

Lawrence Berkeley National Laboratory

LBL Publications

Title

Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils

Permalink

<https://escholarship.org/uc/item/0vb5k8qg>

Journal

PeerJ, 7(7)

ISSN

2167-8359

Authors

Trubl, Gareth

Roux, Simon

Solonenko, Natalie

et al.

Publication Date

2019

DOI

10.7717/peerj.7265

Peer reviewed

Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils

Gareth Trubl^{1,4}, Simon Roux², Natalie Solonenko¹, Yueh-Fen Li¹, Benjamin Bolduc¹, Josué Rodríguez-Ramos^{1,5}, Emiley A. Eloë-Fadrosch², Virginia I. Rich¹ and Matthew B. Sullivan^{1,3}

¹ Department of Microbiology, The Ohio State University, Columbus, OH, United States of America

² United States Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA, United States of America

³ Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, United States of America

⁴ Current affiliation: Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, United States of America

⁵ Current affiliation: Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, United States of America

ABSTRACT

Soils impact global carbon cycling and their resident microbes are critical to their biogeochemical processing and ecosystem outputs. Based on studies in marine systems, viruses infecting soil microbes likely modulate host activities via mortality, horizontal gene transfer, and metabolic control. However, their roles remain largely unexplored due to technical challenges with separating, isolating, and extracting DNA from viruses in soils. Some of these challenges have been overcome by using whole genome amplification methods and while these have allowed insights into the identities of soil viruses and their genomes, their inherent biases have prevented meaningful ecological interpretations. Here we experimentally optimized steps for generating quantitatively-amplified viral metagenomes to better capture both ssDNA and dsDNA viruses across three distinct soil habitats along a permafrost thaw gradient. First, we assessed differing DNA extraction methods (PowerSoil, Wizard mini columns, and cetyl trimethylammonium bromide) for quantity and quality of viral DNA. This established PowerSoil as best for yield and quality of DNA from our samples, though ~1/3 of the viral populations captured by each extraction kit were unique, suggesting appreciable differential biases among DNA extraction kits. Second, we evaluated the impact of purifying viral particles after resuspension (by cesium chloride gradients; CsCl) and of viral lysis method (heat vs bead-beating) on the resultant viromes. DNA yields after CsCl particle-purification were largely non-detectable, while unpurified samples yielded 1–2-fold more DNA after lysis by heat than by bead-beating. Virome quality was assessed by the number and size of metagenome-assembled viral contigs, which showed no increase after CsCl-purification, but did from heat lysis relative to bead-beating. We also evaluated sample preparation protocols for ssDNA virus recovery. In both CsCl-purified and non-purified samples, ssDNA viruses were successfully recovered by using the Accel-NGS 1S Plus Library Kit. While ssDNA viruses were identified in all three soil types, none were identified in the samples that used bead-beating, suggesting this lysis method may impact recovery. Further, 13 ssDNA vOTUs were identified

Submitted 21 March 2019

Accepted 7 June 2019

Published 4 July 2019

Corresponding authors

Virginia I. Rich,
virginia.isabel.rich@gmail.com,
rich.270@osu.edu
Matthew B. Sullivan,
mbsulli@gmail.com

Academic editor

Alexandre Anesio

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.7265

© Copyright
2019 Trubl et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

compared to 582 dsDNA vOTUs, and the ssDNA vOTUs only accounted for ~4% of the assembled reads, implying dsDNA viruses were dominant in these samples. This optimized approach was combined with the previously published viral resuspension protocol into a sample-to-virome protocol for soils now available at protocols.io, where community feedback creates 'living' protocols. This collective approach will be particularly valuable given the high physicochemical variability of soils, which will may require considerable soil type-specific optimization. This optimized protocol provides a starting place for developing quantitatively-amplified viromic datasets and will help enable viral ecogenomic studies on organic-rich soils.

Subjects Ecology, Microbiology, Soil Science, Virology

Keywords Soil viruses, Viromes, DNA extraction, Organics, Microbiology, ssDNA viruses, dsDNA viruses, Viromics

INTRODUCTION

Optimization of experimental methods to generate viral-particle metagenomes (viromes) from aquatic samples has enabled robust ecological analyses of marine viral communities (reviewed in [Brum & Sullivan, 2015](#); [Sullivan, Weitz & Wilhelm, 2017](#); [Hayes et al., 2017](#)). In parallel, optimization of informatics methods to identify and characterize viral sequences has advanced viral sequence recovery from microbial-cell metagenomes, as well as virome analyses ([Edwards & Rohwer, 2005](#); [Wommack et al., 2012](#); [Roux et al., 2015](#); [Brum & Sullivan, 2015](#); [Roux et al., 2016](#); [Bolduc et al., 2017](#); [Ren et al., 2017](#); [Amgarten et al., 2018](#); [Gregory et al., 2019](#)). Application of these methods with large-scale sampling ([Brum et al., 2015](#); [Roux et al., 2016](#)) has revealed viruses as important members of ocean ecosystems acting through host mortality, gene transfer, and direct manipulation of key microbial metabolisms including photosynthesis and central carbon metabolism during infection, via expression of viral-encoded 'auxiliary metabolic genes' (AMGs). More recently, the abundance of several key viral populations was identified as the best predictor of global carbon (C) flux from the surface oceans to the deep sea ([Guidi et al., 2016](#)). This finding suggests that viruses may play a role beyond the viral shunt and help form aggregates that may store C long-term. These discoveries in the oceans have caused a paradigm shift in how we view viruses: no longer simply disease agents, it is now clear that viruses play central roles in ocean ecosystems and help regulate global nutrient cycling.

In soils, however, viral roles are not so clear. Soils contain more C than all the vegetation and the atmosphere combined (between 1500–2400 gigatons; [Lehmann & Kleber, 2015](#)), and soil viruses likely also impact C cycling, as their marine counterparts do. However, our knowledge about soil viruses remains limited due to the dual challenges of separating viruses from the highly heterogeneous soil matrix, while minimizing DNA amplification inhibitors (e.g., humics; reviewed in [Williamson et al., 2017](#)). For these reasons, most soil viral work is limited to direct counts and morphological analyses (i.e., microscopy observations), from which we have learned (i) there are 10^7 – 10^9 virus-like particles/g soil, (ii) viral morphotype richness is generally higher in soils than in aquatic ecosystems,

and (iii) viral abundance correlates with soil moisture, organic matter content, pH, and microbial abundance (reviewed in [Williamson et al., 2017](#); [Narr et al., 2017](#)). The minimal collective metagenomic data for soils suggests that genetic diversity of soil viruses far exceeds that of other environments for which virome data are available and these viral communities are localized in that viruses form habitat-specific groups ([Fierer et al., 2007](#); [Kim et al., 2008](#); [Srinivasiah et al., 2015](#); [Reavy et al., 2015](#); [Zablocki, Adriaenssens & Cowan, 2016](#); [Trubl et al., 2018](#); [Emerson et al., 2018](#); [Green et al., 2018](#)). Thus, while sequencing data for soil viruses is not as robust as it is in aquatic environments, such high particle counts and patterns suggest that viruses also play important ecosystems roles in soils.

The first barrier to obtaining sequence data for soil viruses is simply separating the viral particles from the soil matrix, and then accessing their nucleic acids. Viral resuspension is unlikely to be universally solvable with a single approach due to high variability of soil properties (e.g., mineral content and cation exchange capacity) impacting virus-soil interactions. There have been independent efforts to optimize virus resuspension methods tailored to specific soil types, and employing a range of resuspension methods (reviewed in [Narr et al., 2017](#); [Pratama & Van Elsas, 2018](#)). Once viruses are separated, extraction of their DNA must surmount the additional challenges of co-extracted inhibitors (hampering subsequent molecular biology, as previously described for soil microbes; [Narayan et al., 2016](#); [Zielińska et al., 2017](#)), and low DNA yields.

Extracting viral nucleic acid from soils typically results in very low DNA yields, requiring amplification prior to sequencing. Amplification of viral nucleic acid is necessary because the high heterogeneous nature of soil prevents any meaningful viral ecology if DNA yield is increased by increasing the number of virus extractions and pooling the concentrate (micro-scale variation and the need for smaller-scale sampling reviewed in [Fierer, 2017](#)). Two widely used methods to amplify viral nucleic acid are multiple displacement amplification (MDA; 'whole genome' amplification using the phi29 polymerase) and random priming-mediated sequence-independent single-primer amplification (RP-SISPA). Both allow qualitative observations of viral sequences but preclude quantitative ecological inferences. Specifically, MDA causes dramatic shifts in relative abundances of DNA templates, which impact subsequent estimates of viral populations diversity, and, most dramatically, over-amplify ssDNA viruses ([Binga, Lasken & Neufeld, 2008](#); [Yilmaz, Allgaier & Hugenholtz, 2010](#); [Kim, Whon & Bae, 2013](#); [Marine et al., 2014](#)). RP-SISPA is biased towards the most abundant viruses or largest genomes, and leads to uneven coverage along the amplified genomes ([Karlsson, Belák & Granberg, 2013](#)). More recently, quantitative amplification methods have emerged that use transposon-mediated tagmentation (Nextera, for dsDNA; [Trubl et al., 2018](#); [Segobola et al., 2018](#)) or acoustic shearing to fragment and a custom adaptase (Accel-NGS 1S Plus, for dsDNA and ssDNA; ([Roux et al., 2016](#); [Rosario et al., 2018](#))) to ligate adapters to DNA templates, before PCR amplification is used to obtain enough material for sequencing. These approaches have successfully amplified as little as 1 picogram (Nextera XT; [Rinke et al., 2016](#)) and 100 nanograms (Accel-NGS 1S Plus; [Laurie Kurihara et al., 2014](#)) of input DNA for viromes while maintaining the relative abundances of templates.

