Benchmarking protocols for the metagenomic analysis of stream biofilm viromes

Meriem Bekliz[1], Jade Brandani[1], Massimo Bourquin[1], Tom J. Battin[1], Hannes Peter[1*]

[1] Stream Biofilm and Ecosystem Research Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Corresponding Author:
Hannes Peter
Stream Biofilm and Ecosystem Research Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
Email address: Hannes.peter@epfl.ch

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.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

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 *et al.*, 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 ~~prokaryotic~~ viruses that infect bacteria, also known as bacteriophages, are a~~the~~ main source of ~~prokaryotic~~ bacterial mortality (Suttle, 2007, Brum and Sullivan, 2015, Gregory *et al.*, 2019) and impact ecosystem functions such as the cycling of carbon (~~Guidi *et al.*, 2016~~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 *et al.*, 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 and Breitbart, 2011, Brum and Sullivan, 2015, Roux *et al.*, 2016, Roux *et al.*, 2017). In this context, one may differentiate between untargeted approaches, which haves~~o~~

~~far~~ produced a~~the~~ wealth of viral sequences (Paez-Espino *et al.*, 2016) and virome~~metagenomic~~ sequencing~~ of viromes~~, in which the viral fraction is purified~~which rely on the purification of the viral fraction~~ prior to sequencing, thus ensuring that the sequencing effort is targeted towards viral nucleic acids (Rosario and Breitbart, 2011).

Besides the early recognition of the potential of phages to eradicate bacterial biofilms in medical settings (e.g. Chan and 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 and 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 *et al.*, 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~~ome~~ 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 metage~~mon~~nomics therefore aim at concentrating and purifying viral-like particles (VLPs) and removing~~al of~~ contaminanting DNA, while ~~and at the same time optimizing~~ VLP recovery (e.g. 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. ~~Based on the sturcture of their capsid, s~~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.

The first critical step towards the study of biofilm viruses in streams using virome ~~metagenomic~~ sequencing requires the effective extraction of viral-like particles (VLPs) from the biofilm matrix. The aim of our study was therefore to optimize a sample-to-sequence pipeline ~~laboratory methods for~~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 (~~Vega Thurber *et al.*, 2009,~~e.g. Danovaro *et al.*, 2001, Danovaro and Middelboe, 2010, Hurwitz *et al.*, 2013, Temmam *et al.*, 2015, Trubl *et al.*, 2016, see Thurber *et al.*, 2009 and Hayes *et al.*, 2017 for reviews) ~~Danovaro and Middelboe, 2010, Hurwitz *et al.*, 2013, Trubl *et al.*, 2016)~~, but their rigorous

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

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 & 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 to 15 cm in diameter) using sterile brushes and Milli-Q water. Depending on biofilm thickness, we scrapped biofilms from cobbles with a total surface area ranging from 1.3 to 2.4 m^2 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 (Bransonic Sonifier 450, Branson) on ice (1 min)(Velji and Albright 1993). As previously established (e.g., Besemer *et al.*, 2009), Cells were stained using Syto13 and counted on a flow cytometer

(NovoCyte, ACEA Biosciences)~~; this protocol worked fine for us previously (e.g., Besemer *et al.*, 2009)~~. 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 (Lowry *et al.*, 1951).

Transmission electron microscopy (TEM)

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, 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, USA) operated at 200 kV and equipped with a 4098 x 4098 pixel camera (CETA, ThermoFisher). Magnification ranged from 10~~'~~000 to 29~~'~~000 ×~~x~~, 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, ~~W~~we explored a variety of protocols for the concentration, extraction and purification of ~~stream biofilm~~ 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 ~~at to~~the liberat~~ion of~~e VLP from the biofilm matrix, while purification aims ~~at~~ to reduc~~ing~~ 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 ~~larger~~ organisms contained within the biofilm slurry. We recovered the supernatant for downstream analyses.

