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

Susheel Bhanu Busi [Corresp., Equal first author, 1], **Paraskevi Pramateftaki** [Equal first author, 2], **Jade Brandani** [2], **Stilianos Fodelianakis** [2], **Hannes Peter** [2], **Rashi Halder** [1], **Paul Wilmes** [1], **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

Corresponding Author: Susheel Bhanu Busi
Email address: susheelbhanu@gmail.com

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.

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, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[#]*Co-first authors*

[*]*Corresponding author: Susheel Bhanu Busi*

Address: 7 avenue des hauts-fourneaux, BioTech-I, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, L-4362 Luxembourg

Email: susheel.busi@uni.lu

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.

Introduction

The advent of high-throughput sequencing technologies has brought hitherto inconceivable capacities to characterize the microbial ecology of both well-studied (Jansson and Hofmockel 2018; Nielsen and Ji 2015) and under-explored environments (Hotaling, Hood, and 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 and Sommaruga 2017; Chanudet and Filella 2008; Bogen 1988). 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 and Chen 2012; Porebski, Bailey, and Baum 1997; Zhou, Bruns, and 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, Gao, and Elser 2017). In 2015, Lever *et*

al. elaborately described diverse factors and components that need to be considered for efficient nucleic acid extractions (Lever et al. 2015). 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 and 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, and Clarke 2003; Vishnivetskaya et al. 2014). While studies have concentrated on nucleic acid extraction from glacial ice cores (Dancer, Shears, and Platt 1997) or surface snow (Pei-Ying et al. 2012), none has 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 and 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 and 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, and Elser 2017; Dancer, Shears, and 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 and Zador 2015; Bowers et al. 2015; Thomas, Gilbert, and Meyer 2012; Chafee, Maignien, and Simmons 2015; Peng et al. 2020). Here, we address the utility and efficiency of the “gold” standard phenol-chloroform extraction (Dairawan and Shetty 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 eu- karyotic 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 *meters 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 Brecchia of 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 manual protocols differing primarily in the lysis step (bead-beating and lysis buffer composition; (Roume et al. 2013; Lever et al. 2015; Sambrook and 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 et al. 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 KPO_4, 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 (Supplementary material).

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 0.1mm 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 1h 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 µg/ml LPA, 0.2 vol 0.5 M NaCl, 2.5 vol ethanol and 2h incubation at RT in the dark (Supplementary material). The input weight of 5 g sediment was a modification from previously established protocol and the subsequent reagent volumes were adjusted accordingly.

154

155 Method-3 has been previously applied successfully on sand and clay soils (Hale and Crowley
156 2015). In this protocol, the standard lysis capacity of the DNeasy PowerMax Soil Kit (Qiagen, Cat.
157 No. 112988-10) was modified and enhanced by the addition of phenol:chloroform:isoamyl alcohol
158 along with PowerBeads (kit provided) and C1 solution to 5 g of sediment. Then, the manufacturer-
159 suggested sequence of treatments and rinses with the standard buffers of the kit were followed to
160 reach elution of extracted DNA from silica columns with 6 ml of elution buffer. Further
161 concentration of extracted DNA was carried out with the addition of 240 µl 5M NaCl, 2.5 vol
162 ethanol and 10 µg/ml LPA (Supplementary material). LPA was an additional modification to the
163 original protocol for improved DNA recovery.

164

165 Method-4 involved chemical and enzymatic treatment of samples according (Green and Sambrook
166 2017) with minor modifications. Five g of sample was mixed with 10 ml of lysis buffer,
167 incorporating the SDS as well, (0.1 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8, 1.25 % SDS) and 10
168 µl RNase A (100 mg/ml). Then sediment was vortexed for 15 s and incubated at 37 °C for 1h in a
169 hybridization oven. 100 µl Proteinase K (20 mg/ml) was added in a subsequent step and the mixture
170 was incubated for 10 min at 70 °C. Samples were extracted once with phenol/chloroform/isoamyl
171 alcohol (ratio 25:24:1) and supernatants were extracted subsequently with chloroform/isoamyl
172 alcohol (24:1). More stringent DNA precipitation conditions were applied with the addition of 10
173 µg/ml LPA, and overnight incubation at -20 °C (Supplementary material).

174

175 All DNA extracts were suspended in 100 µl of DNA/DNase–free water (ThermoFisher Scientific).
176 Due to the inadequacy of DNA obtained from Method-1 given the 0.5g input sediment weight, we
177 scaled the extraction to 5 g starting weight prior to sequencing. Extracted DNA was thereafter
178 stored at -20 °C until further use. Due to the low DNA yields, it was necessary to use Qubit dsDNA
179 HS kit (Invitrogen), a fluorescent DNA quantification method with high sensitivity. Quality
180 assessment, with Nanodrop and DNA visualization on 0.8% agarose gel containing GelRed nucleic
181 acid stain, was possible only for DNA extracted with method-4 and for DNA concentrations higher
182 than 0.5 ng/ul. All samples yielded sufficient DNA, i.e. 50 ng (total input), for metagenomic
183 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 2x150 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. 12/2016). 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 (D. Li et al. 2015) . Default IMP parameters were retained for all samples. Subsequently, we used MetaBAT2 (Kang et al. 2019) and MaxBin2 (Wu, Simmons, and 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 (H. 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 and Konstantinidis 2014).

215

216 **Taxonomic classification for metagenomic operational taxonomic units**

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

227

228 **Data analysis**

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

235

236 **Results**

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

238 To ensure native sequencing, by minimizing the number of PCR (polymerase chain reaction)
239 cycles within the library preparation protocols, we tested four protocols for biomolecular
240 extraction, with an aim of acquiring large quantities (>50 ng) DNA from glacier-fed stream benthic
241 sediments. The four methods tested were selected because of their wide applicability on related
242 environmental samples (Method-1 & -2) (Griffiths et al. 2000; Lever et al. 2015; Tatti et al. 2016)
243 and their improved chances of higher yields (Method-3; Qiagen communication). Since method-4
244 is considered the gold standard of DNA extraction in biomedical sciences (Dairawan and Shetty
245 2020) and bacterial cultures (Green and Sambrook 2017), it was included in our study. The four

