

Genomic Characteristics and Evolutionary Analysis of a Rare GI.1 Norovirus Isolate From Beijing, China

Weishan Zhang

Chinese Academy of Sciences

Wei Zhang

Beijing Haidian Hospital

Xinling Hu

Chinese Academy of Sciences

Xintong Zhou

Chinese Academy of Sciences

Guobao Tian

Chinese Academy of Sciences

Yinghua Xu

National Institute for the Control of Pharmaceutical and Biological Products; China National Institute for Food and Drug Control

Kaixia Mi (✉ mik@im.ac.cn)

Chinese Academy of Sciences <https://orcid.org/0000-0002-4399-2930>

Research Article

Keywords: Norovirus, Whole genome sequencing, GI.1, Phylogenetic tree, Evolutionary analysis

Posted Date: September 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-888489/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Genomic characteristics and evolutionary analysis of a rare GI.1 norovirus isolate from Beijing, China

WeiShan Zhang^{1, 2†}, Wei Zhang^{3†}, XinLing Hu¹, Xintong Zhou¹, GuoBao Tian³, Yinghua Xu^{4*},
Kaixia Mi^{1, 2*}

¹ CAS Key Laboratory of Pathogenic and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

² Savaid Medical School, University of Chinese Academy of Sciences, Beijing 101408, China

³ Beijing Haidian Hospital, Beijing 100018, China

⁴ National Institute for Food and Drug Control, Beijing 102629, China

† These authors contributed equally to this work.

Corresponding author:

Yinghua Xu, E-mail: xuyh@nifdc.org.cn

Kaixia Mi, E-mail: mik@im.ac.cn

Abstract

Noroviruses are one of the main pathogens of acute gastroenteritis, causing frequent outbreaks worldwide every year that seriously affect human health. The GII.4 genotype causes most norovirus (NoV) infections and large-scale outbreaks. By contrast, the GI genotype is relatively rare. In this study, the whole genome sequence of a newly isolated ZD strain from a patient in Beijing, China, was sequenced and analyzed. The ZD strain genome consisted of 7,597 nucleotides and contained three open reading frames. Whole-genomic analysis indicated the strain was a GI.1 genotype, and no recombination site was detected in the genome. The histo-blood group antigen (HBGA) binding site associated with invasion of the GI genotype did not change, implying relatively conservative evolution. Phylogenetic analysis indicated the VP1 sequence of GI.1 strains could be divided into three clusters according to time of appearance: older (1968-2011), earlier (2011-2015), and new (2017-2018). Each cluster showed distinctive amino acid substitution characteristics, and the number of substitutions increased with time. The isolated ZD strain was in the new cluster. This study is the first to conduct a phylogenetic analysis of a GI genotype NoV isolated from Beijing. The results improve understanding of NoV diversity in China and can be a reference for further study of nondominant epidemic strains of NoVs as well as epidemic prevention and control.

Keywords

Norovirus; Whole genome sequencing; GI.1; Phylogenetic tree; Evolutionary analysis

Introduction

Noroviruses constitute a genus in the family *Caliciviridae*. Virus particles are round and simple in structure without an envelope and have only a layer of capsid protein with a smooth surface [1]. Formerly known as Norwalk virus, noroviruses were

first collected from fecal samples during a gastroenteritis outbreak in Norwalk, Ohio, in 1929 [2]. In patients infected with norovirus, usual symptoms include vomiting, nausea, abdominal pain, fever, and self-limiting or nonhemorrhagic diarrhea. The incubation period is 24 to 48 h, and the symptoms usually last for 12 to 60 h [3].

The NoV genome is positive-sense single-stranded RNA that is 7.5 to 7.7 kb in length. The 5' end is covalently linked to the virus-encoded protein genome (VPg), and the 3' end is polyadenylated. Except for those of murine NoVs, genomes of most NoVs encode three open reading frames (ORFs) [4]. The first (ORF1) encodes a polyprotein that can be hydrolyzed into six types of nonstructural proteins (NS1-NS7) by protease, including NoV protease and RNA-based RNA polymerase (RdRp) [5]. The second (ORF2) encodes the main structural protein (VP1) of the virus, including shell (S) and protruding (P) domains. The P domain includes P1 and P2 subdomains. The S domain surrounds the virus RNA; whereas the P domain connects to the S domain through flexure hinges. The P2 subdomain is highly variable and represents the maximum exposure region of the protein [6]. It includes the main neutralizing epitope, the binding site of histo-blood group antigen (HBGA), and the location of the main antigenic determinant of the capsid [7]. The third (ORF3) encodes the secondary structural protein (VP2) of the virus, which is within virus particles and may be associated with capsid assembly and genome encapsulation [8]. A complete, mature NoV particle is a dimer composed of 90 major structural proteins (VP1) and one or two minor structural proteins (VP2) symmetrically assembled in an icosahedron.

Because suitable cell culture systems and animal models are lacking, NoVs cannot be cultured *in vitro* and serotyping cannot be used. Therefore, classification of NoVs is mainly based on the amino acid sequences of VP1 and RdRp. Currently, there are six recognized genogroups (GI~GVI) and 40 genotypes. In 2019, Preeti et al. expanded the number of NoV genomes to 10 (GI~GX) on the basis of the diversity of amino acid sequences of VP1, with two of the genomes tentative [9]. They suggested that complete genome sequences are needed to classify NoVs.

Human NoVs (HuNoVs) are one of the main pathogens of acute gastroenteritis (AGE), and they account for a very high proportion of diarrhea outbreaks [10]. Approximately 50% of AGE outbreaks worldwide are associated with NoVs. According to the World Health Organization (WHO), approximately 699 million people suffer annually worldwide from NoV infection, resulting in more than 200,000 deaths [11]. In addition, the average direct cost of NoVs to the health system is approximately 4 billion dollars per year, whereas the socioeconomic burden approaches 60 billion dollars [12]. Noroviruses are one of the leading causes of diarrhea in China, and there were 556 NoV outbreaks from October 2016 to September 2018 [13]. Notably, the primary NoVs infecting patients in those outbreaks were GII, GI, and GIV genotypes. The GII genotypes occur more frequently than GI genotypes; whereas GI genotypes are more frequent than GIV genotypes. Among the genotypes, GII.4 was predominant in all outbreaks, which may be because of its rapid evolution and antigenic diversity achieved through recombination [14]. The frequency of GI incidence is associated with geographic area. In NoV outbreaks in the United States from 2013 to 2016, outbreaks caused by GI.1 were rare, approximately 0.3% [15]. Moreover, sequence information

of GI.1 is relatively scattered.

