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

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

[1] Eco-Systems Biology 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 ecology, 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.

1 **Optimised biomolecular extraction for metagenomic analysis of microbial biofilms from**

2 **high-mountain streams**

3 Susheel Bhanu Busi[1,#,*], Paraskevi Pramateftaki[2,#], Jade Brandani[2], Stylianos Fodelianakis[2],

4 Hannes Peter[2], Rashi Halder[1], Paul Wilmes[1] and Tom J. Battin[2]

6 [1]Eco-Systems Biology group, Luxembourg Centre for Systems Biomedicine, University of

7 Luxembourg, Esch-sur-Alzette, Luxembourg

8 [2]Stream Biofilm and Ecosystems Research group, Ecole Polytechnique Fédérale de Lausanne,

9 Lausanne, Switzerland

11 [#]*Co-first authors*

13 [*]*Corresponding author: Susheel Bhanu Busi*

14 *Address: 7 avenue des hauts-fourneaux, BioTech-I, Luxembourg Centre for Systems Biomedicine,*

15 *University of Luxembourg, Esch-sur-Alzette, L-4362 Luxembourg*

16 *Email: susheel.busi@uni.lu*

18 **Abstract**

19 Glacier-fed streams (GFS) are harsh ecosystems dominated by microbial life organized in benthic

20 biofilms, yet the biodiversity and ecosystem functions provided by these communities remain

21 under-appreciated. To better understand the microbial processes and communities contributing to

22 GFS ecology, it is necessary to leverage high throughput sequencing. Low biomass and high

23 inorganic particle load in GFS sediment samples may affect nucleic acid extraction efficiency

24 using extraction methods tailored to other extreme environments such as deep-sea sediments. Here,

25 we benchmarked the utility and efficacy of four extraction protocols, including an up-scaled

26 phenol-chloroform protocol. We found that established protocols for comparable sample types

27 consistently failed to yield sufficient high-quality DNA, delineating the extreme character of GFS.

28 The methods differed in the success of downstream applications such as library preparation and

29 sequencing. An adapted phenol-chloroform-based extraction method resulted in higher yields and

30 better recovered the expected taxonomic profile and abundance of reconstructed genomes when

31 compared to commercially-available methods. Affordable and straight-forward, this method

32 consistently recapitulated the abundance and genomes of a "mock" community, including
33 eukaryotes. Moreover, by increasing the amount of input sediment, the protocol is readily
34 adjustable to the microbial load of the processed samples without compromising protocol
35 efficiency. Our study provides a first systematic and extensive analysis of the different options for
36 extraction of nucleic acids from glacier-fed streams for high-throughput sequencing applications,
37 which may be applied to other extreme environments.

38

39 **Introduction**

40 The advent of high-throughput sequencing technologies has brought hitherto inconceivable
41 capacities to characterize the microbial ecology of both well-studied (Jansson and Hofmockel
42 2018; Nielsen and Ji 2015) and under-explored environments. Among the latter figure high-
43 mountain and particularly glacier-fed streams (Milner et al. 2017) and the microbial biofilms that
44 colonize their beds (Battin et al. 2016). Today, these streams are changing at an unprecedented
45 pace owing to climate change and the thereby shrinking glaciers, and yet little is known of their
46 microbial diversity (Wilhelm et al. 2013, Milner et al. 2017). Glacier-fed stream (GFS) sediments
47 are extreme habitats characterized by low microbial cell abundance and activities but very high
48 loads of fine mineral particles. In order to understand the diversity and composition of these
49 microbial communities, including both eukaryotes and prokaryotes, and the role that they play, it
50 is essential to extract nucleic acids in sufficient quantity and quality from often complex
51 environmental matrices. After extracting the nucleic acids, downstream applications including
52 molecular biology methods such as PCR and next-generation sequencing of amplicons or
53 metagenomes allow for the compositional, functional and phylogenetic characterization of
54 microbial populations and the communities that they form (Roume et al. 2013).

55

56 While there is no lack of protocols and literature pertaining to the extraction of nucleic acids from
57 a wide variety of environments (Roume et al. 2013; Miller et al. 1999; Xin and Chen 2012;
58 Porebski, Bailey, and Baum 1997; Zhou, Bruns, and Tiedje 1996), few reports dwell on the utility
59 of these methods for biomolecular extractions from sedimentary samples with very low cell
60 abundance as typical for GFS (Wilhelm et al. 2013; Ren, Gao, and Elser 2017). In 2015, Lever *et*
61 *al*. elaborately described diverse factors and components that need to be considered for efficient
62 nucleic acid extractions (Lever et al. 2015). These include, but are not limited to key steps like cell

63   lysis, removal of impurities and inhibitors and of critical additives like carrier DNA molecules to
64   enhance aggregation and thus precipitation of DNA in case of very low concentrations. Since the
65   first extraction of DNA by Swiss medical doctor Friedrich Miescher in 1869 (Dahm 2008),
66   biomolecule extractions have shifted from those performed with solutions prepared primarily in
67   the laboratory (Sambrook and Russell 2006; Miller et al. 1999) to using commercially-available
68   kits. These ready-made options are designed to avoid the use of volatile and toxic chemicals such
69   as phenol and chloroform, and are tailored to various environments including blood, faecal
70   material, plant and soils (Claassen et al. 2013; Psifidi et al. 2015; Smith, Diggle, and Clarke 2003;
71   Vishnivetskaya et al. 2014). While studies have concentrated on nucleic acid extraction from
72   glacial ice cores (Dancer, Shears, and Platt 1997) or surface snow (Pei-Ying et al. 2012), none has
73   demonstrated their utility for GFS sediments. Together with low cell abundance, the complex
74   mineral matrices in GFS - a consequence of the erosion activity of glaciers – may affect nucleic
75   acid extraction efficiency. As we attempt to better understand how nature works at its limits
76   through the study of extreme environments, non-commercial approaches and methodologies need
77   to be revisited and optimized.

78

79   In recent years, several research groups (Besemer et al. 2012; Ren et al. 2017; Ren, Gao, and Elser
80   2017; Dancer, Shears, and Platt 1997) have successfully used kit-based methods for DNA
81   extraction and subsequent 16S ribosomal RNA gene amplicon sequencing on GFS samples.
82   However, the requirements for whole genome shotgun sequencing currently include at least 50 ng
83   of input DNA to minimize bias due to PCR reactions during library preparation. Here, we address
84   the utility and efficiency of the "gold" standard phenol-chloroform extraction (Dairawan and
85   Shetty 2020), and three alternative methodologies to identify the process(es) that yield not only
86   the highest quantity but also quality of DNA, from GFS sediments. We found that for the GFS
87   sediment samples, the phenol-chloroform method yielded the expected diversity and taxonomic
88   profiles, while also enabling reconstruction of metagenome-assembled genomes. Using a mock
89   community, we additionally validated the utility of this method for the extraction of nucleic acids
90   from both pro- and eu- karyotic sources. Overall, our findings provide a framework for the
91   extraction of nucleic acids such as DNA for whole genome shotgun sequencing from GFS
92   sediments, whilst highlighting the potential variability introduced due to the isolation method
93   employed.

94

Methods

Sample origin & collection

97 DNA extraction protocols were benchmarked using three different GFS sediments from the Swiss
98 Alps: Corbassière (CBS, 2444 m a.s.l) and Val Ferret (FE, 1995 m a.s.l) at the glacier snout (up
99 site, FEU) and one kilometer further downstream (down site, FED). CBS differs from FEU and
100 FED in terms of bedrock geology, with clastic sedimentary limestone dominating the catchment
101 of CBS and Brecchia of gneiss dominating in FEU and FED. Sediments generally contain more
102 organic material further downstream from the glacier, which may inhibit DNA extraction.
103 Sediments (0.25 to 3.15 mm) were collected using flame-sterilized sampling equipment. Wet
104 sediments were transferred into 10 ml sterile, DNA/DNase-free tubes and immediately flash-
105 frozen in liquid nitrogen in the field. Samples were transferred to the laboratory and kept at -80 °C
106 until analysis. All necessary measures were taken to ensure contamination-free sampling.