Concentration

We evaluated tangential flow filtration (TFF) and polyethylene glycol (PEG) precipitation to concentrate the viral size fraction. A medium-scale TFF system equipped with a 100 kDa tangential flow filter (GE Healthcare, USA) was used (~~Vega~~ 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 and Peccia, 2013). Aliquot biofilm samples were supplemented with 10% w/v PEG 8000 (Sigma-Aldrich, Germany) and 2.5 M NaCl (Sigma-Aldrich). The supernatant was then agitated by inverting the tubes three times and stored overnight (4°C), followed by centrifugation for 30 minutes at 9432 × g at 4°C (Sorvall RC-5C centrifuge, HS-4 rotor, ThermoFisher). Supernatants were carefully removed and the resulting pellets were eluted with 50 mL sterile H_2O.

Extraction

We tested three different physicochemical treatments (chloroform, tetrasodium pyrophosphate in combination with sonication and dithiothreitol (DTT)) to dislodge viruses from the biofilm matrix. ~~The order – first volume reduction, then chemical detachment – was chosen because the chemicals used for extraction of VLPs may damage TFF filters.~~

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 minutes, vortexed every 5 minutes, and centrifuged at 3234 × g for 15 minutes at 4° C (5810R, Eppendorf, Germany) to recover the supernatant (Marhaver *et al.*, 2008, ~~Vega~~ Thurber *et al.*, 2008).

Following the recommendation of Danovaro and Middelboe (2010), we used sonication in combination with tetrasodium phosphate to separate viruses from biofilms. For this, we added 5 mM of tetrasodium pyrophosphate (final concentration) to another subset of concentrated biofilm samples and incubated them in the dark (15 minutes). Then, all samples were sonicated (frequency: 40 KHz; Bransonic Sonifier 450, Branson) three times for 1 minute with 30-second intervals during which the samples were manually shaken. To prevent heating, samples were put on ice during sonication. Finally, one subset of the samples were treated with 6.5 mM DTT and incubated for 1 hour 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~~ 3234 × g for 15 minutes (5810R, Eppendorf) and the supernatant was sequentially filtered through 0.8 µm and 0.45 µm filters (Whatman, GE Healthcare, USA) to remove debris and cells.

Purification

To eliminate contaminating eukaryotic, prokaryotic and extracellular nucleic acids, DNase I at a final concentration of 2.5U/µL (Life Technologies, Germany) and Tris buffer (100 mM Tris pH 7.5, 5 mM $CaCl_2$ and 25 mM $MgCl_2$) 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) according to the manufacturer's instructions with an initial enzyme activation step (95°C for 2 minutes) followed by 30 cycles of denaturation (95°C for 30 seconds) and hybridization (72°C for 30 seconds) and a final elongation (72°C for 10 minutes). PCR products were visualized under UV light after migration on an agarose gel stained using GelRed (Biotium, USA). However, we did not assess the potential of PCR inhibition.

For final purification of the viral ~~DNA~~ fraction, two protocols of density gradient ultracentrifugation, sucrose and CsCl density gradients, were assessed. The sucrose density gradient was achieved by 3 mL of 0.2-µm filtered 66% sucrose and 7 mL of 0.2-µm filtered 30% sucrose (Temmam *et al.*, 2015). A subset of sample was deposited on top of the~~a~~ sucrose density gradient and centrifuged at 106800 × g (Optima XPN-80 Ultracentrifuge, 32 SWi, Beckman Coulter, 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 1 mL of 1.7, 1.5, 1.35 and 1.2 g/mL CsCl~~,~~ ~~and centrifuged at 106’800 g (Optima XPN-80 Ultracentrifuge, 32 SWi, Beckman Coulter, USA) for 2 h (4°C).~~ 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. ~~Similarly, sucrose~~