246 protocols are largely based on the same principles, *viz.* sample preparation, cell lysis, purification,
247 precipitation and washing (Table 1). From preliminary tests, it became apparent that a small-scale
248 approach (Method-1; 0.5 g input sediment) did not yield sufficient amounts of DNA for
249 metagenomics due to, on average, limited microbial biomass in the samples. Thus, all protocols
250 (aside from Qiagen's - already produced for maxi scale) were scaled up to 5 g of input sediment
251 and a co-precipitant, like linear polyacrylamide, was included in all precipitation steps. This was
252 essential for the quantitative recovery of the small amounts of extracted DNA from high solution
253 volumes (6-10 ml).
254
255 Overall, we found that extractions using the commercial kit from Qiagen (method-3) yielded
256 increased total DNA as compared to a commonly used protocol (method-1; Fig. 1A). Furthermore,
257 method-3 was similar in terms of DNA yield when compared to a generalized protocol (method-
258 2) previously proposed (Lever et al. 2015) (Fig. 1B). On the other hand, the phenol-chloroform
259 based extraction protocol (method-4) was tested against both methods 2 and 3, using sediment
260 samples collected from the three different glacier floodplain streams (CBS, FEU, FED) from
261 Switzerland. Method-1 was omitted from these tests due to insufficient DNA concentrations in the
262 preliminary extractions. We found that for all three GFS, the phenol-chloroform extraction yielded
263 the highest DNA concentrations. In some cases, and notably samples with low cell abundance, we
264 even obtained one order of magnitude more DNA (Fig. 1C).
265
266 Quality assessment of these DNA extracts with Nanodrop showed $OD_{260/280}$ ratios between ~1.4
267 and ~1.6. Agarose gel electrophoresis revealed a high-molecular weight band with no apparent
268 shearing, smearing or residual RNA, indicative of high-quality DNA (Fig. 2A). A secondary effect
269 appearing in certain samples, but without any perceived consequences in the quality of extracted
270 DNA whatsoever, was the development of a pink-red color of varying intensities with the addition
271 of phenol:chloroform:isoamyl alcohol (Fig. 2B). This was pH dependent since samples were
272 decolorized with the addition of sodium acetate pH 5.2 in the precipitation step. This could
273 possibly be due to a ferric-chloride-phenol compound formed when chloride and phenol
274 constituents of the protocol interact reversibly with Fe^{+3} ions contained in certain samples
275 depending on local geology (Banerjee and Haldar 1950). Similar coloration has been previously
276 reported (Lever et al. 2015).

277

278 **Extraction method affects library preparation efficiency**

279 The DNA extractions based on method-3 and using phenol-chloroform methods were subsequently
280 subjected to library preparation for high-throughput whole genome shotgun sequencing. Despite
281 similar quality of DNA across both methods (~1.4-1.6 $OD_{260/280}$), library preparation using the
282 modified commercial kit did not yield any successful libraries (Fig. 3). To assess if any impurities
283 or inhibitors hampered library preparation we tested two clean-up methods for the DNA extracted
284 from the commercial kit: 1) ethanol precipitation and 2) magnetic-bead based clean-up. For
285 magnetic bead clean-up the SPRIselect beads (Beckman Coulter, 23318) were used according to
286 the manufacturer's protocol. We found that the magnetic-bead method leads to a complete loss of
287 sample (i.e., undetectable DNA quantity via Qubit analyses) during the process, especially if
288 starting with a low input DNA concentration. Although we lost six out of twelve samples using
289 the magnetic-bead clean-up, we achieved 100% efficiency as indicated by a concentration of
290 greater than 0.5 ng/ul after library preparation quantified by Qubit, with the remaining six samples.
291 On the other hand, ~20% of the samples cleaned via ethanol precipitation failed library preparation.
292 Contradictory to these methods, DNA extracted using the phenol-chloroform based method
293 (method-4) yielded 100% efficiency in terms of library preparation without any additional clean-
294 up (Fig. 3). Additionally, we found that the distribution of the total yield after library preparation
295 using the phenol-chloroform method was more uniform across samples compared to the other
296 methods (Fig. 3).

297

298 **Whole genome shotgun assembly unaffected by extraction methods**

299 Extraction methods for whole genome shotgun sequencing may affect the sequencing itself,
300 including the quality and assembly of the reads downstream. To assess this, we used the libraries
301 prepared as described above (Fig. 3), and performed whole genome shotgun sequencing on an
302 Illumina NextSeq500. The average quality across all three methods based on short-read sequencing
303 was Q30 after trimming the leading and trailing sequences (described in Methods). We assessed
304 several assembly metrics including the average length of contigs (N50), largest alignment, total
305 aligned length and coverage. We did not find any significant differences among any of these
306 measures across all three methods (Fig. 4A-C, 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 (Wagner Mackenzie, Waite, and Taylor 2015) and library preparation (Bowers 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, 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 (Supplementary fig.1). 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 and Casadevall 2017) affecting lysis and subsequent extraction.

338

339 **Discussion**

340 Improved omic techniques not limited to metagenomics are robust methods for analyzing nucleic
341 acids and the characterisation of microbial communities in various environments (Jansson et al.
342 2012). One way of understanding the impacts of global climate change on GFS includes the
343 establishment of their census of microbial life (Milner et al. 2017). However, methods designed
344 for the extraction of biomolecules including DNA have not been validated for GFS sediments.
345 Although previous glacier-fed streams studies successfully used extracted DNA for 16S rRNA
346 amplicon sequencing (Ren et al. 2017; Ren, Gao, and Elser 2017; Vardhan Reddy et al. 2009;
347 Wilhelm et al. 2013) the input DNA concentration requirements are considerably higher for whole
348 genome shotgun sequencing. In order to pursue a deeper characterisation of the microbial
349 communities within the GFS sediments, increased concentrations of DNA may additionally
350 alleviate PCR biases (Brooks et al. 2015; Kim and Bae 2011). Also, as previously highlighted,
351 several methods exist for extractions from a wide variety of environmental samples, but not for
352 GFS sediments. Here, we systematically tested the utility of four extraction protocols to identify a
353 ubiquitous methodology. We found that a phenol-chloroform based extraction protocol can be
354 used for samples across geographical separations, differences in bedrock, and samples collected at
355 various distances from the glacier.

356

357 Glassing *et al*. demonstrated that inherent DNA contamination may influence microbiota
358 interpretation in low biomass samples (Glassing et al. 2016). Additionally, it is known that certain
359 compounds - polysaccharides, humic acids, may affect PCR reactions (Rådström et al. 2004),
360 requiring the need for additional DNA clean-up. It has been established that DNA losses occur
361 during the purification steps (Roose-Amsaleg, Garnier-Sillam, and Harry 2001), including when
362 using commercial column methods (Howeler, Ghiorse, and Walker 2003; Lloyd et al. 2013), and
363 phenol-chloroform (Ogram, Sayler, and Barkay 1987). Interestingly, we found similar losses when
364 using the magnetic bead clean-up, whereas the ethanol precipitation method was inefficient
365 compared to the phenol-chloroform protocol. Though the kit-based methods are more convenient
366 and safer than phenol-chloroform extractions (Tesena et al. 2017), access to reagents and costs
367 may be a considerable factor. On the other hand, isolation of the aqueous phase from phenol-
368 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 (Wagner Mackenzie, Waite, and Taylor 2015; Carrigg et al. 2007), we found considerable variability when compared to the phenol-chloroform. This is plausible due to the incomplete dissolution of DNA in buffers, especially when using methods involving charged minerals (Vorhies and Gaines 2009; Barton et al. 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.

