

# Benchmarking of a time-saving and scalable protocol for the extraction of DNA from diverse viromes

Michael Shamash<sup>1,2</sup>, Saniya Kapoor<sup>1</sup> and Corinne F. Maurice<sup>1,2</sup>

<sup>1</sup> Department of Microbiology & Immunology, McGill University, Montreal, QC, Canada

<sup>2</sup> McGill Centre for Microbiome Research, Montreal, QC, Canada

## ABSTRACT

The virome, composed of viruses inhabiting diverse ecosystems, significantly influences microbial community dynamics and host health. The phenol-chloroform DNA extraction protocol for viromes, though effective, is time-intensive and requires the use of multiple toxic chemicals. This study introduces a streamlined, scalable protocol for DNA extraction using a commercially-available kit as an alternative, assessing its performance against the phenol-chloroform method across human fecal, mouse fecal, and soil samples. No significant differences in virome diversity or community composition were seen between methods. Most viral operational taxonomic units (vOTUs) were common to both methods, with only a small percentage unique to either approach. Alpha- and beta-diversity analyses showed no significant impact of the extraction method on virome composition, confirming the kit's efficacy and versatility on sample types beyond those officially supported by the manufacturer. While the kit approach offers benefits like reduced toxicity and increased throughput, it has limitations such as higher costs and potential issues reliably capturing low-abundance taxa. This protocol provides a viable option for large-scale virome studies, although the phenol-chloroform approach may still be preferable for specific sample types.

Submitted 1 October 2024

Accepted 10 December 2024

Published 14 January 2025

Corresponding author

Corinne F. Maurice,  
corinne.maurice@mcgill.ca

Academic editor

Alexander Bolshoy

Additional Information and  
Declarations can be found on  
page 8

DOI 10.7717/peerj.18785

© Copyright  
2025 Shamash et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Microbiology

**Keywords** Virome, DNA extraction, Bacteriophage, Human gut microbiota, Soil

## INTRODUCTION

Viromes, the collections of viruses inhabiting diverse ecosystems, play important roles in shaping microbial community composition and function (Reyes *et al.*, 2013; Breitbart *et al.*, 2018; Hsu *et al.*, 2019; Howard-Varona *et al.*, 2020; Graham *et al.*, 2024). In host-associated ecosystems, like the human microbiome, the virome is also a key determinant of host health (Liang & Bushman, 2021).

Despite significant advances in the field in recent years, there remain several challenges associated with studying viromes. Many of these challenges are computational, such as identifying novel and highly divergent viruses, or characterizing the host range of uncultured viruses (Khan Mirzaei *et al.*, 2021). Others are related to sample processing, especially in low biomass sample types, such as amplifying low amounts of purified virome nucleic acids or dealing with high levels of background contamination (usually of human, mouse, or bacterial origin) (Khan Mirzaei *et al.*, 2021).

It is important that virome nucleic acid extraction protocols be efficient and reliable so that we may begin to address some of these analytical challenges. The phenol-chloroform DNA extraction protocol is currently the gold standard for studying viral metagenomes ([Thurber et al., 2009](#); [Shkoporov et al., 2018](#); [Göller et al., 2020](#); [Sinha et al., 2022](#)). However, this protocol has its limitations, including being time-consuming and requiring the use of multiple toxic organic chemicals such as phenol and chloroform.

Given the growing interest in studying viromes across diverse environments, including in human ([Shkoporov & Hill, 2019](#); [Shamash & Maurice, 2021](#)), animal ([Brunse et al., 2021](#); [Shkoporov et al., 2022](#)), and soil ([Roux & Emerson, 2022](#)) communities, there is a need for a faster and scalable extraction protocol which can accommodate increasingly large studies.

In this study, we set out to benchmark a streamlined protocol for virome DNA extractions using a commercially available column-based DNA extraction kit. While the manufacturer recommends its use for plasma, serum, and cell-free body fluids, we apply the kit to extract viral DNA from such as stool and soil samples. Although this kit has been used before for extracting viral DNA from human fecal samples, its performance has not been compared to that of the phenol-chloroform protocol ([Bikel et al., 2022](#)). Here we compare protocols, emphasizing the time savings and avoidance of toxic organic reagents when using the kit. Overall, we find no significant differences in the resulting sequenced viromes between the two methods, with comparable sample yield, viral diversity, and community composition.

## METHODS

Portions of this text were previously published as part of a preprint ([Shamash, Kapoor & Maurice, 2024](#)).

### Sample collection

Human fecal samples were collected with the approval of protocol A04-M27-15B from the McGill University Institutional Review Board. Participants provided informed written consent for the utilization of their samples and met specific inclusion criteria: they were 18 years of age or older, had no diagnosed gastrointestinal disease, and had not used antibiotics in the 3 months prior to sampling. Fresh fecal samples were collected, aliquoted in an anaerobic chamber, and kept at  $-70^{\circ}\text{C}$  until processing.

