## A peer-reviewed version of this preprint was published in PeerJ on 20 December 2019.

<u>View the peer-reviewed version</u> (peerj.com/articles/8187), which is the preferred citable publication unless you specifically need to cite this preprint.

Bekliz M, Brandani J, Bourquin M, Battin TJ, Peter H. 2019. Benchmarking protocols for the metagenomic analysis of stream biofilm viromes. PeerJ 7:e8187 <u>https://doi.org/10.7717/peerj.8187</u>

# Benchmarking protocols for the metagenomic analysis of stream biofilm viromes

Meriem Bekliz<sup>1</sup>, Jade Brandani<sup>1</sup>, Massimo Bourquin<sup>1</sup>, Tom Battin<sup>1</sup>, Hannes Peter<sup>Corresp. 1</sup>

<sup>1</sup> SBER, École Polytechnique Federale de Lausanne, Lausanne, Schweiz

Corresponding Author: Hannes Peter Email address: hannes.peter@epfl.ch

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.

### 1 Benchmarking protocols for the metagenomic analysis of

#### 2 stream biofilm viromes

3 Meriem Bekliz<sup>1</sup>, Jade Brandani<sup>1</sup>, Massimo Bourquin<sup>1</sup>, Tom Battin<sup>1</sup>, Hannes Peter<sup>1\*</sup>

#### 4

- 5 1 Stream Biofilm and Ecosystem Research Laboratory, Ecole Polytechnique Fédérale de
- 6 Lausanne, CH-1015 Lausanne, Switzerland

#### 7

- 8 Corresponding Author:
- 9 Hannes Peter
- 10 Stream Biofilm and Ecosystem Research Laboratory, Ecole Polytechnique Fédérale de
- 11 Lausanne, CH-1015 Lausanne, Switzerland
- 12 Email address: Hannes.peter@epfl.ch
- 13
- 14 Abstract

15 Viruses drive microbial diversity, function and evolution and influence important 16 biogeochemical cycles in aquatic ecosystems. Despite their relevance, we currently lack an 17 understanding of their potential impacts on stream biofilm structure and function. This is 18 surprising given the critical role of biofilms for stream ecosystem processes. Currently, the study 19 of viruses in stream biofilms is hindered by the lack of an optimized protocol for their extraction, 20 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.

31

#### 32 Introduction

33 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, 34 35 2014). Viruses are a large reservoir of genetic diversity (Suttle, 2007, Sullivan et al., 2017) and 36 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 37 38 (Danovaro *et al.*, 2008). The study of viral ecology was pioneered in surface marine systems 39 where prokaryotic viruses, also known as bacteriophages, are the main source of prokaryotic mortality (Suttle, 2007, Brum and Sullivan, 2015, Gregory et al., 2019) and impact ecosystem 40 41 functions such as the cycling of carbon (Guidi et al., 2016). By lysing their hosts, horizontal gene transfer and metabolic reprogramming, bacteriophages play a pivotal role in structuring 42 43 microbial communities (Skvortsov et al., 2016, Silva et al., 2017, Rossum et al., 2018, Daly et 44 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 45 46 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 47 metagenomic sequencing (Rosario and Breitbart, 2011, Brum and Sullivan, 2015, Roux et al., 48 49 2016, Roux et al., 2017).

50 Besides the early recognition of the potential of phages to eradicate bacterial biofilms in medical 51 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 52 to the surface of bacterial cells (Vidakovic et al., 2017), the susceptibility to phage-induced 53 clearance of biofilm has been demonstrated in laboratory experiments (Scanlan and Buckling, 54 55 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., 56 1996, Briandet et al., 2008). Apart from a few examples (Dann et al., 2016, Silva et al., 2017, 57 58 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 59 60 streams, biofilms colonize the sedimentary surfaces of the streambed, are biodiversity hotspots 61 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 62 and biomass turnover in stream biofilms with potential consequences for ecosystem functioning 63 and biogeochemical cycling. However, the heterogeneous matrix of stream biofilms has 64 precluded the study of viromes in stream biofilms. 65

The first critical step towards the study of biofilm viruses in streams requires the effective extraction of viral-like particles (VLPs) from the biofilm matrix. The aim of our study was therefore to optimize laboratory methods for 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