107

DNA extraction methods

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

117

118 Method-1 was based on a previously established method (Griffiths et al. 2000). Sample cell lysis
119 was achieved by adding 0.5 g of sediment into a lysing matrix E tube with beads of variable
120 diameter (SKU 116914050), 500 μl CTAB buffer (5% CTAB, 120 mM KPO_4, pH 8.0) and 500 μl
121 of phenol/chloroform/isoamyl alcohol (ratio 25:24:1). Samples were loaded on a Precellys beater
122 for 45 s at 5.500 r/s. DNA was extracted once more with chloroform/isoamyl alcohol (24:1) and

123   precipitated with 2 vol PEG-6000, 15 µg/ml linear polyacrylamide (LPA) and 2 h incubation on

124   ice (Supplementary material).

125

126   Method-2 was an adaptation to alpine stream sediments of the modular method for DNA extraction

127   previously published (Lever et al. 2015). Samples were prepared by mixing 5 g of sediment, 10-

128   20% of 0.1mm zirconium beads and 1 ml of 100 mM dNTP solution. Cell lysis was achieved with

129   5 ml lysis buffer (30 mM Tris-HCl, 30 mM EDTA, 1% Triton X-100, 800 mM guanidium

130   hydrochloride, pH 10.0) and incubation at 50 °C for 1h with gentle agitation. The supernatant was

131   extracted once with chloroform/isoamyl alcohol (24:1) and DNA was precipitated with 10 µg/ml

132   LPA, 0.2 vol 0.5 M NaCl, 2.5 vol ethanol and 2h incubation at RT in the dark (Supplementary

133   material).

134

135   Method-3 has been previously applied successfully on sand and clay soils (Hale and Crowley

136   2015). In this protocol, the standard lysis capacity of the DNeasy PowerMax Soil Kit is

137   supplemented and enhanced by the addition of phenol:chloroform:isoamyl alcohol along with

138   PowerBeads (kit provided) and C1 solution. Then, an elaborate sequence of treatments and rinses

139   with the standard buffers of the kit followed to reach elution of extracted DNA from silica columns

140   with 6 ml of elution buffer. Further concentration of extracted DNA was carried out with the

141   addition of 240 µl 5M NaCl, 2.5 vol ethanol and 10 µg/ml LPA (Supplementary material).

142

143   Method-4 involves chemical and enzymatic treatment of samples (Sambrook and Russell 2006).

144   Five g of sample was mixed with 10 ml of lysis buffer (0.1 M Tris-HCl pH 7.5, 0.05 M EDTA pH

145   8, 1.25 % SDS) and 10 µl RNase A (100 mg/ml) with mild vortexing for 15 s and incubated at 37

146   °C for 1h in a hybridization oven. 100 µl Proteinase K (20 mg/ml) were added and incubation for

147   10 min at 70 °C. Samples were extracted once with phenol/chloroform/isoamyl alcohol (ratio

148   25:24:1) and supernatants were extracted once more with chloroform/isoamyl alcohol (24:1). DNA

149   was precipitated with 10 µg/ml LPA, 1/10 vol 3M Sodium Acetate pH 5.2 and 1 vol isopropanol

150   and overnight incubation at -20 °C (Supplementary material).

151

152   All DNA extracts were suspended in 100 µl of DNA/DNase–free water (ThermoFisher Scientific).

153   Extracted DNA was thereafter stored at -20 °C until further use. Due to the low DNA yields, it was

154   necessary to use a fluorescent DNA quantification method with high sensitivity like the Qubit

155   dsDNA HS kit (Invitrogen). Quality assessment, with Nanodrop and DNA visualization on 0.8%

156   agarose gel containing GelRed nucleic acid stain, was possible only for DNA extracted with

157   method-4 and for DNA concentrations higher than 2 ng/µl. All samples yielded sufficient DNA

158   for metagenomic sequencing and subsequent analyses.

159

160   **DNA sequencing**

161   All DNA samples were subjected to random shotgun sequencing. The sequencing libraries were

162   prepared using the NEBNext Ultra II FS DNA Library Prep kit for Illumina (Cat.No. E7805) using

163   the protocol provided with the kit. The libraries were prepared considering 350 basepairs (bp)

164   average insert size. Qubit (Invitrogen) was used to quantify the prepared libraries while their

165   quality was assessed on a Bioanalyzer (Agilent). We used the NextSeq500 (Illumina) instrument

166   to perform the sequencing using 2x150 bp read length at the Luxembourg Centre for Systems

167   Biomedicine Sequencing Platform.

168

169   **Genome reconstruction and metagenomic data processing**

170   Paired sequences (i.e., forward and reverse) were processed using the Integrated Meta-omic

171   Pipeline (IMP) (Narayanasamy et al. 12/2016). The metagenomic workflow encompasses pre-

172   processing, assembly, and genome reconstruction in a reproducible manner. The adapter sequences

173   were trimmed in the pre-processing step including the removal of human reads. Thereafter, *de*

174   *novo* assembly was performed using the MEGAHIT (version 2.0) assembler [D. Li et al. 2015]. Default

175   IMP parameters were retained for all samples. Subsequently, we used MetaBAT2 (Kang et al.

176   2019) and MaxBin2 (Wu, Simmons, and Singer 2016) for binning in addition to an in-house

177   binning methodology previously described (Heintz-Buschart et al. 2017). The latter method

178   initially ignores the ribosomal RNA sequences in kmer profiles based on VizBin embedding

179   clusters (Laczny et al. 2015). In this context, VizBin utilises density-based non-hierarchical

180   clustering algorithms and depth of coverage for genome reconstructions. Subsequently we

181   obtained a non-redundant set of metagenome-assembled genomes (MAGs) using DASTool

182   (Sieber et al. 2018) with a score threshold of 0.7 for downstream analyses. The abundance of

183 MAGs in each sample was determined by mapping the reads to the reconstructed genomes using

184 BWA-MEM (H. Li 2013), taking the average coverage across all contigs. Diversity measures from

185 metagenomic sequencing were assessed by determining the abundance-weighted average coverage

186 of all the reads to identify the number of non-redundant read sets (Rodriguez-R and Konstantinidis

187 2014).

188

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

190 We used the trimmed and pre-processed reads from the IMP workflow to determine the microbial

191 abundance and taxonomic profiles based on the mOTU (v2) tool (Milanese et al. 2019). Based on

192 the updated marker genes in the mOTU2 database including those from the TARA Oceans Study

193 (Sunagawa et al. 2015) and recently generated MAGs (Tully, Graham, and Heidelberg 2018),

194 taxonomic profiling was performed on our sequence datasets. We used a minimum alignment

195 length of 140 bp to determine the relative abundances of the mOTUs, including the normalisation

196 of read counts to the gene length, also accounting for the base coverage of the genes. Additionally,

197 we used CheckM (Parks et al. 2015) to assess completeness and contamination. Subsequently,

198 taxonomy for MAGs recovered after the redundancy analyses from DASTool was determined

199 using the GTDB (Genome Taxonomy Database) toolkit (gtdb-tk) (Parks et al. 2018).

200

201 **Data analysis**

202 All figures for the DNA concentrations, library preparation, assembly metrics and supplementary

203 figures were generated using GraphPad Prism (v8.3.0). Taxonomical assessment and diversity

204 measures were created using version 3.6 of the R statistical software package (Team 2013).

205 DESeq2 (Love, Huber, and Anders 2014) with FDR-adjustments for multiple testing were used to

206 assess significant differences in the MAG abundances. The genomic cluster figure for the mock

207 community was obtained as an output from the IMP metagenomic pipeline.

