Complete Genome Sequence of GII.9 Norovirus

Zilong Zhang
shanghai international travel healthcare center
Danlei Liu
Shanghai Jiao Tong University - Minhang Campus: Shanghai Jiao Tong University
Zilei Zhang
Shanghai Jiao Tong University - Minhang Campus: Shanghai Jiao Tong University
Peng Tian
USDA Agricultural Research Service
Qingping Wu
Guangdong Institute of Microbiology
Dapeng Wang
Shanghai Jiao Tong University - Minhang Campus: Shanghai Jiao Tong University
Zhengan Tian (tianzhenganciq@163.com)
shanghai customs district P.R.China https://orcid.org/0000-0001-9835-8438

Research Article

Keywords: Norovirus, genome sequence, Genotype GII.9, acute gastroenteritis outbreaks

Posted Date: June 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-623335/v1

Abstract

Norovirus is recognized as one of the leading causes of acute gastroenteritis outbreaks. Genotype GII.9 was first detected in Norfolk, USA in 1997. However, the complete genome sequence of this genotype was not established yet. In this study, a complete genome sequence of a GII.9[P7] norovirus, marked as SCD1878, from a patient was established using a high-throughput sequencing and rapid amplification of cDNA ends (RACE) technology. The complete genome sequence of SCD1878_GII.9P7 was 7544 nucleotides (nts) in length with a 3' poly (A) tail, including three open reading frames. Homology analysis indicated that SCD1878_GII.9P7 shares 92.1%-92.3% identity with GII.P7 (AB258331 and AB039777) and 96.7%-97.4% identity with GII.9 (AY038599 and DQ379715) sequences. The results suggested that SCD1878 is a member of GII.P7 for P genotypes and GII.9 for genotypes. The viral sequence filled the gap in the whole genome level of the GII.9 genotype.

Introduction

Norovirus (NoV) is recognized as one of the leading causes of acute gastroenteritis outbreaks. NoV belongs to the Caliciviridae family and its positive sense ~7.5 kb RNA genome [20]. Phylogenetically, NoV can be segregated into ten genogroups and further divided into genotypes based on amino acid diversity of the complete VP1. Among the known genogroups, GII is the largest one which consists 26 genotypes, including 23 human NoV genotypes that are responsible for the most epidemics and three porcine NoV (GII.11/18/19) genotypes [2]. As the diversity of NoV increasing through recombination, dual typing of NoV classification was proposed. Partial nucleotide sequences of the RNA-dependent RNA polymerase (RdRp) region of ORF1 is utilized for NoV P type classification independently from genotype. A total of 37 P-types are now included in GII viruses [2].

The first strain of genotype GII.9 virus (VA97207) was detected in Norfolk, VA in USA in 1997 [7]. The partial sequence (a 3290 bp fragment including the complete ORF2 region) of this strain was uploaded to GenBank in 2001 (Accession number: AY038599) [7]. Compared with other genotypes, GII.9 strains were rarely reported. Gelaw et.al detected only one GII.9 strain from 450 clinical samples by RT-PCR and partially sequenced its VP1 gene (300 bp) [4]. The presence were also reported in wastewater in South Africa and oyster samples in Japan [6, 16]. Nevertheless, there was no submission of GII.9 sequence to NoroNet from 2005 to 2016 [19].

Methods

In this study, a rare GII.9[P7] whole genome sequence was obtained from a clinical sample. The anal swab and epidemiological data were collected through the acute gastroenteritis (AGE) outbreak surveillance system monitored by Shanghai Customs. The patient is a Japanese 22-year-old female traveled from India and arrived Shanghai Pudong Airport on March 19^{th,} 2018. The patient had symptoms of diarrhea and vomiting and was diagnosed as AGE.

Majority of this whole viral sequence was generated through RNA-seq and the ends of the viral genome was supplemented by rapid amplification of cDNA ends (RACE) kit (Vazyme, Nanjing, China) [13, 14]. The whole genomic sequence was then assembled and validated using CLC Genomics Workbench (https://digitalinsights.qiagen.com). The assembled viral genome was genotyped using a web-based genotyping tool [10]. Phylogenetic Tree was constructed by MEGA X and blast alignment were displayed by circoletto [3, 11]. The complete sequence, marked as SCD1878_GII.9P7, was deposited in GenBank with the Accession number MZ312111.

A total of 1976 human NoV genome sequences (6400-8500 bp) were fetched from ViPR on March 10th, 2021 [18]. BioAider was used to remove highly similar sequences with sequence identity over 97%. PhyloSuite was used to conduct, manage and streamline the analyses [21]. Sequence was aligned with MAFFT [8]. Best partitioning scheme and evolutionary models for 1 pre-defined partitions were selected using PartitionFinder2 [12] with greedy algorithm and AICc criterion. Maximum likelihood phylogenies were inferred using IQ-TREE [17] under the GTR+I+G4+F model for 20000 ultrafast bootstraps and the Shimodaira–Hasegawa–like approximate likelihood-ratio test [5].