Mouse fecal samples were collected with the approval of animal use protocol MCGL-7999 from McGill University. Fresh fecal samples were collected and kept at  $-70^{\circ}\text{C}$  until processing. All mice had unlimited access to standard chow and water.

Soil samples were collected in sterile 50 mL conical centrifuge tubes from various locations on the downtown campus of McGill University (coordinates: 45.5042 N, 73.5755 W).

### Sample pre-processing

Prior to extracting virome DNA, we first enriched for virus-like particles (VLPs) in the fecal and soil samples. Samples were resuspended in sterile (0.02  $\mu\text{m}$ -filtered) PBS as follows: human fecal samples,  $\sim$ 200–400 mg in 2 mL PBS; mouse fecal samples,  $\sim$ 100 mg in 1 mL PBS; soil samples, 10 mL soil in 10 mL PBS (1:1 volume ratio). Large debris was

pelleted by centrifugation at 1,000 g for 5 min, and the supernatant recovered. Bacterial cells were pelleted by centrifugation at 10,000 g for 10 min. A total of 500  $\mu$ L of the VLP-containing supernatant was added to a sterile Ultrafree-MC centrifugal filter unit (MilliporeSigma, Burlington, MA, USA) with 0.22  $\mu$ m pore size and centrifuged at 12,000 g for 2 min, and 400  $\mu$ L of purified VLPs was recovered. The VLP suspension was cleaned by addition of 100  $\mu$ L chloroform (final chloroform concentration: 20% v/v), followed by thorough vortexing and centrifugation at 21,000 g for 5 min. The VLP-containing supernatant (upper layer) was carefully recovered and transferred to a new tube containing 5  $\mu$ L (10 U) TURBO DNase (Thermo Fisher Scientific, Waltham, MA, USA), 50  $\mu$ L 10X TURBO DNase buffer, and 1  $\mu$ L (approx. 125 U) Benzonase DNase (MilliporeSigma, Burlington, MA, USA). The sample was incubated at 37 °C with mild shaking for 90 min. To stop the DNase digestion reaction, 18  $\mu$ L of a 500 mM EDTA (pH 8) solution was added, followed by heat inactivation of the enzymes at 75 °C for 30 min. Sterile PBS was added to bring the samples up to 800  $\mu$ L in volume: 400  $\mu$ L for phenol/chloroform extraction, and 400  $\mu$ L for kit extraction.

### DNA extraction—phenol/chloroform approach

This protocol is adapted from [Thurber et al. \(2009\)](#), where 40  $\mu$ L of 200X TE buffer (2M tris, 200 mM EDTA, pH 8.5), 440  $\mu$ L formamide, and 10  $\mu$ L UltraPure glycogen (20 mg/mL stock; Thermo Fisher Scientific, Waltham, MA, USA) were added to 400  $\mu$ L purified VLPs and incubated at room temperature for 30 min. Two volume equivalents (approx. 1,780  $\mu$ L) of room temperature 100% ethanol were added to the sample, and DNA was pelleted by centrifugation at 10,000 g for 20 min at 4 °C. The supernatant was carefully removed and discarded, and the crude DNA pellet was washed twice with 1 mL of ice-cold 70% ethanol, centrifuging as above between washes. The final pellet was dried for 5 min at room temperature and resuspended in 567  $\mu$ L of 1X TE buffer (10 mM tris, 1mM EDTA, pH 8). We then added 30  $\mu$ L of 10% (w/v) SDS and 3  $\mu$ L of Proteinase K (20 mg/mL stock; Thermo Fisher Scientific, Waltham, MA, USA) to the sample and briefly vortexed before incubation at 45 °C for 1 h with gentle shaking. After incubation, 100  $\mu$ L of 5M NaCl and 80  $\mu$ L of CTAB buffer (1.1 M NaCl, 450 mM CTAB/cetyltrimethylammonium bromide) were added, the sample vortexed and incubated at 65 °C for 10 min with gentle shaking. One volume equivalent (approx. 780  $\mu$ L) of chloroform:isoamyl alcohol 24:1 (Sigma-Aldrich) was added and mixed well, transferred to a light phase lock gel tube (PLG tube; Quantabio, Beverly, MA, USA), and centrifuged at 12,000 g for 5 min. The aqueous (upper) phase was transferred to a new PLG tube to which another 1 volume equivalent of phenol:chloroform:isoamyl alcohol 25:24:1 (pH 8, Invitrogen) was added, mixed well, and centrifuged as above. After transferring the aqueous (upper) phase to a new PLG tube, this step was repeated with 1 volume equivalent of chloroform:isoamyl alcohol 24:1. The aqueous phase was then recovered, added to a tube containing 550  $\mu$ L (~0.7 volume equivalents) ice-cold 100% isopropanol, and stored overnight at -20 °C for DNA precipitation. The next day, DNA was pelleted by centrifugation at 13,000 g for 15 min at 4 °C. The pellet was washed once with 500  $\mu$ L ice-cold 70% ethanol and pelleted again as above. The final pellet was air-dried at

room temperature, resuspended in 50  $\mu$ L tris buffer (10 mM Tris-Cl, pH 8), and stored in a DNA LoBind tube (Eppendorf, Hamburg, Germany) at  $-20^{\circ}\text{C}$  until library preparation.