~~gradient ultracentrifugation was assessed for final purification of the viral fraction. The sucrose gradient consisted of 3 mL of 0.2 µm filtered 66% sucrose and 7 mL of 0.2 µm filtered 30% sucrose. The viral fraction was harvested from the interface between both layers 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~~Noteworthy,~~ pyrophosphate concentrations >10 mM tend to interfere with VLP counting (Danovaro *et al.*, 2001); however, we used 5 mM pyrophosphate during VLP liberation from biofilms. Subsamples were fixed with formaldehyde, stained with 0.5 µL of 1000 × SYBR ~~gold~~ Gold (Molecular Probes, ThermoScientific, USA) and incubated at room temperature in the dark (30 minutes). After incubation, each sample was filtered onto a 0.02 µm pore size membrane filter (Anodisc, Whatman). The filters were mounted on glass slides with a drop of VECTASHIELD antifade mounting medium (Vector Laboratories, Burlingame, USA). VLP were visualized using blue light (488 nm) excitation and green (512 nm) emission. For each sample, 15 to 20 randomly selected images were acquired with a camera (Axiocam 506 mono, Zeiss) mounted onto the microscope. VLPs were discriminated from bacteria by size (0.015 to 0.2 µm) and enumerated using a custom script in Fiji (Schindelin *et al.*, 2012).

Nucleic acid extraction

Formatted: Font: Italic

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 et al. (1989), ~~purified viral samples were suspended in~~ 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 minutes). Then, two volumes of cold 100% EtOH were added and incubated 30 minutes (4°C). Samples were centrifuged at ~~17'000~~ 17000 × g (4°C, 20 minutes) and pellets washed twice with 70% cold EtOH. Pellets were air-dried and resuspended in 567 µL of TE buffer (10 mM Tris and 1 mM EDTA (pH 8.0)). Thirty µL of warm 10% SDS and 3 µL of proteinase K (20 mg/mL) were added and samples were incubated for 1 h at 37°C. Next, 100 µL of 5M NaCl and 80 µL of warm hexadecyltrimethylammonium bromide (CTAB) were added and incubated for 10 minutes (65°C) (Doyle and 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 minutes) 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 H_2O 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, 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, 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, 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 x 300 bp) was performed on a MiSeq System (Illumina, San Diego, 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/) and then counted the number of reads mapping each contig using samtools (Li et al. 2009). ~~A~~In order to specifically target ssDNA contigs (Trubl *et al.*, 2019), ~~dditionally,~~ we ~~queried~~ 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 ~~to identify ssDNA contigs (Trubl *et al.*, 2019).~~ ~~(Li et al. 2009)~~

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 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 (TEM)

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 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 in DNA yields, which reached 1.22 ng μL^{-1} in VDN, 1.31 ng μL^{-1} in VEV and 18.7 ng μL^{-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 24‘388‘096, 22’459’218 and 23’804’190 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 to 0.25% of the reads, while bacterial contaminant reads accounted for 1.47 to 8.11% of the reads.

We obtained 3698, ~~11'323~~ 11323 and ~~13'591~~ 13591 contigs from de-novo assembly of quality-controlled and deduplicated reads from SNG, VEV and VDN, respectively. The largest contigs were 9493, ~~46'665~~ 46665 and ~~54'492~~ 54492 bp in SNG, VEV and VDN, respectively. Hmmsearch 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*, 153.8 mapped to *Podoviridae* and 115.1 reads mapped to *Myoviridae*, whereas only 29.4 reads mapped to *Siphoviridae*. 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, respectively) than in VDN (2.6 and 2.8) or VEV (36.7 and 2.8).

Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Font: Italic

Discussion

The advent of metagenomic tools has revolutionized the study of the role of viruses in numerous ecosystems (Suttle, 2007, Rosario and Breitbart, 2011, Brum and 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 (Fig. 3, Fig. 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 and 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 m^2 (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 *et al.*, 2008, ~~Vega~~ Thurber *et al.*, 2008, Hewson *et al.*, 2012); however, chloroform may denature the lipid envelopes surrounding viral capsids, internal lipid membranes and nucleo-cytoplasmic large DNA viruses (NCLDV), including *Phycodnaviridae*, which predominantly infect freshwater and marine algae, may be sensitive to chloroform treatment (Feldman and Wang, 1961). ~~Moreover, ssDNA viruses lack a lipid envelope and some tailed bacteriophages may display sensitivity to~~

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

~~chloroform.~~ Similarly, DTT is a reducing agent, which breaks disulfide bonds in proteins. This may explain to the reduced recovery of VLPs from samples treated with PEG, chloroform or DTT.

