



Universiteit  
Leiden  
The Netherlands

## Metagenomic surveillance of viral gastroenteritis in a public health setting

Schmitz, D.; Zwagemaker, F.; Veer, B. van der; Vennema, H.; Laros, J.F.J.; Koopmans, M.P.G.; ... ; Kroneman, A.

### Citation



Schmitz, D., Zwagemaker, F., Veer, B. van der, Vennema, H., Laros, J. F. J., Koopmans, M. P. G., ... Kroneman, A. (2023). Metagenomic surveillance of viral gastroenteritis in a public health setting. *Microbiology Spectrum*, 11(4). doi:10.1128/spectrum.05022-22

Version: Publisher's Version  
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)  
Downloaded from: <https://hdl.handle.net/1887/3753314>

**Note:** To cite this publication please use the final published version (if applicable).



# Metagenomic Surveillance of Viral Gastroenteritis in a Public Health Setting

 Dennis Schmitz,<sup>a,b</sup> Florian Zwagemaker,<sup>a</sup> Bas van der Veer,<sup>a</sup> Harry Vennema,<sup>a</sup> Jeroen F. J. Laros,<sup>a,c</sup> Marion P. G. Koopmans,<sup>b</sup>  
 Miranda De Graaf,<sup>b</sup> Annelies Kroneman<sup>a</sup>

<sup>a</sup>National Institute of Public Health and the Environment, Center for Infectious Disease Control, Bilthoven, The Netherlands

<sup>b</sup>Erasmus Medical Center, Viroscience, Rotterdam, The Netherlands

<sup>c</sup>Leiden University Medical Center, Department of Human Genetics, Leiden, The Netherlands

Miranda De Graaf and Annelies Kroneman contributed equally to this study.

**ABSTRACT** Norovirus is the primary cause of viral gastroenteritis (GE). To investigate norovirus epidemiology, there is a need for whole-genome sequencing and reference sets consisting of complete genomes. To investigate the potential of shotgun metagenomic sequencing on the Illumina platform for whole-genome sequencing, 71 reverse transcriptase quantitative PCR (RT-qPCR) norovirus positive-feces (threshold cycle [ $C_T$ ], <30) samples from norovirus surveillance within The Netherlands were subjected to metagenomic sequencing. Data were analyzed through an in-house next-generation sequencing (NGS) analysis workflow. Additionally, we assessed the potential of metagenomic sequencing for the surveillance of off-target viruses that are of importance for public health, e.g., sapovirus, rotavirus A, enterovirus, parechovirus, aichivirus, adenovirus, and bocaparvovirus. A total of 60 complete and 10 partial norovirus genomes were generated, representing 7 genogroup I capsid genotypes and 12 genogroup II capsid genotypes. In addition to the norovirus genomes, the metagenomic approach yielded partial or complete genomes of other viruses for 39% of samples from children and 6.7% of samples from adults, including adenovirus 41 ( $N = 1$ ); aichivirus 1 ( $N = 1$ ); coxsackievirus A2 ( $N = 2$ ), A4 ( $N = 2$ ), A5 ( $N = 1$ ), and A16 ( $N = 1$ ); bocaparvovirus 1 ( $N = 1$ ) and 3 ( $N = 1$ ); human parechovirus 1 ( $N = 2$ ) and 3 ( $N = 1$ ); Rotavirus A ( $N = 1$ ); and a sapovirus GI.7 ( $N = 1$ ). The sapovirus GI.7 was initially not detected through RT-qPCR and warranted an update of the primer and probe set. Metagenomic sequencing on the Illumina platform robustly determines complete norovirus genomes and may be used to broaden gastroenteritis surveillance by capturing off-target enteric viruses.

**IMPORTANCE** Viral gastroenteritis results in significant morbidity and mortality in vulnerable individuals and is primarily caused by norovirus. To investigate norovirus epidemiology, there is a need for whole-genome sequencing and reference sets consisting of full genomes. Using surveillance samples sent to the Dutch National Institute for Public Health and the Environment (RIVM), we compared metagenomics against conventional techniques, such as RT-qPCR and Sanger-sequencing, with norovirus as the target pathogen. We determined that metagenomics is a robust method to generate complete norovirus genomes, in parallel to many off-target pathogenic enteric virus genomes, thereby broadening our surveillance efforts. Moreover, we detected a sapovirus that was not detected by our validated gastroenteritis RT-qPCR panel, which exemplifies the strength of metagenomics. Our study shows that metagenomics can be used for public health gastroenteritis surveillance, the generation of reference-sets for molecular epidemiology, and how it compares to current surveillance strategies.

**KEYWORDS** next-generation sequencing, norovirus, surveillance, gastroenteritis, enteric viruses, public health, surveillance studies, transmissible gastroenteritis virus

**Editor** Gabriel I. Parra, U.S. Food and Drug Administration

**Copyright** © 2023 Schmitz et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Miranda De Graaf, m.degraaf@erasmusmc.nl.

The authors declare no conflict of interest.

**Received** 8 December 2022

**Accepted** 6 June 2023

**Published** 11 July 2023

Norovirus is the most common cause of viral gastroenteritis (GE) and results in significant morbidity and mortality in the young, elderly, and immunocompromised worldwide (1–6). To reduce the burden of norovirus disease, vaccines are being developed, and the first ones have completed phase II trials, but none have been approved yet (7, 8). Similarly, several drugs and biologicals have been tested off-label, such as ribavirin, immunoglobulins, and nitazoxanide, but conclusive evidence of their effectiveness is lacking (9, 10). These studies are complicated, as treatment and vaccine efficacy may in part be strain or genotype dependent.

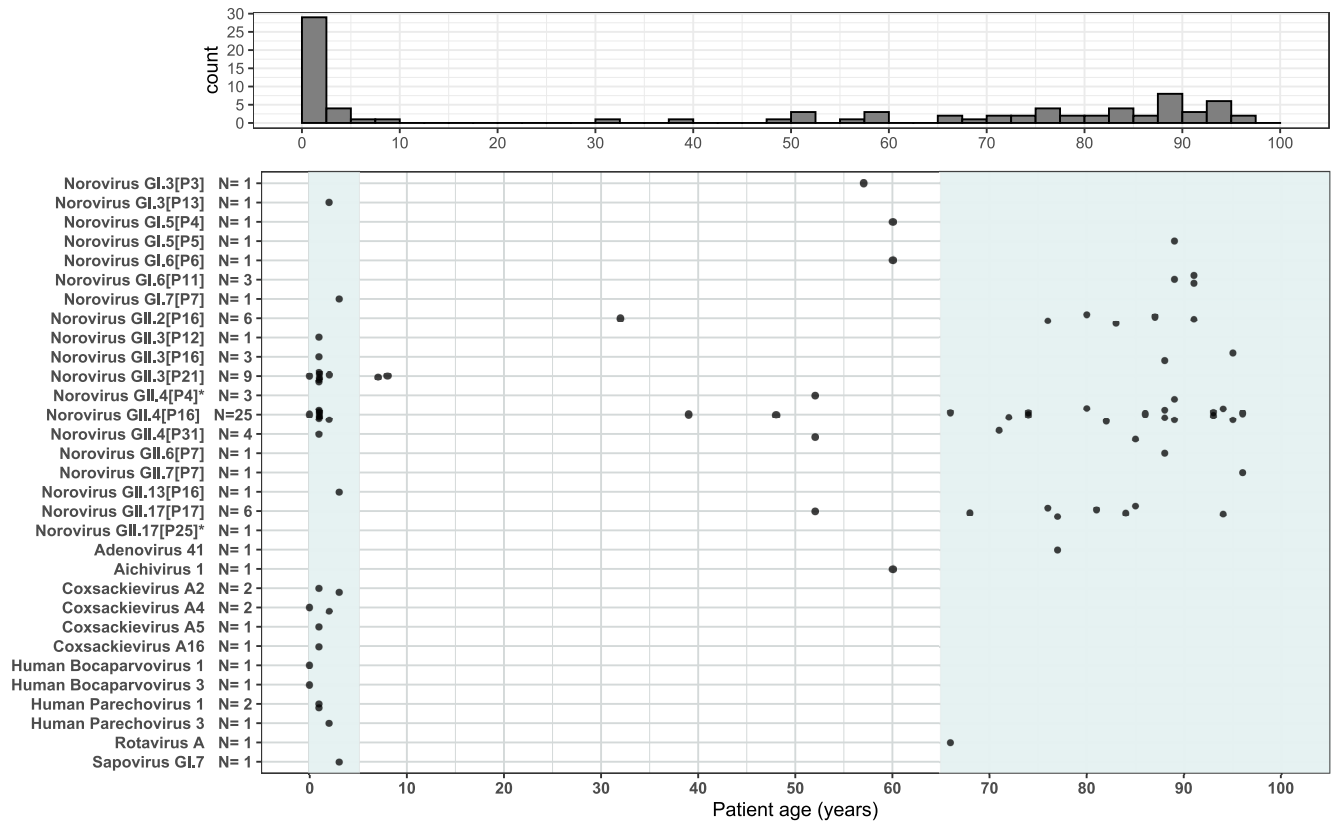
Norovirus is a highly diverse genus and belongs to the *Caliciviridae* family. The *Norovirus* genus is currently divided in 10 genogroups (G), of which viruses belonging to GI, GII, GIV, GVIII, and GIX infect humans. The GI and GII genogroups are further subdivided into at least 48 genotypes (11). Of these, GII genotype 4 (GII.4) is the most prevalent, and due to antigenic drift, new GII.4 variants emerge which can become dominant and replace the previous variant (11). Diversity is further increased by recombination, with the open reading frame 1/2 (ORF1/2) junction as a recombination hot spot. Recombination usually occurs within a genogroup, but some intergenogroup recombinant genomes have also been identified (11–13). Therefore, for comprehensive genomic characterization of noroviruses, both ORF1 and ORF2 are typed.

