Gene Reports 32 (2023) 101786

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Shared clusters between phylogenetic trees for upstream and downstream regions of recombination hotspot in norovirus genomes

Yoshiyuki Suzuki [a,*], Hirokazu Kimura [b], Kazuhiko Katayama [c]

[a] Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya-shi, Aichi-ken 467-8501, Japan
[b] Graduate School of Health Science, Gunma Paz University, 1-7-1 Tonyamachi, Takasaki-shi, Gunma-ken 370-0006, Japan
[c] Omura Satoshi Memorial Institute, Graduate School of Infection Control Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo-to 108-8641, Japan

ARTICLE INFO

Edited by Dominic Voon

Keywords:
Genogroup
Norovirus
P-group
Phylogenetic tree
Recombination
Shared cluster

ABSTRACT

Noroviruses constitute the genus *Norovirus* in the family *Caliciviridae*. Since the recombination hotspot exists at the junction of ORF1 and ORF2, the upstream and downstream regions of recombination hotspot have been independently classified into P-groups and genogroups, respectively. Here, genomic sequences for 339 norovirus strains retrieved from the International Nucleotide Sequence Database were analyzed to identify the shared clusters, which were defined as the clusters consisting of the same sets of multiple strains between the phylogenetic trees constructed for the upstream and downstream regions. P-groups GII, GIV, GVI, and GNA2 as well as genogroups GII, GIV, GVI, GVIII, GIX, and GNA2 were included in a single shared cluster. In contrast, each of P-groups and genogroups GI, GIII and GV was divided into multiple shared clusters. Multiple shared clusters within single host species suggested the existence of incompatibility between the upstream and downstream regions of recombination hotspot in norovirus genomes.

1. Introduction

Noroviruses, which constitute the genus *Norovirus* in the family *Caliciviridae* (Vinje et al., 2019), infect gastrointestinal tract and salivary gland of mammals (Ghosh et al., 2022). The norovirus virion is a non-enveloped icosahedron with the size of 38 nm in diameter (Prasad et al., 1999), containing the genome of a linear, non-segmented, single-stranded RNA of positive polarity with the length of 7.5–8 kb (Lambden et al., 1993). ORF1, ORF2, and ORF3 are located in this order from the 5′-end to the 3′-end of the genome. ORF4 is occasionally located within ORF2 in the +1-shifted reading frame (Thackray et al., 2007).

ORF1 encodes a polyprotein precursor, which is cleaved into six non-structural proteins (NSs); NS1/2 (N-terminal protein, p48), NS3 (nucleoside triphosphatase, NTPase), NS4 (3A-like protein, p22), NS5 (viral protein genome-linked, VPg), NS6 (3CL protease, Pro), and NS7 (RNA-dependent RNA polymerase, RdRp). ORF2, ORF3, and ORF4 encode major capsid protein (viral protein 1, VP1), minor capsid protein (viral protein 2, VP2), and virulence factor 1 (VF1), respectively (McFadden et al., 2011).

In the norovirus genome, the recombination hotspot exists at the junction of ORF1 and ORF2 (Katayama et al., 2002; Bull et al., 2007). Thus, a dual classification system has been established for independently classifying the upstream and downstream regions of recombination hotspot (Bull et al., 2007). The nucleotide sequence for the 3′-terminal 762 positions of ORF1 encoding the C-terminal part of RdRp was used to classify the upstream region into P-groups, which were further divided

Abbreviations: A-A comparison, comparison of phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed both using the amino acid sequence; AIC, Akaike information criterion; A-N comparison, comparison of phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed using the amino acid sequence and the nucleotide sequence, respectively; BIC, Bayesian information criterion; F, equilibrium frequencies of amino acids estimated from data; Γ, gamma distribution; GTR, general time-reversible; I, invariable sites; INSD, International Nucleotide Sequence Database; JTT, Jones-Taylor-Thornton; LG, Le-Gascuel; ML, maximum likelihood; N-A comparison, comparison of phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed using the nucleotide sequence and the amino acid sequence, respectively; NCWG, Norovirus Classification Working Group; NJ, neighbor-joining; N-N comparison, comparison of phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed both using the nucleotide sequence; NS, non-structural protein; NTPase, nucleoside triphosphatase; Pro, protease; RdRp, RNA-dependent RNA polymerase; VF, virulence factor; VP, viral protein; VPg, viral protein genome-linked.

