# Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments

2 3 4

Jonathan H. Raberg<sup>1,2</sup>, David J. Harning<sup>1,2</sup>, Sarah E. Crump<sup>1,\*</sup>, Greg de Wet<sup>1,†</sup>, Aria Blumm<sup>1,‡</sup>, Sebastian Kopf<sup>1</sup>, Áslaug Geirsdóttir<sup>2</sup>, Gifford H. Miller<sup>1</sup>, Julio Sepúlveda<sup>1</sup>

Sebastian Kopf<sup>1</sup>, Áslaug Geirsdóttir<sup>2</sup>, Gifford H. Miller<sup>1</sup>, Julio Sepúlveda<sup>1</sup>
1, Department of Geological Sciences and Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, 80309, USA
Faculty of Earth Sciences, University of Iceland, Revkjavík, Iceland

9 2, Faculty of Earth Sciences, University of Iceland, Reykjavík, Iceland 10

\*Currently at the Ecology and Evolutionary Biology Department, University of California Santa Cruz, Santa Cruz, CA
 95064

13 <sup>†</sup>Currently at the Department of Geosciences, Smith College, Northampton, MA 01063

14 <sup>‡</sup>Currently at the Department of Geosciences, University of Arizona, Tucson, AZ 85721 15

16 Correspondence to: Jonathan H. Raberg (jonathan.raberg@colorado.edu)

17 18

19 Abstract. Distributions of branched glycerol dialkyl glycerol tetraethers (brGDGTs) are frequently employed for 20 reconstructing terrestrial paleotemperatures from lake sediment archives. Although brGDGTs are globally ubiquitous, 21 the microbial producers of these membrane lipids remain unknown, precluding a full understanding of the ways in 22 which environmental parameters control their production and distribution. Here, we advance this understanding in 23 three ways. First, we present 43 new high-latitude lake sites characterized by low mean annual air temperatures 24 (MATs) and high seasonality, filling an important gap in the global dataset. Second, we introduce a new approach for 25 analyzing brGDGT data in which compound fractional abundances (FAs) are calculated within structural groups based 26 on methylation number, methylation position, and cyclization number. Finally, we perform linear and nonlinear 27 regressions of the resulting FAs against a suite of environmental parameters in a compiled global lake sediment dataset 28 (n = 182). We find that our approach deconvolves temperature, conductivity, and pH trends in brGDGTs without 29 increasing calibration errors from the standard approach. We also find that it reveals novel patterns in brGDGT 30 distributions and provides a methodology for investigating the biological underpinnings of their structural diversity. 31 Warm-season temperature indices outperformed MAT in our regressions, with Months Above Freezing yielding the 32 highest-performing model (adjusted  $R^2 = 0.91$ , RMSE =  $1.97^{\circ}$ C, n = 182). The natural logarithm of conductivity had 33 the second-strongest relationship to brGDGT distributions (adjusted  $R^2 = 0.83$ , RMSE = 0.66, n = 143), notably 34 outperforming pH in our dataset (adjusted  $R^2 = 0.73$ , RMSE = 0.57, n = 154) and providing a potential new proxy for 35 paleohydrology applications. We recommend these calibrations for use in lake sediments globally, including at high 36 latitudes, and detail the advantages and disadvantages of each.

37

#### 38 1 Introduction

39 Paleotemperature records from lake sediment archives are highly sought after in studies of terrestrial 40 paleoclimate. Bacterial branched glycerol dialkyl glycerol tetraether (brGDGT) lipids have solidified themselves as 41 an important tool in this pursuit (Fig A1; Schouten et al., 2013). First isolated from peat (Sinninghe Damsté et al., 42 2000), these membrane lipids have since been measured in increasingly diverse settings, from marine, soil, lacustrine,

- 43 and riverine locations (Hopmans et al., 2004; Weijers et al., 2006; Pearson et al., 2011; De Jonge et al., 2014b,
- 44 respectively) to hot springs, fossil bones, groundwater, deep ocean trenches, and methane seeps (Li et al., 2014; Dillon
- 45 et al., 2018; Ding et al., 2018; Xiao et al., 2020; Zhang et al., 2020, respectively). Their ubiquity in nature has given
- 46 them widespread applicability as environmental proxies; brGDGTs have been used to reconstruct temperature in a
- 47 variety of archives including lake sediments (e.g. de Wet et al., 2016), marine sediments (e.g. Dearing Crampton-
- 48 Flood et al., 2018), peat (e.g. Zheng et al., 2017), loess (e.g. Lu et al., 2019), and fossil bone (e.g. Zhao et al., 2020).
- 49 They have additionally been used to reconstruct lake water pH (e.g. Cao et al., 2017). As the microbial producers of
- 50 brGDGTs remain elusive (Sinninghe Damsté et al., 2018), these paleoclimate reconstructions currently rely on
- 51 empirical calibrations at both the regional and global level.

52 At the heart of brGDGT calibrations is the observation that the degree of alkyl-chain methylation and 53 cyclization are correlated to environmental temperature and pH, respectively. These relationships were first quantified 54 by the Methylation index and Cyclization ratio of Branched Tetraethers (MBT and CBT) in a global soil dataset 55 (Weijers et al., 2007). The authors proposed physiological explanations for both connections, positing that an increase 56 in methylation number will enhance membrane fluidity, a desirable trait in cold environments, while a greater number 57 of cyclic mojeties could improve proton permeability, an advantageous adaptation at high pH. These physiological 58 responses have precedent in other bacterial lipid classes (Reizer et al., 1985; Beales, 2004; Yuk and Marshall, 2004) 59 and appear to function for brGDGTs as well. However, genomic analyses of environmental samples (Weber et al., 60 2018; De Jonge et al., 2019; van Bree et al., 2020) have suggested that differences in brGDGT distributions may also 61 stem from shifts in bacterial community composition. Variations in the position of alkyl-chain methylations (Fig. A1; 62 De Jonge et al., 2014a) further complicate the picture, with most studies showing 5-methyl brGDGT isomers to 63 correlate better with temperature than their 6-methyl counterparts (e.g. Russell et al., 2018), but others arriving at the 64 opposite result (e.g. Dang et al., 2018). These isomeric variations have additionally been shown to correlate with pH 65 in lake sediments (Dang et al., 2016). These discoveries highlight the multifaceted nature of the empirical relationship 66 between brGDGTs and environmental gradients and the need for further study.

67 Without a clear mechanistic understanding of the dependencies of brGDGTs on environmental parameters 68 and no brGDGT-producing model organisms currently available for laboratory experimentation, researchers have 69 relied on statistical methods to construct empirical brGDGT calibrations. The majority of recent calibrations have 70 employed a variety of statistical techniques to construct linear or polynomial regressions using brGDGT fractional 71 abundances (FAs; De Jonge et al., 2014a; Martínez-Sosa et al., 2020b; Pérez-Angel et al., 2020). The fractional 72 abundance  $fx_i$  of a compound  $x_i$  in a set of *n* compounds is defined as,

73

$$fx_i = x_i / (x_1 + x_2 + \dots + x_n)$$
(1)

74 where any x is the absolute abundance of the given compound in the set. For brGDGTs, these FAs are traditionally 75 calculated using all 15 commonly measured compounds (Fig. A1),

 $fx_i = x_i/(Ia + Ib + Ic + IIa + IIb + IIc + IIIa + IIIb + IIIc + IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc') (2)$ where  $x_i$  is any given compound in the denominator. By grouping together all 15 common brGDGTs, this approach makes no prior assumptions about the relationships between the compounds themselves and maximizes the degrees

79 of freedom available when exploring a dataset. As relationships between brGDGTs and environment parameters are 80 not yet fully understood, this indiscriminate approach is appropriate. However, by lumping compounds of various 81 types and abundances into the denominator, the approach can also dampen meaningful trends and obscure important 82 relationships, especially for less abundant molecules. This adverse effect has been recognized for other lipid 83 biomarkers and has led to, for example, the exclusion of crenarchaeol from the TEX<sub>86</sub> index (Schouten et al., 2002) 84 and tetra-unsaturated alkenones from the U<sup>K</sup>'<sub>37</sub> index (Prahl and Wakeham, 1987). Numerous ratio-based indices have 85 been developed for brGDGTs that similarly exclude low-abundance (e.g. MBT'; Peterse et al., 2012) or problematic 86 (e.g. MBT'<sub>5Me</sub>; De Jonge et al., 2014a) compounds. However, a selective approach to fractional abundance

87 calculations has been hitherto unexplored.

88 On the other side of the calibration equations are the environmental variables that are regressed against 89 brGDGT indices and FAs. Mean annual air temperature (MAT) has been the traditional target of brGDGT calibrations 90 in lake sediments (e.g. Tierney et al., 2010; Loomis et al., 2012). It was recognized early on, however, that brGDGT-91 derived temperatures in cold regions may more accurately reflect warm-season temperatures (Pearson et al., 2011; 92 Sun et al., 2011), an hypothesis that was strongly supported in high-latitude lake sediments (Shanahan et al., 2013; 93 Peterse et al., 2014; Foster et al., 2016). Since the methodological advances that allowed for the separation of 5- and 94 6-methyl isomers (De Jonge et al., 2014a) and the development of new calibrations, both modern (Hanna et al., 2016; 95 Dang et al., 2018; Cao et al., 2020) and paleo (Super et al., 2018; Thomas et al., 2018; Crump et al., 2019; Harning et 96 al., 2020) studies have continued to support a warm-season bias. Additionally, a recent Bayesian calibration found the 97 mean temperature of Months Above Freezing (MAF) to be the only mode to significantly correlate with brGDGT 98 distributions in a global lake sediment dataset (Martínez-Sosa et al., 2020). However, as the source organisms of 99 brGDGTs remain incompletely known, other measures of warm-season temperature such as peak warmth or 100 cumulative heat may prove to be more biologically relevant. Furthermore, the nature of the warm-season bias has yet 101 to be tested thoroughly in the regions in which it is most pronounced – namely, those with low MAT and high 102 seasonality. As these are the regions that are currently experiencing the most rapid climate change (Landrum and 103 Holland, 2020), their temperature histories are of high interest (Miller et al., 2010) and the quantification of the 104 brGDGT warm-season bias is an important target of study.

Outside of temperature, lake water pH is the most common focus of calibration studies (e.g. Russell et al., 2018). However, numerous other variables including conductivity (Tierney et al., 2010; Shanahan et al., 2013), dissolved oxygen (DO; Colcord et al., 2017; Weber et al., 2018; van Bree et al., 2020; Yao et al., 2020), nutrient availability (Loomis et al., 2014a), and lake mixing regime (Loomis et al., 2014b; van Bree et al., 2020) have been shown to be potentially important controls on brGDGT distributions. Modern calibrations do not currently exist for these environmental variables, largely due to the complexity of the relationships and data limitations.

Uncertainties in the physical sources of brGDGTs on the landscape add an additional challenge for calibration efforts. BrGDGTs produced in soils and rivers can contribute to the lacustrine lipid pool to varying degrees, and the exact proportion of soil- and lake-derived brGDGTs at a particular site requires detailed study to quantify (e.g. Naeher et al., 2014; Guo et al., 2020; van Bree et al., 2020). Further complicating the matter, multiple sources of brGDGTs have been observed within lakes, including at varying depths (e.g. Yao et al., 2020) and within the sediment itself (e.g. Buckles et al., 2014). The intensive studies required to disentangle these various sources are absent for the vast majority of lakes in most calibration studies, undoubtedly contributing to their reported uncertainties. Furthermore, without a clear understanding of where brGDGTs are being produced, it can be difficult to determine which

The without a clear understanding of where of 0101018 are being produced, it can be difficult to determine which

- environmental variables are most appropriate to reconstruct. For example, lacustrine brGDGTs were successfully
- 120 calibrated to lake water temperature at a site in Greenland (Zhao et al., 2021), but such an effort would likely be less
- 121 successful for a lake with substantial soil input.