73 glycol (PEG) precipitation, physicochemically induced biofilm breakup, and ultracentrifugation 74 followed by nucleic acid extraction. Several of these protocols have been described previously (Vega Thurber et al., 2009, Danovaro and Middelboe, 2010, Hurwitz et al., 2013, Trubl et al., 75 76 2016), but their rigorous testing for virome generation from stream biofilms is lacking at present. 77 We benchmark the efficiency of the different protocols using epifluorescence microscopy counts 78 of VLP and DNA yields, and provide first results of the virome structure from biofilms in three 79 streams. We provide a step-by-step version of the optimized protocol at protocols.io, which allows for community participation and continuous protocol development. 80

81

## 82 Materials & Methods83 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 biofilm were randomly collected from stones (5 to 15 cm in diameter) using sterile brushes and Milli-Q water. Slurries were transported on ice to the laboratory pending further processing.

91

#### 92 **Biofilm properties**

93 For bacterial cell counting, samples were fixed with 3.7% formaldehyde (final concentration)
94 and stored at 4°C. Bacterial cells were disintegrated from the biofilm matrix using 0.25 mM
95 tetrasodium pyrophosphate in combination with rigorous shaking (1 h) and sonication (Bransonic
96 Sonifier 450, Branson) on ice (1 min). Cells were stained using Syto13 and counted on a flow

97 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-98 night extraction in 90% EtOH (Lorenzen, 1967). Extracellular polymeric substances (EPS) were 99 extracted from biofilm slurries using 50 mM EDTA and shaking (1 h) (Battin et al., 2003). 100 101 precipitated in 70% EtOH (-20°C, 48 h) Carbohydrates were and measured 102 spectrophotometrically as glucose equivalents (DuBois et al., 1956). Proteins were determined according to Lowry (Lowry et al., 1951). 103

104

#### 105 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.

112

#### 113 Protocols for the extraction of VLP

We explored a variety of protocols for the concentration, extraction and purification of stream biofilm VLP (Fig. 1). The concentration step is required to obtain sufficient genetic material for nucleic acid extraction. Extraction aims at the liberation of VLP from the biofilm matrix, while purification aims at reducing the amount of contaminant nucleic acids and cellular debris. 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 sediment particles and larger organisms contained within the biofilm slurry. We recovered the supernatant for downstream analyses.

124

#### 125 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<sub>2</sub>O.

137

#### 138 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 147 148 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 149 150 samples and incubated them in the dark (15 minutes). Then, all samples were sonicated 151 (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 152 153 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. At the end of the incubation period, samples were chilled on ice. 154

Biofilm samples without chemical treatment served as controls. After each treatment, these samples were first centrifuged (4°C) at 3'234 g for 15 minutes (5810R, Eppendorf) and the supernatant was sequentially filtered through 0.8  $\mu$ m and 0.45  $\mu$ m filters (Whatman, GE Healthcare, USA) to remove debris and cells.

159

#### 160 **Purification**

To eliminate contaminating eukaryotic, prokaryotic and extracellular nucleic acids, DNase I at a final concentration of  $2.5U/\mu L$  (Life Technologies, Germany) and Tris buffer (100 mM Tris pH 7.5, 5 mM CaCl<sub>2</sub> and 25 mM MgCl<sub>2</sub>) 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'-

166 CCTACGGGNGGCWGCAG-3') and 785r (5'-GACTACHVGGGTATCTAAKCC-3'). PCR 167 amplification was performed with a Biometra Thermal Cycler using Taq ALLin polymerase 168 (Axon Lab) according to the manufacturer's instructions with an initial enzyme activation step 169 (95°C for 2 minutes) followed by 30 cycles of denaturation (95°C for 30 seconds) and 170 hybridization (72°C for 30 seconds) and a final elongation (72°C for 10 minutes). PCR products 171 were visualized under UV light after migration on an agarose gel stained using GelRed (Biotium, 172 USA).