* Corresponding author at: Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya, Aichi 467-8501, Japan.
E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp (Y. Suzuki).

https://doi.org/10.1016/j.genrep.2023.101786
Received 24 March 2023; Received in revised form 27 April 2023; Accepted 7 May 2023
Available online 18 May 2023

into P-types, whereas the amino acid sequence for the entire part of VP1 was used to classify the downstream region into genogroups, which were further divided into genotypes (Kroneman et al., 2013).

P-groups and genogroups were defined as the phylogenetic clusters that were distinguished from other clusters according to the 2 × SD criterion (Kroneman et al., 2013). So far, 10 P-groups (GI-GVII, GX, GNA1, and GNA2) and 12 genogroups (GI-GX, GNA1, and GNA2) have been identified (Chhabra et al., 2019), and each NoV strain has been characterized by a P-group and a genogroup. The P-group and the genogroup assigned to a norovirus strain were described as genogroup [P-group]; e.g., GIV[GVI] indicating that the P-group was GVI and the genogroup was GIV. The description about the P-group was omitted when the P-group and the genogroup were the same; e.g., GI indicating that the P-group and the genogroup were both GI.

Recombination is considered to be an important mechanism for creating genetic diversities in noroviruses (Bull et al., 2007; Suzuki, 2021). Reportedly, however, the occurrence of recombination may be restricted by several factors; e.g., limitation in the inter-species transmission for mixed-infection of hosts or co-infection of cells (Ford-Siltz et al., 2019) and incompatibility between the upstream and downstream regions of recombination hotspot (Tohma et al., 2021). The purpose of the present study was to overview the recombination landscape of noroviruses in searching for traces of such restrictions from the phylogenetic analyses of genomic sequences.

2. Materials and methods

2.1. Sequence data

Genomic sequences for norovirus strains with the length of ≥ 6 kb containing the entire coding region of VP1 were retrieved from the International Nucleotide Sequence Database (INSD) on November 12, 2022. After eliminating the sequences that were ≥ 95% identical with other sequences or containing ≥ 1% ambiguous nucleotides (Suzuki et al., 2020), the remaining 339 sequences with the length ranging from 6079 bases to 8092 bases (Supplementary Table S1) were submitted to the Norovirus Genotyping Tool (version 2.0) (Kroneman et al., 2011) for determining P-groups and genogroups. P-groups and genogroups for these sequences were also obtained from the INSD entry files and the literature (Chhabra et al., 2019).

2.2. Data analysis

Genomic sequences for 339 norovirus strains obtained above were divided into the upstream and downstream regions of the junction of ORF1 and ORF2. Multiple alignment of 339 nucleotide sequences was made for the individual regions separately using the computer program MAFFT (version 7.247) (Katoh et al., 2002). After eliminating the nucleotide positions containing gaps, the upstream region consisted of 3338 positions encompassing the coding regions for the C-terminal part of NTPase and the entire parts of p22, VPg, Pro, and RdRp (corresponding to amino acid positions 501–1784 in NSs of the strain Norwalk: INSD accession number M87661). It should be noted that the most variable regions in ORF1 encoding p48 and p22 were not fully included because such regions often contained gaps. The downstream region consisted of 1181 positions encompassing the coding region for the entire part of VP1 (corresponding to amino acid positions 8–530 in VP1 of the strain Norwalk: INSD accession number M87661). After translating nucleotide sequences into amino acid sequences, multiple alignment of 339 amino acid sequences was also made for the individual regions separately using MAFFT (version 7.247) (Katoh et al., 2002).