208

209 **Results**

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

211 To ensure native sequencing, by minimizing the number of PCR (polymerase chain reaction)

212 cycles within the library preparation protocols, we tested four protocols for biomolecular

213 extraction, with an aim of acquiring large quantities (>50 ng) DNA from glacier-fed stream benthic

214 sediments. The four methods tested were selected because of their wide applicability on related
215 environmental samples (Method-1 & -2) (Griffiths et al. 2000; Lever et al. 2015; Tatti et al. 2016)
216 and their improved chances of higher yields (Method-3; Qiagen communication). Since method-4
217 is considered the gold standard of DNA extraction in biomedical sciences (Dairawan and Shetty
218 2020) and of bacterial cultures (Green and Sambrook 2017), it was included in our study. The four
219 protocols are largely based on the same principles, *viz.* sample preparation, cell lysis, purification,
220 precipitation and washing (Table 1). From preliminary tests, it became apparent that a small-scale
221 approach (0.5 g input sediment) did not yield sufficient amounts of DNA for metagenomics due
222 to, on average, limited microbial biomass in the samples. Thus, all protocols (aside from Qiagen's
223 - already produced for maxi scale) were scaled up to 5 g of input sediment and a co-precipitant,
224 like linear polyacrylamide, was included in all precipitation steps. This was essential for the
225 quantitative recovery of the small amounts of extracted DNA from high solution volumes (6-10
226 ml).

227

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

238

239 Quality assessment of these DNA extracts with Nanodrop showed $OD_{260/280}$ ratios in between ~1.4
240 and ~1.6. Agarose gel electrophoresis revealed a high-molecular weight band with no apparent
241 shearing, smearing or residual RNA, indicative of high-quality DNA (Fig. 2). A secondary effect
242 appearing in certain samples, but without any perceived consequences in the quality of extracted
243 DNA whatsoever, was the development of a pink-fuchsia color of varying intensities with the
244 addition of phenol:chloroform:isoamyl alcohol (Fig. 3). This was pH dependent since samples

245 were decolorized with the addition of sodium acetate pH 5.2 in the precipitation step. This could

246 possibly be due to a ferric-chloride-phenol compound formed when chloride and phenol

247 constituents of the protocol interact reversibly with Fe^{+3} ions contained in certain samples

248 depending on local geology (Banerjee and Haldar 1950). Similar coloration has been previously

249 reported (Lever et al. 2015).

250

251 **Extraction method affects library preparation efficiency**

252 The DNA extracted based on method-3 and using phenol-chloroform methods were subsequently

253 subjected to library preparation for high-throughput whole genome shotgun sequencing. Despite

254 similar quality of DNA across both methods (~1.4-1.6 $OD_{260/280}$), library preparation using the

255 modified commercial kit did not yield any successful libraries (Fig. 4). To assess if any impurities

256 or inhibitors hampered library preparation we tested two clean-up methods for the DNA extracted

257 from the commercial kit: 1) ethanol precipitation and 2) magnetic-bead based clean-up. We found

258 that the magnetic-bead method leads to a complete loss of sample (i.e., undetectable DNA quantity

259 via Qubit analyses) during the process, especially if starting with a low input DNA concentration.

260 Although we lost six out of twelve samples using the magnetic-bead clean-up, we achieved 100%

261 efficiency in library preparation with the remaining six samples. On the other hand, ~20% of the

262 samples cleaned via ethanol precipitation failed library preparation. Contradictory to these

263 methods, DNA extracted using the phenol-chloroform based method (method-4) yielded 100%

264 efficiency in terms of library preparation without any additional clean-up (Fig. 4). Additionally,

265 we found that the distribution of the total yield after library preparation using the phenol-

266 chloroform method was more uniform across samples compared to the other methods (Fig. 4).

267

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

269 Extraction methods for whole genome shotgun sequencing may affect the sequencing itself,

270 including the quality and assembly of the reads downstream. To assess this, we used the libraries

271 prepared as described above (Fig. 4), and performed whole genome shotgun sequencing on an

272 Illumina NextSeq500. The average quality across all three methods based on short-read sequencing

273 was Q30 after trimming the leading and trailing sequences (described in Methods). We assessed

274 several assembly metrics including the average length of contigs (N50), largest alignment, total

275 aligned length and coverage. We did not find any significant differences among any of these

276   measures across all three methods (Fig. 5A-C, 5E). Using a diversity index metric, we however

277   found a more uniform distribution across all samples prepared using method-4, albeit no

278   significant differences were observed compared to the commercial kit-based extraction and library

279   preparation (Fig. 5D).

280

281   **Extraction methods influence metagenomic profiles**

282   It is well established that extraction methods (Wagner Mackenzie, Waite, and Taylor 2015) and

283   library preparation (Bowers et al. 2015) protocols affect the taxonomic profiles and genomes

284   recovered after high-throughput sequencing. We determined if the preparation methods affected

285   the overall diversity of taxa recovered and found that phenol-chloroform and the magnetic-bead

286   clean-up methods demonstrated similar levels of diversity (Shannon) as compared to samples

287   precipitated using ethanol (Fig. 6A). Overall, the community profiles of the ethanol precipitation-

288   based method were highly diverse (Fig. 6B). Interestingly, the genomes recovered and their

289   abundances were similar in the phenol-chloroform and magnetic-bead methods as well (Fig. 6C).

290   However, we observed a significant increase ($p<0.001$, FDR-adjusted p-value) in the abundance

291   of a *Ralstonia* genome when prepared with the ethanol precipitation protocol (Supplementary

292   fig.1). Additionally, we found that the number of genomes recovered using the phenol-chloroform

293   was more consistent with previously reported 16S rRNA gene sequencing profiles for GFS from

294   Austria (Wilhelm et al. 2013; Besemer et al. 2012; Wilhelm et al. 2014). Simultaneously, we used

295   an approach to identify metagenomic operational taxonomic units (mOTUs) and found that the

296   phenol-chloroform and magnetic-bead methods showed similar profiles of mOTUs compared to

297   that of ethanol precipitation (Fig. 6D).

298

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

300   To determine whether the phenol-chloroform extraction method is biased against eukaryotes, we

301   used a commercially-available mock community (ZymoBIOMICS Microbial Community

302   Standard #D6300) to assess bias and errors. After sequencing, we recovered high quality (>90%

303   completion, <5% contamination) bacterial genomes (Fig. 7A). Additionally, the abundance of the

304   microbial genomes, including one of the eukaryotes - *Saccharomyces cerevisiae*, were similar to

305   the expected levels in the mock community (Fig. 7B). On the other hand, the protocol enabled the

306   identification and partial recovery of the *Cryptococcus neoformans* genome, albeit at lower levels

307   possibly due to increased melanisation of the cell wall (Grossman and Casadevall 2017) affecting

308   lysis and subsequent extraction.

309

310   **Discussion**

311   Improved omic techniques not limited to metagenomics are robust methods for analyzing nucleic

312   acids and the characterisation of microbial communities in various environments (Jansson et al.

313   2012). One way of understanding the impacts of global climate change on GFS includes the

314   establishment of their census of microbial life (Milner et al. 2017). However, methods designed

315   for the extraction of biomolecules including DNA have not been validated for GFS sediments.

316   Although previous glacier-fed streams studies successfully used extracted DNA for 16S rRNA

317   amplicon sequencing (Ren et al. 2017; Ren, Gao, and Elser 2017; Vardhan Reddy et al. 2009;

318   Wilhelm et al. 2013) the input DNA concentration requirements are considerably higher for whole

319   genome shotgun sequencing. In order to pursue a deeper characterisation of the microbial

320   communities within the GFS sediments, increased concentrations of DNA may additionally

321   alleviate PCR biases (Brooks et al. 2015; Kim and Bae 2011). Also, as previously highlighted,

322   several methods exist for extractions from a wide variety of environmental samples, but not for

323   GFS sediments. Here, we systematically tested the utility of four extraction protocols to identify a

324   ubiquitous methodology. We found that a phenol-chloroform based extraction protocol can be

325   used for samples across geographical separations, differences in bedrock, and samples collected at

326   various distances from the glacier.

