

Benchmarking protocols for the metagenomic analysis of stream biofilm viromes

Meriem Bekliz, Jade Brandani, Massimo Bourquin, Tom J. Battin and Hannes Peter

Stream Biofilm and Ecosystem Research Laboratory, École Polytechnique Federale de Lausanne, Lausanne, Switzerland

ABSTRACT

Viruses drive microbial diversity, function and evolution and influence important biogeochemical cycles in aquatic ecosystems. Despite their relevance, we currently lack an understanding of their potential impacts on stream biofilm structure and function. This is surprising given the critical role of biofilms for stream ecosystem processes. Currently, the study of viruses in stream biofilms is hindered by the lack of an optimized protocol for their extraction, concentration and purification. Here, we evaluate a range of methods to separate viral particles from stream biofilms, and to concentrate and purify them prior to DNA extraction and metagenome sequencing. Based on epifluorescence microscopy counts of viral-like particles (VLP) and DNA yields, we optimize a protocol including treatment with tetrasodium pyrophosphate and ultra-sonication to disintegrate biofilms, tangential-flow filtration to extract and concentrate VLP, followed by ultracentrifugation in a sucrose density gradient to isolate VLP from the biofilm slurry. Viromes derived from biofilms sampled from three different streams were dominated by *Siphoviridae*, *Myoviridae* and *Podoviridae* and provide first insights into the viral diversity of stream biofilms. Our protocol optimization provides an important step towards a better understanding of the ecological role of viruses in stream biofilms.

Submitted 20 August 2019
Accepted 11 November 2019
Published 20 December 2019

Corresponding author
Hannes Peter, hannes.peter@epfl.ch

Academic editor
Christian Rinke

Additional Information and
Declarations can be found on
page 15

DOI [10.7717/peerj.8187](https://doi.org/10.7717/peerj.8187)

© Copyright
2019 Bekliz et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Microbiology, Virology, Freshwater Biology

Keywords Stream biofilm, Phage, Viral metagenomics, Ultracentrifugation, Tangential flow filtration

INTRODUCTION

Viruses are the smallest and most abundant biological entities on Earth, typically outnumbering their prokaryotic and eukaryotic hosts by an order of magnitude (Rohwer, 2003; Sime-Ngando, 2014). Viruses are a large reservoir of genetic diversity (Suttle, 2007; Sullivan, Weitz & Wilhelm, 2017; Angly et al., 2006; Brum et al., 2015) and occur in all habitats (Paez-Espino et al., 2016), including air (Reche et al., 2018; Rosario et al., 2018), soils (Srinivasiah et al., 2008; Williamson et al., 2017) and deep-sea sediments (Danovaro et al., 2008). The study of viral ecology was pioneered in surface marine systems where viruses that infect bacteria, also known as bacteriophages, are the main source of bacterial mortality (Suttle, 2007; Brum & Sullivan, 2015; Gregory et al., 2019) and impact ecosystem functions such as the cycling of carbon (Breitbart et al., 2018). By lysing their hosts,

horizontal gene transfer and metabolic reprogramming, bacteriophages play a pivotal role in structuring microbial communities (Skvortsov *et al.*, 2016; Silva *et al.*, 2017; Rossum *et al.*, 2018; Daly *et al.*, 2019), the flow of energy and matter through food webs (Weitz *et al.*, 2015), the cycling of carbon and nutrients (Dell'Anno, Corinaldesi & Danovaro, 2015; Guidi *et al.*, 2016; Emerson *et al.*, 2018) and the evolution of bacteria (Pal *et al.*, 2007; Rodriguez-Valera *et al.*, 2009; Simmons *et al.*, 2019). Many of these strides have only recently been possible with the advent of molecular tools such as metagenomic sequencing (Rosario & Breitbart, 2011; Brum & Sullivan, 2015; Roux *et al.*, 2016, 2017). In this context, one may differentiate between untargeted approaches, which have produced a wealth of viral sequences (Paez-Espino *et al.*, 2016) and virome sequencing, in which the viral fraction is purified prior to sequencing, thus ensuring that the sequencing effort is targeted towards viral nucleic acids (Rosario & Breitbart, 2011).

Besides the early recognition of the potential of phages to eradicate bacterial biofilms in medical settings (Chan & Abedon, 2015), little is known about the interactions between biofilms and viruses (Sutherland *et al.*, 2004). While the biofilm matrix may impede the access of viruses to the surface of bacterial cells (Vidakovic *et al.*, 2017), the susceptibility to phage-induced clearance of biofilm has been demonstrated in laboratory experiments (Scanlan & Buckling, 2012). Biofilms may also act as a reservoir for virus amplification and viruses may endure periods of unfavorable environmental conditions within the extracellular matrix (Doolittle, Cooney & Caldwell, 1996; Briandet *et al.*, 2008). Apart from a few examples (Dann *et al.*, 2016; Silva *et al.*, 2017; Rossum *et al.*, 2018), we lack an understanding of the ecological role of viruses in streams and rivers in general (Peduzzi, 2016) and in stream biofilms in particular (Battin *et al.*, 2016). In streams, biofilms colonize the sedimentary surfaces of the streambed, are biodiversity hotspots comprising members from all three domains of life and fulfill critical ecosystem processes (Battin *et al.*, 2016). It is reasonable therefore to speculate that viruses also control biodiversity and biomass turnover in stream biofilms with potential consequences for ecosystem functioning and biogeochemical cycling. However, the heterogeneous matrix of stream biofilms has precluded the study of viromes in stream biofilms.

In general, viral metagenomics is complicated by the vast diversity of viruses and lack of universal marker genes, the risk of contamination with non-viral DNA and the underrepresentation of viral sequences in databases (Thurber *et al.*, 2009; Hayes *et al.*, 2017). Protocols for sample preparation for viral metagenomics therefore aim at concentrating and purifying viral-like particles (VLPs) and removing contaminating DNA, while optimizing VLP recovery (Castro-Mejia *et al.*, 2015). However, it is clear that viruses are lost at every step of these protocols. Large viruses, in the size range typical for bacteria, may be removed by filtration. Some viruses are sensitive to chemicals during purification based on the structure of their capsids, while other viruses may be lost because of differences in buoyant density, critical during purification in density gradient ultracentrifugation (Thurber *et al.*, 2009). It is thus clear that no single protocol to extract all viruses exists and modifying as well as rigorous testing of protocols remains a critical task.

Table 1 Sample site and biofilm characteristics.

	VDN	VEV	SNG
Coordinates	46°15'13.5"N 7°06'33.9"E	46°30'46.4"N 6°54'43.7"E	46°33'23.9"N 6°28'55.3"E
Altitude (m a.s.l.)	1,210	766	498
Bacterial abundance (cells m ⁻²)	2.3 × 10 ⁹	1.8 × 10 ¹⁰	4.1 × 10 ¹¹
Chlorophyll-a (µg cm ⁻²)	0.21	0.22	1.21
EPS proteins (µg cm ⁻²)	0.05	0.04	0.24
EPS carbohydrates (µg cm ⁻²)	Below detection	Below detection	0.08

The first critical step towards the study of biofilm viruses in streams using virome sequencing requires the effective extraction of VLPs from the biofilm matrix. The aim of our study was therefore to optimize a sample-to-sequence pipeline including VLP extraction, concentration and purification towards metagenomic analyses. Our effort was specifically tailored to maximize the yield of viral DNA while minimizing DNA contamination from bacteria, eukaryotic hosts or the environment. To this end, we compared a suite of sequential protocols to generate viral metagenomes from stream biofilms including tangential flow filtration (TFF), polyethylene glycol (PEG) precipitation, physicochemically induced biofilm breakup and ultracentrifugation followed by nucleic acid extraction. Several of these protocols have been described previously ([Danovaro et al., 2001](#); [Thurber et al., 2009](#); [Danovaro & Middelboe, 2010](#); [Hurwitz et al., 2013](#); [Temmam et al., 2015](#); [Trubl et al., 2016](#); [Hayes et al., 2017](#) for reviews), but their rigorous testing for virome generation from stream biofilms is lacking at present. We benchmark the efficiency of the different protocols using epifluorescence microscopy counts of VLP and DNA yields and provide first results of the virome structure from biofilms in three streams. We provide a step-by-step version of the optimized protocol at protocols.io, which allows for community participation and continuous protocol development.