122 In this study, we aim to improve lake sediment calibrations for brGDGTs in three ways. First, we extend the 123 global calibration dataset to include high-latitude sites by adding surface sediment from 43 lakes in the Eastern 124 Canadian Arctic, Northern Quebec, and Iceland. Second, we selectively group brGDGTs based on their three structural 125 variables - methylation number, methylation position, and cyclization number - and use FAs calculated within these 126 structural sets to investigate isolated relationships between environmental parameters and each structural variable in 127 turn. Finally, we analyze the relationship between the compiled global dataset and MAT, four warm-season 128 temperature indices, pH, conductivity, DO, and lake geometry and generate empirical calibrations for use in lake 129 sediments globally.

130

# 131 2 Methods

# 132 2.1 Study sites and sample collection

133 Surface sediments (0-0.5, 0-1, or 0-2 cm; Ekman box corer or core-top sediment) were collected from 43 134 lakes (26 from Iceland; 16 from Baffin Island, Arctic Canada; 1 from Northern Quebec; Fig. 1 insets) between 2003 135 and 2019. For 28 of these lakes, water temperature, pH, conductivity, and DO were measured at the time of sampling 136 during the summers of 2017-2020 using a multiparameter probe (HydroLab HL4, OTT HydroMet). These parameters 137 were additionally measured beneath the lake ice in Feb-May of 2018-2020 for 11 lakes. All but one of these lakes 138 experienced depleted bottom water oxygen levels under the ice relative to ice-free conditions. We therefore assume 139 that our ice-free water chemistry profiles do not capture the minimum DO (DOmin) levels experienced by our Canadian 140 and Icelandic lakes and exclude them from our analysis of DOmin, with the exception of the site from Northern Quebec, 141 which contained a summer oxycline. All other water chemistry parameters were averaged across all depths and seasons 142 before being used in calibrations.





Figure 1. Map of sites included in this study. Red and blue: Iceland and Canada (this study; see insets); Orange: Russell
et al. (2018); Green: Dang et al. (2018); Pink: Weber et al. (2018); Yellow: Cao et al. (2020); Light blue: Qian et al. (2019);
Purple: Ning et al. (2019).

147 Previously published data from 36 lakes in Central Europe (Weber et al., 2018), 65 lakes in East Africa 148 (Russell et al., 2018), and 38 lakes in China (Dang et al., 2018; Ning et al., 2019; Qian et al., 2019; Cao et al., 2020), 149 were added to the dataset for a total of 182 data points (Fig. 1). Average brGDGT FAs were used for lakes with 150 multiple surface sediment samples. Lake surface areas (SAs) were taken from published datasets or estimated using 151 DigitalGlobe imagery. Lake volumes were estimated by approximating each lake basin as a hemiellipsoid (volume = 152  $4/3 \times SA \times maximum depth$ ). Water chemistry parameters were taken from the literature where available or else 153 excluded from our analyses. The number of datapoints for each environmental parameter is available in Table S1. 154 Some parameters were correlated in the dataset, most notably conductivity and pH (r = 0.75; Table S2).

155

# 156 2.2 Sample extraction and analysis

157 Roughly 1 g of freeze-dried sediment was extracted using either an accelerated solvent extractor (ASE 200 158 DIONEX: 10 samples) or a modified Bligh and Dyer (BD; 33 samples) method. We provide a comparison of both 159 extraction methods below. For the ASE method, samples were extracted twice using 9:1 (v:v) 160 dichloromethane:methanol (DCM:MeOH) at 100°C and 2,000 psi. Total lipid extracts (TLEs) were redissolved in 161 99:1 (v:v) hexane:isopropanol (Hex:IPA) and filtered (0.45 µm, PTFE) before analysis. The remaining 33 samples 162 were extracted using a modified BD procedure (Wörmer et al., 2013). Briefly, sediment was vortexed and sonicated 163 in Mix A (DCM:MeOH:50mM Phosphate buffer (aq., pH 7.4) [1:2:0.8, v:v:v]). The mixture was then centrifuged at 164 3,000 rpm and 10°C for 10 minutes and the supernatant was collected in a glass separatory funnel. The process was 165 performed twice with Mix A, twice with Mix B (DCM:MeOH:5% Trichloroacetic acid buffer (aq., pH 2) [1:2:0.8, 166 v:v:v]), and once with Mix C (DCM:MeOH [1:5, v:v]). Equal volumes of HPLC-grade water and DCM were added 167 to induce separation. The organic fraction was collected and dried under a nitrogen stream. The aqueous phase was

washed once with DCM and the organic fraction was added to the extract. The TLE was then redissolved in 99:1 (v:v)
 Hex:IPA and filtered (0.45-µm, PTFE) before analysis.

- 170 We analyzed brGDGTs using a Thermo Scientific UltiMate 3000 high-performance liquid chromatography 171 instrument coupled to a Q Exactive Focus Orbitrap-Quadrupole high-resolution mass spectrometer (HPLC-MS) via 172 an atmospheric pressure chemical ionization (APCI). We achieved chromatographic separation using a slightly 173 modified version (Crump et al., 2019; Harning et al., 2019; Pérez-Angel et al., 2020) of the HPLC method described 174 by Hopmans et al. (2016). Due to observed deterioration of chromatography over time, we lowered the initial 175 concentration of eluent B from 18% to 14% to maintain optimal separation of the 5- and 6-methyl isomers. A C46 176 GDGT internal standard (Huguet et al., 2006) was added to the TLE immediately after extraction and was used to 177 quantify brGDGT yields. Peaks with a signal-to-noise ratio of at least 2:1 were considered for analysis.
- 178

# 179 2.3 Comparison of ASE and BD Extraction Methods

180 To ensure that brGDGT distributions were agnostic to our extraction method, we extracted three surface 181 sediments, two suspended particulate matter (SPM) samples (2.5 L lake water filtered onto 0.3 µm glass fiber filters), 182 and three soils from Baffin Island using the ASE and BD methods in parallel (Fig. S1). The mean difference in 183 MBT'<sub>5Me</sub> (Eq. A3) between the two extraction methods was  $0.006 \pm 0.004$ , or  $2 \pm 2\%$ . This translates to a MBT'<sub>5Me</sub>-184 derived temperature difference of  $0.2 \pm 0.2^{\circ}$ C using recent calibrations for soils (Naafs et al., 2017) and lake sediments 185 (Russell et al., 2018), which is well below their respective RMSEs of 5.3°C and 2.14°C (Fig. S2). To test for 186 compound-specific differences, we calculated percent differences in FAs between the two methods. For compounds 187 with FAs > 0.05 in the ASE method, the mean absolute percent difference compared to BD was  $4 \pm 3\%$ . For the lower 188 abundance compounds (FA  $\leq 0.05$ ), this difference was higher (19  $\pm$  17%). No biases in the FA differences were 189 found in either case (difference < standard deviation).

- We further extracted the BD sample residue with the ASE method to determine if any brGDGTs remained after BD extraction. On average, we recovered only an additional  $0.8 \pm 0.6\%$  brGDGTs. These residual brGDGTs had a similar MBT'<sub>5Me</sub> to that of original BD extract (mean difference =  $0.02 \pm 0.01$ , equivalent to  $0.5 \pm 0.4$ °C) and were not present in high enough abundances to significantly affect the overall BD distributions (Fig. S2). We therefore conclude that there are no significant differences between samples extracted with the two methods and treat them identically in the analyses that follow.
- 196

# **197 2.4 Air Temperatures**

Monthly air temperature averages were gathered using the following methods. For nine sites on Baffin Island, one year of *in situ* two-meter air temperature data from five temperature loggers (one- to four-hour resolution, Thermochron iButtons, Maxim Integrated Products) was converted to a 30-year monthly climate normal (1971 to 2000) using a transfer function to relate local data to nearby meteorological stations (Department of Environment, Government of Canada). For the remaining Canadian sites as well as all sites in Iceland, we used the WorldClim database (Fick and Hijmans, 2017) to generate 30-year climate normals for the same time period (1970 to 2000). 204 Monthly temperatures for Central European sites were derived using monthly altitudinal lapse rates constructed from 205 climate normals (1970 to 2013) of 148 meteorological stations (Federal Office of Meteorology and Climatology: 206 MeteoSwiss). Monthly temperature data was not available for the East African lakes. However, the seasonality 207 (standard deviation of monthly temperatures) of these lakes is low ( $0.5 \pm 0.2$  °C in the WorldClim database, or < 2% 208 of range of the dataset). We therefore approximate all monthly temperatures to be equivalent to MAT for these lakes. 209 Monthly temperature data from all other studies were either published or provided by the authors.

210 We used the above monthly air temperatures to calculate MAT and four warm-season temperature indices. 211 Three of these indices represent an average temperature for the warmer portion of the year: mean temperature of 212 Months Above Freezing (MAF), Mean Summer Temperature (MST; mean of June, July, and August in the Northern 213 Hemisphere and December, January, and February in the Southern Hemisphere), and mean Warmest Month 214 Temperature (WMT). These indices capture average temperatures for the warm season, but are unaffected by its 215 duration. We therefore additionally calculate the Summer Warm Index (SWI), defined as the cumulative sum of all 216 monthly temperatures above 0°C. This index represents an important control on vegetation patterns at high latitudes 217 and is a useful alternative to the Growing Degree Days Above 0°C (GDD<sub>0</sub>) index when daily temperature data is not 218 available (e.g. Raynolds et al., 2008). For our five in situ temperature loggers in the Eastern Canadian Arctic for which 219 sub-daily temperature data is available,  $GDD_0$  and SWI are highly correlated ( $R^2 = 0.998$ ). We also quantified shoulder 220 season temperatures by calculating averages of the first and last, or first and last two, positive degree months. 221 However, the monthly resolution of the dataset proved too coarse to apply this method robustly.

#### 223 2.5 Statistical Methods

222

224 To construct calibrations between brGDGT FAs and environmental variables, we used the following method. 225 Each FA was first regressed alone against the environmental variable being investigated. Compounds with a 226 correlation *p*-value  $\geq 0.01$  were considered non-significant and removed from further analysis (Pérez-Angel et al., 227 2020). Fits were then constructed from the remaining compounds using two independent approaches. The first 228 approach was stepwise forward selection/backwards elimination (SFS/SBE) using the MASS package (Venables and 229 Ripley, 2002) in R (R Core Team, 2018). This approach finds the best fit by sequentially adding (SFS) or removing 230 (SBE) terms in a generalized linear model and evaluating the resulting fit using the Bayesian Information Criterion 231 (BIC; Schwarz, 1978). The approach is common for constructing brGDGT calibrations (e.g. Dang et al., 2018; Russell 232 et al., 2018), but it is not exhaustive. We therefore additionally used the leaps package (Lumley, 2020) in R to evaluate 233 all possible linear combinations of fitting variables (the "combinatoric" approach, Pérez-Angel et al., 2020). We again 234 used the BIC to select the best fit. For both methods, we imposed the additional criterion that each of the resulting 235 fitting variables must itself be statistically significant (p < 0.01). To help avoid overfitting, we additionally used the 236 adjusted R<sup>2</sup> to evaluate calibration performances. Some of our variables (conductivity, depth, surface area to depth, 237 and volume) spanned multiple orders of magnitude. We performed regressions against the natural logarithm of these 238 variables, which greatly improved their normality. 239 We applied this calibration procedure to the FAs of each brGDGT structural set and subset defined below

240 (Sect. 3). In the subset-specific calibrations, it was sometimes possible for a single compound to dominate (FA = 1;

- 241 see section S1.3). These samples were generally clear outliers resulting from the low natural abundances of all other
- 242 members of the subset and they were therefore removed from the subset-specific calibration models. We additionally
- 243 tested for linear regressions against a number of previously defined brGDGT indices. A summary of these indices and
- 244 their definitions is provided in the Appendix.
- 245
- All correlations reported in the text and figures were significant (p < 0.01) except those marked with an 246 asterisk or with a *p*-value provided. All  $R^2$  values reported in the text and figures are adjusted  $R^2$ .
- 247

#### 248 **3** Partitioning brGDGTs into structural sets for FA calculations

249 The structure of brGDGTs can vary in three (currently observed) ways: methylation number, methylation 250 position, and cyclization number. These variations result in 15 commonly measured brGDGTs (Fig. A1). The standard 251 approach in brGDGT analysis is to calculate fractional abundances using all 15 of these compounds (Eq. 2). By 252 mathematically mixing brGDGTs with varying methylation number, methylation position, and cyclization number, 253 however, this approach risks convoluting the influences of disparate environmental variables. Here, we present a 254 method of grouping brGDGTs that highlights one type of structural variation (e.g. methylation number) while holding 255 one or both of the others constant. By calculating FAs within each group, we aim to deconvolve the influences of 256 temperature, pH, and other environmental variables on the structural variations of brGDGTs. 257



258

Figure 2. Schematic of the basic (a-c) and combined (d-g) brGDGT structural sets. Fractional abundances are calculated
 within each boxed group independently (Eq. (3-9) and Table A1). Schematic structures highlight the defining alkyl-chain
 moieties, with cyclopentane rings filled in for emphasis and C6 methylations denoted in red. Complete structures are
 available in Fig. A1.