Results And Discussion

The complete genome sequence of SCD1878_GII.9P7 is 7544 nucleotides (nts) in length, with a 3' poly (A) tail. As expected, genome contains three open reading frames (ORFs) (Table 1). The ORF1 can be cleaved into six nonstructural proteins, including denoted p48, NTPase, p22, VPg, Pro and RdRp. The remained two ORFs encode two structural proteins (VP1 and VP2). Genetic similarity and mutations of the sequence against the reference sequence (NC_029646.1, GII) were shown in Table 1.

Homology analysis indicated that SCD1878_GII.9P7 shares 92.1%-92.3% and 96.7%-97.4% identity with GII.P7 (AB258331 and AB039777) and GII.9 (AY038599 and DQ379715) sequences at nucleotide and amino acid levels either in the RdRp gene or the VP1 protein, suggesting that SCD1878 is a member of GII.P7 for P genotypes and GII.9 for G genotypes (Figure 1). We have further explored the possibility in constituting a new GII.P9 genotype. Sequence ranging from 600 nt to 1400 nt was selected for P genotype evolution analysis (data not shown). However, the sequence variation among this sequence and GII.P6 and GII.P7 genotypes was little. The RdRp region of the related GII.P9 sequence could not form a new cluster on phylogenetic tree, and the criteria of 2×SD could not be fulfilled, thus it cannot be a new P-type [1, 9].

Phylogenetic analysis of whole genome sequences showed that SCD1878_GII.9[P7] was clustered into a monophyletic clade with high confidence (bootstrap value, BP =100%, Figure 2), together with three genotypes: GII.6[P7], GII.7[P7] and GII.14[P7]. Within the clade, SCD1878_GII.9[P7] was grouped into a distinct branch alone, proved this sequence to be the first whole genome sequence of GII.9[P7] genotype. Potential recombination within the complete viral genome was screened using SimPlot and no typical similarities exchange of reference sequences was observed [15].

The rapid developing sequencing technology provides great promotion for viruses monitoring. As the second generation and third generation sequencing technology developed, discovering and analyzing longer viral genome become practical. Additional complete RdRp sequences or ideally complete genome sequences for all reference strains will help to improve the robustness of the present classification system [1]. Obtaining whole genome sequences of rare genotype could not only enrich the database but also provide precious information for evolution analysis, reference genome for diversity analysis, and screening for drug and vaccine development.

Nucleotide sequence accession number

The GenBank accession number for norovirus SCD1878_GII.9P7 is MZ312111.

Declarations

Funding: This research was funded by the National Science and Technology Major Project (2018ZX10101003-002), General Administration of Customs Project (2019HK125) and State Key Laboratory of Applied Microbiology Southern China (SKLAM 002-2019).

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

Ethics approval: Ethical approval for this study was obtained from the China CDC Ethical Review Committee (No. M202007) (Beijing, China).

References

1. 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(10):1393-1406. 10.1099/jgv.0.001318
2. Chhabra P, Graaf MD, Parra GI, Chan MC, Green K, Martella V, Wang Q, White PA, Katayama K, Vennema H, Koopmans MPG, Vinjé J (2020) Corrigendum: Updated classification of norovirus genogroups and genotypes. J Gen Virol 101(8):893. 10.1099/jgv.0.001475
3. Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics 26(20):2620-2621. 10.1093/bioinformatics/btq484
4. Gelaw A, Pietsch C, Mann P, Liebert UG (2019) Molecular detection and characterisation of sapoviruses and noroviruses in outpatient children with diarrhoea in Northwest Ethiopia., vol 147, p e218
5. Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307-321. 10.1093/sysbio/syq010
6. Imamura S, Kanezashi H, Goshima T, Suto A, Ueki Y, Sugawara N, Ito H, Zou B, Kawasaki C, Okada T, Uema M, Noda M, Akimoto K (2018) Effect of High Pressure Processing on a Wide Variety of Human Noroviruses Naturally Present in Aqua-Cultured Japanese Oysters., vol 15, pp 621-626
7. Jiang X, Zhong WM, Farkas T, Huang PW, Wilton N, Barrett E, Fulton D, Morrow R, Matson DO (2002) Baculovirus expression and antigenic characterization of the capsid proteins of three Norwalk-like viruses. Arch Virol 147(1):119-130. 10.1007/s705-002-8306-

