Optimised biomolecular extraction for metagenomic analysis of microbial biofilms from high-mountain streams

Susheel Bhanu Busi[1,*], Paraskevi Pramateftaki[2,*], Jade Brandani[2], Stilianos Fodelianakis[2], Hannes Peter[2], Rashi Halder[1], Paul Wilmes[1] and Tom J. Battin[2]

[1] Systems Ecology Research Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2] Stream Biofilm and Ecosystems Research group, École Polytechnique Federale de Lausanne, Lausanne, Switzerland
[*] These authors contributed equally to this work.

ABSTRACT

Glacier-fed streams (GFS) are harsh ecosystems dominated by microbial life organized in benthic biofilms, yet the biodiversity and ecosystem functions provided by these communities remain under-appreciated. To better understand the microbial processes and communities contributing to GFS ecosystems, it is necessary to leverage high throughput sequencing. Low biomass and high inorganic particle load in GFS sediment samples may affect nucleic acid extraction efficiency using extraction methods tailored to other extreme environments such as deep-sea sediments. Here, we benchmarked the utility and efficacy of four extraction protocols, including an up-scaled phenol-chloroform protocol. We found that established protocols for comparable sample types consistently failed to yield sufficient high-quality DNA, delineating the extreme character of GFS. The methods differed in the success of downstream applications such as library preparation and sequencing. An adapted phenol-chloroform-based extraction method resulted in higher yields and better recovered the expected taxonomic profile and abundance of reconstructed genomes when compared to commercially-available methods. Affordable and straight-forward, this method consistently recapitulated the abundance and genomes of a mock community, including eukaryotes. Moreover, by increasing the amount of input sediment, the protocol is readily adjustable to the microbial load of the processed samples without compromising protocol efficiency. Our study provides a first systematic and extensive analysis of the different options for extraction of nucleic acids from glacier-fed streams for high-throughput sequencing applications, which may be applied to other extreme environments.

Subjects Ecosystem Science, Microbiology, Molecular Biology
Keywords Biofilms, Streams, Alpine streams, Glacier fed streams, Glaciers, Metagenomics, Biomolecular extraction

INTRODUCTION

The advent of high-throughput sequencing technologies has brought hitherto inconceivable capacities to characterize the microbial ecology of both well-studied (*Jansson & Hofmockel, 2018*; *Nielsen & Ji, 2015*) and under-explored environments (*Hotaling, Hood & Hamilton,*

2017; *Milner et al., 2017*). Among the latter include high-mountain and particularly glacier-fed streams (*Milner et al., 2017*) and the microbial biofilms that colonize their beds (Battin et al. 2016). Today, these streams are changing at an unprecedented pace owing to climate change and the thereby shrinking glaciers, and yet little is known of their microbial diversity (*Wilhelm et al., 2013*; *Milner et al., 2017*). Glacier-fed stream (GFS) sediments are extreme habitats characterized by low microbial cell abundance and activities but very high loads of fine mineral particles (*Wilhelm et al., 2013*; *Godone, 2017*; *Peter & Sommaruga, 2017*; *Chanudet & Filella, 2008*). In order to understand the diversity and composition of these microbial communities, including both eukaryotes and prokaryotes, and the role that they play, it is essential to extract nucleic acids in sufficient quantity and quality from often complex environmental matrices. After extracting the nucleic acids, downstream applications including molecular biology methods such as PCR and next-generation sequencing of amplicons or metagenomes allow for the compositional, functional and phylogenetic characterization of microbial populations and the communities that they form (*Roume et al., 2013*).

While there is no lack of protocols and literature pertaining to the extraction of nucleic acids from a wide variety of environments (*Roume et al., 2013*; *Miller et al., 1999*; *Xin & Chen, 2012*; *Porebski, Bailey & Baum, 1997*; *Zhou, Bruns & Tiedje, 1996*), few reports dwell on the utility of these methods for biomolecular extractions from sedimentary samples with very low cell abundance as typical for GFS (*Wilhelm et al., 2013*; *Ren et al., 2017*) and in *Lever et al. (2015)* elaborately described diverse factors and components that need to be considered for efficient nucleic acid extractions. These include, but are not limited to key steps like cell lysis, removal of impurities and inhibitors and of critical additives like carrier DNA molecules to enhance aggregation and thus precipitation of DNA in case of very low concentrations. Since the first extraction of DNA by Swiss medical doctor Friedrich Miescher in 1869 (*Dahm, 2008*), biomolecule extractions have shifted from those performed with solutions prepared primarily in the laboratory (*Sambrook & Russell, 2006*; *Miller et al., 1999*) to using commercially-available kits. These ready-made options are designed to avoid the use of volatile and toxic chemicals such as phenol and chloroform, and are tailored to various environments including blood, faecal material, plant and soils (*Claassen et al., 2013*; *Psifidi et al., 2015*; *Smith, Diggle & Clarke, 2003*; *Vishnivetskaya et al., 2014*). While studies have concentrated on nucleic acid extraction from glacial ice cores (*Dancer, Shears & Platt, 1997*) or surface snow (*Pei-Ying et al., 2012*), none have demonstrated their utility for GFS sediments. Together with low cell abundance (*Wilhelm et al., 2014*; *Lever et al., 2015*), the complex mineral matrices in GFS (*Peter & Sommaruga, 2017*)—a consequence of the erosion activity of glaciers (*Bogen, 1988*)—may affect nucleic acid extraction efficiency (*Lever et al., 2015*). As we attempt to better understand how nature works at its limits through the study of extreme environments, non-commercial approaches (*Mukhopadhyay & Roth, 1993*)and methodologies (*Lever et al., 2015*), need to be revisited and optimized.

In recent years, several research groups (*Besemer et al., 2012*; *Ren et al., 2017*; *Ren, Gao & Elser, 2017*; *Dancer, Shears & Platt, 1997*) have successfully used kit-based methods for DNA extraction and subsequent 16S ribosomal RNA gene amplicon sequencing on GFS

samples. However, the requirements for whole genome shotgun sequencing currently include at least 50 ng of input DNA to minimize bias due to PCR reactions during library preparation (*Kebschull & Zador, 2015*; *Bower et al., 2015*; *Thomas, Gilbert & Meyer, 2012*; *Chafee, Maignien & Simmons, 2015*; *Peng et al., 2020*). Here, we address the utility and efficiency of the "gold" standard phenol-chloroform extraction (*Dairawan & Shett, 2020*), and three alternative methodologies to identify the process(es) that yield not only the highest quantity but also quality of DNA, from GFS sediments. Our goal was to address whether the phenol-chloroform method yielded the expected diversity and taxonomic profiles when extracting GFS sediments, while also enabling reconstruction of metagenome-assembled genomes. Simultaneously, we wanted to validate the utility of this method for the extraction of nucleic acids from both pro- and eukaryotic sources. Overall, our findings provide a framework for the extraction of nucleic acids such as DNA for whole genome shotgun sequencing from GFS sediments, whilst highlighting the potential variability introduced due to the isolation method employed.