Although norovirus surveillance to the genotype level is limited, the most common method to genotype norovirus in routine public health and food safety settings is by Sanger sequencing of the 3' ORF1 end of the RNA-dependent RNA polymerase (RdRp; non-structural protein 7, NS7), the 5' ORF2 end of the capsid (viral protein 1, VP1) coding domain regions (CDR), or a larger fragment containing both fragments and spanning the ORF1/2 junction. A classification scheme was developed in which the VP1 genomic sequence and the RdRp sequences are assigned to genogroups and genotypes to capture both antigenic divergence and the presence of recombinants (11, 14). The inferred genotypes and sequences support outbreak investigations and epidemiological surveillance of norovirus circulation (11, 14–16). While these partial genome sequences are informative, the generated sequences are generally too short to conclusively determine transmission chains within an outbreak.

With next-generation sequencing (NGS) techniques becoming more widely available, their application for sewage-based community surveillance and foodborne outbreak surveillance is being explored. NGS of difficult food and environmental sample matrices, which usually only contain low levels of viruses, frequently results in short sequences outside the canonical typing region. The identification and typing of these sequences would require a continuously updated reference set containing the complete genomic information of norovirus strains present in the human population. Full or even partial genomes generated by NGS would also provide a higher genetic resolution to investigate outbreaks and aid in genotype to phenotype studies. In the latest norovirus nomenclature update, the authors anticipated that the next update would be based on complete genome reference sequences, enabling typing of sequences outside the typing regions (11).

NGS generates reads from all taxa present in the sample, and with a metagenomic-focused analysis, reads are reconstituted to (partial) genomes, which are taxonomically classified based on their similarity to known strains in public databases. Therefore, in addition to targeted norovirus surveillance, metagenomics also provides an opportunity to detect several other viruses that cause GE in humans and are, therefore, relevant for public health, including Rotavirus A (RVA), enteroviruses (EV), and a range of other enteric viruses that are less frequent causes of outbreaks (17–21). Currently, a typical diagnostic approach is to use a multiplex reverse transcriptase quantitative PCR (RT-qPCR) to test for the presence of a range of GE viruses. Further characterization through targeted Sanger sequencing has been limited and is usually targeted to a single virus due to technical limitations and cost. However, with metagenomics, other viruses can be detected during norovirus-targeted NGS-based surveillance.

The primary goal of this study is to determine the robustness of a nontargeted NGS approach as a method to determine full norovirus genomes. Secondarily, we aim to



**FIG 1** Patient age per detected virus species or norovirus genotype. The left light-gray rectangle shows young children (<5 years old), and the right one shows the elderly (>65 years old). The histogram depicts the total number of detected viruses of public health relevance per age group, with bins of 2.5 years. \*, Norovirus GII.17[P25] (N = 1) and GII.4[P4] (N = 1) were present in the same sample and could not be plotted due to unknown patient age. N, total number of times a virus genome was observed.

assess its potential to provide off-target molecular characterization of GE pathogens that are relevant from a public health perspective. For this study we selected norovirus RT-qPCR-positive samples that were sent to the Dutch norovirus surveillance program by regional laboratories for referral in outbreak situations.

**RESULTS**

**Complete norovirus genomes were robustly generated using metagenomics.**

To investigate the potential of shotgun metagenomic sequencing on the Illumina platform, 71 RT-qPCR norovirus-positive feces (threshold cycle [C<sub>T</sub>], <30) samples from norovirus surveillance within The Netherlands were subjected to metagenomics. Of the 71 samples, 66 (93%) yielded norovirus genomic sequences, while off-target viruses were identified in 13 (18%) of the samples (Fig. 1; Table 1). Of the 66 samples resulting in norovirus sequences, 58 (82%) yielded a full norovirus genome, and eight (11%) yielded a partial norovirus genome. Two full and two partial minority norovirus strains were additionally generated from double-infection samples, resulting in a total of 60 full and 10 partial norovirus sequences (Table 1). The five (7.9%) remaining samples were possibly negative due to either a low viral input, the presence of inhibitors, or the presence of another highly abundant taxon. Samples R02-07 and R02-09 had high viral loads (C<sub>T</sub>, <25) but contained a highly abundant adenovirus (AdV) (C<sub>T</sub>, 7.1; Table 1) and a 64-kb *Pseudomonas* phage with a 723× depth of coverage, respectively.

A comparison of RT-qPCR values versus genome completeness at a sequencing depth cut-off of 3 showed that above a C<sub>T</sub> value of 27.1, no norovirus contigs were obtained (Fig. 2). For the mixed-strain samples, block-like coverage peaks near the ORF1/2 junction were observed as a result of homologous regions between these strains, and these regions had to be resolved manually (Fig. 3).

