npj | science of food

Article

Published in partnership with Beijing Technology and Business University & International Union of Food Science and Technology

https://doi.org/10.1038/s41538-025-00659-6

# Whole genome sequencing of hepatitis A virus: adapting Illumina protocols for foodborne investigation



**Daseul Yeo[1], Soontag Jung[1], Seongwon Hwang[1], Danbi Yoon[1], Dong Jae Lim[1], Songfeng Jin[1], Jinho Choi[2], Ki Ho Hong[3] & Changsun Choi[1]** ✉

High-throughput sequencing-based whole-genome sequencing (WGS) is highly effective for identifying viral pathogens in microbial research. However, applying WGS directly to foodborne viruses remains challenging because food matrices contain PCR inhibitors and viral titers are typically much lower than those found in clinical specimens. This study aimed to develop a WGS method for analyzing the hepatitis A virus (HAV) genome in clams using the Illumina MiSeq platform. To enhance the HAV WGS method, we applied four approaches to HAV-positive clam field samples: size-exclusion chromatography for sample preparation, a specialized RNA extraction method, optimized cDNA synthesis, and the selection of DNA polymerase. Nine complete HAV genomes were obtained from clams. The obtained HAV genomes and their genetic characteristics were then compared based on phylogeny. Before optimization, only four clam samples yielded detectable amplification; however, following optimization, two additional samples became amplifiable, resulting in six samples suitable for downstream WGS analysis. The developed WGS method was able to sequence low contamination levels of 2.91–3.61 $\log_{10}$ genome copies/mL, achieving coverage of 97.5% and 92.6%. Notably, this study confirmed an average sequencing depth of up to 82.20× and a minimum depth of 25.19×. As a result of sequencing, one HAV-IA, and eight HAV-IB genotypes were identified from six clam samples including the multiple strains. The sequence identity between the strains from clams and serum was 97.80% for HAV-IA and 95.2–97.80% for HAV-IB. This method of viral WGS in food samples may contribute to rapid genotyping, understanding virus evolution, and enhancing epidemiological surveillance in foodborne virus outbreaks.

Hepatitis A virus (HAV), is a single-stranded positive-sense RNA virus classified within the *Picornaviridae* family[1]. There are six HAV genotypes, termed I–VI. Human infections are primarily caused by genotypes I, II, and III, and each are divided into subtypes A and B. Genotypes I and III are the most commonly reported worldwide[2]. The non-specific symptoms of HAV include malaise, loss of appetite, vomiting, and diarrhea. In more severe cases, symptoms may include dark urine, pale stools, and jaundice[3]. HAV is primarily transmitted via the fecal-oral route, either through direct human-to-human contact or by consuming contaminated food[1,2].

Once released into the environment, HAV can persist for weeks to months, either freely in the water column or by attaching to particulate matter and accumulating in sediments[4]. HAV is notably resistant to low pH and requires heating above 98 °C for effective inactivation[5]. Bivalve shellfish, through their filter-feeding activity, can concentrate and retain human enteric pathogens originating from sewage-contaminated waters[4]. Between 1986 and 2012, ~359 shellfish-borne viral outbreaks were reported worldwide, primarily associated with the consumption of oysters, clams, mussels, and cockles[6]. Notably, nearly 50% of these outbreaks occurred in Asia, reflecting the region's high consumption of raw or undercooked seafood[6]. In South Korea, a total of 17,598 hepatitis A cases were reported in 2019, largely linked to the consumption of HAV-contaminated salted clams[7]. Since the implementation of the Hepatitis Mandatory Surveillance System, hepatitis A has exhibited the highest incidence rate among viral hepatitis cases, reaching 33.95 per 100,000 individuals[7]. In Europe, particularly in Italy, from 1997 to 2015, the consumption of raw or insufficiently cooked shellfish was identified as the predominant risk factor for shellfish-associated viral

[1]Department of Food and Nutrition, Chung-Ang University, Anseong-si, Gyeonggi-do, Republic of Korea. [2]Sanigen, Anyang-si, Gyeonggi-do, Republic of Korea. [3]Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea. ✉e-mail: cchoi@cau.ac.kr

infections, accounting for 86.2% (75/87) of reported cases[8]. Similarly, in the United States, a multistate HAV outbreak in 2005 resulted in 39 infections following the consumption of contaminated oysters[6]. Epidemiological evidence consistently indicates that human enteric viruses are the most common etiological agents transmitted through bivalve shellfish.

Despite these large scale HAV outbreaks, detection and surveillance systems were still using the short fragments PCR of the VP1–P2B region (392 bp) or VP3–VP1 region (186 bp)[7,9]. The integration of short gene fragments with epidemiological data and comprehensive HAV characterization remains a major challenge in molecular epidemiology[10,11]. The identification of foodborne HAV outbreaks faces multiple critical challenges: First, the prolonged incubation period of 14–28 days, potentially extending to 50 days, complicates the temporal association between exposure and disease onset. Second, the inability to cultivate wild-type HAV in vitro limits laboratory investigations. Third, despite the high infectivity of HAV (10–100 viral particles sufficient for infection), food samples typically contain minimal viral contamination levels, hampering detection efforts[2,12,13].

Whole-genome sequencing (WGS) of foodborne viruses in food samples using high-throughput sequencing platforms has gained increasing attention. For example, norovirus has been detected in oysters using Illumina platform and optimization of metagenomic analyses using laboratory-prepared oyster samples[14,15]. However, performing WGS directly from food matrices remains challenging, as matrix composition, texture, and solubility can hinder viral nucleic acid extraction. Optimized and matrix-specific extraction methods are essential for maximizing viral recovery while minimizing contamination[16]. To date, no studies have applied high-throughput sequencing for WGS of wild-type foodborne viruses directly from food samples. Most sequencing optimization efforts using high-throughput platforms have focused on samples with high viral concentration and purity[17,18]. In clinical setting, high-throughput sequencing has been utilized for WGS of viruses such as influenza, monkeypox, COVID-19, and Zika in serum and plasma samples[19–23]. Similarly, in the veterinary field, viruses like porcine reproductive and respiratory syndrome virus and parvovirus have been subjected to WGS using high-throughput sequencing[24]. These studies were performed in clinical and veterinary contexts, where virus cultivation or enrichment is feasible, unlike food matrices with low viral loads and complex inhibitors.

In this study, we developed and applied a WGS approach for the direct detection and genetic characterization of wild-type HAV from contaminated food samples. Our research aims to bridge these gaps by developing a WGS method for HAV from clam samples. To optimize the HAV WGS method, we focused on refining key steps in molecular detection, including RNA extraction, cDNA synthesis, and the selection of DNA polymerase. Building on these optimized methods, we developed an improved long PCR method targeting the HAV capsid region and a WGS method using PCR amplicons covering the entire HAV genome on the Illumina platform.

## Results

### RNA extraction, cDNA synthesis, and DNA polymerase optimization

The 260/280 ratio of RNA purity was assessed for various RNA extraction methods, including the RNeasy mini-Kit, RNeasy midi-Kit, NucliSens miniMAG magnetic bead method, and TRIzol + RNA clean and concentrator method. The ratios observed were $2.18 \pm 0.25$, $5.01 \pm 0.35$, $2.79 \pm 0.25$, and $1.13 \pm 0.25$, respectively, indicating variations in RNA purity among the methods. Total RNA yields were also measured, with the RNeasy mini-kit yielding $100.40 \pm 19.23$ ng/50 µL, the RNeasy midi-kit yielding $134.59 \pm 52.13$ ng/150 µL, the NucliSens mini-MAG method yielding $0.01 \pm 0.04$ ng/150 µL, and the TRIzol + RNA clean and concentrator combination yielding $162.10 \pm 8.2$ ng/100 µL. The combination of TRIzol + RNA clean and concentrator exhibited the best performance, achieving high RNA purity and yield, and was deemed most suitable for downstream applications.

To optimize cDNA synthesis conditions, various primer types and concentrations were tested. The HAV strain HM175, known for cell culturable strain, was used as a positive control. Tested samples included HAV detected in clam tissues (strain cau210584, undetectable Ct value by qPCR) and HAV detected in serum (strain cau230022, quantified at 6.29 $\log_{10}$ copies/mL). The efficiency of cDNA synthesis was validated using a long PCR primer set. A combination of 50 µM oligo $(dT)_{18}$ primers with 50 µM random hexamer primers consistently amplified 3208 bp positive bands in both clam and serum samples, demonstrating robust synthesis efficiency, as shown in Fig. 1a.

Building on the optimized RNA extraction and cDNA synthesis protocols, the performance of three different DNA polymerases was evaluated for sensitivity and detection limits (Fig. 1b). Serial dilutions of the HAV HM175 strain (ranging from $2.0 \times 10^5$ to $10^7$ copies/µL) were used as templates. Platinum SuperFi II DNA polymerase exhibited the highest sensitivity, detecting HAV down to $2.0 \times 10^5$ copies/µL. In contrast, *Thermus aquaticus* YT1 Taq polymerase and Q5® High-Fidelity DNA polymerase detected HAV at higher thresholds, with detection limits of up to $2.0 \times 10^7$ copies/µL.