METHODS

Sample origin & collection

DNA extraction protocols were benchmarked using three different GFS sediments from the Swiss Alps: Corbassière (CBS, collection date: 13.11.2018), 2444 m above sea level (m a.s.l) and Val Ferret (FE), 1995 m a.s.l at the glacier snout (up site, FEU, collection date: 23.10.2018) and one kilometer further downstream (down site, FED, collection date: 24.01.2019). Sampling was always performed later in the morning, before noon. Sediments (0.25 to 3.15 mm) were collected using two flame-sterilized metal sieves with a mesh size of 3.15 mm and 0.250 mm respectively. CBS differs from FEU and FED in terms of bedrock geology, with clastic sedimentary limestone dominating the catchment of CBS and brecchiated gneiss dominating in FEU and FED. Sediments generally contain more organic material further downstream from the glacier, which may inhibit DNA extraction. Sediments (0.25 to 3.15 mm) were collected using flame-sterilized sampling equipment. Wet sediments were transferred into 10 ml sterile, DNA/DNase-free tubes and immediately flash-frozen in liquid nitrogen in the field. Samples were transferred to the laboratory and kept at −80 °C until analysis. All necessary measures were taken to ensure contamination-free sampling.

DNA extraction methods

Four different DNA extraction methods were applied to the samples. The key characteristics of the different methods are summarized in Table 1. Method-1, -2 and -4 were non-commercial protocols differing primarily in the lysis step (bead-beating and lysis buffer composition; *Roume et al., 2013*; *Lever et al., 2015*; *Sambrook & Russell, 2006*) while method-3 was a modified protocol of the DNeasy PowerMax Soil Kit (Cat.No. 12988-10) provided by Qiagen (based on communication exchanged with the manufacturer). Due to the very low microbial abundance, additional precautions were taken to establish contamination-free conditions, including daily decontamination of equipment/areas with bleach, using DNA/DNase-free glassware and plasticware, reagents and chemicals.

Additionally, "germ-free" sediment is not a viable option and is hard to remove any and all microorganisms from sediments. So, extraction blanks, i.e., tubes without any sample, were used as controls, which underwent the same extraction protocols along with the other samples. Post-DNA recovery, we assessed whether any of the eluted samples had DNA via both NanoDrop and Qubit, and found them to be below detectable levels. Additionally, the PCR reactions during library preparation did not yield any product confirming serving as a contamination-check. The input weight of sediment ranged from 0.5–5 g and are described in the respective protocols.

Method-1 was based on a previously established method (*Griffiths, Whiteley & Bailey, 2000*). Introduced modifications concerned primarily the step of mechanical lysis and DNA precipitation that was rendered more stringent to improve the recovery of small amounts of DNA. Sample cell lysis was achieved by adding 0.5 g of sediment into a lysing matrix E tube with beads of variable diameter provided by the manufacturer (MP Biomedicals, SKU 116914050), 500 μl CTAB buffer (5% CTAB, 120 mM KPO4, pH 8.0) and 500 μl of phenol/chloroform/isoamyl alcohol (ratio 25:24:1). Samples were loaded on a Precellys beater for 45 s at 5.500 r/s. DNA was extracted once more with chloroform/isoamyl alcohol (24:1) and precipitated with 2 vol PEG-6000, 15 μg/ml linear polyacrylamide (LPA) and 2 h incubation on ice (Supplemental Information).

Method-2 was an adaptation to alpine stream sediments of the modular method for DNA extraction previously published (*Lever et al., 2015*). The appropriate modules of the method, based on the nature of our samples, were put together in our protocol without further modification. Samples were prepared by mixing 5 g of sediment, 10–20% of 0.1 mm zirconium beads and 1 ml of 100 mM dNTP solution. Cell lysis was achieved with 5 ml lysis buffer (30 mM Tris-HCl, 30 mM EDTA, 1% Triton X-100, 800 mM guanidium hydrochloride, pH 10.0) and incubation at 50 °C for 1 h with gentle agitation (Hybridization oven, Labnet Problot L6). The supernatant was extracted once with chloroform/isoamyl alcohol (24:1) and DNA was precipitated with $10_{260/280}$ g/ml LPA, 0.2 vol 0.5 M NaCl, 2.5 vol ethanol and 2 h incubation at RT in the dark (Supplemental Information). The input weight of 5 g sediment was a modification from previously established protocol and the subsequent reagent volumes were adjusted accordingly.

Method-3 has been previously applied successfully on sand and clay soils (*Hale & Crowley, 2015*). In this protocol, the standard lysis capacity of the DNeasy PowerMax Soil Kit (Qiagen, Cat. No. 112988-10) was modified and enhanced by the addition of phenol:chloroform:isoamyl alcohol along with PowerBeads (kit provided) and C1 solution to 5 g of sediment. Then, the manufacturer-suggested sequence of treatments and rinses with the standard buffers of the kit were followed to reach elution of extracted DNA from silica columns with 6 ml of elution buffer. Further concentration of extracted DNA was carried out with the addition of 240 μl 5M NaCl, 2.5 vol ethanol and 10μg/ml LPA (Supplemental Information). LPA was an additional modification to the original protocol for improved DNA recovery.

Method-4 involved chemical and enzymatic treatment of samples according (*Green & Sambrook, 2017*) with minor modifications. Five g of sample was mixed with 10 ml of lysis buffer, incorporating the SDS as well, (0.1 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8, 1.25%

Table 1 Key characteristics of the four selected methods. The table lists the key and specific characteristics of the four extraction methods tested, where *n*, is the total number of times each condition was tested on the material; RT: room temperature.

	Method-1 (*n* = 3)	Method-2 (*n* = 6)	Method-3 (*n* = 30)	Method-4 (*n* = 14)
Sample prep	Lysis matrix E tube (beads diam 1.4, 0.1, 4 mm)	0.1 mm Zirconium beads dNTP solution	PowerBead tubes (carnet 0.7 mm)	–
Cell lysis	Lysis buffer CTAB/KaPO4, pH 8 Phenol:Chloroform:Isoamyl alcohol Bead-beating	Lysis buffer GuHCl/EDTA/Triton X-100, pH 10 Mild vortex\Incubation at 50 °C	PowerBead Buffer+C1 Phenol:Chloroform:Isoamyl alcohol Vortex (MO BIO vortex adapter)	Lysis Buffer Tris-HCl, EDTA, SDS pH 8.0 Mild vortex Incubation at 37 °C Proteinase K addition and incubation at 70 °C
Purification	Chloroform:Isoamyl alcohol (x1)	Chloroform:Isoamyl alcohol (x1)	Inhibitor Removal Technology (C2, C3)	Phenol:Chloroform:Isoamyl alcohol (x1) Chloroform:Isoamyl alcohol (x1)
Precipitation	Linear polyacrylamide PEG-6000 Ice	Linear polyacrylamide NaCl Ethanol RT	1st: Column binding & cleaning C4, C4+EtOH, C5, EtOH Elution: 6 ml EB 2nd: Linear polyacrylamide NaCl, Ethanol	Linear Polyacrylamide Sodium acetate Isopropanol −20 °C

SDS) and 10 μl RNase A (100 mg/ml). Then sediment was vortexed for 15 s and incubated at 37 °C for 1 h in a hybridization oven. 100 μl Proteinase K (20 mg/ml) was added in a subsequent step and the mixture was incubated for 10 min at 70 °C. Samples were extracted once with phenol/chloroform/isoamyl alcohol (ratio 25:24:1) and supernatants were extracted subsequently with chloroform/isoamyl alcohol (24:1). More stringent DNA precipitation conditions were applied with the addition of 10 μg/ml LPA, and overnight incubation at −20 °C (Supplemental Information).

