GIX, previously known as GII, were reclassified through an analysis based on the highly variable VP1 region. Despite being categorized into different genotypes based solely on VP1 sequences, our study confirmed that their genomes closely resemble those of GII strains, as demonstrated in the alignment plot (Figure 2), the phylogenetic tree (Figure 4), and the sequence similarity networks (Figure S5). Notably in Figure 2, the total

(B)

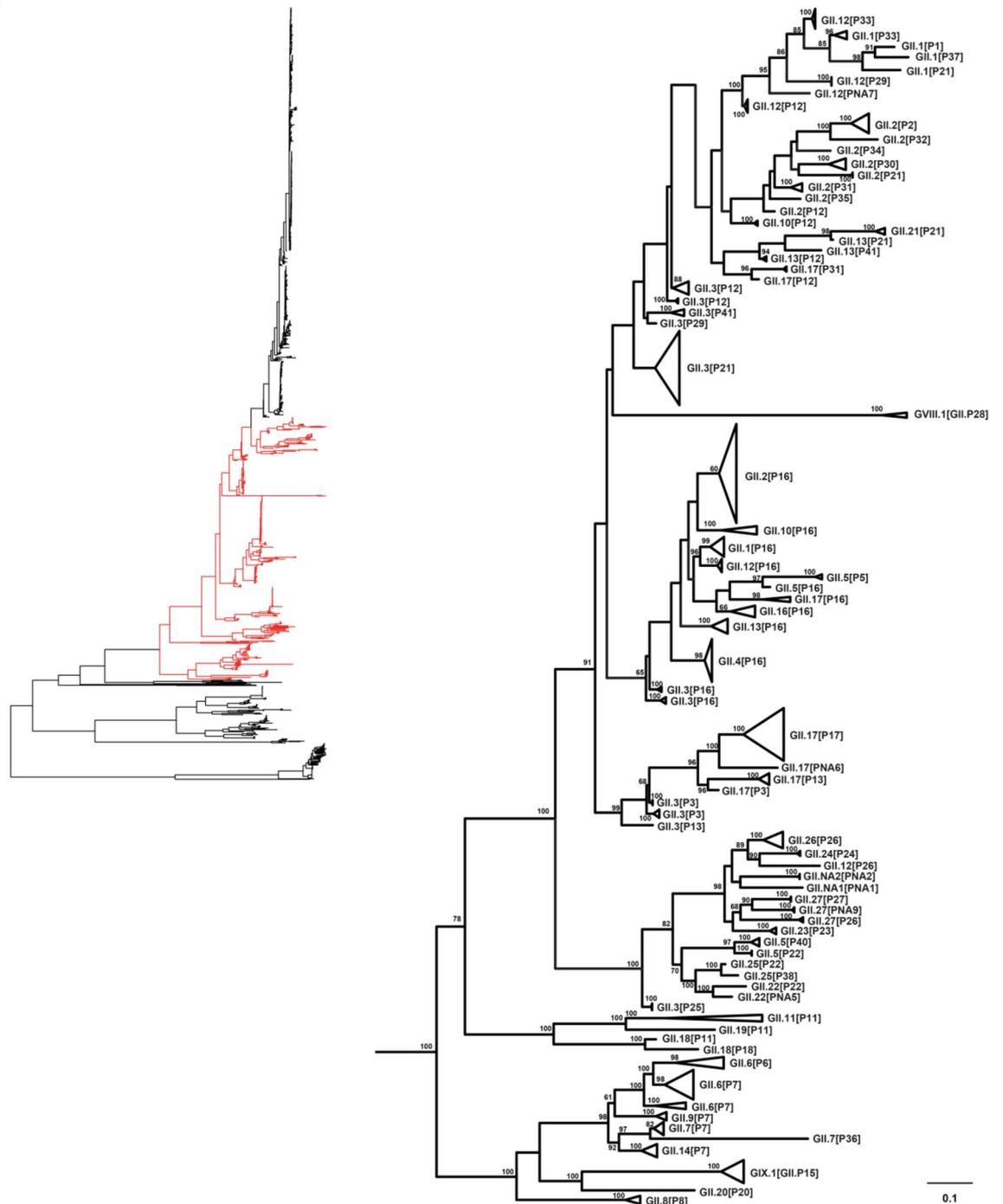


Figure 5. (continued)

conserved base pairs are noticeable, with the GII genogroup sharing over 4800 bp (64% of genome length) with GVIII and GIX, and GIV sharing 4200 bp with GVI. In the GII clade containing GVIII and GIX, the global human pathogen P-types GII.P4, GII.P7, GII.P12, GII.P16, GII.P21, and GII.P31 are exclusively found in GII.A and GII.B⁶⁴. Currently, there are no available drugs or vaccines for treating or preventing norovirus disease in humans⁶⁵. Targeting the GII.A and GII.B groups, which include the globally common P-types, can cover a broad spectrum of norovirus strains, and a heterologous cross-protection in prevention and treatment can be expected.