263 To explore how changes in methylation number alone relate to environmental parameters, we constructed 264 the Methylation (Meth) set (Fig. 2a). This set was generated by grouping brGDGTs with the same number of 265 cyclopentane rings and the same methylation positions. Fractional abundances calculated within the Meth set solely 266 reflect changes in methylation number, while ring number and isomer designation are held constant. Within the Meth 267 set, we defined the Meth-5Me and Meth-6Me subsets as those that contained only 5- and 6-methyl lipids, respectively. 268 As the tetramethylated brGDGTs (Ia, Ib, and Ic) are neither 5-methyl nor 6-methyl isomers, we generated versions of 269 these two subsets that excluded (Meth-5Me, Meth-6Me) and included (Meth-5Me+, Meth-6Me+) these compounds 270 (Fig. S3a-d). We found that the correlation between 5-methyl isomers and temperature was significantly improved 271 when the tetramethylated compounds were included in their FA calculations, while the opposite was true for the 6-272 methyl compounds (see discussion in Sect. 4.2.2). We therefore grouped the tetramethylated compounds with the 5-273 methyl compounds (and not with the 6-methyl compounds) when generating the Meth set. FAs for the Meth set were 274 calculated using Eq. (3),

275 
$$fxy_{Meth} = xy / \sum_{n=1}^{III} ny; fxy'_{Meth} = xy' / \sum_{n=II}^{III} ny'$$
(3)

where fxy and xy are the fractional and absolute abundances of the brGDGT with Roman numeral x (I, II, or III) and alphabet letter y (a, b, or c) and tetramethylated compounds are grouped with 5-methyl isomers.

We next defined the Cyclization (Cyc) set to examine the relationship of brGDGT ring number with environmental variables (Fig. 2b). The Cyc set was formed by grouping brGDGTs with the same number and position of methylations. With these variables held constant, variations in the Cyc FAs reflect only variations in the number of cyclopentane moieties. We defined the Cyc-5Me and Cyc-6Me subsets as those containing only 5- and 6-methyl isomers, respectively (Fig. S3e-f). FAs for the Cyc set were calculated using Eq. (4),

283 
$$fxy^{(\prime)}_{Cyc} = xy^{(\prime)} / \sum_{m=a}^{5} xm^{(\prime)}$$
(4)

where  $f_{xy}$  and  $x_y$  are the fractional and absolute abundances of the 5- or 6-methyl brGDGT with Roman numeral x (I, II, or III) and alphabet letter y (a, b, or c).

- 286 Third, we defined the Isomer (Isom) set to isolate changes in the relative abundances of brGDGT isomers 287 (Fig. 2c). The Isom set was constructed by grouping brGDGTs with the same number of methylations and cyclizations. 288 Its FAs are solely a measure of the relative abundances of 5- and 6-methyl isomers, without the convoluting influence 289 of ring or methylation number variations. The isomeric diversity of brGDGTs is large, however, and there are 290 structural variations that are not controlled for within this set. For example, hexamethylated brGDGTs have two 291 isomers: one with methylations on different alkyl chains (e.g. at C5 and C5') and another with methylations on the 292 same chain (e.g. C5 and C24; De Jonge et al., 2013). As these compounds coelute, they cannot be treated independently 293 at this time. Additionally, it is unclear whether brGDGT-IIIa" (Weber et al., 2015), which contains both a C5 and a 294 C6' methylation, should be grouped with 5- or 6-methyl brGDGTs, or whether the three recently-identified 7-methyl 295 brGDGTs (Ding et al., 2016) should be included in the Isom series. As these compounds are rarely reported, we 296 excluded them from our analysis here, but suggest the possibility of expanding the Isom set to include them in the 297 future should more data become available. The Isom FAs were calculated using Eq. (5),
- 298  $fxy^{(\prime)}_{Isom} = xy^{(\prime)} / \sum_{isomers} xy$ (5)

where fxy and xy are the fractional and absolute abundances of the 5- or 6-methyl brGDGT with Roman numeral x (I, II, or III) and alphabet letter y (a, b, or c) and "isomers" refers to 5- and 6-methyl brGDGTs. The Isom set contained groups of two compounds each, making their FAs redundant. We therefore used only the 6-methyl FAs in our analysis.

The Meth, Cyc, and Isom sets each allow only one structural component of brGDGTs to vary. It is possible, however, that two structural alterations occur in tandem in response to the same environmental variable. We therefore defined three additional sets that hold one variable constant while allowing the other two to vary. The first is the Meth-Isom (MI) combination set (Fig. 2d). In this set, brGDGTs with the same ring number are grouped together, while both methylation number and position are allowed to vary. The Cyc-Isom (CI; Fig. 2e) is analogously constructed by holding methylation number constant, while the Meth-Cyc (MC; Fig. 2f) set holds methylation position constant (again treating tetramethylated brGDGTs as 5-methyl compounds, Fig. S3g-j). Finally, we defined the Full set (Fig. 309 2g), which takes the standard approach of allowing all three structural characteristics to vary freely by grouping all 15

310 commonly measured brGDGTs together. The FAs of the combined sets are calculated using Eq. (6-9),

311 
$$fxy^{(\prime)}{}_{MI} = xy^{(\prime)} / \sum_{isomers} \sum_{n=1}^{III} ny$$
(6)

312 
$$fxy^{(\prime)}{}_{CI} = xy^{(\prime)} / \sum_{isomers} \sum_{m=a}^{c} xm$$
(7)

313 
$$f x y_{MC} = x y / \sum_{n=I}^{III} \sum_{m=a}^{c} nm; \ f x y'_{MC} = x y' / \sum_{n=II}^{III} \sum_{m=a}^{c} nm'$$
(8)

314 
$$fxy^{(\prime)}_{Full} = xy^{(\prime)} / \sum_{isomers} \sum_{n=l}^{ll} \sum_{m=a}^{c} nm$$
(9)

315 where fxy and xy are the fractional and absolute abundances of the 5- or 6-methyl brGDGT with Roman numeral x (I,

316 II, or III) and alphabet letter y (a, b, or c) and tetramethylated compounds are treated as 5-methyl isomers. An expanded

317 guide to FA calculations is provided in Table A1.

318 As a proof of concept, we show that the fractional abundance of just one compound, brGDGT-Ia, can be 319 calculated within different structural sets to provide either a strong temperature or pH correlation, without a strong 320 cross-correlation. When the standard FA is calculated using all 15 compounds (flaFull), a moderate correlation is found 321 with MAF ( $R^2 = 0.61$ ) and none with pH ( $R^2 = 0.04$ , p = 0.014; Fig. 3). When cyclization is held constant by using 322 the MI set, correlation with MAF increases ( $R^2 = 0.75$ ) while that with pH remains uncorrelated ( $R^2 = 0.13$ ). The 323 temperature correlation increases further when isomer designation is controlled for as well ( $fIa_{Meth-5Me^+}$ ,  $R^2 = 0.88$ ), 324 with an  $R^2$  nearly matching that of MBT'<sub>5Me</sub> ( $R^2 = 0.89$ ) and the pH correlation remaining low ( $R^2 = 0.27$ ). In contrast, 325 the correlation with temperature disappears for the analogous 6-methyl subset (fla<sub>Meth-6Me+</sub>,  $R^2 = 0.08$ ). Finally, 326 allowing only ring number to vary (fla<sub>Cvc</sub>) effectively erases the correlation with temperature ( $R^2 = 0.18$ ) and instead 327 provides a correlation with pH that, while modest, is already higher than any reported for a lake sediment calibration 328 to date ( $R^2 = 0.59$ ).

329

330

	E٨	Compounds	MAF R <sup>2</sup>	nH R <sup>2</sup>					
	IA	Compounds		рпп		FΔ	Compounds	MAF R <sup>2</sup>	DH R <sup>2</sup>
	fla	All (15)	0.61	0.04*			compounds		PIIN
_	Full	, (13)	0.01	0.01		fla <sub>Meth-5Me+</sub>	la, lla, llla	0.88	0.27
	fla	la, lla, llla, lla', llla'	0.75	0.13				0.00	0.27
	щ <sub>МI</sub>	14/114/114/114	0.75	0.15		fla	la, lla', llla'	0.08	0.07
	fla	la lb lc	0.18	0.59		Meth-6Me+		0.00	0.07
	Cyc	10,10,10	0.10	0.55					

Figure 3. Adjusted R<sup>2</sup> values for a linear regression of environmental parameters against the fractional abundance (FA) of brGDGT-Ia calculated within different structural sets. Colors denote the strengths of the relationships, from the

minimum to the maximum observed coefficients of determination, with white being the median of the dataset.

331 332 333 334 335 "Compounds" denote all brGDGTs used in the FA calculation. All values are significant (p < 0.01) unless marked with an asterisk.

#### 337 4 Results and Discussion

#### 338 4.1 Distributions of brGDGTs in Icelandic and Canadian lake sediments

339 The new Icelandic and Canadian lake sediments bolster the global dataset on the cold end, extending the 340 lowest MAT from -0.2°C to -18°C and containing 28 of the 30 coldest samples by MAT. The low temperatures are 341 reflected in the brGDGT distributions of these sediments (Fig. 4), which contain on average a higher fIIIaFull and lower 342 fla<sub>Full</sub> than the other samples in this dataset. Additionally, the Canadian and Icelandic datasets provide important end-343 member samples, including those with the highest fIIIaFull, fIIIbFull, and fIIIcFull and the lowest or second-lowest fIaFull, 344 flb<sub>Full</sub>, and flc<sub>Full</sub>. The new samples also contain some of the lowest fractional abundances of 6-methyl isomers, 345 providing low end-member values for fIIa'Full, fIIb'Full, fIIc'Full and below-average values for the hexamethylated 6-346 methyl brGDGTs.

347



349Figure 4. Average fractional abundances (calculated within the standard Full set) of new samples from Canada and350Iceland as compared to the average of the other samples in this study (Others).

351

348

# 352 **4.2** Temperature relationships with brGDGTs

In this section, we show that two adjustments can be made to significantly improve the correlations between brGDGT FAs and temperature in lake sediments: 1) replace MAT with a warm-season temperature index and 2) use FAs calculated within the Meth structural set. The effect of these two adjustments on one representative compound, brGDGT-Ia, is shown from left to right in Fig. 5. The relationship between temperature and fIa is improved from a weak but significant correlation with a large error ( $R^2 = 0.49$ ; RMSE = 7.00°C) to a stronger correlation with a smaller error ( $R^2 = 0.88$ ; RMSE = 2.42°C). The two adjustments are detailed in the following sections.



Figure 5. From left to right, the effects of substituting the Months Above Freezing (MAF) warm-season index for Mean
 Annual Temperature (MAT) and fla<sub>Meth</sub> for fla<sub>Full</sub> on the relationship between the fractional abundance of brGDGT-Ia
 and temperature.