All DNA extracts were suspended in 100 μl of DNA/DNase–free water (ThermoFisher Scientific). Due to the inadequacy of DNA obtained from Method-1 given the 0.5g input sediment weight, we scaled the extraction to 5 g starting weight prior to sequencing. Extracted DNA was thereafter stored at −20 °C until further use. Due to the low DNA yields, it was necessary to use Qubit dsDNA HS kit (Invitrogen), a fluorescent DNA quantification method with high sensitivity. Quality assessment, with Nanodrop and DNA visualization on 0.8% agarose gel containing GelRed nucleic acid stain, was possible only for DNA extracted with method-4 and for DNA concentrations higher than 0.5 ng/ul. All samples yielded sufficient DNA, i.e., 50 ng (total input), for metagenomic sequencing and subsequent analyses. Additionally, a commercially-available microbial mock community (ZymoBIOMICS, Cat.No. D6300) was extracted using Method-4 and used for subsequent sequencing.

DNA sequencing

50 ng of DNA from all samples were subjected to random shotgun sequencing. The sequencing libraries were prepared using the NEBNext Ultra II FS DNA Library Prep kit for Illumina (Cat.No. E7805) using the protocol provided with the kit. The libraries were prepared considering 350 base pairs (bp) average insert size. Qubit (Invitrogen) was used to quantify the prepared libraries while their quality was assessed on a Bioanalyzer (Agilent). We used the NextSeq500 (Illumina) instrument to perform the sequencing using 2×150 bp read length at the Luxembourg Centre for Systems Biomedicine Sequencing Platform.

Genome reconstruction and metagenomic data processing

Paired sequences (i.e., forward and reverse) were processed using the Integrated Meta-omic Pipeline (IMP) (Narayanasamy et al., 2016). The metagenomic workflow encompasses pre-processing (read quality filtering and trimming), assembly, and genome reconstruction in a reproducible manner. The adapter sequences were trimmed in the pre-processing step including the removal of human reads. Thereafter, de novo assembly was performed using the MEGAHIT (version 2.0) assembler (Li et al., 2015). Default IMP parameters were retained for all samples. Subsequently, we used MetaBAT2 (Kang et al., 2019) and MaxBin2 (Wu, Simmons & Singer, 2016) for binning in addition to an in-house binning methodology previously described (Heintz-Buschart et al., 2017). The latter method initially ignores the ribosomal RNA sequences in kmer profiles based on VizBin embedding clusters (Laczny et al., 2015). In this context, VizBin utilises density-based non-hierarchical clustering algorithms and depth of coverage for genome reconstructions. Subsequently we obtained a non-redundant set of metagenome-assembled genomes (MAGs) using DASTool (Sieber et

al., 2018) with a score threshold of 0.7 for downstream analyses. The abundance of MAGs in each sample was determined by mapping the reads to the reconstructed genomes using BWA-MEM (Li, 2013), taking the average coverage across all contigs. Diversity measures from metagenomic sequencing were assessed by determining the abundance-weighted average coverage of all the reads to identify the number of non-redundant read sets (Rodriguez-R & Konstantinidis, 2014).

Taxonomic classification for metagenomic operational taxonomic units

We used the trimmed and pre-processed reads from the IMP workflow to determine the microbial abundance and taxonomic profiles based on the mOTU (v2) tool (Milanese et al., 2019). Based on the updated marker genes in the mOTU2 database including those from the TARA Oceans Study (Sunagawa et al., 2015) and recently generated MAGs (Tully, Graham & Heidelberg, 2018), taxonomic profiling was performed on our sequence datasets. We used a minimum alignment length of 140 bp to determine the relative abundances of the mOTUs, including the normalisation of read counts to the gene length, also accounting for the base coverage of the genes. Additionally, we used CheckM (Parks et al., 2015) to assess completeness and contamination. Subsequently, taxonomy for MAGs recovered after the redundancy analyses from DASTool was determined using the GTDB (Genome Taxonomy Database) toolkit (gtdb-tk) (Parks et al., 2018).

Data analysis

All figures for the DNA concentrations, library preparation, assembly metrics and supplementary figures were generated using GraphPad Prism (v8.3.0). Taxonomical assessment and diversity measures were created using version 3.6 of the R statistical software package (R Core Team, 2013). DESeq2 (Love, Huber & Anders, 2014) with FDR-adjustments for multiple testing were used to assess significant differences in the MAG abundances. The genomic cluster figure for the mock community was obtained as an output from the IMP metagenomic pipeline.

RESULTS

Phenol-chloroform-based extraction method results in higher DNA yields

To ensure native sequencing, by minimizing the number of PCR (polymerase chain reaction) cycles within the library preparation protocols, we tested four protocols for biomolecular extraction, with an aim of acquiring large quantities (>50 ng) DNA from glacier-fed stream benthic sediments. The four methods tested were selected because of their wide applicability on related environmental samples (Method-1 & -2) (Griffiths, Whiteley & Bailey, 2000; Lever et al., 2015; Tatti et al., 2016) and their improved chances of higher yields (Method-3; Qiagen communication). Since method-4 is considered the gold standard of DNA extraction in biomedical sciences (Dairawan & Shett, 2020) and bacterial cultures (Green & Sambrook, 2017), it was included in our study. The four protocols are largely based on the same principles, viz. sample preparation, cell lysis, purification, precipitation and washing (Table 1). From preliminary tests, it became apparent that a

small-scale approach (Method-1; 0.5 g input sediment) did not yield sufficient amounts of DNA for metagenomics due to, on average, limited microbial biomass in the samples. Thus, all protocols (aside from Qiagen's - already produced for maxi scale) were scaled up to 5 g of input sediment and a co-precipitant, like linear polyacrylamide, was included in all precipitation steps. This was essential for the quantitative recovery of the small amounts of extracted DNA from high solution volumes (6–10 ml).

Overall, we found that extractions using the commercial kit from Qiagen (method-3) yielded increased total DNA as compared to a commonly used protocol (method-1; Fig. 1A). Furthermore, method-3 was similar in terms of DNA yield when compared to a generalized protocol (method-2) previously proposed (*Lever et al., 2015*) (Fig. 1B). On the other hand, the phenol-chloroform based extraction protocol (method-4) was tested against both methods 2 and 3, using sediment samples collected from the three different glacier floodplain streams (CBS, FEU, FED) from Switzerland. Method-1 was omitted from these tests due to insufficient DNA concentrations in the preliminary extractions. We found that for all three GFS, the phenol-chloroform extraction yielded the highest DNA concentrations. In some cases, and notably samples with low cell abundance, we even obtained one order of magnitude more DNA (Fig. 1C).

Quality assessment of these DNA extracts with Nanodrop showed OD260/280 ratios between ~1.4 and ~1.6. Agarose gel electrophoresis revealed a high-molecular weight band with no apparent shearing, smearing or residual RNA, indicative of high-quality DNA (Fig. 2A). A secondary effect appearing in certain samples, but without any perceived consequences in the quality of extracted DNA whatsoever, was the development of a pink-red color of varying intensities with the addition of phenol:chloroform:isoamyl alcohol (Fig. 2B). This was pH dependent since samples were decolorized with the addition of sodium acetate pH 5.2 in the precipitation step. This could possibly be due to a ferric-chloride-phenol compound formed when chloride and phenol constituents of the protocol interact reversibly with Fe^{+3} ions contained in certain samples depending on local geology (*Banerjee & Haldar, 1950*). Similar coloration has been previously reported (*Lever et al., 2015*).