Data Availability

The sequencing data generated during the current study are available from NCBI under the BioProject accession number PRJNA624048. A reporting summary for the uploaded data has been included as a metadata file at the accession listed ID. All extraction protocols including the modified commercial methods are available in the *Supplementary Materials*.

Acknowledgments

We are grateful to Laura de Nies, Camille-Martin Gallausiaux, Jean-Pierre Trezzi, Cedric Laczny, Audrey Frachet, Lea Grandmougin, Annegrat Daujemont, Laura Lebrun (LCSB) for discussions and laboratory support. The University of Luxembourg Sequencing Platform and HPC facilities were highly instrumental for the *in-silico* analyses. The present work was supported by NOMIS Foundation and the Swiss National Science Foundation (CRSII5_180241) to Tom J. Battin.

400

401 **Contributions**

402 P.P. performed the biomolecular extractions, including the validation of methods 1-3 alongside
403 quality analyses and quantification. S.B.B. curated and validated the phenol-chloroform extraction
404 method and whole genome shotgun sequencing analyses. T.B. and H.P. collected the glacier-fed
405 stream samples for the experiments. S.F. did the DNA extractions, quantification and
406 qualifications alongside P.P. R.H. handled the library preparation for all samples and the
407 subsequent sequencing. P.W. contributed significantly to the development of method-1 in the
408 manuscript. S.B.B., P.P., S.F., H.P., P.W., and T.J.B. conceived and formulated the experiments.
409 S.B.B. and P.P. developed the manuscript with equal contributions from all authors.

410

411 **Ethical declarations**

412 **Conflicts of interest**

413 The authors do not have any competing interests.

414

415 **Figure legends**

416 *Figure 1. Total DNA concentrations using different extraction protocols*

417 Boxplots represent the total amount of DNA (ng) extracted from 5 g of sediment when comparing
418 (A) method-1 versus the modified-commercial kit-based method-3 and (B) method-2 versus
419 method-3. (C) Boxplots of the DNA quantities isolated from three glacial floodplains (CBS -
420 Corbassière, FEU - Val Ferret up site, FED - Val Ferret down site), using method -2, -3 and -4.
421 Method-1: CTAB buffer lysis (Griffiths *et al.* 2000), Method-2: Modular DNA extraction (Lever
422 *et al.* 2015), Method-3: Qiagen PowerMax Soil DNA extraction kit, Method-4: Chemical and
423 enzymatic extraction. Significance was tested using a Two-Way ANOVA with Student-Neuman
424 Keul's post-hoc analyses. $**p<0.01$, $***p<0.001$

425

426 *Figure 2. Characteristics of DNA extracted with method-4*

427 (A) Agarose gel electrophoresis of DNA extracted with mild vortexing of sediments and
428 incubation in lysis buffer, proteinase K treatment and phenol-chloroform extraction. Lane 1:
429 GeneRuler 1 kb DNA ladder; lanes 2-4: CBS, FED, FEU respectively. (B) Pink-red supernatants
430 developed during phenol:chloroform extraction step.

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.

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.

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.

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.

Supplementary figure 1.

Supplementary figure 1. Relative abundance of Ralstonia sp. AU12-08

The abundance of the *Ralstonia* genome recovered from the samples when processed with method-3 (EtOH and magnetic bead clean-up) and method-4. Significance was tested using One-Way ANOVA with Student-Neuman Keul's post-hoc analyses. ***$p<0.001$, ****$p<0.0001$

References

Banerjee, S., and B. C. Haldar. 1950. "Constitution of Ferri-Phenol Complex in Solution." *Nature* 165 (4208): 1012.

Barton, H. A., N. M. Taylor, B. R. Lubbers, and A. C. Pemberton. 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.

Besemer, Katharina, Hannes Peter, Jürg B. Logue, Silke Langenheder, Eva S. Lindström, Lars J. Tranvik, and Tom J. Battin. 2012. "Unraveling Assembly of Stream Biofilm Communities." *The ISME Journal* 6 (8): 1459–68.

Bogen, Jim. 1988. "Glacial Sediment Production and Development of Hydro-Electric Power in Glacierized Areas" 13 (January). https://doi.org/10.1017/S0260305500007539.

Bowers, Robert M., Alicia Clum, Hope Tice, Joanne Lim, Kanwar Singh, Doina Ciobanu, Chew Yee Ngan, Jan-Fang Cheng, Susannah G. Tringe, and Tanja Woyke. 2015. "Impact of Library Preparation Protocols and Template Quantity on the Metagenomic Reconstruction of a Mock Microbial Community." *BMC Genomics* 16 (October): 856.

Brooks, J. Paul, David J. Edwards, Michael D. Harwich Jr, Maria C. Rivera, Jennifer M. Fettweis, Myrna G. Serrano, Robert A. Reris, Nihar U Sheth, Bernice Huang, Phillippe Girerd, Vaginal Microbiome Consortium, Jerome F Strauss III, Kimberly K Jefferson and Gregory A Buck. 2015. "The Truth about Metagenomics: Quantifying and Counteracting Bias in 16S rRNA Studies." *BMC Microbiology* 15 (March): 66.

Carrigg, Cora, Olivia Rice, Siobhán Kavanagh, Gavin Collins, and Vincent O'Flaherty. 2007. "DNA Extraction Method Affects Microbial Community Profiles from Soils and Sediment." *Applied Microbiology and Biotechnology* 77 (4): 955–64.

Chafee, Meghan, Loïs Maignien, and Sheri L. Simmons. 2015. "The Effects of Variable Sample Biomass on Comparative Metagenomics." *Environmental Microbiology* 17 (7): 2239–53.

Chanudet, Vincent, and Montserrat Filella. 2008. "Size and Composition of Inorganic Colloids in a Peri-Alpine, Glacial Flour-Rich Lake." *Geochimica et Cosmochimica Acta* 72 (5): 1466–79.

Claassen, Shantelle, Elloise du Toit, Mamadou Kaba, Clinton Moodley, Heather J. Zar, and Mark P. Nicol. 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): 103–10.

Dahm, Ralf. 2008. "Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research." *Human Genetics* 122 (6): 565–81.

Dairawan, Mariyam, and Preetha J. Shetty. 2020. "The Evolution of DNA Extraction Methods." *American Journal of Biomedical Science & Research* 8 (1). https://doi.org/10.34297/AJBSR.2020.08.001234.