For final purification of the viral DNA fraction, two protocols of density gradient 173 174 ultracentrifugation were assessed. A subset of sample was deposited on top of a CsCl gradient 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 175 176 (Optima XPN-80 Ultracentrifuge, 32 SWi, Beckman Coulter, USA) for 2 h (4°C). The viral 177 fraction was harvested from the interface between the 1.5 and 1.7 g/mL CsCl layers using an 18G needle. Similarly, sucrose gradient ultracentrifugation was assessed for final purification of 178 179 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 180 both layers using an 18G needle. 181

182

#### 183 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. Subsamples were fixed with formaldehyde, stained with SYBR 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).

196

#### 197 Nucleic acid extraction

To further assess the efficiency of the tested protocols to obtain viral nucleic acids for 198 199 metagenome sequencing, we extracted DNA and quantified its concentration. Following 200 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, and incubated at room temperature 201 202 (30 minutes). Then, two volumes of 100% EtOH were added and incubated 30 minutes ( $4^{\circ}$ C). 203 Samples were centrifuged at 17'000 g (4°C, 20 minutes) and pellets washed twice with 70% cold 204 EtOH. Pellets were air-dried and resuspended in 567 µL of TE buffer (10 mM Tris and 1 mM 205 EDTA (pH 8.0)). Thirty  $\mu$ L of warm 10% SDS and 3  $\mu$ L 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 206 warm CTAB were added and incubated for 10 minutes (65°C). DNA was extracted in a series of 207 208 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 209 overnight in isopropanol ( $-20^{\circ}$ C). DNA was concentrated by centrifugation at 16'000 g (4°C) 210 211 for 20 min, and the pellet washed twice with cold 70% EtOH, air-dried, resuspended in 50  $\mu$ L

212 nuclease-free  $H_2O$  and stored (-20°C). DNA concentration was measured using Qubit and the 213 dsDNA high-sensitivity kit according to the manufacturer's instructions (Life Technologies, 214 Carlsbad, USA).

215

#### 216 Library construction and sequencing

217 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 218 metagenome sequencing. For sequencing library construction, DNA was sheared with an S2 219 220 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, 221 222 USA) which allows low quantities of both single- and double-stranded DNA as input (Roux et 223 al., 2016). Library construction and multiplexing was performed following the manufacturer's instructions for DNA inputs (<1 ng/ $\mu$ L) and 20 cycles of indexing PCR. Paired-end sequencing 224 225 (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 226 Nucleotide Archive under accession number PRJEB33548. 227

228

#### 229 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<sup>®</sup> 1S Plus library preparation kit includes a low complexity 235 adaptase tail, which was clipped (10 bases) according to the manufacturer's instructions. We 236 then used the error correction capability of Tadpole (v. 37.76) to correct for sequencing errors 237 (mode=correct ecc=t prefilter=2). Prior to assembly, we de-duplicated our datasets using 238 239 clumpify (v37.76) with parameters set such that identical reads were identified and only one 240 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 241 kmers set to 21, 33, 55, 77, 99, 127 to assemble contigs. We co-assembled the paired-end reads 242 243 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 244 245 de-replicated reads (forward orientation only) to the web interface of taxonomer (Flygare *et al.*, 246 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 247 "unknown" bins. For identification and classification of viral contigs, we used MetaPhinder 248 (v2.1, Jurtz et al., 2016) through the web interface hosted at the Center for Genomic 249 Epidemiology at the Danish Technical University (DTU). Additionally, we queried the Viral rep 250 and Phage F domains of the PFAM database using hmmsearch (HMMER v3, Eddy, 2011) to 251 identify ssDNA contigs (Trubl et al., 2019). 252

253

254 Results

#### 255 **Biofilm properties**

Bacterial abundance ranged between  $4.1 \times 10^{11}$  cells m<sup>-2</sup> at the lowest stream (SNG) and  $2.3 \times 10^9$  cells m<sup>-2</sup> 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).