## Extraction method affects library preparation efficiency

The DNA extractions based on method-3 and using phenol-chloroform methods were subsequently subjected to library preparation for high-throughput whole genome shotgun sequencing. Despite similar quality of DNA across both methods (~1.4–1.6 OD260/280), library preparation using the modified commercial kit did not yield any successful libraries (Fig. 3). To assess if any impurities or inhibitors hampered library preparation we tested two clean-up methods for the DNA extracted from the commercial kit: (1) ethanol precipitation and (2) magnetic-bead based clean-up. For magnetic bead clean-up the SPRIselect beads (Beckman Coulter, 23318) were used according to the manufacturer's protocol. We found that the magnetic-bead method leads to a complete loss of sample (i.e., undetectable DNA quantity via Qubit analyses) during the process, especially if starting with a low input DNA concentration. Although we lost six out of twelve samples using the magnetic-bead clean-up, we achieved 100% efficiency as indicated by a concentration of greater than
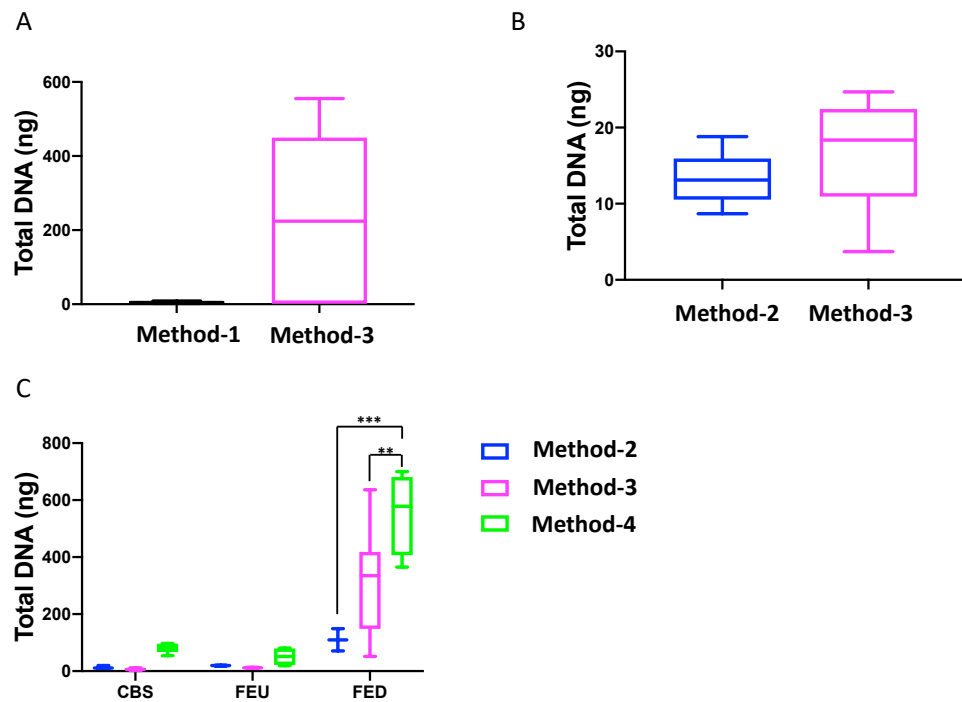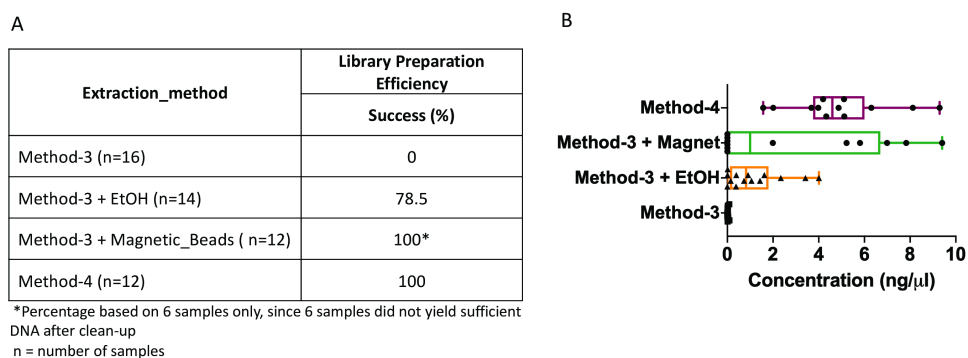
Figure 1 **Total DNA concentrations using different extraction protocols..** Boxplots represent the total amount of DNA (ng) extracted from 5 g of sediment when comparing (A) method-1 versus the modified-commercial kit-based method-3 and (B) method-2 versus method-3. (C) Boxplots of the DNA quantities isolated from three glacial floodplains (CBS - Corbassière, FEU - Val Ferret up site, FED - Val Ferret down site), using method -2, -3 and -4. Method-1: CTAB buffer lysis (*Griffiths, Whiteley & Bailey, 2000*), Method-2: Modular DNA extraction (Lever et al. 2015), Method-3: Qiagen PowerMax Soil DNA extraction kit, Method-4: Chemical and enzymatic extraction. Significance was tested using a Two-Way ANOVA with Student-Neuman Keul's post-hoc analyses. **$p < 0.01$, ***$p < 0.001$.

Full-size 🖼 DOI: 10.7717/peerj.9973/fig-1
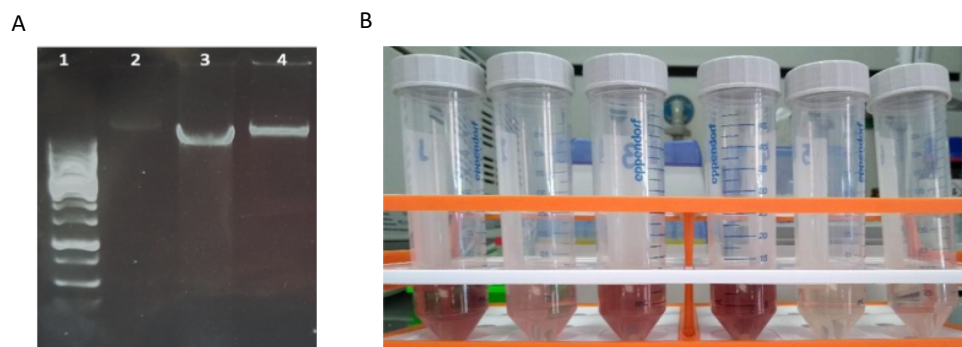
Figure 2 **Characteristics of DNA extracted with method-4.** (A) Agarose gel electrophoresis of DNA extracted with mild vortexing of sediments and incubation in lysis buffer, proteinase K treatment and phenol-chloroform extraction. Lane 1: GeneRuler 1 kb DNA ladder; lanes 2-4: CBS, FED, FEU respectively. (B) Pink-red supernatants developed during phenol:chloroform extraction step.

Full-size 🖼 DOI: 10.7717/peerj.9973/fig-2

A

Extraction_method	Library Preparation Efficiency
	Success (%)
Method-3 (n=16)	0
Method-3 + EtOH (n=14)	78.5
Method-3 + Magnetic_Beads (n=12)	100*
Method-4 (n=12)	100

*Percentage based on 6 samples only, since 6 samples did not yield sufficient DNA after clean-up
n = number of samples

B

Figure 3 **Library preparation efficiency.** The efficiency or success percentage for prepared libraries based on the individual methods is indicated in the table. Boxplots represent concentrations of the prepared libraries.