**TABLE 1** Sample overview of 71 norovirus (NoV) RT-qPCR-positive samples<sup>a</sup>

Sample	Collection date (yr-mo-day)	Cluster no. (SNP distance)	NoV C <sub>v</sub> value <sup>b</sup>		Sanger NoV genotype <sup>c</sup>	WGS NoV genotype <sup>c</sup>	WGS major NoV strain			WGS minor NoV strain Length [bp (%)]; genotype <sup>e</sup> ; avg. depth [x(SD)]	Off-target GE virus C <sub>v</sub> value <sup>d</sup>				
			GI	GII			Length [bp (%)]	Avg. depth [x(SD)]	RVA		AdV	SaV	EV	HPeV	WGS off-target GE virus length [bp (%)]
R01-01	2016-01-25		11.7		GII.4[P4]	GII.4[P4]	7,561 (100)	10,849 (5,790)							
R01-02	2016-04-08		13.9		GII.3[P13]	GII.3[P13]	7,780 (100)	5,906 (2,474)							
R01-03	2016-05-12		17.3		GII.17[P25]	GII.17[P25]	7,530 (100)	25 (10)							
R01-04	2016-06		19.6		GII.5[P5]	GII.5[P5]	7,688 (100)	1,262 (2,886)							
R01-05	2016-02-16		23.8		GII.7[P7]	GII.7[P7]	7,735 (100)	186 (62)							
R01-06	2016-08-16		24.7		GII.3[P21]	GII.3[P21]	7,552 (100)	214 (85)							
R01-07	2016-03-01		27.1		GII.3[P3]	GII.3[P3]	7,745 (100)	6,382 (2,434)							
R01-20	2016-01-13		17.9		GII.17[P17]	GII.17[P17]	7,572 (100)	433 (298)							
R02-06	2015-01-31		17.6		GII.17[P17]	GII.17[P17]	7,561 (100)	58 (30)							
R02-07	2015-02-26		20.6		GII.17[P17]	GII.17[P17]	7,561 (100)	285 (140)							
R02-08	2015-05-15	1 (0)	20.6		GII.17[P17]	GII.17[P17]	7,561 (100)	6,274 (2,445)							
R02-09	2015-05-27		16.8		GII.17[P17]	GII.17[P17]	7,561 (100)	59 (34)							
R02-10	2015-05-27	1 (0)	20.8		GII.17[P17]	GII.17[P17]	7,573 (100)	1,588 (577)							
R02-11	2015-11-20		17.4		GII.17[P17]	GII.17[P17]	7,560 (100)	508 (243)							
R02-12	2015-11-27		20.8		GII.17[P17]	GII.17[P17]	7,560 (100)	376 (306)							
R02-13	2016-03-30		20.7		GII.4[P4]	GII.4[P4]	7,685 (100)	3,830 (17,344)							
R02-14	2016-03-30		22.5		GII.4[P4]	GII.4[P4]	7,685 (100)	9 (6)							
R02-15	2016-10		19.0		GII.6[P11]	GII.6[P11]	7,701 (100)	1,397 (560)							
R02-16	2016-10		17.0		GII.6[P11]	GII.6[P11]	7,701 (100)	213							
R02-17	2016-11-28		19.8		GII.2[P16]	GII.2[P16]	7,540 (93)	1,800 (780)							
R02-18	2017-02-16	2 (0)	23.9		GII.2[P16]	GII.2[P16]	7,536 (100)	41 (15)							
R02-19	2017-02-16	2 (0)	18.9		GII.2[P16]	GII.2[P16]	7,539 (100)	309 (120)							
R02-20	2017-12-27		25.6		GII.2[P16]	GII.2[P16]	7,605 (100)	1,143 (349)							
R02-21	2018-01-11	3 (1)	16.2		GII.4[P16]	GII.4[P16]	7,555 (60)	5 (5)							
R02-22	2018-01-12	3 (1)	21.5		GII.4[P31]	GII.4[P31]	7,560 (100)	586 (247)							
R02-23	2018-08-01		19.6		GII.4[P16]	GII.4[P16]	7,574 (100)	10 (6)							
R02-24	2018-07-03		22.0		GII.4[P16]	GII.4[P16]	7,574 (100)	1,727 (639)							
R03-01	2018-02-14		21.2		GII.4[P16]	GII.4[P16]	7,564 (100)	35 (16)							
R03-02	2018-03-09		17.8		GII.4[P16]	GII.4[P16]	7,564 (100)	8,543 (3017)							
R03-03	2018-03-21		19.9		GII.4[P16]	GII.4[P16]	7,569 (100)	9,561 (3,912)							
R03-04	2018-07-28		18.3		GII.4[P16]	GII.4[P16]	7,569 (100)	262 (97)							
R03-05	2018-07-28		21.1		GII.4[P16]	GII.4[P16]	7,569 (100)	168 (64)							
R03-06	2018-09-18		16.4		GII.4[P16]	GII.4[P16]	7,569 (100)	338 (167)							
R03-07	2018-09-18		15.9		GII.4[P16]	GII.4[P16]	7,569 (100)	103 (44)							
R03-08	2018-09-28		17.3		GII.4[P16]	GII.4[P16]	7,568 (100)	116 (49)							
R03-09	2018-10-03		20.9		GII.4[P16]	GII.4[P16]	7,529 (100)	222 (76)							
R03-10	2018-10-05		17.2		GII.4[P16]	GII.4[P16]	7,569 (100)	224 (91)							
R03-11	2018-08-16		18.7		GII.3[P16]	GII.3[P16]	7,573 (100)	1,093 (628)							
R03-12	2018-10-19		18.1		GII.4[P16]	GII.4[P16]	7,543 (100)	31 (15)							
R03-13	2018-10-22		15.8		GII.4[P16]	GII.4[P16]	7,584 (99)	417 (156)							
R03-14	2018-11-10		20.8		GII.4[P16]	GII.4[P16]	7,569 (100)	1,574 (767)							
R03-15	2018-11-13		16.6		GII.4[P16]	GII.4[P16]	7,568 (100)	1,070 (336)							
R03-16	2018-11-20		20.6		GII.4[P16]	GII.4[P16]	7,529 (100)	50,155 (12,983)							
R03-17	2018-11-23		15.8		GII.4[P16]	GII.4[P16]	7,569 (100)	11 (7)							
R03-18	2018-11-29		18.5		GII.4[P16]	GII.4[P16]	7,569 (100)	4,126 (723)							
R03-20	2018-12-03		16.9		GII.4[P16]	GII.4[P16]	7,569 (100)	150 (54)							
R03-21	2018-12-07	4 (0)	16.1		GII.4[P16]	GII.4[P16]	7,563 (100)	313 (129)							
R03-22	2018-12-07	4 (0)	25.3		GII.4[P16]	GII.4[P16]	7,555 (95)	37 (13)							
R03-23	2018-12-10		18.0		GII.3[P21]	GII.3[P21]	7,546 (100)	10,467 (2,942)							
R03-24	2018-12-02	5 (1)	23.5		GII.3[P16]	GII.3[P16]	7,552 (100)	45 (16)							
R04-01	2018-09-11		13.9		GII.3[P16]	GII.3[P16]	7,569 (99)	531 (186)							
R04-02	2018-12-07		13.6		GII.3[P12]	GII.3[P12]	7,566 (100)	77 (22)							
R04-03	2018-11-03		14.6		GII.4[P16]	GII.4[P16]	7,567 (99)	1,654 (705)							
R04-04	2019-01-03		20.5		GII.4[P16]	GII.4[P16]	7,569 (100)	2,000 (879)							
R04-05	2018-08-30		22.3		GII.4[P16]	GII.4[P16]	7,569 (100)	3 (2)							
R04-06	2018-09-04		23.3		GII.3[P21]	GII.3[P21]	7,543 (99)	8,353 (2742)							
R04-07	2017-12-04		24.8		GII.3[P21]	GII.3[P21]	7,553 (100)	5,926 (1,419)							
R04-08	2018-10-31	6 (2)	22.8		GII.3[P21]	GII.3[P21]	7,553 (100)	13 (6)							
R04-09	2018-10-31	6 (2)	19.7		GII.4[P16]	GII.4[P16]	7,578 (44)								
R04-10	2018-11-30		24.8		GII.4[P16]	GII.4[P16]	7,563 (100)								
R04-11	2018-12-05		20.5		GII.4[P16]	GII.4[P16]	7,563 (100)								
R04-12	2018-12-10		18.8		GII.4[P16]	GII.4[P16]	7,563 (100)								
R04-13	2018-12-10		23.3		GII.3[P21]	GII.3[P21]	7,544 (96)								

(Continued on next page)

**TABLE 1 (Continued)**

Sample	Collection date (yr-mo-day)	Cluster no. (SNP distance)	NoV <sub>C1</sub> value <sup>b</sup>		Sanger NoV genotype <sup>c</sup>	WGS NoV genotype <sup>c</sup>	WGS major NoV strain		WGS minor NoV strain		Off-target GE virus C <sub>1</sub> value <sup>d,e,f</sup>			WGS off-target GE virus length (bp) (%)		
			GI	GII			Length [bp (%)]	Avg. depth [x(SD)]	Length [bp (%)]	depth [x(SD)]	RVA	AdV	SaV	EV	HPeV	CV-A4 (7,437 [1,000])
R04-14	2018-12-21		18.9		GII.4[P16]	GII.4[P16]	7,570 (1,000)	43,377 (12,610)			33.0	25.9	33.6			HBoV1 (5,491 [99]); HBoV3 (5,432 [1,000]) CV-A4 (7,437 [1,000]) AIV-1 (8,252 [1,000])
R04-15	2018-12-12		22.1		G1.5[P4]	G1.5[P4]	7,700 (1,000)	148 (45)	7,688 (99); GI.6[P6]; 19 (22)							
R04-16	2018-09-27		29.3													
R04-17	2018-10-09		25.1		GII.3[P21]	GII.3[P21]	7,536 (99)	42 (17)								
R04-18	2018-11-29		21.7		GII.4[P16]	GII.4[P16]	7,565 (1,000)	343 (87)	27.4							CV-A2 (7,357 [1,000]) CV-A4 (7,427 [1,000])
R04-20	2018-12-13		29.0													
R04-21	2019-03-03		24.3													
R04-22	2018-12-05	5 (1)	19.4			GII.3[P21]	7,549 (1,000)	212 (120)								

<sup>a</sup>RT-qPCR and Sanger sequencing results are shown alongside metagenomic results for NoV and several off-target relevant viruses.

<sup>b</sup>Blank cells represent samples that tested negative.

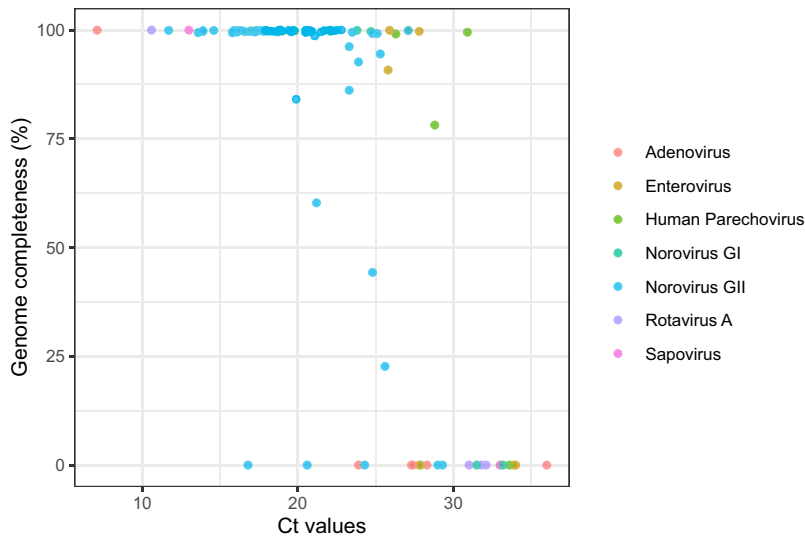
<sup>c</sup>All NoV GI.4 (ORF2/capsid) genotyped sequences were from the Sydney 2012 variant, and all NoV P4 (ORF1/RdRp) genotyped sequences were from the New Orleans 2009 variant.

<sup>d</sup>RVA, Rotavirus A; AdV, adenovirus; SaV, sapovirus; EV, enterovirus; HPeV, human parechovirus; CV, coxsackievirus; HBoV, human bocaparvovirus; AIV, aichivirus.

<sup>e</sup>Not tested due to sample depletion.