263

#### 264 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.

269

#### 270 Virome extraction and purification

In total, we evaluated 16 protocols to concentrate, extract and purify viruses from benthic 271 272 biofilms from three streams (Fig. 1). Based on the number of VLPs and DNA yields retained at 273 the end of each protocol, we observed significant differences in relative extraction efficiencies 274 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 275 combination of TFF for concentration, tetrasodium pyrophosphate and sonication for extraction, 276 and sucrose gradient centrifugation for purification resulted in the highest VLP counts in all three 277 278 samples (two-way ANOVA, p <0.01; Fig. 3). This pipeline also yielded highest DNA concentration (Fig. 4). Protocols involving PEG precipitation generally resulted in a lower 279 recovery of VLPs (on average only 13.6 % compared to the best performing pipeline) and DNA 280 281 yields below detection limit. This may be attributable to the formation of a visible, viscous layer

282 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). 283 Protocols based on TFF and using DTT or chloroform extracted on average 25.0% and 33.2% 284 less VLPs than protocols using TFF followed by tetrasodium pyrophosphate treatment and 285 286 sonication (Fig. 3). Samples without any physicochemical treatment to extract VLP from 287 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 288 CsCl were tested. On average, 1.9 times more VLPs were retained by ultracentrifugation using 289 290 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 291 DNA yields, which reached 1.22 ng  $\mu$ L<sup>-1</sup> in VDN, 1.31 ng  $\mu$ L<sup>-1</sup> in VEV and 18.7 ng  $\mu$ L<sup>-1</sup> in SNG 292 293 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 294 CsCl gradient (paired T-Test, p=0.03) and on average 1.5, 1.7 and 1.9 times higher using 295 296 tetrasodium pyrophosphate and sonication compared to DTT, no physico-chemical detachment 297 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.

302

#### 303 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.

309 We obtained 3698, 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 9493, 310 311 46'665 and 54'492 bp in SNG, VEV and VDN, respectively. Hmmsearch against the PFAM 312 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 313 314 were identified as not further classified Siphoviridae (28.3 - 53.3%), followed by Myoviridae 315 (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 316 317 Twortvirus, Phicbkvirus, Coopervirus and L5virus were only detected in viromes obtained from 318 VEV and VDN but not in SNG.

319

#### 320 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-andpurification-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).

The viromes obtained using the best-performing protocol were dominated by reads of viral or 340 341 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 342 343 viromes. Following an optimized assembly strategy (Roux et al., 2019), the reads assembled into 344 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 345 346 three different stream biofilms, with several contigs classified as T4virus, Cp220virus, Kayvirus 347 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 348 349 freshwater, marine and soil ecosystems (Roux et al., 2016, Trubl et al., 2019) could not be 350 detected in the viromes from stream biofilms. 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.

355 A range of chemical properties may explain the relative differences in virus extraction efficiency 356 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. 357 Previously described methods for isolating viruses involve chloroform (Marhaver et al., 2008, 358 359 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 360 361 DNA viruses (NCLDV), including Phycodnaviridae, which predominantly infect freshwater and 362 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 363 chloroform. Similarly, DTT is a reducing agent, which breaks disulfide bonds in proteins. This 364 365 may explain to the reduced recovery of VLPs from samples treated with PEG, chloroform or DTT. 366

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. This 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

374 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 375 may be too buoyant (Vega Thurber et al., 2009), or degrade the structure of enveloped viruses 376 (Lawrence and Steward, 2010) and therefore reduce their recovery. A critical step of virome 377 preparation concerns the nucleic acid extraction. We chose a derivation of a standard protocol, 378 379 which allows the extraction of both ssDNA and dsDNA viruses (Sambrook et al., 1989). 380 381 **Conclusions** In conclusion, we provide a first protocol for the generation of viromes from stream biofilms. 382 383 This payes 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 384 phage diversity in stream biofilms. 385 386 387 Acknowledgements 388 We acknowledge the help of Davide Demurtas for TEM imaging. 389 References 390 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al. (2012). SPAdes: 391 392 a new genome assembly algorithm and its applications to single-cell sequencing. J393 Comput Biol 19: 455-477. Battin TJ, Kaplan LA, Newbold JD, Cheng X, Hansen C (2003). Effects of current velocity on 394 the nascent architecture of stream microbial biofilms. Appl Environ Microbiol 69: 5443-395 5452. 396 Battin TJ, Besemer K, Bengtsson MM, Romani AM, Packmann AI (2016). The ecology and