327

328   Glassing *et al*. demonstrated that inherent DNA contamination may influence microbiota

329   interpretation in low biomass samples (Glassing et al. 2016). Additionally, it is known that certain

330   compounds - polysaccharides, humic acids, may affect PCR reactions (Rådström et al. 2004),

331   requiring the need for additional DNA clean-up. It has been established that DNA losses occur

332   during the purification steps (Roose-Amsaleg, Garnier-Sillam, and Harry 2001), including when

333   using commercial column methods (Howeler, Ghiorse, and Walker 2003; Lloyd et al. 2013), and

334   phenol-chloroform (Ogram, Sayler, and Barkay 1987). Interestingly, we found similar losses when

335   using the magnetic bead clean-up, whereas the ethanol precipitation method was inefficient

336   compared to the phenol-chloroform protocol. Though the kit-based methods are more convenient

337   and safer than phenol-chloroform extractions (Tesena et al. 2017), access to reagents and costs

338   may be a considerable factor. On the other hand, isolation of the aqueous phase from phenol-

339   chloroform can be user-dependent potentially affecting reproducibility, while kits have been

340   shown to be more consistent across samples (Claassen et al. 2013). Another key feature of our

341   findings was the potential for the kit-based methods to influence the efficiency of genome

342   reconstruction and variability in the taxonomic profiles that were recovered. While this has been

343   reported previously (Wagner Mackenzie, Waite, and Taylor 2015; Carrigg et al. 2007), we found

344   considerable variability when compared to the phenol-chloroform. This is plausible due to the

345   incomplete dissolution of DNA in buffers, especially when using methods involving charged

346   minerals (Vorhies and Gaines 2009; Barton et al. 2006; Vishnivetskaya et al. 2014), which may

347   additionally affect DNA stability. Additionally, we found that method-4 may be used for low

348   biomass sediment samples and can be scaled down in the context of abundant cells per gram (data

349   not shown).

350

351   The utility of extraction methods extends beyond the process itself, impacting downstream

352   applications such as whole genome shotgun sequencing. Our study shows that phenol-chloroform

353   may be an under-appreciated yet powerful method for isolating nucleic acids from glacier-fed

354   stream sediments. While additional steps may be required towards extraction of other biomolecules

355   such as RNA, proteins and metabolites, minor modifications may be sufficient (Toni et al. 2018).

356   Moreover, we report for the first time a systematic assessment of biomolecular extraction methods

357   for GFS sediments. Our findings though fundamental and previously unexplored, may lay the

358   foundations for future in-depth characterisation of GFS microbial communities.

359

360   **Data Availability**

361   The sequencing data generated during the current study are available from NCBI under the

362   BioProject accession number PRJNA624048. A reporting summary for the uploaded data has been

363   included as a metadata file at the accession listed ID. All extraction protocols including the

364   modified commercial methods are available in the *Supplementary Materials*.

365

366   **Acknowledgments**

367   We are grateful to Laura de Nies, Camille-Martin Gallausiaux, Jean-Pierre Trezzi, Cedric Laczny,

368   Audrey Frachet, Lea Grandmougin, Annegrat Daujemont, Laura Lebrun (LCSB) for discussions

369 and laboratory support. The University of Luxembourg Sequencing Platform and HPC facilities

370 were highly instrumental for the *in-silico* analyses. The present work was supported by NOMIS

371 Foundation and the Swiss National Science Foundation (CRSII5_180241) to Tom J. Battin.

372

373 **Contributions**

374 P.P. performed the biomolecular extractions, including the validation of methods 1-3 alongside

375 quality analyses and quantification. S.B.B. curated and validated the phenol-chloroform extraction

376 method and whole genome shotgun sequencing analyses. T.B. and H.P. collected the glacier-fed

377 stream samples for the experiments. S.F. did the DNA extractions, quantification and

378 qualifications alongside P.P. R.H. handled the library preparation for all samples and the

379 subsequent sequencing. P.W. contributed significantly to the development of method-1 in the

380 manuscript. S.B.B., P.P., S.F., H.P., P.W., and T.J.B. conceived and formulated the experiments.

381 S.B.B. and P.P. developed the manuscript with equal contributions from all authors.

382

383 **Ethical declarations**

384 **Conflicts of interest**

385 The authors do not have any competing interests.

386

387 **Figure legends**

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

389 Boxplots represent the total amount of DNA (ng) extracted from 5 g of sediment when comparing

390 (A) method-1 versus the modified-commercial kit-based method-3 and (B) method-2 versus

391 method-3. (C) Boxplots of the DNA quantities isolated from three glacial floodplains (CBS -

392 Corbassière, FEU - Val Ferret up site, FED - Val Ferret down site), using method -2, -3 and -4.

393

394 *Figure 2. Agarose gel electrophoresis of DNA extracted with method-4*

395 Lane 1: GeneRuler 1 kb DNA ladder; lanes 2-4: CBS, FED, FEU respectively.

396

397 *Figure 3. Pink-fuchsia supernatants developed during phenol:chloroform extraction step*

398

399 *Figure 4. Library preparation efficiency*

400   The efficiency or success percentage for prepared libraries based on the individual methods is

401   indicated in the table. Boxplots represent concentrations of the prepared libraries.

402

403   *Figure 5. Estimate of assembly metrics following extraction*

404   Barplots demonstrate the (A) N50 for the sequence assemblies, (B) length of the longest aligned

405   sequence, (C) the total aligned length. Bars indicate standard deviation from the mean. (D) Boxplot

406   showing the nonpareil diversity index across the three groups. (E) Percentage of coverage of the

407   assembled sequences by read-mapping is depicted.

408

409   *Figure 6. Diversity and taxonomic profiles of the metagenomic sequencing*

410   (A) Boxplot showing the Shannon diversity index for the taxonomic profiling for the three groups.

411   Significance was tested using a One-way ANOVA with Student-Neuman Keul's post-hoc analysis.

412   ***p-value<0.001, ****p-value<0.0001. (B) Principal component analyses generated using Bray-

413   Curtis dissimilarity matrix depicts similarities or lack thereof between the three groups. (C)

414   Abundances of the reconstructed genomes are depicted for method-3 + EtOH, method-3 +

415   magnetic bead clean-up and method-4 extraction. (D) Heatmap demonstrating the mOTUs for the

416   three methods is depicted. The hierarchical clustering for the heatmap was generated using Ward's

417   clustering algorithm.

418

419   *Figure 7. Evaluation of phenol-chloroform extraction using a mock community*

420   (A) Scatterplot depicts the clusters of contigs representative of the reconstructed genomes after

421   processing the mock community using the IMP meta-omics pipeline. The taxonomic identity is

422   displayed next to the respective clusters. (B) Barplots indicate the relative abundance of the

423   individual genomes recovered from the mock community sequencing after extraction with the

424   phenol-chloroform method. The upper (black) line represents the expected abundance (12%) of

425   the prokaryotes, while the lower (red) line indicates the expected abundance (2%) of the

426   eukaryotes.

427

428   *Supplementary figure 1.*

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

430 The abundance of the *Ralstonia* genome recovered from the samples when processed with method-

431 3 (EtOH and magnetic bead clean-up) and method-4. Significance was tested using One-Way

432 ANOVA with Student-Neuman Keul's post-hoc analyses. ***$p<0.001$, ****$p<0.0001$

433

434 **Tables**

435 *Table 1: Key characteristics of the four selected methods*

	Method-1	**Method-2**	**Method-3**	**Method-4**
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

436

437 **References**

438 Amalfitano, Stefano, and Stefano Fazi. 2008. "Recovery and Quantification of Bacterial Cells
439   Associated with Streambed Sediments." *Journal of Microbiological Methods* 75 (2): 237–43.
440 Banerjee, S., and B. C. Haldar. 1950. "Constitution of Ferri-Phenol Complex in Solution." *Nature*
441   165 (4208): 1012.
442 Barton, H. A., N. M. Taylor, B. R. Lubbers, and A. C. Pemberton. 2006. "DNA Extraction from
443   Low-Biomass Carbonate Rock: An Improved Method with Reduced Contamination and the
444   Low-Biomass Contaminant Database." *Journal of Microbiological Methods* 66 (1): 21–31.

