Determining the Reliability of Measuring Mercury Cycling Gene Abundance with Correlations with Mercury and Methylmercury Concentrations

Geoffrey A. Christensen[1], Caitlin M. Gionfriddo[1], Andrew J. King[1], James G. Moberly[2], Carrie Miller[3], Anil C. Somenahally[4], Stephen J. Callister[9], Heather Brewer[9], Mircea Podar[1], Steven D. Brown[1], Anthony V. Palumbo[1], Craig C. Brandt[1], Ann M. Wymore[1], Scott C. Brooks[5], Chiachi Hwang[8], Matthew W. Fields[8], Judy D. Wall[7], Cynthia C. Gilmour[6], and Dwayne A. Elias[1,*]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN USA
[2]College of Engineering, University of Idaho, Moscow, ID USA
[3]Department of Biology, Troy University, Troy, AL USA
[4]Department of Soil and Crop Sciences, Texas A&M University, Overton, TX, USA
[5]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN USA
[6]Smithsonian Environmental Research Center, Edgewater, MD USA.
[7]Department of Biochemistry, University of Missouri, Columbia, MO USA.
[8]Department of Microbiology and Immunology, Montana State University, Bozeman, MT USA
[9]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA USA

*Corresponding author
Dwayne A. Elias
Biosciences Division
Oak Ridge National Laboratory
P.O. Box 2008, MS-6036
Oak Ridge, TN, USA
37831-6036
Email: eliasda@ornl.gov
Phone: 1-865-574-0956

31 Fax: 1-865-576-8646
32

33

34 Key Words: microbial community, PCR, qPCR, metagenome, *hgcAB*, 16S sequencing

35

36 ABSTRACT

37 Methylmercury (MeHg) is a bioaccumulative toxic contaminant in many ecosystems, but
38 factors governing its production are poorly understood. Recent work has shown that the
39 anaerobic microbial conversion of mercury (Hg) to MeHg requires the Hg-methylation
40 genes *hgcAB* and that these genes can be used as biomarkers in PCR-based estimators of
41 Hg-methylator abundance. In an effort to determine reliable methods for assessing *hgcA*
42 abundance and diversity and linking them to MeHg concentrations, multiple approaches
43 were compared including metagenomic shotgun sequencing, 16S rRNA gene
44 pyrosequencing and cloning/sequencing *hgcAB* gene products. Hg-methylator
45 abundance was also determined by quantitative *hgcA* qPCR amplification and
46 metaproteomics for comparison to the above measurements. Samples from eight sites
47 were examined covering a range of total Hg (HgT; 0.03–14 mg kg^{-1} dry wt. soil) and
48 MeHg (0.05–27 µg kg^{-1} dry wt. soil) concentrations. In the metagenome and amplicon
49 sequencing of *hgcAB* diversity, the *Deltaproteobacteria* were the dominant Hg-
50 methylators while *Firmicutes* and methanogenic *Archaea* were typically ~50% less
51 abundant. This was consistent with metaproteomics estimates where the

52 *Deltaproteobacteria* were steadily higher. The 16S rRNA gene pyrosequencing did not

53 have sufficient resolution to identify *hgcAB*$^{+}$ species. Metagenomic and *hgcAB* results

54 were similar for Hg-methylator diversity and clade-specific qPCR-based approaches for

55 *hgcA* are only appropriate when comparing the abundance of a particular clade across

56 various samples. Weak correlations between Hg-methylating bacteria and soil Hg

57 concentrations were observed for similar environmental samples, but overall total Hg

58 and MeHg concentrations poorly correlated with Hg-cycling genes.

59

60 INTRODUCTION

61 Methylmercury (MeHg) is a bioaccumulative neurotoxic contaminant[1] that is produced
62 predominantly via a microbially mediated reaction. The ability to methylate Hg is
63 conferred by the gene pair *hgcAB*[2] with gene presence being strain-specific across
64 clades, and methylation efficiencies vary[3] amongst clades.[4] Microbial Hg-methylation
65 primarily occurs in anaerobic environments and is predominantly associated with
66 sulfate/sulfite- and iron-reducing *Deltaproteobacteria* and *Firmicutes*, acetogenic and
67 fermentative *Firmicutes*, and methanogenic *Archaea*[2, 4-7]. Initial studies suggested a
68 predictive relationship for determining MeHg production from total Hg (HgT)[8-10], but
69 was valid only for individual lakes[11, 12]. Hence factors other than Hg(II) concentration
70 are clearly at play in governing MeHg production. In general, the abundance, activity,
71 and diversity of Hg-methylating microorganisms varies within a site due to various
72 temporal, seasonal, and physical factors, including topography, groundwater flow,
73 primary production, organic matter composition and abundance, redox conditions and
74 electron acceptor concentrations[13-18]. Understanding how these conditions influence
75 *hgcAB* diversity and activity presumably could be used to predict MeHg production, and

76 in turn, Hg-cycling at the ecosystem scale [11, 19-21]. The development of genome-based

77 predictive models of net MeHg concentrations in the environment depend on

78 quantitative relationships between the genetic basis of Hg-cycling reactions (e.g.,

79 *hgcAB*, *mer* operon) and rates of MeHg production and MeHg demethylation.

80 Hg-methylator diversity is often linked to redox conditions [13-15] but a comparison of

81 methylator diversity to MeHg and Hg concentrations has failed to show significant

82 correlations [13-15, 21]. This is likely because net MeHg production not only depends on the

83 presence of *hgcAB*+ microorganisms, but also on various biogeochemical controls on

84 microbial Hg uptake and activity including natural organic matter (NOM) and sulfide

85 composition[22, 23 20, 24-26], photolytic or microbial demethylation and Hg(II) reduction[27-29

86 30], and microbial respiration [3, 13, 28, 31-34]. The formation of colloidal and particulate Hg

87 complexes as well as Hg-cell thiol interactions limits the available Hg pool that can be

88 taken up by *hgcAB*+ organisms and converted to MeHg[20, 25, 35]. Methylation rates may

89 also be influenced by cellular activity, such as HgcAB transcription and protein

90 structure[36]. Expression of *hgcAB* is likely not linked to Hg exposure[37, 38], but instead may

91 be connected to an alternative function encoded by *hgcAB*, such as one carbon

92 metabolism [2, 39]. Therefore, factors that govern the activity of anaerobic microorganisms

93 and stimulate syntrophic relationships of *hgcAB* organisms, including availability of

94 labile carbon and terminal electron acceptors may better predict Hg-methylator activity

95 rather than available Hg pools[40-42].

96 Several molecular methods can measure the abundance and diversity of a gene sequence

97 of interest in an environment; cost often increasing with data yield. Since the discovery

98 of *hgcAB* [2], these genes have been used as a biomarker for detecting Hg-methylators in

99 the environment via both gene-specific and metagenomic approaches [13-15, 43-46]. High-

100 throughput metagenomic sequencing and analysis provide a microbial communities'

101 overall genetic capacity, community dynamics and composition, as well as potential

102 metabolic functions, including Hg transformations. In contrast, PCR-based approaches

103 provide a focused, more cost and time effective method to directly target the gene(s) of

104 interest, without the same depth of taxonomic and functional information. It is currently

105 unclear whether the estimates of Hg-methylator diversity are equivalent between the two

106 methods. It is also unclear how well *hgcAB* abundance via metaproteomic analysis

107 correlates with metagenomics but may arguably be more relevant since proteomics

108 measures the enzymatic capacity rather than the genomic potential for a given

109 biochemical reaction.

110 In this study, we compared several approaches to assess the presence, distribution and

111 diversity of Hg-methylating bacteria and determine if the more targeted PCR-based

112 methods yielded similar diversity information as the more involved 16S rRNA and

113 metagenomic sequencing as well as MS-based metaproteomics. Using *hgcA* gene counts

114 from metagenomes we also compared Hg-methylator abundances to *hgcA* qPCR for the

115 three major Hg-methylating clades. Eight different environments were examined,

116 including riverine areas, tidal marshes, and Arctic permafrost with a range of MeHg and

117 HgT concentrations. Finally, we assessed whether Hg-cycling gene abundances (i.e.

118 *hgcA*, *merA*, *merB*) from meta-omic and PCR-based approaches correlated with either

119 HgT or MeHg levels. The focus of this study was on assessing whether a relationship

120 exists between the presence and diversity of potential Hg-cycling organisms and ambient

121 HgT and MeHg concentrations. Assessing microbial activity and Hg bioavailability were

122 beyond the scope of this study and were not addressed.

123 METHODS

124 Sampling locations and sediment characteristics. Eight sampling sites were chosen to

125 include a broad range of HgT (0.03–14.6 mg kg^{-1} dry wt. sediment) and MeHg (0.05–

126 27.3 µg kg^{-1} dry wt. sediment) concentrations (Figure 1, Table 1) while also representing

127 a range of environments (freshwater, estuarine, and organic-rich sediment systems)

128 relevant to biogeochemical Hg-cycling. Two sites were located at the East Fork Poplar

129 Creek watershed area (EFPC) in Oak Ridge, TN. Two sediment cores were collected

130 from the center channel of EFPC at NOAA (upstream) and New Horizon (NH;

131 downstream), about 4 and 21 km, respectively, downstream of the original point-source

132 of contamination. One sediment core was collected from a background site, Hinds Creek

133 (HC), in eastern Tennessee approximately 20 km northeast of EFPC [47, 48]. The

134 Tennessee samples were obtained from intact cores 12 cm deep and the bottom 4 cm

135 was homogenized and used for analysis. Soil cores were also collected from organic

136 carbon-rich permafrost at a low-centered polygon in Barrow Alaska as part of the Next-

137 Generation Ecosystems Experiments-Arctic (NGEE) from 55-82 cm [49]. Salt marshes

138 included in the survey were the Global Change Research Wetland on Chesapeake Bay, at

139 the Smithsonian Environmental Research Center (SERC), in Edgewater, MD, and two

140 marshes on the Penobscot River in Maine. SERC cores were taken from two sites where

141 Hg biogeochemistry had been previously evaluated, an upper marsh site (UM)

142 dominated by *Spartina patens* and *Scirpus olneyi*, and a marsh fringe site (MF), located

143 at the marsh-upland interface dominated by *Typha angustifolia* (MF) [50]. In the

144 Penobscot, soil cores were collected from an oligohaline marsh (Bald Hill Cove, site

145 W10MM) and from mesohaline Mendall Marsh (site W21UM)[51]. The surface 0-3 cm of

146 wetland cores were sectioned. Upon collection all samples were placed on ice and stored

147 at -80°C upon return to the laboratory until further use. For SERC and Penobscot

148 samples, cores were taken by hand in 5 cm core barrels. The top 4 cm of soil was

149 sectioned from cores on the day of collection, homogenized in an O_2-free glove box and

150 frozen. DNA sub-samples were placed immediately on dry ice and kept in a -80°C

151 freezer until analysis.

152 ***Nucleic acid extractions***

153 Community genomic DNA (gDNA) was extracted from frozen (LN_2) and milled

154 samples following established protocols [52, 53]. Between 12 and 26 ng of DNA was

155 recovered from 5 g of sample. (Table 1). Pure cultures of *Geobacter sulfurreducens* PCA

156 DSM-12127 (*Deltaproteobacteria*) [2], *Desulfosporosinus youngiae* JW/YJL-B18 DSM-

157 17734 (*Firmicutes*) [4], and *Methanomethylovorans hollandica* DSM-15978 (*Archaea*) [4]

158 were used as positive controls and cultured as described previously. The gDNA

159 extraction used 1.5 mL of culture processed with a Wizard® Genomic DNA Purification

160 Kit (Promega, Madison, WI) following the gram-positive protocol. All gDNA pellets

161 were resuspended in nuclease-free water and frozen at -20°C until use. Final gDNA

162 concentrations were measured with Qubit® (Thermo Fisher Scientific, Waltham, MA).

163 Nanodrop™ ND-1000 (Thermo Fisher Scientific) measurements of A_{260}/A_{280} and

164 A_{260}/A_{230} ratios were used to assess purity.

165 ***16S rRNA amplicon and metagenomic sequencing, and dataset processing***

166 454 sequencing. Bacterial and Archaeal small subunit (SSU) rRNA genes were amplified

167 using bacterial primer set FD1 (5′-AGAGTTTGATCCTGGCTCAG-3′) and 529r

168 (5′-CGCGGCTGCTGGCAC-3′) and Archaeal primer set 751F (5′-

169 CCGACGGTGAGRGRYGAA-3′) and 1204R (5′-TTMGGGGCATRCNKACCT-

170 3′) [54], respectively, using previously described thermal cycling conditions [55, 56]. For all

171 samples, products from three separate reactions were pooled and analyzed on a 0.8%

172 PCR grade agarose gel. The correctly sized bands were gel extracted and purified via

173 Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) and DNA

174 was quantified with a Qubit fluorometer (Invitrogen, Grand Island, NY, USA). Adaptors

175 for 454 sequencing were ligated to the amplicons and were pyrosequenced on a 454 GS-

176 Junior FLX Titanium™ (454 Life Sciences, Branford, CT, USA). An in-house python

177 script [57] was used for data management and analyses. In brief, the sequences were

178 trimmed to 250 bp, and subjected to Q cutoff at 27% while allowing 15% of the

179 nucleotides to be below the cutoff. Sequences with primer errors or ambiguous

180 nucleotides were also removed. Chimeric sequences were identified and removed using

181 ChimeraSlayer [58]. Taxonomic sequence identification to the genus-level was performed

182 by comparison with the RDP classifier[59] with a confidence threshold cut-off set at 80% .