### DNA extraction—kit approach

DNA was extracted using the QIAGEN MinElute Virus Spin Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions, with the following modifications to incorporate a larger sample input volume. Fifty  $\mu$ L QIAGEN Protease and 400  $\mu$ L Buffer AL were added to 400  $\mu$ L purified VLPs, vortexed, and incubated at  $56^{\circ}\text{C}$  for 15 min. We then added 500  $\mu$ L 100% ethanol to the sample, vortexed, and incubated at room temperature for 5 min. Half of the sample ( $\sim$ 675  $\mu$ L) was added onto a QIAamp MinElute column, and centrifuged at 6,000 g for 1 min. The filtrate was discarded and the remaining sample was added to the same QIAamp MinElute column, centrifuged as above. The remainder of the protocol remained unchanged from the manufacturer's instructions. DNA was eluted from the column using 50  $\mu$ L Buffer EB (tris buffer; 10 mM Tris-Cl, pH 8; 5-min incubation before elution), and stored in a DNA LoBind tube (Eppendorf, Hamburg, Germany) at  $-20^{\circ}\text{C}$  until library preparation. The included Carrier RNA was not used for any of the DNA extractions.

### Library preparation and sequencing

Extracted vDNA (5  $\mu$ L per sample) from both procedures was quantified using the Qubit 1X dsDNA High Sensitivity kit (Invitrogen). The Qubit instrument was re-calibrated before each use using the included standards. Sequencing libraries were prepared using the Illumina DNA Prep kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, maximizing sample input (30  $\mu$ L) and number of PCR cycles (12 cycles). The final libraries were quantified using the Qubit 1X dsDNA High Sensitivity kit and the Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA). An equimolar pool of libraries was created and sequenced on an Illumina MiSeq instrument with 150 bp paired-end reads (SeqCenter, Pittsburgh, PA, USA).

### Bioinformatic analysis of sequence data

Raw reads were trimmed and filtered with fastp (v0.20.1) ([Chen et al., 2018](#)) using the following criteria: `-detect_adapter_for_pe -q 15 -cut_right -cut_window_size 4 -cut_mean_quality 20 -length_required 100`. Trimmed reads were decontaminated for human and mouse genomic DNA with bowtie2 (v2.4.2) ([Langmead & Salzberg, 2012](#)) using the *Homo sapiens* GRCh38 and *Mus musculus* GRCm39 references, respectively. metaSPAdes (v3.15.4) ([Nurk et al., 2017](#)) was used to conduct *de novo* assembly of each sample individually using default settings. VIBRANT (v1.2.1) ([Kieft, Zhou & Anantharaman, 2020](#)) was used on these filtered contigs to identify viral sequences with the following settings: 3 kb minimum length, virome mode. Viral contigs were dereplicated into viral operational taxonomic units (vOTUs) using BLASTN (v2.14.0) ([Camacho et al., 2009](#)) followed by the anicalc.py and aniclust.py scripts from CheckV ([Nayfach et al., 2021](#)) with the following parameters: 95% average nucleotide identity (ANI) over 85% of the shorter contig's length. Bowtie2 (v2.4.2) ([Langmead & Salzberg, 2012](#)) was used to map

decontaminated trimmed/filtered reads to the resulting dereplicated set of vOTUs, and a coverage summary report was generated with Samtools (v1.13) (Li *et al.*, 2009) using the ‘samtools coverage’ command. A standard vOTU detection threshold was applied prior to subsequent our analyses: mean depth of coverage  $\geq 1X$  and breadth of coverage  $\geq 75\%$  (Roux *et al.*, 2017). iPhoP (v1.3.3) (Roux *et al.*, 2023) was used to computationally predict hosts of vOTUs. R (v4.2.2) was used for all diversity analyses with the following packages: phyloseq (v1.42.0) (McMurdie & Holmes, 2013) and vegan (v2.6.4) (Oksanen *et al.*, 2019). Flextable (v0.8.5) (Gohel & Skintzos, 2024) was used to generate tables.