Extracellular DNA is a common component of the biofilm matrix (Flemming and 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 have used nuclease cocktails including DNases, RNases and Benzoases (Temmam *et al.*, 2015, Rosario et al., 2018,) to eliminate various contaminant nucleic acids. Th~~is~~e treatment order~~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 (~~Vega~~ Thurber *et al.*, 2009, Kauffman *et al.*, 2018), or degrade the structure of enveloped viruses (Lawrence and Steward, 2010) and therefore reduce their recovery. ~~A c~~Critical steps of virome preparation concern~~s~~ ~~the~~ 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 *et al.*, 1989). Optimization of DNA extraction protocols, such as done for soil viromes (Trubl *et al.*, 2018), 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.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Finally, there is a plethora of tools available for the bioinformatic analysis of viromes (e.g. those implemented in iVirus (Bolduc *et al.*, 2017). 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.

Formatted: Font: Italic

Formatted: Font: Italic

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, ~~T~~this is a first step towards ~~paves the way for~~ 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). ~~at the EPFL (CIME).~~

References

~~Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19:** 455-477.~~

~~Battin TJ, Kaplan LA, Newbold JD, Cheng X, Hansen C (2003). Effects of current velocity on the nascent architecture of stream microbial biofilms. *Appl Environ Microbiol* **69:** 5443-5452.~~

~~Battin TJ, Besemer K, Bengtsson MM, Romani AM, Packmann AI (2016). The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* **14:** 251-263.~~

Formatted: Left, Line spacing: single

Besemer K, Singer G, Hödl I, Battin TJ (2009). Bacterial Community Composition of Stream Biofilms in Spatially Variable-Flow Environments. *Applied and Environmental Microbiology* **75:** 7189.

Bibby K, Peccia J (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* **47:** 1945-1951.

Briandet R, Lacroix-Gueu P, Renault M, Lecart S, Meylheuc T, Bidnenko E *et al.* (2008). Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages inside biofilms. *Appl Environ Microbiol* **74:** 2135-2143.

Brum JR, Sullivan MB (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* **13:** 147.

Chan BK, Abedon ST (2015). Bacteriophages and their enzymes in biofilm control. *Curr Pharm Des* **21:** 85-99.Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE *et al.* (2019). Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat Microbiol* **4:** 352-361.

Dann LM, Paterson JS, Newton K, Oliver R, Mitchell JG (2016). Distributions of Virus-Like Particles and Prokaryotes within Microenvironments. *PLoS One* **11:** e0146984.

Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C *et al.* (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* **454:** 1084.

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*. . American Society of Limnology and Oceanography: Waco, TX.

Dell'Anno A, Corinaldesi C, Danovaro R (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proc Natl Acad Sci U S A* **112:** E2014-2019.

Doolittle MM, Cooney JJ, Caldwell DE (1996). Tracing the interaction of bacteriophage with bacterial biofilms using fluorescent and chromogenic probes. *J Ind Microbiol* **16:** 331-341.

DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956). Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry* **28:** 350-356.

Eddy SR (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7:** e1002195.

Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB *et al.* (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* **3:** 870-880.

Feldman HA, Wang SS (1961). Sensitivity of Various Viruses to Chloroform. *Proceedings of the Society for Experimental Biology and Medicine* **106:** 736-738.

Flemming HC, Wingender J (2010). The biofilm matrix. *Nat Rev Microbiol* **8:** 623-633.

Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T *et al.* (2016). Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol* **17:** 111.

Gregory AC, Zayed AA, Conceicao-Neto N, Temperton B, Bolduc B, Alberti A *et al.* (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177:** 1109-1123 e1114.

Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S *et al.* (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532:** 465-470.

Hewson I, Brown JM, Burge CA, Couch CS, LaBarre BA, Mouchka ME *et al.* (2012). Description of viral assemblages associated with the Gorgonia ventalina holobiont. *Coral Reefs* **31:** 487-491.