Full-size 🖼 DOI: 10.7717/peerj.9973/fig-3

0.5 ng/ul after library preparation quantified by Qubit, with the remaining six samples. On the other hand, ~20% of the samples cleaned via ethanol precipitation failed library preparation. Contradictory to these methods, DNA extracted using the phenol-chloroform based method (method-4) yielded 100% efficiency in terms of library preparation without any additional clean-up (Fig. 3). Additionally, we found that the distribution of the total yield after library preparation using the phenol-chloroform method was more uniform across samples compared to the other methods (Fig. 3).

Whole genome shotgun assembly unaffected by extraction methods

Extraction methods for whole genome shotgun sequencing may affect the sequencing itself, including the quality and assembly of the reads downstream. To assess this, we used the libraries prepared as described above (Fig. 3), and performed whole genome shotgun sequencing on an Illumina NextSeq500. The average quality across all three methods based on short-read sequencing was Q30 after trimming the leading and trailing sequences (described in Methods). We assessed several assembly metrics including the average length of contigs (N50), largest alignment, total aligned length and coverage. We did not find any significant differences among any of these measures across all three methods (Figs. 4A–4C, 4E). Using a diversity index metric, we however found a more uniform distribution across all samples prepared using method-4, albeit no significant differences to the commercial kit-based extraction and library preparation (Fig. 4D).

Extraction methods influence metagenomic profiles

It is well established that extraction methods (*Mackenzie, Brett & Taylor, 2015*) and library preparation (*Bower et al., 2015*) protocols affect the taxonomic profiles and genomes recovered after high-throughput sequencing. We determined if the preparation methods affected the overall diversity of taxa recovered and found that phenol-chloroform and the magnetic-bead clean-up methods demonstrated similar levels of diversity (Shannon) as compared to samples precipitated using ethanol (Fig. 5A). Overall, the community profiles of the ethanol precipitation-based method were highly diverse (Fig. 5B). Interestingly,
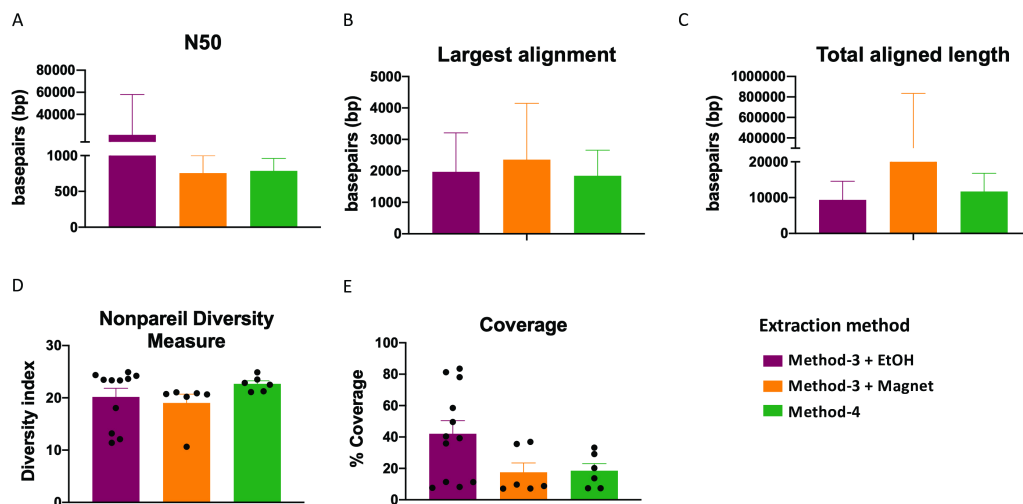
Figure 4 Estimate of assembly metrics following extraction. Barplots demonstrate the (A) N50 for the sequence assemblies, (B) length of the longest aligned sequence, (C) the total aligned length. Bars indicate standard deviation from the mean. (D) Boxplot showing the nonpareil diversity index across the three groups. (E) Percentage of coverage of the assembled sequences by read-mapping is depicted.

Full-size ▲ DOI: 10.7717/peerj.9973/fig-4

the genomes recovered and their abundances were similar in the phenol-chloroform and magnetic-bead methods as well (Fig. 5C). However, we observed a significant increase ($p < 0.001$, FDR-adjusted p-value) in the abundance of a Ralstonia genome when prepared with the ethanol precipitation protocol (Fig. S1). Additionally, we found that the number of genomes recovered using the phenol-chloroform was more consistent with previously reported 16S rRNA gene sequencing profiles for GFS from Austria (*Wilhelm et al., 2013*; *Besemer et al., 2012*; *Wilhelm et al., 2014*). Simultaneously, we used an approach to identify metagenomic operational taxonomic units (mOTUs) and found that the phenol-chloroform and magnetic-bead methods showed similar profiles of mOTUs compared to that of ethanol precipitation (Fig. 5D).

Efficiency of phenol-chloroform extraction on a mock community including eukaryotes

To determine whether the phenol-chloroform extraction method is biased against eukaryotes, we used a commercially-available mock community (ZymoBIOMICS Microbial Community Standard #D6300) to assess bias and errors. After sequencing, we recovered high quality (>90% completion, <5% contamination) bacterial genomes (Fig. 6A). Additionally, the abundance of the microbial genomes, including one of the eukaryotes - *Saccharomyces cerevisiae*, were similar to the expected levels in the mock community (Fig. 6B). On the other hand, the protocol enabled the identification and partial recovery of the Cryptococcus neoformans genome, albeit at lower levels possibly due to increased melanisation of the cell wall (*Grossman & Casadevall, 2017*) affecting lysis and subsequent extraction.
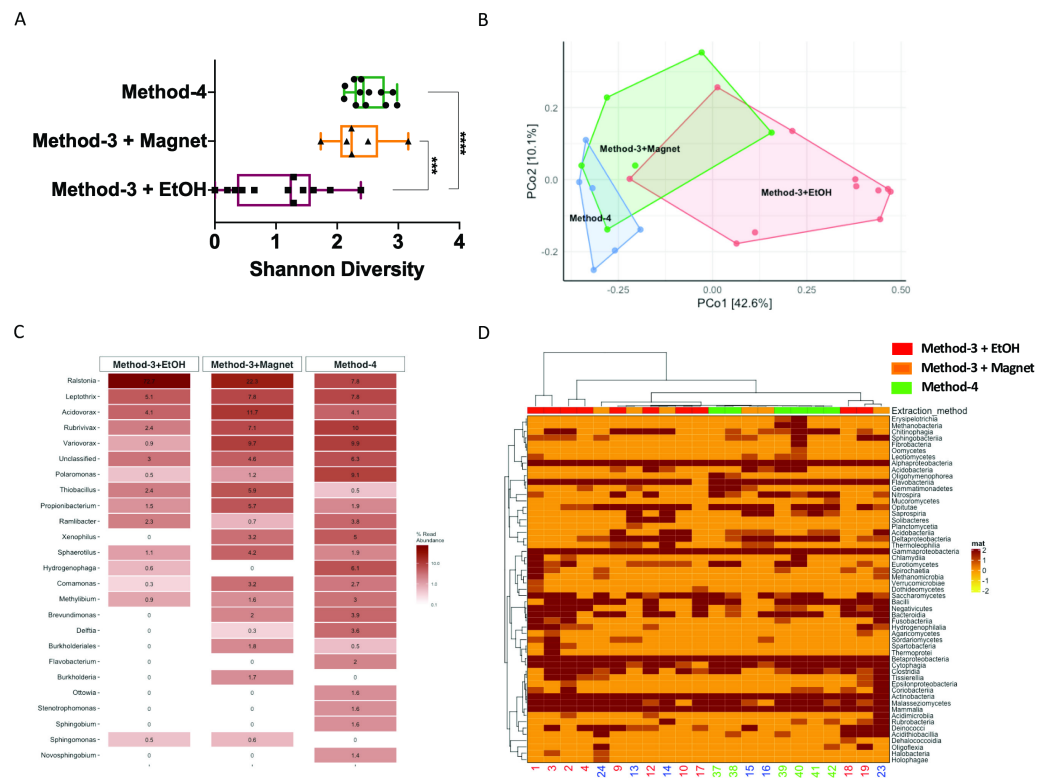
Figure 5 **Diversity and taxonomic profiles of the metagenomic sequencing.** (A) Boxplot showing the Shannon diversity index for the taxonomic profiling for the three groups. Significance was tested using a One-way ANOVA with Student-Neuman Keul's post-hoc analysis. ***p-value <0.001, ****p-value <0.0001. (B) Principal component analyses generated using Bray–Curtis dissimilarity matrix depicts similarities or lack thereof between the three groups. (C) Abundances of the reconstructed genomes are depicted for method-3 + EtOH, method-3 + magnetic bead clean-up and method-4 extraction. (D) Heatmap demonstrating the mOTUs for the three methods is depicted. The hierarchical clustering for the heatmap was generated using Ward's clustering algorithm.