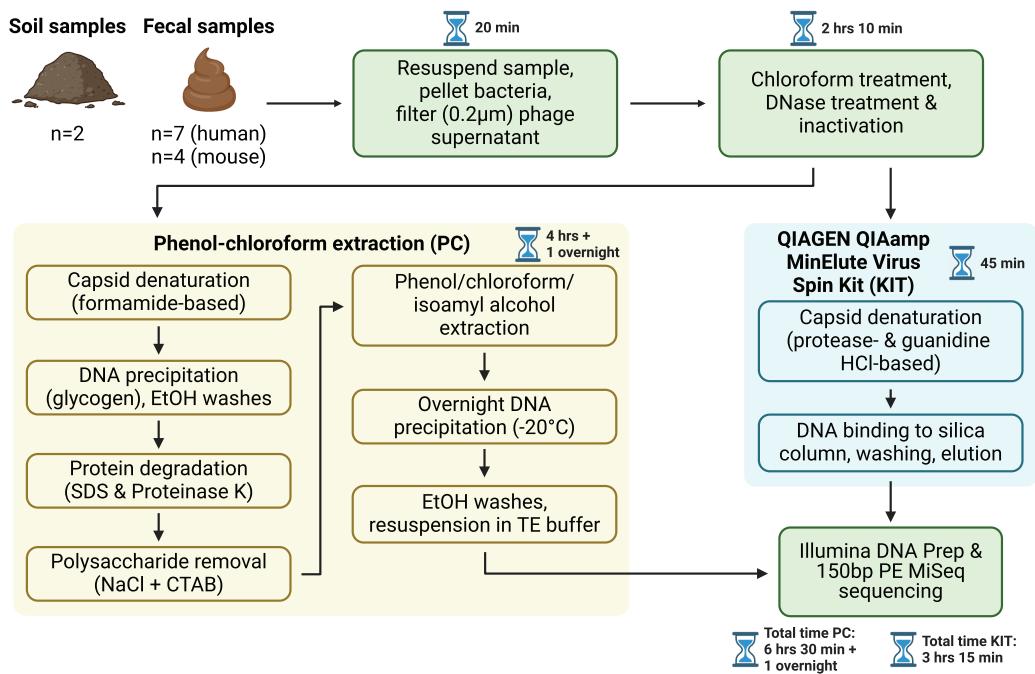
## RESULTS

To compare the kit-based viral DNA extraction protocol (KIT) with the previous standard, phenol-chloroform extractions (PC), we collected samples from a variety of environments which were each processed with both extraction methods: two soil samples, seven human fecal samples, and four mouse fecal samples (Fig. 1). DNA yields were not significantly different across protocols for soil ( $42 \pm 2$  pg/ $\mu$ L KIT,  $47 \pm 32$  pg/ $\mu$ L PC), human fecal ( $249 \pm 141$  pg/ $\mu$ L KIT,  $536 \pm 314$  pg/ $\mu$ L PC) and mouse fecal ( $45 \pm 14$  pg/ $\mu$ L KIT,  $126 \pm 42$  pg/ $\mu$ L PC) samples ( $p > 0.05$ , Wilcoxon signed-rank test).

After sequencing, assembly, and viral detection, a total of 963, 707, and 423 vOTUs were detected in human fecal, mouse fecal, and soil samples, respectively. Of these, most vOTUs were common to both PC and KIT viromes: 875 (91%) of all vOTUs in human fecal viromes, 630 (89%) of all vOTUs in mouse fecal viromes, and 398 (94%) of all vOTUs in soil viromes (Fig. 2A). Few vOTUs (9% of human fecal, 11% of mouse fecal, and 6% of soil vOTUs) were detected exclusively in either PC or KIT viromes (Fig. 2A). Assembly statistics (*e.g.*, total assembly size and N50) and vOTU read recruitment percentages were consistently similar between KIT and PC approaches (Table S1).

We next characterized the alpha- and beta-diversity of our samples, to evaluate compositional differences in viromes which may be due to extraction method. Alpha diversity, measured at the observed richness and Shannon diversity levels, was consistently similar between both extraction methods (mean 255, 317, 200 observed vOTUs, and 3.86, 4.05, 5.10 Shannon index for human fecal, mouse fecal, and soil viromes, respectively, with KIT; and mean 233, 384, 210 observed vOTUs, and 3.82, 4.14, 5.28 Shannon index for human fecal, mouse fecal, and soil viromes, respectively, with PC), with these differences being insignificant (Fig. 2B). Pairwise Bray-Curtis distances were calculated between all samples and plotted on an NMDS plot (Fig. 2C). For each sample, PC and KIT viromes clustered closely together regardless of the environment. A PERMANOVA analysis confirmed that the distances between samples was explained primarily by environment ( $R^2 = 0.363$ ,  $p = 0.001$ ) and the sample itself ( $R^2 = 0.613$ ,  $p = 0.001$ ), rather than the extraction method ( $R^2 = 0.002$ ,  $p = 0.426$ ; Fig. 2C).

Finally, we predicted the bacterial hosts of vOTUs to verify that the predicted hosts align with bacteria associated with the sampled environments. The most abundant soil virome vOTUs with assigned hosts were predicted to infect members of the *Mycobacteriales*, *Flavobacteriales*, and *Streptomycetales*, *Pseudomonadales* orders. In human fecal samples,



**Figure 1** Overview of sample processing pipeline. All samples underwent resuspension in sterile PBS, filtration to remove debris and bacterial or eukaryotic cells, chloroform treatment, and DNase treatment (with subsequent DNase inactivation). Samples were then split to be processed with either the phenol-chloroform extraction (PC) or QIAGEN QIAamp MinElute Virus Spin Kit (KIT) DNA extraction protocols. DNA libraries were prepared with the Illumina DNA Prep kit and sequenced on an Illumina MiSeq instrument with 150 bp paired-end reads. Approximate sample processing times are indicated for each step. The KIT protocol increases sample throughput by reducing protocol complexity and minimizing the number of incubation steps. Created in BioRender (Maurice, 2024).