183 This cut-off was chosen based on the short read length (250 bp) and to minimize the

184 number of unclassified sequences[60]. The denoised and non-chimeric sequences were

185 annotated for *hgcAB*$^{+}$ microorganisms with a cut-off of 97% using BLASTN against the

186 Greengenes 99% OTUs database [61] amended with the exact sequences for all confirmed

187 Hg methylators (Table S1). This cut-off (97%) was chosen to ensure species-level

188 identification of 16S rRNA of known and potential Hg-methylators.

189 Metagenomic sequencing. The eight gDNA extracts were prepared for metagenomic

190 sequencing with the Nextera® XT Sample Prep Kit (Illumina®, La Jolla, CA) according to

191 the manufacturer's instructions through library validation. Samples were tagmented,

192 barcoded, and amplified. Libraries were cleaned with Agencourt AMPure XP beads

193 (Beckman Coulter, Indianapolis, IN). Final libraries were validated with a 2100

194 Bioanalyzer (Agilent) and concentrations measured by Qubit® (Thermo Fisher

195 Scientific). Libraries were prepared for sequencing following the manufacturer's

196 protocols and the Illumina library dilution calculator. The libraries were denatured with

197 0.2 N NaOH and diluted to 15-18 pM. Libraries were loaded into the sequencing

198 cassette (v3) and a paired end (2 x 300) run was completed on an Illumina® MiSeq®

199 System.

200 Fastq files were generated for the forward and reverse reads. Each fastq file was filtered

201 to remove low-quality nucleotide base calling scores at the read ends following the

202 default settings with the software package Extended Randomized Numerical alignEr

203 (ERNE)-filter [62]. Reads from each sample were paired using PEAR [63] with a combined

204 read maximum of 600 bp and minimum of 50 bp. Finally, the nucleotide coded fasta files

205 were translated to amino acids for all possible open reading frames following the transeq

206 default settings [64]. The trimmed reads were assembled into longer contigs using

207 metagenome assemblers MetaVelvet[65] and IDBA-UD[66] using kmer parameters

208 optimized for long sequence reads (300 bp). Quality assessment of the assembled

209 datasets were poor (N_{50} < 600 bp), indicating that sequencing depth was insufficient for

210 sample complexity. Therefore, the PEAR joined reads were used for protein prediction

211 analyses. 16S rRNA genes were identified in the PEAR joined reads using MG-RAST

212 annotations[67], and were annotated for *hgcAB*$^{+}$ microorganisms by BLASTN against the

213 amended Greengenes 99% OTUs database described above.

214 ***Pfam approach for hgcA, mer operon, mcrA and dsrC***

215 Metagenomic data were translated into amino acid sequences using the Pfam database

216 [68], and HgcA and HgcB were identified using custom Pfam models of known HgcA and

217 HgcB sequences [46]. Pfam models of amino acid sequences use protein homology in

218 combination with hidden Markov model (HMM) search algorithms to achieve a higher

219 specificity search for homologous proteins relative to a BLAST search of nucleotide

220 sequences since it eliminates non-HgcA sequences that are within the CFeSP protein

221 family [46]. We also used Pfam models for DsrC (a well conserved gene required for

222 sulfate-reduction), the C-terminal domain of McrA (a well conserved gene required for

223 methanogenesis) and several components of the *mer* operon (a Hg resistance system, [29]),

224 including MerB, two separate orthologs of the *mer* operon regulator MerR (MerR and

225 MerR-HTH), and MerA. The MerR family of transcriptional regulators contains

226 regulators that may be induced by other stressors besides Hg(II), including oxidative

227 stress, other heavy metals, or antibiotics [69]. This broader regulatory family can be mis-

228 annotated in genomes and not all MerR-like proteins are involved in Hg

229 transformations[69]. The MerA active domain is shared across heavy metal resistance

230 proteins and the Pfam model queries against the more generalized heavy-metal-

231 associated (HMA) Pfam. All reads from each metagenome (as amino acids) were queried

232 against each Pfam model using hmmsearch [70] with an Eigen value cutoff of $1x10^{-5}$. For

233 each protein, amino acid sequence outputs were combined with the guide sequences

234 from the Pfam models, aligned using MUSCLE [71] and a phylogenetic tree calculated

235 using FastTree [72]. The tree generated for HgcA contained nodes with a large number of

236 sequences (>50) that lacked a guide sequence and so such nodes were excluded from

237 further analysis and comparisons. The resulting sequence sets underwent a final check so

238 that unpaired forward and reverse reads were only counted once. The HgcA

239 phylogenetic tree in Figure 4 was constructed using hmmsearch [73] and the GAMMA

240 model in MEGA [74].

241 ***PCR (hgcAB) and qPCR (hgcA) analysis for diversity and abundance of Hg-methylators***

242 Hg-methylator diversity was also assessed via cloning and sequencing of *hgcAB* PCR

243 amplicons. Amplicons of ~950 bp were generated using previously described degenerate

244 universal *hgcAB* primers [43] with binding locations in the highly conserved active sites of

245 both genes. Reactions used 10 ng community gDNA per 20 µl. We also used an updated

246 reverse primer (ORNL-HgcAB-uni-R2: CAGGCNCCGCAYTCSATRCA). Amplified

247 products were electrophoresed (90 V, 90 min) through a 1% (wt/vol) agarose gel and

248 the region spanning 800–1100 bp was excised, cleaned with the Wizard® SV Gel and

249 PCR Clean-Up Kit (Promega, Madison, WI) and cloned (TA Cloning® Kit, Thermo

250 Fisher Scientific). One-hundred colonies per sample were picked and screened for

251 appropriately sized inserts using primers M13 Forward (-20) and M13 Reverse

252 (Integrated DNA Technologies). Cultures from all positive screened colonies were

253 grown (37°C in 1 mL LB broth with 50 µg mL^{-1} Kanamycin) and the plasmid extracted

254 with the GeneJET™ Plasmid Miniprep Kit (Thermo Fisher Scientific). Each plasmid

255 was sequenced on both strands with M13 F/R primers following the Sanger protocol at

256 the University of Tennessee Molecular Biology Resource Facility.

257 To determine the abundance of the clade-specific Hg-methylators, qPCR-based hgcA

258 degenerate primers were used [43]. These primers were developed to selectively amplify

259 *hgcA*$^{+}$ organisms in each of the three major clades that contain Hg-methylators.

260 Reactions were performed with 5 ng of gDNA on a C1000 Touch™ Real-Time PCR

261 Detection System (Bio-Rad, Hurcules, CA). Data were analyzed with CFX Manager™

262 (version 3.1, Bio-Rad) with the curve fit to regression [75-77]. Total *hgcA* copy number was

263 determined by comparison to a standard curve (10-fold serial dilutions ranging from 1 x

264 10^6 – 1 x 10^2 genome copies per reaction) generated from gDNA isolated from

265 representative strains *Geobacter sulfurreducens* PCA DSM-12127

266 (*Deltaproteobacteria*), *Desulfosporosinus youngiae* JW/YJL-B18 DSM-17734

267 (*Firmicutes*), and *Methanomethylovorans hollandica* DSM-15978 (methanogenic

268 *Archaea*). The qPCR products from all samples were subject to melt-curve analysis to

269 evaluate potential primer-dimer formation, and amplification of non-specific products.

270 Select qPCR products were also evaluated using agarose gels to check amplicon sizes.

271 The qPCR copy numbers were normalized to genome equivalents for comparison with

272 *hgcA* counts from metagenomic datasets. If the average microbial genome is assumed to

273 be ~5.00 Mbp[78], then 1 ng of DNA by mass would contain ~ 1.85E05 genome

274 equivalents (Table 1).

275

276 ***Mass spectrometry shotgun metaproteomics***

277 All proteomic sample preparation and tandem mass spectra of the resulting peptides

278 were generated at the Environmental Molecular Sciences Laboratory, located at the

279 Pacific Northwest National Laboratory (Richland, WA). Sediment samples (20 g) were

280 treated with binder solution (8 mL; 10% (w:v) arginine and lysine in 1M MES buffer,

281 pH~7) to minimize sediment adsorption and incubated overnight (37°C, 200 rpm

282 shaking). Two grams were bead beaten in desorption buffer (1 ml; 1% CHAPS, 20mM

283 EGTA, 0.1M raffinose, 0.25M NaCl, 1M betaine, 0.5% morpholine in 100mM

284 ammonium bicarbonate, pH 8.5) for 5 min. Samples were centrifuged (15 mL tubes,

285 3,000 rpm, 5 min, 4°C), sediment removed, and sonicated/washed in a 50 mL falcon

286 tube with desorption buffer (6mL), centrifuged (4000 rpm, 5 min, 4°C) and the

287 supernatant collected. Washing was repeated until 15 mL of supernatant was collected.

288 The supernatant was centrifuged (5,000 rpm, 5 min, 4°C) and poured into a 15 mL/3K

289 MWCO spin column (Pierce) and centrifuged to the dead volume. A buffer exchange

290 was performed with 100 mM NH_4HCO_3 three times and the columns centrifuged (4,000

291 x g, 45 min, 4°C).

292 The sample was transferred to a 2.0-mL microcentrifuge tube, the volume determined

293 and powdered urea added to 8 M. Dithiothreitol (500 mM) was added to a final

294 concentration of 5 mM. Samples were incubated (60°C, 30 min) and diluted 10-fold with

295 100 mM NH_4HCO_3. $CaCl_2$ was added to 1 mM and digested (3 hrs) with Trypsin (37°C,

296 1 unit Trypsin/50 units protein) and placed on ice. The sample was cleaned with an C18

297 SPE column (Empore, Sigma Aldrich) (1 mL/100 mg tube was sufficient up to 5 mg of

298 protein). The eluate was concentrated to ~50 uL and the protein concentration

299 determined using a bicinchoninic acid (BCA) kit (Thermofisher).

300 Tandem mass spectra of the resulting peptides were generated according to previously

301 described methods [79]. In brief, samples were randomized for analysis and peptide

302 measurements made using a high-resolution LTQ Orbitrap Velos mass spectrometer

303 (Thermo Scientific, San Jose, CA) via ionization using a single emitter nano-electrospray

304 interface (ESI) coupled to an ISCO online reverse phase HPLC system. The ion transfer

305 tube temperature and spray voltage were kept at 275°C and 2.2 kV respectively. The LC

306 column consisted of 3-μm Jupiter C_{18} bonded particles (Phenomenex, Torrence, CA)

307 slurry packed into a 65-cm long, 75-μm inner diameter fused silica capillary (Polymicro

308 Technologies, Phoenix, AZ).

309 The analytical column was equilibrated with 99% mobile phase A (0.1% formic acid) and

310 1% B (0.1% formic acid in 100% acetonitrile) at 0.3 mL/min prior to sample injection

311 (0.1 μg/μL) onto a 5 μL loop. Sample flow was directed to a trapping column at

312 5 μL/minute for 2 minutes followed by transfer to the analytical column through a

313 switching valve, initiating the chromatographic separation of sample peptides using the

314 gradient profile (min:%B): 0:1, 2:8, 20:12, 75:30, 97:45, 100:95. Acquisition of Orbitrap

315 spectra (AGC $1x10^6$) was triggered 20 minutes after sample injection from 400-2000

316 m/z at a resolution of 30K. The 10 most abundant ions were selected for data dependent

317 ion trap collision-induced dissociation (CID) MS/MS spectra using a 3.0 Da isolation

318 width, 35% normalized collision energy and 30 ms activation time. A dynamic exclusion

319 time of 60 sec was used to discriminate against previously analyzed ions (within -0.60

320 and 1.60 Da).

321 LC-MS/MS raw data were converted into .dta files using Bioworks Cluster 3.2 (Thermo

322 Fisher Scientific, Cambridge, MA), and searched using the MSGF+ algorithm [80] in

323 concert with a translated metagenomic database consisting of 30,628 entries. A search

324 for partial to full-tryptic peptides was included in a standard set of search parameters

325 including a dynamic modification of methionine (oxidation; 15.9949) and a mass error

326 window of 20 ppm for precursor mass spectra. Missed cleavages were allowed. Spectral

327 probabilities were assigned using MSGF+ and only peptides at least 6 amino acid

328 residues in length with a spectral probability of less than $1x10^{-10}$ were retained. A protein

329 was considered positively observed if at least two unique peptides sequences were

330 matched to tandem mass spectra. The resulting sequences were annotated for *hgcAB*$^{+}$

331 microorganisms using BLASTP against a custom database containing HgcAB sequences

332 for all confirmed and potential Hg methylators (Table S1).