### Optimization of amplification for HAV WGS from clam samples

Nine HAV-positive clam samples were utilized to evaluate the optimization of the sample preparation process. The revised protocol incorporated proteinase K treatment, size exclusion chromatography, and ultrafiltration, complementing the optimizations presented in Fig. 1. The efficacy of these enhanced steps was validated using amplicon-based WGS. Following optimization, two additional clam samples, cau210579 (lane E) and cau210589 (lane H), showed positive PCR bands, as presented in Fig. 2a. Moreover, clearer and more defined PCR bands were observed in samples cau210574 (lane A), cau210576 (lane B), cau210584 (lane G), and cau210595 (lane I), further confirming the improvements achieved through the revised preparation steps. The same PCR amplicons were subsequently analyzed with a bioanalyzer, demonstrating sufficient amplification of the 400–700 bp PCR products to levels appropriate for WGS, as shown in Fig. 2b, which is consistent with the results in Fig. 2a. Two clam samples (cau210574 and cau210584) that were initially negative by RT-qPCR yielded positive amplification at the 400 bp region during PCR analysis, as depicted in Fig. 2.

### Amplicon-based WGS using high-throughput sequencing

Among nine clam samples analyzed, four demonstrated RT-qPCR positive results, as presented in Table 1. The HAV viral loads were quantified as 2.91, 3.57, 3.61, and 3.22 $\log_{10}$ copies/mL for samples cau210576, cau210579, cau320589, and cau210595, respectively. Sample cau320589 exhibited the highest viral load at 3.61 $\log_{10}$ copies/mL. Six clam samples (cau210574, cau210576, cau210579, cau210584, cau210589, and cau210595) were mapped to the two genotypes (IA and IB) of the reference strains. Mapping the sequences to the HAJFF-Kan12 strain identified a total of six strains: cau210574-1, cau210576, cau210579, cau210584-1, cau210589, and cau210595-1. The coverage percentages were 97.5% for cau210574-1, 96.7% for cau210576, 93.3% for cau210579, 92.6% for cau210584-1, 93.4% for cau210589, and 87.7% for cau210595-1.

Analysis of sequencing data revealed variability in average sequencing depths across the samples. The highest average depth was observed in sample cau210579 (82.20×), whereas the lowest was noted in cau210584 (25.19×). Intermediate average depths were recorded for samples cau210574 (43.15×), cau210576 (37.45×), cau210589 (50.98×), and cau210595 (49.15×), as depicted in Fig. 3.

### Phylogenetic analysis of the whole-genome sequence of HAV from clam

Figure 4 showed that the sequence of clam sample cau210579 clustered within the MW405349–ON911721 clade, which includes HAV strains from South Korea, Japan, and China. The nucleotide sequence of cau210579 demonstrated a high identity of 99.27% with AB819870 (HAJFF-Kan12 strain from Japan) and OM291913 (HAV KUMC 19-1 strain from South Korea). Among the HAV strains detected in clams, eight were identified as genotype IB. The nucleotide sequence identity analysis revealed that

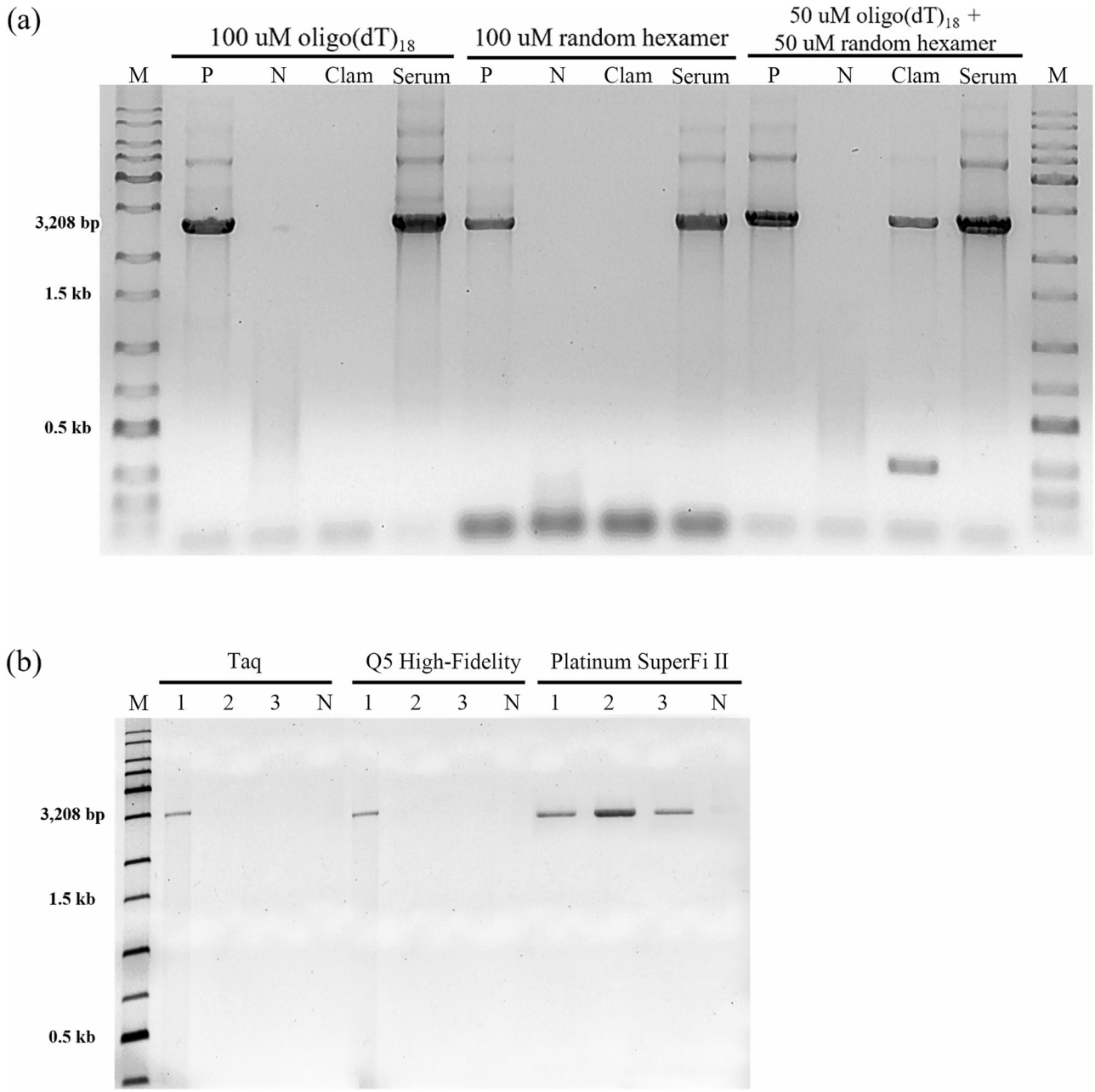


**Fig. 1 | The results of reverse transcription PCR and DNA polymerase optimization. a** Reverse transcription PCR optimization. M: 1.0 kb plus DNA ladder, P: positive control (HM175 strain), N: negative control (RNase-free water), Clam: hepatitis A virus (HAV)-positive clam sample (cau210584 strain), Serum: HAV-positive serum sample. **b** DNA polymerase activity for long PCR. M: 1.0 kb plus DNA ladder, lane 1–3: $2.0 \times 10^7$–$10^5$ copies/μL of HAV HM175 strains. N: negative control (RNase-free water).

cau210574-1 had a 97.80% similarity with cau210595-1, and cau210574-2 had a 95.52% similarity with cau210574-1. The cau210576 strain showed a 98.40% similarity with the reference strains KP879216.1 and KX035096. Additionally, cau210584-1 had a 97.90% similarity with cau210584-2, which in turn had a 97.92% similarity with cau210589. Moreover, the cau210595-1 strain also had a 97.80% similarity with cau210574-1, while cau210595-2 showed a 96.97% similarity with cau210574-1. Regarding clades, these sequences formed the five distinct clades shown in Fig. 4. Strain cau210574-1 and cau210584-1 clustered into one clade. Two sequences, cau210595-1 and cau210595-2, clustered from the same clam sample. Another clade consisted of cau210584-2 and cau210589. Additionally, cau210576 formed a divergence from the reference strain, whereas cau210574-2 formed a separate single node within the IB genotype.

Using the amino acid sequence identity and phylogeny, one IA and eight IB genotypes were identified, as shown in Fig. 5. The cau210579 strain had a 98.68% similarity with ON911721 (DgLn5.14 strain from China), forming a clade from ON911721 to cau210579. The IB genotype had a single clade with the HM175 strain and its adapted strains, ranging from KX343016 to KX035096. Wild-type IB strains clustered into two large clades: one consisting of the ON524433.1 to ON524426 sequences, and another comprising the other strains identified in this study, which formed the cau210576 to cau210574-2 clade. In addition, the amino acid sequence identity analysis of the identified HAV strains showed high similarity among the strains from this study. From cau210574-2 to cau210595-2, each strain had a high similarity of 95.52–98.91%. Additionally, cau210584-1 to cau210589 had a 97.89–97.91% amino acid sequence similarity.