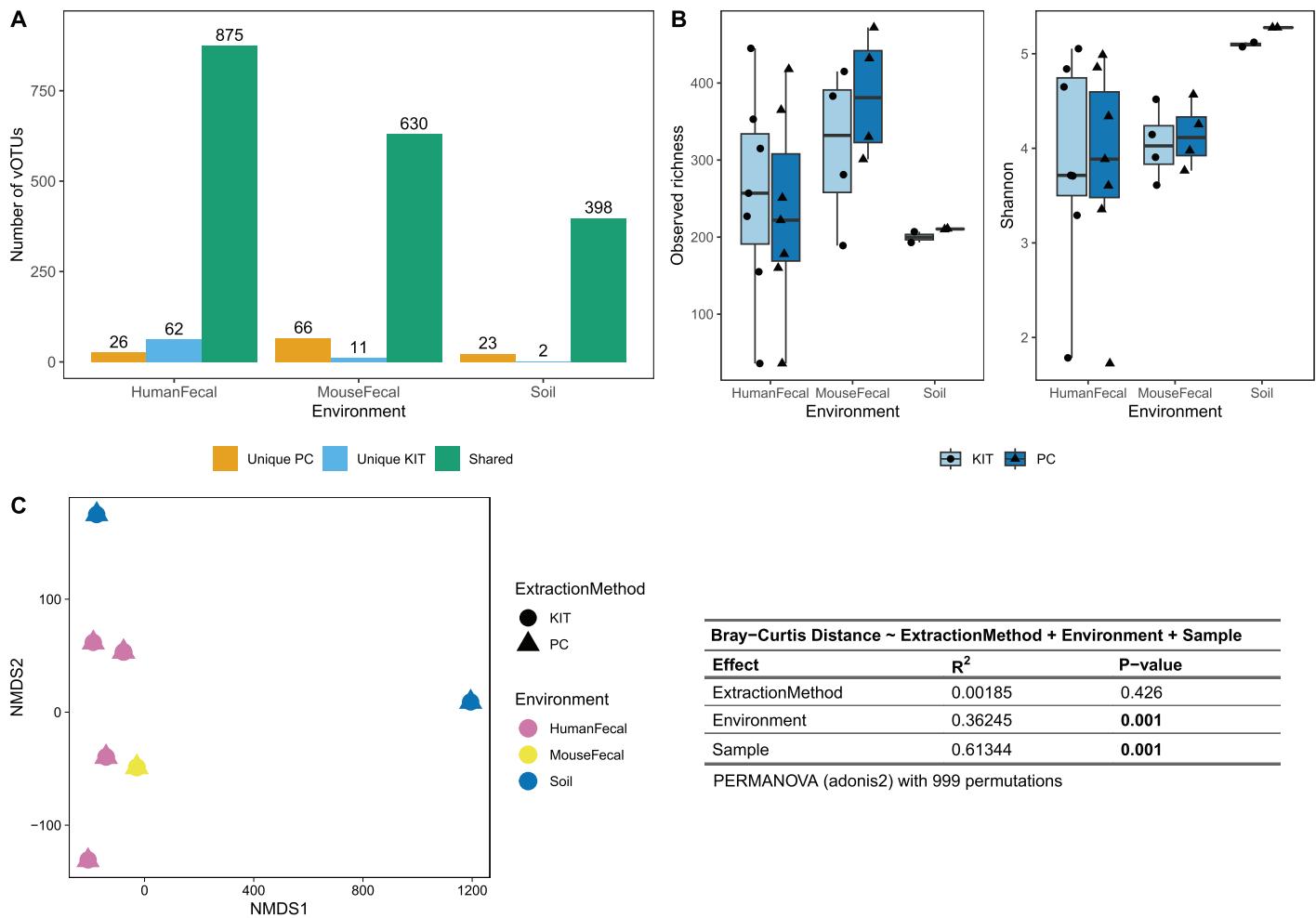
Full-size DOI: 10.7717/peerj.18785/fig-1

the most abundant vOTUs with assigned hosts infect were predicted to infect members of the *Bacteroidales*, *Lactobacillales*, *Oscillospirales*, and *Lachnospirales* orders. In mouse fecal samples, the most abundant vOTUs with assigned hosts were predicted to infect members of the *Lachnospirales*, *Oscillospirales*, *Bacteroidales*, and *Peptostreptococcales* orders.