MATERIALS AND METHODS

Sampling

We sampled benthic biofilms from three streams (Switzerland) draining catchments differing in altitude and land use ([Table 1](#)). The Vallon de Nant (VDN) catchment is pristine with vegetation dominated by alpine forests and meadows. The Veveyse (VEV) catchment is characterized by mixed deciduous forests, but also agricultural and urban land use. The Senoge (SNG) catchment is clearly impacted by agricultural land use. During winter, benthic biofilms were randomly collected from stones (5–15 cm in diameter) using sterile brushes and Milli-Q water. Depending on biofilm thickness, we scraped biofilms from cobbles with a total surface area ranging from 1.3 to 2.4 m² into 10 L Milli-Q water. Slurries were transported on ice to the laboratory pending further processing.

Biofilm properties

For bacterial cell counting, samples were fixed with 3.7% formaldehyde (final concentration) and stored at 4 °C. Bacterial cells were disintegrated from the biofilm

matrix using 0.25 mM tetrasodium pyrophosphate in combination with rigorous shaking (1 h) and sonication (Branson Sonifier 450, Branson, MO, USA) on ice (1 min) (Velji & Albright, 1993). As previously established (Besemer et al., 2009), cells were stained using Syto13 and counted on a flow cytometer (NovoCyte, ACEA Biosciences, San Diego, CA, USA). Chlorophyll *a* content was determined spectrophotometrically after over-night extraction in 90% EtOH (Lorenzen, 1967). Extracellular polymeric substances (EPS) were extracted from biofilm slurries using 50 mM EDTA and shaking (1 h) (Battin et al., 2003). Carbohydrates were precipitated in 70% EtOH (−20 °C, 48 h) and measured spectrophotometrically as glucose equivalents (DuBois et al., 1956). Proteins were determined according to Lowry et al. (1951).

Transmission electron microscopy

A first confirmation of the presence of VLP in biofilms was obtained from TEM. For this, 5 µL of unprocessed sample was adsorbed to a glow-discharged carbon-coated copper grid (Canemco & Marivac, Gore, QC, Canada), washed with deionized water and stained with 5 µL of 2% uranyl acetate. TEM observations were made on an F20 electron microscope (ThermoFisher, Hillsboro, OR, USA) operated at 200 kV and equipped with a 4,098 × 4,098 pixel camera (CETA, ThermoFisher, Waltham, MA, USA). Magnification ranged from 10,000 to 29,000×, using a defocus range of −1.5 to −2.5 µm.

Protocols for the extraction of VLP

In order to establish an optimized stream biofilm sample-to-sequence pipeline, we explored a variety of protocols for the concentration, extraction and purification of VLP (Fig. 1). The pipeline consists of three main parts: concentration, extraction and purification. The concentration step is required to obtain sufficient genetic material for nucleic acid extraction. Extraction aims to liberate VLP from the biofilm matrix, while purification aims to reduce the amount of contaminant nucleic acids and cellular debris. The order—first volume reduction, then chemical detachment—was chosen because chemicals used for extraction of VLP may damage tangential flow filters. We tested all possible combinations of protocols to identify the most promising laboratory pipeline for the preparation of viral metagenomes.

First, we homogenized the biofilm slurries by manual shaking and split samples into aliquots (1 L). Aliquots were centrifuged at 100×g (15 min) (5810R; Eppendorf, Hamburg, Germany) to remove large sediment particles and organisms contained within the biofilm slurry. We recovered the supernatant for downstream analyses.

Concentration

We evaluated TFF and PEG precipitation to concentrate the viral size fraction. A medium-scale TFF system equipped with a 100 kDa tangential flow filter (GE Healthcare, Chicago, IL, USA) was used (Thurber et al., 2009). Viruses were collected in the retentate, whereas water and particles smaller than the pore size were discarded. For each aliquot, the initial volume was reduced to less than 50 mL.

PEG precipitation was performed as described previously (Bibby & Peccia, 2013). Aliquot biofilm samples were supplemented with 10% w/v PEG 8000 (Sigma–Aldrich,

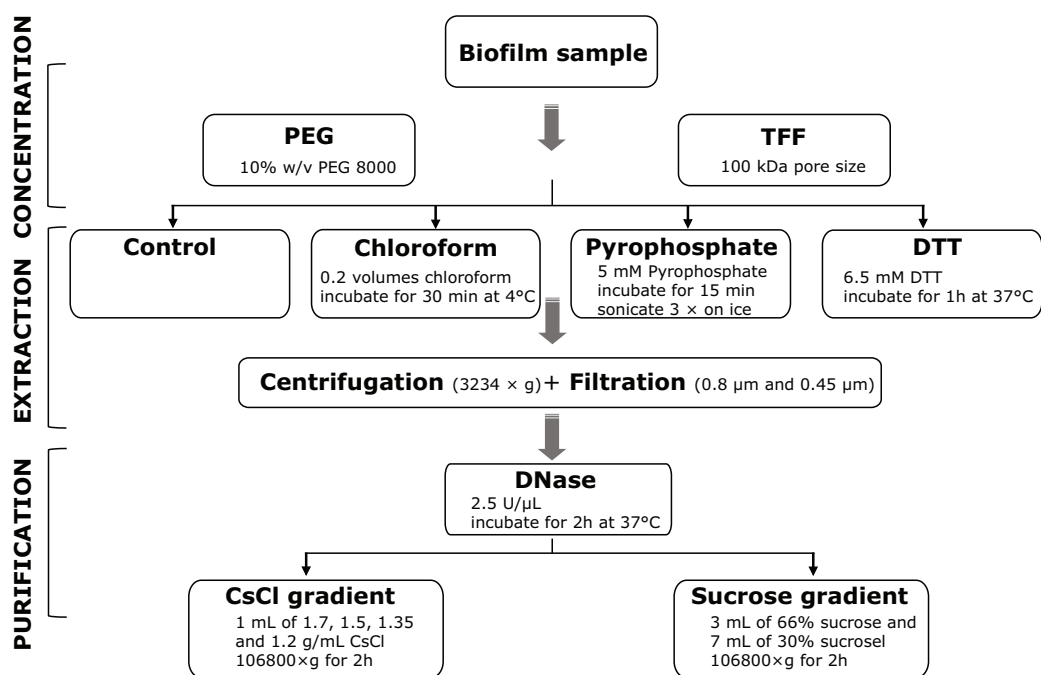


Figure 1 Overview of methods for the extraction and purification of viruses from stream biofilms. First, phages are concentrated using either PEG precipitation or TFF. Different physico-chemical extraction procedures were then evaluated for their efficiency. Prior to DNase I digestion, centrifugation and filtration was used to remove cell debris from all samples. Finally, ultracentrifugation in sucrose or CsCl density gradients was used to purify viruses for downstream molecular analyses. Combinations of all protocols were evaluated for the recovery of VLPs and DNA yield.

Full-size DOI: 10.7717/peerj.8187/fig-1

Steinheim, Germany) and 2.5 M NaCl (Sigma–Aldrich, Steinheim, Germany). The supernatant was then agitated by inverting the tubes three times and stored overnight (4 °C), followed by centrifugation for 30 min at 9,432×g at 4 °C (Sorvall RC-5C centrifuge, HS-4 rotor; ThermoFisher, Waltham, MA, USA). Supernatants were carefully removed and the resulting pellets were eluted with 50 mL sterile H₂O.

Extraction

We tested three different physicochemical treatments (chloroform, tetrasodium pyrophosphate in combination with sonication and dithiothreitol (DTT)) to dislodge viruses from the biofilm matrix. To one subset of the samples, we added 0.2 volumes of chloroform and mixed by inversion. Then, these samples were incubated at 4 °C for 30 min, vortexed every 5 min and centrifuged at 3,234×g for 15 min at 4 °C (5810R; Eppendorf, Hamburg, Germany) to recover the supernatant (*Marhaver, Edwards & Rohwer, 2008; Thurber et al., 2008*).