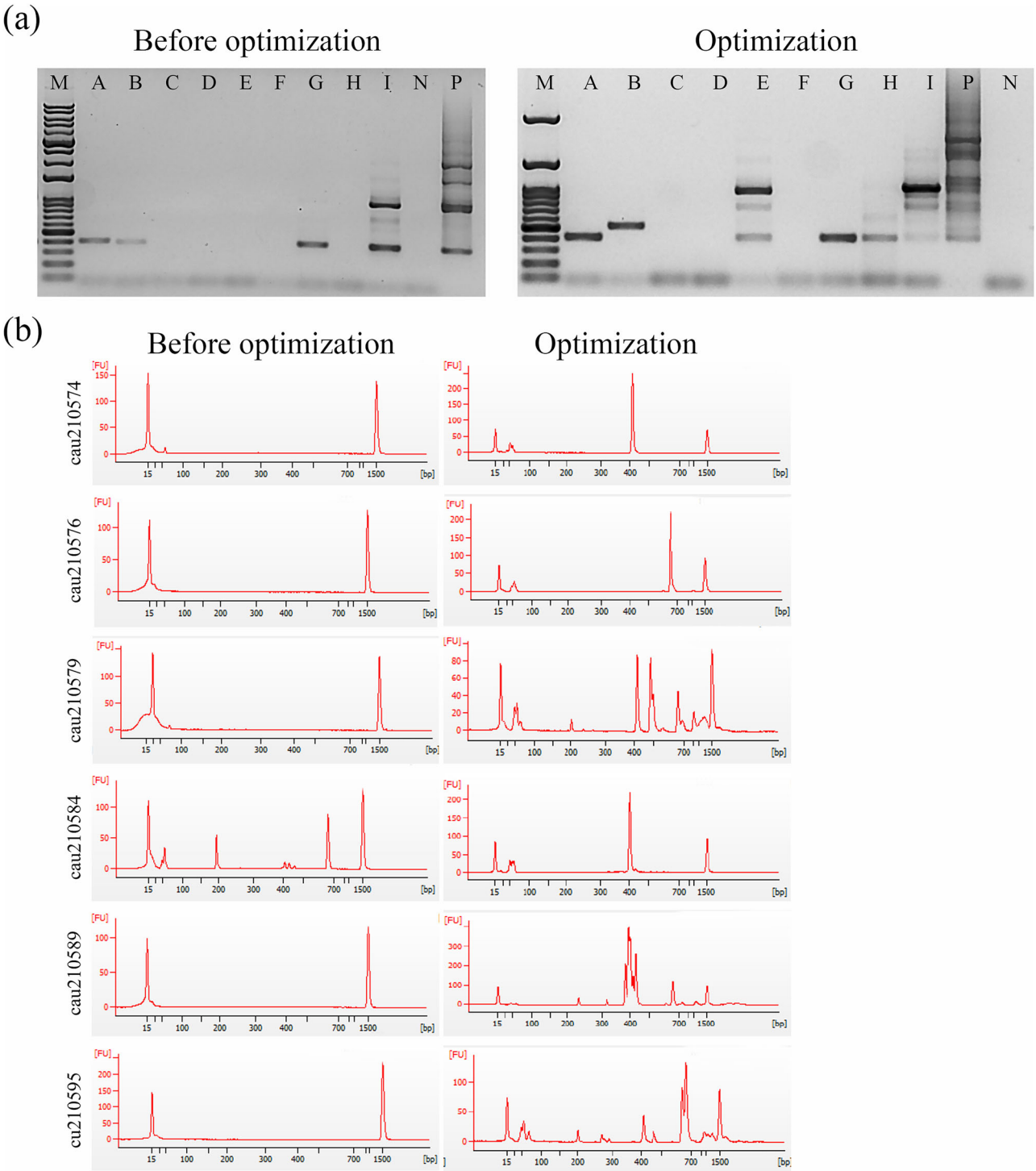


**Fig. 2 | The results of the improvement of sample preparation. a** PCR electrophoresis band image of the amplicon-based whole-genome sequencing (WGS) primer. The lanes are labeled as follows: M: 100 bp plus DNA ladder, P: positive control HM175 strain, N: negative control (RNase-free water), A: cau210574, B: cau210576, C: cau210577, D: cau210578, E: cau210579, F: cau210582, G: cau210584, H: cau210589, I: cau210595. **b** Bioanalyzer peak image of the amplicon-based WGS primer.

The difference between the two clades was 5.9%. However, among the eight HAV strains identified in this study, only the cau210576 strain showed a high similarity of 98.39% with the reference strain M59808.

## Discussion

We developed and optimized methods for HAV WGS from clam samples. To enhance the HAV sequencing method, we applied four approaches to HAV-positive clam field samples: size-exclusion chromatography for sample preparation, a specific RNA extraction method, cDNA synthesis, and the selection of DNA polymerase. This four-step optimization process enabled the development of HAV WGS methods that required neither cell adaptation nor artificial viral spiking.

Clams are known to contain PCR inhibitors such as salts, algae, glycogen, and polysaccharides, with salts and polysaccharides being particularly prominent. Polysaccharides can impede the resuspension of precipitated RNA and disrupt enzymatic processes[25]. Additionally, high

**Table 1 | Hepatitis A virus viral load from clam and virus strain results confirmed by whole-genome sequencing**

| Sample ID | Ct value | HAV viral load[a] | Genotype[b] | Coverage (%) | Average Depth | GenBank number |
|---|---|---|---|---|---|---|
| cau210574 | ND[c] | NA[d] | IB | 97.5 | 43.15 | PQ008999 |
| | | | IB | 96.7 | 43.15 | PQ009000 |
| cau210576 | 38.66 | 2.91 | IB | 96.7 | 37.47 | PQ009001 |
| cau210577 | ND | NA | NA | – | – | – |
| cau210578 | ND | NA | NA | – | – | – |
| cau210579 | 36.19 | 3.57 | IA | 93.3 | 82.2 | PQ009002 |
| cau210582 | ND | NA | NA | – | – | – |
| cau210584 | ND | NA | IB | 92.6 | 25.19 | PQ009003 |
| | | | IB | 88.2 | 25.19 | PQ009004 |
| cau210589 | 36.05 | 3.61 | IB | 93.4 | 50.98 | PQ009005 |
| cau210595 | 37.49 | 3.22 | IB | 87.7 | 49.15 | PQ009006 |
| | | | IB | 93.8 | 49.15 | PQ009007 |

[a]HAV viral load ($log_{10}$ copies/μL).
[b]HAJFF-Kan12, wild-type HM175 2 strains mapped. Duplicate markings are the result of multiple strain identification.
[c]ND: Not detected.
[d]NA: Not available.

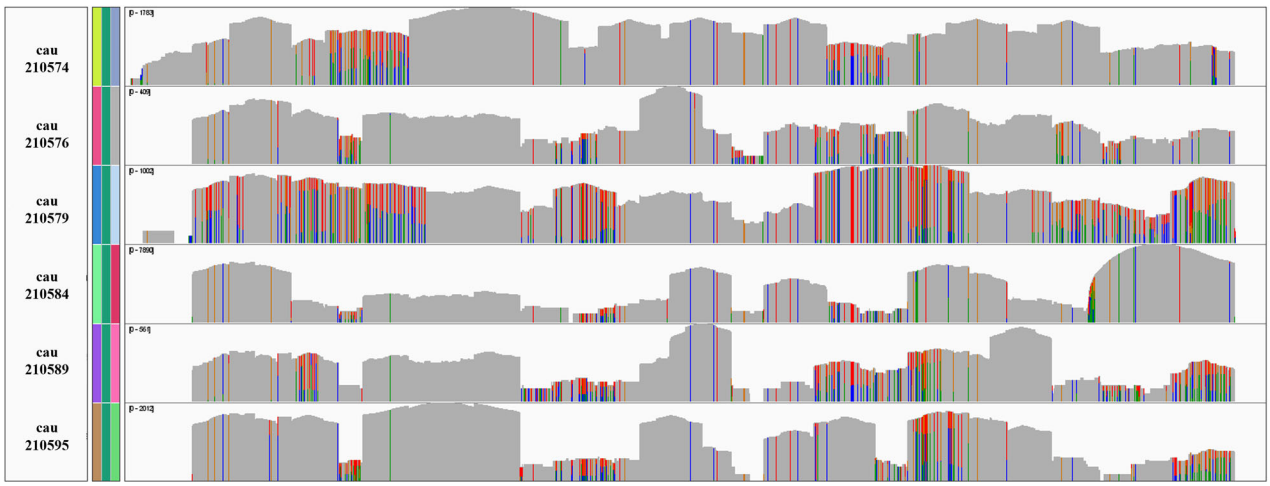


**Fig. 3** | Whole-genome sequencing analysis in integrative genomics viewer of hepatitis A virus from clam samples.

concentrations of salts like calcium may compete for binding with DNA polymerase, further inhibiting the PCR process[25]. Size exclusion chromatography is recognized for its effectiveness in removing free plasma proteins, desalting proteins, and separating large protein molecules[26,27]. Using size exclusion chromatography efficiently removed these PCR and sequencing inhibitors.