Dancer, S. J., P. Shears, and D. J. Platt. 1997. "Isolation and Characterization of Coliforms from Glacial Ice and Water in Canada's High Arctic." *Journal of Applied Microbiology* 82 (5): 597–609.

Glassing, Angela, Scot E. Dowd, Susan Galandiuk, Brian Davis, and Rodrick J. Chiodini. 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.
Godone, Danilo. 2017. *Glacier Evolution in a Changing World*. BoD – Books on Demand.
Green, Michael R., and Joseph Sambrook. 2017. "Isolating DNA from Gram-Negative Bacteria." *Cold Spring Harbor Protocols* 2017 (1). https://doi.org/10.1101/pdb.prot093369.
Griffiths, R. I., A. S. Whiteley, A. G. O'Donnell, and M. J. Bailey. 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): 5488–91.
Grossman, Nina T., and Arturo Casadevall. 2017. "Physiological Differences in Cryptococcus Neoformans Strains In Vitro versus In Vivo and Their Effects on Antifungal Susceptibility." *Antimicrobial Agents and Chemotherapy* 61 (3). https://doi.org/10.1128/AAC.02108-16.
Hale, Lauren, and David Crowley. 2015. "DNA Extraction Methodology for Biochar-Amended Sand and Clay." *Biology and Fertility of Soils* 51 (6): 733–38.
Heintz-Buschart, Anna, Patrick May, Cédric C. Laczny, Laura A. Lebrun, Camille Bellora, Abhimanyu Krishna, Linda Wampach, Jochen G Schneider, Angela Hogan, Carine de Beaufort, and Paul Wilmes. 2017. "Integrated Multi-Omics of the Human Gut Microbiome in a Case Study of Familial Type 1 Diabetes." *Nature Microbiology* 2 (1): 16180.
Hotaling, Scott, Eran Hood, and Trinity L. Hamilton. 2017. "Microbial Ecology of Mountain Glacier Ecosystems: Biodiversity, Ecological Connections and Implications of a Warming Climate." *Environmental Microbiology* 19 (8): 2935–48.
Howeler, Michael, William C. Ghiorse, and Larry P. Walker. 2003. "A Quantitative Analysis of DNA Extraction and Purification from Compost." *Journal of Microbiological Methods* 54 (1): 37–45.
Jansson, Janet K., and Kirsten S. Hofmockel. 2018. "The Soil Microbiome—from Metagenomics to Metaphenomics." *Current Opinion in Microbiology*, Environmental Microbiology * The New Microscopy, 43 (June): 162–68.
Jansson, Janet K., Josh D. Neufeld, Mary Ann Moran, and Jack A. Gilbert. 2012. "Omics for Understanding Microbial Functional Dynamics." *Environmental Microbiology* 14 (1): 1–3.
Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies." *PeerJ* 7 (July): e7359.
Kebschull, Justus M., and Anthony M. Zador. 2015. "Sources of PCR-Induced Distortions in High-Throughput Sequencing Data Sets." *Nucleic Acids Research* 43 (21): e143.
Kim, Kyoung-Ho, and Jin-Woo Bae. 2011. "Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses." *Applied and Environmental Microbiology* 77 (21): 7663–68.
Laczny, Cedric C., Tomasz Sternal, Valentin Plugaru, Piotr Gawron, Arash Atashpendar, Houry Hera Margossian, Sergio Coronado, Laurens van der Maaten, Nikos Vlassis, and Paul Wilmes. 2015. "VizBin - an Application for Reference-Independent Visualization and Human-Augmented Binning of Metagenomic Data." *Microbiome* 3 (1): 1.
Lever, Mark A., Andrea Torti, Philip Eickenbusch, Alexander B. Michaud, Tina Šantl-Temkiv, and Bo Barker Jørgensen. 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.
Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph." *Bioinformatics* 31 (10): 1674–76.
Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.

559 Lloyd, Karen G., Lars Schreiber, Dorthe G. Petersen, Kasper U. Kjeldsen, Mark A. Lever,
560 Andrew D. Steen, Ramunas Stepanauskas, Michael Richter, Sara Kleindienst, Sabine
561 Lennk, Andreas Schramm and Bo Barker Jorgenson. 2013. “Predominant Archaea in
562 Marine Sediments Degrade Detrital Proteins.” *Nature* 496 (7444): 215–18.
563 Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold
564 Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12).
565 https://doi.org/10.1186/s13059-014-0550-8.
566 Milanese, Alessio, Daniel R. Mende, Lucas Paoli, Guillem Salazar, Hans-Joachim Ruscheweyh,
567 Miguelangel Cuenca, Pascal Hingamp, Renato Alves, Paul I Costea, Luis Pedro Coelho,
568 Thomas S. B. Schmidt, Alexandre Almeida, Alex L Mitchell, Robert D. Finn, Jaime Huerta-
569 Cepas, Peer Bork, Georg Zeller and Shinichi Sunagawa. 2019. “Microbial Abundance,
570 Activity and Population Genomic Profiling with mOTUs2.” *Nature Communications* 10 (1):
571 1014.
572 Miller, D. N., J. E. Bryant, E. L. Madsen, and W. C. Ghiorse. 1999. “Evaluation and Optimization
573 of DNA Extraction and Purification Procedures for Soil and Sediment Samples.” *Applied
574 and Environmental Microbiology* 65 (11): 4715–24.
575 Milner, Alexander M., Kieran Khamis, Tom J. Battin, John E. Brittain, Nicholas E. Barrand,
576 Leopold Füreder, Sophie Cauvy-Fraunié, Gísli Már Gíslason, Dean Jacobsen, David M
577 Hannah, Andrew J Hodson, Eran Hood, Valeria Lencioni, Jón S Ólafsson, Christopher T
578 Robinson, Martyn Tranter, Lee E Brown. 2017. “Glacier Shrinkage Driving Global Changes
579 in Downstream Systems.” *Proceedings of the National Academy of Sciences of the United
580 States of America* 114 (37): 9770–78.
581 Mukhopadhyay, T., and J. A. Roth. 1993. “Silicone Lubricant Enhances Recovery of Nucleic
582 Acids after Phenol-Chloroform Extraction.” *Nucleic Acids Research* 21 (3): 781–82.
583 Narayanasamy, Shaman, Yohan Jarosz, Emilie E. L. Muller, Anna Heintz-Buschart, Malte
584 Herold, Anne Kaysen, Cédric C. Laczny, Nicolás Pinel, Patrick May, and Paul Wilmes.
585 12/2016. “IMP: A Pipeline for Reproducible Reference-Independent Integrated
586 Metagenomic and Metatranscriptomic Analyses.” *Genome Biology* 17 (1): 260.
587 Nielsen, Jens, and Boyang Ji. 2015. “New Insight into the Gut Microbiome through
588 Metagenomics.” *Advances in Genomics and Genetics*. https://doi.org/10.2147/agg.s57215.
589 Ogram, Andrew, Gary S. Sayler, and Tamar Barkay. 1987. “The Extraction and Purification of
590 Microbial DNA from Sediments.” *Journal of Microbiological Methods* 7 (2): 57–66.
591 Parks, Donovan H., Maria Chuvochina, David W. Waite, Christian Rinke, Adam Skarshewski,
592 Pierre-Alain Chaumeil, and Philip Hugenholtz. 2018. “A Proposal for a Standardized
593 Bacterial Taxonomy Based on Genome Phylogeny.” *bioRxiv*, January, 256800.
594 Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W.
595 Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from
596 Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55.
597 Pei-Ying, Yan, Hou Shu-Gui, Chen Tuo, Zhang Shu-Hong, and Sun Wei-Jun. 2012. “Methods
598 for Extraction of Microorganism DNA from Glacier Surface Snow.” *Sciences in Cold and
599 Arid Regions* 4 (6): 484.
600 Peng, Zonghui, Xiaolong Zhu, Zhijiao Wang, Xianting Yan, Guangbiao Wang, Meifang Tang,
601 Awei Jiang, and Karsten Kristiansen. 2020. “Comparative Analysis of Sample Extraction
602 and Library Construction for Shotgun Metagenomics.” *Bioinformatics and Biology Insights*
603 14 (June): 1177932220915459.
604 Peter, Hannes, and Ruben Sommaruga. 2017. “Alpine Glacier-Fed Turbid Lakes Are
605 Discontinuous Cold Polymictic rather than Dimictic.” *Inland Waters : Journal of the
606 International Society of Limnology* 7 (1): 45–54.
607 Porebski, Sue, L. Grant Bailey, and Bernard R. Baum. 1997. “Modification of a CTAB DNA
608 Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol
609 Components.” *Plant Molecular Biology Reporter / ISPMB* 15 (1): 8–15.