In this study, a GI.1 NoV strain was isolated from clinical samples of Beijing Haidian Hospital in 2018 and named ZD. Whole sequences of the ZD strain were sequenced and genomic characteristics were analyzed. The complete ZD sequence was compared with sequences of prevalent NoV strains worldwide to obtain the phylogenetic pedigree and its evolutionary characteristics. The analysis showed that recombination did not occur in the ZD strain and that the gene encoding HBGA, a key protein for GI infection, was highly conserved. Phylogenetic analysis of all available GI sequences showed that the prevalent GI strains in different periods were different. In addition, the prevalent GI strains in the same period had similar amino acid mutation characteristics. The number of amino acid mutation sites gradually accumulated and increased over time. The study provides new information on the GI genome as well as insight into understanding why particular genotypes of NoV strains are prevalent.

Materials and methods

Extraction of norovirus RNA

Stool samples containing NoV were obtained in 2018 from Beijing Haidian Hospital, Beijing, China. MagaBio plus virus RNA Purification Kit (BIOER, HangZhou Bioer Technology Co. Ltd., China) was extracted using a kit following the protocol provided by the company. In brief, methods were the following. Stool samples were weighed, approximately 200 mg, and 2 ml of sterilized Phosphate Buffered Saline (PBS) was added. The mixture was centrifuged at 300 ×g for 5 min. The supernatant was removed and centrifuged at 16,260 ×g for 5 min. The supernatant was discarded, and the pellet was resuspended in 1 ml of PBS and centrifuged at 16,260 ×g for 5 min. That pellet was resuspended in 200 μl of PBS. An automatic nucleic acid extraction and purification instrument, Model NPA-32+ (BIOER, HangZhou Bioer Technology Co. Ltd., China), was used to extract virus RNA. Purified RNA was stored in a nuclease-free centrifuge tube at −80 °C.

Construction of norovirus RNA Library

KAPA Stranded RNA-seq Kit (Kapa Biosystems, FL, USA) was used to purify mRNA from the isolated RNA to prepare the RNA library. Samples of RNA, approximately 20 μl, were absorbed at 94 °C for 4 min to deplete ribosomal RNA. The fragmentation samples were prepared for first chain synthesis. PCR reaction conditions were as follow: 4 min at 25 °C, 15 min at 42 °C, and 15 min at 70 °C, followed by 60 min at 16 °C. Agencourt ®AMPure ® XPreagent Magnetic beads (NEW ENGLAND BioLabs, USA) were used to purify samples after synthesis of the second chain. A-tailing was added, and the reaction conditions were as follow: 30 min at 30 °C and 30 min at 60 °C. Then, adapters were ligated by KAPA T4 DNA ligase, and the PCR tube was incubated in a metal bath for 15 min at 20 °C. Then, after two purifications by

magnetic beads, library fragments were amplified by PCR with KAPA HiFi HotStart ReadyMix (2×), and the reaction conditions were as follow: predenaturation for 45 s at 98 °C; 20 cycles of 15 s at 98 °C, 30 s at 60 °C, and 30 s at 72 °C; and then 5 min at 72 °C. Magnetic beads were used to purify the amplified fragments. The purified library was used to detect the concentration of Qubit, and then, the mixed samples of the library were sequenced on a computer. Samples were sequenced by double-terminal 150 bp (PE150) with Illumina® (<https://www.illumina.com/>) NoVaseq. There was 4.1 Gb of clean ZD data.

Bioinformatics analysis

ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to search for ZD ORFs and then compare those ORFs with corresponding ORFs in the reference genome (NC_001959.2). The Pfam sequence analysis tool (<http://pfam.xfam.org/>) was used to correct ORF positions [16]. The BLAST tool (<http://www.ncbi.nlm.nih.gov/BLAST/>) was used to search for similarity of nucleotide sequences, and BioAider (v1.3) was used to compare selected sequences in different genomes.

BioAider [17] was used to compare the ZD strain with the 2018 Shanghai strain (MT008453.1), which had the highest homology with ZD. The Norovirus Typing Tool (<https://www.rivm.nl/mpf/typingtool/norovirus/>) [18] was used to determine the genotypes. ClustalW in BioEdit (v7.0.9) was used for multiple alignment of selected sequences. After alignment, the most suitable algorithm model to analyze the entire virus genome and different ORF sequences was obtained by using Find Best DNA/Protein Models in Molecular Evolutionary Genetic Analysis software (v10.0) (MEGA X). Whole genome sequences were analyzed using the GTR+G model, and ORF sequences were analyzed using the GTR+G+I model [19]. Maximum likelihood (ML) was used to construct phylogenetic trees, and trees were visualized using PhyML 3.0 with the previous optimal model parameters. A bootstrap test (100 replicates) was used. After multiple DNA sequences alignment with BioEdit, amino acid translation results were obtained by MEGA X, and changes in amino acid sites of the main structural protein VP1 of related strains were compared simultaneously.

Reference norovirus strains

GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) accession numbers (country/genotype) of the NoV strains used in this study are as follow:

NC_001959.2(USA/GI.1), MT372465.1(USA/GI.2), MZ223425.1(USA/GI.2), NC_044854.1(JP/GI.6), MT372477.1(USA/GI.6), MT357996.1(USA/GI.6), MZ229330.1(USA/GI.4), MT372474.1(USA/GI.9), MZ215728.1(USA/GI.9), MT372469.1(USA/GI.7), MZ215727.2(USA/GI.7), MT372476.1(USA/GI.8), KP407450.1(CHN/GI.8), KJ196298.1(USA/GI.8), MT031988.1(USA/GI.3), MT372470.1(USA/GI.3), MN448477.1(USA/GI.3), NC_039897.1(JP/GI.3), KJ196292.1(JP/GI.3), MZ203488.1(USA/GI.3).