Full-size 🖼 DOI: 10.7717/peerj.9973/fig-5

DISCUSSION

Improved omic techniques not limited to metagenomics are robust methods for analyzing nucleic acids and the characterisation of microbial communities in various environments (*Jansson et al., 2012*). One way of understanding the impacts of global climate change on GFS includes the establishment of their census of microbial life (*Milner et al., 2017*). However, methods designed for the extraction of biomolecules including DNA have not been validated for GFS sediments. Although previous glacier-fed streams studies successfully used extracted DNA for 16S rRNA amplicon sequencing (Ren et al. 2017; Ren, Gao, and Elser 2017; Vardhan (*Reddy & Vishnu, 2009*; *Wilhelm et al., 2013*) the input DNA concentration requirements are considerably higher for whole genome shotgun sequencing. In order to pursue a deeper characterisation of the microbial communities within the GFS sediments, increased concentrations of DNA may additionally alleviate PCR biases (*Brooks et al., 2015*; *Kim & Bae, 2011*). Also, as previously highlighted, several methods exist for extractions from a wide variety of environmental samples, but not for
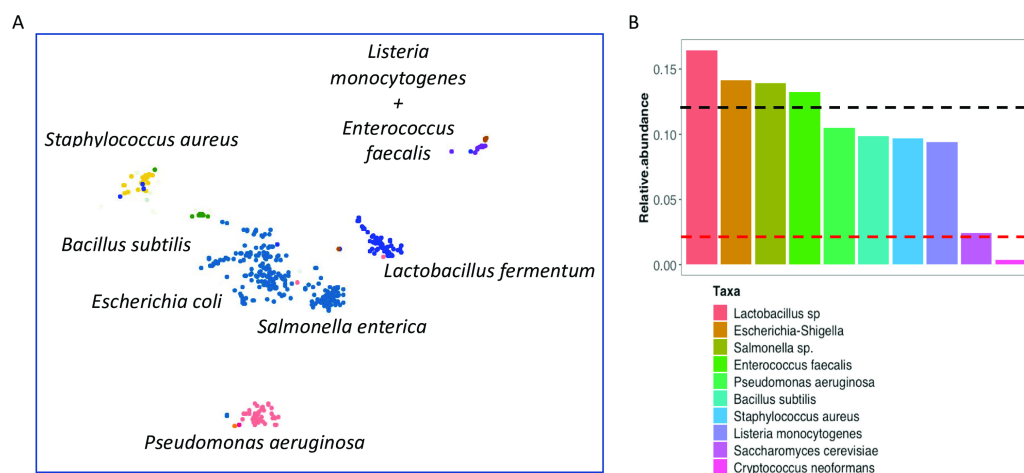
Figure 6 **Evaluation of phenol-chloroform extraction using a mock community.** (A) Scatterplot depicts the clusters of contigs representative of the reconstructed genomes after processing the mock community using the IMP meta-omics pipeline. The taxonomic identity is displayed next to the respective clusters. (B) Barplots indicate the relative abundance of the individual genomes recovered from the mock community sequencing after extraction with the phenol-chloroform method. The upper (black) line represents the expected abundance (12%) of the prokaryotes, while the lower (red) line indicates the expected abundance (2%) of the eukaryotes.

Full-size 🖾 DOI: 10.7717/peerj.9973/fig-6

GFS sediments. Here, we systematically tested the utility of four extraction protocols to identify a ubiquitous methodology. We found that a phenol-chloroform based extraction protocol can be used for samples across geographical separations, differences in bedrock, and samples collected at various distances from the glacier.

Glassing et al. (2016) demonstrated that inherent DNA contamination may influence microbiota interpretation in low biomass samples. Additionally, it is known that certain compounds—polysaccharides, humic acids, may affect PCR reactions *Rådström et al., 2004*, requiring the need for additional DNA clean-up. It has been established that DNA losses occur during the purification steps (*Roose-Amsaleg, Garnier-Sillam & Harry, 2001*), including when using commercial column methods (*Howeler, Ghiorse & Walker, 2003*; *Lloyd et al., 2013*), and phenol-chloroform (*Ogram, Sayler & Barkay, 1987*). Interestingly, we found similar losses when using the magnetic bead clean-up, whereas the ethanol precipitation method was inefficient compared to the phenol-chloroform protocol. Though the kit-based methods are more convenient and safer than phenol-chloroform extractions (*Tesena et al., 2017*), access to reagents and costs may be a considerable factor. On the other hand, isolation of the aqueous phase from phenol-chloroform can be user-dependent potentially affecting reproducibility, while kits have been shown to be more consistent across samples (*Claassen et al., 2013*). Another key feature of our findings was the potential for the kit-based methods to influence the efficiency of genome reconstruction and variability in the taxonomic profiles that were recovered. While this has been reported previously (*Mackenzie, Brett & Taylor, 2015*; *Carrigg et al., 2007*), we found considerable variability when compared to phenol-chloroform. This is plausible due to the incomplete dissolution of DNA in buffers, especially when using methods involving charged minerals

(*Vorhies & Gaines, 2009*; *Barton, Taylor & Pemberton, 2006*; Vishnivetskaya et al. 2014), which may additionally affect DNA stability.

CONCLUSIONS

The utility of extraction methods extends beyond the process itself, impacting downstream applications such as whole genome shotgun sequencing. Our study shows that phenol-chloroform may be an under-appreciated yet powerful method for isolating nucleic acids from glacier-fed stream sediments. While additional steps may be required towards extraction of other biomolecules such as RNA, proteins and metabolites, minor modifications may be sufficient (*Toni et al., 2018*). Moreover, we report for the first time a systematic assessment of biomolecular extraction methods for GFS sediments. Our findings though fundamental and previously unexplored, may lay the foundations for future in-depth characterisation of GFS microbial communities.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Susheel Bhanu Busi and Paraskevi Pramateftaki conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jade Brandani, Stilianos Fodelianakis, Hannes Peter and Rashi Halder performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Paul Wilmes and Tom J. Battin conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Data is available at NCBI: PRJNA624048.