333

334 ***THg and MeHg Determination***

335 HgT and MeHg concentrations were determined using modifications of EPA methods

336 1630 and 1631, with the details described previously [21, 50, 81, 82]. Lab-specific

337 modifications, necessary equipment, and laboratory QC summaries are described

338 elsewhere [21, 83]. Detection limits for total Hg and MeHg in soils were about 0.1 and 0.05

339 ng/g dry weight, respectively.

340 ***Data Availability***

341 Metagenomic data are available from MG-RAST (https://www.mg-rast.org/) and at

342 KBase (https://kbase.us/). MG-RAST dataset accession numbers are

343 KirkPatrickSERC_Paired mgs269152, NOAA_Paired mgs269161,

344 PenobscotW21UM_Paired mgs269167, PenobscottW10MM_Paired mgs269164,

345 New_Horizon_Paired mgs269155, HC_Paired mgs269170, KirkpatrickFringe_Paired

346 mgs269149, NGA73_Paired mgs269158. The datasets at KBase are located under

347 ws.36658.obj.1 (https://narrative.kbase.us/narrative/ws.36658.obj.1).

348 RESULTS AND DISCUSSION

349 *Comparison among Hg-methylator identification methods*

350 Rapid, inexpensive and accurate estimates for Hg-methylator abundance and diversity

351 would facilitate research on the links between microbial community structure and MeHg

352 production. Being able to accurately predict the environmental conditions as well as the

353 microbial factors that facilitate MeHg production are needed in order to develop

354 methodologies aimed at curbing MeHg production and accumulation, thus facilitating

355 remediation at the ecosystem-scale. Current biogeochemical models for net MeHg

356 production are generally weak predictors, especially when used across ecosystems [83-89].

357 This is partially because no model has factored in the abundance or activity of *hgcAB*$^{+}$

358 organisms (or of demethylating organisms). One of our goals is to evaluate whether

359 inclusion of the biological information would improve the prediction of net MeHg

360 accumulation in ecosystems. To do so, one of the first steps is to identify reliable

361 methods for determining Hg-methylator abundance and diversity. While shotgun

362 metagenomic sequencing of environmental DNA may be considered the gold standard
363 for community genetic structure (given that all genes are sequenced), it remains costly,
364 labor intensive and requires specialized bioinformatics skills. Additionally, it remains
365 difficult to evaluate biases in metagenomic data (experimental and bioinformatic) that
366 confound identification and quantification of individual genes [90]. A less costly approach
367 for identification of *hgcAB* gene sequences uses gene-specific primers to generate
368 amplicons for cloning and sequencing, as previously described [43]. Gene-specific primers
369 can also provide a potentially more quantitative method for evaluating *hgcA* copy
370 number in natural samples using qPCR.

371 In this study we compared these different techniques (16S rRNA-based phylogeny,
372 metagenomic sequencing, and gene-specific PCR/qPCR) to evaluate their relative ability
373 to quantify Hg-methylator abundance and diversity in eight natural sediment and soil
374 samples from diverse settings.

375 Amplicon-specific 16S rRNA pyrosequencing identified fewer than 15 total sequences
376 per sample as 16S rRNA genes from potential methylators; equating to only a small
377 fraction (<0.55E-6) of total reads (Table 1) from 866 Bacterial and 92 Archaeal OTUs

378 per ng of DNA. Identifications were made based on the current Greengenes 16S rRNA

379 database, amended with 16S rRNA sequences from all additional sequenced Hg-

380 methylating species. Hg-methylating organisms are commonly found in anaerobic niches

381 in nature, but even there they are relatively rare with, on average, ~0.5% of the genomes

382 from sediments and wetland soils containing *hgcAB* [46]. The depth of 16S rRNA

383 pyrosequencing and the overall inability to identify organisms beyond family or genus

384 level did not adequately allow for positive identification of Hg-methylators since some

385 members of a given genus possess *hgcAB* and others do not, which is consistent with

386 previous work [21].

387 Conversely, high-throughput shotgun metagenomes provided deeper sequencing and a

388 wealth of information on the genetic potential for biochemical activities including Hg

389 cycling and related metabolic activity (e.g. *merA*, *merB*, *dsrC*, *mcrA*; Table 2). Overall,

390 paired-end Illumina MiSeq® shotgun metagenomic analyses of our samples generated an

391 average 15 Gbp and 5.3×10^7 reads per sample (Table S2). We extracted the 16S rRNA

392 sequences from the metagenomes, and evaluated the potential for identifying Hg-

393 methylators, mimicking the information that would be obtained from high-throughput

394 16S rRNA amplicon-specific sequencing. Of all 16S rRNA genes identified in the

395 metagenomic read sets, only a fraction (7.54E-4 – 7.61E-3) could be putatively identified

396 as Hg-methylators using the same process used for the 16S rRNA pyrosequencing (Table

397 1). We also evaluated *hgcA* and *hgcB* sequences in the metagenomes, using a custom

398 Pfam model of known HgcA and HgcB sequences, after translating metagenomic data

399 into amino acid sequences [46]. The Pfam/HMM model identifies the corrinoid and

400 transmembrane binding domains of HgcA and the two [4Fe-4S] cluster-binding motifs

401 in HgcB, allowing specific identification even in previously unidentified organisms

402 (unlike 16S rRNA identifications). By this approach, 381–1,021 HgcA and 1,101–3,232

403 HgcB homologs were identified in metagenomic reads, equating to 0.23–0.82 and 0.80–

404 2.47 copies of *hgcA* and *hgcB* per genome equivalent (Table 2). The number of genome

405 equivalents encompassed from the metagenomic sequencing was calculated from

406 dividing the size of the paired metagenomic read set (Gbp) by the average size of a

407 microbial genome (0.005 Gbp)[78]. The higher abundance of potential *hgcB* genes

408 identified in metagenomes compared to *hgcA*, is likely due to the shared homology of

409 HgcB with other 4Fe-4S ferredoxins. Even with a threshold cut-off of 1E-5, in short read

410 sequences (<600 bp) without the presence of *hgcA* on the contig, this shared homology

411 makes it difficult to confidently differentiate *hgcB* from other ferredoxin-encoding

412 genes. Conversely, HgcA can be confidently differentiated from paralog proteins, carbon

413 monoxide dehydrogenase and acetyl-coA synthase, from non-Hg-methylating bacteria

414 through phylogenetic analyses. Although both genes are needed to encode Hg-

415 methylation capability, only the metagenomic *hgcA* values were used for estimating Hg-

416 methylator abundance and diversity. The clade-distribution of Hg-methylators using

417 Pfam was similar to Hg-methylator identification based on 16S rRNA reads, both derived

418 from metagenomics (Figure 2), with *Deltaproteobacteria* dominating most of the

419 samples, but with a higher diversity. Methanogens and *Firmicutes* were better

420 represented in the metagenomic *hgcA* Pfam data and so using the 16S rRNA genes

421 derived from the metagenomics may have over-estimated the *Deltaproteobacteria*. The

422 diversity observed in the metagenomic *hgcA* data from these sites was generally

423 consistent with published metagenomes from diverse habitats (Figures 3,4) [46].

424 Last, *hgcAB* diversity was evaluated from clone libraries. Clones were produced using

425 degenerate universal PCR primers with binding locations in the highly conserved active

426 sites of both genes [43]. About 100 clones from each sample were sequenced (Table 1).

427 Among the eight sites, *hgcAB* amplicon diversity was highly similar to the metagenome

428 *hgcA* data sets (Figure 2). Clone libraries also estimated a broader depth in Hg-

429 methylator diversity than did 16S rRNA sequences, with higher representation of

430 *Firmicutes* and *Methanomicrobia* in *hgcAB* clones libraries compared to 16S rRNA

431 analyses. Although the sequencing depth from clone libraries was much less than in the

432 metagenomes, the clone libraries provide highly accurate *hgcAB* identification because

433 of the long reads (~950 bp), and because both genes are sequenced.

434 For each of the sequencing approaches, *Deltaproteobacteria* were the most abundantly

435 identified Hg methylators at all sites (~40–70%), except for Hinds Creek where the

436 methanogenic *Archaea* were more abundant. *Firmicutes* were slightly more abundant

437 overall in the metagenomes than the clone libraries (~10–40%) while the methanogenic

438 *Archaea* (~0–40%) were similar. Unidentified or "other" sequences comprised <10% of

439 *hgcA* sequences by either method (Table 1, Figures 2,4).

440 To summarize, we found that community structure information based on 16S rRNA

441 gene amplicons were insufficient to adequately identify Hg-methylator diversity. Clade-

442 level diversity estimates based on 16S rRNA were quite different from either of the

443 approaches that involved direct *hgcAB* sequencing. The use of 16S rRNA data alone to

444 identify Hg-methylators is problematic because a) Hg-methylation is a strain-specific

445 trait, i.e. *hgcAB*$^{+}$ strains within groups of closely related organisms (even species) may

446 or may not encode *hgcAB*, b) species identification based on 16S is inherently difficult

447 even with OTU calling at 97%, and c) the database of 16S sequences for Hg methylators

448 remains small (<300 species). Translated *hgcA* sequences from metagenomic data

449 provided very similar diversity information to clone libraries. Overall, assessing the Hg-

450 methylating community by identifying *hgcAB* from metagenomic or PCR-based methods

451 showed highly similar results. We conclude that direct *hgcAB* sequencing is required to

452 accurately identify Hg-methylating organisms.

453 ***Quantitative determination of Hg-methylators***

454 To more accurately quantify the Hg-methylator community, *hgcA* qPCR was conducted

455 for each of the three major Hg-methylator clades at all sites (Table 1). The qPCR

456 estimates were compared to metagenome derived estimates for Hg-methylator

457 abundance (Figure S1). While both methods have the potential to be used quantitatively,

458 they are both subject to experimental and computational biases [91, 92]. Steps were taken to

459 minimize these concerns for qPCR by including a large standard dynamic range ($1x10^6$–

460 10^0 genome copies) for standard curves and use of low template concentration to dilute

461 inhibitors and limit non-specific and non-clade specific amplification. Both methods

462 were carried out using the same DNA extracts.

463 The clade-specific qPCR primers yielded information on *hgcA* clade abundance that was

464 quite similar to metagenome data (Figure S1). For *Deltaproteobacteria*, methanogens,

465 and total *hgcA* reads, abundances from both methods were significantly correlated

466 (r^2=0.67, p =0.013) across sites, after normalization for the mass of DNA extracted per g

467 of sediment from each sample. *Firmicute* total counts via qPCR were low across all

468 samples (<~100 *hgcA* copies per ng DNA). This may be due in part to the *Firmicute*

469 qPCR protocol not targeting fermentative Hg-methylating *Firmicutes* [43]. The exception

470 was Arctic permafrost which showed a markedly higher *Firmicute hgcA* abundance than

471 all other samples tested and was consistent in both the qPCR and metagenomics. The

472 qPCR data provides a direct estimate of *hgcA* copy number in sediments, while the

473 metagenomic data have undergone an amplification step during library preparation, so

474 the absolute magnitude of copy numbers is not comparable.

475 ***Comparison of hgcA to key metabolic genes.*** The metagenomic data allow for a rough

476 comparison of *hgcA* abundance to the abundance of key metabolic genes in major clades

477 of Hg-methylators, and to the abundance of potential MeHg-demethylating

478 microorganisms. The genes evaluated included *dsrC* targeting sulfate-reducers;[37, 93],

479 *mcrA* (targeting methanogens) [94] and several components of the *mer* Hg resistance

480 system, [29], including *merB* (which encodes for organomercurial lyase; [32, 95], two separate

481 orthologs of the *mer* operon regulator *merR* (*merR* and *merR-HTH*), and *merA*

482 (mercury reductase). For most sites, the *dsrC* copy number exceeded the *hgcA* copy

483 number at least ten-fold (Figure S2), highlighting the fact that only a small percentage of

484 sulfate-reducing organisms encode *hgcA*. A key methanogen gene, *mcrA*, was found in

485 5-10 fold less abundance than *hgcA.* Further, the *hgcA* gene copy number measured

486 using the clade-specific qPCR primers reflected the abundance of the marker genes for

487 each clade in the samples (Figure S3). Similarly, Bravo et al. [42] found a correlation

488 between the transcript abundance of *dsrC* and *hgcA* in Hg-contaminated river

489 sediments.

490 The Hg reductase (*merA*) and mer operon regulator (*merR*) genes were found in about

491 equal copy numbers, both about ten-times as abundant as the methylmercury lyase gene

492 *merB* (Figure S2). The abundances of all three genes were co-correlated as would be

493 expected since they are in the same operon (Figure S4). Total Hg concentration was

494 predictive of the *merR* gene copy number (Figure S5). Interestingly, *merA* and *merR*

495 were much more abundant than marker genes for sulfate-reduction and methanogenesis,

496 but *merB,* the methylmercury lyase, was found in roughly equal copy numbers with

497 *hgcA*, the MeHg production gene (Figure S2).