445 Besemer, Katharina, Hannes Peter, Jürg B. Logue, Silke Langenheder, Eva S. Lindström, Lars
446   J. Tranvik, and Tom J. Battin. 2012. "Unraveling Assembly of Stream Biofilm Communities."
447   *The ISME Journal* 6 (8): 1459–68.
448 Bowers, Robert M., Alicia Clum, Hope Tice, Joanne Lim, Kanwar Singh, Doina Ciobanu, Chew
449   Yee Ngan, Jan-Fang Cheng, Susannah G. Tringe, and Tanja Woyke. 2015. "Impact of
450   Library Preparation Protocols and Template Quantity on the Metagenomic Reconstruction
451   of a Mock Microbial Community." *BMC Genomics* 16 (October): 856.
452 Brooks, J. Paul, David J. Edwards, Michael D. Harwich Jr, Maria C. Rivera, Jennifer M.
453   Fettweis, Myrna G. Serrano, Robert A. Reris, et al. 2015. "The Truth about Metagenomics:
454   Quantifying and Counteracting Bias in 16S rRNA Studies." *BMC Microbiology* 15 (March):
455   66.
456 Carrigg, Cora, Olivia Rice, Siobhán Kavanagh, Gavin Collins, and Vincent O'Flaherty. 2007.
457   "DNA Extraction Method Affects Microbial Community Profiles from Soils and Sediment."
458   *Applied Microbiology and Biotechnology* 77 (4): 955–64.
459 Claassen, Shantelle, Elloise du Toit, Mamadou Kaba, Clinton Moodley, Heather J. Zar, and
460   Mark P. Nicol. 2013. "A Comparison of the Efficiency of Five Different Commercial DNA
461   Extraction Kits for Extraction of DNA from Faecal Samples." *Journal of Microbiological*
462   *Methods* 94 (2): 103–10.
463 Dahm, Ralf. 2008. "Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid
464   Research." *Human Genetics* 122 (6): 565–81.
465 Dairawan, Mariyam, and Preetha J. Shetty. 2020. "The Evolution of DNA Extraction Methods."
466   *American Journal of Biomedical Science & Research* 8 (1).
467   https://doi.org/10.34297/AJBSR.2020.08.001234.
468 Dancer, S. J., P. Shears, and D. J. Platt. 1997. "Isolation and Characterization of Coliforms from
469   Glacial Ice and Water in Canada's High Arctic." *Journal of Applied Microbiology* 82 (5):
470   597–609.
471 Glassing, Angela, Scot E. Dowd, Susan Galandiuk, Brian Davis, and Rodrick J. Chiodini. 2016.
472   "Inherent Bacterial DNA Contamination of Extraction and Sequencing Reagents May Affect
473   Interpretation of Microbiota in Low Bacterial Biomass Samples." *Gut Pathogens* 8 (May):
474   24.
475 Green, Michael R., and Joseph Sambrook. 2017. "Isolating DNA from Gram-Negative Bacteria."
476   *Cold Spring Harbor Protocols* 2017 (1). https://doi.org/10.1101/pdb.prot093369.
477 Griffiths, R. I., A. S. Whiteley, A. G. O'Donnell, and M. J. Bailey. 2000. "Rapid Method for
478   Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA-
479   and rRNA-Based Microbial Community Composition." *Applied and Environmental*
480   *Microbiology* 66 (12): 5488–91.
481 Grossman, Nina T., and Arturo Casadevall. 2017. "Physiological Differences in Cryptococcus
482   Neoformans Strains In Vitro versus In Vivo and Their Effects on Antifungal Susceptibility."
483   *Antimicrobial Agents and Chemotherapy* 61 (3). https://doi.org/10.1128/AAC.02108-16.
484 Hale, Lauren, and David Crowley. 2015. "DNA Extraction Methodology for Biochar-Amended
485   Sand and Clay." *Biology and Fertility of Soils* 51 (6): 733–38.
486 Hammes, Frederik, Michael Berney, Yingying Wang, Marius Vital, Oliver Köster, and Thomas
487   Egli. 2008. "Flow-Cytometric Total Bacterial Cell Counts as a Descriptive Microbiological
488   Parameter for Drinking Water Treatment Processes." *Water Research* 42 (1-2): 269–77.
489 Heintz-Buschart, Anna, Patrick May, Cédric C. Laczny, Laura A. Lebrun, Camille Bellora,
490   Abhimanyu Krishna, Linda Wampach, et al. 2017. "Integrated Multi-Omics of the Human
491   Gut Microbiome in a Case Study of Familial Type 1 Diabetes." *Nature Microbiology* 2 (1):
492   16180.
493 Howeler, Michael, William C. Ghiorse, and Larry P. Walker. 2003. "A Quantitative Analysis of
494   DNA Extraction and Purification from Compost." *Journal of Microbiological Methods* 54 (1):
495   37–45.

496   Jansson, Janet K., and Kirsten S. Hofmockel. 2018. "The Soil Microbiome—from Metagenomics
497   to Metaphenomics." *Current Opinion in Microbiology*, Environmental Microbiology * The
498   New Microscopy, 43 (June): 162–68.
499   Jansson, Janet K., Josh D. Neufeld, Mary Ann Moran, and Jack A. Gilbert. 2012. "Omics for
500   Understanding Microbial Functional Dynamics." *Environmental Microbiology* 14 (1): 1–3.
501   Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong
502   Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome
503   Reconstruction from Metagenome Assemblies." *PeerJ* 7 (July): e7359.
504   Kim, Kyoung-Ho, and Jin-Woo Bae. 2011. "Amplification Methods Bias Metagenomic Libraries
505   of Uncultured Single-Stranded and Double-Stranded DNA Viruses." *Applied and
506   Environmental Microbiology* 77 (21): 7663–68.
507   Laczny, Cedric C., Tomasz Sternal, Valentin Plugaru, Piotr Gawron, Arash Atashpendar, Houry
508   Hera Margossian, Sergio Coronado, Laurens van der Maaten, Nikos Vlassis, and Paul
509   Wilmes. 2015. "VizBin - an Application for Reference-Independent Visualization and
510   Human-Augmented Binning of Metagenomic Data." *Microbiome* 3 (1): 1.
511   Lever, Mark A., Andrea Torti, Philip Eickenbusch, Alexander B. Michaud, Tina Šantl-Temkiv,
512   and Bo Barker Jørgensen. 2015. "A Modular Method for the Extraction of DNA and RNA,
513   and the Separation of DNA Pools from Diverse Environmental Sample Types." *Frontiers in
514   Microbiology* 6 (May): 476.
515   Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015.
516   "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics
517   Assembly via Succinct de Bruijn Graph." *Bioinformatics* 31 (10): 1674–76.
518   Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-
519   MEM." *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.
520   Lloyd, Karen G., Lars Schreiber, Dorthe G. Petersen, Kasper U. Kjeldsen, Mark A. Lever,
521   Andrew D. Steen, Ramunas Stepanauskas, et al. 2013. "Predominant Archaea in Marine
522   Sediments Degrade Detrital Proteins." *Nature* 496 (7444): 215–18.
523   Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold
524   Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12).
525   https://doi.org/10.1186/s13059-014-0550-8.
526   Milanese, Alessio, Daniel R. Mende, Lucas Paoli, Guillem Salazar, Hans-Joachim Ruscheweyh,
527   Miguelangel Cuenca, Pascal Hingamp, et al. 2019. "Microbial Abundance, Activity and
528   Population Genomic Profiling with mOTUs2." *Nature Communications* 10 (1): 1014.
529   Miller, D. N., J. E. Bryant, E. L. Madsen, and W. C. Ghiorse. 1999. "Evaluation and Optimization
530   of DNA Extraction and Purification Procedures for Soil and Sediment Samples." *Applied
531   and Environmental Microbiology* 65 (11): 4715–24.
532   Milner, Alexander M., Kieran Khamis, Tom J. Battin, John E. Brittain, Nicholas E. Barrand,
533   Leopold Füreder, Sophie Cauvy-Frauné, et al. 2017. "Glacier Shrinkage Driving Global
534   Changes in Downstream Systems." *Proceedings of the National Academy of Sciences of
535   the United States of America* 114 (37): 9770–78.
536   Narayanasamy, Shaman, Yohan Jarosz, Emilie E. L. Muller, Anna Heintz-Buschart, Malte
537   Herold, Anne Kaysen, Cédric C. Laczny, Nicolás Pinel, Patrick May, and Paul Wilmes.
538   12/2016. "IMP: A Pipeline for Reproducible Reference-Independent Integrated
539   Metagenomic and Metatranscriptomic Analyses." *Genome Biology* 17 (1): 260.
540   Nielsen, Jens, and Boyang Ji. 2015. "New Insight into the Gut Microbiome through
541   Metagenomics." *Advances in Genomics and Genetics*. https://doi.org/10.2147/agg.s57215.
542   Ogram, Andrew, Gary S. Sayler, and Tamar Barkay. 1987. "The Extraction and Purification of
543   Microbial DNA from Sediments." *Journal of Microbiological Methods* 7 (2): 57–66.
544   Parks, Donovan H., Maria Chuvochina, David W. Waite, Christian Rinke, Adam Skarshewski,
545   Pierre-Alain Chaumeil, and Philip Hugenholtz. 2018. "A Proposal for a Standardized
546   Bacterial Taxonomy Based on Genome Phylogeny." *bioRxiv*, January, 256800.