5

8. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33(2):511-518
9. 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(10):2059-2068. 10.1007/s00705-013-1708-5
10. Kroneman A, Vennema H, Deforche K, Avoort HVD, Peñaranda S, Oberste MS, Vinjé J, Koopmans M (2011) An automated genotyping tool for enteroviruses and noroviruses. J Clin Virol 51(2):121-125. 10.1016/j.jcv.2011.03.006
11. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547-1549
12. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34(3):772-773. 10.1093/molbev/msw260
13. Liu D, Zhang Z, Li S, Wu Q, Tian P, Zhang Z, Wang D (2020) Fingerprinting of human noroviruses co-infections in a possible foodborne outbreak by metagenomics. Int J Food Microbiol 333:108787. 10.1016/j.ijfoodmicro.2020.108787
14. Liu D, Zhang Z, Wu Q, Tian P, Geng H, Xu T, Wang D (2020) Redesigned Duplex RT-qPCR for the Detection of GI and GII Human Noroviruses. Engineering-Prc 6(4):442-448. https://doi.org/10.1016/j.eng.2019.08.018
15. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73(1):152-160
16. Mabasa VV, Meno KD, Taylor MB, Mans J (2018) Environmental Surveillance for Noroviruses in Selected South African Wastewaters 2015-2016: Emergence of the Novel GII.17., vol 10, pp 16-28
17. Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30(5):1188-1195. 10.1093/molbev/mst024
18. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH (2012) ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res 40(D1):D593-D598. 10.1093/nar/gkr859
19. van Beek J, de Graaf M, Al-Hello H, Allen DJ, Ambert-Balay K, Botteldoorn N, Brytting M, Buesa J, Cabrerizo M, Chan M, Cloak F, Di Bartolo I, Guix S, Hewitt J, Iritani N, Jin M, Johne R, Lederer I, Mans J, Martella V, Maunula L, Mcallister G, Niendorf S, Niesters HG, Podkolzin AT, Poljsak-Prijatelj M, Rasmussen LD, Reuter G, Tuite G, Kroneman A, Vennema H, Koopmans M (2018) Molecular surveillance of norovirus, 2005-16: an epidemiological analysis of data collected from the NoroNet network. Lancet Infect Dis 18(5):545-553. 10.1016/S1473-3099(18)30059-8
20. Vinjé J, Estes MK, Esteves P, Green KY, Katayama K, Knowles NJ, L Homme Y, Martella V, Vennema H, White PA (2019) ICTV Virus Taxonomy Profile: Caliciviridae. J Gen Virol 100(11):1469-1470. 10.1099/jgv.0.001332
21. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour 20(1):348-355. 10.1111/1755-0998.13096

Table

Table 1 Detailed information of sequence SCD1878_GII.9P7 genetic diversity analysis†

†: genetic diversity was analyzed based on reference sequence NC_029646.1

*: Inserts / Deletes / Misaligned / Frameshifts

Figures

	Begin	End	Coverage	Score	Concordance	Matches	Identities	I/D/M/F*	Stop Codons
NT	1	7518	100%	4816	32.50%	7479(99.1%)	4987(66.1%)	27/39	
CDS									
ORF1	1	1700	100%	9172	78.30%	1692(99.2%)	1261(73.9%)	6/8/0/0	1
ORF2	1	536	100%	2712	71.10%	535 (99.3%)	351 (65.1%)	3/1/0/0	1
ORF3	1	260	100%	1086	66.70%	256 (98.5%)	159 (61.2%)	0/4/0/0	1
Proteins									
nonstructural polyprotein (YP_009237897.1)	1	1700	100%	9172	78.30%	1692(99.2%)	1261(73.9%)	6/8/0/0	1
p48 (YP_009238492.1)	1	330	100%	1541	65.90%	328 (97.6%)	209 (62.2%)	6/2/0/0	0
NTPase (YP_009238487.1)	1	366	100%	2126	87.60%	366 (100%)	299 (81.7%)	0/0/0/0	0
p22 (YP_009238488.1)	1	179	100%	536	46.10%	173 (96.6%)	78 (43.6%)	0/6/0/0	0
VPg (YP_009238489.1)	1	133	100%	832	92.10%	133 (100%)	120 (90.2%)	0/0/0/0	0
Pro (YP_009238490.1)	1	181	100%	1108	86.00%	181 (100%)	144 (79.6%)	0/0/0/0	0
RdRp (YP_009238491.1)	1	510	100%	3028	84.30%	510 (100%)	410 (80.4%)	0/0/0/0	0
VP1 (YP_009237898.1)	1	536	100%	2712	71.10%	535 (99.3%)	351 (65.1%)	3/1/0/0	1
VP2 (YP_009237899.1)	1	260	100%	1086	66.70%	256 (98.5%)	159 (61.2%)	0/4/0/0	1