Phylogenetic trees for the upstream and downstream regions of recombination hotspot were constructed using both the nucleotide and amino acid sequences. The best fit models of nucleotide and amino acid substitutions were obtained based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) using MEGA (version 10.1.7) (Kumar et al., 2018); the general time-reversible (GTR) model with rate-heterogeneity among sites following a gamma distribution (Γ) including invariable sites (I) (GTR + Γ + I) for the nucleotide sequences of the upstream and downstream regions, the Jones-Taylor-Thornton (JTT) model (Jones et al., 1992) with Γ and I (JTT + Γ + I) for the amino acid sequence of the upstream region, and the Le-Gascuel (LG) model (Le and Gascuel, 2008) with Γ and I assuming equilibrium frequencies of amino acids estimated from data (F) (LG + Γ + I + F) for the amino acid sequence of the downstream region. The maximum likelihood (ML) method (Felsenstein, 1981) with the best fit model was adopted using MEGA (version 10.1.7) (Kumar et al., 2018). In addition, the neighbor-joining (NJ) method (Saitou and Nei, 1987) with the p distance, which was known to produce reliable topologies when a large number of closely related sequences were analyzed (Nei and Kumar, 2000; Yoshida and Nei, 2016), was also adopted using MEGA (version 10.1.7) (Kumar et al., 2018). Reliabilities of interior branches were assessed by the bootstrap method with 1000 re-samplings.

Phylogenetic trees constructed for the upstream and downstream regions of recombination hotspot were compared to identify recombining groups, such that norovirus strains from the same group may recombine but those from different groups may not. Recombination between norovirus strains may cause topological differences between the phylogenetic trees for the upstream and downstream regions. However, each recombining group should be clustered in both phylogenetic trees, although topologies within the cluster may be different due to recombination. Therefore, the recombining groups may be identified as the shared clusters, which were defined as the clusters consisting of the same sets of multiple strains between the phylogenetic trees for the upstream and downstream regions. Phylogenetic trees were visualized using FigTree (version 1.4.4) (Rambaut, 2018).

3. Results

3.1. Phylogenetic trees

In the ML tree for the upstream region of recombination hotspot constructed using the nucleotide sequence, GIV, GIV[GVI], GVI, GVIII [GII], GIX[GII], and GNA2 strains were included in the cluster of GII strains (Supplementary Fig. S1). In contrast, GI, GIII, GV, GVII and GX strains made individual clusters. The single GNA1 strain was located outside the cluster of GI strains. These clustering patters were not always consistent with the P-group classification, which was also determined for the upstream region using the nucleotide sequence (Chhabra et al., 2019). It should be noted, however, that P-groups were determined through the phylogenetic analysis based on 762 nucleotide positions (Kroneman et al., 2013), whereas the phylogenetic analysis conducted in the present study was based on a greater number (3338) of nucleotide positions, which may produce more reliable clustering patterns. Similar topologies were obtained when the ML tree was constructed using the amino acid sequence, except that GVII strains were additionally included in the cluster of GII strains (Supplementary Fig. S2).

In the ML tree for the downstream region of recombination hotspot constructed using the nucleotide sequence, GIX[GII] and GNA2 strains were included in the cluster of GII strains (Supplementary Fig. S3). The single GVIII[GII] strain was located outside the cluster of these strains. In contrast, GI, GIII, GV, GVI, GVII, and GX strains made individual clusters. GIV and GIV[GVI] stains jointly made a cluster. The single GNA1 strain was located outside the cluster of GI strains. When the ML tree was constructed using the amino acid sequence, the cluster of GIX [GII] strains and the single GNA2 strain were not included in the cluster of GII strains (Supplementary Fig. S4). The joint cluster of GIV and GIV [GVI] stains as well as single strains of GVIII[GII] and GNA2 was included in the cluster of GVI strains. Although the clustering patterns of GIV, GIV[GVI], GVI, GVIII[GII], and GNA2 strains were slightly different from the genogroup classification (Chhabra et al., 2019), which was also determined for the entire part of VP1 using the amino acid sequence

(Kroneman et al., 2013), they were not supported with high bootstrap probabilities.

Similar results were obtained in the NJ trees constructed for the upstream and downstream regions of recombination hotspot using the nucleotide and amino acid sequences (Supplementary Figs. S5–S8). The only exception was that the single strains of GVIII[GII] and GNA2 were not included in the cluster of GVI strains in the NJ tree constructed for the downstream region using the amino acid sequence, which was consistent with the genogroup classification (Chhabra et al., 2019).