610 Psifidi, Androniki, Chrysostomos I. Dovas, Georgios Bramis, Thomai Lazou, Claire L. Russel,
611 Georgios Arsenos, and Georgios Banos. 2015. “Comparison of Eleven Methods for
612 Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-
613 Term DNA Banking Using Blood Samples.” *PloS One* 10 (1): e0115960.
614 Rådström, Peter, Rickard Knutsson, Petra Wolffs, Maria Lövenklev, and Charlotta Löfström.
615 2004. “Pre-PCR Processing: Strategies to Generate PCR-Compatible Samples.” *Molecular*
616 *Biotechnology* 26 (2): 133–46.
617 Ren, Ze, Hongkai Gao, and James J. Elser. 2017. “Longitudinal Variation of Microbial
618 Communities in Benthic Biofilms and Association with Hydrological and Physicochemical
619 Conditions in Glacier-Fed Streams.” *Freshwater Science* 36 (3): 479–90.
620 Ren, Ze, Hongkai Gao, James J. Elser, and Qiudong Zhao. 2017. “Microbial Functional Genes
621 Elucidate Environmental Drivers of Biofilm Metabolism in Glacier-Fed Streams.” *Scientific*
622 *Reports* 7 (October). https://doi.org/10.1038/s41598-017-13086-9.
623 Rodriguez-R, Luis M., and Konstantinos T. Konstantinidis. 2014. “Nonpareil: A Redundancy-
624 Based Approach to Assess the Level of Coverage in Metagenomic Datasets.”
625 *Bioinformatics* 30 (5): 629–35.
626 Roose-Amsaleg, C. L., E. Garnier-Sillam, and M. Harry. 2001. “Extraction and Purification of
627 Microbial DNA from Soil and Sediment Samples.” *Applied Soil Ecology: A Section of*
628 *Agriculture, Ecosystems & Environment* 18 (1): 47–60.
629 Roume, Hugo, Anna Heintz-Buschart, Emilie E. L. Muller, and Paul Wilmes. 2013. “Sequential
630 Isolation of Metabolites, RNA, DNA, and Proteins from the Same Unique Sample.” *Methods*
631 *in Enzymology* 531: 219–36.
632 Sambrook, Joseph, and David W. Russell. 2006. “Purification of Nucleic Acids by Extraction
633 with Phenol:chloroform.” *CSH Protocols* 2006 (1). https://doi.org/10.1101/pdb.prot4455.
634 Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess,
635 Susannah G. Tringe, and Jillian F. Banfield. 2018. “Recovery of Genomes from
636 Metagenomes via a Dereplication, Aggregation and Scoring Strategy.” *Nature Microbiology*
637 3 (7): 836–43.
638 Smith, K., M. A. Diggle, and S. C. Clarke. 2003. “Comparison of Commercial DNA Extraction
639 Kits for Extraction of Bacterial Genomic DNA from Whole-Blood Samples.” *Journal of*
640 *Clinical Microbiology* 41 (6): 2440–43.
641 Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie,
642 Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R Mende, Adriana Alberti,
643 Francisco M Cornejo-Castillo, Paul I Costea, Corinne Cruaud, Francesco d'Ovidio, Stefan
644 Engelen, Isabel Ferrera, Josep M Gasol, Lionel Guidi, Falk Hildebrand, Florian Kokoszka,
645 Cyrille Lepoivre, Gipsi Lima-Mendez, Julie Poulain, Bonnie T Poulos, Marta Royo-Llonch,
646 Hugo Sarmento, Sara Vieira-Silva, Céline Dimier, Marc Picheral, Sarah Searso, Stefanie
647 Kandels-Lewis, Tara Oceans coordinators; Chris Bowler, Colomban de Vargas, Gabriel
648 Gorsky, Nigel Grimsley, Pascal Hingamp, Daniele Iudicone, Olivier Jaillon, Fabrice Not,
649 Hiroyuki Ogata, Stephane Pesant, Sabrina Speich, Lars Stemmann, Matthew B Sullivan,
650 Jean Weissenbach, Patrick Wincker, Eric Karsenti, Jeroen Raes, Silvia G Acinas, Peer
651 Bork. 2015. “Ocean Plankton. Structure and Function of the Global Ocean Microbiome.”
652 *Science* 348 (6237): 1261359.
653 Tatti, Enrico, Boyd A. McKew, Corrine Whitby, and Cindy J. Smith. 2016. “Simultaneous DNA-
654 RNA Extraction from Coastal Sediments and Quantification of 16S rRNA Genes and
655 Transcripts by Real-Time PCR.” *Journal of Visualized Experiments: JoVE*, no. 112 (June).
656 https://doi.org/10.3791/54067.
657 Team, R. Core. 2013. “R: A Language and Environment for Statistical Computing.”
658 Tesena, Parichart, Wasamon Korchunjit, Jane Taylor, and Tuempong Wongtawan. 2017.
659 “Comparison of Commercial RNA Extraction Kits and qPCR Master Mixes for Studying
660 Gene Expression in Small Biopsy Tissue Samples from the Equine Gastric Epithelium.”