<sup>f</sup>All 11 segments (genotype, length, BoC at a depth of coverage of  $\geq 3$ ): segment 1 (R1, 3,302, 100%), segment 2 (C1, 2,729, 100%), segment 3 (M1, 2,591, 100%), segment 4 (P8, 2,359, 100%), segment 5 (A1, 1,566, 100%), segment 6 (I1, 1,356, 100%), segment 7 (T1, 1,032, 100%), segment 8 (N1, 1,059, 100%), segment 9 (G9, 1,061, 100%), segment 10 (E1, 750, 100%), segment 11 (H1, 664, 100%).

<sup>g</sup>Contained several *Enterovirus A* contigs that could not be reconstituted into a single taxon.



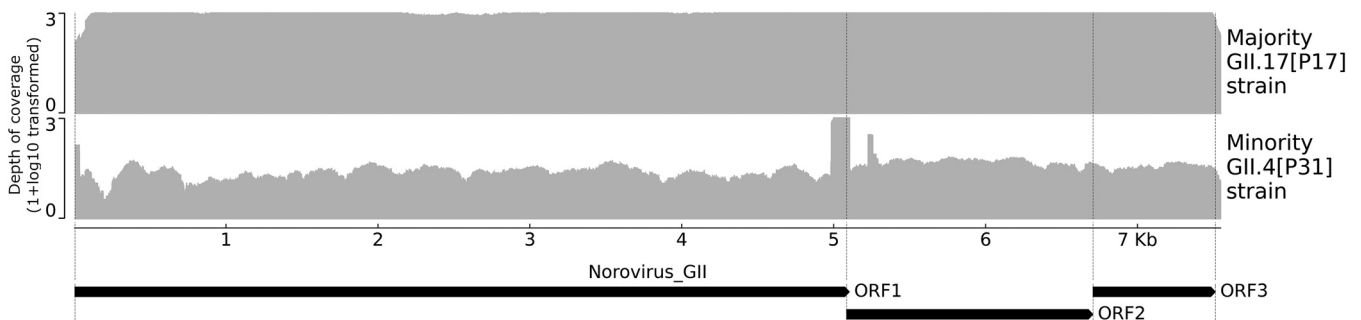
**FIG 2** Percentage of genome completeness of norovirus and off-target GE viruses versus  $C_T$ -values. Genome completeness was determined for a minimum  $3\times$  depth of coverage.

**Metagenomic sequencing of norovirus has higher sensitivity than Sanger sequencing.**

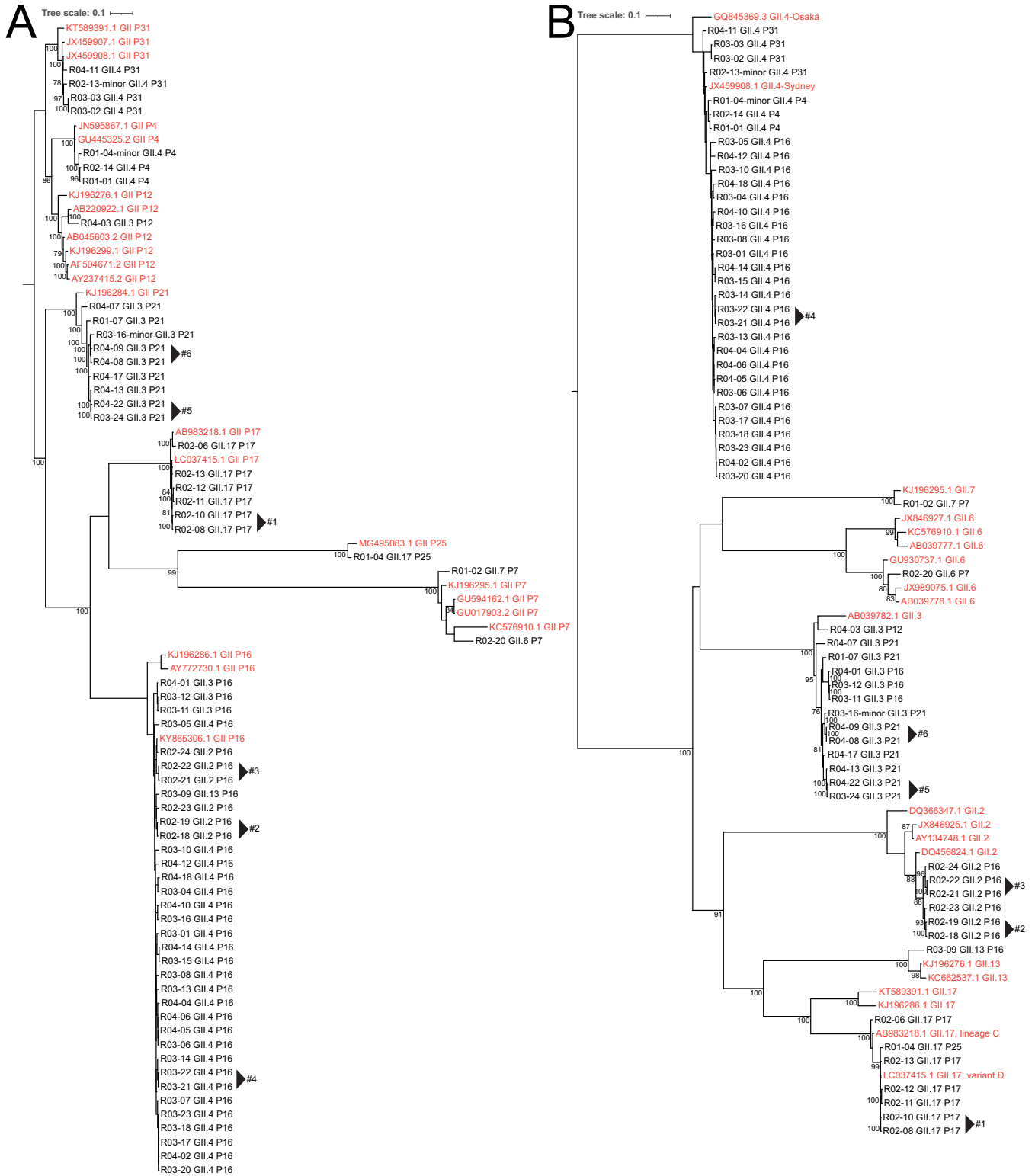
For 60 out of 71 samples, a Sanger norovirus sequence could be retrieved, compared to 66 for metagenomic sequences. To investigate the concordance and resolution of NGS versus Sanger sequencing, sequences generated by both methods were compared using a pairwise single nucleotide polymorphism (SNP) distance method with a pairwise deletion option that removes ambiguous positions. In total, 50 high-quality Sanger sequences were compared to NGS sequences. SNP differences were identified for six samples: R03-06, R03-07, R03-12, R03-13, and R04-07 had 1 SNP difference, while R03-09 had 2 SNP differences over a 969-nucleotide (nt) region (see Fig. S5 in the supplemental material). Although ambiguous positions were ignored by this method, NGS could clearly resolve ambiguous nucleotides that arose due to overlapping fluorescent peaks in the Sanger trace files.

**Phylogenetic analysis of norovirus strains.** To investigate the genetic diversity of the obtained norovirus sequences, we inferred maximum likelihood trees of the ORF1 and ORF2 sequences for norovirus GII (Fig. 4) and GI (Fig. 5). We included references from the latest nomenclature update described in Chhabra et al. (11). For samples that were linked in location and time (sampled within 1 week of each other), we investigated if there were differences in single nucleotide polymorphisms (SNPs). We identified six linked sample pairs, that had up to two SNP distances between them over the entire combined ORF1 to ORF3 coding domain sequence (CDS) (Table 1; Fig. 4).

GII.17[P25] strains are not often detected; therefore, we further investigated the origin of ORF2 and ORF1. For norovirus GII.17 strains, there are several lineages (A to D), of which