Hurwitz BL, Deng L, Poulos BT, Sullivan MB (2013). Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* **15:** 1428-1440.

Jurtz VI, Villarroel J, Lund O, Voldby Larsen M, Nielsen M (2016). MetaPhinder-Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PLoS One* **11:** e0163111.

Lawrence JE, Steward GF (2010). Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. ASLO. pp 166-181.

Lorenzen CJ (1967). Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology and Oceanography* **12:** 343-346.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951). Protein measurement with the Folin phenol reagent. *The 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:** 2277-2286.

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N *et al.* (2016). Uncovering Earth's virome. *Nature* **536:** 425-430.

Pal C, Macia MD, Oliver A, Schachar I, Buckling A (2007). Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450:** 1079-1081.

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. *Nat Protoc* **2:** 269-276.

Peduzzi P (2016). Virus ecology of fluvial systems: a blank spot on the map? *Biol Rev Camb Philos Soc* **91:** 937-949.

Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA (2018). Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* **12:** 1154-1162.

Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F *et al.* (2009). Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7:** 828-836.

Rohwer F (2003). Global Phage Diversity. *Cell* **113:** 141.

Rosario K, Breitbart M (2011). Exploring the viral world through metagenomics. *Curr Opin Virol* **1:** 289-297.

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. *Environ Sci Technol* **52:** 1014-1027.

Rossum TV, Uyaguari-Diaz MI, Vlok M, Peabody MA, Tian A, Cronin KI *et al.* (2018). Spatiotemporal dynamics of river viruses, bacteria and microeukaryotes. *bioRxiv*: 259861.

Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB *et al.* (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4:** e2777.

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.

Roux S, Trubl G, Goudeau D, Nath N, Couradeau E, Ahlgren NA *et al.* (2019). Optimizing *de novo* genome assembly from PCR-amplified metagenomes. *PeerJ* **7:** e6902.

Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press: New York.

Formatted: Font: Italic

Scanlan PD, Buckling A (2012). Co-evolution with lytic phage selects for the mucoid phenotype of Pseudomonas fluorescens SBW25. *ISME J* **6:** 1148-1158.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9:** 676-682.
Silva BSO, Coutinho FH, Gregoracci GB, Leomil L, de Oliveira LS, Froes A *et al.* (2017). Virioplankton Assemblage Structure in the Lower River and Ocean Continuum of the Amazon. *mSphere* **2**.
Sime-Ngando T (2014). Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Frontiers in microbiology* **5:** 355-355.
Simmons M, Bond MC, Drescher K, Bucci V, Nadell CD (2019). Evolutionary dynamics of phage resistance in bacterial biofilms. *bioRxiv*: 552265.
Skvortsov T, de Leeuwe C, Quinn JP, McGrath JW, Allen CC, McElarney Y *et al.* (2016). Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest Freshwater Lake in Ireland. *PLoS One* **11:** e0150361.
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:** 349-357.
Sullivan MB, Weitz JS, Wilhelm S (2017). Viral ecology comes of age. *Environmental Microbiology Reports* **9:** 33-35.
Sutherland IW, Hughes KA, Skillman LC, Tait K (2004). The interaction of phage and biofilms. *FEMS Microbiol Lett* **232:** 1-6.
Suttle CA (2007). Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5:** 801-812.
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.
Trubl G, Roux S, Solonenko N, Li Y-F, Bolduc B, Rodríguez-Ramos J *et al.* (2019). Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *PeerJ Preprints* **7:** e27640v27641.
Vega Thurber R, Haynes M, Breitbart M, Wegley L, Rohwer F (2009). Laboratory procedures to generate viral metagenomes. *Nat Protoc* **4:** 470-483.
Vega Thurber RL, Barott KL, Hall D, Liu H, Rodriguez-Mueller B, Desnues C *et al.* (2008). Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral Porites compressa. *Proc Natl Acad Sci U S A* **105:** 18413-18418.
Vidakovic L, Singh PK, Hartmann R, Nadell CD, Drescher K (2017). Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiology*.
Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A *et al.* (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J* **9:** 1352-1364.
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.
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, and Rohwer F. 2006. The marine viromes of four oceanic regions. *PLoS Biol* 4:e368. 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, and Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455-477. 10.1089/cmb.2012.0021

Battin TJ, Besemer K, Bengtsson MM, Romani AM, and Packmann AI. 2016. The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* 14:251-263. 10.1038/nrmicro.2016.15

Battin TJ, Kaplan LA, Newbold JD, Cheng X, and Hansen C. 2003. Effects of current velocity on the nascent architecture of stream microbial biofilms. *Appl Environ Microbiol* 69:5443-5452.