We previously optimized a viral resuspension method for three peat soil habitats (palsa, bog, and fen, spanning a permafrost thaw gradient; [Trubl et al., 2016](#)). Given emerging quantitative low-input DNA library construction options, we sought here to characterize how the choice of methods for viral particle purification, lysis and DNA extraction impacted viral DNA yield and quality, and resulting virome diversity. The objectives of this work were to (1) optimize the generation of viromes from soils and (2) evaluate the capability of the Accel-NGS 1S Plus kit to quantitatively amplify ssDNA and dsDNA viruses from soils. We conducted two independent experiments testing three different DNA extraction methods (Experiment 1), and then two virion lysis methods with and without further particle purification (Experiment 2). Because microscopy is not sufficient for assessing the presence of non-viral particles, we employed a combination of qPCR and virus-specific bioinformatics to evaluate the success of this protocol to yield genuine viral genomes. Quantitative soil viromes for both ssDNA and dsDNA viruses were generated, enabling a robust comparison of the different protocols tested.

METHODS

Field site and sampling

Stordalen Mire (68.35°N, 19.05°E) is a peat plateau in Arctic Sweden in a zone of discontinuous permafrost. Peat depth ranges from 1–3 m ([Johansson et al., 2006](#); [Normand et al., 2017](#)). Habitats broadly span three stages of permafrost thaw: palsa (drained soil, dominated by small shrubs, and underlain by intact permafrost; pH ~6.50), bog (partially inundated peat, dominated by Sphagnum moss, and underlain by partially thawed permafrost; pH ~4.10), and fen (fully inundated peat, dominated by sedges, and with no detectable permafrost at <1 m; pH ~5.70) (further described in [Hodgkins et al., 2014](#)). These soils vary chemically ([Hodgkins et al., 2014](#); [Normand et al., 2017](#); [Wilson et al., 2017](#)), hydraulically ([Christensen et al., 2004](#); [Malmer et al., 2005](#); [Olefeldt et al., 2012](#); [Jonasson et al., 2012](#)), and biologically ([Mondav et al., 2014](#); [McCalley et al., 2014](#); [Mondav et al., 2017](#); [Woodcroft et al., 2018](#)), creating three distinct habitats with increasing organic matter lability with permafrost thaw. Soil was collected with an 11 cm-diameter custom circular push corer at palsa sites, and with a 10 cm × 10 cm square Wardenaar corer (Eijkelkamp, The Netherlands) at the bog and fen sites. Three cores from each habitat were processed using clean techniques described previously ([Trubl et al., 2016](#)) and cut in five-centimeter increments from 1–40 cm for palsa and 1–80 cm for bog and fen cores. Samples were flash-frozen in liquid nitrogen and kept at –80 °C until processing. The sampled palsa, bog, and fen habitats were directly adjacent, such that all cores were collected within a 120 m radius. For this work, viruses were analyzed from 20–24 cm deep peat, from three cores at each of the three habitats. For Experiment 1 (DNA extraction), 18 samples were used (9 bog and 9 fen), with 10 ± 1 g of soil per sample. For Experiment 2 (virion lysis and purification), 36 samples were used (12 palsa, 12 bog, and 12 fen) with 7.5 ± 1 g of soil per sample.

Experiment 1: optimizing DNA extraction

Viruses were resuspended using a previously optimized method for these soils (Trubl *et al.*, 2016) with minor adjustments. Briefly, 10 ml of a 1% potassium citrate resuspension buffer amended with 10% phosphate buffered-saline and 150 mM magnesium sulfate was added to 10 ± 0.5 g peat (AKC' buffer). Viruses were physically dispersed via 1 min of vortexing, 30 s of manual shaking, and then 15 min of shaking at 400 rpm at 4 °C. The samples were then centrifuged for 20 min at 1, 500×g at 4 °C to pellet debris, and the supernatant was transferred to new tubes. The resuspension steps above were repeated two more times and the supernatants were combined, and then filtered through a 0.2 μm polyethersulfone membrane filter to remove particles and cells and transferred into a new 50 ml tube. The filtrate was then purified via overnight treatment with DNase I (Kunitz units; ThermoFisher, Waltham, Massachusetts) at a 1:10 dilution at 4 °C, inactivated by adding a final concentration of 10 mM EDTA and EGTA and mixing for 1 h. All viral particles were further purified by CsCl density gradients, established with five CsCl density layers of ρ 1.2, 1.3, 1.4, 1.5, and 1.65 g/cm³; we included a 1.3 g/cm³ CsCl layer to collect ssDNA viruses (Thurber *et al.*, 2009). After density gradient centrifugation of the viral particles, we collected and pooled the 1.3–1.52 g/cm³ range from the gradient for viral DNA extraction. The viral DNA was extracted (same elution volume) using one of three methods: Wizard mini columns (Wizard; Promega, Madison, WI, products A7181 and A7211), cetyl trimethylammonium bromide (CTAB; Porebski, Bailey & Baum, 1997), or modified DNeasy PowerSoil DNA extraction kit (C3 reagent was 1/3 of working volume and C4 reagent was 1.5× working volume) with heat lysis (10 min incubation at 70 °C, vortexing for 5 s, and 5 min more of incubation at 70 °C) (PowerSoil; Qiagen, Hilden, Germany, product 12888). The extracted DNA was further cleaned up with AMPure beads (Beckman Coulter, Brea, CA, product A63881). DNA purity was assessed with a Nanodrop 8000 spectrophotometer (Implen GmbH, Germany) by the reading of A260/A280 and A260/A230, and quantified using a Qubit 3.0 fluorometer (Invitrogen, Waltham, Massachusetts). DNA sequencing libraries were prepared using Swift Accel-NGS 1S Plus DNA Library Kit (Swift BioSciences, Washtenaw County, Michigan), and libraries were determined to be 'successful' if there was a smooth peak on the Bioanalyzer with average fragment size of <1kb (200–800 bp ideal) and minimal-to-no secondary peak at ~200 bp (representing concatenated adapters) (Fig. S1), and <20 PCR cycles were required for sequencing. Six libraries were successful (two from bog and four from fen) and required 15 PCR cycles. The successful libraries were sequenced using Illumina HiSeq (300 million reads, 2 × 100 bp paired-end) at JP Sulzberger Columbia Genome Center.

Experiment 2: optimizing particle lysis and purification

Viromes were generated as in Experiment 1 with minor changes. First, viruses were resuspended as described for Experiment 1, except half of the samples were not purified with CsCl density gradient centrifugation. This was to follow-up on our previous work that suggested CsCl resulted in potentially a major loss of viruses (Trubl *et al.*, 2016). Second, DNA was extracted from all samples using the PowerSoil method, but the physical method of particle lysis was tested by half of the samples undergoing the standard heat

lysis as above and the other half undergoing the alternative PowerSoil bead-beating step (with 0.7 mm garnet beads). Third, the extracted DNA was further cleaned up with DNeasy PowerClean Pro Cleanup Kit (Qiagen, Hilden, Germany, product 12997), instead of AMPure beads. Assessment of microbial contamination was done via qPCR (pre and post-cleanup) with primer sets 1406f (5'-GYACWCACCGCCCGT-3') and 1525r (5'-AAGGAGGTGWTCCARCC-3') on 5 μ l of sample input to amplify bacterial and archaeal 16S rRNA genes as previously described ([Woodcroft et al., 2018](#)). Finally, the 12 *palsa* samples were sequenced at the Joint Genome Institute (JGI; Walnut creek, CA), where library preparation was performed using the Accel-NGS 1S Plus kit. All viromes required 20 PCR cycles, except –CsCl, bead-beating which required 18. All libraries were sequenced using the Illumina HiSeq-2000 1TB platform (2 \times 151 bp paired-end).

Bioinformatics and statistics

The same informatics and statistics approaches were applied to viromes from Experiments 1 and 2. The sequences were quality-controlled using Trimmomatic ([Bolger, Lohse & Usadel, 2014](#)), adaptors were removed, reads were trimmed as soon as the average per-base quality dropped below 20 on 4 nt sliding windows, and reads shorter than 50 bp were discarded, with an additional 10 bp removed from the beginning of read pair one and the end of read pair two to remove the low complexity tail specific to the Accel-NGS 1S Plus kit, per the manufacturer's instruction. Reads were assembled using SPAdes ([Bankevich et al., 2012](#); single-cell option, and k-mers 21, 33, and 55), and the contigs were processed with VirSorter to distinguish viral from microbial contigs (virome decontamination mode; [Roux et al., 2015](#)).

Contigs that were selected as VirSorter categories 1 and 2 were used to identify dsDNA viral contigs (as in [Trubl et al., 2018](#)). ssDNA viruses, due to short genomes and highly divergent hallmark genes, can frequently be missed by automatic viral sequence identification tools (e.g., VirSorter from [Roux et al., 2015](#) or VirFinder in [Ren et al., 2017](#)). We therefore applied a two-step approach to ssDNA identification. First, we identified circular contigs that matched ssDNA marker genes from the PFAM database (Viral_Rep and Phage_F domains), using hmmsearch ([Eddy, 2009](#); HMMER v3; cutoffs: score \geq 50 and e-value \leq 0.001). This identified four Phage_F-encoding and five Viral_Rep-encoding circular contigs, i.e., presumed complete genomes. Second, 2 new HMM profiles were generated, using the protein sequences from the nine identified circular viral contigs, and used to search (hmmsearch with the same cutoffs) the viromes' predicted proteins. This resulted in a final set of 23 predicted ssDNA contigs identified across nine viromes ([Table S1](#)).