498 The abundance of *mer* genes was not correlated with *hgcA* abundance (Figure S5), nor

499 with the abundance of marker genes for methanogenesis or sulfate-reduction (Figure

500 S6). The abundance of *merB* ranged from ~1% (permafrost) to 10% (Tennessee

501 freshwater sediments) of *merA*. However, the ratio of *merB* to *merA*, which might

502 reasonably predict the relative level of MeHg demethylation compared to Hg(II)

503 reduction, was not predictive of MeHg concentration or %MeHg in samples (Figure S4).

504 Last, we speculated that the ratio of *merB* to *hgcA* could be predictive of the net percent

505 of total Hg methylated in samples (%MeHg), but it was not (Figure S5).

506 ***Gene and protein abundance considerations***

507 Predicting MeHg generation from genomic information is difficult in part because the

508 presence of a gene conveys the genomic potential for a given activity, but does not

509 ensure a functional protein. Hence, we employed metaproteomics on the EFPC and

510 Hinds Creek, TN sites for direct comparison to clade-specific metagenomic abundance

511 estimates and trends against HgT and MeHg concentrations, and %MeHg (Table 1,

512 Figures S7, S8).

513 As expected, *Deltaproteobacteria* were the most abundant followed by *Firmicutes* and

514 *Archaea* (Figure S7), suggesting again that the *Deltaproteobacteria* were likely the

515 primary Hg-methylators at this site as previously described [47, 48]. The number of HgcA

516 proteins per g of sediment averaged 20X higher than the qPCR-based copy number per g

517 sediment for *Deltaproteobacteria,* 90X higher for *Firmicutes*, and only ~1.5X higher for

518 methanogens. This might be an indication of the relative expression among clades,

519 suggesting that *hgcA* was not as highly expressed by methanogens as by

520 *Deltaproteobacteria* in these samples. It may also reflect the incomplete phylogenetic

521 coverage of the *Firmicute* qPCR primer.

522 Expression of *hgcA* into proteins was relatively constant regardless HgT, MeHg or

523 %MeHg (Figure S8) across the three Tennessee creek sites, which comprise a large

524 gradient of total Hg concentration (Figure S8). It is unclear how *hgcA* protein expression

525 is maintained or regulated, however it is not likely to be induced by Hg concentrations

526 tested, and expression is possibly constitutive since it is likely involved in C1 metabolism

527 [46]. This is corroborated given the one known study on *hgcAB* expression where

528 *Desulfovibrio dechloroacetivorans* BerOc1 was grown in various conditions followed by

529 reverse-transcription PCR (RT-PCR) analysis with negligible change in *hgcAB* mRNA

530 even though Hg-methylation varied greatly [38], agreeing with our proteomics findings.

531 ***Relating gene abundance to THg and MeHg concentrations***

532 Abundance estimates of *hgcA* were compared to THg and MeHg concentration, and

533 %MeHg (Figures 3,4). No correlation was observed between copy number estimates

534 using either qPCR or metagenomic-based methods with sediment total Hg concentration

535 (Figure 3). If *hgcA* gene copy number could be used as a proxy for Hg-methylator

536 abundance and in turn the methylation capacity of the microbial community, then a

537 positive correlation between *hgcA* and MeHg production would be expected.

538 Interestingly, *hgcA* copy number declined with MeHg for total *hgcA* copies derived from

539 metagenomic data, and for *Deltaproteobacteria hgcA* estimated from either method,

540 with no correlation observed across the eight environmental samples, regardless of

541 method, between *hgcA* abundance and %MeHg. These relationships were not changed

542 or improved when *hgcA* copy numbers were normalized to copies per g of sediment.

543 Other biogeochemical controls on Hg methylation besides Hg-methylator abundance

544 include Hg-speciation, microbial activity, and Hg-methylator diversity.

545 Not all *hgcAB*+ organisms methylate Hg at the same rate[6,63], therefore Hg-methylator

546 diversity and not just abundance may be an important control on MeHg generation.

547 Phylogenetic analyses of *hgcA* genes pulled from metagenomic datasets revealed a

548 breadth in diversity of potential Hg-methylating organisms across the eight sites (Figure

549 4). Taxonomic classifications were assigned to *hgcA* reads from the metagenomic

550 datasets based on phylogenetic distances to *hgcA* genes from known and predicted Hg-

551 methylators (Table S1), resulting in 77 distinct OTUs, 11 of which were closely related

552 to known methylators (Figure 4). The *hgcA* genes related to known and potential Hg-

553 methylators from *Deltaproteobacteria*, *Firmicutes*, *Euryarchaeota*, *Lentisphaerae*,

554 *Bacteroidetes*, *Chloroflexi*, *Spirochaetes*, *Chlorobi*, and *Nitrospirae* were present in the

555 metagenomes from all eight sites. Within these clades, a large proportion (6–16%) of

556 *hgcA* genes could be classified as *Geobacter*, *Desulfovibrio*, *Dehalococcoides*,

557 *Methanoregula,* or *Chloroflexi bacterium RIFOXYD12_Full_57_15* (Figure 4). The *hgcA*

558 genes related to *Elusimicrobia*, and Candidatus phyla *Atribacteria* and *Aminicenantes*

559 were less prevalent in the metagenomes. *Deltaproteobacterial hgcA* were the dominant

560 OTU in all metagenomes except Arctic permafrost (NGA73), where the most abundant

561 OTUs were related to *Bacteroidetes* and *Elusimicrobia*. The OTUs related to *Nitrospirae*

562 and *Chloroflexi* were also prevalent in EFPC and Penobscot metagenomes. A large

563 proportion (4–22%) of *hgcA* genes in all eight metagenomes could not be classified

564 (Figure 4), indicating that a large proportion of potential Hg-methylators in the

565 environment are not closely related to organisms in culture or genomic sequencing

566 collections. Furthermore, only 10% of *hgcAB*+ organisms have had their methylation

567 capability assessed (Table S1). This represents a significant gap in our understanding

568 between the potential of a community to methylate Hg and the production of MeHg.

569 A lack of correlation between the number of *hgcA* (and *merA*) genes in sediment or soil,

570 and HgT, MeHg, or %MeHg levels across diverse environments is perhaps unsurprising,

571 given the known influence of geochemistry on Hg and MeHg bioavailability for

572 methylation and demethylation. Within constrained conditions, correlations have been

573 observed between *hgcA* gene copy number and MeHg concentration [15, 42, 96]. Bravo et al.

574 [42] found no relationship between *hgcA* transcripts on sediment MeHg or %MeHg

575 content in Hg-contaminated river sediments. They attributed the lack of predictive

576 ability of *hgcA* to changing Hg bioavailability across their sampling gradient. For our

577 sites where multiple samples were taken within an ecosystem, the trends between *hgcA*

578 copy number and MeHg or %MeHg varied in direction (Figure S9). Using

579 metagenomics or PCR to identify and quantify *hgcA* locally is reasonable to assess the

580 Hg-methylator community, but not necessarily to determine Hg methylation potential

581 without additional information, including gene expression levels [97], abundance of HgcA

582 and HgcB proteins, Hg-bioavailability and hydrobiogeochemical properties [20]. While

583 metagenomics yielded the highest fidelity and simultaneously captured Hg-methylator
584 and demethylator diversity, to-date the cost for processing hundreds of samples for
585 metagenomics is still cost prohibitive, particularly at depths high enough to capture the
586 diversity of Hg-methylators. Our comparison of *hgcAB* PCR and metagenomics to assess
587 Hg-methylator diversity yielded similar results, suggesting that environmental studies
588 targeting MeHg production can be enhanced by the low-cost *hgcAB* PCR and *hgcA*
589 qPCR. To this end, the advent of reliable techniques for direct sequencing of *hgcAB*
590 amplicons would be most desirable to eliminate the laborious cloning steps. However,
591 when combined with selective metagenomics, a powerful and comprehensive picture of
592 Hg-methylation potential can emerge. The inclusion of geochemical and functional
593 (meta- transcriptomics and proteomics) data is likely required to gain a comprehensive
594 understanding of these relationships for predicting net MeHg activity. While the
595 benchmarking of this new tool is an advantage, more comprehensive characterizations of
596 HgcAB activity level and bioavailability studies may allow for the elucidation the critical
597 factors controlling Hg-methylation.

598 **Supplemental Information**

599 The Supporting Information is available free of charge on the ACS Publications website at

600 http://pubs.acs.org and contains figures S1-S9 detailing the relationship between PCR/qPCR and

601 metagenomic sequencing.

Acknowledgments

We thank R. W. Clauss for metaproteomics data generation and analysis, and Ally Soren and Alyssa McBurney for laboratory support. This research was sponsored in part by the Office of Biological and Environmental Research, Office of Science, US Department of Energy (DOE) as part of the Mercury Science Focus Area at Oak Ridge National Laboratory, which is managed by UT-Battelle LLC for the DOE under contract DE-AC05-00OR22725. A portion this research was performed using EMSL (Ringgold ID 130367), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

613 REFERENCES

614 1. UNEP *UNEP Global Mercury Assessment 2013: Sources, Emissions, Releases and Environmental*
615 *Transport*; United Nations Environment Programme Chemicals Branch: Geneva, Switzerland, 2013.
616 2. Parks, J. M.; Johs, A.; Podar, M.; Bridou, R.; Hurt, R. A., Jr.; Smith, S. D.; Tomanicek, S. J.; Qian, Y.;
617 Brown, S. D.; Brandt, C. C.; Palumbo, A. V.; Smith, J. C.; Wall, J. D.; Elias, D. A.; Liang, L., The genetic basis
618 for bacterial mercury methylation. *Science* **2013,** *339*, (6125), 1332-1335.
619 3. Miskimmin, B. M.; Rudd, J. W.; Kelly, C. A., Influence of dissolved organic carbon, pH, and
620 microbial respiration rates on mercury methylation and demethylation in lake water. *Canadian Journal*
621 *of Fisheries and Aquatic Sciences* **1992,** *49*, (1), 17-22.
622 4. Eddy, S. R., A new generation of homology search tools based on probabilistic inference. In
623 *Genome Informatics 2009*, Published by Imperial College Press and distributed by World Scientific
624 Publishing Co.: 2012; pp 205-211.
625 5. Compeau, G. C.; Bartha, R., Sulfate-reducing bacteria: principal methylaors of mercury in anoxic
626 estuarine sediment. *Applied and Environmental Microbiology* **1985,** *50*, (2), 498-502.
627 6. Yu, R.-Q.; Reinfelder, J. R.; Hines, M. E.; Barkay, T., Mercury methylation by the methanogen
628 Methanospirillum hungatei. *Applied and Environmental Microbiology* **2013,** *79*, (20), 6325-6330.
629 7. Gilmour, C. C.; Podar, M.; Bullock, A. L.; Graham, A. M.; Brown, S. D.; Somenahally, A. C.; Johs,
630 A.; R. A. Hurt, J.; Bailey, K. L.; Elias, D. A., Mercury methylation by novel microorganisms from new
631 environments. *Environ. Sci. Technol.* **2013,** *47*, (20), 11810-11820.
632 8. Hecky, R. E.; Ramsey, D. J.; Bodaly, R. A.; Strange, N. E., Increased Methylmercury Contamination
633 in Fish in Newly Formed Fresh-Water Reservoirs. *Roch S Env* **1991**, 33-52.
634 9. Parks, J. W.; Lutz, A.; Sutton, J. A., Water Column Methylmercury in the Wabigoon-English River-
635 Lake System - Factors Controlling Concentrations, Speciation, and Net Production. *Can J Fish Aquat Sci*
636 **1989,** *46*, (12), 2184-2202.
637 10. Rudd, J. W. M.; Turner, M. A.; Furutani, A.; Swick, A. L.; Townsend, B. E., The English Wabigoon
638 River System .1. A Synthesis of Recent Research with a View Towards Mercury Amelioration. *Can J Fish*
639 *Aquat Sci* **1983,** *40*, (12), 2206-2217.
640 11. Kelly, C. A.; Rudd, J. W. M.; Louis, V. L.; Heyes, A., Is Total Mercury Concentration a Good
641 Predictor of Methyl Mercury Concentration in Aquatic Systems. *Water Air and Soil Pollution* **1995,** *80*, (1-
642 4), 715-724.
643 12. Rothenberg, S. E.; Ambrose, R. F.; Jay, J. A., Mercury cycling in surface water, pore water and
644 sediments of Mugu Lagoon, CA, USA. *Environmental pollution* **2008,** *154*, (1), 32-45.
645 13. Schaefer, J. K.; Kronberg, R. M.; Morel, F. M. M.; Skyllberg, U., Detection of a key Hg methylation
646 gene, hgcA, in wetland soils. *Environ. Microbiol. Rep.* **2014,** *6*, (5), 441-447.
647 14. Bae, H.; Dierberg, F. E.; Ogram, A., Syntrophs dominate sequences associated with the mercury
648 methylation-related gene hgcA in the water conservation areas of the Florida everglades. *Appl. Environ.*
649 *Microbiol.* **2014,** *80*, (20), 6517-6526.
650 15. Liu, Y.; Yu, R.; Zheng, Y.; He, J., Analysis of the microbial community structure by monitoring an
651 Hg methylation gene (hgcA) in paddy soils along an Hg gradient *Appl. Environ. Microbiol.* **2014,** *80*, (9),
652 2874-2879.
653 16. Horvat, M.; Nolde, N.; Fajon, V.; Jereb, V.; Logar, M.; Lojen, S.; Jacimovic, R.; Falnoga, I.; Liya, Q.;
654 Faganeli, J.; Drobne, D., Total mercury, methylmercury and selenium in mercury polluted areas in the
655 province Guizhou, China. *The Science of the total environment* **2003,** *304*, (1-3), 231-56.
656 17. StLouis, V. L.; Rudd, J. W. M.; Kelly, C. A.; Beaty, K. G.; Flett, R. J.; Roulet, N. T., Production and
657 loss of methylmercury and loss of total mercury from boreal forest catchments containing different
658 types of wetlands. *Environ Sci Technol* **1996,** *30*, (9), 2719-2729.