In addition, GI.1 NoV strains with complete VP1 sequence in this study are listed in Table 1.

Table 1 Description of GI.1 NoV strains with complete VP1 sequences.

Accession No.	Collection Region	Year	Accession No.	Collection Region	Year
JX023285.1	USA	1968	MH638228.1	USA	2017
KF429765.1	USA	1968	MH638229.1	USA	2017
KF429761.1	USA	1972	MT357991.1	USA	2018
KF429770.1	USA	1973	MT357993.1	USA	2018
KF429783.1	USA	1979	MT357992.1	USA	2018
AY502016.1	USA	2001	MK956175.1	France	2018
KF039735.1	USA	2008	MK956173.1	France	2018
KF039728.1	USA	2008	MK956174.1	France	2018
KT943503.1	USA	2008	FJ384783.1	Sweden	2007
KF039734.1	USA	2009	EU085522.1	Sweden	2015
KF039726.1	USA	2009	EU085524.1	Sweden	2015
KF039729.1	USA	2010	EU085528.1	Sweden	2015
KF039725.1	USA	2010	EF547392.1	Japanese	2003
KF039733.1	USA	2011	MT008453.1	China*	2018
MZ220330.1	USA	2015	KT732279.1	China	2015

* isolated from Shanghai.

Results

Complete genome of the ZD strain

The whole nucleotide sequence of the ZD strain genome was 7,597 bp and included three ORFs spanning from 537 to 5,363 (ORF1), 5,347 to 6,939 (ORF2), and 5363 to 7,577 (ORF3) (Fig. 1). A 16-bp nucleotide overlap occurred between ORF1 and ORF2. The nucleotide sequence of the ZD strain genome was used in BLAST search and comparison, and the query coverage rate was 100%. The highest homology (99.14%) was with the Shanghai strain isolated in 2018 (MT008453.1). Compared with the MT008453.1 nucleotide sequence, the ZD genome had 64 single nucleotide polymorphisms (SNPs) (Table S1). Among the SNPs, 57 were in coding regions, with 36 in ORF1, 16 in ORF2, and 5 in ORF3 (Table 2). Mutated amino acids corresponding to the SNPs are also listed in Table 2. There were seven nonsynonymous mutations, with four in ORF1, one in ORF2, and two in ORF3. In addition, G:C A:T transition > G:C A:T transversion, and the transition–transversion ratio estimate was 20.3 (Table 3).

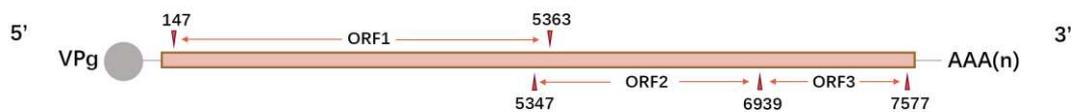


Fig. 1 Composition of the genome of the ZD strain of norovirus

Table 2 Comparison of ORFs between ZD strain and MT008453.1 noroviruses

ZD	Position Length (nt/aa)	Pairwise Identity (nt)%	Mutation nt	Pairwise Identity (aa)%	Mutation aa
ORF1	5216/1739	99.21	36	99.77	$n = 4$; A31V, I191V, P737S, S1653V
ORF2	1592/531	99.18	16	99.81	$n = 1$; P349L
ORF3	638/213	98.44	5	99.06	$n = 2$; A41V, R50K

Table 3 Mutation types of DNA between ZD strain and MT008453.1 noroviruses

Mutation type	Amount	
Transition	T-C	37
	A-G	24
	T-A	1
Transversion	T-G	2
	C-G	0
	Ratio ^a	20.3

^aRatio = Transition–Transversion bias ratio.

Phylogenetic analysis of the ZD strain

The Online Norovirus Typing Tool v2.0 predicted the ZD strain had the GI.1 genotype. To understand the genetic relations between the ZD strain and other NoV genotypes, a phylogenetic tree based on whole genomes was constructed (Fig. 2). Consistent with the prediction, the ZD strain was classified as genotype GI.1. As shown in Fig. 2, the ZD strain was most closely related to the Shanghai strain isolated in 2018, and the identity between the two strains was 99.14%, as obtained by BLAST.

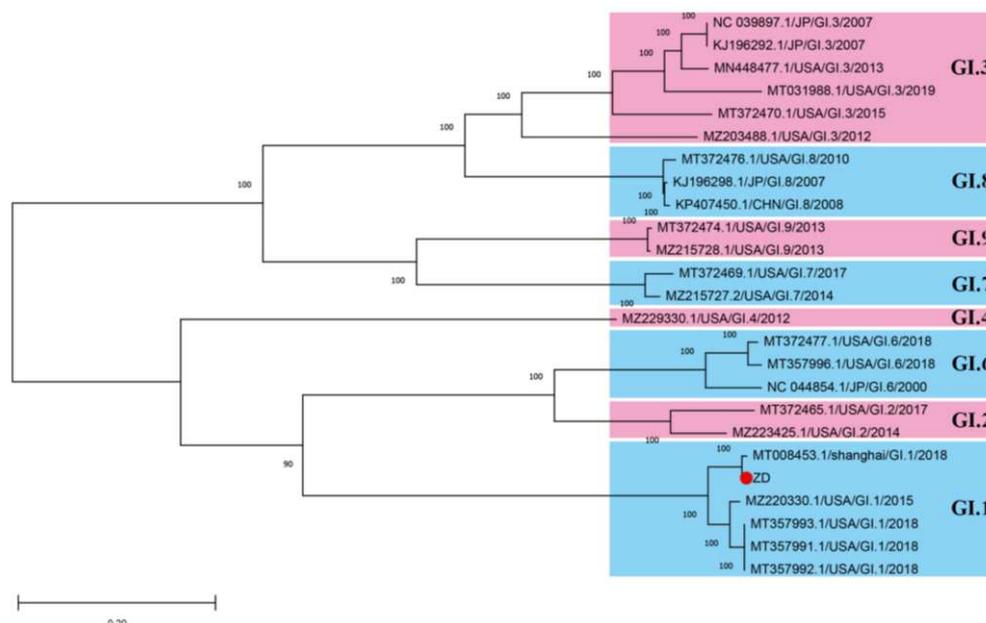


Fig. 2 Phylogenetic tree based on whole genomes of noroviruses. Phylogenetic analysis was performed using the maximum likelihood method (GTR+G model). Results were validated with a

bootstrap test (100 replicates). The red circle identifies the ZD strain, and background colors separate NoV genotypes