The viral contigs were clustered at 95% average nucleotide identify (ANI) across 85% of the contig ([Roux et al., 2019a](#)) using nucmer ([Delcher, Salzberg & Phillippy, 2003](#)). The same contigs were also compared by BLAST to a pool of potential laboratory contaminants (i.e., Enterobacteria phage PhiX17, Alpha3, M13, Cellulophaga baltica phages, and Pseudoalteromonas phages), and any contigs matching a potential contaminant at more than 95% ANI across 80% of the contig were removed. Viral operational taxonomic units (vOTUs) were defined as non-redundant (i.e., post-clustering) viral contigs >10

kb for dsDNA viruses (from VirSorter categories 1 or 2; *Roux et al., 2015*) and circular contigs from 4–8 kb for Microviridae viruses or 1–5 kb for circular replication-associated protein (Rep)-encoding ssDNA (CRESS DNA) viruses. The vOTUs represent populations that are likely species-level taxa and there is extensive literature context supporting this new standard terminology, which is summarized in a recent consensus paper (*Roux et al., 2019a; Roux et al., 2019b*). The relative abundance of vOTUs was estimated based on post-QC reads mapping at $\geq 90\%$ ANI and covering $>10\%$ of the contig (*Paez-Espino et al., 2016; Roux et al., 2019a; Roux et al., 2019b*) using Bowtie2 (*Langmead & Salzberg, 2012*). Figures were generated with R, using packages Vegan for diversity (*Oksanen et al., 2016*) and ggplot2 (*Wickham, 2016*) or pheatmap (*Kolde, 2015*) for heatmaps. Hierarchical clustering (function pvcust; method.dist="euclidean" and method.hclust="complete") was conducted on Bray-Curtis dissimilarity matrices using 1,000 bootstrap iterations and only the approximately unbiased (AU) bootstrap values were reported.

Data availability

The 18 viromes from Experiments 1 and 2 are available at the IsoGenie project database under data downloads at <https://isogenie.osu.edu/> and at CyVerse (<https://www.cyverse.org/>) file path /iplant/home/shared/iVirus/Trubl_Soil_Viromes. Data was processed using The Ohio Supercomputer Center (*Ohio Supercomputer Center, 1987*). The final optimized protocol can be accessed here: <https://www.protocols.io/view/soil-viral-extraction-protocol-for-ssdna-amp-dsdna-tzzep76>.

RESULTS AND DISCUSSION

Two independent experiments were performed to optimize the generation of quantitatively-amplified viromes from soil samples (*Fig. 1*). Experiment 1 evaluated three different DNA extraction methods for DNA yield, purity, and successful virome generation on the challenging humic-laden bog and fen soils. Experiment 2 compared two viral particle purification methods (with or without CsCl) and two virion lysis methods (heat vs bead-beating), for DNA yield, microbial DNA contamination, and successful virome generation for all three site habitats (palsa, bog and fen). An optimized virome generation protocol was determined for these palsa, bog and fen soils.

Experiment 1: different DNA extraction methods display variable efficiencies and recover distinct vOTUs

In Experiment 1, three DNA extraction methods were evaluated for DNA yield and purity: PowerSoil DNA extraction kits, Wizard mini columns, and a classic molecular biological approach using cetyl trimethylammonium bromide (CTAB). The PowerSoil kit was designed for humic-rich soils, which dominate our site (*Hodgkins et al., 2014; Normand et al., 2017*), and has performed well previously for viral samples (*Iker et al., 2013*). Wizard mini columns were used previously to generate viromes from these soils (*Trubl et al., 2018*). CTAB performs well on polysaccharide-rich samples (*Porebski, Bailey & Baum, 1997*), such as our site's peat soils.

Overall, the PowerSoil kit performed best, with the highest DNA yields and increased purity which led to more successful libraries and identification of more vOTUs in the

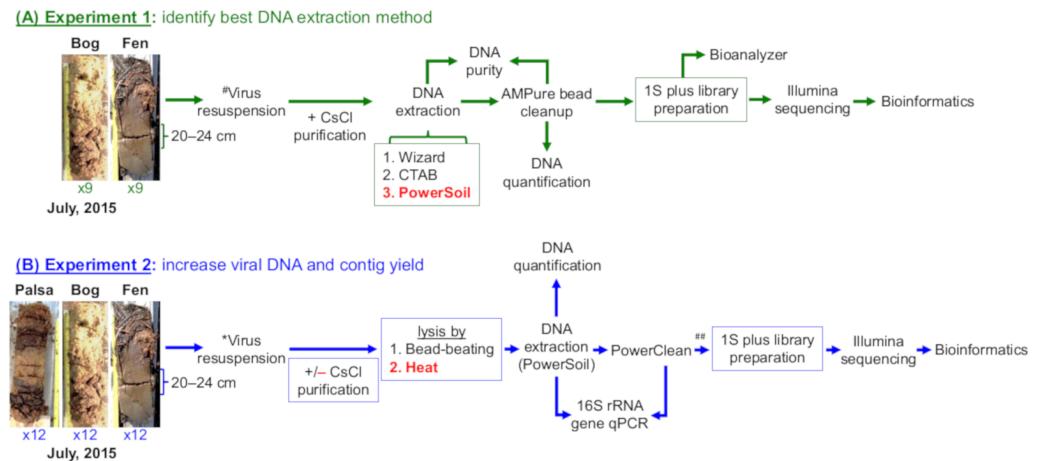


Figure 1 Overview of experiments to optimize methods for virome generation. Two experiments evaluated three DNA extraction methods (A, Experiment 1 in green), two different virion lysis methods, and CsCl virion purification (B, Experiment 2 in blue), for optimizing virome generation from three peats soils along a permafrost thaw gradient. Nine soil cores were collected in July 2015, three from each habitat, and used to create 18 samples (9 bog and 9 fen) with 10 ± 1 g of soil in each sample for Experiment 1 and 36 samples (12 palsa, 12 bog, and 12 fen) with 7.5 ± 1 g of soil in each sample for Experiment 2; representative photos of cores were taken by Gary Trubl. Viruses were resuspended as previously described in *Trubl et al. (2016)*, but with the addition of a DNase step and a 1.3 g/ml layer for CsCl purification. Red font color indicates the best-performing option within each set. # denotes adapted protocol from *Trubl et al. (2016)*. ## indicates that only 12 palsa samples proceeded to library preparation.

Full-size DOI: 10.7717/peerj.7265/fig-1

soils tested (bog and fen). Specifically, the PowerSoil kit generally yielded the most DNA (6.34 ± 0.94 in bog and 13.64 ± 4.95 in fen), although the increase was only significant in the fen habitat (one-way ANOVA, α 0.05, and Tukey's test with p -value < 0.05 ; *Fig. 2A*). DNA purity, which is also essential to virome generation (since proteins, phenols, and organics can inhibit amplification; reviewed in *Alaeddini, 2012*), was examined via A260:280 (*Fig. 2B*; for proteins and phenol contamination; (*Maniatis, Fritsch & Sambrook, 1982*)) and A260:230 ratios (*Fig. S2*; for carbohydrates and phenols; *Maniatis, Fritsch & Sambrook, 1982*; *Tanveer, Yadav & Yadav, 2016*). We posited that A260:280 is a more robust predictor of virome success, since previous work showed that A260:230 of DNA extracts had limited correlation to amplification success (*Costa et al., 2010*; *Ramos-Gómez et al., 2014*), although both are highly variable for low DNA concentrations typical for soil viral extracts. For bog samples, at least one replicate from each DNA extraction method had a clean sample based on A260:280 (defined as 1.6–2.1). For the fen, both the Wizard and PowerSoil samples were considered clean. One bog PowerSoil sample, and one fen CTAB sample, had unusually high A260:280 ratios, suggesting the presence of leftover extraction reagents in the sample.

Soil microbial metagenome protocols commonly include further DNA clean-up after extraction to remove inhibitory substances commonly seen in soil (summarized in *Roose-Amsaleg, Garnier-Sillam & Harry, 2001*; *Roslan, Mohamad & Omar, 2017*), therefore we evaluated the potential improvement in viral DNA purity from clean-up by AMPure beads. Purity (measured via A260:280) improved significantly in the bog PowerSoil + AMPure samples and was best in the CTAB + AMPure samples, while in the fen, only PowerSoil