3.2. Shared clusters

There were 92 shared clusters identified in the comparison of ML trees for the upstream and downstream regions of recombination hotspot constructed both using the nucleotide sequence (N-N comparison) (Table 1; Supplementary Figs. S1 and S3). GI, GIII, GV, GVII, GIX[GII], and GX strains constituted individual shared clusters. GII strains constituted a shared cluster with GIV, GIV[GVI], GVI, GVIII[GII], GIX [GII], and GNA2 strains. The single GNA1 strain constituted a shared cluster with GI strains. GI, GIII, and GV strains were divided into two, two, and three shared clusters, respectively.

The number of shared clusters decreased to 55 in the comparison of ML trees for the upstream and downstream regions of recombination hotspot constructed both using the amino acid sequence (A-A comparison) (Table 1; Supplementary Figs. S2 and S4). Compared to the case of N-N comparison, GVII strains were included in the shared cluster of GII, GIV, GIV[GVI], GVI, GVIII[GII], GIX[GII], and GNA2 strains. In addition, GIII strains were not divided into two shared clusters but constituted a single shared cluster.

When the ML tree for the upstream region of recombination hotspot was constructed using the nucleotide sequence and that for the downstream region was constructed using the amino acid sequence (N-A comparison), 66 clusters were identified as the shared clusters (Table 1; Supplementary Figs. S9 and S10). Compared to the case of N-N comparison, GIII strains were not divided into two shared clusters but constituted a single shared cluster.

In the comparison of the ML tree for the upstream region of recombination hotspot constructed using the amino acid sequence and that for the downstream region constructed using the nucleotide sequence (A-N comparison), 68 clusters were identified as the shared clusters (Table 1; Supplementary Figs. S11 and S12). Compared to the case of N-N comparison, GVII strains were included in the shared cluster of GII, GIV, GIV [GVI], GVI, GVIII[GII], GIX[GII], and GNA2 strains.

When the NJ trees were used for the analysis, the numbers of shared clusters identified in the N-N, A-A, N-A, and A-N comparisons were 95 (Supplementary Figs. S5 and S7), 56 (Supplementary Figs. S6 and S8), 64 (Supplementary Figs. S13 and S14), and 70 (Supplementary Figs. S15 and S16), respectively, which were slightly greater than those identified using the ML trees in total (Table 1). The relationships of shared clusters with P-groups and genogroups observed using the NJ trees were the same as those observed using the ML trees described above.

Table 1
Relationships of shared clusters with P-groups and genogroups.

Comparison	ML[e]	NJ[f]	Relationship[g]
N-N[a]	92	95	{{{{GI/2[h]}, GNA1}, {GIII/2[i]}}, {{GII, GIV, GIV[GVI], GVI, GVIII[GII], {GIX[GII]}, GNA2}, {GVII}}, {GV/3[j]}, {GX}}
A-A[b]	55	56	{{{{GI/2[h]}, GNA1}, {GIII}}, {GII, GIV, GIV[GVI], GVI, {GVII}, GVIII[GII], {GIX[GII]}, GNA2}, {GV/3[j]}, {GX}}
N-A[c]	66	64	{{{{GI/2[h]}, GNA1}, {GIII}}, {{GII, GIV, GIV[GVI], GVI, GVIII[GII], {GIX[GII]}, GNA2}, {GVII}}, {GV/3[j]}, {GX}}
A-N[d]	68	70	{{{{GI/2[h]}, GNA1}, {GIII/2[i]}}, {GII, GIV, GIV[GVI], GVI, {GVII}, GVIII[GII], {GIX[GII]}, GNA2}, {GV/3[j]}, {GX}}

[a] Phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed both using the nucleotide sequence were compared.
[b] Phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed both using the amino acid sequence were compared.
[c] Phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed using the nucleotide sequence and the amino acid sequence, respectively, were compared.
[d] Phylogenetic trees for the upstream and downstream regions of recombination hotspot constructed using the amino acid sequence and the nucleotide sequence, respectively, were compared.
[e] Number of shared clusters identified between the ML trees.
[f] Number of shared clusters identified between the NJ trees.
[g] P-groups and genogroups of norovirus strains included in each shared cluster are listed in curly brackets.
[h] GI strains were divided into two shared clusters.
[i] GIII strains were divided into two shared clusters.
[j] GV strains were divided into three shared clusters.