Previous studies show that gene recombination occurs frequently near the junction of ORF1 and ORF2 in human NoV, which is related to antigenic shift [20]. Recombination at the junction of ORF2 and ORF3[14] is less common. Because antigen evolution is related to recombination [21], phylogenetic analyses also were performed to compare recombination in each ORF of the ZD strain with that in selected NoV genomes (Fig.3). The results showed that the ZD strain was nonrecombinant on the basis of whole genome or single ORF analysis.

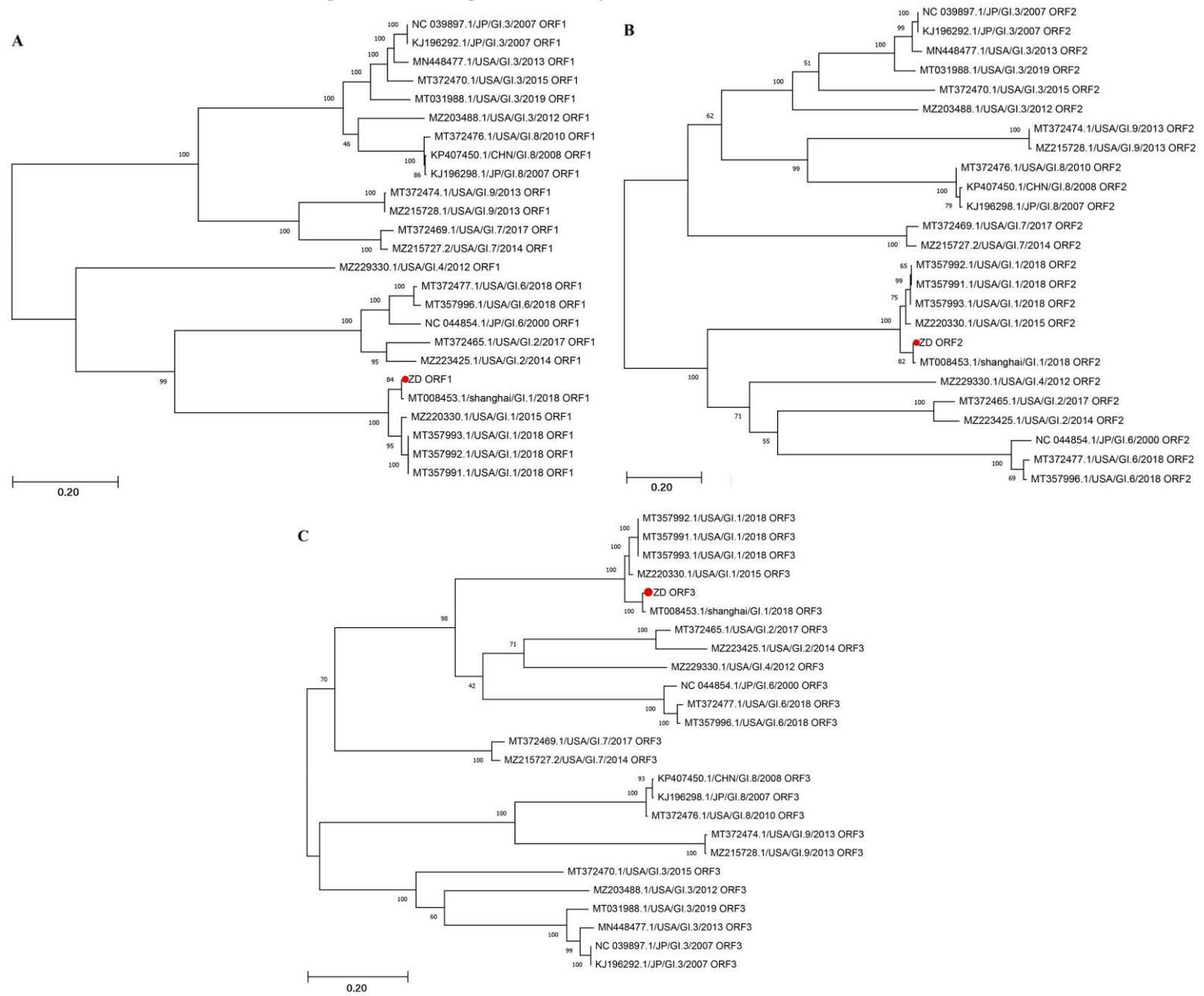


Fig. 3 Phylogenetic trees based on the nucleotide sequences of (A) ORF1, (B) ORF2, and (C) ORF3 in noroviruses. Phylogenetic analysis was performed using the maximum likelihood method

(GTR+G+I model). Results were validated with a bootstrap test (100 replicates). The red circle in each tree indicates the ZD strain. ORF, open reading frame

Comparison of the capsid protein VP1 sequence of the ZD strain with that of other GI.1 strains

Changes in the encoding sequences of the viral capsid protein VP1 can reveal potential variations in antigenicity and receptors for binding receptors [21]. In addition, VP1 is also an important part in norovirus vaccines [22]. Therefore, full-length VP1 proteins of all GI.1 genotype strains in GenBank were obtained. Notably, phylogenetic results of alignments of capsid protein VP1 sequences showed that all GI.1 strains, including the ZD strain, were divided into three clusters according to time of appearance. The GI.1 strains collected from 1968 to 2011 were clustered in an “older” cluster; those collected from 2011 to 2015 were clustered in an “earlier” cluster; and those collected from 2017 to 2018 were clustered in a “new” cluster (Fig. 4). The ZD strain collected in 2018 was in the “new” cluster.