Following the recommendation of *Danovaro & Middelboe (2010)*, we used sonication in combination with tetrasodium phosphate to separate viruses from biofilms. For this, we added five mM of tetrasodium pyrophosphate (final concentration) to another subset of concentrated biofilm samples and incubated them in the dark (15 min). Then, all samples were sonicated (frequency: 40 kHz; Bransonic Sonifier 450, Branson, MO, USA) three

times for 1 min with 30-s intervals during which the samples were manually shaken. To prevent heating, samples were kept on ice during sonication. Finally, a subset of the samples was treated with 6.5 mM DTT and incubated for 1 h at 37 °C (Lim *et al.*, 2014). At the end of the incubation period, samples were chilled on ice. Biofilm samples without chemical treatment served as controls. After each treatment, these samples were first centrifuged (4 °C) at 3,234×g for 15 min (5810R; Eppendorf, Hamburg, Germany) and the supernatant was sequentially filtered through 0.8 µm and 0.45 µm filters (Whatman, Maidstone, UK; GE Healthcare, Chicago, IL, USA) to remove debris and cells.

Purification

To eliminate contaminating eukaryotic, prokaryotic and extracellular nucleic acids, DNase I at a final concentration of 2.5 U/µL (Life Technologies, Darmstadt, Germany) and Tris buffer (100 mM Tris pH 7.5, five mM CaCl₂ and 25 mM MgCl₂) were added to the clarified supernatant and incubated at 37 °C (2 h). To confirm the removal of contaminant DNA, polymerase chain reaction (PCR) targeting the 16S rRNA gene was performed using the universal primer pair 341f (5'-CCTACGGGNGGCWGCAG-3') and 785r (5'-GACTACHVGGGTATCTAAKCC-3') (Thijs *et al.*, 2017). PCR amplification was performed with a Biometra Thermal Cycler using Taq ALLin polymerase (Axon Lab, Baden, Switzerland) according to the manufacturer's instructions with an initial enzyme activation step (95 °C for 2 min) followed by 30 cycles of denaturation (95 °C for 30 s) and hybridization (72 °C for 30 s) and a final elongation (72 °C for 10 min). PCR products were visualized under UV light after migration on an agarose gel stained using GelRed (Biotium, Hayward, CA, USA). We did not assess the potential of PCR inhibition.

For final purification of the viral fraction, two protocols of density gradient ultracentrifugation, sucrose and CsCl density gradients, were assessed. The sucrose density gradient was achieved by three mL of 0.2-µm filtered 66% sucrose and seven mL of 0.2 µm filtered 30% sucrose (Temmam *et al.*, 2015). A subset of sample was deposited on top of the sucrose density gradient and centrifuged at 106,800×g (Optima XPN-80 Ultracentrifuge, 32 SWi; Beckman Coulter, Indianapolis, IN, USA) for 2 h at 4 °C. The viral fraction was harvested by retrieving 1.5 mL from the interface between both layers using an 18G needle. Similarly, subsamples were deposited onto CsCl density gradients composed of layers of one mL of 1.7, 1.5, 1.35 and 1.2 g/mL CsCl. The purified viral fraction was harvested from just below the interface between the 1.5 and 1.7 g/mL CsCl layers by retrieving 1.5 mL using an 18G needle.

Enumerating VLPs using epifluorescence microscopy

To assess the relative viral extraction efficiency of the various protocols, we counted VLP under an epifluorescence microscopy (AxioImager Z2, Zeiss, Germany) as described by Patel *et al.* (2007). Because of high background noise owing to the biofilm matrix constituents, samples could only be counted after the purification steps. It should be noted that pyrophosphate concentrations >10 mM tend to interfere with VLP counting (Danovaro *et al.*, 2001); however, we used five mM pyrophosphate during VLP liberation from biofilms. Subsamples were fixed with formaldehyde, stained with 0.5 µL of

1,000 × SYBR Gold (Molecular Probes; ThermoScientific, Waltham, MA, USA) and incubated at room temperature in the dark (30 min). After incubation, each sample was filtered onto a 0.02 µm pore size membrane filter (Anodisc; Whatman, Maidstone, UK). The filters were mounted on glass slides with a drop of VECTASHIELD antifade mounting medium (Vector Laboratories, Burlingame, CA, USA). VLP were visualized using blue light (488 nm) excitation and green (512 nm) emission. For each sample, 15–20 randomly selected images were acquired with a camera (Axiocam 506 mono, Zeiss, Germany) mounted onto the microscope. VLPs were discriminated from bacteria by size (0.015–0.2 µm) and enumerated using a custom script in Fiji ([Schindelin et al., 2012](#)).

Nucleic acid extraction

To further assess the efficiency of the tested protocols to obtain viral nucleic acids for metagenome sequencing, we extracted DNA and quantified its concentration. Following [Sambrook, Fritsch & Maniatis \(1989\)](#), 0.1 volume of TE buffer, 0.01 volume of 0.5 M EDTA (pH 8) and 1 volume of formamide was added to purified viral samples, and incubated at room temperature (30 min). Then, two volumes of cold 100% EtOH were added and incubated for 30 min (4 °C). Samples were centrifuged at 17,000×g (4 °C, 20 min) and pellets washed twice with 70% cold EtOH. Pellets were air-dried and resuspended in 567 µL of TE buffer (10 mM Tris and one mM EDTA (pH 8.0)). Thirty µL of warm 10% SDS and three µL of proteinase K (20 mg/mL) were added and samples were incubated for 1 h at 37 °C. Next, 100 µL of five M NaCl and 80 µL of warm hexadecyltrimethylammonium bromide (CTAB) were added and incubated for 10 min (65 °C) ([Doyle & Doyle, 1987](#)). DNA was extracted in a series of chloroform, phenol: chloroform:isoamyl alcohol (25:24:1) and chloroform treatments, with centrifugation at 16,000×g (10 min) for phase separation. Finally, DNA was precipitated overnight in isopropanol (–20 °C). DNA was concentrated by centrifugation at 16,000×g (4°C) for 20 min, and the pellet washed twice with cold 70% EtOH, air-dried, resuspended in 50 µL nuclease-free H₂O and stored (–20 °C). DNA concentration was measured using Qubit and the dsDNA high-sensitivity kit according to the manufacturer's instructions (Life Technologies, Carlsbad, CA, USA).

Library construction and sequencing

Based on the epifluorescence microscopy counts of VLP and the DNA concentration from purified biofilm samples, we selected samples processed with the best performing pipeline for metagenome sequencing. For sequencing library construction, DNA was sheared with an S2 focused ultrasonicator (Covaris, Woburn, MA, USA) to achieve a target size of DNA fragments of around 350 bp. We opted for the ACCEL-NGS[®] 1S PLUS DNA library kit (Swift Biosciences, Ann Arbor, MI, USA) which allows low quantities of both single- and double-stranded DNA as input ([Roux et al., 2016](#)). Library construction and multiplexing was performed following the manufacturer's instructions for DNA inputs (<1 ng/µL) and 20 cycles of indexing PCR. Paired-end sequencing (2 × 300 bp) was performed on a MiSeq System (Illumina, San Diego, CA, USA) at the Lausanne Genomic

Technologies Facilities. Raw sequences have been submitted to the European Nucleotide Archive under accession number [PRJEB33548](#).

Bioinformatic analyses

The number of reads and GC content of each sample before and after quality control were calculated using a custom python script. For virome classification, we followed the recommendations to assemble viral contiguous sequences (contigs) according to [Roux et al. \(2019\)](#). First, BBDuk (v35.79) was used to remove Illumina adapters, filtering and trimming (trimq = 12). Next, reads with >93% similarity to a human reference genome were discarded (using BBmap). The ACCEL-NGS[®] 1S Plus library preparation kit includes a low complexity adaptase tail, which was clipped (10 bases) according to the manufacturer's instructions. We then used the error correction capability of Tadpole (v. 37.76) to correct for sequencing errors (mode = correct ecc = t prefilter = 2). Prior to assembly, we de-duplicated our datasets using clumpify (v37.76) with parameters set such that identical reads were identified and only one copy was retained (dedupe subs = 0). Finally, we used the SPAdes assembler (v3.13.0, [Bankevich et al., 2012](#)) in single-cell (—sc) mode, with error correction disabled (—only-assembler) and kmers set to 21, 33, 55, 77, 99, 127 to assemble contigs. We co-assembled the paired-end reads from both sequencing runs for each sample individually. To obtain an overview of the potential contaminant sequences (e.g., human, bacterial and phiX), we uploaded the quality-trimmed and de-replicated reads (forward orientation only) to the web interface of taxonomer ([Flygare et al., 2016](#)). Taxonomer is an ultrafast taxonomy assigner, which assigns and classifies reads to human, bacterial, viral, phage, fungal, phix, ambiguous (i.e., reads fit to more than one bin) and “unknown” bins. For identification and classification of viral contigs, we used MetaPhinder (v2.1, [Jurtz et al., 2016](#)) through the web interface hosted at the Center for Genomic Epidemiology at the Danish Technical University (DTU). We mapped the reads to the contigs using BWA-MEM (<http://bio-bwa.sourceforge.net/>) using default setting (e.g., a mismatch penalty of four and a minimal alignment score of 30) and then counted the number of reads mapping each contig using samtools ([Li et al., 2009](#)). The average number of reads per contig were adjusted by contig length and normalized by library size to obtain the average percent of reads mapping to contigs. In order to specifically target ssDNA contigs ([Trubl et al., 2019](#)), we matched the contigs to the Viral_rep and Phage_F domains of the PFAM database using hmmsearch (HMMER v3, [Eddy, 2011](#)) with cutoff scores ≥ 50 and e -values ≤ 0.001 .