Besemer K, Hödl I, Singer G, and Battin TJ. 2009. Architectural differentiation reflects bacterial community structure in stream biofilms. *ISME J* 3:1318. 10.1038/ismej.2009.73

Bibby K, and Peccia J. 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* 47:1945-1951. 10.1021/es305181x

Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, and Sullivan MB. 2017. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J* 11:7-14. 10.1038/ismej.2016.89

Breitbart M, Bonnain C, Malki K, and Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3:754-766. 10.1038/s41564-018-0166-y

Briandet R, Lacroix-Gueu P, Renault M, Lecart S, Meylheuc T, Bidnenko E, Steenkeste K, Bellon-Fontaine MN, and Fontaine-Aupart MP. 2008. Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages inside biofilms. *Appl Environ Microbiol* 74:2135-2143. 10.1128/aem.02304-07

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier 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, and Sullivan MB. 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. 10.1126/science.1261498

Brum JR, and Sullivan MB. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* 13:147. 10.1038/nrmicro3404

Castro-Mejia JL, Muhammed MK, Kot W, Neve H, Franz CM, Hansen LH, Vogensen FK, and Nielsen DS. 2015. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* 3:64. 10.1186/s40168-015-0131-4

Chan BK, and Abedon ST. 2015. Bacteriophages and their enzymes in biofilm control. *Curr Pharm Des* 21:85-99.

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, and Wilkins MJ. 2019. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat Microbiol* 4:352-361. 10.1038/s41564-018-0312-6

Dann LM, Paterson JS, Newton K, Oliver R, and Mitchell JG. 2016. Distributions of Virus-Like Particles and Prokaryotes within Microenvironments. *PLOS ONE* 11:e0146984. 10.1371/journal.pone.0146984

Danovaro R, Dell'anno A, Trucco A, Serresi M, and Vanucci S. 2001. Determination of virus abundance in marine sediments. *Applied and Environmental Microbiology* 67:1384-1387.

Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, and Weinbauer M. 2008. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* 454:1084. 10.1038/nature07268

Danovaro R, and Middelboe M. 2010. Separation of free virus particles from sediments in aquatic systems. In: Wilhelm SW, Weinbauer MG, and Suttle CA, eds. *Manual of Aquatic Viral Ecology* Waco, TX: American Society of Limnology and Oceanography.

Dell'Anno A, Corinaldesi C, and Danovaro R. 2015. Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proc Natl Acad Sci U S A* 112:E2014-2019. 10.1073/pnas.1422234112

Doolittle MM, Cooney JJ, and Caldwell DE. 1996. Tracing the interaction of bacteriophage with bacterial biofilms using fluorescent and chromogenic probes. *J Ind Microbiol* 16:331-341.

Doyle JJ, and 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, and Smith F. 1956. Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry* 28:350-356. 10.1021/ac60111a017

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* 7:e1002195. 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, and Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* 3:870-880. 10.1038/s41564-018-0190-y

Feldman HA, and Wang SS. 1961. Sensitivity of Various Viruses to Chloroform. *Proceedings of the Society for Experimental Biology and Medicine* 106:736-738. 10.3181/00379727-106-26459

Flemming HC, and Wingender J. 2010. The biofilm matrix. *Nat Rev Microbiol* 8:623-633. 10.1038/nrmicro2415

Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T, Graf EH, Tardif KD, Kapusta A, Rynearson 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, and Schlaberg R. 2016. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol* 17:111. 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, and Sullivan MB. 2019. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* 177:1109-1123 e1114. 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, and Gorsky G. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465-470. 10.1038/nature16942