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

404	Briandet R, Lacroix-Gueu P, Renault M, Lecart S, Meylheuc T, Bidnenko E et al. (2008).
405	Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages
406	inside biofilms. Appl Environ Microbiol 74: 2135-2143.
407	Brum JR, Sullivan MB (2015). Rising to the challenge: accelerated pace of discovery transforms
408	marine virology. Nature Reviews Microbiology 13: 147.
409	Chan BK, Abedon ST (2015). Bacteriophages and their enzymes in biofilm control. Curr Pharm
410	<i>Des</i> <b>21</b> : 85-99.
411	Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE et al. (2019). Viruses
412	control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic
413	fracturing. Nat Microbiol 4: 352-361.
414	Dann LM, Paterson JS, Newton K, Oliver R, Mitchell JG (2016). Distributions of Virus-Like
415	Particles and Prokaryotes within Microenvironments. PLoS One 11: e0146984.
416	Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C et al. (2008).
417	Major viral impact on the functioning of benthic deep-sea ecosystems. Nature 454: 1084.
418	Danovaro R, Middelboe M (2010). Separation of free virus particles from sediments in aquatic
419	systems. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). Manual of Aquatic Viral
420	Ecology American Society of Limnology and Oceanography: Waco, TX.
421	Dell'Anno A, Corinaldesi C, Danovaro R (2015). Virus decomposition provides an important
422	contribution to benthic deep-sea ecosystem functioning. Proc Natl Acad Sci U S A 112:
423	E2014-2019.
424	Doolittle MM, Cooney JJ, Caldwell DE (1996). Tracing the interaction of bacteriophage with
425	bacterial biofilms using fluorescent and chromogenic probes. J Ind Microbiol 16: 331-
426	341.
427	DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956). Colorimetric Method for
428	Determination of Sugars and Related Substances. Analytical Chemistry 28: 350-356.
429	Eddy SR (2011). Accelerated Profile HMM Searches. PLoS Comput Biol 7: e1002195.
430	Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB et al. (2018). Host-linked soil
431	viral ecology along a permafrost thaw gradient. <i>Nature Microbiology</i> <b>3:</b> 870-880.
432	Feldman HA, Wang SS (1961). Sensitivity of Various Viruses to Chloroform. Proceedings of the
433	Society for Experimental Biology and Medicine <b>106:</b> 736-738.
434	Flemming HC, Wingender J (2010). The biofilm matrix. Nat Rev Microbiol 8: 623-633.
435	Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T et al. (2016). Taxonomer: an
436	interactive metagenomics analysis portal for universal pathogen detection and host

- 437 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.
- 457 Marhaver KL, Edwards RA, Rohwer F (2008). Viral communities associated with healthy and
  458 bleaching corals. *Environmental microbiology* 10: 2277-2286.
- 459 Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N
  460 *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.
- 466 Peduzzi P (2016). Virus ecology of fluvial systems: a blank spot on the map? *Biol Rev Camb*467 *Philos Soc* 91: 937-949.
- 468 Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA (2018). Deposition rates of viruses and
  469 bacteria above the atmospheric boundary layer. *ISME J* 12: 1154-1162.
- 470 Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer
  471 F *et al.* (2009). Explaining microbial population genomics through phage predation.
  472 *Nature Reviews Microbiology* 7: 828-836.
- 473 Rohwer F (2003). Global Phage Diversity. *Cell* **113:** 141.
- 474 Rosario K, Breitbart M (2011). Exploring the viral world through metagenomics. *Curr Opin*475 *Virol* 1: 289-297.
- 476 Rosario K, Fierer N, Miller S, Luongo J, Breitbart M (2018). Diversity of DNA and RNA
  477 Viruses in Indoor Air As Assessed via Metagenomic Sequencing. *Environ Sci Technol*478 52: 1014-1027.
- 479 Rossum TV, Uyaguari-Diaz MI, Vlok M, Peabody MA, Tian A, Cronin KI *et al.* (2018).
  480 Spatiotemporal dynamics of river viruses, bacteria and microeukaryotes. *bioRxiv*:
  481 259861.