547   Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W.
548   Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered from
549   Isolates, Single Cells, and Metagenomes." *Genome Research* 25 (7): 1043–55.
550   Pei-Ying, Yan, Hou Shu-Gui, Chen Tuo, Zhang Shu-Hong, and Sun Wei-Jun. 2012. "Methods
551   for Extraction of Microorganism DNA from Glacier Surface Snow." *Sciences in Cold and*
552   *Arid Regions* 4 (6): 484.
553   Porebski, Sue, L. Grant Bailey, and Bernard R. Baum. 1997. "Modification of a CTAB DNA
554   Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol
555   Components." *Plant Molecular Biology Reporter / ISPMB* 15 (1): 8–15.
556   Psifidi, Androniki, Chrysostomos I. Dovas, Georgios Bramis, Thomai Lazou, Claire L. Russel,
557   Georgios Arsenos, and Georgios Banos. 2015. "Comparison of Eleven Methods for
558   Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-
559   Term DNA Banking Using Blood Samples." *PloS One* 10 (1): e0115960.
560   Rådström, Peter, Rickard Knutsson, Petra Wolffs, Maria Lövenklev, and Charlotta Löfström.
561   2004. "Pre-PCR Processing: Strategies to Generate PCR-Compatible Samples." *Molecular*
562   *Biotechnology* 26 (2): 133–46.
563   Ren, Ze, Hongkai Gao, and James J. Elser. 2017. "Longitudinal Variation of Microbial
564   Communities in Benthic Biofilms and Association with Hydrological and Physicochemical
565   Conditions in Glacier-Fed Streams." *Freshwater Science* 36 (3): 479–90.
566   Ren, Ze, Hongkai Gao, James J. Elser, and Qiudong Zhao. 2017. "Microbial Functional Genes
567   Elucidate Environmental Drivers of Biofilm Metabolism in Glacier-Fed Streams." *Scientific*
568   *Reports* 7 (October). https://doi.org/10.1038/s41598-017-13086-9.
569   Rodriguez-R, Luis M., and Konstantinos T. Konstantinidis. 2014. "Nonpareil: A Redundancy-
570   Based Approach to Assess the Level of Coverage in Metagenomic Datasets."
571   *Bioinformatics* 30 (5): 629–35.
572   Roose-Amsaleg, C. L., E. Garnier-Sillam, and M. Harry. 2001. "Extraction and Purification of
573   Microbial DNA from Soil and Sediment Samples." *Applied Soil Ecology: A Section of*
574   *Agriculture, Ecosystems & Environment* 18 (1): 47–60.
575   Roume, Hugo, Anna Heintz-Buschart, Emilie E. L. Muller, and Paul Wilmes. 2013. "Sequential
576   Isolation of Metabolites, RNA, DNA, and Proteins from the Same Unique Sample." *Methods*
577   *in Enzymology* 531: 219–36.
578   Sambrook, Joseph, and David W. Russell. 2006. "Purification of Nucleic Acids by Extraction
579   with Phenol:chloroform." *CSH Protocols* 2006 (1). https://doi.org/10.1101/pdb.prot4455.
580   Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess,
581   Susannah G. Tringe, and Jillian F. Banfield. 2018. "Recovery of Genomes from
582   Metagenomes via a Dereplication, Aggregation and Scoring Strategy." *Nature Microbiology*
583   3 (7): 836–43.
584   Smith, K., M. A. Diggle, and S. C. Clarke. 2003. "Comparison of Commercial DNA Extraction
585   Kits for Extraction of Bacterial Genomic DNA from Whole-Blood Samples." *Journal of*
586   *Clinical Microbiology* 41 (6): 2440–43.
587   Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie,
588   Guillem Salazar, Bardya Djahanschiri, et al. 2015. "Ocean Plankton. Structure and Function
589   of the Global Ocean Microbiome." *Science* 348 (6237): 1261359.
590   Tatti, Enrico, Boyd A. McKew, Corrine Whitby, and Cindy J. Smith. 2016. "Simultaneous DNA-
591   RNA Extraction from Coastal Sediments and Quantification of 16S rRNA Genes and
592   Transcripts by Real-Time PCR." *Journal of Visualized Experiments: JoVE*, no. 112 (June).
593   https://doi.org/10.3791/54067.
594   Team, R. Core. 2013. "R: A Language and Environment for Statistical Computing."
595   Tesena, Parichart, Wasamon Korchunjit, Jane Taylor, and Tuempong Wongtawan. 2017.
596   "Comparison of Commercial RNA Extraction Kits and qPCR Master Mixes for Studying
597   Gene Expression in Small Biopsy Tissue Samples from the Equine Gastric Epithelium."

598   *Journal of Equine Science* 28 (4): 135–41.

599 Toni, Lee S., Anastacia M. Garcia, Danielle A. Jeffrey, Xuan Jiang, Brian L. Stauffer, Shelley D.
600   Miyamoto, and Carmen C. Sucharov. 2018. "Optimization of Phenol-Chloroform RNA
601   Extraction." *MethodsX* 5 (May): 599–608.

602 Tully, Benjamin J., Elaina D. Graham, and John F. Heidelberg. 2018. "The Reconstruction of
603   2,631 Draft Metagenome-Assembled Genomes from the Global Oceans." *Scientific Data* 5
604   (January): 170203.

605 Vardhan Reddy, Puram Vishnu, Singireesu Soma Shiva Nageswara Rao, Mambatta
606   Shankaranarayanan Pratibha, Buddhi Sailaja, Bakka Kavya, Ravoori Ruth Manorama, Shiv
607   Mohan Singh, Tanuku Naga Radha Srinivas, and Sisinthy Shivaji. 2009. "Bacterial Diversity
608   and Bioprospecting for Cold-Active Enzymes from Culturable Bacteria Associated with
609   Sediment from a Melt Water Stream of Midtre Lovenbreen Glacier, an Arctic Glacier."
610   *Research in Microbiology* 160 (8): 538–46.

611 Vishnivetskaya, Tatiana A., Alice C. Layton, Maggie C. Y. Lau, Archana Chauhan, Karen R.
612   Cheng, Arthur J. Meyers, Jasity R. Murphy, et al. 2014. "Commercial DNA Extraction Kits
613   Impact Observed Microbial Community Composition in Permafrost Samples." *FEMS
614   Microbiology Ecology* 87 (1): 217–30.

615 Vorhies, John S., and Robert R. Gaines. 2009. "Microbial Dissolution of Clay Minerals as a
616   Source of Iron and Silica in Marine Sediments." *Nature Geoscience* 2 (3): 221–25.