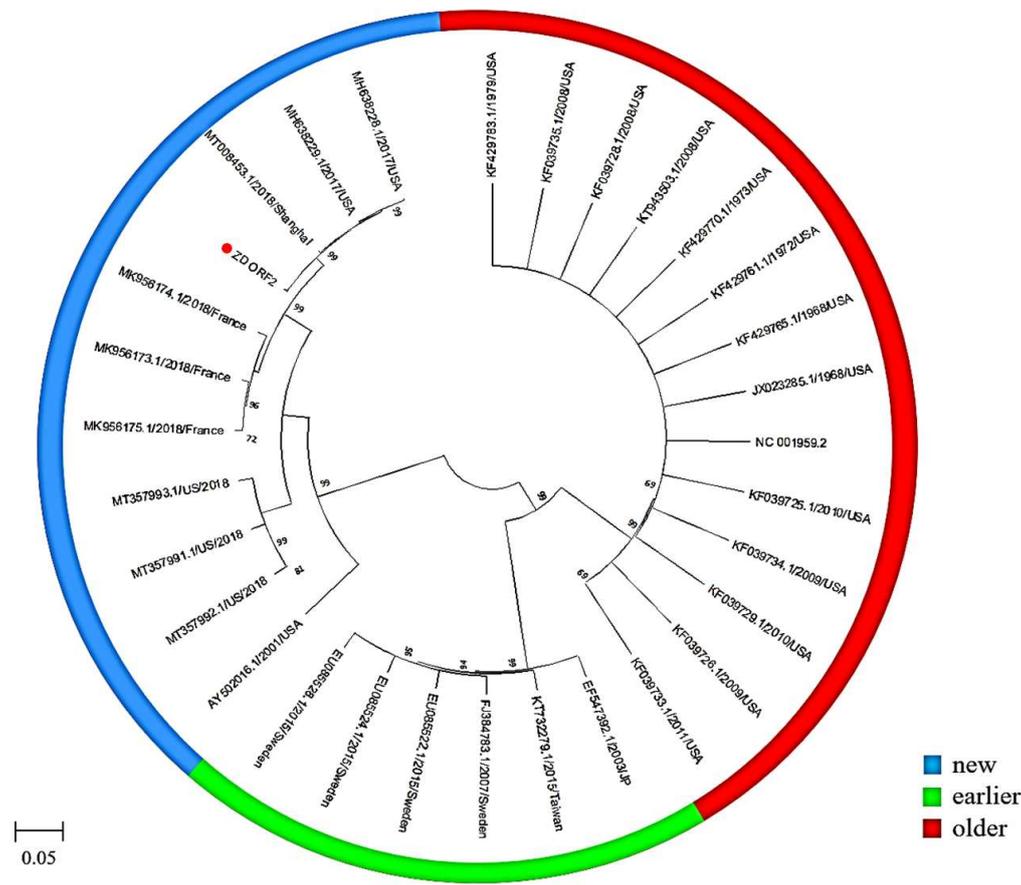


Fig. 4 Phylogenetic analysis of GI.1 noroviruses based on the nucleotide sequences of VP1. Phylogenetic analysis was performed using the maximum likelihood method (GTR+G+I model). Results were validated with a bootstrap test (100 replicates). Bootstrap values greater than 69% are shown on the branches. The red circle indicates the ZD strain. Older cluster (1968-2011) is indicated by red; earlier cluster (2011-2015) is indicated by green; and new cluster (2017-2018) is indicated

by blue

On the basis of multiple sequence alignments of the GI.1 capsid protein VP1, 30 amino acid sites of the 531 amino acids encoding VP1 changed, compared with the G1 reference strain (NC_001959.2). “Older,” “earlier,” and “new” clusters were identified, consistent with those of the phylogenetic tree (Table 4). The results verified that all collected GI.1 NoVs could be clustered according to the time of collection. In the “older” cluster, there were no significant changes in amino acids compared with the reference strain. In the “earlier” cluster, the GI.1 strain began to change, but there were fewer change sites, mainly concentrated at 252 (S252G). However, in the “new” cluster, multiple amino acid positions of the GI.1 strain changed regularly, namely S11N, V25A, A27T, A142S, S251G, and H286Q. There were also several special strains that were collected in earlier years but appeared in more recent clusters, including a Swedish strain collected in 2007 (FJ384783.1) and a Japanese strain collected in 2003 (EF547392.1) in the “earlier” group and an American strain collected in 2001 (EF547392.1) in the “new” group. Prevalent NoVs tend to be replaced periodically, and thus, it was not surprising that older strains became prevalent again.

Table 4 Changes in amino acids sites of GI.1 noroviruses

Note: Changes in amino acid sites of the capsid protein (VP1) of the ZD strain based on alignment with other GI.1 strains, as performed in MEGA X.