A comparative analysis of four RNA extraction methods for WGS optimization revealed that combining TRIzol + RNA clean and concentrator methods delivered superior results. The selected approach produced purified RNA with a 260/280 ratio of 1.13 ± 0.25 and a total RNA yield of 162.10 ± 8.2 ng/100 μL, confirming its optimal performance. The findings aligned with previous research on HAV nanopore sequencing[28], which identified the TRIzol LS method as optimal for producing higher purity RNA, achieving a 260/280 ratio of 2.19.

Considering that HAV genomes possess a 3' UTR followed by a poly A-tail[29], it was hypothesized that a combination of random hexamer primers and oligo $(dT)_{18}$ primers would provide more efficient amplification. This combined approach successfully generated 3208 bp amplicons in clam and serum samples, as shown in Fig. 1a. The results corroborate previous studies[30], which demonstrated increased specificity and sensitivity in cDNA synthesis from long non-coding RNA when employing initial poly A-tailing followed by random hexamer primers.

Selection of an appropriate DNA polymerase is crucial for amplifying long viral genomes, particularly in complex food matrices where inhibitors are present. High-fidelity DNA polymerases are commonly used in long PCR due to their high accuracy[31]. The fidelity of DNA polymerases varies by type, with low-fidelity enzymes such as Taq exhibiting fidelity values around $10^{-5}$, whereas high-fidelity enzymes typically have values closer to $10^{-6}$[31]. Previous studies have demonstrated the suitability of *Pyrococcus furiosus* pfu DNA polymerase for long PCR[32,33], a finding confirmed in the current investigation. Experimental results showed that Platinum SuperFi II DNA polymerase exhibited 100-fold higher sensitivity, likely due to the consistent buffer composition maintained throughout the cDNA synthesis and long PCR steps. Given the lack of prior research addressing this specific aspect, further investigation is required to elucidate the underlying mechanisms.

Previous analyses of HAV outbreaks from contaminated salted clam shellfish were limited to the VP1–2B region (392 bp)[7]. Efforts to establish epidemiological links between import sources, contaminated food, and patients were inconclusive, primarily due to variations in sequence lengths and target regions. In contrast, the optimized WGS protocol developed in this study successfully amplified HAV from six clam samples through refined food preprocessing and HAV amplification techniques. This novel WGS methodology demonstrated superior sensitivity for detecting low-titer HAV, achieving successful sequencing at contamination levels ranging from

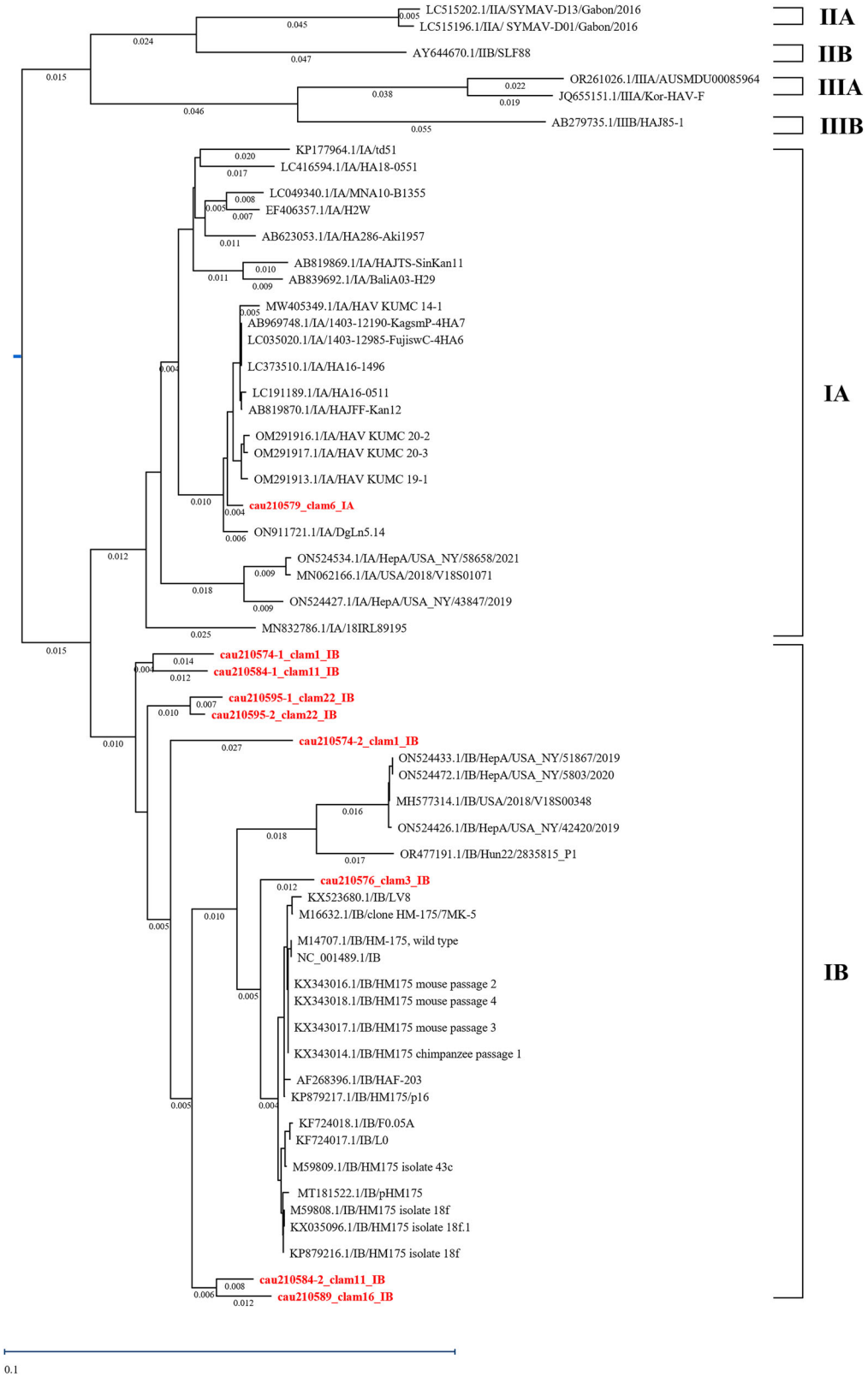


**Fig. 4 | The phylogenetic tree of hepatitis A virus (HAV) nucleotide whole-genome sequence.** Red leaves indicate clam HAV strains identified in this research. The numbers of compressed sequence distances are shown in parentheses.

2.91 to 3.61 $\log_{10}$ copies/mL. Notably, even with titers below the detection limit, samples cau210574 and cau210584 achieved high coverage rates of 97.5% and 92.6%, respectively. While earlier studies were limited by restricted target regions and sequence lengths, this study offers a more comprehensive and sensitive approach for detecting and analyzing HAV.

Another search achieved wild-type HAV WGS from frozen berries using PCR amplicons on the Illumina MiSeq platform[34], employing a similar methodological approach. However, this approach had the limitation of requiring additional HAV isolation steps using Frp/3 cells[35], which hindered its practical application due to the rare successful propagation of


**Fig. 5 | Phylogenetic tree of hepatitis A virus (HAV) amino acid sequence whole-genome sequence.** Red leaves indicate clam HAV strains identified in this research. The numbers of compressed sequence distances are shown in parentheses.

wild-type HAV, necessitating cell adaptation. In contrast, this study enables direct WGS without requiring cell culture adaptation. Additionally, whereas earlier studies offered limited sequencing quality data, the methodology presented here demonstrated robust metrics, achieving average depths of up to 82.20× and minimum depths of 25.19×.

Our study identified multiple strains within the same clam sample, with mixed contamination detected in three samples: cau210574, cau210584, and cau210595. As filter feeders, shellfish are known to filter fine particles from the water[36], which can concentrate pathogens. Various bacterial and viral pathogens have been reported in shellfish. By transferring

oysters free of specific pathogens to saline water, the progression of different infections has been studied[37]. Additionally, mixed calicivirus infections in shellfish have been linked to illnesses from consuming raw oysters[38].

In the IB cluster, the strains mainly used as reference genes, KX523680 (LV8 strain) to KP879216 (HM175/18f strain), formed a single clade in Fig. 5. The ON911721 (isolate DgLn5.14), which was reportedly isolated from the serum of a patient with hepatitis A in China in 2014, has a high sequence identity with the IB type strain identified in this study. The high sequence identity indicates that the data are unlikely to result from sequencing errors caused by interference with the mapping reference sequence or PCR primers. The clade containing the IA type cau210579 strain was identified in the current study, with reference strains comprising isolates from Japan, China, and South Korea. These results confirm the endemic nature of HAV, as previously documented[1]. While earlier research identified IA and IIIA as the predominant genotypes in South Korea[7], the current HAV WGS analysis revealed that IB genotypes were most frequently detected in clam samples.