The GIV and GVI strains were subdivided into two clades based on not the capsid sequences but their infection hosts. GIV.1 and GIV.3, which infect humans, possessed the RdRP and VP1 of GIV, whereas GIV.2 and GVI strains, which are the carnivore noroviruses, regardless of the capsid protein, had the RdRP of GVI. Moreover, the predicted cleavage sites for the ORF1 polyproteins of GIV and GVI viruses demonstrated conservation in both location and amino acid sequence by host, rather than genogroup³⁶. Furthermore, a structure analysis revealed that the VP1 of GIV.2 has a large loop insertion in the P-domain, a characteristic

animals, it may ultimately lead to the emergence of a strain capable of infecting humans. Our inference regarding the interactions between human and animal viruses leads us to assert the potential of zoonotic transmission.

Conclusions

In conclusion, we conducted a comprehensive analysis to enhance the phylogenetic interpretation of norovirus evolution. As a result, we identified their genomic characteristics and the thresholds for the identity range of inter-genogroup and intra-genogroup in the current classification system. Thereafter, we reconstructed a phylogenomic tree of norovirus strains to compare the evolutionary relationships between gene-based and the whole genome-based study. Genome-based classification can be used to detect norovirus dual types accurately from environmental samples and identify emerging recombinants. Overall, our study marks a significant initial step towards the phylogenomic classification of the genus *Norovirus*, valuable not only for interpreting the evolutionary relationships among norovirus strains but also for antiviral targeting.

Data availability

The raw sequencing data for this study are available from the NCBI Sequence Read Archive (BioProject: PRJ-NA1054470, SRA: SRR27336526–SRR27336535).