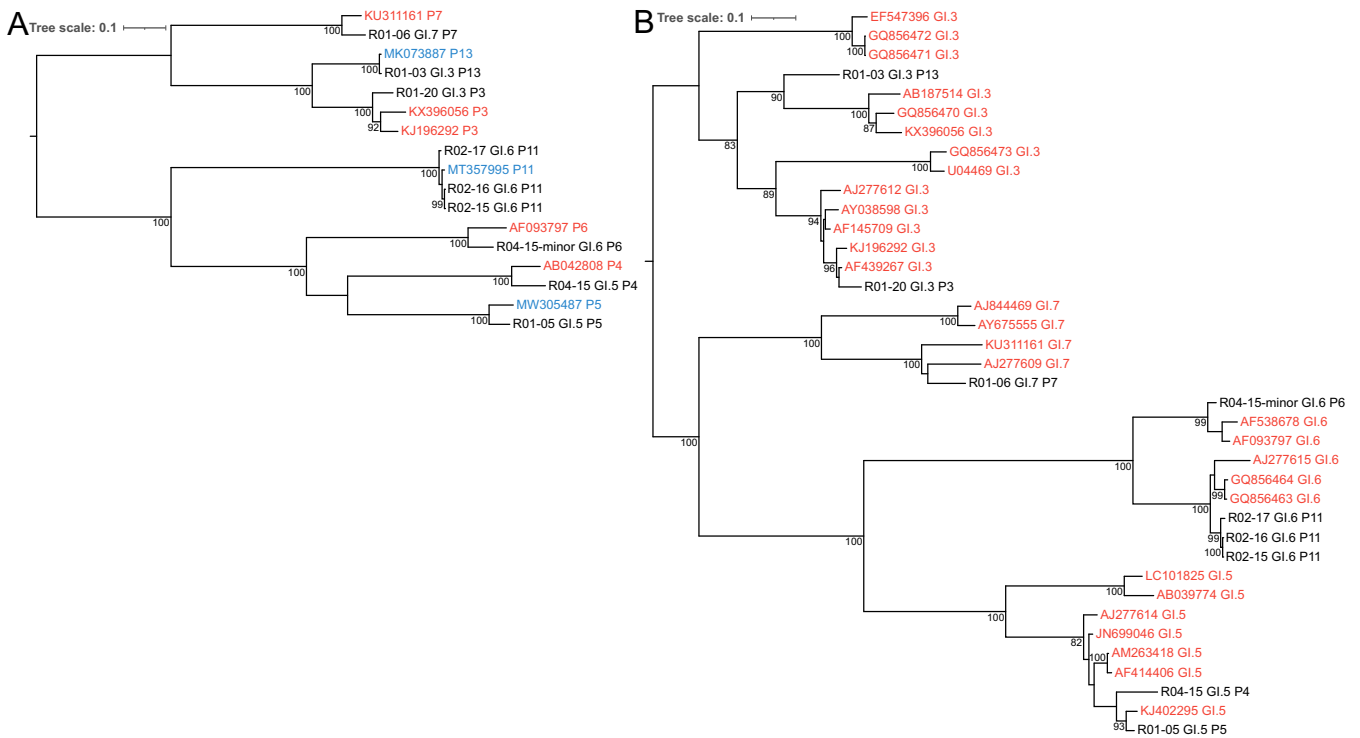


**FIG 3** Example depth of coverage plot of norovirus (NoV) GII strains in sample R02-13. Two strains were present in this sample, a GII.17[P17] norovirus strain (top) with high depth of coverage (DoC) and a GII.4[P31] norovirus strain (bottom) with lower DoC, denoted “major” and “minor,” respectively. The characteristic block-like high coverage peaks near the ORF1/2 junction of the minority strain are due to homology between the strains in these regions. The aligned reads of the majority strain adversely affect the consensus sequence of the minority strain, necessitating manual correction. A full overview of all norovirus GI, GII, and off-target viruses with public health relevance is shown in Fig. S1 to S4. The coverage is shown as a  $1 + \log_{10}$  transformation with a maximum  $999\times$  depth of coverage.



**FIG 4** Maximum likelihood tree of 61 GII strains. (A) Maximum likelihood trees were inferred for ORF1 based on 5,166 nucleotides using the GTR best-fit model. (B) ORF2 based on 1,689 nucleotides using the TIM2 best-fit model. Reference strains from Chhabra et al. (11) are shown in red; clusters are annotated with numbered arrows (Table 1). The scale bar represents nucleotide substitutions per site, and selected bootstrap values greater than 70 are shown. For samples containing multiple strains, the lowest depth of coverage strain is denoted with “minor” in the tip label (e.g., “R01-04-minor”). All GII.4 and GII.P4 strains were Sydney and New Orleans variants, respectively.





**FIG 5** Maximum likelihood tree of 9 GI strains. (A and B) Maximum likelihood trees were inferred for (A) ORF1 based on 5,412 nucleotides using the TIM2 best-fit model and (B) ORF2 based on 1,665 nucleotides using the TIM2 best-fit model. Genotype reference strains from Chhabra et al. (11) are shown in red, while additional references from NCBI GenBank are shown in blue. The scale bar represents nucleotide substitutions per site, and selected bootstrap values greater than 70 are shown. In sample R04-15, two strains were detected, with the strain of lower coverage denoted as "minor" in the tip label.

variants C (GenBank accession number [AB983218](#)) and D (accession number [LC037415](#)) are contemporary strains that recently caused large outbreaks in Asia (22, 23). All GII.17 strains in this study belonged to variant D, with the exception of R02-06, which belonged to variant C (Fig. 4). The GII.17[P25] strain (R01-04) belonged to GII. 17 variant D (Fig. S6B), and it had a single amino acid change from the reference strain, which suggests a recent recombination event. Additionally, its polymerase region clustered with GII.P25 (Fig. S6A). This strain represents the second complete GII.P25 ORF1 sequence alongside [MG495083](#).

**Children were frequently coinfecting with another GE virus.** From the samples that tested positive for norovirus by RT-qPCR, 18% yielded off-target GE viruses, resulting in 13 full and two partial genomes (Table 1; Fig. 2 and 5). Children were more often coinfecting with another GE virus (39%) than adults (6.7%; Fig. 1). Aichivirus 1 (AiV-1), RVA, and Adv41 were detected in individuals of 60, 66, and 71 years old, respectively, while all other off-target GE viruses were detected in samples from individuals younger than 5 years (Fig. 1). All detected EV were coxsackieviruses (CV). For several of the GE viruses, the number of sequences generated in this study provided a marked increase in the number of sequences uploaded to NCBI between 2017 and 2022 (Table S1).

To investigate the sensitivity of metagenomic sequencing for the off-target GE viruses, (RT)-qPCRs were performed for RVA, Adv, sapovirus (SaV), EV, and human parechovirus (HPeV) on 71 samples. For AiV and human bocaparvovirus (HBoV), (RT)-qPCR assays were not available (Table 1). The sample size was limited, but overall, no sequences were obtained with a  $C_T$  value of  $>28$  except for two HPeV genomes, while RT-qPCR detected several off-target viruses with a  $C_T$  of  $>28$ , which were not detected with metagenomics, metagenomics allowed for the reliable detection and genotyping of off-target viruses in a single assay.

**Metagenomics identified a primer/probe mismatch for SaV GI.7.** An SaV GI.7 was initially identified by metagenomics but not by RT-qPCR. Based on the obtained SaV GI.7 sequence, the SaV RT-qPCR was updated (Table 1). The forward primer remained unchanged (5'-GAYCWGGCYCTGCCACCT-3'), the probe was truncated from 5'-TGY ACCACTATRAACCAVGG-3' to 5'-TGYACCACCTATRAACCA-3', and the reverse primer

was truncated at the 3' end and extended on the 5' end: 5'-GCCCTCCATYTCAAACACTAWTTTG-3' to 5'-CATTGCCCTCCATYTCAAACACTA-3'. This updated primer set allowed for the retrospective detection of SaV with a  $C_T$  of 13.

## DISCUSSION

As previously reported by others and shown in this study (24), Illumina-based metagenomic sequencing is a robust method for generating norovirus sequences. We were also able to determine norovirus genomes from minority strains in double-infection samples in which Sanger sequencing could not differentiate the minor from the major strain.

Although this study did not primarily investigate outbreaks, we identified several pairs of samples that were collected from the same location and time. These pairs had a pairwise SNP distance of  $\leq 2$  across the combined ORF1 to ORF3 CDS. However, to identify norovirus clusters, further investigation of outbreaks with well-characterized transmission chains is necessary to determine an appropriate SNP cutoff.

When comparing genome completeness against qPCR  $C_T$  values, a proxy for the viral load, an overall lower genome completeness at higher  $C_T$  values was observed, as reported by others (25). The genome ends had a lower coverage depth, but the reads were evenly distributed across the rest of the genome (26–29). For three samples with low  $C_T$  value, no norovirus sequences could be obtained, and we hypothesize that this was caused by other abundant taxa outcompeting norovirus nucleic acids for sequencing space or the presence of inhibitors negatively affecting amplification efficiency (30). This limitation highlights one of the challenges of using agnostic, metagenomics-based virus typing.

In assessing the off-target coinfections, CV-A16 (B1a), CV-A2, CV-A4, and CV-A5 genomes were identified in infants. CV is the causative agent of hand, foot, and mouth disease (HFMD), which sporadically causes neurological symptoms, is linked to intrauterine fetal demise (31–33), and has recently caused large HFMD outbreaks in China (34–38). HPeV 1 and 3 were also identified in infants and can result in GE, and sporadically in neurological symptoms and death, in young children (39–43).

Moreover, an RVA coinfection, which causes severe GE (2), was identified in an elderly individual, and all 11 segments were genotyped. Importantly, this study was performed in The Netherlands, where RVA will be included in the National vaccination program in 2024. Metagenomic GE surveillance can help assess the effect of vaccination on the epidemiology of RVA in The Netherlands. The other off-target viruses that were detected were an AiV1, which is linked to predominantly subclinical GE (44, 45), AdV41, a known cause of GE (2), an HBoV1, and an HBoV3, the former of which is a respiratory virus, while the latter is a GE-causing virus (17, 19, 46). Lastly, an SaV GI.7 was identified in an infant, which belongs to the same viral family as norovirus with similar clinical and epidemiological characteristics (47, 48). These off-target findings provide a reference set to assess silent transmission and to facilitate genotype to phenotype studies to determine if genetic or genotype-specific factors contribute to severe pathogenesis. However, it should be noted that studies have shown that 7 to 15% of norovirus infections remains asymptomatic, meaning that some of these off-target viruses could be the causative agent of the reported gastroenteritis (49).