Clusters	Strains	Amino acid position																																
		11	25	27	43	46	142	194	200	251	252	286	288	291	330	343	346	349	350	364	365	370	376	381	399	410	425	432	445	488	521			
older	NC_001959.2	S	V	A	V	A	A	G	F	S	S	H	A	I	I	Y	D	P	D	I	G	V	I	H	I	Y	M	N	L	A	S			
	JX023285.1/1968/USA		
	KF429765.1/1968/USA		
	KF429761.1/1972/USA		
	KF429770.1/1973/USA		
	KF429783.1/1979/USA		
	KF039735.1/2008/USA	T		
	KF039728.1/2008/USA		
	KT943503.1/2008/USA		
	KF039734.1/2009/USA	R		
	KF039726.1/2009/USA	N		
	KF039729.1/2010/USA	V	
	KF039725.1/2010/USA	
	KF039733.1/2011/USA	R	
earlier	▲ FJ384783.1/2007/Sweden	G	N	I	V	.	.	F	I	S	.	.	T			
	▲ EF547392.1/2003/JP	G	V	S	I	V	.	.	.	I	S	.	.	T		
	EU085522.1/2015/Sweden	S	.	.	G	N	I	V	.	.	F	I	S	.	.	T		
	EU085524.1/2015/Sweden	G	N	I	V	.	.	F	I	S	.	.	T		
	EU085528.1/2015/Sweden	G	N	I	V	.	.	F	I	S	.	.	T		
KT732279.1/2015/Taiwan	L	.	G	N	I	V	.	.	F	I	S	.	.	T			
new	▲ AY502016.1/2001/USA	N	A	T	.	.	S	Q	.	.	V	F	.	.	E	V	.	I	.	.	.	F	I	.	F	.	.			
	MH638228.1/2017/USA	N	A	T	.	.	S	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MH638229.1/2017/USA	N	A	T	.	.	S	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MK956173.1/2018/France	N	.	T	G	.	S	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MK956174.1/2018/France	N	.	T	.	S	.	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MK956175.1/2018/France	N	A	T	G	.	S	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MT008453.1/2018/Shanghai	N	A	T	.	S	.	.	.	G	.	.	Q	.	.	V	F	.	.	E	.	I	.	.	.	V	F	.	F	.	.			
	MT357991.1/2018/USA	N	A	T	.	S	.	.	.	G	.	Q	G	V	.	F	.	.	E	V	.	I	.	.	.	V	F	I	.	F	S	T		
	MT357992.1/2018/USA	N	A	T	.	S	.	.	.	G	.	Q	G	.	.	F	.	.	E	V	.	I	.	.	.	V	F	I	.	F	S	T		
	MT357993.1/2018/USA	N	A	T	.	S	.	.	.	G	.	Q	G	.	.	F	.	.	E	V	.	I	.	.	.	V	F	I	.	F	S	T		
	ZD	N	A	T	.	S	.	.	.	G	.	Q	.	.	.	V	F	.	L	E	.	.	I	.	.	.	V	F	.	F	.	.		

Notably, the P2 region of VP1 is exposed on the surface of the capsid protein and is the main epitope for immune recognition and contains the binding domain of HBGA. As a highly variable region in NoVs, the occurrence and accumulation of mutations in the P2 region is often the main driving force for the evolution of epidemic strains of NoV (such as GII.4) [23]. Such mutations are also important in developing new strains with new immune binding characteristics and antigenicity [24]. In this study, although

the GI.1 strains had multiple mutation sites, the amino acid residue sites related to HBGA binding did not change.

Discussion

Noroviruses are one of the major pathogens of AGE and diarrhea and are an important cause of morbidity and mortality in children under 5 years of age [25]. They are highly contagious and environmentally stable and are usually viable for several weeks after infection. Probability of symptomatic disease from a single NoV particle may be as high as 0.5. Norovirus infection causes severe vomiting and diarrhea. There are 106 to 109 stable, nonenveloped virus particles per milliliter of excreta, creating unlimited opportunities for further infection transmission and outbreaks [26]. Noroviruses are primarily transmitted via fecal–oral and vomit–oral routes and either person-to-person contact or contaminated food or water, as well as environmental contamination [27].

Because NoVs cannot be cultured *in vitro*, it is important to use whole genome sequencing for classification and analysis of epidemic trends. In this study, sequencing techniques were used to obtain the whole genome of the ZD strain of norovirus, isolated from clinical samples of Beijing Haidian Hospital. According to the NoV online typing tool (v2.0) and MEGA X, the ZD strain had a GI.1 genotype. An online search found that whole genome and partial sequences data of GI.1 were scattered, with links to sporadic cases worldwide. There were only about 20 GI.1 whole genome sequences in National Center for Biotechnology Information (NCBI), indicating that epidemics caused by GI.1 are extremely rare [15].

In the phylogenetic analysis, the ZD strain was clustered with the strain collected in Shanghai in the same year (data from NCBI online) (Fig. 2 and 3). The high similarity suggested the two strains derived from the same origin. Recent studies indicate that NoV outbreaks are often closely associated with consumption of low-processed shellfish [28,29]. Moreover, several studies show that genogroup GI strains are more often associated with waterborne transmission than GII strains [30]. In general, GI strains have high environmental stability in seawater and bivalves and can usually be isolated from contaminated oysters [31]. Cheng et al. assessed NoV contamination of oysters imported from different countries for consumption in Hong Kong using RT-PCR and found the total positive rate was 10.5% [32]. Therefore, in countries that consume seafoods such as oysters, the ZD strain may come from contaminated seafood, indicating the need for kits to detect GI NoVs within seafood.

In the phylogenetic analyses of the whole genome and each ORF sequence of the ZD strain (Fig. 3), recombination [20] within the ZD genome was not detected. This result suggests the genome of genotype GI is conserved and stable. By contrast, in prevalent strains, such as the GII.4 strain, genomes are unstable and variable because of recombination. Changes in the sequence of NoV capsid proteins, especially VP1, can, at least partially, reveal the antigenicity and corresponding receptor binding functions of NoV variants [21]. In norovirus genotypes that cause epidemics, such as GII.4, evolution of capsid proteins can lead to evasion of host immune defenses and evolution

of new subtypes. Therefore, changes in capsid proteins of NoVs need to be analyzed in the study of virus evolution. After narrowing the scope of the VP1 sequence of GI.1 from 1968 to 2018, a linear relationship with time from 1968 to 2018 was detected. On the basis of phylogenetic analysis and changes in amino acids encoding VP1, the GI.1 genotype was divided into three clusters related to time of appearance: “older,” “earlier,” and “new” (Fig. 4). Each cluster had unique amino acid substitution characteristics, and the number of amino acid substitutions increased with the timeline. Consistent with a previous study [33], of the several amino acid substitutions identified in GI.1, most were in the highly variable P2 region. According to previous studies, the GI.1 amino acid residue sites associated with HBGA binding were 327, 329, 338, 342, 344, 375, 377, 378, and 380[34]. However, in GI.1 strains analyzed in this study, amino acid substitutions within the P2 region did not occur in the epitopes that are highly associated with NoV invasion and infection (Table 4). Comparison of evolution rates between non-GII.4 and GII.4 strains [21] suggests non-GII.4 strains are subject to less adaptive pressure. In this study, the low recombination frequency could explain the low prevalence and low incidence of GI.1.