Journal of Equine Science 28 (4): 135–41.
Thomas, Torsten, Jack Gilbert, and Folker Meyer. 2012. “Metagenomics - a Guide from Sampling to Data Analysis.” *Microbial Informatics and Experimentation* 2 (1): 3.
Toni, Lee S., Anastacia M. Garcia, Danielle A. Jeffrey, Xuan Jiang, Brian L. Stauffer, Shelley D. Miyamoto, and Carmen C. Sucharov. 2018. “Optimization of Phenol-Chloroform RNA Extraction.” *MethodsX* 5 (May): 599–608.
Tully, Benjamin J., Elaina D. Graham, and John F. Heidelberg. 2018. “The Reconstruction of 2,631 Draft Metagenome-Assembled Genomes from the Global Oceans.” *Scientific Data* 5 (January): 170203.
Vardhan Reddy, Puram Vishnu, Singireesu Soma Shiva Nageswara Rao, Mambatta Shankaranarayanan Pratibha, Buddhi Sailaja, Bakka Kavya, Ravoori Ruth Manorama, Shiv Mohan Singh, Tanuku Naga Radha Srinivas, and Sisinthy Shivaji. 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): 538–46.
Vishnivetskaya, Tatiana A., Alice C. Layton, Maggie C. Y. Lau, Archana Chauhan, Karen R. Cheng, Arthur J. Meyers, Jasity R. Murphy, Alexandra W Rogers, Geetha S Saarunya, Daniel E Williams, Susan M Pfiffner, John P Biggerstaff, Brandon T Stackhouse, Tommy J Phelps, Lyle Whyte, Gary S Sayler, and Tullis C Onstott. 2014. “Commercial DNA Extraction Kits Impact Observed Microbial Community Composition in Permafrost Samples.” *FEMS Microbiology Ecology* 87 (1): 217–30.
Vorhies, John S., and Robert R. Gaines. 2009. “Microbial Dissolution of Clay Minerals as a Source of Iron and Silica in Marine Sediments.” *Nature Geoscience* 2 (3): 221–25.
Wagner Mackenzie, Brett, David W. Waite, and Michael W. Taylor. 2015. “Evaluating Variation in Human Gut Microbiota Profiles due to DNA Extraction Method and Inter-Subject Differences.” *Frontiers in Microbiology* 6 (February): 130.
Wilhelm, Linda, Katharina Besemer, Christina Fasching, Tim Urich, Gabriel A. Singer, Christopher Quince, and Tom J. Battin. 2014. “Rare but Active Taxa Contribute to Community Dynamics of Benthic Biofilms in Glacier-Fed Streams.” *Environmental Microbiology* 16 (8): 2514–24.
Wilhelm, Linda, Gabriel A. Singer, Christina Fasching, Tom J. Battin, and Katharina Besemer. 2013. “Microbial Biodiversity in Glacier-Fed Streams.” *The ISME Journal* 7 (8): 1651–60.
Wu, Yu-Wei, Blake A. Simmons, and Steven W. Singer. 2016. “MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets.” *Bioinformatics* 32 (4): 605–7.
Xin, Zhanguo, and Junping Chen. 2012. “A High Throughput DNA Extraction Method with High Yield and Quality.” *Plant Methods*. https://doi.org/10.1186/1746-4811-8-26.
Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. “DNA Recovery from Soils of Diverse Composition.” *Applied and Environmental Microbiology* 62 (2): 316–22.

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 *et al.* 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. $^{**}p<0.01$, $^{***}p<0.001$

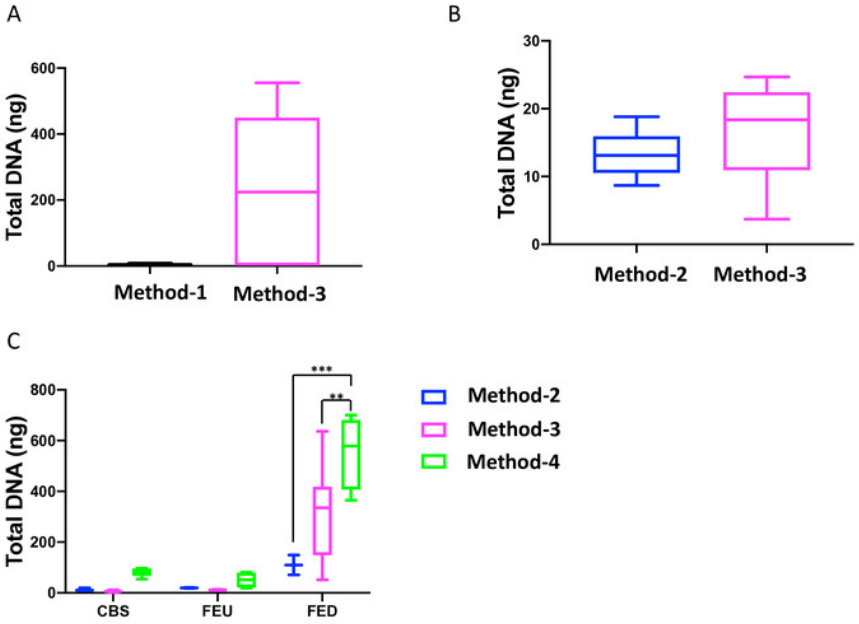
A
Total DNA (ng)
600
400
200
0
Method-1
Method-3
B
Total DNA (ng)
30
20
10
0
Method-2
Method-3
C
Total DNA (ng)
800
600
400
200
0
***
**
CBS
FEU
FED
Method-2
Method-3
Method-4

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.

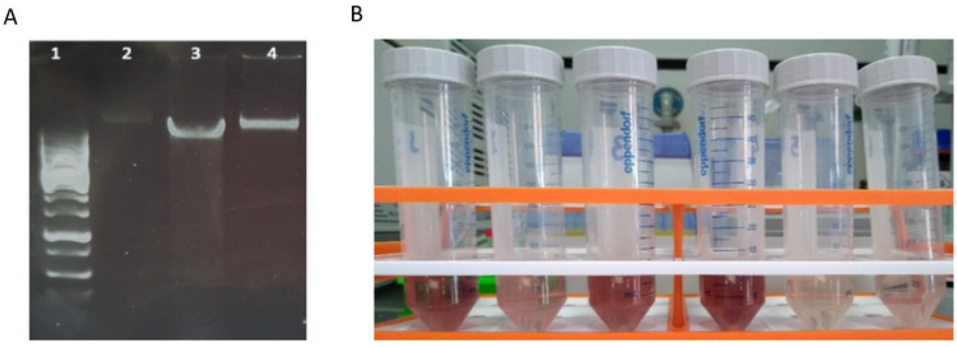

Figure 3

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.