Interestingly, the SaV GI.7-positive sample that initially tested negative with RT-qPCR was identified by metagenomics. Evaluation of the primers and probe showed that these were not compatible with this genotype. Therefore, they were updated, and now they are similar to the frequently used primer and probe sets of Oka et al. (50). Since Okada et al. first detected SaV GI.7 in 2006, only a few GI.7 sequences have been submitted to NCBI, and the first complete genome was sequenced in 2018 (51, 52). The SaV GI.7 complete genome generated in this study most closely resembles this complete genome (GenBank accession number [AB522390](https://www.ncbi.nlm.nih.gov/nucl/AB522390)) with only 89% nucleotide identity, while SaV GI is shown to have low intragenotype diversity (51, 53). A recent large-scale review of SaV prevalence (48) did not report the GI.7 genotype in surveillance programs. Therefore, our results raise questions about where this genotype circulated without being detected in surveillance programs.

Even though this study had a limited sample size, it generated a marked addition of contemporary and publicly available full genomes. Compared to full-genome sequences

uploaded to NCBI over a 4-year period, the genomes generated in this study represent up to a 33% increase in contemporary sequences for some of these off-target viruses (Table S1). This can help contextualize molecular epidemiology and outbreak tracing.

While Illumina metagenomic NGS is a more costly technique than RT-qPCR or Sanger sequencing, it has broad surveillance potential. Conventional methods require continuous updates of primers and probe sets to account for antigenic drift and novel genotypes. Without metagenomic-based surveillance, the SaV GI.7 in this study would not have been identified, and similar results are reported by others (54). Primer mismatches can also occur for viruses of chronically infected patients, where prolonged replication can result in mutations in the otherwise conserved domains targeted by Sanger sequencing assays.

In-depth analysis of NGS data to identify GE viruses in a public health setting is challenging and time-consuming. Negative controls and detection thresholds have to be optimized to account for lab-specific contamination levels to mitigate false-positives. Likewise, a mixed infection with shared homologous regions requires careful manual curation. Here, it helps to have standardized pipelines, with parameter settings tailored to the local laboratory methods and visualization of genomic regions for manual in-depth inspection.

In conclusion, this study shows the potential of NGS-based norovirus surveillance for generating a full-genome reference set for a broad range of public health-relevant pathogens and for high-resolution outbreak detection. By describing the caveats and strengths of metagenomic-based surveillance, and by sharing these complete genome sequences, we hope to help other researchers contextualize their outbreak investigations and improve molecular epidemiology.

## MATERIALS AND METHODS

**Sample selection.** A total of 71 norovirus RT-qPCR positive fecal samples obtained between 2015 and 2019 with a  $C_T$  of  $<30$  for at least one norovirus genogroup were selected from GE outbreak-related samples sent to the Dutch National Institute for Public Health and the Environment (RIVM) norovirus surveillance program by medical microbiology laboratories for referral. The samples originated predominantly from young children ( $<5$  years old,  $N = 23$ ) and elderly ( $>65$  years old  $N = 38$ ). For one sample, the patient age was not reported.  $C_T$  values ranged from 17.0 to 33.2 ( $N = 10$ ) and 11.7 to 29.3 ( $N = 63$ ) for GI and GII, respectively. Feces samples were stored at 4°C until processing, which is the standard procedure at the RIVM for norovirus-positive feces.

**qPCR and Sanger sequencing.** Feces sample was suspended in 1,000  $\mu\text{L}$  modified Eagle medium (MEM) medium containing gentamicin (Thermo Fisher, Bleiswijk, The Netherlands) to give a 10 to 20% vol/vol suspension and centrifuged at 16,100 relative centrifugal force (RCF) for 1 min at room temperature. Total nucleic acid (TNA) was extracted using the MagNA Pure 96 system (Roche, Almere, The Netherlands) and eluted in 50  $\mu\text{L}$  elution buffer. All samples were subjected to a qPCR panel targeting norovirus GI, norovirus GII, RVA, sapovirus (SaV), and adenovirus (AdV). Norovirus Sanger sequences were generated as described previously (3, 15). Sanger sequences were analyzed using BioNumerics (AppliedMaths, Sint-Martens-Latem, Belgium) and genotyped via the NoroNet norovirus typing-tool (16). Additional viruses detected by NGS in this study, which are of relevance for public health, were confirmed by RT-qPCR: EV and human parechovirus (HPEV) via RT-qPCR as described by Benschop et al. (55, 56) and SaV via RT-qPCR with modified primers (as described above).

**Virus enrichment, NGS pretreatment, and NGS.** Feces was suspended in 1,000  $\mu\text{L}$  Eagle MEM containing gentamicin (Thermo Fisher, Bleiswijk, The Netherlands) to give a 10 to 20% vol/vol suspension and centrifuged at 16,100 RCF for 1 min at room temperature. The supernatant was filtered using Costar Spin-X 0.45- $\mu\text{m}$  CA membrane centrifuge tube filters (Corning, Amsterdam, The Netherlands). Afterward, 200  $\mu\text{L}$  filtrate was supplemented with 25  $\mu\text{L}$  25 mM  $\text{MgCl}_2$  and treated with 1.25  $\mu\text{L}$  200 U/ $\mu\text{L}$  OmniCleave endonuclease (EpiCentre, Leusden, The Netherlands) for 1 h at 37°C.

Total nucleic acid (TNA) was extracted using the MagNA Pure 96 system (Roche, Almere, The Netherlands) and eluted in 50  $\mu\text{L}$  elution buffer. First-strand cDNA synthesis was performed using SuperScript III (Invitrogen, Bleiswijk, The Netherlands) using 11  $\mu\text{L}$  TNA eluate as input. The second cDNA strand was synthesized using the NEBNext mRNA second-strand synthesis module (New England Biolabs, Leusden, The Netherlands) using 20  $\mu\text{L}$  as input, as per the manufacturer's instructions, resulting in 160  $\mu\text{L}$  output. Next, the sample was purified using DNA Clean & Concentrator-5 capped columns (Zymo Research, Leiden, The Netherlands), as per the manufacturer's instructions, resulting in 15  $\mu\text{L}$  double-stranded DNA (dsDNA) per sample.

Library preparation was performed with the Nextera XT DNA library prep kit (Illumina, Eindhoven, The Netherlands). Samples were processed in four independent NGS runs, denoted with the prefixes R01 to R04. Prefixes R02, R03, and R04 were 150-nucleotide paired-end sequenced on an Illumina NextSeq instrument per the manufacturer's specifications, and samples starting with R01 were 300-nucleotide paired-end sequenced on an Illumina MiSeq instrument by BaseClear B.V. (Leiden, The Netherlands). On average, 4.7 million reads (postfiltering) were generated per sample.

**Data analysis.** All NGS data were analyzed using an in-house NGS analysis workflow called Jovian (v1.01, <https://github.com/DennisSchmitz/Jovian>) using default settings. Briefly, this workflow removes human and poor-quality reads (i.e., nucleotides with a Phred score of  $<20$  are trimmed, and reads with a length of

<50 nt are discarded), assembles reads into contigs and annotates contigs of  $\geq 250$  nt via megaBLAST against the NCBI NT database, determines the lowest common ancestor (LCA) of the BLAST results, and finally, performs genotyping of several clinically important virus families and genera via their respective typing tools (16).

Norovirus sequences were manually curated and subsequently scaffolded and assembled based on same-genotype public reference sequences or by extending the 5' and 3' ends using soft-clipped overhangs. This manually assembled draft genome was then assessed and corrected by performing Minimap2 (57) alignment and LoFreq (58) SNP-calling, with manual curation of the genome ends, to produce a final sequence. Homologous regions of mixed-strain samples were manually corrected and curated. Any reference nucleotide with <3 reads coverage was masked with an N in the final genome sequence. Considering a minimum Phred score of 20 and minimum depth of coverage of 3, a sequencing error in the consensus requires  $\geq 2/3$  simultaneous errors which, by binomial distribution, has a 0.03% chance. This is a lower bound since the depth of coverage generally was  $\geq 1$  order of magnitude higher than 3 (Table 1).

Virus genomes were considered complete when their entire CDS was characterized with  $\leq 100$  N's, and partial when it contained >100 N's or could not be scaffolded to a full CDS sequence. If fewer than 50 total reads aligned against same-taxon scaffolds with a length  $\geq 250$  nt, they were considered negative. Norovirus scaffolds with a different ORF1 and ORF2 genotype than their majority strain were considered minority strains.

Norovirus sequences were aligned with reference strains from Chhabra et al. (11) using MUSCLE v3 (59). Maximum likelihood trees were inferred with IQ-TREE 2 (60), using 100 Felsenstein bootstraps (61) with best-fit models as identified by ModelFinder (62) and were edited with iTOL v5 (63). Results were visualized via ggplot2 and deeptools (64, 65).