Hayes S, Mahony J, Nauta A, and van Sinderen D. 2017. Metagenomic Approaches to Assess Bacteriophages in Various Environmental Niches. *Viruses* 9. 10.3390/v9060127

Hewson I, Brown JM, Burge CA, Couch CS, LaBarre BA, Mouchka ME, Naito M, and Harvell CD. 2012. Description of viral assemblages associated with the Gorgonia ventalina holobiont. *Coral Reefs* 31:487-491. 10.1007/s00338-011-0864-x

Hurwitz BL, Deng L, Poulos BT, and Sullivan MB. 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* 15:1428-1440. 10.1111/j.1462-2920.2012.02836.x

Jurtz VI, Villarroel J, Lund O, Voldby Larsen M, and Nielsen M. 2016. MetaPhinder-Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PLOS ONE* 11:e0163111. 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, and Polz MF. 2018. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* 554:118. 10.1038/nature25474

Lawrence JE, and Steward GF. 2010. Purification of viruses by centrifugation. In: Wilhelm SW, Weinbauer MG, and Suttle CA, eds. *Manual of Aquatic Viral Ecology*: ASLO, 166–181.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079. 10.1093/bioinformatics/btp352

Lim YW, Haynes M, Furlan M, Robertson CE, Harris JK, and Rohwer F. 2014. Purifying the Impure: Sequencing Metagenomes and Metatranscriptomes from Complex Animal-associated Samples. *JoVE*:e52117. doi:10.3791/52117

Lorenzen CJ. 1967. Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology and Oceanography* 12:343-346. 10.4319/lo.1967.12.2.0343

Lowry OH, Rosebrough NJ, Farr AL, and Randall RJ. 1951. Protein measurement with the Folin phenol reagent. *J Biol Chem* 193:265-275.

Marhaver KL, Edwards RA, and Rohwer F. 2008. Viral communities associated with healthy and bleaching corals. *Environmental microbiology* 10:2277-2286. 10.1111/j.1462-2920.2008.01652.x

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, and Kyrpides NC. 2016. Uncovering Earth's virome. *Nature* 536:425-430. 10.1038/nature19094

Pal C, Macia MD, Oliver A, Schachar I, and Buckling A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450:1079-1081. 10.1038/nature06350

Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, and Fuhrman JA. 2007. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat Protoc* 2:269-276. 10.1038/nprot.2007.6

Peduzzi P. 2016. Virus ecology of fluvial systems: a blank spot on the map? *Biol Rev Camb Philos Soc* 91:937-949. 10.1111/brv.12202

Reche I, D'Orta G, Mladenov N, Winget DM, and Suttle CA. 2018. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* 12:1154-1162. 10.1038/s41396-017-0042-4

Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, and Mira A. 2009. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* 7:828-836.

Rohwer F. 2003. Global Phage Diversity. *Cell* 113:141. https://doi.org/10.1016/S0092-8674(03)00276-9

Rosario K, and Breitbart M. 2011. Exploring the viral world through metagenomics. *Curr Opin Virol* 1:289-297. 10.1016/j.coviro.2011.06.004

Rosario K, Fierer N, Miller S, Luongo J, and Breitbart M. 2018. Diversity of DNA and RNA Viruses in Indoor Air As Assessed via Metagenomic Sequencing. *Environ Sci Technol* 52:1014-1027. 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, Prystajecky NA, Suttle CA, and Brinkman FSL. 2018. Spatiotemporal dynamics of river viruses, bacteria and microeukaryotes. *bioRxiv*:259861. 10.1101/259861

Roux S, Emerson JB, Eloe-Fadrosh EA, and Sullivan MB. 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. 10.7717/peerj.3817

Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, Coleman ML, Breitbart M, and Sullivan MB. 2016. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* 4:e2777. 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, and Eloe-Fadrosh EA. 2019. Optimizing de novo genome assembly from PCR-amplified metagenomes. *PeerJ* 7:e6902. 10.7717/peerj.6902

Sambrook J, Fritsch EF, and Maniatis T. 1989. *Molecular Cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

Scanlan PD, and Buckling A. 2012. Co-evolution with lytic phage selects for the mucoid phenotype of Pseudomonas fluorescens SBW25. *ISME J* 6:1148-1158.