18. Mikac, N.; Picer, M.; Stegnar, P.; Tusekznidaric, M., Mercury Distribution in a Polluted Marine Area, Ratio of Total Mercury, Methyl Mercury and Selenium in Sediments, Mussels and Fish. *Water Res* **1985,** *19*, (11), 1387-1392.
19. Kannan, K.; Smith, R. G., Jr.; Lee, R. F.; Windom, H. L.; Heitmuller, P. T.; Macauley, J. M.; Summers, J. K., Distribution of total mercury and methyl mercury in water, sediment, and fish from south Florida estuaries. *Archives of environmental contamination and toxicology* **1998,** *34*, (2), 109-18.
20. Hsu-Kim, H.; Kucharzyk, K. H.; Zhang, T.; Deshusses, M. A., Mechanisms regulating mercury bioavailability for methylating microorganisms in the aquatic environment: a critical review. *Environmental science & technology* **2013,** *47*, (6), 2441-2456.
21. Christensen, G. A.; Somenahally, A. C.; Moberly, J. G.; Miller, C. M.; King, A. J.; Gilmour, C. C.; Brown, S. D.; Podar, M.; Brandt, C. C.; Brooks, S. C.; Palumbo, A. V.; Wall, J. D.; Elias, D. A., Carbon amendments alter microbial community structure and net mercury methylation potential in sediments *Appl. Environ. Microbiol.* **2017,** *84*, (3), e01049-17.
22. Bravo, A. G.; Bouchet, S.; Tolu, J.; Björn, E.; Mateos-Rivera, A.; Bertilsson, S., Molecular composition of organic matter controls methylmercury formation in boreal lakes. *Nature Communications* **2017,** *8*, 14255.
23. Mangal, V.; Stenzler, B. R.; Poulain, A. J.; Guéguen, C., Aerobic and anaerobic bacterial mercury uptake is driven by algal organic matter composition and molecular weight. *Environmental science & technology* **2018,** *53*, (1), 157-165.
24. Benoit, J. M.; Gilmour, C. C.; Heyes, A.; Mason, R. P.; Miller, C. L., Geochemical and Biological Controls Over Methylmercury Production and Degradation in Aquatic Ecosystems. In *Biogeochemistry of Environmentally Important Trace Elements*, Y., C.; Braids, O. C., Eds. American Chemical Society: Washington, DC., 2003; pp 262-297.
25. Graham, A. M.; Aiken, G. R.; Gilmour, C. C., Dissolved Organic Matter Enhances Microbial Mercury Methylation Under Sulfidic Conditions. *Environ Sci Technol* **2012,** *46*, (5), 2715-2723.
26. Schaefer, J. K.; Rocks, S. S.; Zheng, W.; Liang, L.; Gu, B.; Morel, F. M. M., Active transport, substrate specificity, and methylation of Hg(II) in anaerobic bacteria. *Proc. Natl. Acad. Sci. USA* **2011,** *108*, (21), 8714-8719.
27. Kronberg, R.-M.; Schaefer, J. K.; Björn, E.; Skyllberg, U., Mechanisms of Methyl Mercury Net Degradation in Alder Swamps: The Role of Methanogens and Abiotic Processes. *Environmental Science & Technology Letters* **2018,** *5*, (4), 220-225.
28. Barkay, T.; Miller, S. M.; Summers, A. O., Bacterial mercury resistance from atoms to ecosystems. *FEMS microbiology reviews* **2003,** *27*, (2-3), 355-84.
29. Boyd, E. S.; Barkay, T., The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. *Front. Microbiol.* **2012,** *3:349*.
30. Suda, I.; Suda, M.; Hirayama, K., Degradation of methyl and ethyl mercury by singlet oxygen generated from sea water exposed to sunlight or ultraviolet light. *Arch. Toxicol.* **1993,** *67*, (5), 365-368.
31. Oremland, R. S.; Culbertson, C. W.; Winfrey, M. R., Methylmercury Decomposition in Sediments and Bacterial Cultures: Involvement of Methanogens and Sulfate Reducers in Oxidative Demethylation. *Applied and Environmental Microbiology* **1991,** *57*, (1), 130-137.
32. Schaefer, J. K.; Yagi, J.; Reinfelder, J. R.; Cardona, T.; Ellickson, K. M.; Tel-Or, S.; Barkay, T., Role of the bacterial organomercury lyase (MerB) in controlling methylmercury accumulation in mercury-contaminated natural waters. *Environ Sci Technol* **2004,** *38*, (16), 4304-11.
33. Marvin-DiPasquale, M.; Agee, J.; McGowan, C.; Oremland, R. S.; Thomas, M.; Krabbenhoft, D.; Gilmour, C. C., Methyl-mercury degradation pathways: A comparison among three mercury-impacted ecosystems. *Environ Sci Technol* **2000,** *34*, (23), 4908-4916.
34. Schaefer, J. K.; Letowski, J.; Barkay, T., Mer-mediated resistance and volatilization of Hg(II) under anaerobic conditions. *Geomicrobiology Journal* **2002,** *19*, (1), 87-102.

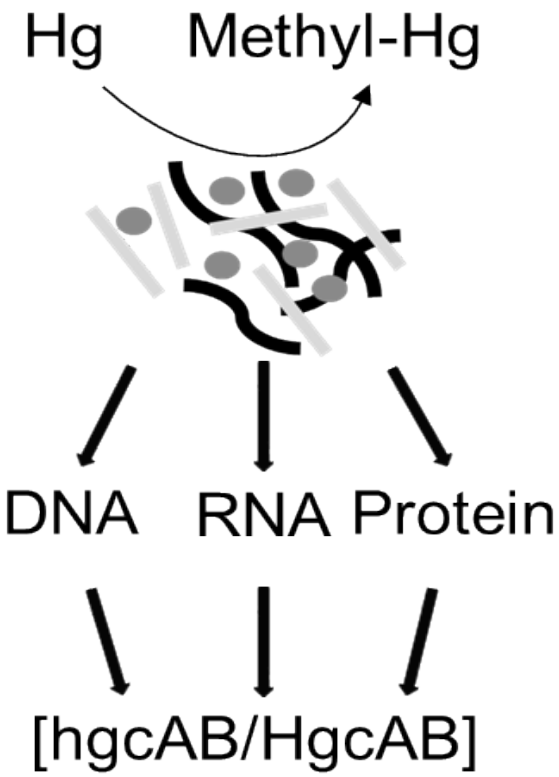
35. Ndu, U.; Christensen, G. A.; Rivera, N. A.; Gionfriddo, C. M.; Deshusses, M. A.; Elias, D. A.; Hsu-Kim, H., Quantification of Mercury Bioavailability for Methylation Using Diffusive Gradient in Thin-Film Samplers. *Environmental Science & Technology* **2018 ,**52 15, 8521-8529.
36. Smith, S. D.; Bridou, R.; Johs, A.; Parks, J. M.; Elias, D. A.; Hurt, R. A.; Brown, S. D.; Podar, M.; Wall, J. D., Site-directed mutagenesis of HgcA and HgcB reveals amino acid residues important for mercury methylation. *Applied and environmental microbiology* **2015,** *81*, (9), 3205-3217.
37. Gilmour, C. C.; Elias, D. A.; Kucken, A. M.; Brown, S. D.; Palumbo, A. V.; Schadt, C. W.; Wall, J. D., Sulfate-Reducing Bacterium Desulfovibrio desulfuricans ND132 as a Model for Understanding Bacterial Mercury Methylation. *Applied and Environmental Microbiology* **2011,** *77*, (12), 3938-3951.
38. Goñi-Urriza, M.; Corsellis, Y.; Lanceleur, L.; Tessier, E.; Gury, J.; Monperrus, M.; Guyoneaud, R., Relationships between bacterial energetic metabolism, mercury methylation potential, and *hgcA*/*hgcB* gene expression in *Desulfovibrio dechloroacetivorans* BerOc1. *Environ Sci Pollut Res* **2015,** *22*, (18), 13764-13771.
39. Qian, C.; Chen, H.; Johs, A.; Lu, X.; An, J.; Pierce, E. M.; Parks, J. M.; Elias, D. A.; Hettich, R. L.; Gu, B., Quantitative Proteomic Analysis of Biological Processes and Responses of the Bacterium Desulfovibrio desulfuricans ND132 upon Deletion of Its Mercury Methylation Genes. *Proteomics* **2018,** *18*, (17), 1700479.
40. Christensen, G. A.; Moon, J.; Veach, A. M.; Mosher, J. J.; Wymore, A. M.; van Nostrand, J. D.; Zhou, J.; Hazen, T. C.; Arkin, A. P.; Elias, D. A., Use of in-field bioreactors demonstrate groundwater filtration influences planktonic bacterial community assembly, but not biofilm composition. *PloS one* **2018,** *13*, (3), e0194663.
41. Xu, J.; Buck, M.; Eklof, K.; Osman, O. A.; Schaefer, J. K.; Bishop, K.; Björn, E.; Skyllberg, U.; Bertilsson, S.; Bravo, A. G., Mercury methylating microbial communities of boreal forest soils. *Scientific Reports* **2019**, 9,518.
42. Bravo, A. G.; Loizeau, J. L.; Dranguet, P.; Makri, S.; Bjorn, E.; Ungureanu, V. G.; Slaveykova, V. I.; Cosio, C., Persistent Hg contamination and occurrence of Hg-methylating transcript (hgcA) downstream of a chlor-alkali plant in the Olt River (Romania). *Environmental Science and Pollution Research* **2016,** *23*, (11), 10529-10541.
43. Christensen, G. A.; Wymore, A. M.; King, A. J.; Podar, M.; R. A. Hurt, J.; Santillan, E. U.; Soren, A.; Brandt, C. C.; Brown, S. D.; Palumbo, A. V.; Wall, J. D.; Gilmour, C. C.; Elias, D. A., Development and validation of broad-range qualitative and clade-specific quantitative molecular probes for assessing Mercury methylation in the environment. *Appl. Environ. Microbiol.* **2016,** *82*, (19), 6068-6078.
44. Gionfriddo, C. M.; Tate, M. T.; Wick, R. R.; Schultz, M. B.; Zemla, A.; Thelen, M. P.; Schofield, R.; Krabbenhoft, D. P.; Holt, K. E.; Moreau, J. W., Microbial mercury methylation in Antarctic sea ice. *Nature microbiology* **2016,** *1*, (10), 16127.
45. Christensen, G. A.; Moon, J.; Veach, A. M.; Mosher, J. J.; Wymore, A. M.; van Nostrand, J. D.; Zhou, J. Z.; Hazen, T. C.; Arkin, A. P.; Elias, D. A., Use of in-field bioreactors demonstrate groundwater filtration influences planktonic bacterial community assembly, but not biofilm composition. *PLoS ONE* **2018,** *13*, (3), e0194663.
46. Podar, M.; Gilmour, C. C.; Brandt, C. C.; Soren, A.; Brown, S. D.; Crable, B. R.; Palumbo, A. V.; Somenahally, A. C.; Elias, D. A., Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci. Adv.* **2015,** *1*, (9), e1500675.
47. Mosher, J. J.; Vishnivetskaya, T. A.; Elias, D. A.; Podar, M.; Brooks, S. C.; Brown, S. D.; Brandt, C. C.; Palumbo, A. V., Characterization of the deltaproteobacteria in contaminated and uncontaminated stream sediments and identification of potential mercury methylators. *Aquatic Microbial Ecology* **2012,** *66*, 271-282.
48. Vishnivetskaya, T. A.; Mosher, J. J.; Palumbo, A. V.; Yang, Z. K.; Podar, M.; Brown, S. D.; Brooks, S. C.; Gu, B.; Southworth, G. R.; Drake, M. M.; Brandt, C. C.; Elias, D. A., Mercury and other heavy metals