RESULTS

Biofilm properties

Bacterial abundance ranged between 4.1×10^{11} cells m^{-2} at the lowest stream (SNG) and 2.3×10^9 cells m^{-2} at the uppermost stream (VDN; [Table 1](#)). This pattern was mirrored by an increase in chlorophyll *a* content and the protein and carbohydrate concentrations in EPS that were higher in the lowland than the high-altitude stream ([Table 1](#)). Despite these differences in biofilm properties, no differences in extraction efficiency among the different protocols were detected. In fact, the average number of

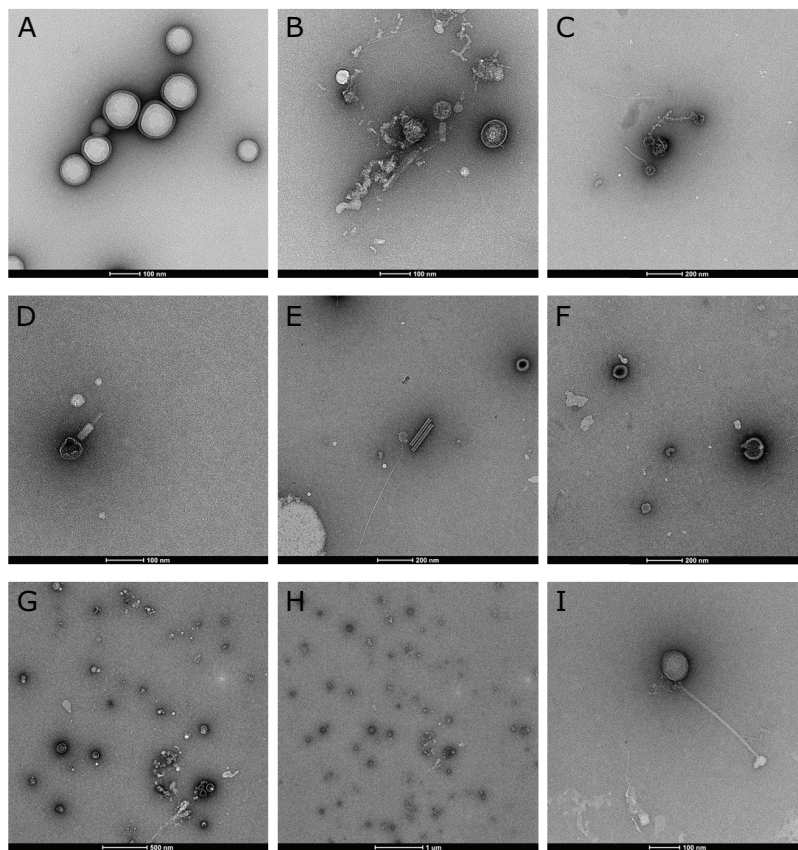


Figure 2 Electron microscopic evidence of virus-like particles in stream biofilm samples. (A) large morphological diversity of VLPs, including tailed bacteriophages (B, C, D, I), lemon-shaped, polyhedral (e.g. B, H), spherical and filamentous viruses was observed using TEM. Some morphologies resembled plant viruses (E) while amorphous structures (A, F, G, H) could also be membrane vesicles.

Full-size  DOI: [10.7717/peerj.8187/fig-2](https://doi.org/10.7717/peerj.8187/fig-2)

VLPs extracted by the various protocols were strongly correlated among the different samples (pairwise Spearman's r_s ranging between 0.94 and 0.98).

Transmission electron microscopy

Direct electron microscopic observations of raw biofilm samples revealed the presence and morphological diversity of virions, including tailed bacteriophages and lemon-shaped viruses, in biofilms from all three streams (Fig. 2). Polyhedral, spherical and filamentous VLPs were also observed, which may include untailed bacteriophages or viruses infecting eukaryotes. Some amorphous structures may also be interpreted as membrane vesicles.

Virome extraction and purification

In total, we evaluated 16 protocols to concentrate, extract and purify viruses from benthic biofilms from three streams (Fig. 1). Based on the number of VLPs and DNA yields retained at the end of each protocol, we observed significant differences in relative extraction efficiencies among protocols. Across all samples, average VLP counts and DNA yields were correlated (Spearman's $r_s = 0.74$), suggesting conformity among these two

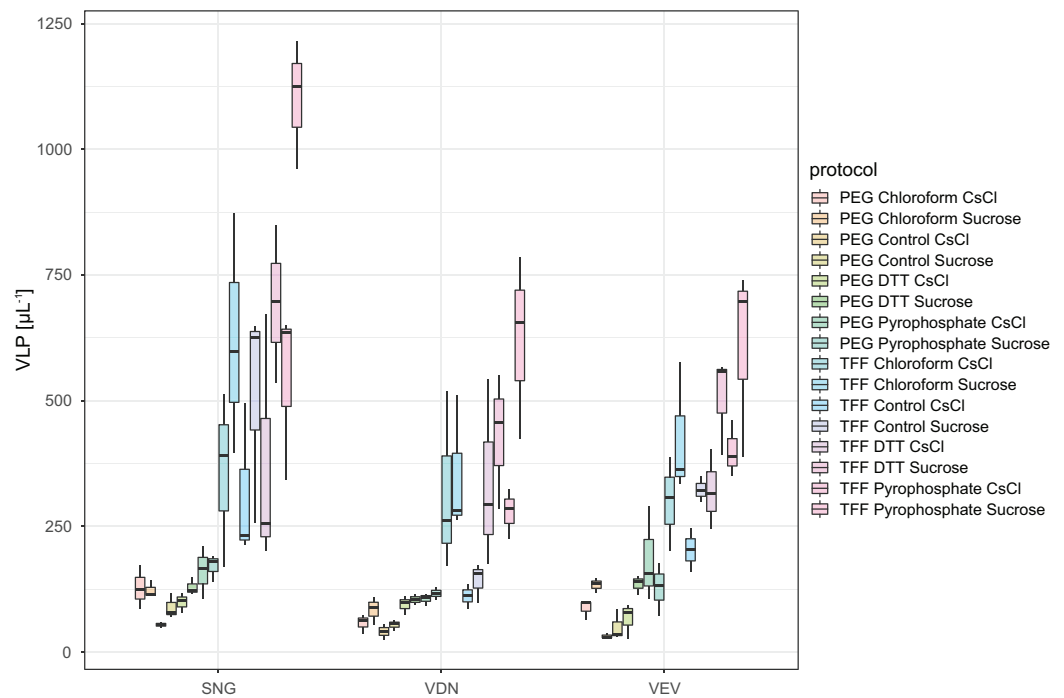


Figure 3 VLP yields of the different combination of methods for the extraction and purification of viruses from stream biofilms. The boxes show the number of VLPs in replicated subsamples processed with each protocol, the median is given as a horizontal line, hinges correspond to the first and third quartile while whiskers extend to the largest and smallest values counted. In each of the three streams tested, a protocol based on TFF, pyrophosphate in combination with sonication and ultracentrifugation in sucrose gradient yielded significantly higher VLP counts than any other combination of protocols.