To summarize, whole-genome sequencing and bioinformatic analysis demonstrated the ZD strain was a rare GI.1 genotype. Phylogenetic analysis indicated a conservative evolutionary pattern in GI.1, with high homology among strains isolated from different regions in the same period. These results improve understanding of the GI genotype, the nondominant epidemic NoV strain. The results also suggest possible correlation between GI NoVs and seafood. This study provides reference for future genetic analyses and studies of evolutionary patterns of nondominant epidemic strains.

Funding: This work was supported by grants from the Ministry of Science and Technology of China (2018YFC1603900, 2017YFA0505901 to K.M.), National Natural Science Foundation of China (31970136 to K.M.), International Joint Research Project of the Institute of Medical Science, University of Tokyo (Extension-2019-K3006 to K.M.).

Conflicts of Interest: The authors declare no conflict of interest.

Availability of data and material: Data of ZD strain is deposited in China National Microbiology Data Center (NMDC) with accession numbers NMDCN0000Q78.

Code availability: Not applicable.

Authors' contributions: KM conceived and designed the experiments; WSZ, WZ and GBT collected the samples and performed the experiments; WSZ, XLH and XTZ analyzed the data; WSZ and KM wrote the paper; KM and YHX did the editing and proofreading. All authors have read and agreed to the published version of the manuscript.

Ethics approval: This article is in compliance with ethical standards for research.

Consent to participate: Informed consent was obtained from the study subjects themselves.

References

1. Smith HQ, Smith TJ (2019) The Dynamic Capsid Structures of the Noroviruses. *Viruses* 11:235. <https://doi.org/10.3390/v11030235>
2. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM (1972) Visualization by Immune Electron Microscopy of a 27-nm Particle Associated with Acute Infectious Nonbacterial Gastroenteritis. *Journal of Virology* 10:1075-1081. <https://doi.org/10.1128/jvi.10.5.1075-1081.1972>
3. Atmar RL, Estes MK (2001) Diagnosis of noncultivable gastroenteritis viruses, the human caliciviruses. *Clin Microbiol Rev* 14:15-37. <https://doi.org/10.1128/CMR.14.1.15-37.2001>
4. Thorne LG, Goodfellow IG (2014) Norovirus gene expression and replication. *Journal of General Virology* 95:278-291. <https://doi.org/10.1099/vir.0.059634-0>
5. Hardy ME (2005) Norovirus protein structure and function. *FEMS Microbiology Letters* 253: 1-8. <https://doi.org/10.1016/j.femsle.2005.08.031>
6. Bányai K, Estes MK, Martella V, Parashar UD (2018) Viral gastroenteritis. *The Lancet* 392: 175-186. [https://doi.org/10.1016/S0140-6736\(18\)31128-0](https://doi.org/10.1016/S0140-6736(18)31128-0)
7. Hutson AM, Atmar RL, Marcus DM, Estes MK (2003) Norwalk Virus-Like Particle Hemagglutination by Binding to H Histo-Blood Group Antigens. *JOURNAL OF VIROLOGY* 77: 405-415. <https://doi.org/10.1128/JVI.77.1.405-415.2003>
8. Robilotti E, Deresinski S, Pinsky BA (2015) Norovirus. *Clin Microbiol Rev* 28:134-64. <https://doi.org/10.1128/CMR.00075-14>
9. Chhabra P, de Graaf M, Parra GI, Chan MC, Green K, Martella V, Wang Q, White PA, Katayama K, Vennema H, Koopmans MPG, Vinjé J (2019) Updated classification of norovirus genogroups and genotypes. *J Gen Virol* 100:1393-1406. <https://doi.org/10.1099/jgv.0.001318>
10. Lopman BA, Steele D, Kirkwood CD, Parashar UD (2016) The Vast and Varied Global Burden of Norovirus: Prospects for Prevention and Control. *PLoS Med* 13: e1001999. <https://doi.org/10.1371/journal.pmed.1001999>
11. Netzler NE, Enosi Tuipulotu D, White PA (2019) Norovirus antivirals: Where are we now? *Med Res Rev* 39: 860-886. <https://doi.org/10.1002/med.21545>
12. Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY (2016) Global Economic Burden of Norovirus Gastroenteritis. *PLoS One* 11:1-16. <https://doi.org/10.1371/journal.pone.0151219>
13. Jin M, Wu S, Kong X, Xie H, Fu J, He Y, Feng W, Liu N, Li J, Rainey JJ, Hall AJ, Vinjé J, Duan Z (2020) Norovirus Outbreak Surveillance, China, 2016-2018. *Emerg Infect Dis* 26:437-445. <https://doi.org/10.3201/eid2603.191183>
14. Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA (2013) Recombination within the pandemic norovirus GII.4 lineage. *J Virol* 87:6270-6282. <https://doi.org/10.1128/JVI.03464-12>
15. Pogan R, Schneider C, Reimer R, Hansman G, Utrecht C (2018) Norovirus-like VP1 particles exhibit isolate dependent stability profiles. *Journal of Physics: Condensed Matter* 30: 064006.
16. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* 49: D412-D419. <https://doi.org/10.1093/nar/gkaa913>
17. Zhou ZJ, Qiu Y, Pu Y, Huang X, Ge XY (2020) BioAider: An efficient tool for viral genome analysis and its application in tracing SARS-CoV-2 transmission Sustainable Cities and Society. *Sustain Cities Soc* 63:1-13. <https://doi.org/10.1016/j.scs.2020.102466>
18. Kroneman A, Vega E, Vennema H, Vinjé J, White PA, Hansman G, Green K, Martella V,