359

364

#### 4 4.2.1 Warm-season temperatures outperform MAT

To assess the possibility of a warm-season bias in brGDGT temperature reconstructions, we tested the relationships between brGDGTs and five air temperature indices: MAT, MAF, MST, WMT, and SWI (Sect. 2.4). The correlation between these variables and the FA of one representative compound, fla<sub>Meth</sub>, is shown in Fig. 6. In all cases, substituting MAT with a warm-season temperature variable draws samples with strong seasonality into the main body of data. MAF performs best for brGDGT-Ia ( $R^2 = 0.88$ ), with high-seasonality lakes falling progressively out of alignment for SWI ( $R^2 = 0.84$ ), MST ( $R^2 = 0.64$ ), and WMT ( $R^2 = 0.60$ ). This result was upheld when performing temperature calibrations; MAF outperformed all other measures of temperature examined in this study (Sect. 4.4.1).

372 A possible explanation for the success of the MAF temperature index is that the activity of brGDGT-373 producing microbes may be heavily depressed under lake ice and/or in frozen soils (Pearson et al., 2011; Shanahan et 374 al., 2013; Peterse et al., 2014; Cao et al., 2020). However, studies employing sub-seasonal sampling of sediment traps 375 and suspended particulate matter in two mid-latitude lakes have shown that brGDGTs are produced within the water 376 column throughout the year, despite the presence of ice cover (Woltering et al., 2012; Loomis et al., 2014b). These 377 and other studies employing similar sampling techniques (Hu et al., 2016; Miller et al., 2018; Weber et al., 2018; van 378 Bree et al., 2020) have additionally found production of brGDGTs to be dependent on the degree and timing of lake 379 mixing versus stratification. Though heightened biological activity may still be the underlying driver of the observed 380 warm-season bias, these depth- and time-resolved studies paint a complex picture of brGDGT production in lakes that 381 precludes a simple explanation. Unfortunately, as knowledge of the timing, extent, and temperature of ice cover and 382 mixing events is lacking for the vast majority of lakes in this study, these effects cannot be tested here. We therefore 383 stress the empirical nature of our MAF calibrations and the need for further study.





Figure 6. Relationships between fla<sub>Meth</sub> and the five air temperature variables tested in this study, a) mean annual
 temperature (MAT), b) months above freezing (MAF), c) mean summer temperature (MST), d) warmest month
 temperature (WMT), and e) summer warmth index (SWI). Colors indicate air temperature seasonality (standard
 deviation of mean monthly temperatures). Dashed lines indicate mean or cumulative temperatures of 0°C for reference.
 R<sup>2</sup> values for the relationship between fla<sub>Meth</sub> and each temperature variable are provided in each subplot.

# **392 4.2.2** Temperature and the Methylation set

393 The Methylation set provided the strongest relationships between brGDGT FAs and MAF (Fig. 7). This set 394 both strengthened existing correlations (e.g. flaMeth, Fig. 7a) compared to the standard approach (Full set, R<sup>2</sup> values in 395 parentheses in Fig. 7) and generated new ones (e.g. fIlb<sub>Meth</sub>, Fig. 7e). Furthermore, many relationships between FAs 396 and temperature were revealed to be qualitatively similar regardless of ring number. For example, the FAs of all 397 tetramethylated compounds (fla<sub>Meth</sub>, flb<sub>Meth</sub>, flc<sub>Meth</sub>) had a strong positive linear relationship with temperature (Fig. 7a-398 c), while those of the 5-methyl pentamethylated compounds all had a noisier negative relationship (Fig. 7d-f). The 399 hexamethylated FA trends were less clear in part due to their lower abundances, but all were negatively correlated 400 with temperature and nonlinearities were apparent (Fig. 7g-l). These analogous trends show that the number of 401 methylations responds similarly to temperature regardless of the number of cyclopentane rings. Within a paleoclimate 402 lens, this observation opens up the possibility of independent temperature calibrations for un-, mono-, and bicyclized 403 brGDGTs (Sect. 4.4.1). From a biological standpoint, it could imply that methylation and cyclization play their 404 biological roles independently, as the former appears to vary more or less freely of the latter. 405



Figure 7. Relationships between the average air temperature of months above freezing (MAF) and brGDGT FAs
 calculated within the Meth set. R<sup>2</sup> values are provided for each subplot, with R<sup>2</sup> values for the standard Full FAs given in
 parentheses for comparison. *P-values* were < 0.01 except where marked with an asterisk. Note that plots of IIa', IIb', and</li>

410 IIc' are redundant because they exactly mirror those of IIIa', IIIb', and IIIc', respectively, and are therefore not shown.

411 The brGDGT temperature response appears to be agnostic to methylation position as well (Fig. 7g-i versus 412 j-l), but only when the tetramethylated brGDGTs (Ia, Ib, and Ic) are excluded from 6-methyl FA calculations. The 413 Meth-5Me+ subset (Fig. S3a) showed strong relationships between 5-methyl brGDGTs and MAF ( $R^2 \le 0.88$ ; Fig. S4). 414 On the other hand, the analogous Meth-6Me+ subset (Fig. S3c) was broadly uncorrelated with MAF ( $R^2 \le 0.29$ ; Fig. 415 S5). At present, there is no known mechanism whereby an additional methylation at the C5 position would have an 416 influence on membrane physiology in a way that a methylation at C6 would not (though the magnitude of the effect 417 may be stronger closer to the less fluid membrane surface; see discussion in Section S3). At first glance, then, the 418 markedly different responses of the Meth-5Me+ and Meth-6Me+ subsets to changes in temperature do not appear to 419 support a physiological basis for the empirical relationship between temperature and brGDGT methylation number. 420 However, when tetramethylated brGDGTs were excluded from the FA calculations for these compounds (Meth-5Me 421 and Meth-6Me subsets; Figs. S3b and d), statistically significant and qualitatively similar temperature relationships 422 did become visible for both isomer types (Figs. S6-7). An analogous result was found for the MC set (Figs. S3g-j and 423 S8-11). It is not clear at this time why the inclusion of tetramethylated brGDGTs improved temperature correlations 424 for 5-methyl compounds but weakened them for 6-methyl compounds. The discrepancy may imply one or a 425 combination of the following: 1) the isomers are produced by different organisms; 2) the isomers serve distinct 426 biological functions, either in addition to or apart from a temperature response; or 3) the currently measured 427 tetramethylated brGDGTs (Ia, Ib, and Ic) are not the precursors of 6-methyl brGDGTs. Regardless, this result allows 428 us to combine the higher-performing Meth-5Me+ and Meth-6Me subsets to generate the Meth set (Fig. 2a), which 429 maximizes the temperature responses of all 15 commonly measured brGDGT (Fig. 7).

430 While the Meth set highlights the relationships between brGDGT FAs and temperature, it simultaneously 431 weakens those with other environmental variables. FAs calculated in the standard Full set contain conductivity and 432 pH dependencies ( $R^2 \le 0.66$  and 0.50, respectively; Figs S25 and S32) that are greatly reduced in the Meth set ( $R^2 \le$ 433 0.40 and 0.28, respectively; Figs S19 and S26). This is evidence that many of the conductivity and pH relationships 434 visible in the Full FAs are in fact due to the mathematical mixing of brGDGTs with different cyclization numbers and 435 isomer designations. Holding these variations constant in the Meth FAs largely removes the effects of these 436 environmental variables (e.g. fIIaMeth in Fig. S19d). DO dependencies are weak in both the Full and Meth sets, but 437 slightly weaker in the latter ( $R^2 \le 0.40$  and 0.35, respectively, Figs S39 and S33). The Methylation set thus improves 438 compound-specific correlations with temperature while decreasing their dependencies on other environmental 439 variables.

440

#### 441 **4.3** Conductivity and pH relationships with brGDGTs

While pH is the traditional secondary target of brGDGT calibrations after temperature, numerous works have
suggested that conductivity plays an important role in controlling brGDGT distributions (Tierney et al., 2010;
Shanahan et al., 2013). Conductivity and pH often plot nearly colinearly in principal component analyses (Shanahan

- 445 et al., 2013; Dang et al., 2018; Russell et al., 2018) and were moderately correlated (r = 0.75) in this dataset (Table
- 446 S2). This correlation is not surprising, as pH will affect conductivity both directly (through the concentration of
- 447 H<sub>3</sub>O<sup>+</sup>/OH<sup>-</sup>) and indirectly (by altering the solubility of various ions). Furthermore, both pH and conductivity can be
- 448 affected by the same factors, such as catchment lithology and precipitation chemistry (cf. Wetzel, 2001). BrGDGT
- 449 FAs in this study showed similar relationships with pH and conductivity, but also displayed potentially important
- 450 differences. We therefore treat them separately in our analyses but discuss them together in this section.
- 451

# 452 4.3.1 The Isomer set and conductivity

453 Conductivity provided the strongest compound-specific correlations with brGDGT FAs after temperature 454 (Figs. 8 and S23). Of the basic brGDGT sets (Meth, Cyc, and Isom), the Isom set had the highest statistical 455 performance (Fig. 8). The FAs of all 6-methyl brGDGTs showed a positive linear correlation with conductivity in this 456 set, with coefficients of determination as high as  $R^2 = 0.70$  (Fig. 8d). Furthermore, this positive correlation was broadly 457 independent of both methylation number and cyclization number, indicating that methylation position varies with 458 conductivity irrespective of other structural properties. The 6-methyl brGDGTs were also all positively correlated 459 with pH, but more weakly so ( $R^2 \le 0.54$ , Fig. S28). A relationship between brGDGT isomers and pH in lake sediments 460 has been previously observed and quantified by the isomerization of branched tetraethers (IBT, Eq. A14; Ding et al., 461 2015) and the Isomer Ratio of 6-methyl isomers (IR<sub>6Me</sub>, Eq. A8; Dang et al., 2016). However, these indices were more 462 closely tied to conductivity (IBT  $R^2 = 0.65$ ,  $IR_{6Me} R^2 = 0.66$ ) than pH (IBT  $R^2 = 0.55$ ,  $IR_{6Me} R^2 = 0.49$ ) in our dataset. 463 These results indicate that isomer abundances are primarily dependent on conductivity, but have some relation to pH 464 as well. Temperature correlations were also present in the Isom subset ( $R^2 \le 0.57$  with MAF, Fig. S14). Further study 465 is needed to elucidate whether temperature acts as a secondary control on isomer abundances or whether these 466 correlations are simply due to the inherent connection between MAF and conductivity in our dataset (r = 0.70, Table S1). DO provided little to no correlation in the Isom set ( $R^2 \le 0.38$ , Fig. S35). 467 468



469

Figure 8. Relationships between the natural logarithm of conductivity and brGDGT FAs calculated within the Isom set.
 R<sup>2</sup> values are provided for each subplot, with R<sup>2</sup> values for the standard Full FAs given in parentheses for comparison. *P-values* were < 0.01 except where marked with an asterisk. Note that plots of 5-methyl compounds are redundant because they exactly mirror those of their 6-methyl counterparts; only the 6-methyl FAs are shown.</li>

# 475 4.3.2 The Cyclization set and pH

476 The Cyclization set highlighted the relationship between brGDGTs and pH. As pH increased, all cyclized 477 brGDGTs were found in greater relative abundance (Fig. 9). This result reinforces previous observations that higher 478 ring number is associated with higher pH, as has been quantified by the CBT (Weijers et al., 2007), #ringstetra 479 (Sinninghe Damsté, 2016), Degree of Cyclization (DC; Baxter et al., 2019), and related indices (see Appendix). 480 However, this pH relationship has thus far been demonstrated primarily in soils; the single-compound FAs of the Cyc 481 subset already provide the strongest pH correlations ( $R^2 = 0.61$ , Fig. 9b) yet reported in lake sediments. The Cyc set 482 also reveals that compounds with structural similarities exhibit analogous responses to pH. For example, all 483 monocyclized brGDGTs show a nonlinear increase with pH (Fig. 9b, e, h, k, n), while uncyclized brGDGTs all exhibit 484 a nonlinear decrease (Fig. 9a, d, g, j, m). These trends are apparent regardless of methylation number or methylation 485 position, suggesting that ring number is broadly independent of both. This independence may imply that alkyl-chain 486 cyclization serves its biological function(s) regardless of the number and position of methylations present. It also

487 allows for the construction of independent pH calibrations for tetra-, penta-, and hexamethylated brGDGTs (Sect.488 4.4.2).