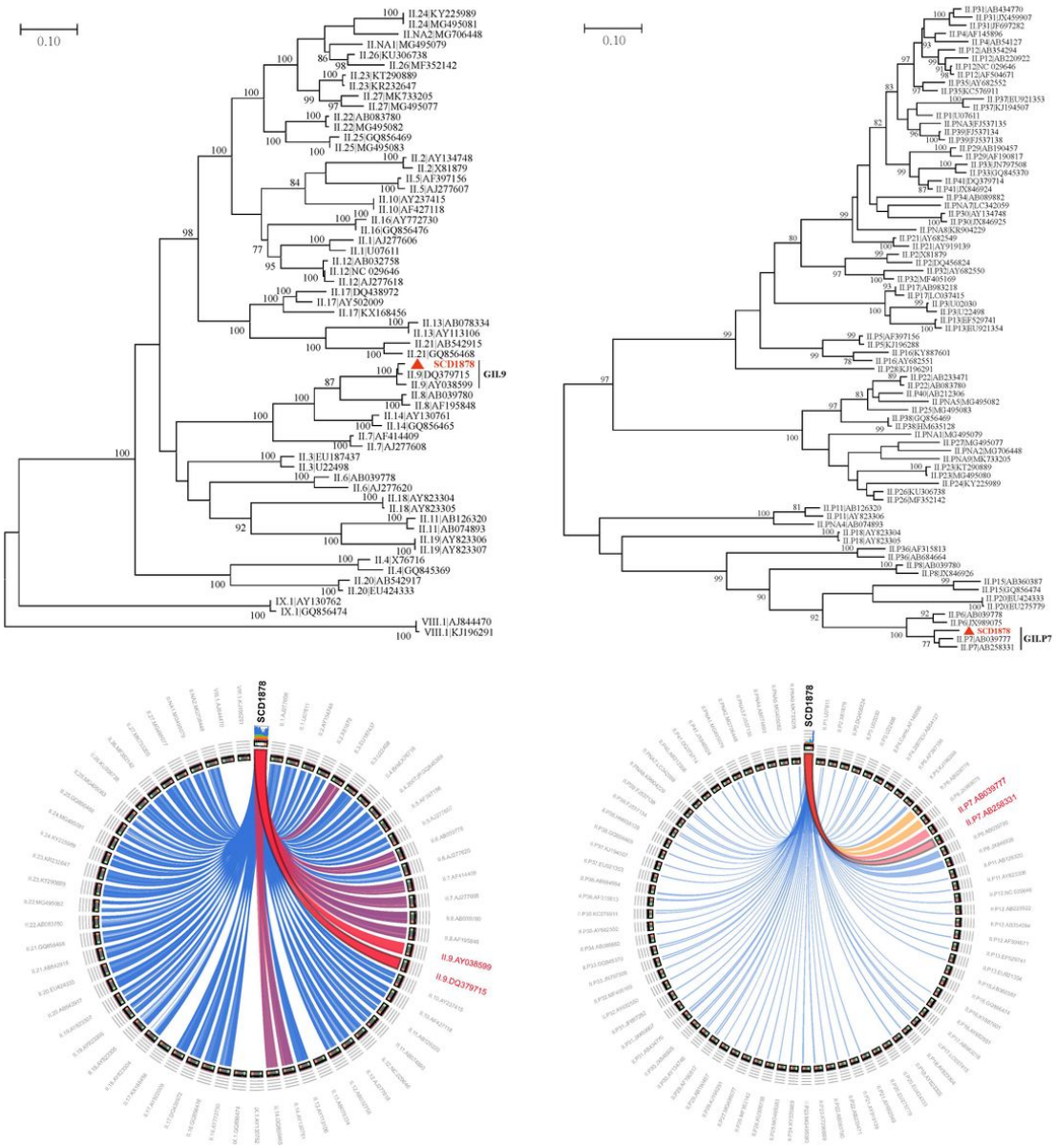

Figure 1

Phylogenetic Tree of genotypes (left) and P-types (right) based on amino acid diversity of the complete VP1 and nucleotide diversity of the RNA-dependent RNA polymerase (RdRp) region respectively. The percentage of replicate trees (>75%) in the bootstrap test (500 replicates) were shown next to the branches. The blast comparison results were displayed below the corresponding phylogenetic tree. Sequences with high identity were connected with red line and corresponding Accession numbers were highlighted in bold red while low identity sequences were connected with blue line.

Figure 2

Maximum likelihood phylogenies for human NoV genome sequences (6400-8500 bp). A partial enlargement of SCD1878_GII.9P7 related sequences was zoomed in yellow box. Ultrafast bootstraps and the Shimodaira–Hasegawa–like approximate likelihood-ratio were included in node labels.

Supplementary Files

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

- support.pdf