755 influence bacterial community structure in contaminated Tennessee streams. *Appl. Environ. Microbiol.*
756 **2011,** *77*, (1), 302-311.
757 49. Roy_Chowdhury, T.; Herndon, E. K.; Phelps, T. J.; Elias, D. A.; Gu, B.; Liang, L.; Wullschleger, S. D.;
758 Graham, D. E., Stoichiometry and temperature sensitivity of methanogenesis and CO_2 production from
759 saturated polygonal tundra in Barrow, Alaska. *Glob. Change Biol.* **2015,** *21*, 722-737.
760 50. Mitchell, C. P. J.; Gilmour, C. C., Methylmercury production in a Chesapeake Bay salt marsh. *J.*
761 *Geophys. Res. Biogeo.* **2008,** *113,* GOOCO4, 1-14.
762 51. Gilmour, C. C.; Bell, T.; Soren, A.; Riedel, G.; Riedel, G.; Kopec, D.; Bodaly, D.; Ghosh, U.,
763 Activated carbon thin-layer placement as an in situ mercury remediation tool in a Penobscot River salt
764 marsh. *Sci Tot Environ* **2018,** *621*, 839-848.
765 52. Hurt, R. A.; Brown, S. D.; Podar, M.; Palumbo, A. V.; Elias, D. A., Sequencing intractable DNA to
766 close microbial genomes. *PloS One* **2012,** *7*, (7), e41295.
767 53. Hurt, R. A.; Robeson, M. S.; Shakya, M.; Moberly, J. G.; Vishnivetskaya, T. A.; Gu, B.; Elias, D. A.,
768 Improved yield of high molecular weight DNA coincides with increased microbial diversity access from
769 iron oxide cemented sub-surface clay environments. *PLoS ONE* **2014,** *9*, (7), e102826.
770 54. Baker, G. C.; Smith, J. J.; Cowan, D. A.; .55(3):541-55, Review and re-analysis of domain-specific
771 16S primers. *J Microbiol Meth* **2003,** *55*, (3), 541-555.
772 55. Hwang, C. C.; Wu, W. M.; Gentry, T. J.; Carley, J.; Corbin, G. A.; Carroll, S. L.; Watson, D. B.;
773 Jardine, P. M.; Zhou, J. Z.; Criddle, C. S.; Fields, M. W., Bacterial community succession during in situ
774 uranium bioremediation: spatial similarities along controlled flow paths. *Isme J.* **2009,** *3*, (1), 47-64.
775 56. León, K. B. D.; Gerlach, R.; Peyton, B. M.; Fields, M. W., Archaeal and bacterial communities in
776 three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. *Front Microbiol* **2013,**
777 *4*, 330.
778 57. León, K. B. D.; Ramsay, B. D.; Fields, M. W., Quality-score refinement of SSU rRNA gene
779 pyrosequencing differs across gene region for environmental samples. *Microb Ecol* **2012,** *64*, (2), 499-
780 508.
781 58. Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.;
782 Tabbaa, D.; Highlander, S. K.; Sodergren, E.; Methe, B.; DeSantis, T. Z.; Petrosino, J. F.; Knight, R.; Birren,
783 B. W.; Consortium, H. M., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-
784 pyrosequenced PCR amplicons. *Genome Res* **2011,** *21*, (3), 494-504.
785 59. Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R., Naive Bayesian classifier for rapid assignment
786 of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* **2007,** *73*,
787 (16), 5261-5267.
788 60. Claesson, M. J.; O'Sullivan, O.; Wang, Q.; Nikkilä, J.; Marchesi, J. R.; Smidt, H.; de Vos, W. M.;
789 Ross, R. P.; O'Toole, P. W., Comparative analysis of pyrosequencing and a phylogenetic microarray for
790 exploring microbial community structures in the human distal intestine. *PloS one* **2009,** *4*, (8), e6669.
791 61. DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.;
792 Hu, P.; Andersen, G. L., Greengenes, a chimera-checked 16S rRNA gene database and workbench
793 compatible with ARB. *Appl. Environ. Microbiol.* **2006,** *72*, (7), 5069-5072.
794 62. Prezza, N.; Vezzi, F.; Kaller, M.; Policriti, A., Fast, accurate, and lightweight analysis of BS-treated
795 reads with ERNE 2. *BMC Bioinformatics* **2016,** *17*.
796 63. Zhang, J. J.; Kobert, K.; Flouri, T.; Stamatakis, A., PEAR: a fast and accurate Illumina Paired-End
797 reAd mergeR. *Bioinformatics* **2014,** *30*, (5), 614-620.
798 64. Goujon, M.; McWilliam, H.; Li, W. Z.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R., A new
799 bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **2010,** *38*, W695-W699.
800 65. Namiki, T.; Hachiya, T.; Tanaka, H.; Sakakibara, Y., MetaVelvet: an extension of Velvet assembler
801 to de novo metagenome assembly from short sequence reads. *Nucleic acids research* **2012,** *40*, (20),
802 e155-e155.

803 66. Peng, Y.; Leung, H. C.; Yiu, S.-M.; Chin, F. Y., IDBA-UD: a de novo assembler for single-cell and
804 metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012,** *28*, (11), 1420-1428.
805 67. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E. M.; Kubal, M.; Paczian, T.; Rodriguez,
806 A.; Stevens, R.; Wilke, A., The metagenomics RAST server–a public resource for the automatic
807 phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* **2008,** *9*, (1), 386.
808 68. Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta,
809 M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G. A.; Tate, J.; Bateman, A., The Pfam protein families
810 database: towards a more sustainable future. *Nucleic Acids Res.* **2016,** *44*, (D1), D279-D285.
811 69. Brown, N. L.; Stoyanov, J. V.; Kidd, S. P.; Hobman, J. L., The MerR family of transcriptional
812 regulators. *FEMS microbiology reviews* **2003,** *27*, (2-3), 145-163.
813 70. Wheeler, T. J.; Eddy, S. R., nhmmer: DNA homology search with profile HMMs. *Bioinformatics*
814 **2013,** *29*, (19), 2487-2489.
815 71. Edgar, R. C., MUSCLE: a multiple sequence alignment method with reduced time and space
816 complexity. *BMC Bioinformatics* **2004,** *5*, 1-19.
817 72. Price, M. N.; Dehal, P. S.; Arkin, A. P., FastTree: computing large minimum evolution trees with
818 profiles instead of a distance matrix. *Molec. Biol. Evol.* **2009,** *26*, (7), 1641-1650.
819 73. Johnson, L. S.; Eddy, S. R.; Portugaly, E., Hidden Markov model speed heuristic and iterative
820 HMM search procedure. *BMC Bioinformatics* **2010,** *11*, (1), 431.
821 74. Kumar, S.; Stecher, G.; Tamura, K., MEGA7: molecular evolutionary genetics analysis version 7.0
822 for bigger datasets. *Molecular biology and evolution* **2016,** *33*, (7), 1870-1874.
823 75. Pfaffl, M. W., A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic*
824 *Acids Res* **2001,** *29*, (9), e45.
825 76. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F.,
826 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple
827 internal control genes. *Genome biology* **2002,** *3*, (7), RESEARCH0034.
828 77. Hellemans, J.; Mortier, G.; De Paepe, A.; Speleman, F.; Vandesompele, J., qBase relative
829 quantification framework and software for management and automated analysis of real-time
830 quantitative PCR data. *Genome biology* **2007,** *8*, (2), R19.
831 78. Land, M.; Hauser, L.; Jun, S.-R.; Nookaew, I.; Leuze, M. R.; Ahn, T.-H.; Karpinets, T.; Lund, O.;
832 Kora, G.; Wassenaar, T., Insights from 20 years of bacterial genome sequencing. *Functional & integrative*
833 *genomics* **2015,** *15*, (2), 141-161.
834 79. Robidart, J.; Callister, S. J.; Song, P.; Nicora, C. D.; Wheat, C. G.; Girguis, P. R., Characterizing
835 microbial community and geochemical dynamics at hydrothermal vents using osmotically driven
836 continuous fluid samplers. *Environ Sci Technol* **2013,** *47*, (9), 4399-4407.
837 80. Kim, S.; Gupta, N.; Pevzner, P. A., Spectral probabilities and generating functions of tandem
838 mass spectra: a strike against decoy databases. *Proteome Res* **2008,** *7*, (8), 3354-3363.
839 81. Mathews, T. J.; Looney, B. B.; Bryan, A. L.; Smith, J. G.; Miller, C. L.; Southworth, G. R.; Peterson,
840 M. J., The effects of a stannous chloride-based water treatment system in a mercury contaminated
841 stream. *Chemosphere* **2015,** *138*, 190-196.
842 82. Moberly, J. G.; Brown, S. D.; Kucken, A. M.; Gilmour, C. C.; Brandt, C. C.; Palumbo, A. V.; Wall, J.
843 D.; Elias, D. A., Role of morphological growth state and gene expression in *Desulfovibrio africanus* strain
844 Walvis Bay mercury methylation. *Environ Sci Technol* **2012,** *46*, (9), 4926–4932.
845 83. Gilmour, C. C.; Bullock, A. L.; McBurney, A.; Podar, M.; Elias, D. A., Robust Mercury Methylation
846 across Diverse Methanogenic Archaea. *mBio* **2018,** *9*, (2), e02403-17.
847 84. Bailey, L. T.; Mitchell, C. P. J.; Engstrom, D. R.; Berndt, M. E.; Wasik, J. K. C.; Johnson, N. W.,
848 Influence of porewater sulfide on methylmercury production and partitioning in sulfate-impacted lake
849 sediments. *Science of the Total Environment* **2017,** *580*, 1197-1204.

850 85. Olsen, T. A.; Muller, K. A.; Painter, S. L.; Brooks, S. C., Kinetics of Methylmercury Production
851 Revisited. *Environ Sci Technol* **2018,** *52*, (4), 2063-2070.
852 86. Jonsson, S.; Skyllberg, U.; Nilsson, M. B.; Westlund, P. O.; Shchukarev, A.; Lundberg, E.; Bjorn, E.,
853 Mercury Methylation Rates for Geochemically Relevant Hg-II Species in Sediments. *Environ Sci Technol*
854 **2012,** *46*, (21), 11653-11659.
855 87. Hsu-Kim, H.; Eckley, C. S.; Acha, D.; Feng, X. B.; Gilmour, C. C.; Jonsson, S.; Mitchell, C. P. J.,
856 Challenges and opportunities for managing aquatic mercury pollution in altered landscapes. *Ambio*
857 **2018,** *47*, (2), 141-169.
858 88. Schartup, A. T.; Balcom, P. H.; Mason, R. P., Sediment-Porewater Partitioning, Total Sulfur, and
859 Methylmercury Production in Estuaries. *Environ Sci Technol* **2014,** *48*, (2), 954-960.
860 89. Tjerngren, I.; Karlsson, T.; Bjorn, E.; Skyllberg, U., Potential Hg methylation and MeHg
861 demethylation rates related to the nutrient status of different boreal wetlands. *Biogeochemistry* **2012,**
862 *108*, (1-3), 335-350.
863 90. Nayfach, S.; Rodriguez-Mueller, B.; Garud, N.; Pollard, K. S., An integrated metagenomics
864 pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome*
865 *Res.* **2016,** *26*, (11), 1612-1625.
866 91. Acinas, S. G.; Sarma-Rupavtarm, R.; Klepac-Ceraj, V.; Polz, M. F., PCR-induced sequence artifacts
867 and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample.
868 *Appl. Environ. Microbiol.* **2005,** *71*, 8966-8969.
869 92. Nayfach, S.; Pollard, K. S., Toward Accurate and Quantitative Comparative Metagenomics. *Cell*
870 **2016,** *166*, (5), 1103-1116.
871 93. Gilmour, C. C.; Elias, D. A.; Kucken, A.; Brown, S. D.; Palumbo, A. V.; Wall, J. D., The sulfate-
872 reducing bacterium Desulfovibrio desulfuricans ND132 as a model for understanding bacterial mercury
873 methylation. *Geochim Cosmochim Ac* **2010,** *74*, (12), A333-A333.
874 94. Luton, P. E.; Wayne, J. M.; Sharp, R. J.; Riley, P. W., The *mcrA* gene as an alternative to 16S rRNA
875 in the phylogenetic analysis of methanogen populations in landfill. *Microbiol.* **2002,** *148*, (11), 3521-
876 3530.
877 95. Lu, X.; Liu, Y. R.; Johs, A.; Zhao, L. D.; Wang, T. S.; Yang, Z. M.; Lin, H.; Elias, D. A.; Pierce, E. M.;
878 Liang, L. Y.; Barkay, T.; Gu, B. H., Anaerobic Mercury Methylation and Demethylation by Geobacter
879 bemidjiensis Bem. *Environ Sci Technol* **2016,** *50*, (8), 4366-4373.
880 96. Du, H. X.; Ma, M.; Sun, T.; Dai, X. Z.; Yang, C. Y.; Luo, F.; Wang, D. Y.; Igarashi, Y., Mercury-
881 methylating genes dsrB and hgcA in soils/sediments of the Three Gorges Reservoir. *Environmental*
882 *Science and Pollution Research* **2017,** *24*, (5), 5001-5011.
883 97. Bravo, A. G.; Zopfi, J.; Buck, M.; Xu, J.; Bertilsson, S.; Schaefer, J. K.; Poté, J.; Cosio, C.,
884 *Geobacteraceae* are important members of mercury-methylating microbial communities of sediments
885 impacted by waste water releases. *The ISME journal* **2018,** *12*, (3), 802-812.
886
887