## DISCUSSION

In this study, we describe a streamlined protocol for virome DNA extractions using a commercially available column-based DNA extraction kit, with some additional upstream steps. Unlike the manufacturer's recommended use for cell-free body fluids, plasma, or serum, we applied this protocol to more complex sample types, including human fecal, mouse fecal, and soil samples. This approach offers several advantages over the gold standard phenol-chloroform protocol, including reduced dependency on toxic organic chemicals, increased throughput, and improved ease-of-use. We compared this kit's performance to that of the phenol-chloroform protocol to evaluate its efficacy across these diverse sample types (Fig. 1). Our results show that the choice of protocol did not significantly affect virome diversity or community composition (Fig. 2).

Most of the assembled vOTUs were detected in both KIT and PC viromes, with few vOTUs unique to either approach (Fig. 2A). The presence of unique vOTUs to either KIT



**Figure 2** PC and KIT extraction methods result in equivalent virome communities. (A) The number of vOTUs detected only within the phenol-chloroform (PC), or kit (KIT) extraction method, or shared between both methods. (B) Observed richness (left) and Shannon diversity (right) of viromes according to environment and extraction method. No statistical differences were observed between extraction methods within any environment (Wilcoxon signed-rank test,  $p > 0.05$ ). (C) Non-metric multidimensional scaling (NMDS) of Bray-Curtis distances of vOTUs between samples, colors representing sample environment and shapes representing extraction method (NMDS stress =  $9.4 \times 10^{-5}$ ). Tests for significant differences in Bray-Curtis dissimilarity were conducted using PERMANOVA (adonis2) with 999 permutations, and summary statistics are reported in the table.

Full-size DOI: 10.7717/peerj.18785/fig-2

or PC viromes did not have a significant effect on community alpha- or beta- diversity (Figs. 2B, 2C). The unique vOTUs were low in abundance (mean unique vOTU abundance of  $0.13 \pm 0.42\%$ ) and made up a small fraction of the overall viromes (mean cumulative unique vOTU abundance of  $1.75 \pm 0.26\%$ ), indicating that both extraction methods were equally good at capturing high-abundance vOTUs. vOTUs host prediction was consistent with descriptions of soil (Graham *et al.*, 2024), adult human gut (Camarillo-Guerrero *et al.*, 2021), and adult mouse gut (Ritz *et al.*, 2024) viromes.

While this kit has been used to extract viral DNA from human fecal samples in other published work, there has been no direct comparison against the phenol-chloroform gold standard, or any other extraction protocol (Bikel *et al.*, 2022). In addition, there are several different experimental steps between both studies. Specifically, we suggest using two

DNases to maximize removal of bacterial and host DNA, inactivating the nucleases with EDTA and a higher incubation temperature (75 °C for 30 min) before proceeding directly to extraction with the kit, and allowing the column-bound DNA to sit with the elution buffer for 5 min. All these steps aim to maximize DNA yield and remove non-viral DNA.

Despite its advantages, the kit approach has a few limitations. This kit still includes a chloroform step upstream of extraction which destroys the membrane of enveloped viruses ([Callanan et al., 2021](#)). Cost may be another factor, as kits are generally more expensive than the different reagents used in the phenol-chloroform method. Furthermore, reliance on proprietary reagents may be an issue if the manufacturer changes their formulation. Finally, while kits offer high consistency, they may not always result in the highest yield for all sample types, as we report here for soil samples. Researchers specifically targeting low-abundance virome members may still wish to use the phenol-chloroform method. For researchers working with low biomass samples, it is important to test and validate the kit's performance with this sample type and include reagent controls.

## ACKNOWLEDGEMENTS

ChatGPT (OpenAI Inc., San Francisco, CA, USA) was used for editing purposes to improve clarity of a few sentences. The sentence meaning was not altered, and there were no changes to any of our presented data, facts, or conclusions by using this tool.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was funded by a Canadian Institutes of Health Research Canada Project grant (PJT-175065) and an NSERC Discovery grant (RGPIN 2023-04216) to Corinne F. Maurice. Corinne F. Maurice is a Tier 2 Canada Research Chair in gut microbial interactions. Michael Shamash is supported by the Canadian Institutes of Health Research Canada Graduate Scholarship to Honor Nelson Mandela (CIHR CGS-D; #DF2-187718), and the Fonds de recherche du Québec-Santé: Bourse de formation au doctorat (FRQS; #311071). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Canadian Institutes of Health Research Canada: PJT-175065, CIHR CGS-D; #DF2-187718.

NSERC Discovery: RGPIN 2023-04216.

Fonds de recherche du Québec-Santé: Bourse de formation au doctorat: FRQS; #311071.

### Competing Interests

Corinne F. Maurice is an Academic Editor for PeerJ.

## Author Contributions

- Michael Shamash conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Saniya Kapoor conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.
- Corinne F. Maurice conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

## Human Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

Research Ethics Office (IRB) of the Faculty of Medicine and Health Sciences at McGill University.

## Animal Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

Comparative Medicine and Animal Resources Centre Facility Animal Care Committee (CMARC FACC) at McGill University.

## Data Availability

The following information was supplied regarding data availability:

The virome sequencing reads (human and mouse sequences removed) are available at NCBI SRA: [PRJNA1125394](https://www.ncbi.nlm.nih.gov/sra/PRJNA1125394).