489 Though the Cyc set FAs were most strongly correlated with pH ( $R^2 \le 0.61$ ), they also exhibited robust 490 relationships with conductivity ( $R^2 \le 0.57$ ; Fig. S20). All cyclized compounds showed positive correlations with 491 conductivity, and this increase was largely independent of methylation number or position. These results indicate that

- 492 ring number is primarily dependent on pH, but is correlated with conductivity in a similar manner. The brGDGT
- 493 indices showed an analogous result; all cyclization indices (CBT and related indices, #ringstetra and related indices,
- 494 and DC; see Appendix) correlated most strongly with pH, but also exhibited weaker relationships with conductivity.
- 495 The Cyc set exhibited little to no correlation with either temperature ( $R^2 \le 0.25$ ) or DO ( $R^2 \le 0.08$ ), suggesting that
- 496 neither of these environmental variables plays an important role in controlling brGDGT cyclization.



Figure 9. Relationships between pH and brGDGT FAs calculated within the Cyc set. R<sup>2</sup> values are provided for each
 subplot, with R<sup>2</sup> values for the standard Full FAs given in parentheses for comparison. *P-values* were < 0.01 except where</li>
 marked with an asterisk.

#### 503 4.3.3 The Combined Cyclization-Isomer set strengthens both conductivity and pH trends

504 Though the strongest conductivity trends were displayed by the Isom set ( $R^2 \le 0.70$ ), correlations were also 505 present in the Cyc FAs ( $R^2 \le 0.57$ , Fig. S20). Similarly, the Cyc set contained the highest pH dependencies ( $R^2 \le$ 506 0.61), but notable relationships were visible in the Isom set as well ( $R^2 \le 0.54$ , Fig. S28). To take advantage of all of 507 these conductivity and pH relationships, we therefore used the combined Cyc-Isom set, which holds only methylation 508 number constant while allowing both cyclization number and methylation position to vary (Fig. 2). Both conductivity 509 and pH trends were strengthened in this combination set ( $R^2 \le 0.73$  and 0.62, Fig. S23 and S30, respectively), 510 especially for the uncyclized compounds. However, temperature correlations were also increased in the CI set ( $R^2 \leq$ 511 0.60, Fig. S16), a potentially convoluting influence that may not be desired.

512

# 513 4.4 Calibrations

514 For each set, combined set, and subset defined in Sect. 3, we performed linear and quadratic regressions 515 against temperature, conductivity, pH, dissolved oxygen, and lake geometry variables using SFS/SBE and 516 combinatoric fitting methods. We found temperature and conductivity to provide the strongest empirical calibrations 517 with brGDGTs, followed by pH and dissolved oxygen, and discuss our recommended calibrations below.

518

#### 519 4.4.1 Temperature Calibrations

520 For each structural set and subset defined in Section 3, we performed regressions of subset-specific brGDGT 521 FAs against five temperature variables to generate multiple global-scale calibrations. Of the temperature indices that 522 we tested, MAF provided the fit with the highest statistical significance ( $R^2 = 0.91$ ), followed closely by MST ( $R^2 =$ 523 0.90), SWI ( $R^2 = 0.89$ ), and WMT ( $R^2 = 0.88$ ). MAT provided calibrations with high statistical performance as well 524  $(R^2 \le 0.87)$ . However, these fits showed clear seasonality biases in their residuals, resulting in substantial over-525 estimations of MAT for cold sites (Fig. S40). Furthermore, they often relied heavily on low-abundance compounds as 526 fitting variables (IIc, IIb', IIIc, IIIc', IIIb, and IIIb'). We therefore do not recommend a brGDGT MAT calibration and 527 focus our discussion on the warm-season temperature indices, especially MAF. We note, however, that MAF is equal 528 to MAT in warmer climates or lower latitudes and that MAF calibrations may be interpreted as reconstructing MAT 529 in these cases.

530 Methylation number was the single most important structural variable for temperature calibrations. The Meth 531 set, which allowed only methylation number to vary, provided a MAF calibration ( $R^2 = 0.90$ , Fig. 10b) that was on 532 par with other recent global and regional lake sediment calibrations ( $R^2 = 0.85$  to 0.94; Dang et al., 2018; Russell et 533 al., 2018; Martínez-Sosa et al., 2020b). The MI, MC, and Full sets, which additionally allowed for changes in 534 cyclization number and/or methylation position, added little to the calibration performance ( $R^2 = 0.90$  to 0.91. Fig. 535 10a). Furthermore, sets which held methylation number constant - Cyc, Isom, and CI - performed markedly worse 536  $(R^2 = 0.51, 0.63, and 0.67, respectively, Fig. 10a)$ . Analogous results were found using RMSE to compare fit 537 performance (Fig. S41a). These results indicate that effectively all of the temperature dependence of the 15 commonly 538 measured brGDGTs is captured by methylation number alone.



Figure 10. a) Performance (adjusted R<sup>2</sup>) of all linear and quadratic fits for the mean air temperature of months above
freezing (MAF) and brGDGT fractional abundances (FAs) calculated within the basic (Meth, Cyc, Isom; left of dashed
line) and combined (Meth-Isom (MI), Cyc-Isom (CI), Meth-Cyc (MC), and Full sets; right of dashed line) structural sets.
Results of both the SFS/SBE and combinatoric fitting methods are plotted. The fit we suggest for general use (Meth set,
quadratic, SFS/SBE; Eq. 10) is bolded and marked with an asterisk in a) and plotted in b). "Est. MAF" is the MAF
temperature estimated using this suggested fit.

547 Within the Meth set, independent calibrations were generated for the Metha, Methb, and Methc subsets as 548 well. These subsets consist of only un-, mono-, and bicylcized compounds, respectively (Fig. 2a). As the FAs of these 549 subsets were calculated independently, we were able to test for temperature calibrations of each subset alone. The 550 Meth<sub>a</sub> subset provided the strongest of the subset MAF fits ( $R^2 = 0.88$ ). This calibration has the notable advantage of 551 employing only the 3 most abundant brGDGTs typically found in nature, Ia, IIa, and IIIa, which may allow for 552 temperatures to be reproduced with high fidelity from even organic-lean samples. Furthermore, since it uses only non-553 cyclized, 5-methyl brGDGTs, it may be less subject to influence by the environmental factors that impact cyclization 554 numbers and isomer ratios. MAF calibrations were also obtained using only the monocyclized (Meth<sub>b</sub>,  $R^2 = 0.79$ ) and 555 bicyclized (Meth<sub>c</sub>,  $R^2 = 0.74$ ) brGDGTs. These fits represent, to our knowledge, the first calibrations that make use of 556 only cyclized brGDGTs. This is a noteworthy result as it shows conclusively what can be seen by eye in Fig. 7; the 557 relationship between temperature and methylations is a broad feature of brGDGTs that is present regardless of the 558 number of rings on the carbon backbone. These un-, mono-, and bicyclized calibrations may find use in the case that 559 one or more brGDGTs are suspected to be influenced by variables other than temperature.

560 Of the brGDGT temperature indices that we tested (see Appendix), MBT'<sub>5Me</sub> performed best. This index 561 correlated better with MAF ( $R^2 = 0.89$ ) than any other warm-season variable (SWI  $R^2 = 0.84$ ; MST  $R^2 = 0.70$ ; MAT 562  $R^2 = 0.70$ ; WMT  $R^2 = 0.66$ ). The slope and intercept of the MAF/MBT'<sub>5Me</sub> calibration (MAF = -0.5 (± 0.4) + 30.4 (± 563 0.8) \* MBT'<sub>5Me</sub>) were similar to the MAT/MBT'<sub>5Me</sub> calibration presented by Russell et al. (2018) (MAT = -1.21 + 564 32.42 \* MBT'<sub>5Me</sub>). This may suggest that MBT'<sub>5Me</sub>-derived temperatures using the Russell et al. (2018) calibration in

- 565 cold regions are best considered to reconstruct MAF rather than MAT. Though MBT'6Me was previously found to
- 566 correlate well with temperature in highly alkaline lakes on a regional scale ( $R^2 = 0.75$ ; Dang et al., 2018), it was not
- 567 correlated with any temperature variable in our global dataset ( $R^2 \le 0.12$ ). Finally, we note that the Community Index

568 (De Jonge et al., 2019), which was associated with bacterial community changes in geothermally heated Icelandic

- 569 soils, is identical to our fla<sub>Meth</sub>. This may suggest that the strong connection between fla<sub>Meth</sub> and MAF could also be
- 570 driven at least in part by changes in microbial community composition. However, genomic data is not currently
- 571 available for the majority of the sites in this study to test this hypothesis.
- 572 MAF provided the strongest temperature calibrations at the global level in this study. Unlike for MAT, these 573 calibrations yielded unbiased residuals at the cold end (Fig. 10b), encouraging their application in cold regions with 574 high seasonality in particular. However, we note that the range of MAF values in our combined Canadian and Icelandic 575 dataset is small (6.6°C, or 3.2°C without the warm and cold endmembers) compared to that of MAT (22.9°C), 576 especially in relation to the MAF calibration error (RMSE =  $2.14^{\circ}$ C, Fig. 10b). While the global-scale MAF 577 calibrations presented here are valid for application at high latitudes, they may therefore prove unable to resolve 578 smaller temperature variations in paleoclimate archives from those regions. Regional high-latitude calibrations aimed 579 at improving this temperature resolution are the subject of ongoing work.
- 580

# 581 4.4.2 Conductivity and pH Calibrations

- 582 Conductivity outperformed pH in our calibrations ( $R^2 = 0.83$  versus 0.74) and was the second most important 583 predictor of brGDGT distributions in our dataset after temperature. Both cyclization number and methylation position 584 were important to the success of these calibrations. The Cyc set, which allowed only cyclization number to change, 585 provided a conductivity fit with  $R^2 = 0.73$  (Fig. 11a). The Isom set, which isolated trends in isomer abundances, 586 generated a slightly stronger calibration with  $R^2 = 0.76$ . When both of these structural properties were allowed to vary 587 together in the CI set, the calibration was markedly improved ( $R^2 = 0.83$ , Fig. 11b). In contrast, methylation number 588 was a poorer predictor of conductivity alone (Meth set  $R^2 = 0.65$ ) and did not improve upon the CI correlation in the 589 combined sets (MI  $R^2 = 0.80$ ; MC  $R^2 = 0.83$ ; Full  $R^2 = 0.81$ ). Similar results were found when using RMSE to compare 590 fit performance (Fig. S41b). The CI<sub>III</sub> and CI<sub>II</sub> subsets also provided conductivity calibrations with relatively high 591 statistical performance ( $R^2 = 0.75, 0.73$ ). The CI<sub>I</sub> subset performed worse ( $R^2 = 0.65$ ), likely due to the fact that no 592 isomer variations are present in these FAs. These subset-specific fits are the first of their kind, and their success 593 indicates that the relationship between brGDGTs and conductivity is present regardless of methylation number.
- 594 In addition to conductivity, the CI set also best captured the relationship between brGDGTs and pH ( $R^2 =$ 595 0.73; Fig. 11c and d). The addition of methylation number variations in the Full set did not substantially improve the 596 calibration ( $R^2 = 0.74$ ), indicating that the majority of the relationship between brGDGTs and pH is captured by 597 cyclization number and methylation position. Of these two structural variables, cyclization number was more 598 important for pH; the Cyc set calibration ( $R^2 = 0.67$ ) outperformed those from the Isom and Meth sets ( $R^2 = 0.59$  for 599 both). Analogous results were found using RMSE (Fig. S41c). The CI<sub>I</sub>, CI<sub>II</sub>, and CI<sub>III</sub> subsets provided weaker, but 600 significant calibrations with comparable performances to one another ( $R^2 = 0.67, 0.68$ , and 0.62, respectively), 601 indicating again that pH relationships are more or less independent of methylation number.