Received: 28 August 2024; Accepted: 21 January 2025

Published online: 13 February 2025

References

- Ahmed, S. M. et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect. Dis.* **14**, 725–730. [https://doi.org/10.1016/s1473-3099\(14\)70767-4](https://doi.org/10.1016/s1473-3099(14)70767-4) (2014).
- Glass, R. I., Parashar, U. D. & Estes, M. K. Norovirus gastroenteritis. *N. Engl. J. Med.* **361**, 1776–1785. <https://doi.org/10.1056/NEJMra0804575> (2009).
- Kapikian, A. Z. et al. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J. Virol.* **10**, 1075–1081. <https://doi.org/10.1128/jvi.10.5.1075-1081.1972> (1972).
- Lambden, P. R., Caul, E. O., Ashley, C. R. & Clarke, I. N. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259**, 516–519. <https://doi.org/10.1126/science.8380940> (1993).
- Thorne, L. G. & Goodfellow, I. G. Norovirus gene expression and replication. *J. Gen. Virol.* **95**, 278–291. <https://doi.org/10.1099/vir.0.059634-0> (2014).
- Jiang, X., Wang, M., Wang, K. & Estes, M. K. Sequence and genomic organization of Norwalk virus. *Virology* **195**, 51–61. <https://doi.org/10.1006/viro.1993.1345> (1993).
- Green, S. M., Lambden, P. R., Caul, E. O., Ashley, C. R. & Clarke, I. N. Capsid diversity in small round-structured viruses: molecular characterization of an antigenically distinct human enteric calicivirus. *Virus Res.* **37**, 271–283. [https://doi.org/10.1016/0168-1702\(95\)00041-n](https://doi.org/10.1016/0168-1702(95)00041-n) (1995).
- Ando, T. et al. Detection and differentiation of antigenically distinct small round-structured viruses (Norwalk-like viruses) by reverse transcription-PCR and southern hybridization. *J. Clin. Microbiol.* **33**, 64–71. <https://doi.org/10.1128/jcm.33.1.64-71.1995> (1995).
- Ando, T., Noel, J. S. & Fankhauser, R. L. Genetic classification of Norwalk-like viruses. *J. Infect. Dis.* **181** (Suppl 2), 336–348. <https://doi.org/10.1086/315589> (2000).
- Lewis, D., Ando, T., Humphrey, C. D., Monroe, S. S. & Glass, R. I. Use of solid-phase immune electron microscopy for classification of Norwalk-like viruses into six antigenic groups from 10 outbreaks of gastroenteritis in the United States. *J. Clin. Microbiol.* **33**, 501–504. <https://doi.org/10.1128/jcm.33.2.501-504.1995> (1995).
- Fankhauser, R. L. et al. Epidemiologic and molecular trends of Norwalk-like viruses associated with outbreaks of gastroenteritis in the United States. *J. Infect. Dis.* **186**, 1–7. <https://doi.org/10.1086/341085> (2002).
- Oliver, S. L. et al. Molecular characterization of bovine enteric caliciviruses: a distinct third genogroup of noroviruses (Norwalk-like viruses) unlikely to be of risk to humans. *J. Virol.* **77**, 2789–2798. <https://doi.org/10.1128/jvi.77.4.2789-2798.2003> (2003).
- Koopmans, M. et al. Molecular epidemiology of human enteric caliciviruses in the Netherlands. *J. Infect. Dis.* **181** (Suppl 2), 262–269. <https://doi.org/10.1086/315573> (2000).
- Zheng, D. P. et al. Norovirus classification and proposed strain nomenclature. *Virology* **346**, 312–323. <https://doi.org/10.1016/j.virol.2005.11.015> (2006).
- Kroneman, A. et al. Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* **158**, 2059–2068. <https://doi.org/10.1007/s00705-013-1708-5> (2013).
- Vinjí, J. Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* **53**, 373–381. <https://doi.org/10.1128/jcm.01535-14> (2015).
- Chhabra, P. et al. Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* **100**, 1393–1406. <https://doi.org/10.1099/jgv.0.001318> (2019).
- Rouhani, S. et al. Norovirus infection and acquired immunity in 8 countries: results from the MAL-ED study. *Clin. Infect. Dis.* **62**, 1210–1217. <https://doi.org/10.1093/cid/ciw072> (2016).
- Lo, M. et al. Genetic characterization and evolutionary analysis of norovirus genotypes circulating among children in eastern India during 2018–2019. *Arch. Virol.* **166**, 2989–2998. <https://doi.org/10.1007/s00705-021-05197-6> (2021).
- Navarro-Lleó, N. et al. Recombinant noroviruses circulating in Spain from 2016 to 2020 and proposal of two novel genotypes within Genogroup I. *Microbiol. Spectr.* **10**, e0250521. <https://doi.org/10.1128/spectrum.02505-21> (2022).
- Degnan, J. H. & Rosenberg, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340. <https://doi.org/10.1016/j.tree.2009.01.009> (2009).
- Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H. & Jansen, R. K. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Mol. Phylogenet. Evol.* **138**, 219–232. <https://doi.org/10.1016/j.ympev.2019.05.022> (2019).
- Medici, M. C. et al. Novel recombinant GII.P16_GII.13 and GII.P16_GII.3 norovirus strains in Italy. *Virus Res.* **188**, 142–145. <https://doi.org/10.1016/j.virusres.2014.04.005> (2014).
- Mahar, J. E., Bok, K., Green, K. Y. & Kirkwood, C. D. The importance of intergenic recombination in norovirus GII.3 evolution. *J. Virol.* **87**, 3687–3698. <https://doi.org/10.1128/jvi.03056-12> (2013).
- 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.* **87**, 6270–6282. <https://doi.org/10.1128/jvi.03464-12> (2013).
- Mans, J. et al. Norovirus diversity in children with gastroenteritis in South Africa from 2009 to 2013: GII.4 variants and recombinant strains predominate. *Epidemiol. Infect.* **144**, 907–916. <https://doi.org/10.1017/S0950268815002150> (2016).