TOC ART



894 Table 1: Site Hg Concentrations and *hgcA(B)* Diversity and Abundance by Various
895 Methods

	Tennessee (Riverine area)			Maryland (Tidal Marsh)		Maine (Tidal Marsh)		Alaska (Arctic Permafrost)
	EFPC NOAA Upstream	EFPC NH Downstream	Hinds Creek Background	SERC Marsh Fringe	SERC Upper Marsh	Penobscot W10MM 10 km	Penobscot W21UM 21 km	Barrow Permafrost
HgT (mg kg^{-1} dry wt.)	14.60	5.02	0.03	0.07	0.09	0.14	0.23	0.05
MeHg (µg kg^{-1} dry wt.)	2.77	0.13	0.56	0.05	0.55	4.17	27.30	0.13
% MeHg	0.02	0.06	1.74	0.06	0.60	3.07	11.90	0.38
Pyrosequencing (16S rRNA genes per total read pairs x 10^6)[a]								
Deltaproteobacteria	0.55	0.35	0	0	0	0	0.12	0.08
Firmicutes	0	0	0	0	0.069	0	0.31	0
Archaea	0	0	0	0	0	0	0	0
Other	0	0	0	0	0	0	0	0
Total	0.55	0.35	0	0	0.069	0	0.44	0.08
Metagenome (16S rRNA genes per total 16S read pairs x 10^6)[b]								
Deltaproteobacteria	603	1,002	913	628	608	2,763	7,611	763
Firmicutes	75	0	228	0	101	0	0	85
Archaea	75	501	1,141	209	0	102	0	678
Other	0	0	0	105	101	0	0	0
Total	754	1,503	2,283	941	811	2,866	7,611	1,525
Clone (*hgcAB*, ~100 clones sequenced)[c]								
Deltaproteobacteria	46	38	23	69	63	61	75	60
Firmicutes	24	23	29	24	11	13	17	10
Archaea	25	29	41	0	9	24	8	20
Other	5	9	8	7	17	1	1	10
Total	100	99	101	100	100	99	101	100
Metagenome (*hgcA* genes per genome equiv.)[d]								
Deltaproteobacteria	0.140	0.245	0.292	0.341	0.344	0.302	0.133	0.400
Firmicutes	0.078	0.140	0.166	0.116	0.113	0.122	0.040	0.248
Archaea	0.041	0.113	0.161	0.115	0.093	0.099	0.027	0.125
Other	0.014	0.037	0.055	0.145	0.126	0.047	0.028	0.048
Total	0.272	0.536	0.674	0.716	0.675	0.570	0.229	0.821
qPCR (*hgcA* copies per ng DNA)[e]								
Deltaproteobacteria	208 ± 54	1390 ± 193	553 ± 39	1238 ± 162	782 ± 33	483 ± 44	202 ± 81	686 ± 138
Firmicutes	28 ± 3	90 ± 9	33 ± 2	22 ± 2	121 ± 1	97 ± 3	70 ± 14	1208 ± 93
Archaea	2407 ± 444	1080 ± 105	2894 ± 107	N.D.	5030 ± 992	1560 ± 205	N.D.	1913 ± 395
qPCR (*hgcA* copies per genome equiv)[e]								
Deltaproteobacteria	0.0011 ± 0.00029	0.0075 ± 0.0010	0.0030 ± 0.0002	0.0067 ± 0.00087	0.0042 ± 0.00018	0.0026 ± 0.00024	0.0011 ± 0.00044	0.0037 ± 0.0007
Firmicutes	0.0002 ± 0.00002	0.0005 ± 0.00005	0.0002 ± 0.00001	0.0001 ± 0.00001	0.0007 ± 0.00001	0.0005 ± 0.00002	0.0004 ± 0.00008	0.0065 ± 0.0005
Archaea	0.0130 ± 0.00240	0.0058 ± 0.00057	0.0156 ± 0.00058		0.0272 ± 0.0054	0.0084 ± 0.00111		0.0103 ± 0.0021
Proteomics (HgcA)[f]								
Deltaproteobacteria	8,269	8,640	9,180	N.A.	N.A.	N.A.	N.A.	N.A.
Firmicutes	3,752	3,714	4,015	N.A.	N.A.	N.A.	N.A.	N.A.
Archaea	2890	2796	3244	N.A.	N.A.	N.A.	N.A.	N.A.
Other	359	361	413	N.A.	N.A.	N.A.	N.A.	N.A.
Proteomics (HgcB)[f]								
Deltaproteobacteria	5,403	2,895	6,060	N.A.	N.A.	N.A.	N.A.	N.A.
Firmicutes	1,823	1,009	2,063	N.A.	N.A.	N.A.	N.A.	N.A.
Archaea	4718	2550	5378	N.A.	N.A.	N.A.	N.A.	N.A.
Other	737	357	712	N.A.	N.A.	N.A.	N.A.	N.A.
Genome equivalents	1418.4	1262.2	565.5	1133.0	1511.8	1299.3	1567.5	1182.7
genome equiv./ 1 ng DNA	185245							

896
897 [a] Identified by previously established protocol[21]
898 [b] Metagenomes mined for 16S rRNA gene to confidently identify Hg-methylators.
899 [c] Identified by qualitative PCR protocol[43], followed by TA cloning of the amplicon.
900 [d] Identified by HMM protocol[46]
901 [e] Identified by clade-specific quantitative PCR-based protocol, technical triplicates[43]
902 [f] Identified by mining metaproteomes for amino acid sequences that can be confidently identified as hgcA.
903 Protein copies per 0.4 g sediment.

904

905

Table 2: Site Hg Concentrations and Potential MeHg demethylator Diversity and Abundance

	Tennessee (Riverine area)			Maryland (Tidal Marsh)		Maine (Tidal Marsh)		Alaska
	EFPC NOAA Upstream	EFPC NH Downstream	Hinds Creek Background	SERC Marsh Fringe	SERC Upper Marsh	Penobscot W10MM 10 km	Penobscot W21UM 21 km	(Arctic Permafrost) Barrow Permafrost
HgT (mg kg^{-1} dry wt.)	14.60	5.02	0.03	0.07	0.09	0.14	0.23	0.05
MeHg (µg kg^{-1} dry wt.)	2.77	0.13	0.56	0.05	0.55	4.17	27.30	0.13
% MeHg	0.02	0.06	1.74	0.06	0.60	3.07	11.90	0.38
Metagenome (reads per genome equiv)[a]								
HgcA	0.27	0.54	0.67	0.72	0.68	0.57	0.23	0.82
HgcB	0.99	1.69	1.95	2.47	2.14	1.44	0.80	1.65
MerB	0.19	0.12	0.11	0.07	0.08	0.05	0.05	0.01
MerA	1.80	1.43	1.02	0.92	0.99	0.93	0.78	1.39
MerR	2.12	1.51	1.25	0.92	1.28	0.98	1.18	1.39
MerR_HTH	2.83	2.25	1.87	1.59	1.97	1.65	1.84	1.97
DsrC	0.78	0.67	0.56	0.77	0.89	0.70	0.63	0.45
McrA_C_Domain	0.02	0.03	0.03	0.02	0.00	0.01	0.00	0.12
Genome equivalents	1418.4	1262.2	565.5	1133.0	1511.8	1299.3	1567.5	1182.7

906

907 [a] Refers to number of protein homologs from paired metagenomic read sets that could be confidently identified above threshold (1E-5) normalized
908 to number of genome equivalents per metagenome

909

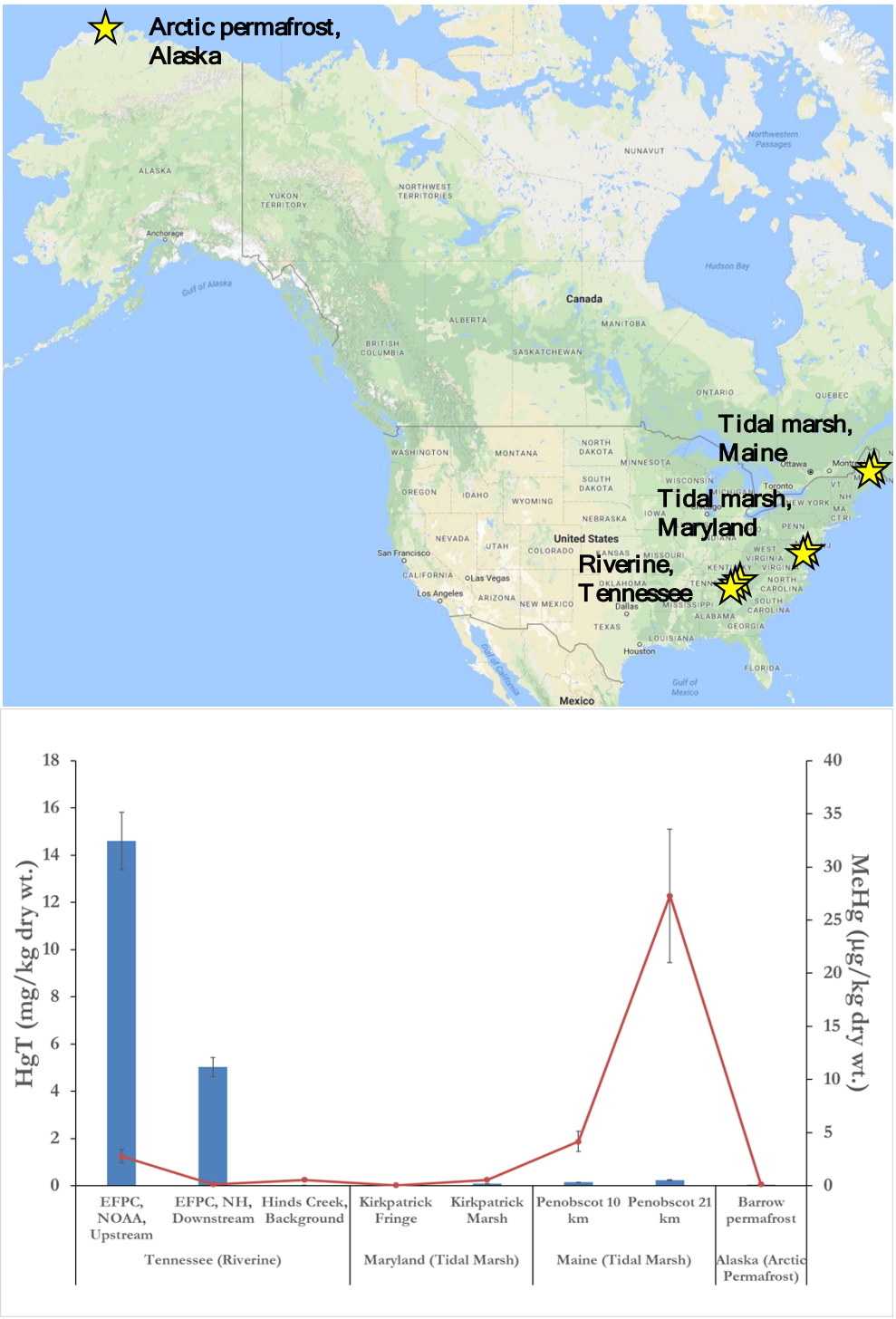


Figure 1: Geographic locations of the eight sites used in this study and their HgT and MeHg concentrations. (Top) yellow stars represent sample locations. (Bottom) relative HgT (blue bar, left axis) and MeHg (red line, right axis) concentrations. The map was obtained from google maps **https://www.google.com/maps/@44.8947428,-65.2920883,3.77Z**.