4. Discussion

4.1. Identification of shared clusters

In the present study, recombining groups of norovirus strains were identified as the shared clusters between the phylogenetic trees for the upstream and downstream regions of recombination hotspot in norovirus genomes. Although the occurrence of recombination may be restricted by limitation in the inter-species transmission for mixed-infection of hosts or co-infection of cells (Ford-Siltz et al., 2019) and incompatibility between the upstream and downstream regions of recombination hotspot (Tohma et al., 2021), it may also be influenced by the stochastic nature of infection as well as recombination. Therefore, some of the shared clusters identified in the present study may be merged in the future through discovery of new recombinants.

4.2. Relationships of shared clusters with P-groups and genogroups

GI, GIII, and GV strains were divided into two, two, and three shared clusters, respectively. Consistently, it has been reported that GI strains were composed of two co-circulating populations evolving independently without recombination in humans (Tohma et al., 2021). In the phylogenetic trees, GV strains were largely separated into two lineages, one composed of a single shared cluster and the other composed of two shared clusters. The former lineage consisted of rat strains, whereas the latter lineage consisted of mouse strains, suggesting that there may be host species specificity between rat and mouse noroviruses (Li et al., 2022). It should be noted that the existence of host species specificity has also been reported for GIV, GIV[GVI], and GVI strains between humans, canines, and felines (Ford-Siltz et al., 2019). However, two shared clusters of mouse strains may have evolved independently without recombination. In addition, two shared clusters of GIII strains, which were mostly isolated from bovines, may also have been maintained without recombination. Multiple shared clusters within single host species suggested the existence of incompatibility between the upstream and downstream regions of recombination hotspot in norovirus genomes (Tohma et al., 2021).

4.3. Possibility of using shared clusters for classification of noroviruses

The dual classification system has been widely used for characterizing the upstream and downstream regions of recombination hotspot in norovirus genomes (Chhabra et al., 2019). However, as the P-group and genogroup classifications are not necessarily consistent, the dual classification system may not be suitable for building the norovirus taxonomy. According to the biological species concept, members of a species were defined by their ability to exchange genetic information (Bobay and Ochman, 2018). Since the exchange of genetic information between

norovirus strains occurs through recombination, shared clusters may be used for developing a singular classification system for noroviruses; e.g., norovirus strains may be classified into a maximum number of shared clusters, which may be called the shared cluster groups. However, further discussion on this issue may be subject to the Norovirus Classification Working Group (NCWG) (Kroneman et al., 2013; Chhabra et al., 2019).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genrep.2023.101786.

CRediT authorship contribution statement

Yoshiyuki Suzuki: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft. **Hirokazu Kimura:** Conceptualization, Validation, Writing – review & editing. **Kazuhiko Katayama:** Conceptualization, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank anonymous reviewers for their valuable comments. This work was supported by JSPS KAKENHI grant number JP19K12221 and AMED grant number JP21fk0108120 to Y.S.