Full-size DOI: 10.7717/peerj.8187/fig-3

means of evaluation. The combination of TFF for concentration, tetrasodium pyrophosphate and sonication for extraction, and sucrose gradient centrifugation for purification resulted in the highest VLP counts in all three samples (two-way ANOVA, $p < 0.01$; Fig. 3). This pipeline also yielded the highest DNA concentration (Fig. 4). Protocols involving PEG precipitation generally resulted in a lower recovery of VLPs (on average only 13.6% compared to the best performing pipeline) and DNA yields below detection limit. This may be attributable to the formation of a visible, viscous layer upon addition of PEG to the biofilm samples. The dissociation of VLPs from the biofilm matrix was most effective using tetrasodium pyrophosphate and sonication (two-way ANOVA, $p < 0.01$). Protocols based on TFF and using DTT or chloroform extracted on average 25.0% and 33.2% less VLPs than protocols using TFF followed by tetrasodium pyrophosphate treatment and sonication (Fig. 3). Samples without any physicochemical treatment to extract VLP from biofilms yielded on average 54.8% less VLPs than the tetrasodium pyrophosphate and sonication treatment. To further purify viruses, two discontinuous gradient formulations with sucrose and CsCl were tested. On average, 1.9 times more VLPs were retained by ultracentrifugation using the sucrose gradient than using the CsCl gradient in samples concentrated using TFF and treated with tetrasodium pyrophosphate and sonication (paired-*T*-Test, $p < 0.01$). This is also reflected

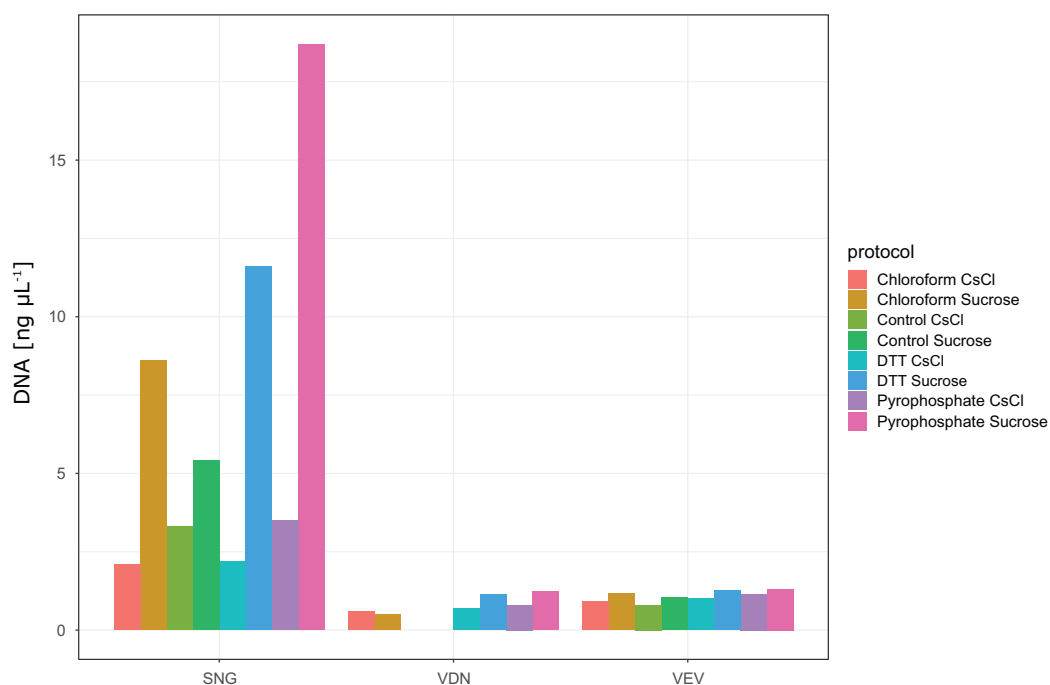


Figure 4 DNA yields from samples processed with different protocols. None of the protocols involving PEG precipitation resulted in detectable DNA yields (see DOI: 10.6084/m9.figshare.8943341). Protocols using TFF and either no treatment (control), chloroform, DTT, pyrophosphate in combination with sonication for extraction and either CsCl or Sucrose gradient ultracentrifugation for purification yielded detectable DNA concentrations. Highest DNA yields were obtained using a combination of TFF, followed by pyrophosphate and sonication and sucrose density gradient ultracentrifugation. Note that although VLP counts were only 1.8 times higher in the best performing protocol in SNG as compared to the other two samples (VDN, VEV), DNA yields in this sample were more than 14 times higher.

Full-size DOI: 10.7717/peerj.8187/fig-4

in DNA yields, which reached $1.22 \text{ ng } \mu\text{L}^{-1}$ in VDN, $1.31 \text{ ng } \mu\text{L}^{-1}$ in VEV and $18.7 \text{ ng } \mu\text{L}^{-1}$ in SNG using TFF, pyrophosphate and sonication followed by sucrose gradient centrifugation. DNA yields were on average 3 times higher using ultracentrifugation in the sucrose compared to the CsCl gradient (paired *T*-Test, $p = 0.03$) and on average 1.5, 1.7 and 1.9 times higher using tetrasodium pyrophosphate and sonication compared to DTT, no physico-chemical detachment and chloroform, respectively.

Given that stream biofilms contain abundant prokaryotic, eukaryotic and extracellular DNA, it is crucial to verify the absence of DNA potentially contaminating the samples. Negative PCR results from samples treated with DNase I confirmed the absence of DNA contamination from prokaryotic cells.

Stream biofilm viromes

From the two sequencing runs we obtained 24388096, 22459218 and 23804190 paired-end reads from SNG, VEV and VDN, respectively (Table 2). After quality control, error correction and deduplication, on average 97.5% of the reads remained. Initial screening using taxonomer showed that human contaminant reads accounted for 0.13–0.25% of the reads, while bacterial contaminant reads accounted for 1.47–8.11% of the reads.

Table 2 Virome dataset statistics.

	Raw reads		Quality trimmed		Contigs	Average contig length (bp)
	Paired-end sequences	GC content (%)	Paired-end sequences	GC content (%)		
SNG	24 388 096	40.36	24 172 084	40.16	3,698	452
VEV	22 459 218	40.51	22 153 818	40.24	11,323	645
VDN	23 804 190	40.68	23 422 408	40.21	13,591	676

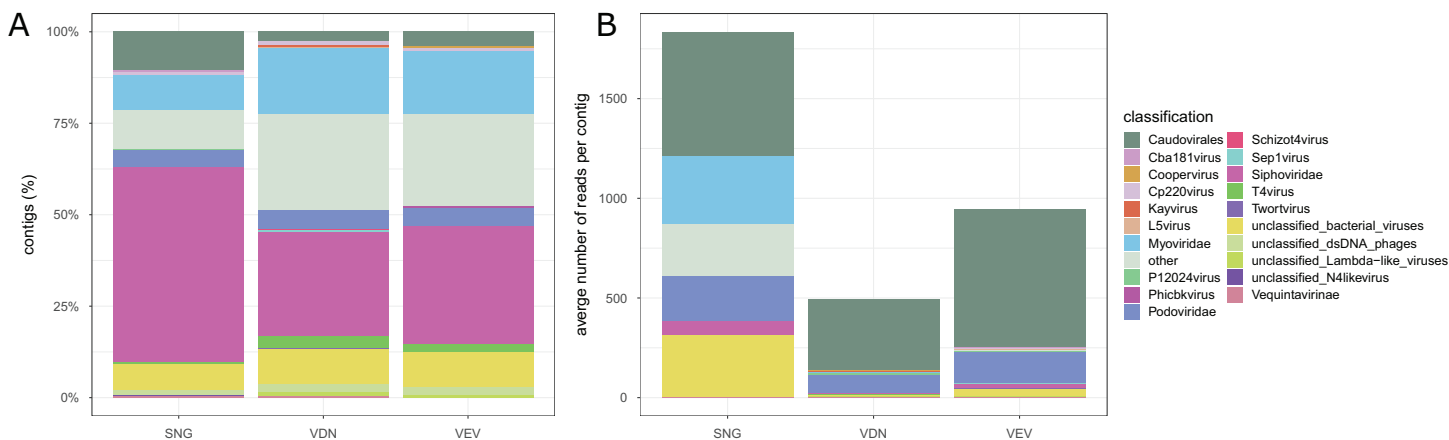


Figure 5 Stream biofilm virome taxonomic composition. The relative number of classified contigs among the viromes are shown (A). Stream biofilm viromes were dominated by not further classified members of *Siphoviridae* and *Myoviridae*. The relative composition of contigs was similar among the three samples, however, markedly fewer contigs were obtained from SNG than from the other two stream biofilm samples. ssDNA viruses could not be identified among contigs from any of the streams. The average number of reads mapped to classified contigs in the three samples (B).