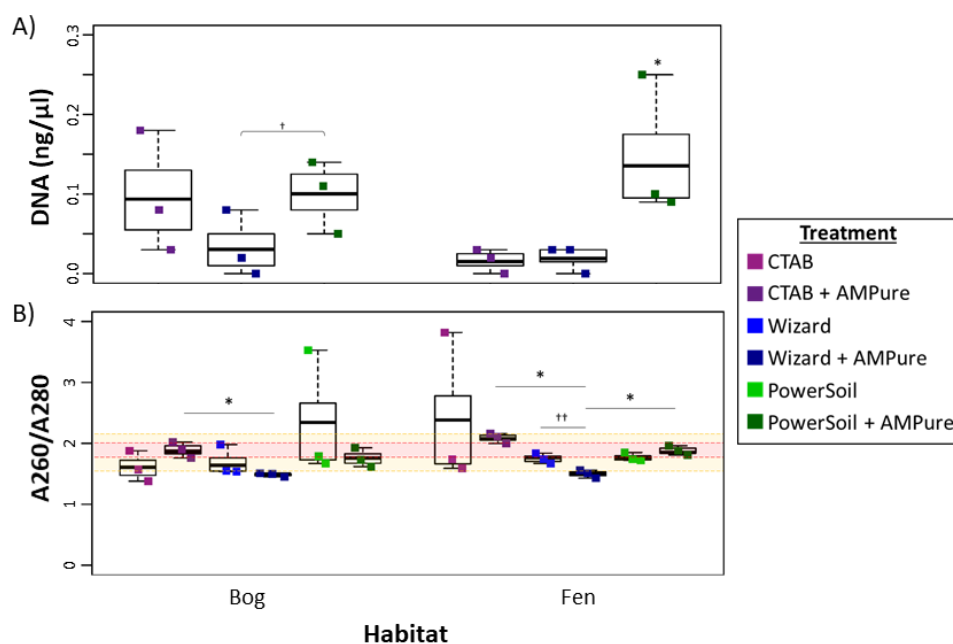


Figure 2 Impact of extraction methods on DNA yields and purity (Experiment 1). Bog samples are shown on the left of each panel, fen samples on the right. DNA extraction methods are color-coded: purple for CTAB, blue for Wizard, and green for PowerSoil. * denotes significant difference via one-way ANOVA, α 0.05, and Tukey's test with p -value < 0.05. † denotes significant difference for t test, p -value < 0.05; †† = p -value < 0.01; ††† = p -value < 0.001. (A) The DNA concentration (ng/μl) after AMPure purification for the three DNA extraction methods. (B) DNA extract purity via A260/A280. Dotted lines are purity thresholds: Acceptable range in yellow shading and preferred range in red shading.

Full-size [DOI: 10.7717/peerj.7265/fig-2](https://doi.org/10.7717/peerj.7265/fig-2)

extracts showed improvement. For A260:230, all post-clean-up DNAs were still below the standard minimum threshold (1.6–2.2, Fig. S2).

Although DNA extract yield and purity metrics are useful indicators of extract quality, the goal is successful library preparation and sequencing. Thus, we used the cleaned-up DNA to attempt virome generation, which revealed that PowerSoil-derived DNA was more amenable to library construction than the other extracts. Specifically, five of six PowerSoil extracts successfully generated libraries, whereas only one of the Wizard and none of the CTAB extracts led to successful library construction (threshold for success described in methods). Presumably, the success of the PowerSoil extraction methods was increased due to the kit having been optimized for humic-laden soils (specific reagents proprietary to Qiagen).

Where sequencing library construction was successful, we then sequenced and analyzed the resultant viromes to assess whether the vOTUs captured varied across replicate PowerSoil viromes and between the PowerSoil and Wizard viromes. In total, the 6 viromes produced 1,311 dsDNA viral contigs (VirSorter categories 1 and 2; Roux *et al.*, 2015), which clustered into 516 vOTUs (see methods; Roux *et al.*, 2019a; Roux *et al.*, 2019b). There were dramatic changes in the presence and relative abundance of vOTUs across the two DNA extraction kits evaluated, the biological replicates, and the soil habitats, which is partially

the result of uneven coverage due to the 15 rounds of PCR performed to amplify the DNA (Fig. S3). While PCR amplification is a powerful tool that permits ecological interpretation of resulting viral data (Duhaime & Sullivan, 2012; Solonenko & Sullivan, 2013; Solonenko et al., 2013), library amplification can lead to an enrichment in short inserts, resulting in uneven coverage, a bias that scales with the number of PCR cycles performed (Roux et al. 2019). The differences in vOTU presence/absence among viromes decreased but remained noticeable even when using the most sensitive thresholds proposed for the detection of a vOTU in a metagenome (Roux et al., 2019a, Fig. S3). This suggests bias from the DNA extraction method (as reported previously for microbial populations; Delmont et al., 2011; Zielińska et al., 2017), and/or haphazard detection of low-abundance vOTUs due to inadequate sampling and/or sequencing depth.

Experiment 2: heat-based lysis of non-CsCl-purified virus particles provides the most comprehensive viromes

The results of Experiment 1 identified PowerSoil as the optimal DNA extraction kit (yielding the most successful viromes), so we conducted a second experiment (Experiment 2), independent of the first, to evaluate whether density-based particle purification and/or alternative virion lysis methods could increase viral DNA yield, as previously suggested (Delmont et al., 2011; Zielińska et al., 2017). We reasoned that purification by cesium-chloride (CsCl) density gradients could result in viral loss (as previously described in Trubl et al., 2016), but also lead to reduced microbial DNA and particulate (e.g., clay or organic material) contamination by removing ultra-small (<0.2 μm) cells, known to be present in these soils (Emerson et al., 2018; Trubl et al., 2018) or material that passes the filtration step. For lysis methods, we compared the two suggested in the PowerSoil protocol and posited that heat lysis would work better because it has been used previously on viruses (reviewed in McCance, 1996) and the bead-beating method was previously shown to cause ~27% more viral loss than not using beads with PowerSoil extraction kit on diverse soils (Iker et al., 2013).

To assess this, viruses were resuspended from three palsa, bog, and fen samples as previously described (Trubl et al., 2016), and then the samples were split with half undergoing particle purification via CsCl gradients and half not, and each purification treatment lysed by each of the two lysis methods (heat and bead beating) for a total of 4 treatments, all followed by PowerSoil extraction (Fig. 1). We found significant differences in DNA yield due to purification and lysis method choice (Fig. 3, one-way ANOVA, α 0.05, and Tukey's test with p -value <0.05). CsCl purification had the most impact: yield was higher without it than with it for all but one sample (Bog, -CsCl[BB]). Lysis method also mattered, with heat producing significantly higher DNA yield than bead-beating (t test, p -value <0.05), for the -CsCl samples in the palsa and fen samples (not significant in the bog) (Fig. 3). These findings suggest that DNA yields were highest when CsCl density gradients were omitted and viral particles were lysed using heat.

Higher DNA yields could result from contaminating (i.e., non-viral) DNA, so we quantified microbial DNA in all extracts via 16S rRNA gene qPCR (Fig. 4). Surprisingly, we generally observed higher microbial contamination in the CsCl-purified samples

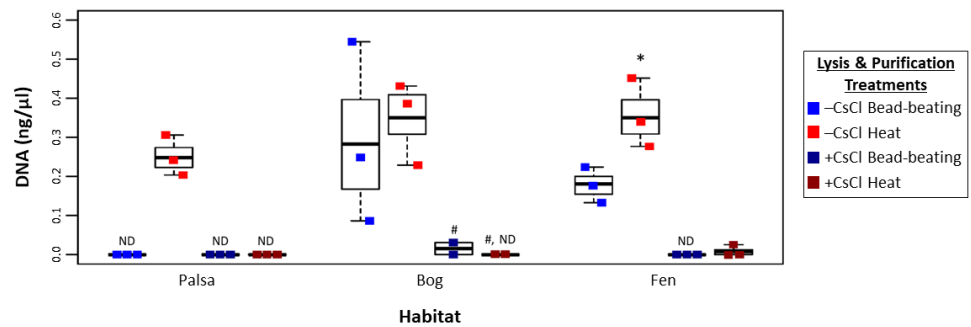


Figure 3 Impact of lysis and purification methods on DNA yields (Experiment 2). The DNA concentration (ng/ μ l) is given for the two virion lysis methods used, with or without CsCl purification, for all three habitats. The four treatments are color coded with blue for bead-beating, red for heat lysis and a darker shade if also purified with CsCl. * denotes significant difference via one-way ANOVA, α 0.05, and Tukey's test with p -value < 0.05. # denotes $n = 2$. N/D denotes non-detectable DNA concentration.

Full-size DOI: [10.7717/peerj.7265/fig-3](https://doi.org/10.7717/peerj.7265/fig-3)

(Fig. 4, one-way ANOVA, α 0.05, and Tukey's test with p -value < 0.05), and this varied along the thaw gradient with palsa contamination being higher than that of bog and fen samples. Since residual soil organics can interfere with PCR (Kontanis & Reed, 2006), we repeated the qPCR assay after DNA purification with the PowerClean kit. Generally, microbial contamination increased for –CsCl samples (Fig. 4), suggesting that their previously low microbial contamination was due to PCR inhibition, and +CsCl samples had mixed results, but in each habitat +CsCl[BB] samples had a significant increase in measurable contamination (Fig. 4). All treatments had higher qPCR-based microbial contamination after PowerClean, except +CsCl[H] samples which averaged a 1.5–26-fold reduction. Overall there was still no consistent, or significant, improvement in microbial contamination from inclusion of a CsCl purification step, even after PowerClean treatment.

Since we sequenced bog and fen viromes to characterize treatment effects on the viral signal in Experiment 1, we opted in Experiment 2 to do this evaluation on the 12 palsa samples, which were all sequenced. We found that the higher DNA yields in the –CsCl samples led to \sim 3-fold more viral contigs, which were also an average of 2.3-fold larger than +CsCl samples (Fig. 5A). The results from heat-lysis samples were more modest as they resulted in only \sim 33% more viral contigs, and statistically indistinguishable contig sizes across treatments (Fig. 5B; unequal variance t -test, p -value > 0.05). These findings suggest that the optimal combination for recovering virus genomes from these soils may be to skip CsCl purification, but still using some form of purification method (DNase used here), and lyse the resultant viral particles using heat.