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

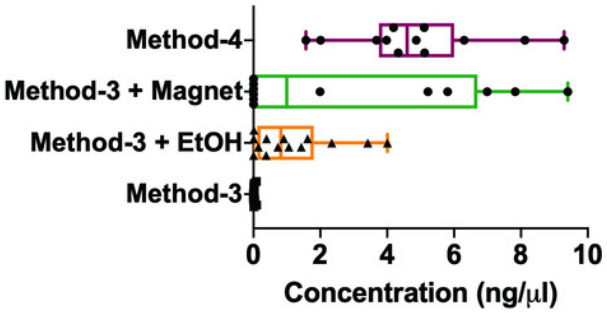

Figure 4

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.

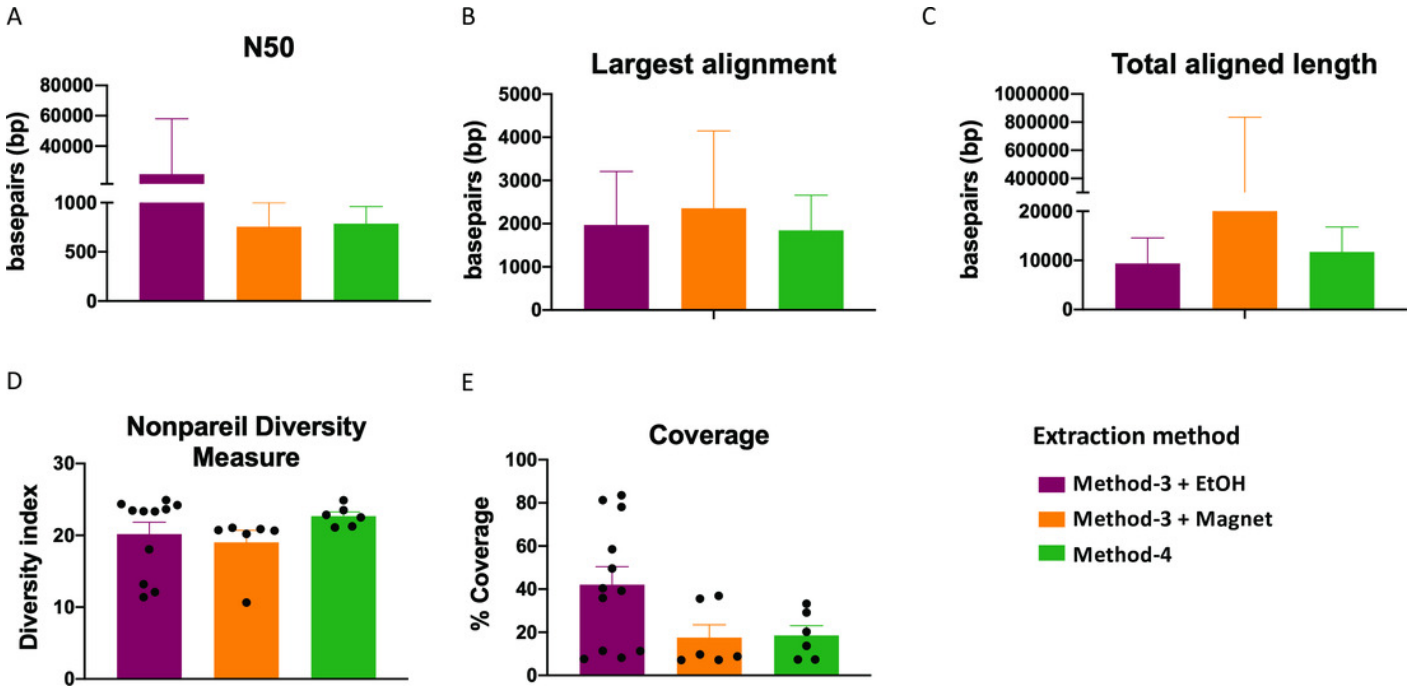

Figure 5

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.