Currently, WGS of HAV directly from food, and water, samples has not been reported. The methodology established in this study provides a critical framework for future HAV WGS research and enables systematic data accumulation for more comprehensive epidemiological investigations. For efficient sample preparation, additional size exclusion chromatography and ultrafiltration proved effective for inhibitor removal and concentration after proteinase K treatment. In the subsequent steps, optimizations included using TRIzol and RNA clean and concentrator, employing a mixture of two types of primers for cDNA synthesis, and utilizing Platinum SuperFi II DNA polymerase. The improved WGS method successfully sequenced very low contamination levels of 2.91–3.61 $\log_{10}$ copies/mL, achieving coverage of 97.5% and 92.6%. The analysis confirmed an average sequencing depth of up to 82.20× and a minimum depth of 25.19×. Future studies should include a broader range of shellfish samples and HAV genotypes. Extending this method to food and environmental samples may also enable the identification of additional HAV genotypes.

## Methods

### Preparing the positive control virus

The reference HAV strain HM175 (ATCC VR-1402) and fetal rhesus monkey kidney cells FRhK-4 (ATCC CRL-1688) were obtained from the ATCC. FRhK-4 cells were maintained in DMEM supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco, Grand Island, NY, USA) at 37 °C[39]. For virus propagation, FRhK-4 cell monolayers at 80–90% confluence were infected with HAV and incubated at 37 °C for 5 days. The virus-containing supernatant was harvested after three freeze-thaw cycles (–80 °C/37 °C) and clarified by centrifugation at 2500 × *g* for 15 min at 4 °C[40]. The HM175 strain served exclusively as a positive control for subsequent molecular analyses.

### Prescreening of HAV from clam

The clam samples were collected from 2019 to 2020 through the Ministry of Food and Drug Safety of the Republic of Korea. Clam samples were prescreened for HAV positivity following the ISO 15216-1:2017 standard[41]. Digestive tissues were dissected from five clams, treated with 3 U/mL proteinase K, and then pooled. The tissue homogenate was incubated at 100 rpm for 60 min at 37 °C, followed by clarification via centrifugation at 4000 × *g* for 5 min at 4 °C. The resulting supernatant was used for RNA extraction[41].

RNA extraction was performed using the NucliSens miniMAG system (NSmM; Biomerieux, Marcy l'Étoile, France). Subsequently, cDNA synthesis was carried out with the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc., Cincinnati, OH, USA). The primer sequences used for HAV prescreening and quantification were HAVfor (GGT AGG CTA CGG GTG AAA C), HAVrev (AAC TCA CCA ATA TCC GC), and HAVpro (CTT AGG CTA ATA CTT CTA TGA AGA GAT GC)[42]. The qPCR process was carried out by initially heating to 95 °C for 10 min, followed by 45 cycles of heating at 95 °C for 15 sec, 55 °C for 20 sec, and then 72 °C for 15 sec using a CFX96™ Real-Time PCR system (Bio-Rad, Hercules, CA, USA) and Premix Ex Taq (2X)™ (Takara Bio Inc., Shiga, Japan). A standard curve for the HAV quantitative genomic DNA (VR-3257SD, ATCC) was used. The viral copy number was determined by linear regression analysis, which showed a high correlation with an $R^2$ value of 0.99. The same clam sample was used in subsequent processes.

### Optimization of clam sample preparation

The workflow for sample preparation optimization appears in Fig. 6. The protocol began with homogenizing raw clam (*Venerupis philippinarum*) digestive tissues (3.0 g) combined with 3 U/mL proteinase K solution. The homogenized clam sample underwent a two-step incubation process, initially at 37 °C for 60 min with shaking at 100 rpm, followed by a second incubation at 60 °C for 15 min. Afterward, the sample was centrifuged at 4000 × *g* for 5 min at 4 °C. Following centrifugation, 5 mL of supernatant was collected, and the remaining homogenate was stored at −80 °C for RNA extraction[41].

The optimization protocol incorporated a PD-10 desalting column and ultrafiltration to improve sample preparation and RNA recovery. For sample preparation, 3.0 g of clam digestive tissues were homogenized in a solution containing 3 U/mL of proteinase K to digest proteins and enhance viral RNA release. The homogenate was incubated at 37 °C with gentle shaking at 100 rpm for 60 min to allow sufficient enzymatic activity. The mixture was then incubated at 60 °C for 15 min to inactivate potential contaminants and enhance protein digestion. After incubation, the homogenized sample was centrifuged at 4000 × *g* for 5 min at 4 °C to separate the supernatant from the debris. A total of 5 mL of clarified supernatant was carefully collected, and the pellet was discarded[43].

For desalting, PD-10 desalting columns packed with Sephadex G-25 resin (Cytiva, Marlborough, MA, USA) were used to remove small molecules and impurities from the supernatant. The columns were equilibrated by sequentially adding deionized water (~10 mL) five times to ensure thorough resin preparation. After equilibration, the columns were placed into clean 50 mL centrifuge tubes to prevent cross-contamination during sample processing. A measured volume of 2.25 mL of the clarified supernatant (minimum 1.75 mL, maximum 2.5 mL) was carefully pipetted onto the center of the resin bed within the PD-10 column to ensure even distribution. The column was then centrifuged at 1000 × *g* for 2 min at 4 °C to collect the eluted supernatant. To improve desalting efficiency, the process was repeated twice, ensuring maximum removal of impurities.

The eluted sample was then subjected to ultrafiltration for further concentration using Vivaspin-20 centrifugal concentrators with a nominal molecular weight cutoff of 10 kDa (Sartorius, Göttingen, Germany). This step retained viral RNA while excluding smaller contaminants. Ultrafiltration was performed by centrifuging the sample at 8000 × *g* for 50 min at 4 °C to achieve optimal concentration levels[44]. The final concentrated sample was aliquoted and stored at −80 °C to preserve RNA integrity for subsequent extraction and downstream analyses.

### Optimization of the RNA extraction method from clam samples

The RNA extraction optimization workflow appears in Fig. 6. Initial RNA extraction employed the NucliSens miniMAG magnetic bead method[41]. A comparative analysis evaluated four RNA extraction methods: NucliSens miniMAG (NSmM; Biomerieux) magnetic bead method, RNeasy mini-kit (Qiagen, Hilden, Germany), RNeasy midi-kit (Qiagen), and a combination of TRIzol with RNA clean and concentration-5 kits (Zymo Research, Irvine, CA, USA). The extraction protocols adhered to manufacturer guidelines, incorporating a modified approach for the combination of TRIzol and RNA clean and concentrator-5 kits[28]. The modification included mixing the RNA-containing aqueous phase obtained after TRIzol LS treatment with an equal volume of absolute ethanol. The resulting mixture underwent

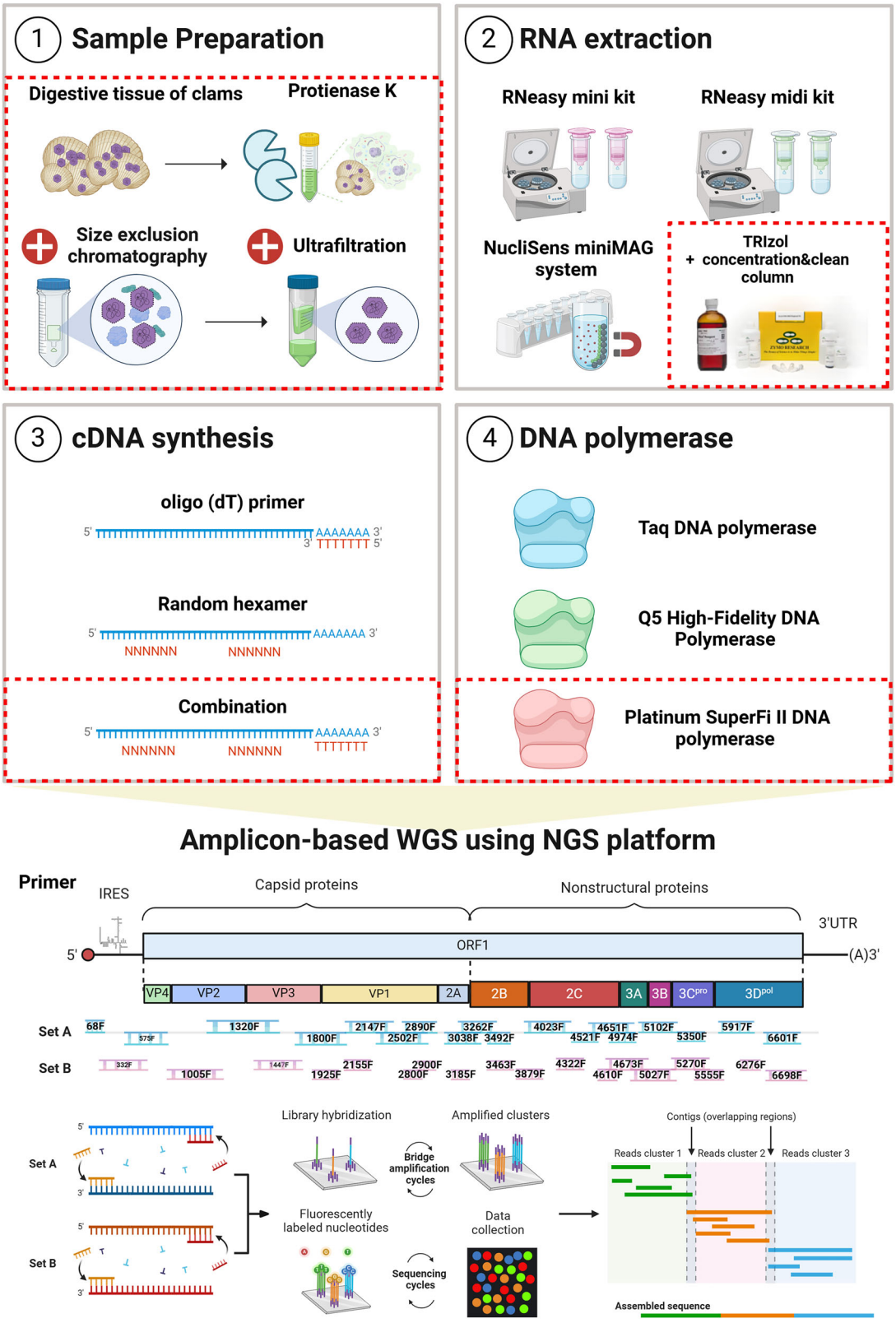


**Fig. 6 | Schematic diagram of this study.** The red box in the diagram indicates the optimized method. Created in BioRender. Jeong, S. (2025) https://BioRender.com/g94o791.