The code and scripts are available at Zenodo: Michael Shamash. (2024). mshamash/vdna\_protocol\_manuscript: Manuscript\_v2 (Version v2). Zenodo. <https://doi.org/10.5281/zenodo.14236083>.

## Supplemental Information

Supplemental information for this article can be found online at [http://dx.doi.org/10.7717/peerj.18785#supplemental-information](https://dx.doi.org/10.7717/peerj.18785#supplemental-information).

## REFERENCES

**Bikel S, Gallardo-Becerra L, Cornejo-Granados F, Ochoa-Leyva A. 2022.** Protocol for the isolation, sequencing, and analysis of the gut phageome from human fecal samples. *STAR Protocols* **3**(1):101170 DOI [10.1101/j.xpro.2022.101170](https://doi.org/10.1101/j.xpro.2022.101170).

**Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018.** Phage puppet masters of the marine microbial realm. *Nature Microbiology* **3**(7):754–766 DOI [10.1038/s41564-018-0166-y](https://doi.org/10.1038/s41564-018-0166-y).

**Brunse A, Deng L, Pan X, Hui Y, Castro-Mejía JL, Kot W, Nguyen DN, Secher JB-M, Nielsen DS, Thymann T. 2021.** Fecal filtrate transplantation protects against necrotizing enterocolitis. *The ISME Journal* **16**(3):686–694 DOI [10.1038/s41396-021-01107-5](https://doi.org/10.1038/s41396-021-01107-5).

**Callanan J, Stockdale SR, Shkorporov A, Draper LA, Ross RP, Hill C. 2021.** Biases in viral metagenomics-based detection, cataloguing and quantification of bacteriophage genomes in human faeces, a review. *Microorganisms* **9**(3):524 DOI [10.3390/microorganisms9030524](https://doi.org/10.3390/microorganisms9030524).

**Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421 DOI [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421).

**Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. 2021.** Massive expansion of human gut bacteriophage diversity. *Cell* **184**(4):1098–1109 DOI [10.1016/j.cell.2021.01.029](https://doi.org/10.1016/j.cell.2021.01.029).

**Chen S, Zhou Y, Chen Y, Gu J. 2018.** fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17):i884–i890 DOI [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560).

**Gohel D, Skintzos P. 2024.** flextable: functions for tabular reporting. Available at <https://davidgohel.github.io/flextable/>.

**Göller PC, Haro-Moreno JM, Rodriguez-Valera F, Loessner MJ, Gómez-Sanz E. 2020.** Uncovering a hidden diversity: optimized protocols for the extraction of dsDNA bacteriophages from soil. *Microbiome* **8**(1):17 DOI [10.1186/s40168-020-0795-2](https://doi.org/10.1186/s40168-020-0795-2).

**Graham EB, Camargo AP, Wu R, Neches RY, Nolan M, Paez-Espino D, Kyrpides NC, Jansson JK, McDermott JE, Hofmockel KS, the Soil Virosphere Consortium, Blanchard JL, Liu XJA, Rodrigues JLM, Freedman ZB, Baldrian P, Stursova M, DeAngelis KM, Lee S, Godoy-Vitorino F, Yeoh YK, Cadillo-Quiroz H, Tringe SG, Chauhan A, Cowan DA, Van Goethem MW, Woyke T, Dove NC, Konstantinidis KT, Juenger TE, Hart SC, Myrold DD, Onstott TC, Bohannan BJM, Schmer MR, Palmer NA, Nüsslein K, Makhalanyane TP, Dynarski KA, Taş N, Nicol GW, Hazard C, Scully ED, Jain KR, Madamwar D, Bissett A, Constant P, Oliveira RS, Takacs-Vesbach C, Cregger MA, Carrell AA, Klingeman DM, Pietrasiak N. 2024.** A global atlas of soil viruses reveals unexplored biodiversity and potential biogeochemical impacts. *Nature Microbiology* **9**(7):1873–1883 DOI [10.1038/s41564-024-01686-x](https://doi.org/10.1038/s41564-024-01686-x).

**Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, Jang HB, Andreopoulos B, Brewer HM, Glavina del Rio T, Adkins JN, Paul S, Sullivan MB, Duhaime MB. 2020.** Phage-specific metabolic reprogramming of virocells. *ISME Journal* **14**(4):881–895 DOI [10.1038/s41396-019-0580-z](https://doi.org/10.1038/s41396-019-0580-z).

**Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019.** Dynamic modulation of the gut microbiota and metabolome by bacteriophages in a mouse model. *Cell Host and Microbe* **25**(6):803–814.e5 DOI [10.1016/j.chom.2019.05.001](https://doi.org/10.1016/j.chom.2019.05.001).

**Khan Mirzaei M, Xue J, Costa R, Ru J, Schulz S, Taranu ZE, Deng L. 2021.** Challenges of studying the human virome—relevant emerging technologies. *Trends in Microbiology* **29**(2):171–181 DOI [10.1016/j.tim.2020.05.021](https://doi.org/10.1016/j.tim.2020.05.021).