We next evaluated whether vOTU representation and diversity estimates from the same samples varied across the purification and lysis methods tested here. DNA quantification of 9 out of the 12 viromes showed non-detectable amounts of DNA, but we identified vOTUs in each of the 12 palsa viromes, suggesting the Accel-NGS 1S Plus kit amplifies DNA from the picogram range. In total, 66 vOTUs were identified with 100% of the vOTUs identified in –CsCl samples, 89% (59) identified in the +CsCl samples, and vOTUs identified by both datasets displaying an average of 30-fold more coverage (Fig. 6) in –CsCl viromes. This

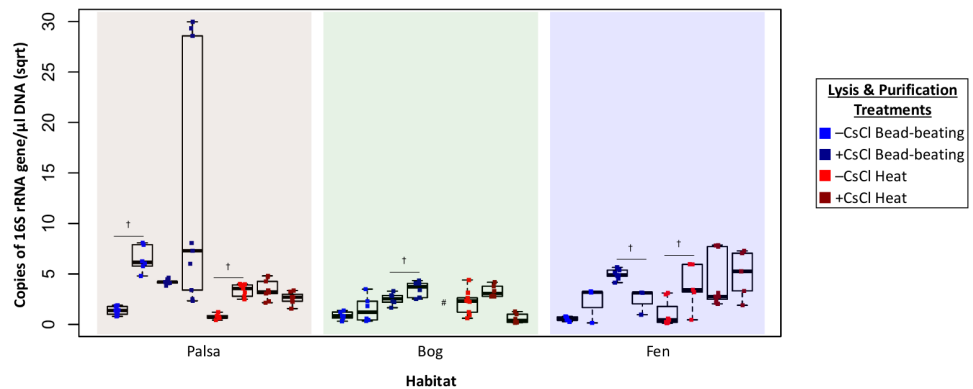


Figure 4 Evaluation of microbial contamination (Experiment 2). The 16S rRNA gene contamination (square root) is indicated for each virome grouped by habitat before (left) and after (right) clean up with PowerClean. The four treatments are color coded with blue for bead-beating and red for heat lysis and a darker shade after CsCl purification. # denotes no data available. 16S qPCR primers were 1406F-1525R, from Woodcroft et al. (2018) † denotes significant difference for *t* test, p -value < 0.05; ††, p -value < 0.01; †††, p -value < 0.001.

Full-size [DOI: 10.7717/peerj.7265/fig-4](https://doi.org/10.7717/peerj.7265/fig-4)

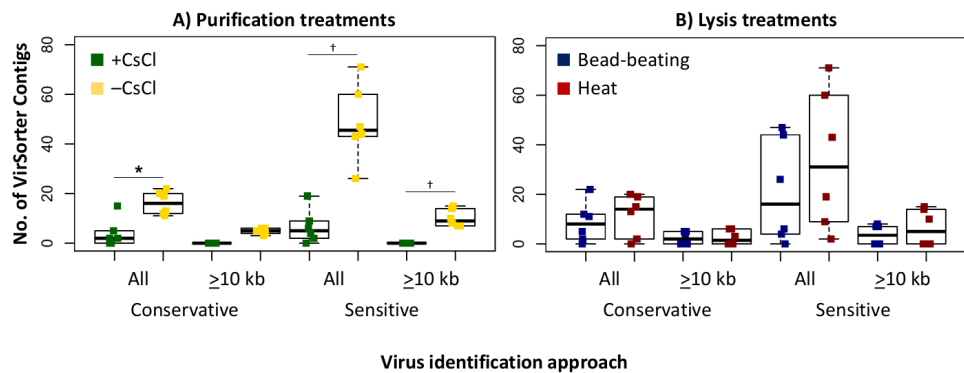


Figure 5 Number and size of assembled viral contigs (Experiment 2). Boxplots show the number of viral contigs assembled, and those > 10 kb, for each treatment. Viral contigs were identified by two approaches: the “conservative” one included only contigs in VirSorter categories 1 & 2 for which a viral origin is very likely, while the “sensitive” one also included contigs in VirSorter category 3, for which a viral origin is possible but unsure.

Full-size [DOI: 10.7717/peerj.7265/fig-5](https://doi.org/10.7717/peerj.7265/fig-5)

indicates that the CsCl purification step reduced the samples to a subset of the initial viral community, it did not help recover virus genomes that would be missed otherwise, and confirmed that the 16S rRNA gene copies identified from the qPCR analyses were likely microbial contamination and not the result of 16S rRNA gene copies carried by viruses (Ghosh et al., 2008). Profiles of the recovered communities clustered first by soil core (AU branch supports >76), then mostly by purification (AU branch supports >66), and lastly by lysis, and did not change after varying the threshold for considering a lineage present (Fig. S4). Collectively this suggests that differences introduced by sample preparation were outweighed by the distinctiveness of each core’s viral community. We proceeded to use

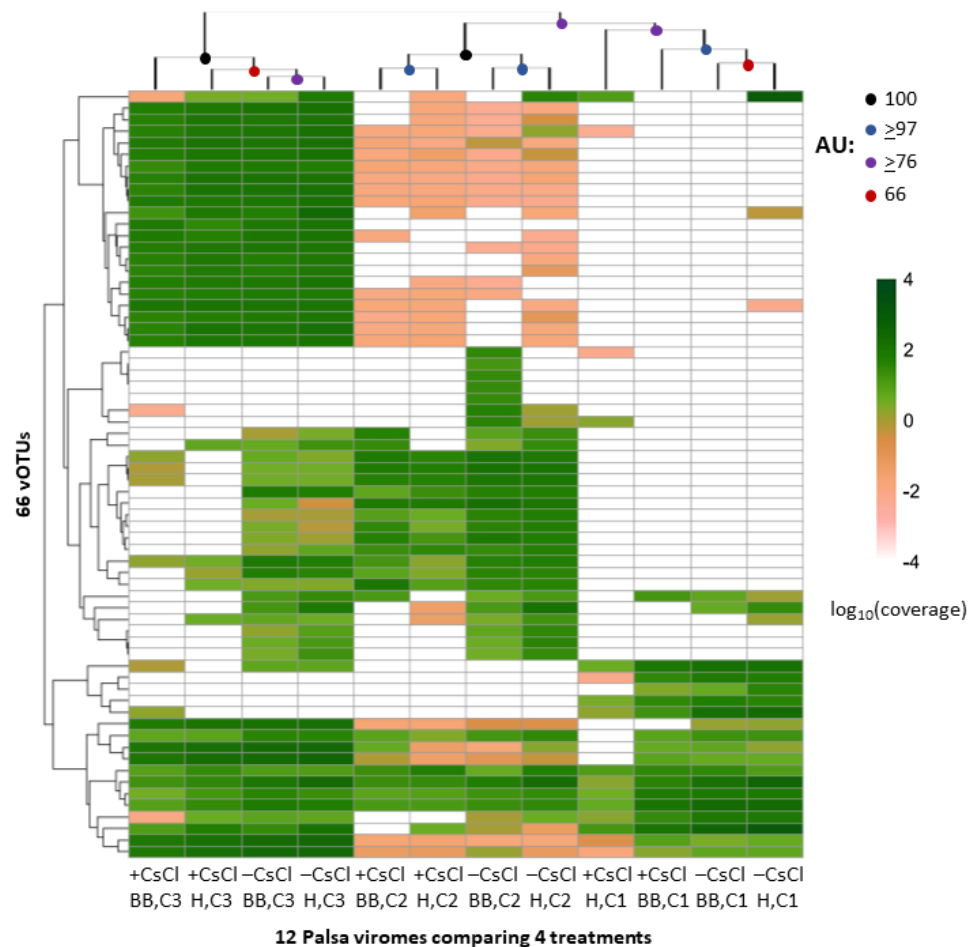


Figure 6 Relative abundance of vOTUs across 12 palsa viromes (Experiment 2). A heatmap showing the Euclidean-based hierarchical clustering of a Bray-Curtis dissimilarity matrix calculated from vOTU relative abundances within each virome with an approximately unbiased (AU) bootstrap value ($n = 1,000$). The relative abundances were normalized by contig length and per Gbp of metagenome and were \log_{10} transformed. Reads were mapped to contigs at $\geq 90\%$ nucleotide identity and the relative abundance was set to 0 if reads covered $< 10\%$ of the contig. Heatmaps with alternative genome coverage thresholds are presented in Fig. S3. Abbreviations: H, heat lysis; BB, bead-beating; +/- CsCl, with or without cesium chloride purification; C, soil core.

Full-size DOI: 10.7717/peerj.7265/fig-6

diversity metrics to evaluate the different methods' impacts. The alpha diversity metrics paralleled treatment DNA yields where $-CsCl$ samples were on average 56% more diverse than the $+CsCl$ samples, and heat samples were on average 83% more diverse than the bead-beating samples (Fig. S5A). A comparison of dissimilarities among samples suggested the lysis method had more of an impact, although this effect was variable between samples and thus not statistically significant overall (Fig. S5B).

ssDNA viruses are recovered in all 3 habitats

Previous viromic studies have been limited to describing dsDNA viruses or using MDA to describe ssDNA viruses, but with the onset of the Accel-NGS 1S Plus kit, we leveraged

the quantitatively-amplified viromics data produced here to investigate the diversity and relative abundance of ssDNA viruses in our soil samples. Culture collections have revealed ssDNA viruses commonly infect plants as opposed to bacteria, but their distributions in soils remain poorly explored outside a handful of papers which suggest they are highly diverse (*Kim et al., 2008; Reavy et al., 2015; Green et al., 2018*). Notably, the first quantitative ssDNA/dsDNA viromes suggested that identifiable ssDNA viruses represent a few percent of the viruses observed in marine and freshwater systems (*Roux et al., 2016*).