centrifugation at 16,000 × *g* for 30 s through a Zymo-spin IC column. Subsequent processing steps followed the manufacturer's protocol. Final RNA elution was performed in 15 µL, with storage at −80 °C until further analysis.

### Optimization of the reverse transcription PCR and DNA polymerase

Reverse transcription for cDNA synthesis was performed using 2.0 ng of total RNA and the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo

**Table 2 | Reverse transcription PCR reaction composition and primer concentrations**

| Component | Final volumes and concentrations of RT-PCR components | | |
|---|---|---|---|
| | Primer 1 | Primer 2 | Primer 3 |
| 5X PCR buffer | 1X | 1X | 1X |
| RTase[a] | 1 µL | 1 µL | 1 µL |
| RNase inhibitor (20U) | 1 µL | 1 µL | 1 µL |
| Primer (tested) | Oligo(dT)$_{18}$ 100 nM | Random hexamer 100 nM | Oligo(dT)$_{18}$ 50 nM + Random hexamer 50 nM |
| 10 mM dNTP | 2 µL | 2 µL | 2 µL |
| RNA | 2 µL (2 ng) | 2 µL (2 ng) | 2 µL (2 ng) |
| Nuclease-free water | to 20 µL | to 20 µL | to 20 µL |
| Total volume | 20 µL | 20 µL | 20 µL |

[a]RTase: SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific).

**Table 3 | PCR reaction composition and component concentrations for each DNA polymerase**

| Component | | Final volumes and concentrations of PCR components | | |
|---|---|---|---|---|
| | | Taq DNA polymerase[a] | Q5®[b] | SuperFi II[c] |
| PCR buffer | | 1X | 1X | 1X |
| dNTP | | 0.2 mM | 0.2 mM | 0.2 mM |
| Forward primer | | 0.25 µM | 0.25 µM | 0.25 µM |
| Reverse primer | | 0.25 µM | 0.25 µM | 0.25 µM |
| DNA polymerase | | 1 unit | 1 unit | 1 unit |
| cDNA | | 2 µL | 2 µL | 2 µL |
| Nuclease-Free water | | to 25 µL | to 25 µL | to 25 µL |
| Initial denaturation | | 94 °C, 1 min | 98 °C, 30 sec | 98 °C, 30 sec |
| Denaturation | 40 cycles | 94 °C, 20 sec | 98 °C, 10 sec | 98 °C, 10 sec |
| Annealing | | 53 °C, 1 min | 53 °C, 1 min | 53 °C, 1 min |
| Extension | | 72 °C, 1 min | 72 °C, 1 min | 72 °C, 1 min |
| Final extension | | 72 °C, 10 min | 72 °C, 10 min | 72 °C, 10 min |

[a]Thermus aquaticus YT1 Taq polymerase (Bioneer).
[b]Q5® High-Fidelity DNA Polymerase (New England Biolabs).
[c]Platinum SuperFi II DNA polymerase (Thermo Fisher Scientific).

Fisher Scientific). Three different primer sets were evaluated for efficiency and specificity, as summarized in Table 2 and Fig. 6. The reverse transcription PCR (RT-PCR) reaction was conducted under the following thermal conditions: 25 °C for 5 min, 42 °C for 1 hr, and 70 °C for 5 min using a T100 Thermal Cycler (Bio-Rad)[45]. The final reagent concentrations and reaction volumes used for RT-PCR are also detailed in Table 2. To optimize DNA amplification, three commercially available DNA polymerases were compared for their efficiency and sensitivity in detecting HAV genomic DNA. The tested DNA polymerases included: (1) Thermus aquaticus YT1 Taq polymerase (Bioneer, Daejeon, South Korea), (2) Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA), and (3) Platinum SuperFi II DNA Polymerase (Thermo Fisher Scientific). To evaluate the detection limits of each polymerase, HAV quantitative genomic DNA (VR-3257SD, ATCC) was serially diluted to concentrations ranging from $10^5$ to $10^7$ copies/µL. PCR reactions were prepared under standardized conditions, with final concentrations, reaction mixtures, and thermal cycling parameters specified in Table 3. The comparison of DNA polymerases aimed to determine their sensitivity, specificity, and robustness for HAV detection, providing critical insights into their suitability for WGS and genotyping applications.

### Long PCR

From serial diluted HAV HM175 strain, the long PCR was performed for the sensitivity test. HAV HM175 strain was RNA extracted, and cDNA synthesis using TRIzol and RNA clean and concentration-5 kits (Zymo Research) and a combination of 50 µM oligo dT primers and 50 µM random hexamer primers. The developed long PCR method targeted the 5'UTR to the 2A region of the HAV genome, covering nucleotide positions 78–3286 (a total of 3208 base pairs). The forward primer HAV-F1 (GCC TAG GCT ATA GGC TAA AT) was newly designed, while the reverse primer BR9 (AGT CAC ACC TCT CCA GGA AAA CTT) was slightly modified from a previously reported sequence[46]. Long PCR was performed under the SuperFi II conditions outlined in Table 3, using a T100 Thermal Cycler (Bio-Rad).

### HAV genome amplification and Illumina sequencing

Primer design employed the Primal Scheme tool (http://primal.zibraproject.org), utilizing multiple reference genomes: HAJFF-Kan12 strain (AB819870.1), HM175 wild type strain (M14707), SYMAV-D13/Gabon/2016 (LC515202), SLF88 strain (AY644670), Kor-HAV-F strain (JQ655151), and HAJ85-1 strain (AB279735). The amplicon genome sequencing strategy enabled WGS from clam samples. PCR amplification utilized two primer sets (A and B): set A comprising 18 forward and 18 reverse primers, and set B containing 17 forward and 17 reverse primers. These primers generated overlapping fragments of 200–500 bp, providing complete HAV genome coverage[47]. Table 4 details the amplification size for each primer pair.

The optimized protocol for HAV whole-genome amplification (Fig. 6) utilized a 25 µL reaction mixture containing 12.5 µL of 2X Platinum SuperFi II PCR Master mix (Thermo Fisher Scientific), 50 µM of primer set A (pools 1 and 2) or set B (pools 3 and 4), 6.20 µL of RNase-free water, and 2 µL of cDNA template. The PCR thermal cycling conditions consisted of initial denaturation at 98 °C for 30 sec, followed by 40 cycles of denaturation at 98 °C for 15 sec and annealing/extension at 60 °C for 5 min, with a final extension at 72 °C for 1 min. QC of PCR products was performed using a Bioanalyzer[48].

The DNA library was prepared with the TruSeq Nano DNA Library Prep Kit (Illumina) following the manufacturer's protocol. After genome pooling, sequencing was performed on a MiSeq platform (Illumina) using 300 bp paired-end reads. Post-sequencing processing involved removing adapters and reads below a Q-score threshold of 30 using Trimmomatic v0.38. Control DNA and duplicate reads were eliminated with Picard v2.20.2, and coverage normalization was carried out using BBMap 38.84[49].

### Phylogenetic analysis

The complete HAV whole-genome sequences were aligned with 49 reference HAV sequences to identify conserved and variable regions using ClustalW in DNASTAR Lasergene MegAlign Pro (ver 17.2.1; DNASTAR)[50]. ClustalW was chosen for its robust pairwise and multiple sequence alignment capabilities, ensuring high alignment quality by balancing sequence similarity and evolutionary distance.

Phylogenetic analysis of nucleotide sequences was conducted using the Neighbor-Joining method, which calculates evolutionary distances to construct an unrooted tree, revealing relationships among HAV genotypes and strains. For whole-genome amino acid sequence analysis, alignment was performed with the MAFFT program using the G-INS-i algorithm[50]. This alignment ensured the precise identification of homologous residues essential for accurate phylogenetic reconstruction. For the phylogenetic tree of the amino acid sequences, the maximum likelihood method was applied using RAxML-NG (ver. 1.1.0)[51]. The GTR + F + I + G4 model was selected to account for varying evolutionary rates across sites. maximum likelihood-based tree construction was performed with 1000 rapid bootstrap replicates to provide high statistical confidence in the tree topology[44].