617 Wagner Mackenzie, Brett, David W. Waite, and Michael W. Taylor. 2015. "Evaluating Variation
618   in Human Gut Microbiota Profiles due to DNA Extraction Method and Inter-Subject
619   Differences." *Frontiers in Microbiology* 6 (February): 130.

620 Wilhelm, Linda, Katharina Besemer, Christina Fasching, Tim Urich, Gabriel A. Singer,
621   Christopher Quince, and Tom J. Battin. 2014. "Rare but Active Taxa Contribute to
622   Community Dynamics of Benthic Biofilms in Glacier-Fed Streams." *Environmental
623   Microbiology* 16 (8): 2514–24.

624 Wilhelm, Linda, Gabriel A. Singer, Christina Fasching, Tom J. Battin, and Katharina Besemer.
625   2013. "Microbial Biodiversity in Glacier-Fed Streams." *The ISME Journal* 7 (8): 1651–60.

626 Wu, Yu-Wei, Blake A. Simmons, and Steven W. Singer. 2016. "MaxBin 2.0: An Automated
627   Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets."
628   *Bioinformatics* 32 (4): 605–7.

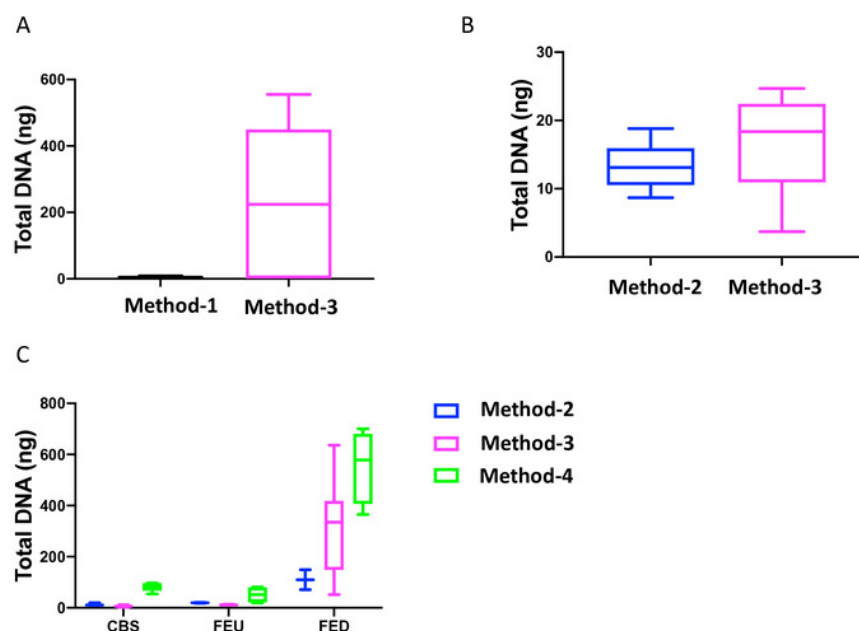
629 Xin, Zhanguo, and Junping Chen. 2012. "A High Throughput DNA Extraction Method with High
630   Yield and Quality." *Plant Methods*. https://doi.org/10.1186/1746-4811-8-26.

631 Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. "DNA Recovery from Soils of Diverse
632   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.

Figure 2

Agarose gel electrophoresis of DNA extracted with method-4

Lane 1: GeneRuler 1 kb DNA ladder; lanes 2-4: CBS, FED, FEU respectively.

Figure 2. Agarose gel electrophoresis of DNA extracted with method-4

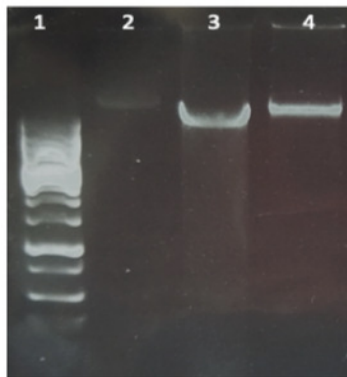



Figure 3

Pink-fuchsia supernatants developed during phenol:chloroform extraction step

Figure 3. Pink-fucshia supernatants developed during phenol:chloroform extraction step

Figure 4

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. Library preparation efficiency

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

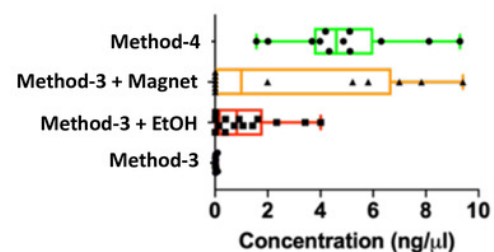
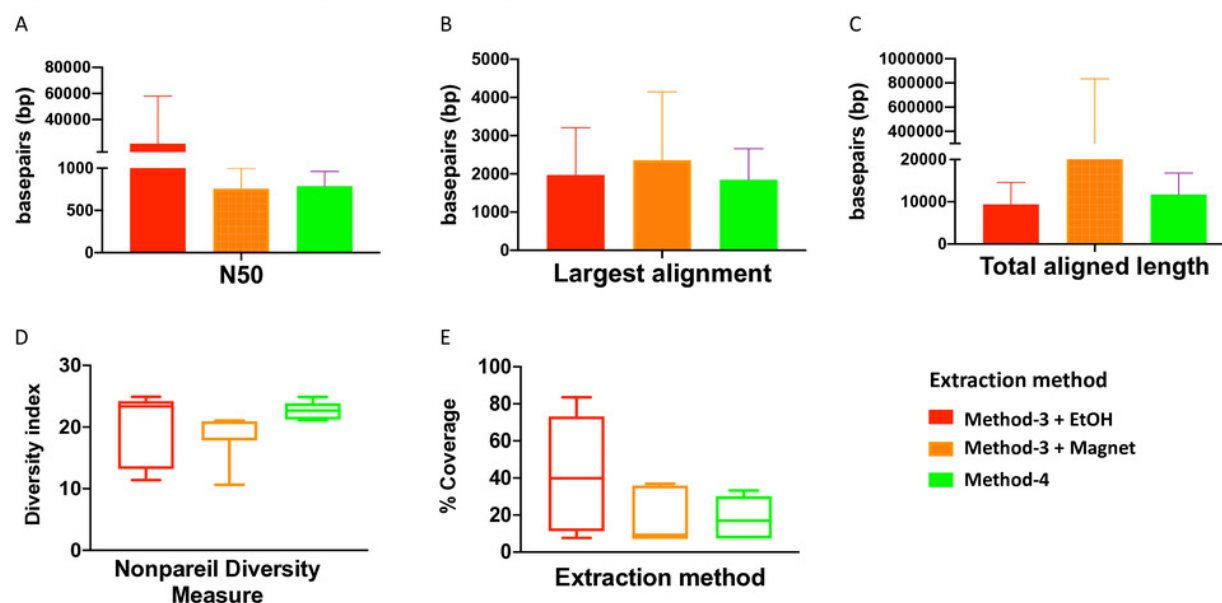