To assess this biological signal in soils, we investigated the recovery and relative abundance of ssDNA viruses across our different soil habitats and sample preparations. Overall, we identified 35 putative ssDNA viruses, 11 from the Microviridae family and 24 CRESS DNA viruses (*Fig. 7*), which clustered into 13 vOTUs (3 Microviridae and 10 CRESS DNA). These ssDNA vOTUs were only a small fraction of the total vOTUs identified in each habitat (1% in bog and fen, and 8% in palsa) and only bog and fen samples included both types (Microviridae and CRESS-DNA), while palsa samples included exclusively CRESS-DNA viruses (*Table S1*). This suggests that, as for dsDNA viruses, the composition of the ssDNA virus community varies along the thaw gradient, potentially as a result of known changes in the host communities (*Trubl et al., 2018*), both microbial (*Mondav et al., 2017; Woodcroft et al., 2018*) and plant (*Hodgkins et al., 2014; Normand et al., 2017*). Notably, bead-beating-lysis samples did not include any ssDNA viruses. We posit that this was likely due to the heterogeneity of soil, because ssDNA viruses have previously been identified from experiments that used a bead-beating lysis (*Hopkins et al., 2014*). Finally, ssDNA viruses represented on average 4% of the community in the samples where ssDNA and dsDNA viruses were detected, which suggests that ssDNA viruses are not the dominant type of virus in these soils.

CONCLUSIONS

The development of a sample-to-sequence pipeline for ssDNA and dsDNA viruses in soils is crucial for characterizing viruses and their impact in these ecosystems. Our work here built upon previous work that optimized virus resuspension from peatland soils by evaluating DNA extraction and lysis methods to increase DNA yields and purity. Additionally, this is the first evaluation of the Accel-NGS 1S Plus kit to capture ssDNA viruses in soils and our data suggests it is also capable of amplifying DNA down to the picogram range. Although these efforts have made inroads towards characterizing the soil virosphere, several challenges remain. Initial challenges arise from resuspension and enumeration of “fake” virus particles (*Ackermann & Tiekotter, 2012; Forterre et al., 2013*), the lack of data on what fraction of the free virus particles are being recovered from soils, and how to achieve a holistic sampling of the virus community (i.e., dsDNA, ssDNA, and RNA viruses). After viruses are resuspended from soils and their nucleic acid is extracted, there is still a need for amplification which can cause downstream issues (e.g., uneven coverage). Beyond these, the presence of non-viral DNA in capsids or vesicles, e.g., gene transfer agents, can dilute the viral signal in metagenomes and complicate interpretation (reviewed in *Roux et al., 2013; Hurwitz, Hallam & Sullivan, 2013; Lang & Beatty, 2010*), although new methods

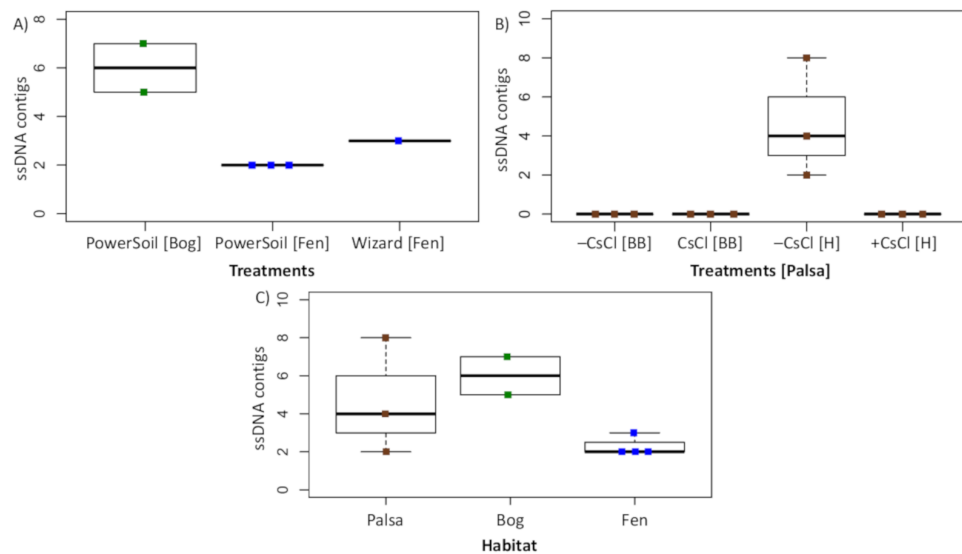


Figure 7 Recovery of ssDNA viruses across habitats and methods. (A) ssDNA viral contigs from viromes in Experiment 2. The PowerSoil bog samples are grouped, as are the PowerSoil fen samples. The single Wizard virome from the fen habitat is also shown. (B) ssDNA viral contigs from viromes in Experiment 2 grouped by the four treatments: +/- CsCl and bead-beating [BB] or heat [H] virion lysis method. (C) ssDNA viruses from both Experiments are shown and grouped by habitat.

Full-size DOI: [10.7717/peerj.7265/fig-7](https://doi.org/10.7717/peerj.7265/fig-7)

are being developed to identify and characterize these contaminating agents (reviewed in [Lang, Westbye & Beatty, 2017](#)). Given all the known contaminants that can pass through filtration and their unknown densities or impact on DNA extraction and amplification, we caution the removal of the CsCl purification without further assessment on additional soils.

In addition to optimization of methods to characterize soil viruses, there are many techniques that can be implemented that will greatly advance our knowledge of viruses in soils. The advent of long-read sequencing technologies have recently been applied to viromics and can improve contig generation for regions of genome with high similarity or complexity (summarized in [Roux et al., 2017](#); [Karamitros et al., 2018](#)) and prevent formation of chimeric contigs. Longer-read viromes can thereby not only increase vOTU recovery but also provide resolution of hypervariable genome regions with niche-defining genes, and help capture micro-diverse populations missed by short-read assemblies ([Warwick-Dugdale et al., 2019](#)). Next, inferences of viral impacts on microbial communities and C cycling will require predicting hosts both *in silico* ([Edwards et al., 2015](#); [Paez-Espino et al., 2017](#)) and *in vitro* ([Deng et al., 2014](#); [Brum & Sullivan, 2015](#); [Cenens et al., 2015](#)), approaches to which are emerging. Finally, identification of the active viral community and characterization of their roles in biogeochemical processes can be better resolved with techniques like stable isotope-based approaches linked with nanoscale secondary ion mass spectrometry (NanoSIP; [Pacton et al., 2014](#); [Pasulka et al., 2018](#); [Gates et al., 2018](#)). Application of these and other approaches to soil viromics will increase and diversify

publicly available viral datasets, advance our understanding of soil viral ecology, and improve our knowledge of viral roles in soil ecosystems.

ACKNOWLEDGEMENTS

We thank Michelle Carlson from Qiagen for her support and technical advice. We thank Olivier Zablocki for his suggestions and comments on the manuscript. We thank Moira Hough, Sky Dominguez, and Nicole Raab for collecting the soil cores and geochemical data, and the Abisko Naturvetenskapliga Station for field support.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research (grants DE-SC0010580 and DE-SC0016440), The Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program, and by the Gordon and Betty Moore Foundation Investigator Award (GBMF#3790 to Matt Sullivan). The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under contract number DE-SC0014664. The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. Bioinformatics were supported by The Ohio Supercomputer Center and by the National Science Foundation under Award Numbers DBI-0735191 and DBI-1265383; URL: www.cyverse.org. 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:

Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research: DE-SC0010580, DE-SC0016440.

The Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR).

Gordon and Betty Moore Foundation Investigator Award: GBMF#3790.

Oak Ridge Institute for Science and Education (ORISE): DE-SC0014664.

Office of Science of the US Department of Energy: DE-AC02-05CH11231.

Ohio Supercomputer Center and by the National Science Foundation under Award Numbers: DBI-073519, DBI-1265383.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Gareth Trubl, Yueh-Fen Li conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Simon Roux, Benjamin Bolduc and Josué Rodríguez-Ramos analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Natalie Solonenko conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Emiley A. Eloë-Fadrosh contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Virginia I. Rich and Matthew B. Sullivan conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data is available in the IsoGenieDB, in the Viral Sequencing Data section at https://isogenie-db.asc.ohio-state.edu/datasources#viral_sequencing. All datasets named Trubl Soil Viromes were used in this study.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7265#supplemental-information>.

REFERENCES

- Ackermann HW, Tiekotter KL. 2012.** Murphy's law—If anything can go wrong, it will: problems in phage electron microscopy. *Bacteriophage* **2**(2):122–129 DOI [10.4161/bact.20693](https://doi.org/10.4161/bact.20693).
- Alaeddini R. 2012.** Forensic implications of PCR inhibition—a review. *Forensic Science International: Genetics* **6**(3):297–305 DOI [10.1016/j.fsigen.2011.08.006](https://doi.org/10.1016/j.fsigen.2011.08.006).
- Amgarten DE, Braga LPP, Da Silva AM, Setubal JC. 2018.** MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Frontiers in Genetics* **9**:304 DOI [10.3389/fgene.2018.00304](https://doi.org/10.3389/fgene.2018.00304).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5):455–477 DOI [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- Binga EK, Lasken RS, Neufeld JD. 2008.** Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *The ISME journal* **2**(3):233 DOI [10.1038/ismej.2008.10](https://doi.org/10.1038/ismej.2008.10).
- Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017.** iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *The ISME Journal* **11**(1):7.

- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120.
- Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, De Vargas C, Gasol JM, Gorsky G. 2015.** Patterns and ecological drivers of ocean viral communities. *Science* **348**(6237):1261498.
- Brum JR, Sullivan MB. 2015.** Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* **13**(3):147 DOI [10.1038/nrmicro3404](https://doi.org/10.1038/nrmicro3404).
- Cenens W, Makumi A, Govers SK, Lavigne R, Aertsen A. 2015.** Viral transmission dynamics at single-cell resolution reveal transiently immune subpopulations caused by a carrier state association. *PLOS Genetics* **11**(12):e1005770 DOI [10.1371/journal.pgen.1005770](https://doi.org/10.1371/journal.pgen.1005770).
- Christensen TR, Johansson T, Malmer N, Åkerman J, Friberg T, Crill P, Svensson B. 2004.** Changes in Climate, permafrost and vegetation-effects on subarctic methane emission. *Geophysical Research Letters* **31**(4).
- Costa J, Mafra I, Amaral JS, Oliveira MBP. 2010.** Detection of genetically modified soybean DNA in refined vegetable oils. *European Food Research and Technology* **230**(6):915–923 DOI [10.1007/s00217-010-1238-2](https://doi.org/10.1007/s00217-010-1238-2).
- Delcher AL, Salzberg SL, Phillippy AM. 2003.** Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics* (1):10–13 DOI [10.1002/0471250953.bi1003s00](https://doi.org/10.1002/0471250953.bi1003s00).
- Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM. 2011.** Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology* **77**(4):1315–1324 DOI [10.1128/AEM.01526-10](https://doi.org/10.1128/AEM.01526-10).
- Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. 2014.** Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**(7517):242 DOI [10.1038/nature13459](https://doi.org/10.1038/nature13459).
- Duhaime MB, Sullivan MB. 2012.** Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434**(2):181–186 DOI [10.1016/j.virol.2012.09.036](https://doi.org/10.1016/j.virol.2012.09.036).
- Eddy SR. 2009.** A new generation of homology search tools based on probabilistic inference. *Genome informatics. International Conference on Genome Informatics* **23**(1):205–211 DOI [10.1142/9781848165632_0019](https://doi.org/10.1142/9781848165632_0019).
- Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2015.** Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews* **40**(2):258–272.
- Edwards RA, Rohwer F. 2005.** Viral metagenomics. *Nature Reviews Microbiology* **3**(6):504.
- Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB. 2018.** Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* **3**(8):870.
- Fierer N. 2017.** Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15**(10):579 DOI [10.1038/nrmicro.2017.87](https://doi.org/10.1038/nrmicro.2017.87).
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R. 2007.** Metagenomic and small-subunit rRNA

- analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology* **73**(21):7059–7066.
- Forterre P, Soler N, Krupovic M, Marguet E, Ackermann HW. 2013.** Fake virus particles generated by fluorescence microscopy. *Trends in Microbiology* **21**(1):1–5.
- Gates SD, Condit RC, Moussatche N, Stewart BJ, Malkin AJ, Weber PK. 2018.** High initial sputter rate found for vaccinia virions using isotopic labeling, nanoSIMS, and AFM. *Analytical Chemistry* **90**(3):1613–1620 DOI [10.1021/acs.analchem.7b02786](https://doi.org/10.1021/acs.analchem.7b02786).
- Ghosh D, Roy K, Williamson KE, White DC, Wommack KE, Sublette KL, Radosevich M. 2008.** Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and trzN genes in viral-community DNA. *Applied and Environmental Microbiology* **74**(2):495–502.
- Green JC, Rahman F, Saxton MA, Williamson KE. 2018.** Quantifying aquatic viral community change associated with stormwater runoff in a wet retention pond using metagenomic time series data. *Aquatic Microbial Ecology* **81**(1):19–35 DOI [10.3354/ame01856](https://doi.org/10.3354/ame01856).
- Gregory A, Zayed A, Conceição Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmicheal M, Cruaud C, Dimier C. 2019.** Marine DNA viral macro- and micro-diversity from pole to pole. *Cell* **177**(5):1109–1123 DOI [10.1016/j.cell.2019.03.040](https://doi.org/10.1016/j.cell.2019.03.040).
- Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP. 2016.** Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**(7600):465.
- Hayes S, Mahony J, Nauta A, Van Sinderen D. 2017.** Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* **9**(6):127 DOI [10.3390/v9060127](https://doi.org/10.3390/v9060127).
- Hodgkins SB, Tfaily MM, McCalley CK, Logan TA, Crill PM, Saleska SR, Rich VI, Chanton JP. 2014.** Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences of the United States of America* **111**(16):5819–5824 DOI [10.1073/pnas.1314641111](https://doi.org/10.1073/pnas.1314641111).
- Hopkins M, Kailasan S, Cohen A, Roux S, Tucker KP, Shevenell A, Agbandje-McKenna M, Breitbart M. 2014.** Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. *The ISME journal* **8**(10):2093 DOI [10.1038/ismej.2014.43](https://doi.org/10.1038/ismej.2014.43).
- Hurwitz BL, Hallam SJ, Sullivan MB. 2013.** Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology* **14**(11):R123 DOI [10.1186/gb-2013-14-11-r123](https://doi.org/10.1186/gb-2013-14-11-r123).
- Iker BC, Bright KR, Pepper IL, Gerba CP, Kitajima M. 2013.** Evaluation of commercial kits for the extraction and purification of viral nucleic acids from environmental and fecal samples. *Journal of Virological Methods* **191**(1):24–30 DOI [10.1016/j.jviromet.2013.03.011](https://doi.org/10.1016/j.jviromet.2013.03.011).
- Johansson T, Malmer N, Crill PM, Friborg T, Aakerman JH, Mastepanov M, Christensen TR. 2006.** Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Global Change Biology* **12**(12):2352–2369 DOI [10.1111/j.1365-2486.2006.01267.x](https://doi.org/10.1111/j.1365-2486.2006.01267.x).

- Jonasson C, Sonesson M, Christensen TR, Callaghan TV. 2012.** Environmental monitoring and research in the Abisko area—an overview. *Ambio* **41**(3):178–186.
- Karamitros T, Van Wilgenburg B, Wills M, Klenerman P, Magiorkinis G. 2018.** Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. *BMC Genomics* **19**(1):577 DOI [10.1186/s12864-018-4949-6](https://doi.org/10.1186/s12864-018-4949-6).
- Karlsson OE, Belák S, Granberg F. 2013.** The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* **11**(S1):S227–S23 DOI [10.1089/bsp.2013.0008](https://doi.org/10.1089/bsp.2013.0008).
- Kim KH, Chang HW, Nam YD, Roh SW, Kim MS, Sung Y, Jeon CO, Oh HM, Bae JW. 2008.** Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and Environment Microbiology* **74**(19):5975–5985 DOI [10.1128/AEM.01275-08](https://doi.org/10.1128/AEM.01275-08).
- Kim MS, Whon TW, Bae JW. 2013.** Comparative viral metagenomics of environmental samples from Korea. *Genomics & Informatics* **11**(3):121–128 DOI [10.5808/GI.2013.11.3.121](https://doi.org/10.5808/GI.2013.11.3.121).
- Kolde R. 2015.** Pheatmap: pretty heatmaps. R package version, 61. Available at <https://cran.r-project.org/package=pheatmap>.
- Kontanis EJ, Reed FA. 2006.** Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors. *Journal of Forensic Sciences* **51**(4):795–804 DOI [10.1111/j.1556-4029.2006.00182.x](https://doi.org/10.1111/j.1556-4029.2006.00182.x).
- Lang AS, Beatty JT. 2010.** Gene transfer agents and defective bacteriophages as sources of extracellular prokaryotic DNA. In: *Extracellular nucleic acids*. Berlin, Heidelberg: Springer, 15–24.
- Lang AS, Westbye AB, Beatty JT. 2017.** The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annual Review of Virology* **4**:87–104 DOI [10.1146/annurev-virology-101416-041624](https://doi.org/10.1146/annurev-virology-101416-041624).
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**(4):357 DOI [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).
- Laurie Kurihara, Banks L, Chupreta S, Couture C, Kelchner V, Laliberte J, Sandhu S, Spurbeck R, Makarov V. 2014.** A new method for preparation of low-input, PCR-free next generation sequencing libraries [Abstract 3564]. *Cancer Research* **74**(19 Suppl) DOI [10.1158/1538-7445.AM2014-3564](https://doi.org/10.1158/1538-7445.AM2014-3564).
- Lehmann J, Kleber M. 2015.** The contentious nature of soil organic matter. *Nature* **528**(7580):60 DOI [10.1038/nature16069](https://doi.org/10.1038/nature16069).
- Malmer N, Johansson T, Olsrud M, Christensen TR. 2005.** Vegetation, climatic changes and net carbon sequestration in a North-Scandinavian subarctic mire over 30 years. *Global Change Biology* **11**(11):1895–1909.
- Maniatis T, Fritsch EF, Sambrook J. 1982.** *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory.
- Marine R, McCarren C, Vorrassane V, Nasko D, Crowgey E, Polson SW, Wommack KE. 2014.** Caught in the middle with multiple displacement amplification: the myth of

- pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* 2(1):3 DOI 10.1186/2049-2618-2-3.
- McCalley CK, Woodcroft BJ, Hodgkins SB, Wehr RA, Kim EH, Mondav R, Crill PM, Chanton JP, Rich VI, Tyson GW, Saleska SR. 2014.** Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* 514(7523):478 DOI 10.1038/nature13798.
- McCance DJ. 1996.** DNA viruses: DNA extraction, purification and characterization. In: *Virology methods manual*. Academic Press, 191–230.
- Mondav R, McCalley CK, Hodgkins SB, Frolking S, Saleska SR, Rich VI, Chanton JP, Crill PM. 2017.** Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. *Environmental Microbiology* 19(8):3201–3218 DOI 10.1111/1462-2920.13809.
- Mondav R, Woodcroft BJ, Kim EH, McCalley CK, Hodgkins SB, Crill PM, Chanton J, Hurst GB, VerBerkmoes NC, Saleska SR, Hugenholtz P. 2014.** Discovery of a novel methanogen prevalent in thawing permafrost. *Nature Communications* 5:3212.
- Narayan A, Jain K, Shah AR, Madamwar D. 2016.** An efficient and cost-effective method for DNA extraction from athalassohaline soil using a newly formulated cell extraction buffer. *3 Biotech* 6(1):62.
- Narr A, Nawaz A, Wick LY, Harms H, Chatzinotas A. 2017.** Soil viral communities vary temporally and along a land use transect as revealed by virus-like particle counting and a modified community fingerprinting approach (fRAPD). *Frontiers in Microbiology* 8:1975 DOI 10.3389/fmicb.2017.01975.
- Normand AE, Smith AN, Clark MW, Long JR, Reddy KR. 2017.** Chemical composition of soil organic matter in a subarctic peatland: influence of shifting vegetation communities. *Soil Science Society of America Journal* 81(1):41–49 DOI 10.2136/sssaj2016.05.0148.
- Ohio Supercomputer Center. 1987.** *Ohio supercomputer center*. Columbus OH: Ohio Supercomputer Center.
- Oksanen J, Blanchet F, Kindt R, Legendre P, O'Hara R. 2016.** Vegan: community ecology package. R package 2.3-3. Available at <https://cran.r-project.org/package=vegan>.
- Olefeldt D, Roulet NT, Bergeron O, Crill P, Bäckstrand K, Christensen TR. 2012.** Net carbon accumulation of a high-latitude permafrost tundra mire similar to permafrost-free peatlands. *Geophysical Research Letters* 39(3) DOI 10.1029/2011GL050355 .
- Pacton M, Wacey D, Corinaldesi C, Tangherlini M, Kilburn MR, Gorin GE, Danovaro R, Vasconcelos C. 2014.** Viruses as new agents of organomineralization in the geological record. *Nature Communications* 5:4298.
- Paez-Espino D, Eloë-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016.** Uncovering Earth's virome. *Nature* 536(7617):425 DOI 10.1038/nature19094.
- Paez-Espino D, Pavlopoulos GA, Ivanova NN, Kyrpides NC. 2017.** Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. *Nature Protocols* 12(8):1673 DOI 10.1038/nprot.2017.063.

- Pasulka AL, Thamatrakoln K, Kopf SH, Guan Y, Poulos B, Moradian A, Sweredoski MJ, Hess S, Sullivan MB, Bidle KD, Orphan VJ. 2018.** Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environmental Microbiology* **20**(2):671–692 DOI [10.1111/1462-2920.13996](https://doi.org/10.1111/1462-2920.13996).
- Porebski S, Bailey LG, Baum BR. 1997.** Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**(1):8–15 DOI [10.1007/BF02772108](https://doi.org/10.1007/BF02772108).
- Pratama AA, Van Elsas JD. 2018.** The ‘neglected’ soil virome-potential role and impact. *Trends in Microbiology* **26**(8):649–662.
- Ramos-Gómez S, Busto MD, Perez-Mateos M, Ortega N. 2014.** Development of a method to recovery and amplification DNA by real-time PCR from commercial vegetable oils. *Food Chemistry* **158**:374–383 DOI [10.1016/j.foodchem.2014.02.142](https://doi.org/10.1016/j.foodchem.2014.02.142).
- Reavy B, Swanson MM, Cock PJ, Dawson L, Freitag TE, Singh BK, Torrance L, Mushegian AR, Taliany M. 2015.** Distinct circular single-stranded DNA viruses exist in different soil types. *Applied and Environment Microbiology* **81**(12):3934–3945 DOI [10.1128/AEM.03878-14](https://doi.org/10.1128/AEM.03878-14).
- Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017.** VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**(1):69 DOI [10.1186/s40168-017-0283-5](https://doi.org/10.1186/s40168-017-0283-5).
- Rinke C, Low S, Woodcroft BJ, Raina JB, Skarshewski A, Le XH, Butler MK, Stocker R, Seymour J, Tyson GW, Hugenholtz P. 2016.** Validation of picogram-and femtogram-input DNA libraries for microscale metagenomics. *PeerJ* **4**:e2486.
- Roose-Amsaleg CL, Garnier-Sillam E, Harry M. 2001.** Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology* **18**(1):47–60 DOI [10.1016/S0929-1393\(01\)00149-4](https://doi.org/10.1016/S0929-1393(01)00149-4).
- Rosario K, Fierer N, Miller S, Luongo J, Breitbart M. 2018.** Diversity of DNA and RNA viruses in indoor air as assessed via metagenomic sequencing. *Environmental Science & Technology* **52**(3):1014–1027 DOI [10.1021/acs.est.7b04203](https://doi.org/10.1021/acs.est.7b04203).
- Roslan MAM, Mohamad MAN, Omar SM. 2017.** High quality DNA from peat soil for metagenomic studies a minireview on dna extraction methods. *Science* **1**(2):01–06.
- Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MS, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. 2019a.** Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology* **37**(1):29–37.
- Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017.** Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**:e3817 DOI [10.7717/peerj.3817](https://doi.org/10.7717/peerj.3817).

- Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985 DOI 10.7717/peerj.985.
- Roux S, Krupovic M, Debroas D, Forterre P, Enault F. 2013. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biology* 3(12):130160 DOI 10.1098/rsob.130160.
- Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, Coleman ML, Breitbart M, Sullivan MB. 2016. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* 4:e2777 DOI 10.7717/peerj.2777.
- Roux S, Trubl G, Goudeau D, Nath N, Couradeau E, Ahlgren NA, Zhan Y, Marsan D, Chen F, Fuhrman JA, Northen TR. 2019b. Optimizing de novo genome assembly from PCR-amplified metagenomes. *PeerJ* 7:e6902.
- Segobola J, Adriaenssens E, Tsekoa T, Rashamuse K, Cowan D. 2018. Exploring viral diversity in a unique South African soil habitat. *Scientific Reports* 8(1):111 DOI 10.1038/s41598-017-18461-0.
- Solonenko SA, Ignacio-Espinoza JC, Alberti A, Cruaud C, Hallam S, Konstantinidis K, Tyson G, Wincker P, Sullivan MB. 2013. Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* 14(1):320 DOI 10.1186/1471-2164-14-320.
- Solonenko SA, Sullivan MB. 2013. Preparation of metagenomic libraries from naturally occurring marine viruses. In: *Methods in enzymology*. Vol. 531. Academic Press, 143–165.
- Srinivasiah S, Lovett J, Ghosh D, Roy K, Fuhrmann JJ, Radosevich M, Wommack KE. 2015. Dynamics of autochthonous soil viral communities parallels dynamics of host communities under nutrient stimulation. *FEMS Microbiology Ecology* 91(7).
- Sullivan MB, Weitz JS, Wilhelm S. 2017. Viral ecology comes of age. *Environmental Microbiology Reports* 9(1):33–35.
- Tanveer A, Yadav S, Yadav D. 2016. Comparative assessment of methods for metagenomic DNA isolation from soils of different crop growing fields. *3 Biotech* 6(2):220 DOI 10.1007/s13205-016-0543-2.
- Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nature Protocols* 4(4):470.
- Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, Solden L, Ellenbogen J, Runyon AT, Bolduc B, Woodcroft BJ, Saleska SR, Tyson GW, Wrighton KC, Sullivan MB, Rich VI. 2018. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* 3:e00076-18 DOI 10.1128/mSystems.00076-18.
- Trubl G, Solonenko N, Chittick L, Solonenko SA, Rich VI, Sullivan MB. 2016. Optimization of viral resuspension methods for carbon-rich soils along a permafrost thaw gradient. *PeerJ* 4:e1999.
- Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, Temperton B. 2019. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7:e6800 DOI 10.7717/peerj.6800.

- Wickham H.** 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York. Available at <https://ggplot2.tidyverse.org>.
- Wilson RM, Fitzhugh L, Whiting GJ, Frolking S, Harrison MD, Dimova N, Burnett WC, Chanton JP.** 2017. Greenhouse gas balance over thaw-freeze cycles in discontinuous zone permafrost. *Journal of Geophysical Research: Biogeosciences* 122(2):387–404.
- Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M.** 2017. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annual Review of Virology* 4:201–219.
- Wommack KE, Bhavsar J, Polson SW, Chen J, Dumas M, Srinivasiah S, Furman M, Jamindar S, Nasko DJ.** 2012. VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Standards in Genomic Sciences* 6:427–439 DOI 10.4056/sigs.2945050.
- Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AA, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM.** 2018. Genome—centric view of carbon processing in thawing permafrost. *Nature* 560(7716):49.
- Yilmaz S, Allgaier M, Hugenholtz P.** 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* 7(12):943 DOI 10.1038/nmeth1210-943.
- Zablocki O, Adriaenssens EM, Cowan D.** 2016. Diversity and ecology of viruses in hyperarid desert soils. *Applied and Environment Microbiology* 82(3):770–777 DOI 10.1128/AEM.02651-15.
- Zielińska S, Radkowski P, Blendowska A, Ludwig-Gałęzowska A, Łoś JM, Łoś M.** 2017. The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *MicrobiologyOpen* 6(4):e00453 DOI 10.1002/mbo3.453.