**Table 4 | Primers for PCR-based high-throughput whole-genome sequencing of hepatitis A virus**

| Primer name | Multiple PCR set | Pool | Sequence (5' → 3') | Size | Primer name | Multiple PCR set | Pool | Sequence (5' → 3') | Size |
|---|---|---|---|---|---|---|---|---|---|
| 68F | A | 1 | TGC TTG TAA ATA TTA ATT CCT GCA GGT | 475 | 332F | B | 3 | ATA GGG TAA CAG CGG CGG ATA T | 526 |
| 542R | A | 2 | ATT CCC TCA ATG CAT CCA CTG G | | 857R | B | 4 | ACC AGT CAC TGC AGT CCT ATC A | |
| 575F | A | 1 | AGG TAC TCA GGG GCA TTT AGG T | 450 | 1005F | B | 3 | TGG CTC ACT ACA CAT GCT CTT T | 401 |
| 1024R | A | 2 | CAA CCC TTG AAC AGC AAA CTG C | | 1405R | B | 4 | TGA AGT GTA AGC TGA AGT TCC TGT | |
| 1320F | A | 1 | AAA GAT CCA CAA TAC CCA GTT TGG | 399 | 1447F | B | 3 | GAT CAG GAA GAT TGG AAA TCT GAT CC | 531 |
| 1718R | A | 2 | TGG GTC AAC TGG AAT AAC TTT GAT CT | | 1977R | B | 4 | TCA AGG TTG ATT GCA CTC CTG T | |
| 1800F | A | 1 | GTT TTG TTC CTG GCA ATG AGC T | 488 | 1925F | B | 3 | CTG ATG TGG ATG GTA TGG CCT G | 206 |
| 2287R | A | 2 | GTT GTT ATG CCG ACT TGG GGA T | | 2130R | B | 4 | TCC CAA GCC ATC CAA TGC TC | |
| 2147F | A | 1 | GCA AGC ACC TGT GGG AGC TA | 394 | 2155F | B | 3 | CAG GTT GGA GAT GAT TCW GGA GGT | 422 |
| 2540R | A | 2 | AGC TGA CTC CTT YTC YAC CCA AG | | 2576R | B | 4 | GGT CCT CTA TAY AAC TGA AAC AGG T | |
| 2502F | A | 1 | ACC TTC AAT TCA AAC AAT AAA GAG TAC ACA | 409 | 2800F | B | 3 | TCA GAT TAG ATT GCC ATG GTA TTC TTA TT | 141 |
| 2910R | A | 2 | TTG CAA TCT GAA TAG AAA CCA ATC CA | | 2940R | B | 4 | GCA TTT GAG TTC AAT GGA GCT CT | |
| 2890F | A | 1 | CTG GTT TCT ATT CAG ATT GCA AAT TAC A | 208 | 2900F | B | 3 | GGA GGA CAG AAG ATT TGA GAG TCA | 281 |
| 3097R | A | 2 | CTG TCC TCC TCT GAT CTG GGA T | | 3180R | B | 4 | CCA AGA AAA TTT CAT TAT TTC ATG CTC CT | |
| 3038F | A | 1 | AGT CAT ATA GAA TGC AGG AAG CCA | 385 | 3185F | B | 3 | AGA GTT GTC AAA TGA AGT GCT TCC | 201 |
| 3422R | A | 2 | TGT GGG AAA TTC ACT TTG GAC CA | | 3385R | B | 4 | GCA TCC ATC TCA AGA GTC CAC A | |
| 3262F | A | 1 | ATG GAT GCT GGG GTT CTT ACT G | 459 | 3463F | B | 3 | GAT TTG TGT TTC TTG CTG CAT TGG | 518 |
| 3720R | A | 2 | AGC ACA GTG TTT ATA ATT TCA ACA GTC A | | 3980R | B | 4 | CCA TCA TTC TGG AGT CCA TTT GC | |
| 3492F | A | 1 | TGG TCT AAA GTG AAT TTT CCA CAT GG | 280 | 3879F | B | 3 | TGA ATT ATG CAG ATA TTG GTT GTT CAG T | 294 |
| 3771R | A | 2 | CAA CAG TCA CAG AAT GAT GAA ACC C | | 4172R | B | 4 | ACC TGA GCC ATT GAA TGA ACA GT | |
| 4023F | A | 1 | GAG AAA GCY MTT GAR GAA GCC GAT | 455 | 4322F | B | 3 | AGA TGT GAG CCA GTT GTT TGC T | 276 |
| 4477R | A | 2 | TCT GAC CAR TCC TCA TCT GTT GTG T | | 4597R | B | 4 | GTT GTG TTT TGG CCA ATA TCA TCA AT | |
| 4521F | A | 1 | AGA GTA CAC ATT TCC AAT AAC TCT GTC | 290 | 4610F | B | 3 | GTC AGA TTT TTG TCA ATT AGT GTC AGG A | 140 |
| 4810R | A | 2 | GAC TCC TTT TCT ACC CAA GGC G | | 4749R | B | 4 | GGA TTT TTG AAA AAT GAA GCA GGT TTA AC | |
| 4651F | A | 1 | GGC CAA AAC YCA ACA GAT GAR GAT | 422 | 4673F | B | 3 | AAG GAA GCA ATT GAY CGY AGR CT | 383 |
| 5072R | A | 2 | AYT CCC TGW GAC CAC AAC TCC A | | 5055R | B | 4 | GCT CCC ACA GCA ACC CAC TT | |
| 4974F | A | 1 | AGT GAA TTC ATG GAG TTG TGG TCT | 200 | 5027F | B | 3 | CAG TTG GCA TTC TTG GAG TGC T | 513 |
| 5173R | A | 2 | AAC CAT CCT CCC ACA AGC AC | | 5539R | B | 4 | TGA AAC CCC ACA TCC AAA GAT TGA | |
| 5102F | A | 1 | TGT TAC TAA TCA CAA GTG GGT TGC T | 207 | 5270F | B | 3 | TGG GAG TGA AAG ATG ATT GGC T | 401 |
| 5308R | A | 2 | GCT ATT TCC AAA GTT GAC TGA GAC T | | 5670R | B | 4 | TTC ACC TTT TCC TCT CCA TGC C | |
| 5350F | A | 1 | AGA AAA ATG GAT GTG TGA GAT GGG | 202 | 5555F | B | 3 | TGT TAA TTT CTG AGG GCC CAC T | 458 |
| 5551R | A | 2 | AGA ACA ACA TCT TGA AAC CCC AC | | 6012R | B | 4 | GCT TTA GAA AAG GGC ATA GCT GC | |
| 5917F | A | 1 | TGT GGT CTC CAA AAC GCT TTT T | 365 | 6276F | B | 3 | TCT CCT GGG TTT CCT TAT GTC CA | 413 |
| 6281R | A | 2 | ACA GTA TTG AAT AAG ATT CTC TGA GCC | | 6688R | B | 4 | GCG GAA AAA TCT AAA TCA AGA CCA AC | |
| 6601F | A | 1 | TGG CAT AGA TCC TGA TAG ACA GTG | 397 | 6698F | B | 3 | AGT CTT TTC CAG AGA TGT TCA AAT TGA | 600 |
| 6997R | A | 2 | ACT ATC AAA ACA TCA TCT CCA TAA CAG AG | | 7297R | B | 4 | CCA AAT TGT CTT TTC TGA AAT TGC AGG | |

*F* forward primer, *R* reverse primer.

### Ethical approval

This study was approved by the Chung-Ang University Bioethics Committee (approval no. 1041078-202007-BR-179-01). Serum samples, including the cau230022 strain (GenBank accession no. PP389036), were prepared in accordance with the ethical standards of the Committee.

## Data availability

Accession numbers PQ008999 to PQ009007 are listed in Table 1 of the article.