482	Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB et al. (2016).
483	Towards quantitative viromics for both double-stranded and single-stranded DNA
484	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.
- 488 Roux S, Trubl G, Goudeau D, Nath N, Couradeau E, Ahlgren NA *et al.* (2019). Optimizing de
  489 novo genome assembly from PCR-amplified metagenomes. *PeerJ* 7: e6902.
- 490 Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: a laboratory manual*. Cold
  491 Spring Harbor Laboratory Press: New York.
- 492 Scanlan PD, Buckling A (2012). Co-evolution with lytic phage selects for the mucoid phenotype
   493 of Pseudomonas fluorescens SBW25. *ISME J* 6: 1148-1158.
- 494 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T *et al.* (2012). Fiji: an
  495 open-source platform for biological-image analysis. *Nat Methods* 9: 676-682.
- 496 Silva BSO, Coutinho FH, Gregoracci GB, Leomil L, de Oliveira LS, Froes A *et al.* (2017).
  497 Virioplankton Assemblage Structure in the Lower River and Ocean Continuum of the
  498 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.
- 503 Skvortsov T, de Leeuwe C, Quinn JP, McGrath JW, Allen CC, McElarney Y *et al.* (2016).
  504 Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest
  505 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.
- 511 Sutherland IW, Hughes KA, Skillman LC, Tait K (2004). The interaction of phage and biofilms.
  512 *FEMS Microbiol Lett* 232: 1-6.
- 513 Suttle CA (2007). Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 5:
  514 801-812.
- 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.
- 526 Vidakovic L, Singh PK, Hartmann R, Nadell CD, Drescher K (2017). Dynamic biofilm
   527 architecture confers individual and collective mechanisms of viral protection. *Nature* 528 *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.

535

### Table 1(on next page)

Sample site and biofilm characteristics.

### NOT PEER-REVIEWED

	VDN	VEV	SNG
Coordinates	46°15'13.5"N	46°30'46.4"N	46°33'23.9"N
	7°06'33.9"E	6°54'43.7"E	6°28'55.3"E
Altitude (m a.s.l.)	1210	766	498
Bacterial abundance [cells m <sup>-2</sup> ]	2.3 x 10 <sup>9</sup>	1.8 x 10 <sup>10</sup>	4.1 x 10 <sup>11</sup>
Chlorophyll-a [µg cm <sup>-2</sup> ]	0.21	0.22	1.21
EPS proteins [µg cm <sup>-2</sup> ]	0.05	0.04	0.24
EPS carbohydrates [µg cm <sup>-2</sup> ]	below detection	below detection	0.08

1

### Table 2(on next page)

Virome dataset statistics

	raw reads		quality trimmed			
	paired-end sequences	GC		GC		. <i>.</i> .
		content (%)	paired-end sequences	content	Contigs	Average contig
				(%)		
SNG	24 388 096	40.36	24 172 084	40.16	3698	452
VEV	22 459 218	40.51	22 153 818	40.24	11323	645
VDN	23 804 190	40.68	23 422 408	40.21	13591	676

1

#### Figure 1(on next page)

Overview of methods for the extraction and purification of viruses from stream biofilms.

First, phages are concentrated using either PEG precipitation or TFF. Different physicochemical 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 CsCl or sucrose gradients was used to purify viruses for downstream molecular analyses. Combinations of all protocols were evaluated for the recovery of VLPs and DNA yield.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27914v1 | CC BY 4.0 Open Access | rec: 23 Aug 2019, publ: 23 Aug 2019

## Figure 2

Electron microscopic evidence of virus-like particles in stream biofilm samples.

A large morphological diversity of VLPs, including tailed bacteriophages, lemon-shaped, polyhedral, spherical and filamentous viruses was observed using TEM.

#### NOT PEER-REVIEWED



#### Figure 3(on next page)

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, tetrasodium pyrophosphate in combination with sonication and ultracentrifugation in sucrose gradient yielded significantly higher VLP counts than any other combination of protocols.



#### Figure 4(on next page)

DNA yields from samples processed with different protocols.

None of the protocols involving PEG precipitation resulted in detectable DNA yields (not shown). Protocols using TFF and either no treatment (control), chloroform, DTT, tetrasodium pyrophosphate in combination with sonication for extraction and either CsCl or Sucrose gradient ultracentrifugation for purification yielded between 0.5 and 18.7 ng DNA  $\mu$ L<sup>-1</sup>. 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.



#### Figure 5(on next page)

Viral contig taxonomic composition.

Taxonomic classification after contig assembly showed that 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.