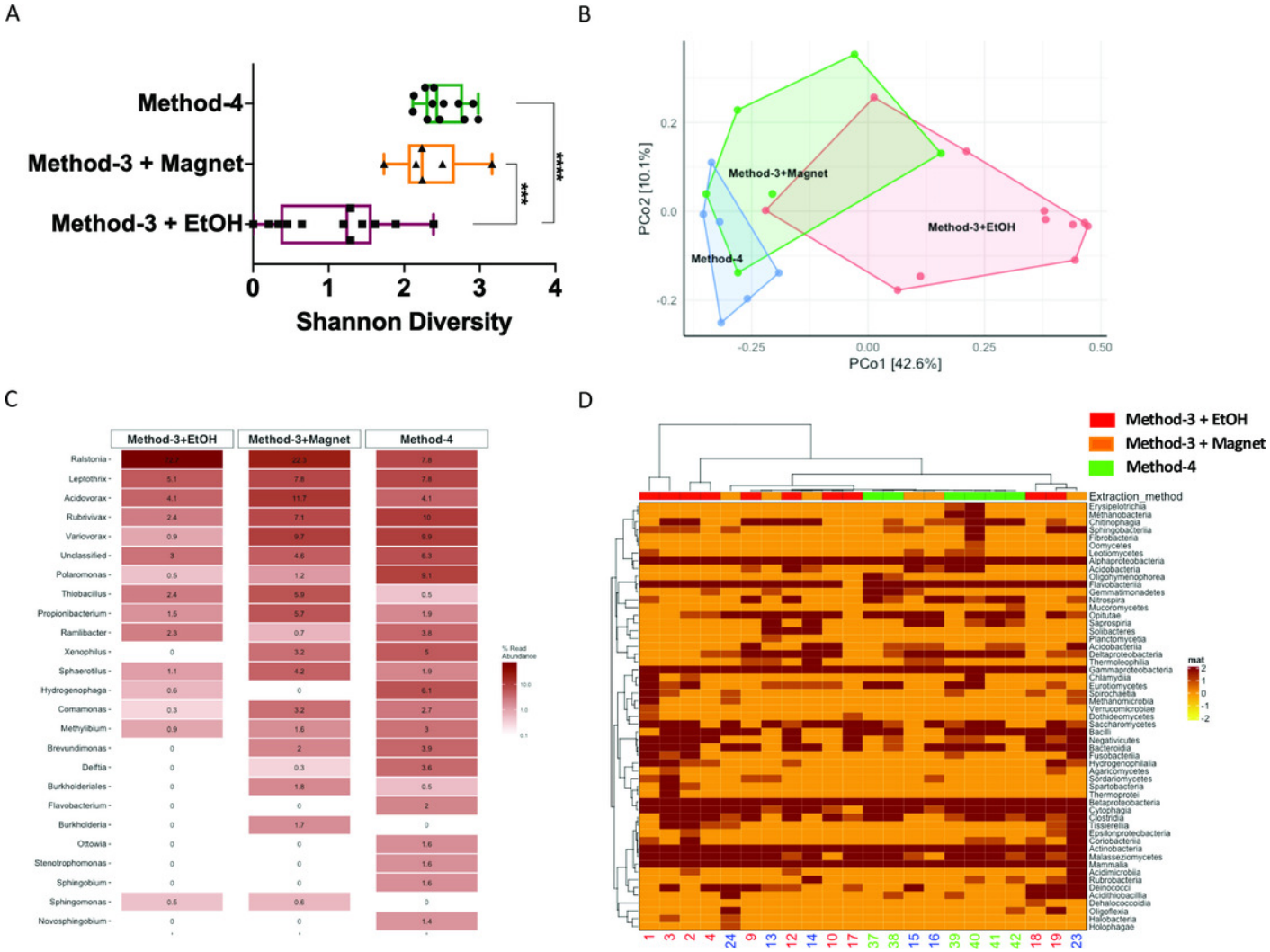
A
Method-4
Method-3 + Magnet
Method-3 + EtOH
***
****
0
1
2
3
4
Shannon Diversity
B
Method-3+Magnet
Method-3+EtOH
Method-4
PCo2 [10.1%]
PCo1 [42.6%]
0.2
0.0
-0.2
-0.25
0.00
0.25
0.50
C
Method-3+EtOH
Method-3+Magnet
Method-4
Ralstonia
Leptothrix
Acidovorax
Rubrivivax
Variovorax
Unclassified
Polaromonas
Thiobacillus
Propionibacterium
Ramlibacter
Xenophilus
Sphaerotilus
Hydrogenophaga
Comamonas
Methylibium
Brevundimonas
Delftia
Burkholderiales
Flavobacterium
Burkholderia
Ottowia
Stenotrophomonas
Sphingobium
Sphingomonas
Novosphingobium
% Read Abundance
D
Method-3 + EtOH
Method-3 + Magnet
Method-4
Extraction_method
mat
Erysipelotrichia
Methanobacteria
Chitinophagia
Sphingobacteriia
Fibrobacteria
Oomycetes
Leotiomycetes
Alphaproteobacteria
Acidobacteria
Oligohymenophorea
Flavobacteria
Gemmatimonadetes
Nitrospira
Mucoromycetes
Opitutae
Saprospiria
Solibacteres
Planctomycetia
Acidobacteriia
Deltaproteobacteria
Thermoleophilia
Gammaproteobacteria
Chlamydiia
Eurotiomycetes
Spirochaetia
Methanomicrobia
Verrucomicrobiae
Dothideomycetes
Saccharomycetes
Bacilli
Negativicutes
Bacteroidia
Fusobacteriia
Hydrogenophilalia
Agaricomycetes
Sordariomycetes
Spartobacteria
Thermoprotei
Betaproteobacteria
Cytophagia
Clostridia
Tissierellia
Epsilonproteobacteria
Coriobacteriia
Actinobacteria
Malassezziomycetes
Mammalia
Acidimicrobiia
Rubrobacteria
Deinococci
Acidithiobacillia
Dehalococcoidia
Oligoflexia
Halobacteria
Holophagae
1 3 2 4 24 9 13 12 14 10 17 37 38 15 16 39 40 41 42 18 19 23

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.

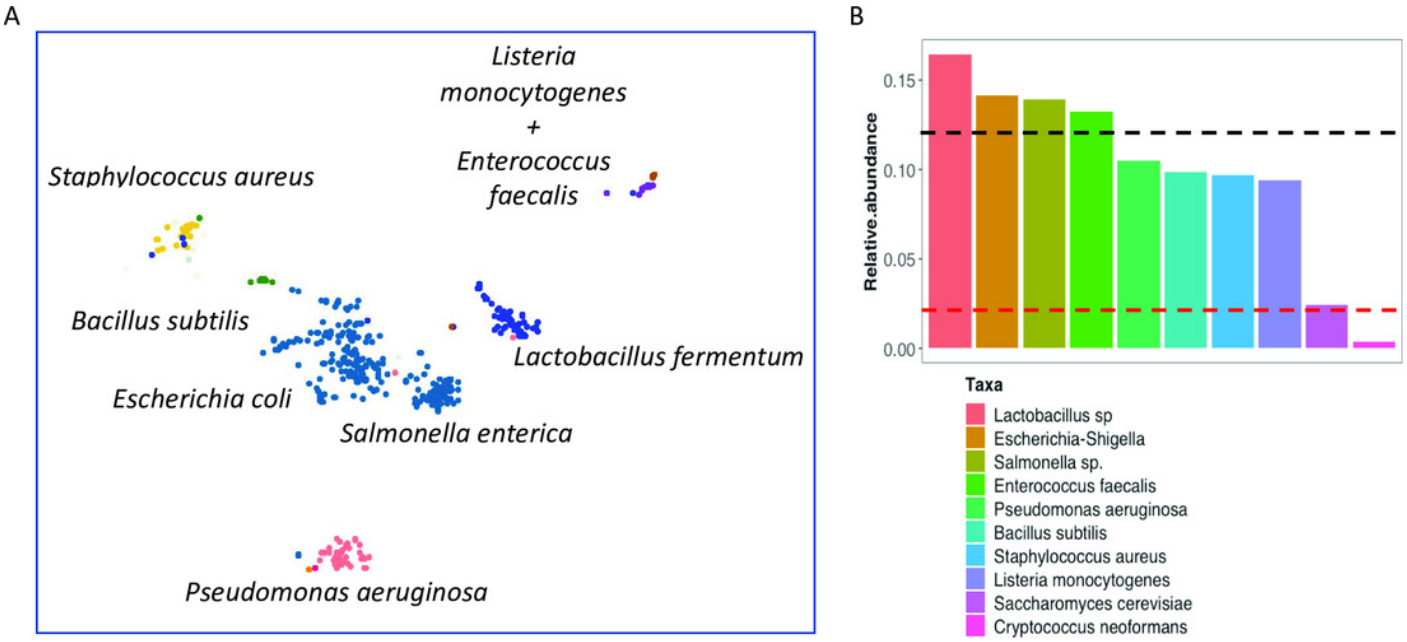

Table 1(on next page)

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)	0.1mm Zirconium beads dNTP solution	PowerBead tubes (carnet 0.7mm)	-
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

1