Full-size [DOI: 10.7717/peerj.8187/fig-5](https://doi.org/10.7717/peerj.8187/fig-5)

We obtained 3,698, 11,323 and 13,591 contigs from de-novo assembly of quality-controlled and deduplicated reads from SNG, VEV and VDN, respectively. The largest contigs were 9,493, 46,665 and 54,492 bp in SNG, VEV and VDN, respectively. Hmsearch against the PFAM databases did not yield ssDNA contigs in the three viromes. Between 726 and 2613 contigs were classified as of viral origin in the three viromes (Fig. 5). In all samples, the majority of contigs were identified as not further classified *Siphoviridae* (28.3–53.3%), followed by *Myoviridae* (9.6–18.0%) and *Podoviridae* (4.7–5.2%). Contigs classified as *T4virus*, *Cp220virus*, *Kayvirus* and *P12024virus* were common in all three biofilm viromes. However, contigs classified as *Twortvirus*, *Phicbkvirus*, *Coopervirus* and *L5virus* were only detected in viromes obtained from VEV and VDN but not in SNG. Across the three samples, on average 554.6 reads mapped to contigs classified as *Caudovirales*, averaging 63.4, 63.8 and 89.1% accounting for contig size in SNG, VEV and VDN, respectively. On average, 153.8 reads mapped to *Podoviridae* (range 2.2–30.8% of bp) and 115.1 reads mapped to *Myoviridae* (range 0.4–6.1% of bp), whereas only 29.4 reads mapped to *Siphoviridae* (range 1.8–17.7% of bp). There was substantial variation in the number of reads mapped to contigs in the different samples. For instance, more reads mapped on average to unclassified bacterial viruses and *Myoviridae* in SNG (311.1 and 339.7 reads accounting for 9.3 and 6.1% of bp) than in VDN

(2.6 and 2.8 reads accounting for 0.6 and 0.8% bp) or VEV (36.7 and 2.8 reads accounting for 1.5 and 0.4% bp).

DISCUSSION

The advent of metagenomic tools has revolutionized the study of the role of viruses in numerous ecosystems (Suttle, 2007; Rosario & Breitbart, 2011; Brum & Sullivan, 2015; Trubl et al., 2019). For biofilms in general, and particularly for biofilms in streams and rivers, however, an optimized protocol for viral metagenomics has been missing. Here, we establish an optimized sample-to-sequence pipeline for the concentration and purification of viruses from stream biofilms. This protocol is publicly available at <https://www.protocols.io/view/extraction-and-purification-of-viruses-from-stream-32qgqdw>. We used two metrics, the number of VLP retained and the amount of DNA extracted from the samples to evaluate the performance of the different combinations of protocols (Figs. 3 and 4). This approach allowed us to obtain a relative comparison among the tested protocols. However, it does not permit the quantification of extraction efficiencies since it was not possible to enumerate VLPs without prior extraction and purification.

The best performing sample-to-sequence pipeline involves TFF for sample concentration, pyrophosphate and sonication for the detachment of viruses from the biofilm matrix and DNase I treatment followed by sucrose gradient ultracentrifugation for purification. This is similar to protocols for other complex samples, such as soils (Trubl et al., 2016) or marine and freshwater sediments (Danovaro & Middelboe, 2010). The suggested combination of protocols was consistently the best performing pipeline for biofilms obtained from three streams differing in trophic state and with different biofilm properties (Table 1). However, depending on biofilm biomass, sampling efforts should be tailored towards obtaining sufficient nucleic material for virome metagenomic sequencing. For instance, in the oligotrophic mountain stream, sampling biofilms from 0.5 m² (approximately 10–12 pebbles of 5–10 cm diameter) and using 1 L of Milli-Q water suffices to generate viromes using the optimized protocols and an appropriate library preparation strategy.

The viromes obtained using the best-performing protocol were dominated by reads of viral or unknown origin (presumably reflecting the lack of viral sequences in public databases), and contaminant sequences (i.e., of bacterial or human origin) contributed only marginally to the viromes. Following an optimized assembly strategy (Roux et al., 2019), the reads assembled into a large number of contigs, however, of small average contig size. Still many of the contigs were classified as of viral origin. Viral community composition was remarkably similar across the three different stream biofilms, with several contigs classified as *T4virus*, *Cp220virus*, *Kayvirus* and *P12024virus* found in all samples. Despite the use of the ACCEL-NGS[®] 1S PLUS DNA kit (Roux et al., 2016), ssDNA viruses, which account for <5% of DNA viral communities in other freshwater, marine and soil ecosystems (Roux et al., 2016; Trubl et al., 2019) could not be detected in the viromes from stream biofilms. This may be related to the choice of the density gradient harvested during the final purification step (Kauffman et al., 2018). Due to the low buoyant density of ssDNA viruses (Thurber et al., 2009), additionally sampling from, for

instance, a 1.3 g/mL CsCl density layer (Trubl *et al.*, 2019) may be advisable to specifically target ssDNA viruses. Strikingly, the virome obtained from the most eutrophic stream (SNG) resulted in the lowest number of viral contigs and lacked contigs classified as *Twortvirus*, *Phicbkvirus*, *Coopervirus* and *L5virus*. However, given the low number of samples, we caution against concluding an anthropogenic effect on stream biofilm viral communities.

A range of chemical properties may explain the relative differences in virus extraction efficiency from stream biofilms. PEG precipitation generally failed to concentrate viruses from biofilm slurries, potentially due to the formation of a viscous layer that impaired PEG removal. Previously described methods for isolating viruses involve chloroform (Marhaver, Edwards & Rohwer, 2008; Thurber *et al.*, 2008; Hewson *et al.*, 2012); however, chloroform may denature the lipid envelopes surrounding viral capsids, internal lipid membranes. Nucleo-cytoplasmic large DNA viruses (NCLDV), including Phycodnaviridae, which predominantly infect freshwater and marine algae, may thus be sensitive to chloroform treatment (Feldman & Wang, 1961). Similarly, DTT is a reducing agent, which breaks disulfide bonds in proteins. This may explain the reduced recovery of VLPs from samples treated with PEG, chloroform or DTT.

Extracellular DNA is a common component of the biofilm matrix (Flemming & Wingender, 2010), which together with DNA from damaged eukaryotic and prokaryotic cells needs to be removed prior to virome sequencing. We propose a DNase I treatment as an efficient way to digest extracellular DNA from biofilms, followed by sucrose density gradient ultracentrifugation for further purification. However, DNase I will not degrade ssDNA or RNA and recent studies used nuclease cocktails including DNases, RNases and Benzoases (Temmam *et al.*, 2015; Rosario *et al.*, 2018) to eliminate various contaminant nucleic acids. The treatment order is important because otherwise DNA from viral particles damaged during ultracentrifugation may be digested by the DNase digestion. Compared to ultracentrifugation in a CsCl gradient, sucrose gradient ultracentrifugation probably maximized the purification of a wider range of viruses because of the larger gradient of densities recovered with this method. Moreover, CsCl density gradient separation may exclude viral particles as that may be too buoyant (Thurber *et al.*, 2009; Kauffman *et al.*, 2018), or degrade the structure of enveloped viruses (Lawrence & Steward, 2010) and therefore reduce their recovery. Critical steps of virome preparation concerns nucleic acid extraction and library preparation. We chose a derivation of a standard extraction protocol, which allows the extraction of both ssDNA and dsDNA viruses (Sambrook, Fritsch & Maniatis, 1989). Optimization of DNA extraction protocols, such as done for soil viromes (Trubl *et al.*, 2016), may be necessary for biofilms sampled from environments with high humic acid concentrations such as in boreal streams or streams draining wetlands, potentially causing inhibition during library preparation.

Finally, there is a plethora of tools available for the bioinformatic analysis of viromes (e.g., those implemented in iVirus (Bolduc *et al.*, 2017) and available through powerful computational infrastructures such as CyVerse (www.cyverse.org) or KBase (www.kbase.us). Here, we opted for contig assembly, classification and read mapping to assess the capability of the laboratory procedures to generate diverse viromes from stream

biofilm. Clearly, the choice of bioinformatic analyses depends on the specific questions regarding the composition and role of viruses in stream biofilms.