Figure 5

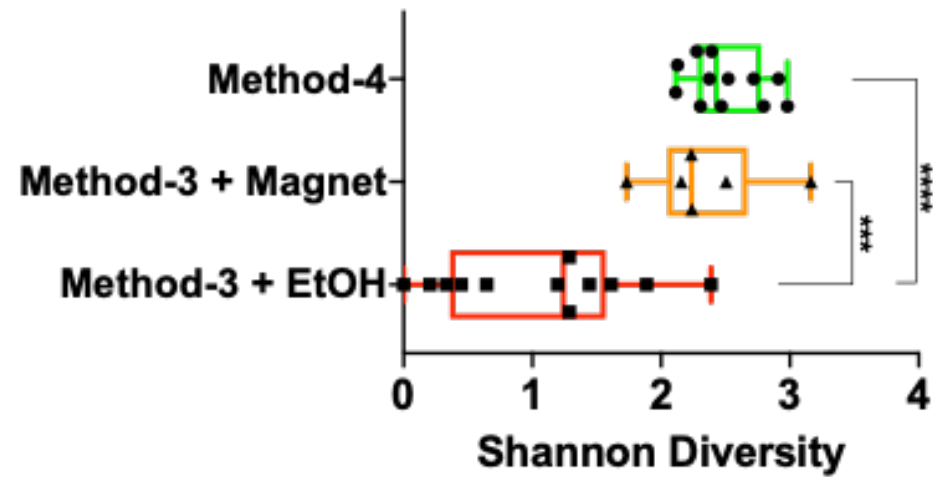
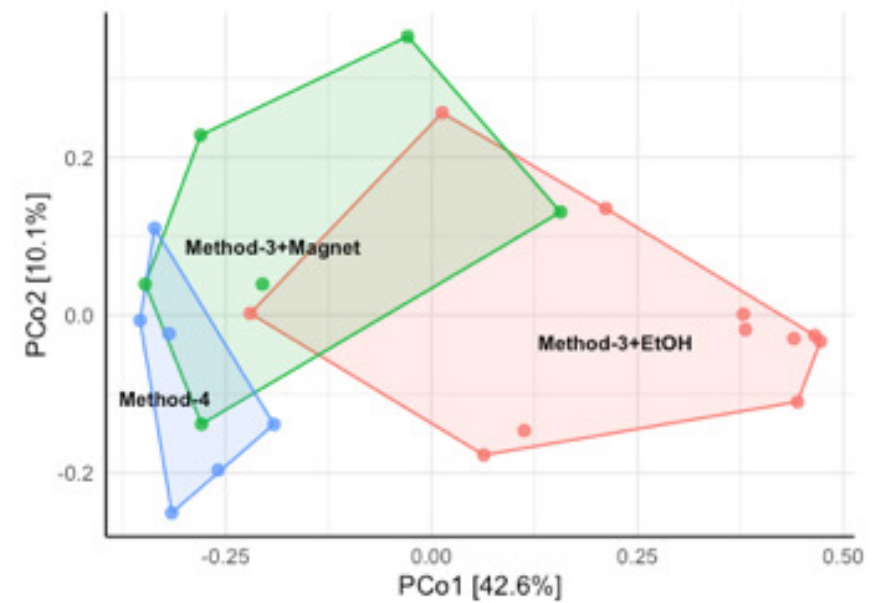
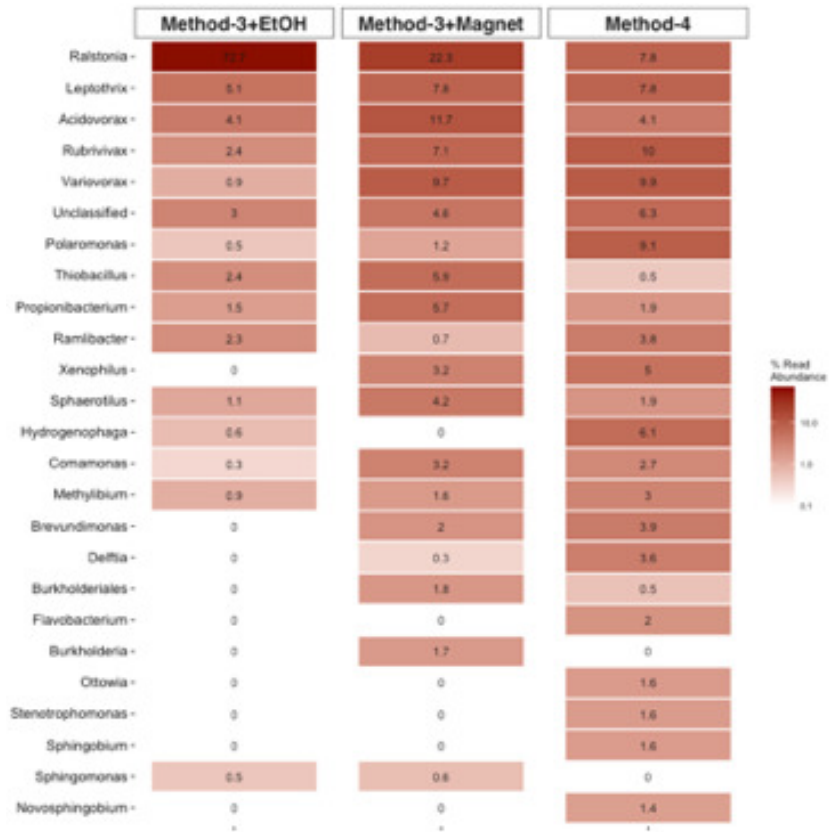
Estimate of assembly metrics following extraction

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. Estimate of assembly metrics following extraction

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. Diversity and taxonomic profiles from metagenomic sequencing

A

B

C

D

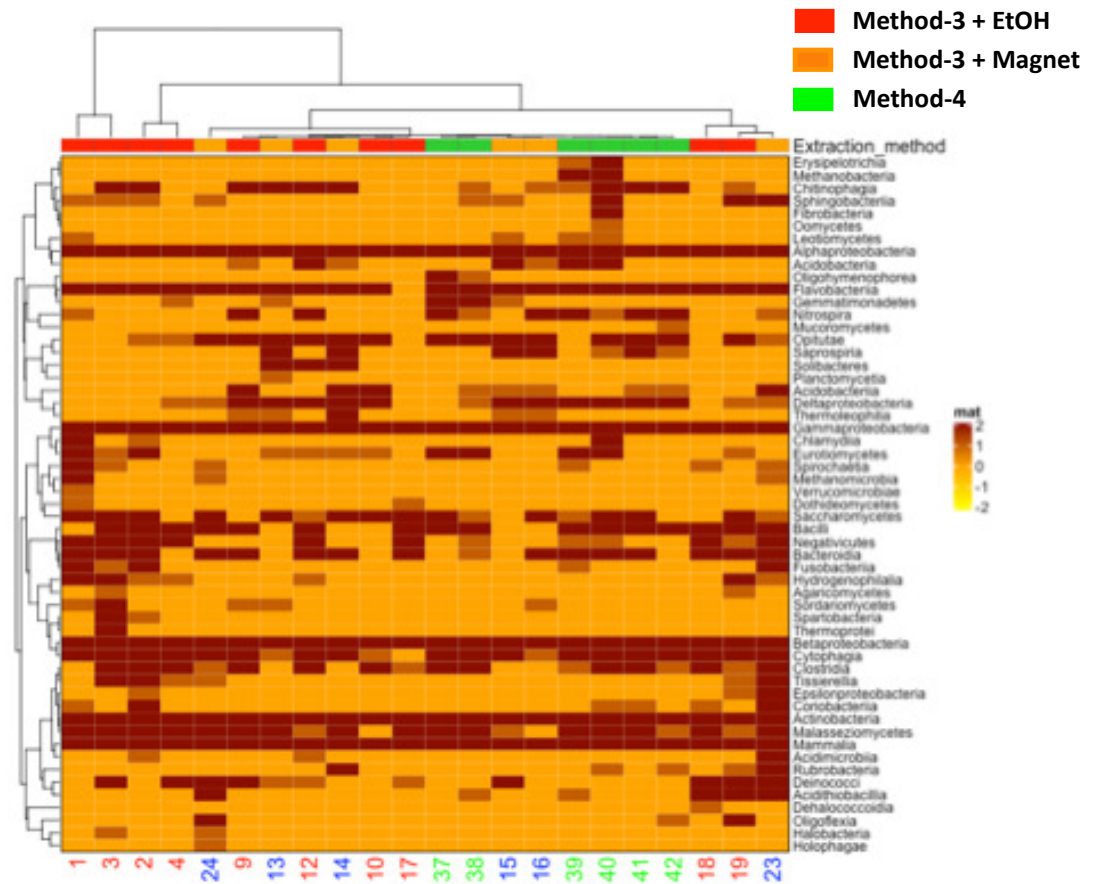



Figure 7

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.

Figure 7. Evaluation of phenol:chloroform extraction using a mock community