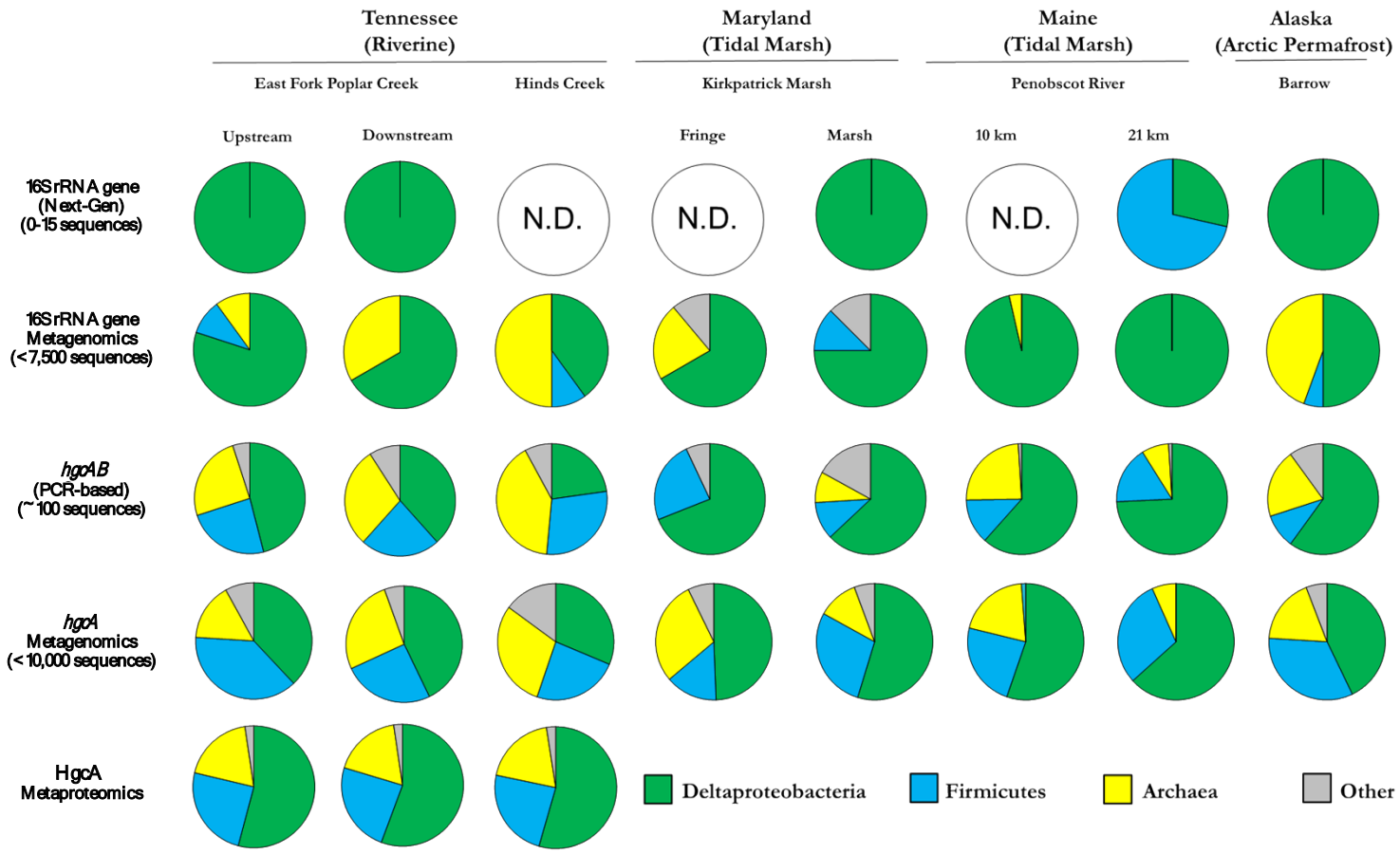
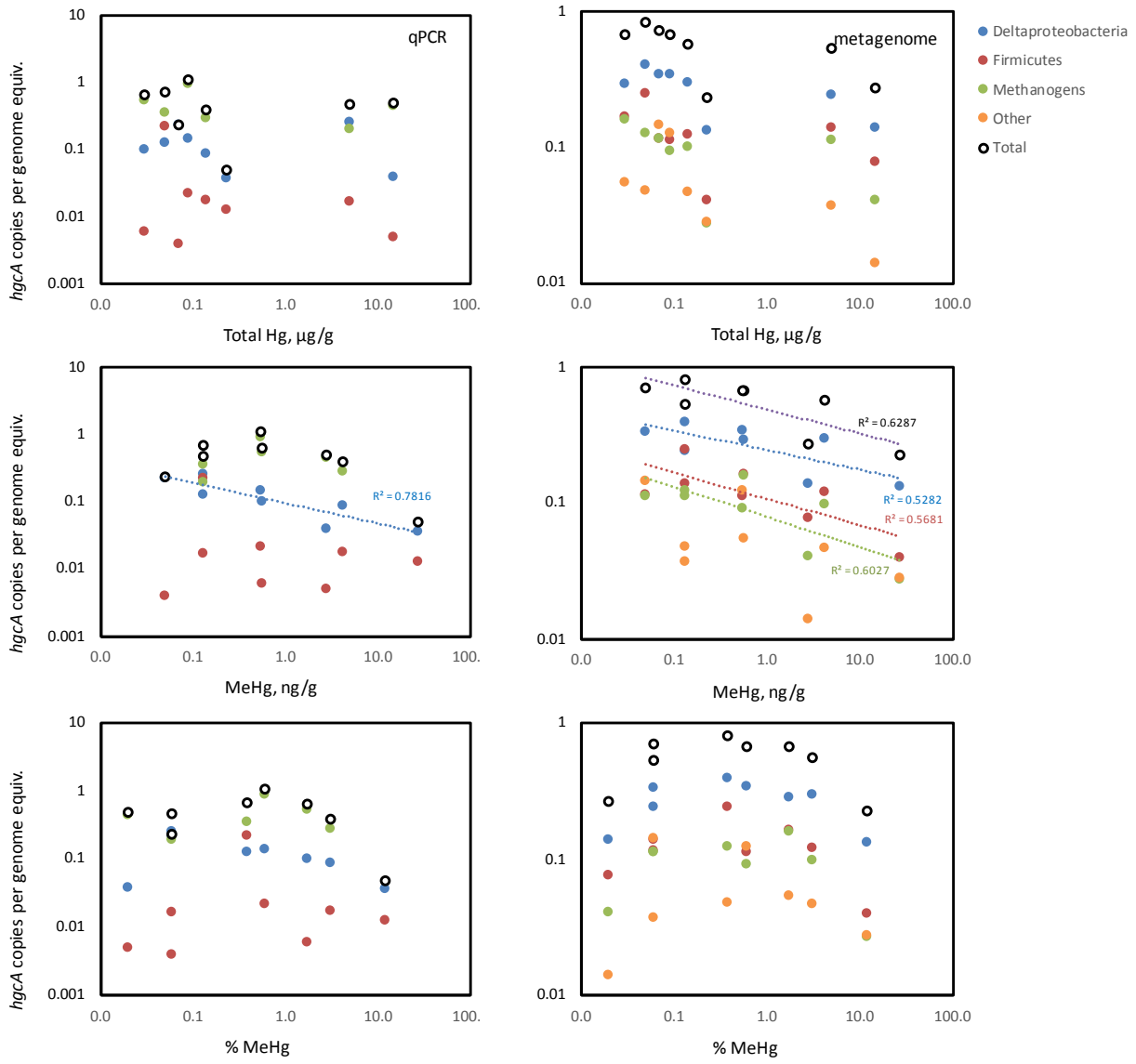


Figure 2: Comparison of Hg-methylator community estimates from the eight different environmental sites by different methods: (top row) 16S rRNA pyrosequencing; (second row) 16S rRNA gene counts from metagenomes; (third row) *hgcAB* gene counts from PCR amplification and sequencing of ~100 clones; (fourth row) *hgcA* gene counts from metagenomes, (bottom row) HgcA clade distribution based on metaproteomics. N.D. not detected.

947



948

949

950 **Figure 3:** The relationships between qPCR (left) or metagenome (right) based estimates for
951 *hgcA* copy numbers per genome equivalent from the three major Hg-methylator clades and
952 environmental Hg levels in sediments for each of the eight sites. Several relationships were
953 significantly but negatively correlated with MeHg ($p<0.05$); these are shown as dotted lines.
954 Correlations were evaluated using log transformed data to normalize distributions. Relationships
955 were not improved when *hgcA* copy numbers were normalized to copies per g of sediment.

956

957

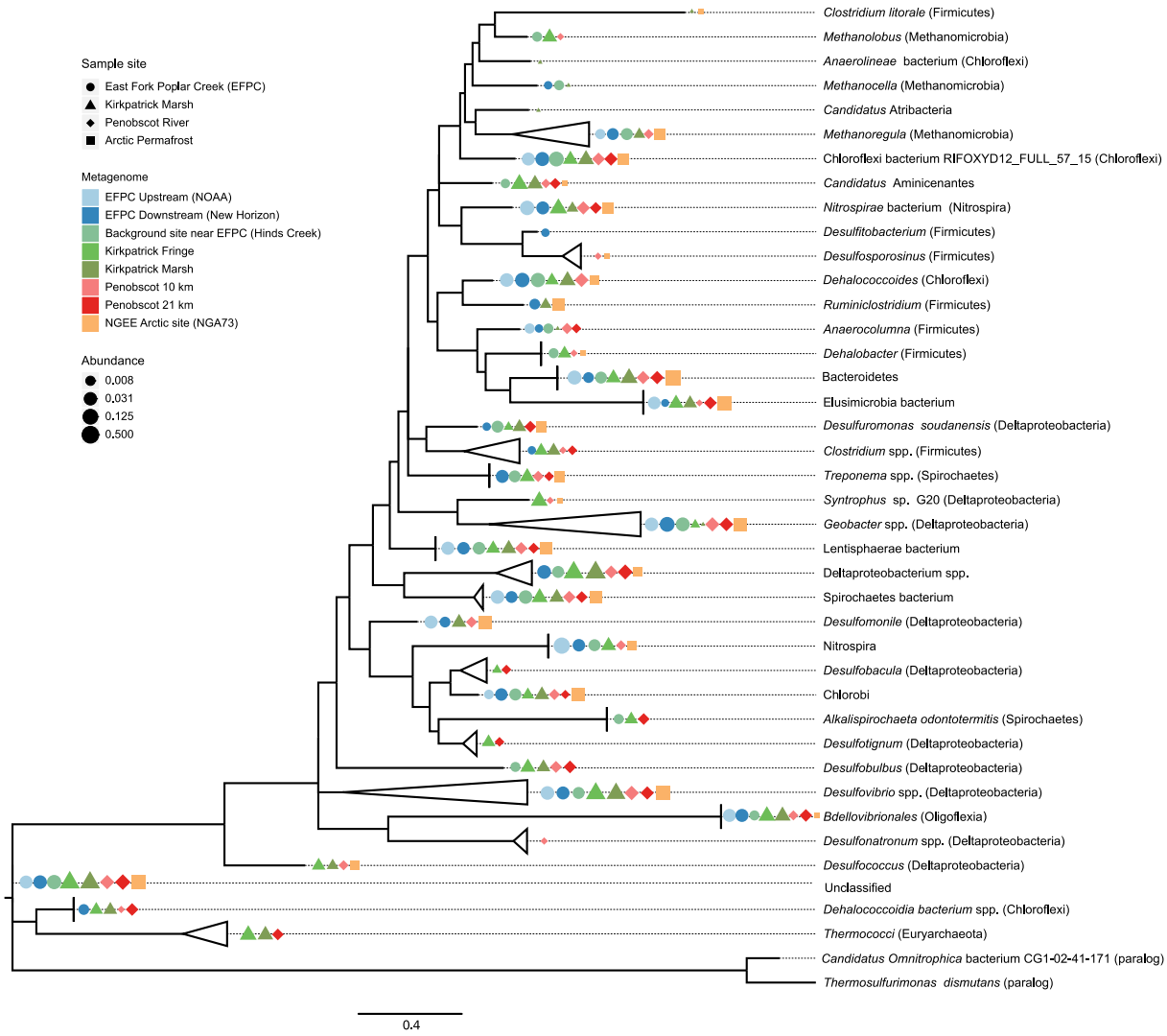


Figure 4: Relative abundance and phylogenetic relationship of HgcA OTUs from environmental metagenomes. HgcA sequences were extracted from metagenomic datasets using 'hmmsearch' and reads that aligned with the cap-helix region of HgcA were kept for taxonomic classification, using an inclusion threshold of $1x10^{-7}$. Taxonomy was assigned to each HgcA read by phylogenetic placement of sequences onto a reference tree using the python program 'taxtastic' with an in-house HgcA reference database and classification cut-off of 90%. Reads were merged into OTUs based on phylogenetic distances, resulting in 77 taxonomic units. ***Tree construction:*** Representative HgcA sequences, based on 'nearest neighbor' classification, were chosen for each OTU and used for tree construction. The tree represents a maximum likelihood (ML) phylogenetic distribution of HgcA estimated using the GAMMA model of rate heterogeneity in MEGA with 100 bootstrap replicates. Tree is rooted by paralog HgcA sequences (i.e. carbon monoxide dehydrogenases) from non-Hg-methylators *Candidatus Omnitrophica* and *Thermosulfurimonas dismutans*. Branch labels indicate lowest taxonomic classification for OTUs. Abundance values are based on the fraction of metagenomic reads that clustered to each OTU relative to total HgcA reads from each metagenome that aligned to the cap-helix of HgcA.

974