CONCLUSIONS

In conclusion, we provide a first protocol for the generation of viromes from stream biofilms. The sample-to-sequence pipeline generates diverse viromes, however the purification scheme may select against viruses with different buoyant densities, such as ssDNA viruses or viruses containing lipids. Similarly, filtration may discriminate against large or filamentous viruses and some viruses may be sensitive to chemicals used during biofilm breakup. However, this is a first step towards a better understanding of the roles viruses may play in stream ecology. By providing a step-by-step protocol on protocols.io, we hope to further stimulate research on phage diversity in stream biofilms.

ACKNOWLEDGEMENTS

We acknowledge the help of Davide Demurtas for TEM imaging at EPFL (CIME).

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Swiss National Science Foundation (No. 169651) to Tom J. Battin and Hannes Peter. 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:
Swiss National Science Foundation: 169651.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Meriem Bekliz conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Jade Brandani performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Massimo Bourquin analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Tom J. Battin conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Hannes Peter conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The raw sequences are available at the European Nucleotide Archive: [ERP116349](https://www.ebi.ac.uk/ena/record/ERP116349).

Data Availability

The following information was supplied regarding data availability:

Peter, Hannes; Bekliz, Meriem (2019): Stream biofilm viromes. figshare. Dataset.
DOI [10.6084/m9.figshare.8943341.v2](https://doi.org/10.6084/m9.figshare.8943341.v2).

REFERENCES

- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. 2006. The marine viromes of four oceanic regions. *PLOS Biology* 4(11):e368 DOI [10.1371/journal.pbio.0040368](https://doi.org/10.1371/journal.pbio.0040368).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 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).
- Battin TJ, Besemer K, Bengtsson MM, Romani AM, Packmann AI. 2016. The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology* 14(4):251–263 DOI [10.1038/nrmicro.2016.15](https://doi.org/10.1038/nrmicro.2016.15).
- Battin TJ, Kaplan LA, Newbold JD, Cheng X, Hansen C. 2003. Effects of current velocity on the nascent architecture of stream microbial biofilms. *Applied Environmental Microbiology* 69(9):5443–5452 DOI [10.1128/AEM.69.9.5443-5452.2003](https://doi.org/10.1128/AEM.69.9.5443-5452.2003).
- Besemer K, Hödl I, Singer G, Battin TJ. 2009. Architectural differentiation reflects bacterial community structure in stream biofilms. *ISME Journal* 3(11):1318–1324 DOI [10.1038/ismej.2009.73](https://doi.org/10.1038/ismej.2009.73).
- Bibby K, Peccia J. 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental Science & Technology* 47(4):1945–1951 DOI [10.1021/es305181x](https://doi.org/10.1021/es305181x).
- 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. *ISME Journal* 11(1):7–14 DOI [10.1038/ismej.2016.89](https://doi.org/10.1038/ismej.2016.89).
- 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).
- Briandet R, Lacroix-Gueu P, Renault M, Lecart S, Meylheuc T, Bidnenko E, Steenkeste K, Bellon-Fontaine MN, Fontaine-Aupart MP. 2008. Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages inside biofilms. *Applied Environmental Microbiology* 74(7):2135–2143 DOI [10.1128/AEM.02304-07](https://doi.org/10.1128/AEM.02304-07).
- Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, De Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348(6237):1261498 DOI [10.1126/science.1261498](https://doi.org/10.1126/science.1261498).

- Brum JR, Sullivan MB. 2015.** Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* **13**(3):147–159 DOI [10.1038/nrmicro3404](https://doi.org/10.1038/nrmicro3404).
- Castro-Mejia JL, Muhammed MK, Kot W, Neve H, Franz CM, Hansen LH, Vogensen FK, Nielsen DS. 2015.** Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* **3**(1):64 DOI [10.1186/s40168-015-0131-4](https://doi.org/10.1186/s40168-015-0131-4).
- Chan BK, Abedon ST. 2015.** Bacteriophages and their enzymes in biofilm control. *Current Pharmaceutical Design* **21**(1):85–99 DOI [10.2174/1381612820666140905112311](https://doi.org/10.2174/1381612820666140905112311).
- Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, Hoyt DW, Meulia T, Wolfe RA, Hanson AJ, Mouser PJ, Moore JD, Wunch K, Sullivan MB, Wrighton KC, Wilkins MJ. 2019.** Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nature Microbiology* **4**(2):352–361 DOI [10.1038/s41564-018-0312-6](https://doi.org/10.1038/s41564-018-0312-6).
- Dann LM, Paterson JS, Newton K, Oliver R, Mitchell JG. 2016.** Distributions of virus-like particles and prokaryotes within microenvironments. *PLOS ONE* **11**(1):e0146984 DOI [10.1371/journal.pone.0146984](https://doi.org/10.1371/journal.pone.0146984).
- Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, Weinbauer M. 2008.** Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* **454**(7208):1084–1087 DOI [10.1038/nature07268](https://doi.org/10.1038/nature07268).
- Danovaro R, Dell'Anno A, Trucco A, Serresi M, Vanucci S. 2001.** Determination of virus abundance in marine sediments. *Applied and Environmental Microbiology* **67**(3):1384–1387 DOI [10.1128/AEM.67.3.1384-1387.2001](https://doi.org/10.1128/AEM.67.3.1384-1387.2001).
- Danovaro R, Middelboe M. 2010.** Separation of free virus particles from sediments in aquatic systems. In: Wilhelm SW, Weinbauer MG, Suttle CA, eds. *Manual of Aquatic Viral Ecology*. Waco: American Society of Limnology and Oceanography.
- Dell'Anno A, Corinaldesi C, Danovaro R. 2015.** Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proceedings of the National Academy of Sciences of the United States of America* **112**(16):E2014–E2019 DOI [10.1073/pnas.1422234112](https://doi.org/10.1073/pnas.1422234112).
- Doolittle MM, Cooney JJ, Caldwell DE. 1996.** Tracing the interaction of bacteriophage with bacterial biofilms using fluorescent and chromogenic probes. *Journal of Industrial Microbiology* **16**(6):331–341 DOI [10.1007/BF01570111](https://doi.org/10.1007/BF01570111).
- Doyle JJ, Doyle JL. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**:11–15.
- DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. 1956.** Colorimetric method for determination of sugars and related substances. *Analytical Chemistry* **28**(3):350–356 DOI [10.1021/ac60111a017](https://doi.org/10.1021/ac60111a017).
- Eddy SR. 2011.** Accelerated profile HMM searches. *PLOS Computational Biology* **7**(10):e1002195 DOI [10.1371/journal.pcbi.1002195](https://doi.org/10.1371/journal.pcbi.1002195).
- Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018.** Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* **3**(8):870–880 DOI [10.1038/s41564-018-0190-y](https://doi.org/10.1038/s41564-018-0190-y).
- Feldman HA, Wang SS. 1961.** Sensitivity of various viruses to chloroform. *Proceedings of the Society for Experimental Biology and Medicine* **106**(4):736–738 DOI [10.3181/00379727-106-26459](https://doi.org/10.3181/00379727-106-26459).
- Flemming HC, Wingender J. 2010.** The biofilm matrix. *Nature Reviews Microbiology* **8**(9):623–633 DOI [10.1038/nrmicro2415](https://doi.org/10.1038/nrmicro2415).

- Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T, Graf EH, Tardif KD, Kapusta A, Ryneerson S, Stockmann C, Queen K, Tong S, Voelkerding KV, Blaschke A, Byington CL, Jain S, Pavia A, Ampofo K, Eilbeck K, Marth G, Yandell M, Schlaberg R. 2016. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biology* 17(1):111 DOI 10.1186/s13059-016-0969-1.
- Gregory AC, Zayed AA, Conceicao-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Dominguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay JE, Vik D, Babin M, Bowler C, Culley AI, De Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177(5):1109–1123.e14 DOI 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 J, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, De Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532(7600):465–470 DOI 10.1038/nature16942.
- 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.
- Hewson I, Brown JM, Burge CA, Couch CS, LaBarre BA, Mouchka ME, Naito M, Harvell CD. 2012. Description of viral assemblages associated with the *Gorgonia ventalina* holobiont. *Coral Reefs* 31(2):487–491 DOI 10.1007/s00338-011-0864-x.
- Hurwitz BL, Deng L, Poulos BT, Sullivan MB. 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology* 15(5):1428–1440 DOI 10.1111/j.1462-2920.2012.02836.x.
- Jurtz VI, Villarreal J, Lund O, Larsen MV, Nielsen M. 2016. Metaphinder-identifying bacteriophage sequences in metagenomic data sets. *PLOS ONE* 11(9):e0163111 DOI 10.1371/journal.pone.0163111.
- Kauffman KM, Hussain FA, Yang J, Arevalo P, Brown JM, Chang WK, VanInsberghe D, Elsherbini J, Sharma RS, Cutler MB, Kelly L, Polz MF. 2018. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* 554(7690):118–122 DOI 10.1038/nature25474.
- Lawrence JE, Steward GF. 2010. Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA, eds. *Manual of Aquatic Viral Ecology*. Waco: ASLO, 166–181.
- 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.
- Lim YW, Haynes M, Furlan M, Robertson CE, Harris JK, Rohwer F. 2014. Purifying the impure: sequencing metagenomes and metatranscriptomes from complex animal-associated samples. *Journal of Visualized Experiments* 94:e52117 DOI 10.3791/52117.
- Lorenzen CJ. 1967. Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology and Oceanography* 12(2):343–346 DOI 10.4319/lo.1967.12.2.0343.
- Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. 1951. Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry* 193:265–275.

- Marhaver KL, Edwards RA, Rohwer F. 2008. Viral communities associated with healthy and bleaching corals. *Environmental Microbiology* **10**(9):2277–2286 DOI [10.1111/j.1462-2920.2008.01652.x](https://doi.org/10.1111/j.1462-2920.2008.01652.x).
- 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):425–430 DOI [10.1038/nature19094](https://doi.org/10.1038/nature19094).
- Pal C, Macia MD, Oliver A, Schachar I, Buckling A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**(7172):1079–1081 DOI [10.1038/nature06350](https://doi.org/10.1038/nature06350).
- Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, Fuhrman JA. 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nature Protocols* **2**(2):269–276 DOI [10.1038/nprot.2007.6](https://doi.org/10.1038/nprot.2007.6).
- Peduzzi P. 2016. Virus ecology of fluvial systems: a blank spot on the map? *Biological Reviews of the Cambridge Philosophical Society* **91**(4):937–949 DOI [10.1111/brv.12202](https://doi.org/10.1111/brv.12202).
- Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. 2018. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME Journal* **12**(4):1154–1162 DOI [10.1038/s41396-017-0042-4](https://doi.org/10.1038/s41396-017-0042-4).
- Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7**(11):828–836 DOI [10.1038/nrmicro2235](https://doi.org/10.1038/nrmicro2235).
- Rohwer F. 2003. Global phage diversity. *Cell* **113**(2):141 DOI [10.1016/S0092-8674\(03\)00276-9](https://doi.org/10.1016/S0092-8674(03)00276-9).
- Rosario K, Breitbart M. 2011. Exploring the viral world through metagenomics. *Current Opinion in Virology* **1**(4):289–297 DOI [10.1016/j.coviro.2011.06.004](https://doi.org/10.1016/j.coviro.2011.06.004).
- 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).
- Rossum TV, Uyaguari-Diaz MI, Vlok M, Peabody MA, Tian A, Cronin KI, Chan M, Croxen MA, Hsiao WWL, Isaac-Renton J, Tang PKC, Prystajek NA, Suttle CA, Brinkman FSL. 2018. Spatiotemporal dynamics of river viruses, bacteria and microeukaryotes. *bioRxiv* **29**:170 DOI [10.1101/259861](https://doi.org/10.1101/259861).
- Roux S, Emerson JB, Eloë-Fadrosh EA, 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).
- 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](https://doi.org/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, Sullivan MB, Rich VI, Malmstrom RR, Eloë-Fadrosh EA. 2019. Optimizing de novo genome assembly from PCR-amplified metagenomes. *PeerJ* **7**(5879):e6902 DOI [10.7717/peerj.6902](https://doi.org/10.7717/peerj.6902).
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.
- Scanlan PD, Buckling A. 2012. Co-evolution with lytic phage selects for the mucoid phenotype of *Pseudomonas fluorescens* SBW25. *ISME Journal* **6**(6):1148–1158 DOI [10.1038/ismej.2011.174](https://doi.org/10.1038/ismej.2011.174).
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7):676–682 DOI [10.1038/nmeth.2019](https://doi.org/10.1038/nmeth.2019).

- Silva BSO, Coutinho FH, Gregoracci GB, Leomil L, De Oliveira LS, Froes A, Tschoeke D, Soares AC, Cabral AS, Ward ND, Richey JE, Krusche AV, Yager PL, De Rezende CE, Thompson CC, Thompson FL. 2017. Virioplankton assemblage structure in the lower river and ocean continuum of the Amazon. *mSphere* 2(5):e00366–17 DOI 10.1128/mSphere.00366-17.
- Sime-Ngando T. 2014. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Frontiers in Microbiology* 5:355 DOI 10.3389/fmicb.2014.00355.
- Simmons M, Bond MC, Drescher K, Bucci V, Nadell CD. 2019. Biofilm structure promotes coexistence of phage-resistant and phage-susceptible bacteria. *bioRxiv* 42(2):385 DOI 10.1101/552265.
- Skvortsov T, De Leeuwe C, Quinn JP, McGrath JW, Allen CC, McElarney Y, Watson C, Arkhipova K, Lavigne R, Kulakov LA. 2016. Metagenomic characterisation of the viral community of lough Neagh, the largest freshwater lake in Ireland. *PLOS ONE* 11(2):e0150361 DOI 10.1371/journal.pone.0150361.
- Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE. 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Research in Microbiology* 159(5):349–357 DOI 10.1016/j.resmic.2008.04.010.
- Sullivan MB, Weitz JS, Wilhelm S. 2017. Viral ecology comes of age. *Environmental Microbiology Reports* 9(1):33–35 DOI 10.1111/1758-2229.12504.
- Sutherland IW, Hughes KA, Skillman LC, Tait K. 2004. The interaction of phage and biofilms. *FEMS Microbiology Letters* 232(1):1–6 DOI 10.1016/S0378-1097(04)00041-2.
- Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nature Reviews Microbiology* 5(10):801–812 DOI 10.1038/nrmicro1750.
- Temmam S, Monteil-Bouchard S, Robert C, Pascalis H, Michelle C, Jardot P, Charrel R, Raoult D, Desnues C. 2015. Host-associated metagenomics: a guide to generating infectious RNA viromes. *PLOS ONE* 10(10):e0139810 DOI 10.1371/journal.pone.0139810.
- Thijs S, Op De Beeck M, Beckers B, Truyens S, Stevens V, Van Hamme JD, Weyens N, Vangronsveld J. 2017. Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Frontiers in Microbiology* 8:494 DOI 10.3389/fmicb.2017.00494.
- Thurber VR, Barott KL, Hall D, Liu H, Rodriguez-Mueller B, Desnues C, Edwards RA, Haynes M, Angly FE, Wegley L, Rohwer FL. 2008. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites compressa*. *Proceedings of the National Academy of Sciences of the United States of America* 105(47):18413–18418 DOI 10.1073/pnas.0808985105.
- 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.
- Trubl G, Roux S, Solonenko N, Li Y-F, Bolduc B, Rodríguez-Ramos J, Eloe-Fadrosch EA, Rich VI, Sullivan MB. 2019. Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *PeerJ* 7:e7265 DOI 10.7717/peerj.7265.
- 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(12):e1999 DOI 10.7717/peerj.1999.
- Velji MI, Albright LJ. 1993. Improved sample preparation for enumeration of aggregated aquatic substrate bacteria. In: Kemp PF, Sherr BF, Sherr EB, Cole JJ, eds. *Handbook of Methods in Aquatic Microbial Ecology*. Boca Raton: Lewis Publishers, 139–142.
- Vidakovic L, Singh PK, Hartmann R, Nadell CD, Drescher K. 2017. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiology* 3(1):26–31 DOI 10.1038/s41564-017-0050-1.

- Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderegger DL, Suttle CA, Taylor BP, Frede Thingstad T, Wilson WH, Eric Wommack K. 2015.** A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME Journal* **9(6)**:1352–1364 DOI [10.1038/ismej.2014.220](https://doi.org/10.1038/ismej.2014.220).
- 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(1)**:201–219 DOI [10.1146/annurev-virology-101416-041639](https://doi.org/10.1146/annurev-virology-101416-041639).