27. Philippe, H. Opinion: long branch attraction and protist phylogeny. *Protist* **151**, 307–316. [https://doi.org/10.1078/s1434-4610\(04\)70029-2](https://doi.org/10.1078/s1434-4610(04)70029-2) (2000).
28. Nickrent, D. L., Parkinson, C. L., Palmer, J. D. & Duff, R. J. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* **17**, 1885–1895. <https://doi.org/10.1093/oxfordjournals.molbev.a026290> (2000).
29. Rosenberg, M. S. & Kumar, S. Taxon sampling, bioinformatics, and phylogenomics. *Syst. Biol.* **52**, 119–124. <https://doi.org/10.1080/10635150390132894> (2003).
30. Rosenberg, M. S. & Kumar, S. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proc. Natl. Acad. Sci. USA* **98**, 10751–10756 (2001). <https://doi.org/10.1073/pnas.191248498>
31. Johnston, P. R. et al. A multigene phylogeny toward a new phylogenetic classification of Leotiomycetes. *IMA Fungus*. **10**, 1. <https://doi.org/10.1186/s43008-019-0002-x> (2019).
32. Okada, M., Ogawa, T., Kaiho, I. & Shinozaki, K. Genetic analysis of noroviruses in Chiba Prefecture, Japan, between 1999 and 2004. *J. Clin. Microbiol.* **43**, 4391–4401. <https://doi.org/10.1128/jcm.43.9.4391-4401.2005> (2005).
33. Matsushima, Y. et al. Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro. Surveill.* **20** <https://doi.org/10.2807/1560-7917.es2015.20.26.21173> (2015).
34. Mathijs, E. et al. Novel norovirus recombinants and of GII.4 sub-lineages associated with outbreaks between 2006 and 2010 in Belgium. *Virolog. J.* **8**, 310. <https://doi.org/10.1186/1743-422x-8-310> (2011).
35. Kim, Y. E. et al. Phylogenetic characterization of norovirus strains detected from sporadic gastroenteritis in Seoul during 2014–2016. *Gut Pathog.* **10**, 36. <https://doi.org/10.1186/s13099-018-0263-8> (2018).
36. Ford-Siltz, L. A. et al. Genomics analyses of GIV and GVI noroviruses reveal the distinct clustering of human and animal viruses. *Viruses* **11** <https://doi.org/10.3390/v11030204> (2019).
37. Le Guyader, F. S. et al. Detection and quantification of noroviruses in shellfish. *Appl. Environ. Microbiol.* **75**, 618–624. <https://doi.org/10.1128/aem.01507-08> (2009).
38. Analysis of the European. Baseline survey of norovirus in oysters. *Efsa j.* **17**, e05762. <https://doi.org/10.2903/j.efsa.2019.5762> (2019).
39. Lowther, J. A., Gustar, N. E., Powell, A. L., Hartnell, R. E. & Lees, D. N. Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom. *Appl. Environ. Microbiol.* **78**, 5812–5817. <https://doi.org/10.1128/aem.01046-12> (2012).
40. Webby, R. J. et al. Internationally distributed frozen oyster meat causing multiple outbreaks of norovirus infection in Australia. *Clin. Infect. Dis.* **44**, 1026–1031. <https://doi.org/10.1086/512807> (2007).
41. Cheng, P. K., Wong, D. K., Chung, T. W. & Lim, W. W. Norovirus contamination found in oysters worldwide. *J. Med. Virol.* **76**, 593–597. <https://doi.org/10.1002/jmv.20402> (2005).
42. Carrillo, H. & Lipman, D. The multiple sequence Alignment Problem in Biology. *SIAM J. Appl. Math.* **48**, 1073–1082. <https://doi.org/10.1137/0148063> (1988).
43. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556. <https://doi.org/10.1093/bioinformatics/13.5.555> (1997).
44. Brudno, M. et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731. <https://doi.org/10.1101/gr.926603> (2003).
45. Deval, J., Jin, Z., Chuang, Y. C. & Kao, C. C. Structure(s), function(s), and inhibition of the RNA-dependent RNA polymerase of noroviruses. *Virus Res.* **234**, 21–33. <https://doi.org/10.1016/j.virusres.2016.12.018> (2017).
46. Belliot, G. et al. Norovirus proteinase-polymerase and polymerase are both active forms of RNA-dependent RNA polymerase. *J. Virol.* **79**, 2393–2403. <https://doi.org/10.1128/jvi.79.4.2393-2403.2005> (2005).
47. Sharp, T. M., Guix, S., Katayama, K., Crawford, S. E. & Estes, M. K. Inhibition of cellular protein secretion by Norwalk virus nonstructural protein p22 requires a mimic of an endoplasmic reticulum export signal. *PLoS One* **5**, e13130. <https://doi.org/10.1371/journal.pone.0013130> (2010).
48. Cotten, M. et al. Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. *J. Virol.* **88**, 11056–11069. <https://doi.org/10.1128/jvi.01333-14> (2014).
49. Hong, X., Xue, L., Gao, J., Jiang, Y. & Kou, X. Epochal coevolution of minor capsid protein in norovirus GII.4 variants with major capsid protein based on their interactions over the last five decades. *Virus Res.* **319**, 198860. <https://doi.org/10.1016/j.virusres.2022.198860> (2022).