**Data availability.** All filtered, General Data Protection Regulation-compliant, FASTQ files and full genomes were submitted to NCBI SRA under study accession number PRJEB54724 (accession numbers OP162334 to OP162342, OP205527 to OP205585 and OP255971 to OP255995). Sequences that contained too many N's could not be uploaded to NCBI but are available on request. All complete and partial norovirus genomes were also uploaded to NoroNet (<https://www.rivm.nl/en/noronet>).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.5 MB.

## ACKNOWLEDGMENTS

We are very grateful for the assistance of Thijs Bosch, Dirk Eggink, Jeroen Cremer, Anne-Marie van den Brandt, Ryanne Jaarsma, Elsa Poorter, Kim Freriks, Robert Verhagen, Erwin van Wieringen, and Mark Kroon, as well as the regional laboratories for their participation in the national norovirus surveillance program. This study was financed under the European Union's Horizon H2020 grant VEO (grant number 874735) and the NWO Stevin prize (Koopmans).

## REFERENCES

- Simpson R, Aliyu S, Iturriza-Gómara M, Desselberger U, Gray J. 2003. Infantile viral gastroenteritis: on the way to closing the diagnostic gap. *J Med Virol* 70:258–262. <https://doi.org/10.1002/jmv.10386>.
- Bányai K, Estes MK, Martella V, Parashar UD. 2018. Viral gastroenteritis. *Lancet* 392:175–186. [https://doi.org/10.1016/S0140-6736\(18\)31128-0](https://doi.org/10.1016/S0140-6736(18)31128-0).
- Svraka S, Duizer E, Vennema H, de Bruin E, van der Veer B, Dorresteyn B, Koopmans M. 2007. Etiological role of viruses in outbreaks of acute gastroenteritis in The Netherlands from 1994 through 2005. *J Clin Microbiol* 45:1389–1394. <https://doi.org/10.1128/JCM.02305-06>.
- Kambhampati A, Koopmans M, Lopman BA. 2015. Burden of norovirus in healthcare facilities and strategies for outbreak control. *J Hospital Infection* 89:296–301. <https://doi.org/10.1016/j.jhin.2015.01.011>.
- Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, de Silva NR, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleeschauwer B, Adegoke GO, Afshari R, Alasfoor D, Baines J, Balakrishnan K, Hamza W b, Black RE, Bolger PM, Chaicumpa W, Cravioto A, Döpfer D, Ehiri JE, Fazil A, Ferreccio C, Fèvre EM, Hall G, Kasuga F, Keddy KH, Lanata CF, Lei H, Liu X, Manyindo B, Nasinyama G, Ongolo-Zogo P, Pitt J, Rokni MB, Sripa B, van Leeuwen R, Verger P, Willingham AL, Zhou XN, Aspinall W, Buchanan R, Budke C, on behalf of World Health Organization Foodborne Disease Burden Epidemiology Reference Group, et al. 2015. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLoS Med* 12:e1001923. <https://doi.org/10.1371/journal.pmed.1001923>.
- de Graaf M, van Beek J, Koopmans MPG. 2016. Human norovirus transmission and evolution in a changing world. *Nat Rev Microbiol* 14:421–433. <https://doi.org/10.1038/nrmicro.2016.48>.
- Sherwood J, Mendelman PM, Lloyd E, Liu M, Boslego J, Borkowski A, Jackson A, Faix D, US Navy study team. 2020. Efficacy of an intramuscular bivalent norovirus GI.1/GII.4 virus-like particle vaccine candidate in healthy US adults. *Vaccine* 38:6442–6449. <https://doi.org/10.1016/j.vaccine.2020.07.069>.
- Tan M. 2021. Norovirus vaccines: current clinical development and challenges. *Pathogens* 10:1641. <https://doi.org/10.3390/pathogens10121641>.
- Netzer 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>.
- Nussbaum EZ, Azar MM, Cohen E, McManus D, Topal JE, Malinis M. 2020. Orally administered human immunoglobulin therapy for norovirus enteritis in solid organ transplant recipients: a case series at a single academic transplant center. *Clin Infect Dis* 71:E206–E209. <https://doi.org/10.1093/cid/ciaa093>.
- Chhabra P, de Graaf M, Parra GI, Chan MCW, 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>.
- Bull RA, Tanaka MM, White PA. 2007. Norovirus recombination. *J Gen Virol* 88:3347–3359. <https://doi.org/10.1099/vir.0.83321-0>.
- Ludwig-Begall LFL, Mauroy A, Thiry E. 2018. Norovirus recombinants: Recurrent in the field, recalcitrant in the lab: a scoping review of recombination and recombinant types of noroviruses. *J General Virology* 99:970–988. <https://doi.org/10.1099/jgv.0.001103>.
- 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>.