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, and Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676-682. 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, and Thompson FL. 2017. Virioplankton Assemblage Structure in the Lower River and Ocean Continuum of the Amazon. *mSphere* 2. 10.1128/mSphere.00366-17

Sime-Ngando T. 2014. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. *Frontiers in microbiology* 5:355-355. 10.3389/fmicb.2014.00355

Simmons M, Bond MC, Drescher K, Bucci V, and Nadell CD. 2019. Evolutionary dynamics of phage resistance in bacterial biofilms. *bioRxiv*:552265. 10.1101/552265

Skvortsov T, de Leeuwe C, Quinn JP, McGrath JW, Allen CC, McElarney Y, Watson C, Arkhipova K, Lavigne R, and Kulakov LA. 2016. Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest Freshwater Lake in Ireland. *PLOS ONE* 11:e0150361. 10.1371/journal.pone.0150361

Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, and Wommack KE. 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Research in Microbiology* 159:349-357. https://doi.org/10.1016/j.resmic.2008.04.010

Sullivan MB, Weitz JS, and Wilhelm S. 2017. Viral ecology comes of age. *Environmental Microbiology Reports* 9:33-35. 10.1111/1758-2229.12504

Sutherland IW, Hughes KA, Skillman LC, and Tait K. 2004. The interaction of phage and biofilms. *FEMS Microbiol Lett* 232:1-6.

Suttle CA. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 5:801-812. 10.1038/nrmicro1750

Temmam S, Monteil-Bouchard S, Robert C, Pascalis H, Michelle C, Jardot P, Charrel R, Raoult D, and Desnues C. 2015. Host-Associated Metagenomics: A Guide to Generating Infectious RNA Viromes. *PLOS ONE* 10:e0139810. 10.1371/journal.pone.0139810

Thijs S, Op De Beeck M, Beckers B, Truyens S, Stevens V, Van Hamme JD, Weyens N, and Vangronsveld J. 2017. Comparative Evaluation of Four Bacteria-Specific Primer Pairs for 16S rRNA Gene Surveys. *Frontiers in microbiology* 8:494-494. 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, and Rohwer FL. 2008. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral Porites compressa. *Proc Natl Acad Sci U S A* 105:18413-18418. 10.1073/pnas.0808985105

Thurber VR, Haynes M, Breitbart M, Wegley L, and Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4:470-483. 10.1038/nprot.2009.10

Trubl G, Roux S, Solonenko N, Li Y-F, Bolduc B, Rodríguez-Ramos J, Eloe-Fadrosh EA, Rich VI, and Sullivan MB. 2019. Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *PeerJ Preprints* 7:e27640v27641. 10.7287/peerj.preprints.27640v1

Trubl G, Solonenko N, Chittick L, Solonenko SA, Rich VI, and Sullivan MB. 2016. Optimization of viral resuspension methods for carbon-rich soils along a permafrost thaw gradient. *PeerJ* 4:e1999. 10.7717/peerj.1999

Velji MI, and Albright LJ. 1993. Improved sample preparation for enumeration of aggregated aquatic substrate bacteria. In: Kemp PF, Sherr BF, Sherr EB, and Cole JJ, eds. *Handbook of Methods in Aquatic Microbial Ecology*. Boca Raton, Florida: Lewis Publishers, 139-142.

Vidakovic L, Singh PK, Hartmann R, Nadell CD, and Drescher K. 2017. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiology*.

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, and Eric Wommack K. 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J* 9:1352-1364. 10.1038/ismej.2014.220

Williamson KE, Fuhrmann JJ, Wommack KE, and Radosevich M. 2017. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annual Review of Virology* 4:201-219. 10.1146/annurev-virology-101416-041639

Formatted: Indent: Left: 0", Hanging: 0.5", Line spacing: single