- Katayama K, Koopmans M (2013) Proposal for a unified norovirus nomenclature and genotyping. *Arch Virol* 158:2059-2068. <https://doi.org/10.1007/s00705-013-1708-5>
19. Xue L, Wu Q, Kou X, Cai W, Zhang J, Guo W (2015) Genome characterization of a GII.6 norovirus strain identified in China. *Infect Genet Evol* 31:110-7. <https://doi.org/10.1016/j.meegid.2015.01.027>
 20. Bull RA, Hansman GS, Clancy LE, Tanaka MM, Rawlinson WD, White PA (2005) Norovirus Recombination in ORF1/ORF2 Overlap. *Emerg Infect Dis* 11:1079-1085. <https://doi.org/10.3201/eid1107.041273>
 21. de Graaf M, van Beek J, Koopmans MP (2016) Human norovirus transmission and evolution in a changing world. *Nat Rev Microbiol* 14: 421-433. <https://doi.org/10.1038/nrmicro.2016.48>
 22. Mattison CP, Cardemil CV, Hall AJ (2018) Progress on norovirus vaccine research: public health considerations and future directions. *Expert Rev Vaccines* 17:773-784. <https://doi.org/10.1080/14760584.2018.1510327>
 23. Ge L, Chen X, Liu J, Zheng L, Chen C, Luo S, Guo P, Kong J, Song Y, Huo Y (2012) Genomic and biological characterization of a pandemic norovirus variant GII.4 Sydney 2012. *Virus Genes* 56:174-181. <https://doi.org/10.1007/s11262-019-01729-0>
 24. Carmona-Vicente N, Vila-Vicent S, Allen D, Gozalbo-Rovira R, Iturriza-Gómara M, Buesa J, Rodríguez-Díaz J (2016) Characterization of a Novel Conformational GII.4 Norovirus Epitope: Implications for Norovirus-Host Interactions. *J Virol* 90:7703-7714. <https://doi.org/10.1128/JVI.01023-16>
 25. Mateo R, Lindesmith LC, Garg SJ, Gottlieb K, Lin K, Said S, Leon JS, Sims AC, Weber DJ, Baric RS, Tucker SN, Taylor DN (2020) Production and Clinical Evaluation of Norwalk GI.1 Virus Lot 001-09NV in Norovirus Vaccine Development. *J Infect Dis* 221:919-926. <https://doi.org/10.1093/infdis/jiz540>
 26. Teunis PF, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, Le Pendu J, Calderon RL (2008) Norwalk virus: how infectious is it? *J Med Virol*, 2008. 80:1468-1476. <https://doi.org/10.1002/jmv.21237>
 27. Randazzo W, D'Souza DH, Sanchez G (2018) Norovirus: The Burden of the Unknown. *Adv Food Nutr Res* 86:13-53. <https://doi.org/10.1016/bs.afnr.2018.02.005>
 28. Ha JH, Choi C, Ha SD (2014) Evaluation of Immunomagnetic Separation Method for the Recovery of Hepatitis A Virus and GI.1 and GII.4 Norovirus Strains Seeded on Oyster and Mussel. *Food Environ Virol* 6: 290-296. <https://doi.org/10.1007/s12560-014-9156-2>
 29. Le Guyader FS, Bon F, DeMedici D, Parnaudeau S, Bertone A, Crudeli S, Doyle A, Zidane M, Suffredini E, Kohli E, Maddalo F, Monini M, Gallay A, Pommepuy M, Pothier P, Ruggeri FM (2006) Detection of multiple noroviruses associated with an international gastroenteritis outbreak linked to oyster consumption. *J Clin Microbiol* 44: 3878-3882. <https://doi.org/10.1128/JCM.01327-06>
 30. Lysén M, Thorhagen M, Brytting M, Hjertqvist M, Andersson Y, Hedlund KO (2009) Genetic diversity among food-borne and waterborne norovirus strains causing outbreaks in Sweden. *J Clin Microbiol* 47: 2411-2418. <https://doi.org/10.1128/JCM.02168-08>
 31. Nenonen NP, Hannoun C, Olsson MB, Bergström T (2009) Molecular analysis of an oyster-related norovirus outbreak. *J Clin Virol* 45:105-108. <https://doi.org/10.1016/j.jcv.2009.04.011>
 32. Cheng PK, Wong DK, Chung TW, Lim WW (2005) Norovirus contamination found in oysters worldwide. *J Med Virol* 76: 593-597. <https://doi.org/10.1002/jmv.20402>

33. Sukhrie FH, Teunis P, Vennema H, Bogerman J, van Marm S, Beersma MF, Koopmans M (2013) P2 domain profiles and shedding dynamics in prospectively monitored norovirus outbreaks. *J Clin Virol* 56: 286-292. <https://doi.org/10.1016/j.jcv.2012.12.006>
34. van Loben Sels JM, Green KY (2019) The Antigenic Topology of Norovirus as Defined by B and T Cell Epitope Mapping: Implications for Universal Vaccines and Therapeutics. *Viruses* 11: 432. <https://doi.org/10.3390/v11050432>

Table S1 Comparison of mutation sites between the ZD strain and MT008453.1 noroviruses

Strain	Site															
	ORF1															
MT008453.1 ZD	33	42	63	92	255	384	525	571	678	969	978	1053	1131	1167	1236	1290
	T	C	G	C	T	T	T	A	C	T	G	T	A	A	A	A
	C	T	A	T	C	A	C	G	T	C	A	C	G	G	G	G
MT008453.1 ZD	1551	1890	2061	2076	2209	2589	2640	3147	3403	3495	3705	3813	3972	4056	4200	4236
	A	T	C	G	C	T	T	G	C	C	G	G	G	C	A	G
	G	C	T	A	T	C	C	A	T	T	A	A	A	T	G	A
												ORF2				
MT008453.1 ZD	4299	4347	4518	4687	4758	4779	5073	5133	5187	5308	5350	5635	5731	5761	5872	5995
	T	C	G	T	T	A	C	T	G	G	T	T	G	C	T	T
	C	T	A	G	C	G	T	C	A	A	C	C	A	T	G	C
												ORF3				
MT008453.1 ZD	6235	6246	6452	6484	6595	6715	6819	6833	6837	6888	6933	6941	7200	7260	7296	7383
	G	C	T	T	C	T	T	C	C	T	G	G	A	T	C	C
	A	T	C	C	T	C	C	T	T	C	A	A	G	C	T	T

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)