15. Svraka S, van der Veer B, Duizer E, Dekkers J, Koopmans M, Vennema H. 2009. Novel approach for detection of enteric viruses to enable syndrome surveillance of acute viral gastroenteritis. *J Clin Microbiol* 47:1674–1679. <https://doi.org/10.1128/JCM.00307-09>.
16. Kroneman A, Vennema H, Deforche K, Avoort H, Peñaranda S, Oberste MS, Vinjé J, Koopmans M. 2011. An automated genotyping tool for enteroviruses and noroviruses. *J Clin Virol* 51:121–125. <https://doi.org/10.1016/j.jcv.2011.03.006>.
17. Oude Munnink BB, van der Hoek L. 2016. Viruses causing gastroenteritis: the known, the new and those beyond. *Viruses* 8:42. <https://doi.org/10.3390/v8020042>.
18. Wilhelmi I, Roman E, Sánchez-Fauquier A. 2003. Viruses causing gastroenteritis. *Clin Microbiol Infect* 9:247–262. <https://doi.org/10.1046/j.1469-0691.2003.00560.x>.
19. Mietzsch M, Kailasan S, Garrison J, Ilyas M, Chipman P, Kantola K, Janssen ME, Spear J, Sousa D, McKenna R, Brown K, Söderlund-Venermo M, Baker T, Agbandje-McKenna M. 2017. Structural insights into human bocaparvoviruses. *J Virol* 91:e00261-17. <https://doi.org/10.1128/JVI.00261-17>.
20. Kumthip K, Khamrin P, Ushijima H, Maneekarn N. 2017. Multiple enterovirus genotypes circulating in children hospitalized with acute gastroenteritis in Thailand. *Infect Genet Evol* 55:324–331. <https://doi.org/10.1016/j.meegid.2017.10.002>.
21. Pham NTK, Thongprachum A, Trinh QD, Okitsu S, Komine-Aizawa S, Shimizu H, Hayakawa S, Ushijima H. 2018. Detection and genetic characterization of enterovirus strains circulating among children with acute gastroenteritis in Japan during 2014–2016. *Infect Genet Evol* 61:16–19. <https://doi.org/10.1016/j.meegid.2018.03.009>.
22. Chan MCW, Lee N, Hung TN, Kwok K, Cheung K, Tin EKY, Lai RWM, Nelson EAS, Leung TF, Chan PKS. 2015. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat Commun* 6:1:10061. <https://doi.org/10.1038/ncomms10061>.
23. Chan MCW, Roy S, Bonifacio J, Zhang LY, Chhabra P, Chan JCM, Celma C, Igoy MA, Lau SL, Mohamad KN, Vinjé J, Vennema H, Breuer J, Koopmans M, de Graaf M, for NOROPATROL2. 2021. Detection of norovirus variant GII.4 Hong Kong in Asia and Europe, 2017–2019. *Emerg Infect Dis* 27:289–293. <https://doi.org/10.3201/eid2701.203351>.
24. Petronella N, Ronholm J, Suresh M, Harlow J, Mykytczuk O, Corneau N, Bidawid S, Naseri N. 2018. Genetic characterization of norovirus GII.4 variants circulating in Canada using a metagenomic technique. *BMC Infect Dis* 18:1–11. <https://doi.org/10.1186/s12879-018-3419-8>.
25. Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, Burfin G, Scholtes C, Morfin F, Valette M, Lina B, Bal A, Josset L. 2020. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol* 6:veaa075. <https://doi.org/10.1093/ve/veaa075>.
26. Batty EM, Wong THN, Trebes A, Argoud K, Attar M, Buck D, Ip CLC, Golubchik T, Cule M, Bowden R, Manganis C, Klenerman P, Barnes E, Walker AS, Wyllie DH, Wilson DJ, Dingle KE, Peto TEA, Crook DW, Piazza P. 2013. A modified RNA-seq approach for whole genome sequencing of RNA Viruses from faecal and blood samples. *PLoS One* 8:e66129. <https://doi.org/10.1371/journal.pone.0066129>.
27. Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* 38:e131. <https://doi.org/10.1093/nar/gkq224>.
28. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628. <https://doi.org/10.1038/nmeth.1226>.
29. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349. <https://doi.org/10.1126/science.1158441>.
30. Schrader C, Schielke A, Ellerbroek L, Johne R. 2012. PCR inhibitors: occurrence, properties and removal. *J Appl Microbiol* 113:1014–1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x>.
31. Robinson CR, Doane FW, Rhodes AJ. 1958. Report of an outbreak of febrile illness with pharyngeal lesions and exanthem: Toronto, summer 1957— isolation of group A coxsackie virus. *Can Med Assoc J* 79:615.
32. Zhang Y, Wang D, Yan D, Zhu S, Liu J, Wang H, Zhao S, Yu D, Nan L, An J, Chen L, An H, Xu A, Xu W. 2010. Molecular evidence of persistent epidemic and evolution of subgenotype B1 coxsackievirus A16-associated hand, foot, and mouth disease in China. *J Clin Microbiol* 48:619–622. <https://doi.org/10.1128/JCM.02338-09>.
33. Yu W, Tellier R, Wright JR Jr. 2015. Coxsackie virus a16 infection of placenta with massive perivillous fibrin deposition leading to intrauterine fetal demise at 36 weeks gestation. *Pediatr Dev Pathol* 18:331–334. <https://doi.org/10.2350/15-01-1603-CR.1>.
34. Yu Y, Luo Z, Jin W, Mai J, Qian S, Lu J, Wei Z, Meng S, Wang Z, Guan X, Tong Y, Shen S. 2021. Emergence of a novel recombinant of CV-A5 in HFMD epidemics in Xiangyang, China. *BMC Med Genomics* 14:279. <https://doi.org/10.1186/s12920-021-01107-6>.
35. Ji T, Guo Y, Lv L, Wang J, Shi Y, Yu Q, Zhang F, Tong W, Ma J, Zeng H, Zhao H, Zhang Y, Han T, Song Y, Yan D, Yang Q, Zhu S, Zhang Y, Xu W. 2019. Emerging recombination of the C2 sub-genotype of HFMD-associated CV-A4 is persistently and extensively circulating in China. *Sci Rep* 9:13668. <https://doi.org/10.1038/s41598-019-49859-7>.
36. Hu YF, Yang F, Du J, Dong J, Zhang T, Wu ZQ, Xue Y, Jin Q. 2011. Complete genome analysis of coxsackievirus A2, A4, A5, and A10 strains isolated from hand, foot, and mouth disease patients in China revealing frequent recombination of human enterovirus A. *J Clin Microbiol* 49:2426–2434. <https://doi.org/10.1128/JCM.00007-11>.
37. Nagai T, Hanaoka N, Katano H, Konagaya M, Tanaka-Taya K, Shimizu H, Mukai T, Fujimoto T. 2021. A fatal case of acute encephalopathy in a child due to coxsackievirus A2 infection: a case report. *BMC Infect Dis* 21:1167. <https://doi.org/10.1186/s12879-021-06858-2>.
38. Yang Q, Zhang Y, Yan D, Zhu S, Wang D, Ji T, Li X, Song Y, Gu X, Xu W. 2016. Two genotypes of coxsackievirus A2 associated with hand, foot, and mouth disease circulating in China since 2008. *PLoS One* 11:e0169021. <https://doi.org/10.1371/journal.pone.0169021>.
39. Boivin G, Abed Y, Boucher FD. 2005. Human parechovirus 3 and neonatal infections. *Emerg Infect Dis* 11:103–105. <https://doi.org/10.3201/eid1101.040606>.
40. Benschop KSM, Schinkel J, Minnaar RP, Pakrj D, Spanjerberg L, Kraakman HC, Berkhout B, Zaaijer HL, Beld MGHM, Wolthers KC. 2006. Human parechovirus infections in Dutch children and the association between serotype and disease severity. *Clin Infect Dis* 42:204–210. <https://doi.org/10.1086/498905>.
41. Harvala H, Robertson I, Chieochansin T, McWilliam Leitch EC, Templeton K, Simmonds P. 2009. Specific association of human parechovirus type 3 with sepsis and fever in young infants, as identified by direct typing of cerebrospinal fluid samples. *J Infect Dis* 199:1753–1760. <https://doi.org/10.1086/599094>.
42. Koskiniemi M, Paetau R, Linnavuori K. 1989. Severe encephalitis associated with disseminated echovirus 22 infection. *Scand J Infect Dis* 21:463–466. <https://doi.org/10.3109/0036548909167453>.
43. Figueroa JP, Ashley D, King D, Hull B. 1989. An Outbreak of acute flaccid paralysis in Jamaica associated with echovirus type 22. *J Med Virol* 29:315–319. <https://doi.org/10.1002/jmv.1890290418>.
44. Yip CCY, Lo KL, Que TL, Lee RA, Chan KH, Yuen KY, Woo PCY, Lau SKP. 2014. Epidemiology of human parechovirus, Aichi virus and salivirus in fecal samples from hospitalized children with gastroenteritis in Hong Kong. *Virol J* 11:182. <https://doi.org/10.1186/1743-422X-11-182>.
45. Rivadulla E, Romalde JL. 2020. A comprehensive review on human aichi virus. *Virol Sin* 35:501–516. <https://doi.org/10.1007/s12250-020-00222-5>.
46. Shao L, Shen W, Wang S, Qiu J. 2021. Recent advances in molecular biology of human bocavirus 1 and its applications. *Front Microbiol* 12:696604. <https://doi.org/10.3389/fmicb.2021.696604>.
47. Oka T, Wang Q, Katayama K, Saif LJ. 2015. Comprehensive review of human sapoviruses. *Clin Microbiol Rev* 28:32–53. <https://doi.org/10.1128/CMR.00011-14>.
48. Becker-Dreps S, Gonzalez F, Bucardo F. 2020. Sapovirus: an emerging cause of childhood diarrhea. *Curr Opin Infect Dis* 33:388–397. <https://doi.org/10.1097/QCO.0000000000000671>.
49. Qi R, Huang Y, Liu J, Sun Y, Sun X, Han H-J, Qin X-R, Zhao M, Wang L, Li W, Li J, Chen C, Yu X-J. 2018. Global prevalence of asymptomatic norovirus infection: a meta-analysis. *EclinicalMedicine* 2–3:50–58. <https://doi.org/10.1016/j.eclinm.2018.09.001>.
50. Oka T, Katayama K, Hansman GS, Kageyama T, Ogawa S, Wu FT, White PA, Takeda N. 2006. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. *J Med Virol* 78:1347–1353. <https://doi.org/10.1002/jmv.20699>.
51. Oka T, Iritani N, Okada M, Ogawa T, Izuka S, Tatsumi C, Harada S, Haga K, Doan YH. 2018. First complete genome sequences of human sapovirus strains classified as GI.3, GI.4, GI.6, GI.7, and GI.7. *Genome Announc* 6:e00168-18. <https://doi.org/10.1128/genomeA.00168-18>.
52. Okada M, Yamashita Y, Oseto M, Shinozaki K. 2006. The detection of human sapoviruses with universal and genogroup-specific primers. *Arch Virol* 151:2503–2509. <https://doi.org/10.1007/s00705-006-0820-1>.
53. Tohma K, Kulka M, Coughlan S, Green KY, Parra GI. 2020. Genomic analyses of human sapoviruses detected over a 40-year period reveal disparate patterns of evolution among genotypes and genome regions. *Viruses* 12:516. <https://doi.org/10.3390/v12050516>.

54. Zhou Y, Qiu Q, Luo K, Liao Q, Li Y, Cui P, Liang L, Cheng Y, Wang L, Wang K, Van Tan L, Van Doorn HR, Yu H. 2020. Molecular strategy for the direct detection and identification of human enteroviruses in clinical specimens associated with hand, foot and mouth disease. *PLoS One* 15:e0241614. <https://doi.org/10.1371/journal.pone.0241614>.
55. Benschop K, Molenkamp R, van der Ham A, Wolthers K, Beld M. 2008. Rapid detection of human parechoviruses in clinical samples by real-time PCR. *J Clin Virol* 41:69–74. <https://doi.org/10.1016/j.jcv.2007.10.004>.
56. Hayes A, Nguyen D, Andersson M, Antón A, Bailly JL, Beard S, Benschop KSM, Berginc N, Blomqvist S, Cunningham E, Davis D, Dembinski JL, Diedrich S, Dudman SG, Dyrdak R, Eltringham GJA, Gonzales-Goggia S, Gunson R, Howson-Wells HC, Jääskeläinen AJ, López-Labrador FX, Maier M, Majumdar M, Midgley S, Mirand A, Morley U, Nordbø SA, Oikarinen S, Osman H, Papa A, Pellegrinelli L, Piralla A, Rabella N, Richter J, Smith M, Söderlund Strand A, Templeton K, Vipond B, Vuorinen T, Williams C, Wollants E, Zakikhany K, Fischer TK, Harvala H, Simmonds P. 2020. A European multicentre evaluation of detection and typing methods for human enteroviruses and parechoviruses using RNA transcripts. *J Med Virol* 92:1065–1074. <https://doi.org/10.1002/jmv.25659>.
57. Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
58. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 40:11189–11201. <https://doi.org/10.1093/nar/gks918>.
59. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
60. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
61. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>.
62. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
63. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>.
64. Wickham H. 2016. ggplot2: elegant graphics for data analysis. <https://ggplot2.tidyverse.org>.
65. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9:189. <https://doi.org/10.1038/s41467-017-02525-w>.