Figure 11. Performance (adjusted R<sup>2</sup>) of all linear and quadratic fits for lake water conductivity (a) and pH (c) and
brGDGT fractional abundances (FAs) calculated within the basic (Meth, Cyc, Isom; left of dashed lines) and combined
(Meth-Isom (MI), Cyc-Isom (CI), Meth-Cyc (MC), and Full sets; right of dashed lines) structural sets. Results of both the
SFS/SBE and combinatoric fitting methods are plotted. The fits we suggest for general use (CI set, quadratic,
combinatoric, for both variables; Eqs. 12-13) is bolded and marked with an asterisk in a) and c) and plotted in b) and d).
"Est. In(Conductivity)" and "Est. pH" are the natural logarithm of lake water conductivity and pH estimated using these

- 610 suggested fits.

## 613 4.4.3 Dissolved Oxygen and Lake Geometry Calibrations

There is increasing evidence that oxygen availability strongly affects lacustrine brGDGT distributions (Colcord et al., 2017; Weber et al., 2018; van Bree et al., 2020; Yao et al., 2020). We therefore tested for calibrations within our dataset with mean and minimum dissolved oxygen concentration (DO<sub>mean</sub> and DO<sub>min</sub>). As lake morphology can be an important predictor of lake oxygen levels (Hutchinson, 1938; Nürnberg, 1995), we also tested the natural logarithms of maximum water depth (Depth), the ratio of lake surface area to maximum depth (SA/D), and approximate lake volume. None of the DO or lake geometry variables generated strong correlations with brGDGT distributions (Table S3; section S5.1).

621 In light of increasing evidence that oxygen availability strongly affects lacustrine brGDGT distributions, it is 622 perhaps surprising that we do not find a stronger correlation between brGDGTs and DO. However, the effects of DO 623 on brGDGT distributions appear to be highly site-specific. For example, some detailed studies have found elevated 624 levels of brGDGT-IIIa in low oxygen conditions (Weber et al., 2018; Yao et al., 2020), but another found all of the 625 most common brGDGTs except IIIa in abundance in the oxygen-depleted hypolimnion (van Bree et al., 2020). A third 626 detailed study found no correlation between brGDGTs and oxygen at all (Loomis et al., 2014b), and no calibration 627 study to date has found regional or global trends. The wide range of possible drivers of DO - mixing regimes, 628 eutrophication state, and ice cover, to name a few - may play a role in the incoherent relationship between brGDGTs 629 and DO in these studies and our own. Additionally, DO measurements are often taken at the time of sampling and are 630 most likely not representative of the annual range. For the Canadian and Icelandic lakes in this study, for example, 631 DO was depleted under lake ice relative to ice-free conditions in 10 out of 11 cases. Few of the lakes in our study have 632 continuous DO monitoring data available, and most are from Central Europe. Therefore, while our study does find 633 some significant correlations between brGDGTs and DO, we do not recommend a calibration for general use and 634 instead highlight the need for further study.

635

# 636 4.4.4 Recommended Calibrations

The Meth, MC, MI, and Full subsets all provided MAF temperature calibrations with comparable R<sup>2</sup> (0.90 to 0.91) and RMSE (1.97 to 2.14°C) values (Fig. 10a, Table S3). However, the Meth set provided the FAs with the strongest compound-specific relationships with MAF (R<sup>2</sup>  $\leq$  0.88) in the modern dataset and exhibited little influence (R<sup>2</sup>  $\leq$  0.49) from any other variable examined in this study. In contrast, the MC, MI, and Full sets all exhibited stronger relationships with pH and conductivity (Fig. 11a and c). We therefore recommend the highest-performing Meth calibration (Fig. 10b; Eq. 10; n = 182, R<sup>2</sup> = 0.90, RMSE = 2.14°C) for general use in lake sediments:

$$\begin{array}{ll} 643 & MAF(^{\circ}C) = 92.9(\pm 15.98) + 63.84(\pm 15.58) \times fIb_{Meth}^2 - 130.51(\pm 30.73) \times fIb_{Meth} \\ 644 & - 28.77(\pm 5.44) \times fIIa_{Meth}^2 - 72.28(\pm 17.38) \times fIIb_{Meth}^2 - 5.88(\pm 1.36) \times fIIc_{Meth}^2 \\ 645 & + 20.89(\pm 7.69) \times fIIIa_{Meth}^2 - 40.54(\pm 5.89) \times fIIIa_{Meth} \\ 646 & - 80.47(\pm 19.19) \times fIIIb_{Meth} \end{array}$$

647 Importantly, Eq. (10) contains terms for  $fIIc_{Meth}$  and  $fIIIb_{Meth}$ , which may be equal to zero in organic-lean 648 samples (e.g. Fig. 7f and h) and could thus result in unrealistic values for MAF. In these cases, we recommend applying 649 subset-specific calibrations that exclude bicyclized (Eq. S1; n = 182, R<sup>2</sup> = 0.89, RMSE = 2.19°C) or both mono- and  $\begin{array}{ll} 650 & \mbox{bicyclized (Eq. S2; n = 182, R^2 = 0.88, RMSE = 2.33^{\circ}C) \mbox{ compounds. These calibrations offer flexibility in downcore} \\ 651 & \mbox{applications with minimal sacrifice of statistical performance. Subset-specific calibrations are also available for mono-} \\ 652 & \mbox{and bicyclized compounds alone (Eq. S3-4).} \end{array}$ 

- The Full set MAF calibration provided the highest  $R^2$  and lowest RMSE in our dataset. This fit may be applicable in settings with good conductivity or pH control and may be useful for comparison with previous calibrations. It is therefore provided in Eq. (11) (n = 182,  $R^2 = 0.91$ , RMSE = 1.97°C):
- $656 \qquad MAF (^{\circ}C) = -8.06(\pm 1.56) + 37.52(\pm 2.35) \times fIa_{Full} 266.83(\pm 98.61) \times fIb_{Full}^{2}$   $657 \qquad + 133.42(\pm 19.51) \times fIb_{Full} + 100.85(\pm 9.27) \times fIIa'_{Full}^{2} + 58.15(\pm 10.09) \times fIIIa'_{Full}^{2}$   $658 \qquad + 12.79(\pm 2.89) \times fIIIa_{Full} \qquad (11)$

The calibrations presented in Eqs. (10), (11), and (S1-4) allow for the quantitative reconstruction of warmseason air temperatures from lake sediment archives, including those at high latitudes. The statistical performance of these fits is comparable to recently published calibrations (Russell et al. (2018):  $R^2 = 0.94$ , RMSE = 2.14°C, n = 65; Martínez-Sosa et al. (2020):  $R^2 = 0.85$ , RMSE = 2.8°C, n = 261; Dang et al. (2018):  $R^2 = 0.91$ , RMSE = 1.10°C, n = 39). We do not recommend an independent MAT calibration due to residual seasonality biases (Fig. S40), though we note again that MAT is often identical to MAF in warm or low-seasonality settings.

- Independent calibrations were generated for lake water conductivity and pH. However, we re-emphasize that these two variables are fundamentally related (Section 4.3), which may cause them to vary in tandem in a paleo record. While we provide separate calibrations for conductivity and pH here, an examination of downcore trends in both reconstructed variables is recommended, along with an understanding of their relationship at the modern site.
- The CI set generated the highest-performing conductivity calibration ( $R^2 = 0.83$ , RMSE = 0.66; Fig. 11b, Table S3). The Full, MC, and MI sets provided calibrations that were statistically comparable ( $R^2 = 0.80$  to 0.83, RMSE = 0.65 to 0.70), but their FAs contained marked temperature dependencies ( $R^2 \le 0.70$ , 0.77, and 0.75, respectively; Figs S18, S17 and S15). We therefore recommend the top CI fit for use (Eq. 12; n = 143,  $R^2 = 0.83$ , RMSE = 0.66), and provide subset-specific calibrations for CI<sub>I</sub>, CI<sub>II</sub>, and CI<sub>III</sub> in the Supplement (Eq. S5-7):
- 674  $ln(Cond.) = 6.62(\pm 1.01) + 8.87(\pm 1.24) \times fIb_{CI} + 5.12(\pm 1.54) \times fIIa'_{CI}^2 + 10.64(\pm 1.88) \times fIIa_{CI}^2$
- 675  $-8.59(\pm 2.21) \times fIIa_{CI} 4.32(\pm 1.46) \times fIIIa'_{CI} 5.31(\pm 0.95) \times fIIIa_{CI}^{2}$
- 676

 $-142.67(\pm 36.08) \times fIIIb_{Cl}^2$ 

A brGDGT-based paleoconductivity reconstruction has yet to be attempted, but diatom-inferred conductivity records show the potential for this variable to provide valuable insight into changes in lake hydrology. These records have reconstructed changes in precipitation and evaporation balance, lake level fluctuations, meltwater influx events, and the isolation of a lake from the sea (Ng and King, 1999; Yang et al., 2004; Stager et al., 2013). The conductivity calibration in Eq. (12) thus enables brGDGTs to be tested as a new alternative or complementary proxy in paleohydrology reconstructions.

The Full and CI sets provided calibrations with pH that were statistically comparable to one another ( $R^2 = 0.74$  and 0.73, Fig. 11c, Table S3). Given the stronger temperature dependencies of the Full set, we recommend the CI calibration (Fig. 11d; Eq. 13; n = 154,  $R^2 = 0.73$ , RMSE = 0.57) for use and provide its subset-specific calibrations (CI<sub>1</sub>, CI<sub>11</sub>, and CI<sub>111</sub>) in the Supplement (Eq. S8-10):

(12)

$$687 pH = 8.93(\pm 0.21) - 3.84(\pm 0.25) \times f Ia_{CI}^2 + 2.63(\pm 0.35) \times f IIa_{CI}' (13)$$

overestimate pH for acidic samples and underestimate it for alkaline ones. A similar bias in pH calibrations has been

688 The residuals of the fit in Eq. (13) have a weak but significant correlation with pH ( $R^2 = 0.17$ ), causing it to 689

690 previously observed in soils (De Jonge et al., 2014a). We therefore caution the use of this calibration in acidic (pH <

691 5) or alkaline (pH > 9) conditions.

692 Previous work has demonstrated the value of brGDGT-derived pH records in studies of terrestrial 693 paleoclimate (Tyler et al., 2010; Cao et al., 2017; Fastovich et al., 2020). However, these studies relied on calibrations 694 generated from soils and/or analyses in which the 5- and 6-methyl isomers were not separated. The fit presented in 695 Eq. (13) may improve such studies by providing a globally-distributed pH calibration in lake sediments using the latest 696 chromatographic methods which improves upon the error of previously available calibrations (RMSE = 0.80, Russell 697 et al., 2018).

698 The structural set calibrations for MAF, conductivity, and pH presented above offer potential advantages 699 over traditional calibrations constructed using the standard Full set. The structural set FAs separate the effects of 700 temperature from pH and conductivity in the modern dataset. The application of calibrations such as Eqs. (10), (12), 701 and (13) could therefore provide some protection against the unwanted influences of other environmental variables 702 downcore. Importantly, this advantage is gained without sacrificing the statistical performance of the calibration. 703 Furthermore, the subset-specific calibrations such as Eqs. (S1) and (S2) use only the most abundant compounds, 704 making it easier to apply brGDGT proxies to organic-lean samples where only some compounds are typically present 705 above the detection limit.

#### 706 **5** Conclusions

707 We have shown that brGDGT structural sets and warm-season temperature indices improve compound-708 specific correlations with environmental parameters while advancing our biological understanding of the lipids 709 themselves. Grouping brGDGTs into structural sets based on methylation number, methylation position, and 710 cyclization number elucidated the relationships between environmental variables and brGDGT structures. These sets 711 revealed that methylation number fully captures the relationship between brGDGT distributions and temperature. They 712 also showed the relative abundance of 5- and 6-methyl isomers to be dependent on conductivity and cyclization 713 number to be primarily tied to pH. The deconvolved relationships provided by these subsets allowed for the generation 714 of calibrations with temperature and pH that relied on fewer compounds with robust modern trends, without sacrificing 715 statistical performance. They additionally revealed conductivity to be the second-most important variable in 716 controlling brGDGT distributions and provided a calibration for this oft-overlooked variable, which may find use as 717 a proxy for precipitation/evaporation balances or hydrologic changes.