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA624048/.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.9973#supplemental-information.

REFERENCES

Banerjee S, Haldar BC. 1950. Constitution of ferri-phenol complex in solution. *Nature* **165(4208)**:1012 DOI 10.1038/1651012b0.

Barton HA, Taylor NM, Lubbers BR, Pemberton AC. 2006. DNA extraction from low-biomass carbonate rock: an improved method with reduced contamination and the low-biomass contaminant database. *Journal of Microbiological Methods* **66(1)**:21–31 DOI 10.1016/j.mimet.2005.10.005.

Besemer K, Peter H, Logue JB, Langenheder S, Lindström ES, Tranvik LJ, Battin TJ. 2012. Unraveling assembly of stream biofilm communities. *The ISME Journal* **6(8)**:1459–1468 DOI 10.1038/ismej.2011.205.

Bogen B. 1988. Glacial sediment production and development of hydro-electric power in glacierized areas 13 (January). DOI 10.1017/S0260305500007539.

Bower RM, Clum A, Tice H, Lim J, Singh K, Ciobanu D, Yee Ngan C, Cheng J-F, Tringe SG, Woyke T. 2015. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. *BMC Genomics* **16(October)**:856 DOI 10.1186/s12864-015-2063-6.

Brooks JP, Edwards DJ, Harwich Jr MD, Rivera MC, Fettweis JM, Serrano MG, Reris , Ni har U Sheth RA, Huang B, Girerd P, Vaginal Microbiome Consortium, Strauss III JF, Jefferson KK, Buck GA. 2015. The truth about metagenomics: quantifying and counteracting bias in 16s rrna studies. *BMC Microbiology* **15(March)**:66 DOI 10.1186/s12866-015-0351-6.

Carrigg C, Rice O, Kavanagh S, Collins G, O'Flaherty V. 2007. DNA extraction method affects microbial community profiles from soils and sediment. *Applied Microbiology and Biotechnology* **77(4)**:955–964 DOI 10.1007/s00253-007-1219-y.

Chafee M, Maignien L, Simmons SL. 2015. The effects of variable sample biomass on comparative metagenomics. *Environmental Microbiology* **17(7)**:2239–2253 DOI 10.1111/1462-2920.12668.

Chanudet V, Filella M. 2008. Size and composition of inorganic colloids in a Peri-Alpine. *Glacial Flour-Rich Lake. Geochimica et Cosmochimica Acta* **72(5)**:1466–1479 DOI 10.1016/j.gca.2008.01.002.

Claassen S, Du Toit E, Kaba M, Clinton Moodley C, Zar HJ, Nicol MP. 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *Journal of Microbiological Methods* **94(2)**:103–110 DOI 10.1016/j.mimet.2013.05.008.

Dahm R. 2008. Discovering DNA: friedrich miescher and the early years of nucleic acid research. *Human Genetics* **122(6)**:565–581 DOI 10.1007/s00439-007-0433-0.

Dairawan M, Shett PJ. 2020. The evolution of DNA extraction methods. *American Journal of Biomedical Science & Research* **8(1)** DOI 10.34297/AJBSR.2020.08.001234.

Dancer SJ, Shears P, Platt DJ. 1997. Isolation and characterization of coliforms from glacial ice and water in canada's high arctic. *Journal of Applied Microbiology* **82(5)**:597–609 DOI 10.1111/j.1365-2672.1997.tb03590.x.

Glassing A, Dowd SE, Galandiuk S, Davis , Chiodini RJ. 2016. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens* **8(May)**:24 DOI 10.1186/s13099-016-0103-7.

Godone D. 2017. *Glacier evolution in a changing world. BoD –books on demand.*

Green MR, Sambrook J. 2017. Isolating DNA from gram-negative bacteria. *Cold Spring Harbor Protocols* **2017(1)** DOI 10.1101/pdb.prot093369.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Applied and Environmental Microbiology* **66(12)**:5488–5491 DOI 10.1128/AEM.66.12.5488-5491.2000.

Grossman NT, Casadevall A. 2017. Physiological differences in cryptococcus neoformans strains in vitro versus in vivo and their effects on antifungal susceptibility. *Antimicrobial Agents and Chemotherapy* **61(3)** DOI 10.1128/AAC.02108-16.

Hale L, Crowley D. 2015. DNA extraction methodology for biochar-amended sand and clay. *Biology and Fertility of Soils* **51(6)**:733–738 DOI 10.1007/s00374-015-1020-5.

Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider JG, Hogan A, De Beaufort C, Wilmes P. 2017. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nature Microbiology* **2(1)**:16180 DOI 10.1038/nmicrobiol.2016.180.

Hotaling S, Hood E, Hamilton TL. 2017. Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. *Environmental Microbiology* **19(8)**:2935–2948 DOI 10.1111/1462-2920.13766.

Howeler M, Ghiorse WC, Walker LP. 2003. A quantitative analysis of DNA extraction and purification from compost. *Journal of Microbiological Methods* **54(1)**:37–45 DOI 10.1016/S0167-7012(03)00006-X.

Jansson JK, Hofmockel KS. 2018. The Soil microbiome—from metagenomics to metaphenomics. *Current Opinion in Microbiology, Environmental Microbiology * The New Microscopy* **43(June)**:162–168.

Jansson JK, Neufeld JD, Moran AN, Gilbert JA. 2012. Omics for understanding microbial functional dynamics. *Environmental Microbiology* **14(1)**:1–3 DOI 10.1111/j.1462-2920.2011.02518.x.

Kang DD, Thomas A, Egan R, An H, Wang Z, Hong A, Zhong W. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7(July)**:e7359 DOI 10.7717/peerj.7359.

Kebschull JM, Zador AM. 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research* **43(21)**:e143 DOI 10.1093/nar/gku1263.

Kim K-H, Bae J-W. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Applied and Environmental Microbiology* **77(21)**:7663–7668 DOI 10.1128/AEM.00289-11.

Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Hera Margossian H, Coronado S, Van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3(1)**:1 DOI 10.1186/s40168-014-0066-1.

Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Frontiers in Microbiology* **6(May)**:476 DOI 10.3389/fmicb.2015.00476.

Li H. 2013. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv [q-bio.GN]. arXiv. Available at http://arxiv.org/abs/1303.3997*.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* **31(10)**:1674–1676 DOI 10.1093/bioinformatics/btv033.

Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lennk S, Schramm A, Jorgenson BB. 2013. Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496(7444)**:215–218 DOI 10.1038/nature12033.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biology* **15(12)** DOI 10.1186/s13059-014-0550-8.

Mackenzie W, Brett DWW, Taylor MW. 2015. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Frontiers in Microbiology* **6(February)**:130.

Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, Hingamp P, Alves R, Costea PI, Coelho LP, Schmidt TSB, Almeida A, Mitchell AL, Finn RD, Huerta-Cepas J, Bork P, Zeller G, Sunagawa S. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications* **10(1)**:1014 DOI 10.1038/s41467-019-08844-4.

Miller DN, Bryant JE, Madsen EL, Ghiorse WC. 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology* **65(11)**:4715–4724 DOI 10.1128/AEM.65.11.4715-4724.1999.