References

Bobay, L.-M., Ochman, H., 2018. Biological species in the viral world. Proc. Natl. Acad. Sci. U. S. A. 115, 6040–6045.
Bull, R.A., Tanaka, M.M., White, P.A., 2007. Norovirus recombination. J. Gen. Virol. 88, 3347–3359.
Chhabra, P., de Graaf, M., Parra, G.I., Chan, M.C.-W., Green, K., Martella, V., Wang, Q., White, P.A., Katayama, K., Vennema, H., Koopmans, M.P.G., Vinje, J., 2019. Updated classification of norovirus genogroups and genotypes. J. Gen. Virol. 100, 1393–1406.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Ford-Siltz, L.A., Mullis, L., Sanad, Y.M., Tohma, K., Lepore, C.J., Azevedo, M., Parra, G.I., 2019. Genomics analyses of GIV and GVI noroviruses reveal the distinct clustering of human and animal viruses. Viruses 11, 204.
Ghosh, S., Kumar, M., Santiana, M., Mishra, A., Zhang, M., Labayo, H., Chibly, A.M., Nakamura, H., Tanaka, T., Henderson, W., Lewis, E., Voss, O., Su, Y., Belkaid, Y., Chiorini, J.A., Hoffman, M.P., Altan-Bonnet, N., 2022. Enteric viruses replicate in salivary glands and infect through saliva. Nature 607, 345–350.
Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.
Katayama, K., Shirato-Horikoshi, H., Kojima, S., Kageyama, T., Oka, T., Hoshino, F.B., Fukushi, S., Shinohara, M., Uchida, K., Suzuki, Y., Gojobori, T., Takeda, N., 2002. Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. Virology 299, 225–239.
Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kroneman, A., Vennema, H., Deforche, K., Avoort, H.V., Penaranda, S., Oberste, M.S., Koopmans, V.J., 2011. An automated genotyping tool for enteroviruses and noroviruses. J. Clin. Virol. 51, 121–125.
Kroneman, A., Vega, E., Vennema, H., Vinje, J., White, P.A., Hansman, G., Green, K., Martella, V., Katayama, K., Koopmans, M., 2013. Proposal for a unified norovirus nomenclature and genotyping. Arch. Virol. 158, 2059–2068.
Kumar, S., Stecher, G., Li, M., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Lambden, P.R., Caul, E.O., Ashley, C.R., Clarke, I.N., 1993. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. Science 259, 516–519.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Li, M., Li, K., Lan, H., Hao, X., Liu, Y., Zhou, C., 2022. Investigation of genotype diversity of 7,804 norovirus sequences in humans and animals of China. Open Life Sci. 17, 1429–1435.
McFadden, N., Bailey, D., Carrara, G., Benson, A., Chaudhry, Y., Shortland, A., Heeney, J., Yarovinsky, F., Simmonds, P., Macdonald, A., Goodfellow, I., 2011. Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. PLoS Pathog. 7, e1002413.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.
Prasad, B.V.V., Hardy, M.E., Dokland, T., Bella, J., Rossmann, M.G., Estes, M.K., 1999. X-ray crystallographic structure of the Norwalk virus capsid. Science 286, 287–290.
Rambaut, A., 2018. FigTree. http://tree.bio.ed.ac.uk/software/figtree.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Suzuki, Y., 2021. Effect of recombinations on changes in genotype proportions between norovirus seasons in Japan. Meta Gene 29, 100934.
Suzuki, Y., Nishimura, M., Inoue, T., Kobayashi, Y., 2020. Methods for reducing the number of sequences in molecular evolutionary analyses. Meta Gene 23, 100629.
Thackray, L.B., Wobus, C.E., Chachu, K.A., Liu, B., Alegre, E.R., Henderson, K.S., Kelley, S.T., Virgin, H.W.I.V., 2007. Murine noroviruses comprising a single genogroup exhibit biological diversity despite limited sequence divergence. J. Virol. 81, 10460–10473.
Tohma, K., Lepore, C.J., Martinez, M., Degiuseppe, J.I., Khamrin, P., Saito, M., Mayta, H., Nwaba, A.U.A., Ford-Siltz, L.A., Green, K.Y., Galeano, M.E., Zimic, M., Stupka, J.A., Gilman, R.H., Maneekarn, N., Ushijima, H., Parra, G.I., 2021. Genome-wide analyses of human noroviruses provide insights on evolutionary dynamics and evidence of coexisting viral populations evolving under recombination constraints. PLoS Pathog. 17, e1009744.
Vinje, J., Estes, M.K., Esteves, P., Green, K.Y., Katayama, K., Knowles, N.J., L'Homme, Y., Martella, V., Vennema, H., White, P.A., ICTV Report Consortium, 2019. ICTV virus taxonomy profile: *Caliciviridae*. J. Gen. Virol. 100, 1469–1470.
Yoshida, R., Nei, M., 2016. Efficiencies of the NJp, maximum likelihood, and Bayesian methods of phylogenetic construction for compositional and noncompositional genes. Mol. Biol. Evol. 33, 1618–1624.