718 The structural sets also provided insight into the biological underpinnings of brGDGT structural diversity. 719 The Meth, Cyc, and Isom sets gave evidence that methylation number, cyclization number, and methylation position 720 vary more or less independently of one another across environmental gradients. They further revealed that the 721 inclusion of tetramethylated compounds (Ia, Ib, Ic) enhances the temperature dependencies of 5-methyl compounds 722 but erases those of their 6-methyl counterparts. Finally, the results linking methylation number to temperature and 723 cyclization number to pH are consistent with a physiological explanation of brGDGT distributions (Weijers et al.,

2007). However, an explanation for the connection between conductivity and methylation position has, to our
knowledge, not yet been proposed, and encourages further study. As the microbial producers of brGDGTs have yet to
be identified and cultured, the structural set approach thus provides a valuable tool for investigating controls on
brGDGT diversity with a biological lens.

728 Warm-season temperatures outperformed MAT as the most important predictors of brGDGT distributions, 729 particularly in high-latitude environments. We introduced 43 new lake sediment samples from sites with low MAT 730 and high seasonality. In conjunction with a global dataset, these samples showed a clear warm-season bias in brGDGT 731 temperature relationships, with MAF providing the strongest fits in our dataset. The warm-season bias may suggest a 732 direct or indirect connection to heightened primary productivity in the summer. Alternatively, it may be the result of 733 a more complex relationship with dynamic lake processes such as mixing events. While further study is needed to 734 unravel these complications, the strong empirical calibrations presented here support the use of the brGDGT 735 paleotemperature proxy to quantitatively reconstruct warm-season air temperatures from high-latitude lake sediments. 736 In summary, the use of brGDGT structural sets and warm-season temperature indices deconvolved 737 relationships between brGDGT structure and environmental gradients, revealed trends with biological implications, 738 and tied brGDGT distributions to warm-season temperatures. Furthermore, they allowed for the construction of 739 temperature and pH calibrations that rely on a smaller number of compounds with clearer modern trends compared to 740 the traditional approach, without sacrificing statistical performance. Finally, we used the new structural set framework 741 to generate the first brGDGT conductivity calibration in a global lake sediment dataset. Our results thus allow 742 brGDGTs to be used to quantitatively reconstruct warm-season air temperatures and lake water conductivity and pH

from lake sediment archives and provide a new methodology for the study of brGDGTs in the future.

# 744 Appendix



<sup>746</sup> 

Figure A1. Structures of the 15 commonly measured brGDGTs along with their schematic representations, with C6

<sup>748</sup> methylations denoted in red in the schematics.

- Full and schematic structures of the 15 commonly measured brGDGTs are provided in Fig. A1. Table A1
- 750 details equations for calculating FAs within the structural sets.
- 751

(Sub)set name	(Sub)set compounds	Fractional abundance equation
Full	$S = \{Ia, Ib, Ic, IIa, IIb, IIc, IIIa, IIIb, IIIc, IIa',$	
	IIb', IIc', IIIa', IIIb', IIIc'}	
MC-5Me+	$S = {Ia, Ib, Ic, IIa, IIb, IIc, IIIa, IIIb, IIIc}$	
MC-6Me+	$S = {Ia, Ib, Ic, IIa', IIb', IIc', IIIa', IIIb', IIIc'}$	$fxy_{[(sub)set name]} = xy / \sum S$
MC-5Me	$S = {IIa, IIb, IIc, IIIa, IIIb, IIIc}$	
MC-6Me	$S = {IIa', IIb', IIc', IIIa', IIIb', IIIc'}$	
MC	Use MC-5Me+ and MC-6Me FAs	
MI	$S_a = \{Ia, IIa, IIIa, IIa', IIIa'\}$	
	$S_b = \{Ib, IIb, IIIb, IIb', IIIb'\}$	
	$S_c = \{Ic, IIc, IIIc, IIc', IIIc'\}$	
Meth-5Me+	$S_a = \{Ia, IIa, IIIa\}$	
	$S_b = \{Ib, IIb, IIIb\}$	
	$S_c = \{Ic, IIc, IIIc\}$	
Meth-6Me+	$S_a = \{Ia, IIa', IIIa'\}$	
	$S_b = \{Ib, IIb', IIIb'\}$	$f_{rry} = rry / \sum s$
	$S_c = \{Ic, IIc', IIIc'\}$	$Ix y_{[(sub)set name]} - x y / \sum S_y$
Meth-5Me	$S_a = {IIa, IIIa}$	
	$S_b = {IIb, IIIb}$	
	$S_c = {IIc, IIIc}$	
Meth-6Me	$S_a = \{IIa', IIIa'\}$	
	$S_b = {IIb', IIIb'}$	
	$S_c = \{IIc', IIIc'\}$	
Meth	<i>Use Meth-5Me+ and Meth-6Me FAs</i>	
CI	$S_{I} = \{Ia, Ib, Ic\}$	
	$S_{II} = \{IIa, IIb, IIc, IIa', IIb', IIc'\}$	
	$S_{III} = \{IIIa, IIIb, IIIc, IIIa', IIIb', IIIc'\}$	
Cyc-5Me	$S_{I} = \{Ia, Ib, Ic\}$	
	$S_{II} = {IIa, IIb, IIc}$	$frv_{res} = rv/\sum s$
	$S_{III} = \{IIIa, IIIb, IIIc\}$	$[xy]((sub)set name] = xy/ \sum 3x$
Cyc-6Me	$S_{I} = \{Ia, Ib, Ic\}$	
	$S_{II} = \{IIa', IIb', IIc'\}$	
	$S_{III} = \{IIIa', IIIb', IIIc'\}$	
Cyc	Use Cyc-5Me and Cyc-6Me FAs	
Isom	$S_{IIa} = \{IIa, IIa'\}; S_{IIIa} = \{IIIa, IIIa'\}$	$\sim$
	$S_{IIb} = {IIb, IIb'}; S_{IIIb} = {IIIb, IIIb'}$	$fxy_{[(sub)set name]} = xy / \sum S_{xy}$
	$S_{IIc} = \{IIc, IIc'\}; S_{IIIc} = \{IIIc, IIIc'\}$	

## 753 Table A1. Equations for calculating FAs within brGDGT subsets, where fxy and xy are the fractional and absolute 754 abundances of the 5- or 6-methyl brGDGT with Roman numeral x (I, II, or III) and alphabet letter y (a, b, or c).

We used the following previously-defined brGDGT indices in this study: CBT (Weijers et al., 2007); MBT' (Peterse et al., 2012); MBT'<sub>5Me</sub>, MBT'<sub>6Me</sub>, CBT<sub>5Me</sub>, CBT', and Index1 (De Jonge et al., 2014a); IR<sub>6Me</sub> (Dang et al., 2016); #rings<sub>tetra</sub>, #rings<sub>penta 5Me</sub>, and #rings<sub>penta 6Me</sub> (Sinninghe Damsté, 2016); Degree of Cyclization (DC) as

# reformulated by (Baxter et al., 2019); HP5 (Yao et al., 2020), isomerization of branched tetraethers (IBT; Ding et al.,

759 2015), community index (CI; De Jonge et al., 2019). Their equations are given below:

760 
$$CBT = -\log\left(\frac{Ib + IIb + IIb'}{Ia + IIa + IIa'}\right)$$
(A1)

761 
$$MBT' = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa + IIb + IIc + IIIa + IIa' + IIb' + IIc' + IIIa')}$$
(A2)

762 
$$MBT'_{5Me} = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa + IIb + IIc + IIIa)}$$
(A3)

763 
$$MBT'_{6Me} = \frac{(Ia + Ib + Ic)}{(Ia + Ib + Ic + IIa' + IIb' + IIc' + IIIa')}$$
(A4)

764 
$$CBT_{5Me} = -\log\left(\frac{lb + llb}{la + lla}\right) \tag{A5}$$

765 
$$CBT' = -\log\left(\frac{Ic + IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc'}{Ia + IIa + IIIa}\right)$$
(A6)

766 
$$Index1 = log\left(\frac{Ia + Ib + Ic + IIa' + IIIa'}{Ic + IIa + IIc + IIIa + IIIa'}\right)$$
(A7)

767 
$$IR_{6Me} = \frac{(IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc')}{(IIa' + IIb' + IIc' + IIIa' + IIIb' + IIIc' + IIa + IIb + IIc + IIIa + IIIb + IIIc)}$$
(A8)

$$\#rings_{tetra} = \frac{(Ib+2 * Ic)}{(Ia+Ib+Ic)}$$
(A9)

769 
$$\#rings_{penta\ 5Me} = \frac{(IIb + 2 * IIc)}{(IIa + IIb + IIc)}$$
(A10)

770 
$$\#rings_{penta\ 6Me} = \frac{(IIb' + 2 * IIc')}{(IIa' + IIb' + IIc')}$$
(A11)

771 
$$DC = \frac{(Ib + 2 * Ic + IIb + IIb')}{(Ia + Ib + Ic + IIa + IIa' + IIb + IIb')}$$
(A12)

$$HP5 = \frac{IIIa}{(IIa + IIIa)}$$
(A13)

773 
$$IBT = -log\left(\frac{IIa' + IIIa'}{IIa + IIIa}\right)$$
(A14)

774 
$$CI = \frac{Ia}{(Ia + IIa + IIIa)} (= fIa_{Meth})$$
(A15)

775

768

776

# 777 Code and data availability

Biomarker and associated metadata is provided in the Supplement and will be archived at the PANGAEA data
 repository. Code for generating and plotting set-specific FAs in R is provided in the Supplement. Additional code,

- 780 data, and calibration equations will be available upon request.
- 781

## 782 Author contribution

- 783 GHM, JS, ÁG, JHR, SEC, DJH, and GdW designed the study and carried out the sampling. GHM, JS, ÁG, and SEC
- funded the research. JHR, AB, and DJH performed the laboratory work and processed the HPLC-MS data under the
- supervision of JS. JHR and SK generated the R code. JHR analyzed the data and interpreted the results. JHR wrote
- the manuscript with input from all authors. All authors contributed to the article and approved the submitted version.
- 787

#### 788 Competing interests

- 789 The authors declare that they have no conflict of interest.
- 790

#### 791 Acknowledgements

792 This work was supported by the National Science Foundation (OPP-1737712 to GHM and JS; OPP-1836981 to GHM, 793 ÁG, and JS; DDRI-1657743 to GHM and SEC; SK acknowledges support from NSF EAR-1945484), a Doctoral Grant 794 from the University of Iceland and a project grant from the University of Iceland Research Fund to AG, a National 795 Geographic Society Early Career Grant (#CP-019ER-17 to SEC), and the University of Colorado Boulder. We thank 796 the Inuit of Nunavut for permitting access to their land and to sample soils and lake sediment (Scientific Research 797 Licenses 01022 17R-M, 02034 18R-M, 02038 19R-M) and the Qikiqtaani Inuit of Qikiqtarjuaq and Clyde River for 798 assistance in the field. We thank the Nunavut Research Institute for logistical assistance and Polar Continental Shelf 799 Project for air support. Field research in Iceland benefited from the assistance of Ian Holmen, Sveinbjörn 800 Steinthorsson, and Thor Blöndahl. We thank Yuki Weber and Shucheng Xie for providing additional temperature 801 data, James Russell for a valuable discussion, and Lina Pérez-Angel for assistance with the calibration code. We 802 additionally thank Nadia Dildar, Sebastian Cantarero, and Katie Rempfert for laboratory assistance and Chris Florian 803 and Kate Zalzal for providing additional samples. The manuscript benefited from field assistance and/or discussions 804 with Martha Raynolds, Shawnee Gowan, Helga Bueltmann, Elizabeth Thomas, Devon Gorby, Kayla Hollister, Kurt 805 Lindberg, Nicolò Ardenghi, and Jamie McFarlin.