**Kieft K, Zhou Z, Anantharaman K. 2020.** VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**(1):90 DOI [10.1186/s40168-020-00867-0](https://doi.org/10.1186/s40168-020-00867-0).

**Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**(4):357–359 DOI [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16):2078–2079 DOI [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).

**Liang G, Bushman FD. 2021.** The human virome: assembly, composition and host interactions. *Nature Reviews Microbiology* **19**(8):514–527 DOI [10.1038/s41579-021-00536-5](https://doi.org/10.1038/s41579-021-00536-5).

**Maurice CF.** 2024. Figure 1. Overview of sample processing pipeline. Created in BioRender. Available at <https://Biorender.com/f45n257>.

**McMurdie PJ, Holmes S.** 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* **8**(4):e61217 DOI [10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217).

**Nayfach S, Camargo AP, Schulz F, Eloe-fadros E, Roux S, Kyrpides NC.** 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* **39**(5):578–585 DOI [10.1038/s41587-020-00774-7](https://doi.org/10.1038/s41587-020-00774-7).

**Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* **27**(5):824–834 DOI [10.1101/gr.213959.116](https://doi.org/10.1101/gr.213959.116).

**Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H.** 2019. vegan: community ecology package. Available at <https://github.com/vegandevs/vegan/>.

**Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI.** 2013. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proceedings of the National Academy of Sciences of the United States of America* **110**(50):20236–20241 DOI [10.1073/pnas.1319470110](https://doi.org/10.1073/pnas.1319470110).

**Ritz NL, Draper LA, Bastiaanssen TFS, Turkington CJR, Peterson VL, Van De Wouw M, Vlckova K, Fülling C, Guzzetta KE, Burokas A, Harris H, Dalmasso M, Crispie F, Cotter PD, Shkoporov AN, Moloney GM, Dinan TG, Hill C, Cryan JF.** 2024. The gut virome is associated with stress-induced changes in behaviour and immune responses in mice. *Nature Microbiology* **9**(2):359–376 DOI [10.1038/s41564-023-01564-y](https://doi.org/10.1038/s41564-023-01564-y).

**Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A.** 2023. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLOS Biology* **21**(4):e3002083 DOI [10.1371/journal.pbio.3002083](https://doi.org/10.1371/journal.pbio.3002083).

**Roux S, Emerson JB.** 2022. Diversity in the soil virosphere: to infinity and beyond? *Trends in Microbiology* **30**(11):1025–1035 DOI [10.1016/j.tim.2022.05.003](https://doi.org/10.1016/j.tim.2022.05.003).

**Roux S, Emerson JB, Eloe-Fadros E, Sullivan MB.** 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**(1):e3817 DOI [10.7717/peerj.3817](https://doi.org/10.7717/peerj.3817).

**Shamash M, Kapoor S, Maurice CF.** 2024. Optimization of an improved, time-saving, and scalable, protocol for the extraction of DNA from diverse viromes. *BioRxiv* DOI [10.1101/2024.06.26.599709](https://doi.org/10.1101/2024.06.26.599709).

**Shamash M, Maurice CF.** 2021. Phages in the infant gut: a framework for virome development during early life. *The ISME Journal* **16**(2):323–330 DOI [10.1038/s41396-021-01090-x](https://doi.org/10.1038/s41396-021-01090-x).

**Shkoporov AN, Hill C.** 2019. Bacteriophages of the Human Gut: the “Known Unknown” of the microbiome. *Cell Host and Microbe* **25**(2):195–209 DOI [10.1016/j.chom.2019.01.017](https://doi.org/10.1016/j.chom.2019.01.017).

**Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, McDonnell SA, Nolan JA, Sutton TDS, Dalmasso M, McCann A, Ross RP, Hill C.** 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**(1):68 DOI [10.1186/s40168-018-0446-z](https://doi.org/10.1186/s40168-018-0446-z).

**Shkoporov AN, Stockdale SR, Lavelle A, Kondova I, Heuston C, Upadrasta A, Khokhlova EV, van der Kamp I, Ouwerling B, Draper LA, Langermans JAM, Paul Ross R, Hill C.** 2022. Viral biogeography of the mammalian gut and parenchymal organs. *Nature Microbiology* **7**(8):1301–1311 DOI [10.1038/s41564-022-01178-w](https://doi.org/10.1038/s41564-022-01178-w).

**Sinha A, Li Y, Mirzaei MK, Shamash M, Samadfam R, King IL, Maurice CF. 2022.**  
Transplantation of bacteriophages from ulcerative colitis patients shifts the gut bacteriome and exacerbates the severity of DSS colitis. *Microbiome* **10**(1):105 DOI [10.1186/s40168-022-01275-2](https://doi.org/10.1186/s40168-022-01275-2).

**Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009.** Laboratory procedures to generate viral metagenomes. *Nature Protocols* **4**(4):470–483 DOI [10.1038/nprot.2009.10](https://doi.org/10.1038/nprot.2009.10).