50. Zhou, N., Li, M., Zhou, L. & Huang, Y. Genetic characterizations and molecular evolution of human norovirus GII.6 genotype during the past five decades. *J. Med. Virol.* **95**, e28876. <https://doi.org/10.1002/jmv.28876> (2023).
51. McFadden, N. et al. Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. *PLoS Pathog.* **7**, e1002413. <https://doi.org/10.1371/journal.ppat.1002413> (2011).
52. Borg, C. et al. Murine norovirus virulence factor 1 (VF1) protein contributes to viral fitness during persistent infection. *J. Gen. Virol.* **102**. <https://doi.org/10.1099/jgv.0.001651> (2021).
53. Nichols, R. Gene trees and species trees are not the same. *Trends Ecol. Evol.* **16**, 358–364. [https://doi.org/10.1016/S0169-5347\(01\)02203-0](https://doi.org/10.1016/S0169-5347(01)02203-0) (2001).
54. Maddison, W. P. Gene Trees in Species Trees. *Syst. Biol.* **46**, 523–536. <https://doi.org/10.1093/sysbio/46.3.523> (1997).
55. Duffy, S., Shackleton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276. <https://doi.org/10.1038/nrg2323> (2008).
56. Pamilo, P. & Nei, M. Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583. <https://doi.org/10.1093/oxfordjournals.molbev.a040517> (1988).
57. Di Martino, B. & Marsilio, F. Feline calicivirus VP2 is involved in the self-assembly of the capsid protein into virus-like particles. *Res. Vet. Sci.* **89**, 279–281. <https://doi.org/10.1016/j.rvsc.2010.03.011> (2010).
58. Zhu, S. et al. Identification of immune and viral correlates of norovirus protective immunity through comparative study of intra-cluster norovirus strains. *PLoS Pathog.* **9**, e1003592. <https://doi.org/10.1371/journal.ppat.1003592> (2013).
59. Lin, Y., Fengling, L., Lianzhu, W., Yuxiu, Z. & Yanhua, J. Function of VP2 protein in the stability of the secondary structure of virus-like particles of genogroup II norovirus at different pH levels: function of VP2 protein in the stability of NoV VLPs. *J. Microbiol.* **52**, 970–975. <https://doi.org/10.1007/s12275-014-4323-6> (2014).
60. Chan, M. C. et al. Covariation of major and minor viral capsid proteins in norovirus genogroup II genotype 4 strains. *J. Virol.* **86**, 1227–1232. <https://doi.org/10.1128/jvi.00228-11> (2012).
61. Siebenga, J. J. et al. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001–2007. *J. Infect. Dis.* **200**, 802–812. <https://doi.org/10.1086/605127> (2009).
62. Motomura, K. et al. Divergent evolution of norovirus GII/4 by genome recombination from May 2006 to February 2009 in Japan. *J. Virol.* **84**, 8085–8097. <https://doi.org/10.1128/jvi.02125-09> (2010).
63. Lam, T. T. et al. The recombinant origin of emerging human norovirus GII.4/2008: intra-genotypic exchange of the capsid P2 domain. *J. Gen. Virol.* **93**, 817–822. <https://doi.org/10.1099/vir.0.039057-0> (2012).
64. Cannon, J. L. et al. Global trends in Norovirus genotype distribution among children with Acute Gastroenteritis. *Emerg. Infect. Dis.* **27**, 1438–1445. <https://doi.org/10.3201/eid2705.204756> (2021).
65. Estes, M. K. et al. Norwalk virus vaccines: challenges and progress. *J. Infect. Dis.* **181**(Suppl 2), 367–373. <https://doi.org/10.1086/315579> (2000).

Acknowledgements

We thank members of the Eyun lab for their valuable discussions. Profs. Tae Jung Park (Chung-Ang University, Korea) provided helpful comments and suggestions on an earlier draft of this manuscript. This work was supported by the grant (22192MFDS022) from Ministry of Food and Drug Safety in 2022, the National Research Foundation of Korea (2022R1A2C4002058), and Korea Institute of Marine Science & Technology Promotion (RS-2022-KS221676) funded by the Ministry of Oceans and Fisheries.

Author contributions

Huijeong Doh: Writing – original draft, Validation, Software, Methodology, Visualization, Formal analysis, Conceptualization. Chanhyeon Lee: Resources, Data curation, Formal analysis. Nam Yee Kim: Investigation, Resources. Yun-Yong Park: Data curation, Validation. Eun-jeong Kim: Data curation, Validation. Changsun Choi: Resources, Data curation. Seong-il Eyun: Writing – original draft, Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study was approved by the Chung-Ang University Bioethics Committee (Approval No: 1041078-202007-BR-179-01), which waived the requirement for informed consent.

DECLARATION OF COMPETING INTEREST

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-87719-9>.

Correspondence and requests for materials should be addressed to S.-i.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025