Milner AM, Kieran K, Tom JB, John EB, Nicholas EB, Leopold F, Sophie C-F. 2017. Glacier shrinkage driving global changes in downstream systems. *Proceedings of the National Academy of Sciences of the United States of America* **114(37)**:9770–9778 DOI 10.1073/pnas.1619807114.

Mukhopadhyay T, Roth JA. 1993. Silicone lubricant enhances recovery of nucleic acids after phenol-chloroform extraction. *Nucleic Acids Research* **21(3)**:781–782 DOI 10.1093/nar/21.3.781.

Narayanasamy S, Jarosz Y, Muller EEL, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Pinel N, May P, Paul W. 2016. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biology* **17(1)**:260 DOI 10.1186/s13059-016-1116-8.

Nielsen J, Ji B. 2015. New insight into the gut microbiome through metagenomics. *Advances in Genomics and Genetics* DOI 10.2147/agg.s57215.

Ogram A, Sayler GS, Barkay T. 1987. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods* **7(2)**:57–66 DOI 10.1016/0167-7012(87)90025-X.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A proposal for a standardized bacterial taxonomy based on genome phylogeny. *bioRxiv*. 256800.

Parks DH, Imelfort M, Skennerton , Philip Hugenholtz CT, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25(7)**:1043–1055 DOI 10.1101/gr.186072.114.

Pei-Ying Y, Shu-Gui H, Tuo C, Shu-Hong Z, Wei-Jun S. 2012. Methods for extraction of microorganism dna from glacier surface snow. *Sciences in Cold and Arid Regions* **4(6)**:484 DOI 10.3724/SP.J.1226.2012.00484.

Peng Z, Zhu X, Wang Z, Yan X, Wang G, Tang M, Jiang A, Kristiansen K. 2020. Comparative analysis of sample extraction and library construction for shotgun metagenomics. *Bioinformatics and Biology Insights* **14(June)**:1177932220915459.

Peter H, Sommaruga R. 2017. Alpine glacier-fed turbid lakes are discontinuous cold polymictic rather than dimictic. Inland Waters. *Journal of the International Society of Limnology* **7(1)**:45–54.

Porebski S, Bailey G, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter / ISPMB* **15(1)**:8–15 DOI 10.1007/BF02772108.

Psifidi A, Dovas CI, Bramis G, Lazou T, Russel LR, Arsenos G, Banos G. 2015. Comparison of eleven methods for genomic dna extraction suitable for large-scale whole-genome genotyping and long-term DNA Banking Using Blood Samples. *PLOS ONE* **10(1)**:e0115960 DOI 10.1371/journal.pone.0115960.

R Core Team. 2013. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/.*

Ren Z, Gao H, Elser JJ, Zhao Q. 2017. Microbial functional genes elucidate environmental drivers of biofilm metabolism in glacier-fed streams. *Scientific Reports* **7(October)** DOI 10.1038/s41598-017-13086-9.

Roume H, Heintz-Buschart A, Muller EEL, Wilmes P. 2013. Sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample. *Methods in Enzymology* **531**:219–236 DOI 10.1016/B978-0-12-407863-5.00011-3.

Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protocols* **2006(1)**:pdb.prot4455 DOI 10.1101/pdb.prot4455.

Toni LS, Garcia AM, Jeffrey DA, Jiang X, Stauffer BL, Miyamoto SD, Sucharov CC. 2018. Optimization of phenol-chloroform RNA extraction. *Methods* **5(May)**:599–608.

Vorhies JS, Gaines RR. 2009. Microbial dissolution of clay minerals as a source of iron and silica in marine sediments. *Nature Geoscience* **2(3)**:221–225 DOI 10.1038/ngeo441.

Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA, Quince C, Battin TJ. 2014. Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environmental Microbiology* **16(8)**:2514–2524 DOI 10.1111/1462-2920.12392.

Wilhelm L, Singer GA, Fasching C, Battin TJ, Besemer K. 2013. Microbial biodiversity in glacier-fed streams. *The ISME Journal* **7(8)**:1651–1660 DOI 10.1038/ismej.2013.44.

Rådström P, Knutsson R, Wolffs P, Lövenklev M, Löfström C. 2004. Pre-PCR processing: strategies to generate PCR-compatible samples. *Molecular Biotechnology* **26(2)**:133–146 DOI 10.1385/MB:26:2:133.

Reddy V, Vishnu P. 2009. Bacterial diversity and bioprospecting for cold-active enzymes from culturable bacteria associated with sediment from a melt water stream of midtre lovenbreen glacier, an arctic glacier. *Research in Microbiology* **160(8)**:538–546 DOI 10.1016/j.resmic.2009.08.008.

Ren Z, Gao H, Elser JJ. 2017. Longitudinal variation of microbial communities in benthic biofilms and association with hydrological and physicochemical conditions in glacier-fed streams. *Freshwater Science* **36(3)**:479–490 DOI 10.1086/693133.

Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30(5)**:629–635 DOI 10.1093/bioinformatics/btt584.

Roose-Amsaleg CL, Garnier-Sillam E, Harry M. 2001. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology: A Section of Agriculture, Ecosystems & Environment* **18(1)**:47–60 DOI 10.1016/S0929-1393(01)00149-4.

Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* **3(7)**:836–843 DOI 10.1038/s41564-018-0171-1.

Smith K, Diggle MA, Clarke SC. 2003. Comparison of commercial DNA extraction kits for extraction of bacterial genomic DNA from whole-blood samples. *Journal of Clinical Microbiology* **41(6)**:2440–2443 DOI 10.1128/JCM.41.6.2440-2443.2003.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searso S, Kandels-Lewis S, Tara Oceans coordinators, Bowler C, De Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P. 2015.

Ocean plankton, structure and function of the global ocean microbiome. *Science* **348(6237)**:1261359 DOI 10.1126/science.1261359.

Tatti E, McKew BA, Whitby C, Smith CJ. 2016. Simultaneous DNA-RNA extraction from coastal sediments and quantification of 16S rRNA genes and transcripts by real-time PCR. *Journal of Visualized Experiments* **112(June)** DOI 10.3791/54067.

Tesena P, Korchunjit W, Taylor J, Wongtawan T. 2017. Comparison of commercial RNA extraction kits and qPCR master mixes for studying gene expression in small biopsy tissue samples from the equine gastric epithelium. *Journal of Equine Science* **28(4)**:135–141 DOI 10.1294/jes.28.135.

Thomas T, Gilbert J, Meyer F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2(1)**:3 DOI 10.1186/2042-5783-2-3.

Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2, 631 draft metagenome-assembled genomes from the global oceans. *Scientific Data* **5(January)**:170203 DOI 10.1038/sdata.2017.203.

Vishnivetskaya TA, Layton AC, Lau MCY, Chauhan A, Cheng KR, Meyers AJ, Murphy JR, Rogers AW, Saarunya GS, Williams DE, Pfiffner SM, Biggerstaff JP, Stackhouse BT, Phelps TJ, Whyte L, Sayler GS, Onstott TC. 2014. Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiology Ecology* **87(1)**:217–230 DOI 10.1111/1574-6941.12219.

Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32(4)**:605–607 DOI 10.1093/bioinformatics/btv638.

Xin Z, Chen J. 2012. A high throughput DNA extraction method with high yield and quality. *Plant Methods* DOI 10.1186/1746-4811-8-26.

Zhou J, Bruns MA, Tiedje JM. 1996. DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology* **62(2)**:316–322 DOI 10.1128/AEM.62.2.316-322.1996.