Received: 9 January 2025; Accepted: 29 November 2025;
Published online: 22 December 2025

## References

1. Shouval, D. The history of hepatitis A. *Clin. Liver Dis.* **16**, 12–23 (2020).
2. Di Cola, G., Fantilli, A. C., Pisano, M. B. & Ré, V. E. Foodborne transmission of hepatitis A and hepatitis E viruses: a literature review. *Int. J. Food Microbiol.* **338**, 108986 (2021).
3. Herzog, C., Van Herck, K. & Van Damme, P. Hepatitis A vaccination and its immunological and epidemiological long-term effects–a review of the evidence. *Hum. Vaccines Immunotherapeutics* **17**, 1496–1519 (2021).
4. Formiga-Cruz, M. et al. Distribution of human virus contamination in shellfish from different growing areas in Greece, Spain, Sweden, and the United Kingdom. *Appl. Environ. Microbiol.* **68**, 5990–5998 (2002).
5. Noele Nelson, M. W. *Travel-Associated Infections & Diseases*. *Hepatitis A*. (Centers for Disease Control and Prevention, 2024). https://www.cdc.gov/yellow-book/hcp/travel-associated-infections-diseases/hepatitis-a.html.
6. Normanno, G. Foodborne viruses associated with consumption of shellfish. *EC Microbiol.* **17**, 45–57 (2021).
7. Jeong, H. et al. Hepatitis A virus strains identified in jogaejeot associated with outbreaks in Seoul, South Korea. *Lett. Appl. Microbiol.* **73**, 107–112 (2021).
8. Costantino, A. et al. Hepatitis A virus strains circulating during 1997-2015 in Campania, a Southern Italy region with periodic outbreaks. *J. Med. Virol.* **89**, 1931–1936 (2017).
9. Park, S. H. et al. A study on the detection rate of hepatitis A from gastroenteritis patients and the genotype analysis of hepatitis A virus in Busan. *J. Bacteriol. Virol.* **53**, 74–80 (2023).
10. Takeuchi, S. et al. Identification of potential pathogenic viruses in patients with acute myocarditis using next-generation sequencing. *J. Med. Virol.* **90**, 1814–1821 (2018).
11. Zufan, S. E. et al. High-performance enrichment-based genome sequencing to support the investigation of hepatitis A virus outbreaks. *Microbiol. Spectr.* **12**, e02834–02823 (2024).
12. Pradhan, A. K., Pang, H. & Mishra, A. In *Safety and Practice for Organic Food.* 135–150 (Elsevier, 2019).
13. WHO (World Health Organization). Weekly Epidemiological Record, 2022, vol. 97, 40 [full issue]. *Weekly Epidemiological Record.* **97**, 493–512 (2022).
14. Dirks, R. A. et al. A metagenomic survey of virological hazards in market-ready oysters. *Food Environ. Virol.* **17**, 16 (2025).
15. Imamura, S. et al. Next-generation sequencing analysis of the diversity of human noroviruses in Japanese oysters. *Foodborne Pathog. Dis.* **14**, 465–471 (2017).
16. Mahony, J. & van Sinderen, D. Virome studies of food production systems: time for 'farm to fork'analyses. *Curr. Opin. Biotechnol.* **73**, 22–27 (2022).
17. Misu, M. et al. Rapid whole genome sequencing methods for RNA viruses. *Front. Microbiol.* **14**, 1137086 (2023).
18. Sabatier, M. et al. Comparison of nucleic acid extraction methods for a viral metagenomics analysis of respiratory viruses. *Microorganisms* **8**, 1539 (2020).
19. Bang, E. et al. Zika virus infection during research vaccine development: investigation of the laboratory-acquired infection via nanopore whole-genome sequencing. *Front. Cell. Infect. Microbiol.* **12**, 819829 (2022).
20. Chauhan, R. P. & Gordon, M. L. Review of genome sequencing technologies in molecular characterization of influenza A viruses in swine. *J. Vet. Diagn. Investig.* **34**, 177–189 (2022).
21. Isabel, S. et al. Targeted amplification-based whole genome sequencing of Monkeypox virus in clinical specimens. *Microbiol. Spectr.* **12**, e02979–02923 (2024).
22. Wolf, J. M. et al. Molecular evolution of the human monkeypox virus. *J. Med. Virol.* **95**, e28533 (2023).
23. Yang, Z. et al. Application of next generation sequencing toward sensitive detection of enteric viruses isolated from celery samples as an example of produce. *Int. J. Food Microbiol.* **261**, 73–81 (2017).
24. Lalonde, C., Provost, C. & Gagnon, C. A. Whole-genome sequencing of porcine reproductive and respiratory syndrome virus from field clinical samples improves the genomic surveillance of the virus. *J. Clin. Microbiol.* **58**, e00097-20 (2020).
25. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. P. C. R. inhibitors–occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
26. Nagy, K. & Vékey, K. Separation methods. In *Medical Applications of Mass Spectrometry* (eds Vékey, K., Telekes, A. & Vertes, A.) 61–92 (Elsevier, 2008).
27. Pallares-Rusiñol, A. et al. Advances in exosome analysis. *Adv. Clin. Chem.* **112**, 69–117 (2023).
28. Batista, F. M. et al. Whole genome sequencing of hepatitis A virus using a PCR-free single-molecule nanopore sequencing approach. *Front. Microbiol.* **11**, https://doi.org/10.3389/fmicb.2020.00874 (2020).
29. Kozak, R. A. et al. Development and evaluation of a molecular hepatitis A virus assay for serum and stool specimens. *Viruses* **14**, 159 (2022).
30. Kolenda, T. et al. Quantification of long non-coding RNAs using qRT-PCR: comparison of different cDNA synthesis methods and RNA stability. *Arch. Med. Sci.* **17**, 1006 (2021).
31. McInerney, P., Adams, P. & Hadi, M. Z. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol. Biol. Int.* **2014**, https://doi.org/10.1155/2014/287430 (2014).
32. DeMarino, C. et al. HIV-1 RNA in extracellular vesicles is associated with neurocognitive outcomes. *Nat. Commun.* **15**, 4391 (2024).
33. Gyllö Lind, N. *Understanding Sequence Errors in Analysis of Short Tandem Repeats Through Optimisation of a Library Preparation Protocol Utilising Unique Molecular Indices*. Master's thesis, Uppsala University (2022).
34. Chiapponi, C. et al. Isolation and genomic sequence of hepatitis A virus from mixed frozen berries in Italy. *Food Environ. Virol.* **6**, 202–206 (2014).
35. Kanda, T. et al. Cell culture systems and drug targets for hepatitis A virus infection. *Viruses* **12**, 533 (2020).
36. Burge, C. A. et al. The use of filter-feeders to manage disease in a changing world. *Integr. Comp. Biol.* **56**, 573–587 (2016).
37. Petton, B. et al. Crassostrea gigas mortality in France: the usual suspect, a herpes virus, may not be the killer in this polymicrobial opportunistic disease. *Front. Microbiol.* **6**, 686 (2015).
38. Wang, Y., Zhang, J. & Shen, Z. The impact of calicivirus mixed infection in an oyster-associated outbreak during a food festival. *J. Clin. Virol.* **73**, 55–63 (2015).
39. Hossain, M. I. et al. Comparison of the virucidal efficacy of essential oils (cinnamon, clove, and thyme) against hepatitis A virus in suspension and on food-contact surfaces. *Curr. Res. Food Sci.* **7**, 100634 (2023).
40. Lee, M. H. et al. Reduction of hepatitis A virus on FRhK-4 cells treated with Korean red ginseng extract and ginsenosides. *J. Food Sci.* **78**, M1412–M1415 (2013).

41. ISO, E. 15216-1: 2017—Microbiology of the Food Chain—Horizontal Method for Determination of Hepatitis A Virus and Norovirus Using Real-Time RT-PCR—Part 1: Method for Quantification. (International Organization for Standardization, 2017). https://www.iso.org/standard/65681.html.
42. Jothikumar, N., Cromeans, T., Sobsey, M. & Robertson, B. Development and evaluation of a broadly reactive TaqMan assay for rapid detection of hepatitis A virus. *Appl. Environ. Microbiol.* **71**, 3359–3363 (2005).
43. Tunyakittaveeward, T. et al. Norovirus monitoring in oysters using two different extraction methods. *Food Environ. Virol.* **11**, 374–382 (2019).
44. Yeo, D. et al. Prevalence of foodborne viruses and influenza A virus from poultry processing plants to retailed chickens. *Front. Sustain. Food Syst.* **7**, https://doi.org/10.3389/fsufs.2023.1113743 (2023).
45. Jung, S. et al. Cross-species transmission and histopathological variation in specific-pathogen-free minipigs infected with different hepatitis E virus strains. *Vet. Res.* **55**, 87 (2024).
46. Park, S. H. et al. Molecular characterization of hepatitis A virus isolated from acute gastroenteritis patients in the Seoul region of Korea. *Eur. J. Clin. Microbiol. Infect. Dis.* **28**, 1177–1182 (2009).
47. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
48. Park, S. Y. et al. Utilizing cost-effective portable equipment to enhance COVID-19 variant tracking both on-site and at a large scale. *J. Clin. Microbiol.* **62**, e01558–01523 (2024).
49. Pillay, S. et al. Whole genome sequencing of SARS-CoV-2: adapting Illumina protocols for quick and accurate outbreak investigation during a pandemic. *Genes* **11**, 949 (2020).
50. Katoh, K. et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
51. Togkousidis, A. et al. Adaptive RAxML-NG: accelerating phylogenetic inference under maximum likelihood using dataset difficulty. *Mol. Biol. Evol.* **40**, https://doi.org/10.1093/molbev/msad227 (2023).

## Acknowledgements

This work was supported by the Ministry of Food and Drug Safety (grant number 20162MFDS033) from 2020 to 2021. This research was supported by the Chung-Ang University Graduate Research Scholarship (Academic scholarship for College of Biotechnology and Natural Resources) in 2024.

## Author contributions

C.C.: Conceptualization. S.J.: Methodology. S.J.: Validation. D.J.L., and J.C.: Formal Analysis. D.Y., and S.H.: Investigation. K.H.H.: Resources. D.Y. and J.C.: Data Curation. D.Y.: Writing – Original. D.Y. and C.C.: Writing – Review & Editing. S.J.: Visualization. C.C.: Supervision. C.C.: Project Administration. C.C.: Funding Acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Changsun Choi.