806

#### 807 Literature Cited

- Baxter, A.J., Hopmans, E.C., Russell, J.M., Sinninghe Damsté, J.S., 2019. Bacterial GMGTs in East African lake
   sediments: Their potential as palaeotemperature indicators. Geochimica et Cosmochimica Acta 259, 155–169.
- Beales, N., 2004. Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and
   Osmotic Stress: A Review. Comprehensive Reviews in Food Science and Food Safety 3, 1–20.
- Buckles, L.K., Weijers, J.W.H., Verschuren, D., Sinninghe Damsté, J.S., 2014. Sources of core and intact branched
  tetraether membrane lipids in the lacustrine environment: Anatomy of Lake Challa and its catchment,
  equatorial East Africa. Geochimica et Cosmochimica Acta 140, 106–126.
- 815 Cao, J., Rao, Z., Jia, G., Xu, Q., Chen, F., 2017. A 15 ka pH record from an alpine lake in north China derived from
- 816 the cyclization ratio index of aquatic brGDGTs and its paleoclimatic significance. Organic Geochemistry 109,

817 31-46.

- 818 Cao, J., Rao, Z., Shi, F., Jia, G., 2020. Ice formation on lake surfaces in winter causes warm-season bias of
- 819 lacustrine brGDGT temperature estimates. Biogeosciences 17, 2521–2536.
- Colcord, D.E., Pearson, A., Brassell, S.C., 2017. Carbon isotopic composition of intact branched GDGT core lipids
  in Greenland lake sediments and soils. Organic Geochemistry 110, 25–32.
- Crump, S.E., Miller, G.H., Power, M., Sepúlveda, J., Dildar, N., Coghlan, M., Bunce, M., 2019. Arctic shrub
  colonization lagged peak postglacial warmth: Molecular evidence in lake sediment from Arctic Canada.
  Global Change Biology 1–13.
- Bang, X., Ding, W., Yang, H., Pancost, R.D., Naafs, B.D.A., Xue, J., Lin, X., Lu, J., Xie, S., 2018. Different
  temperature dependence of the bacterial brGDGT isomers in 35 Chinese lake sediments compared to that in
  soils. Organic Geochemistry 119, 72–79.
- Bang, X.Y., Xue, J.T., Yang, H., Xie, S.C., 2016. Environmental impacts on the distribution of microbial tetraether
   lipids in Chinese lakes with contrasting pH: Implications for lacustrine paleoenvironmental reconstructions.
   Science China Earth Sciences.
- Be Jonge, C., Hopmans, E.C., Stadnitskaia, A., Rijpstra, W.I.C., Hofland, R., Tegelaar, E., Sinninghe Damsté, J.S.,
  2013. Identification of novel penta- and hexamethylated branched glycerol dialkyl glycerol tetraethers in peat
  using HPLC-MS2, GC-MS and GC-SMB-MS. Organic Geochemistry 54, 78–82.
- Be Jonge, C., Hopmans, E.C., Zell, C.I., Kim, J.H., Schouten, S., Sinninghe Damsté, J.S., 2014a. Occurrence and
   abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate
   reconstruction. Geochimica et Cosmochimica Acta 141, 97–112.
- Barton Bernet State
   Barton Bernet State<
- Be Jonge, C., Stadnitskaia, A., Hopmans, E.C., Cherkashov, G., Fedotov, A., Sinninghe Damsté, J.S., 2014b. In situ
  produced branched glycerol dialkyl glycerol tetraethers in suspended particulate matter from the Yenisei
- River, Eastern Siberia. Geochimica et Cosmochimica Acta 125, 476–491.
- de Wet, G.A., Castañeda, I.S., DeConto, R.M., Brigham-Grette, J., 2016. A high-resolution mid-Pleistocene
  temperature record from Arctic Lake El'gygytgyn: A 50 kyr super interglacial from MIS 33 to MIS 31? Earth
  and Planetary Science Letters 436, 56–63.
- Bearing Crampton-Flood, E., Peterse, F., Munsterman, D., Sinninghe Damsté, J.S., 2018. Using tetraether lipids
  archived in North Sea Basin sediments to extract North Western European Pliocene continental air
  temperatures. Earth and Planetary Science Letters 490, 193–205.
- Dillon, J.T., Lash, S., Zhao, J., Smith, K.P., van Dommelen, P., Scherer, A.K., Huang, Y., 2018. Bacterial tetraether
  lipids in ancient bones record past climate conditions at the time of disposal. Journal of Archaeological
  Science 96, 45–56.
- Bing, S., Kohlhepp, B., Trumbore, S., Küsel, K., Totsche, K.U., Pohnert, G., Gleixner, G., Schwab, V.F., 2018. In
  situ production of core and intact bacterial and archaeal tetraether lipids in groundwater. Organic

854 Geochemistry 126, 1–12.

- Ding, S., Schwab, V.F., Ueberschaar, N., Roth, V.-N., Lange, M., Xu, Y., Gleixner, G., Pohnert, G., 2016.
  Identification of novel 7-methyl and cyclopentanyl branched glycerol dialkyl glycerol tetraethers in lake
  sediments. Organic Geochemistry 102, 52–58.
- Ding, S., Xu, Y., Wang, Y., He, Y., Hou, J., Chen, L., He, J.S., 2015. Distribution of branched glycerol dialkyl
  glycerol tetraethers in surface soils of the Qinghai-Tibetan Plateau: Implications of brGDGTs-based proxies in
  cold and dry regions. Biogeosciences 12, 3141–3151.
- Fastovich, D., Russell, J.M., Jackson, S.T., Williams, J.W., 2020. Deglacial temperature controls on no-analog
   community establishment in the Great Lakes Region. Quaternary Science Reviews 234, 106245.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas.
  International Journal of Climatology 37, 4302–4315.
- 865 Foster, L.C., Pearson, E.J., Juggins, S., Hodgson, D.A., Saunders, K.M., Verleyen, E., Roberts, S.J., 2016.
- Bevelopment of a regional glycerol dialkyl glycerol tetraether (GDGT)-temperature calibration for Antarctic
  and sub-Antarctic lakes. Earth and Planetary Science Letters 433, 370–379.
- Guo, J., Glendell, M., Meersmans, J., Kirkels, F., Middelburg, J.J., Peterse, F., 2020. Assessing branched tetraether
  lipids as tracers of soil organic carbon transport through the Carminowe Creek catchment (southwest
  England). Biogeosciences 17, 3183–3201.
- Hanna, A.J.M., Shanahan, T.M., Allison, M.A., 2016. Distribution of branched GDGTs in surface sediments from
  the Colville River, Alaska: Implications for the MBT'/CBT paleothermometer in Arctic marine sediments.
  Journal of Geophysical Research: Biogeosciences 121, 1762–1780.
- Harning, D.J., Andrews, J.T., Belt, S.T., Cabedo-Sanz, P., Geirsdóttir, Á., Dildar, N., Miller, G.H., Sepúlveda, J.,
  2019. Sea Ice Control on Winter Subsurface Temperatures of the North Iceland Shelf During the Little Ice
  Age: A TEX86 Calibration Case Study. Paleoceanography and Paleoclimatology 34, 1006–1021.
- 877 Harning, D.J., Curtin, L., Geirsdóttir, Á., D'Andrea, W.J., Miller, G.H., Sepúlveda, J., 2020. Lipid Biomarkers
- 878 Quantify Holocene Summer Temperature and Ice Cap Sensitivity in Icelandic Lakes. Geophysical Research
  879 Letters 47, 1–11.
- Hopmans, E.C., Schouten, S., Sinninghe, J.S., 2016. The effect of improved chromatography on GDGT-based
  palaeoproxies. Organic Geochemistry 93, 1–6.
- Hopmans, E.C., Weijers, J.W., Schefuß, E., Herfort, L., Sinninghe Damsté, J.S., Schouten, S., 2004. A novel proxy
  for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. Earth and
  Planetary Science Letters 224, 107–116.
- Hu, J., Zhou, H., Peng, P., Spiro, B., 2016. Seasonal variability in concentrations and fluxes of glycerol dialkyl
  glycerol tetraethers in Huguangyan Maar Lake, SE China: Implications for the applicability of the MBT-CBT
  paleotemperature proxy in lacustrine settings. Chemical Geology 420, 200–212.
- Huguet, C., Hopmans, E.C., Febo-Ayala, W., Thompson, D.H., Sinninghe Damsté, J.S., Schouten, S., 2006. An
  improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids.
  Organic Geochemistry 37, 1036–1041.

- Hutchinson, G.E., 1938. On the Relation between the Oxygen Deficit and the productivity and Typology of Lakes.
  Internationale Revue der gesamten Hydrobiologie und Hydrographie 36, 336–355.
- Landrum, L., Holland, M.M., 2020. Extremes become routine in an emerging new Arctic. Nature Climate Change
   10. doi:10.1038/s41558-020-0892-z
- Li, F., Zhang, C.L., Wang, S., Chen, Y., Sun, C., Dong, H., Li, W., Klotz, M.G., Hedlund, B.P., 2014. Production of
  branched tetraether lipids in Tibetan hot springs: A possible linkage to nitrite reduction by thermotolerant or
  thermophilic bacteria? Chemical Geology 386, 209–217.
- Loomis, S.E., Russell, J.M., Eggermont, H., Verschuren, D., Sinninghe Damsté, J.S., 2014a. Effects of temperature,
   pH and nutrient concentration on branched GDGT distributions in East African lakes: Implications for
   paleoenvironmental reconstruction. Organic Geochemistry 66, 25–37.
- Loomis, S.E., Russell, J.M., Heureux, A.M., D'Andrea, W.J., Sinninghe Damsté, J.S., 2014b. Seasonal variability of
   branched glycerol dialkyl glycerol tetraethers (brGDGTs) in a temperate lake system. Geochimica et
   Cosmochimica Acta 144, 173–187.
- Loomis, S.E., Russell, J.M., Ladd, B., Street-Perrott, F.A., Sinninghe Damsté, J.S., 2012. Calibration and application
   of the branched GDGT temperature proxy on East African lake sediments. Earth and Planetary Science Letters
   357–358, 277–288.
- Lu, H., Liu, W., Yang, H., Wang, H., Liu, Z., Leng, Q., Sun, Y., Zhou, W., An, Z., 2019. 800-kyr land temperature
   variations modulated by vegetation changes on Chinese Loess Plateau. Nature Communications 10.
   doi:10.1038/s41467-019-09978-1
- 910 Lumley, T., 2020. leaps: Regression Subset Selection.
- Martínez-Sosa, P., Tierney, J.E., Stefanescu, I.C., Dearing Crampton-Flood, E., Shuman, B.N., Routson, C., 2020. A
   global Bayesian temperature calibration for lacustrine brGDGTs. EarthArXiv [preprint].
- 913 doi:https://doi.org/10.31223/X5PS3P
- Miller, D.R., Habicht, M.H., Keisling, B.A., Castañeda, I.S., Bradley, R.S., 2018. A 900-year New England
   temperature reconstruction from in situ seasonally produced branched glycerol dialkyl glycerol tetraethers
   (brGDGTs). Clim. Past 14, 1653–1667.
- Miller, G.H., Alley, R.B., Brigham-Grette, J., Fitzpatrick, J.J., Polyak, L., Serreze, M.C., White, J.W.C., 2010.
  Arctic amplification: Can the past constrain the future? Quaternary Science Reviews 29, 1779–1790.
- Naafs, B.D.A., Gallego-Sala, A. V., Inglis, G.N., Pancost, R.D., 2017. Refining the global branched glycerol dialkyl
   glycerol tetraether (brGDGT) soil temperature calibration. Organic Geochemistry 106, 48–56.
- Naeher, S., Peterse, F., Smittenberg, R.H., Niemann, H., Zigah, P.K., Schubert, C.J., 2014. Sources of glycerol
  dialkyl glycerol tetraethers (GDGTs) in catchment soils, water column and sediments of Lake Rotsee
  (Switzerland) Implications for the application of GDGT-based proxies for lakes. Organic Geochemistry 66,
  164–173.
- Ng, S.L., King, R.H., 1999. Development of a diatom-based specific conductivity model for the glacioisostatic lakes
   of Truelove Lowland: Implications for paleoconductivity and paleoenvironmental reconstructions in Devon
   Island lakes, N.W.T., Canada. Journal of Paleolimnology 22, 367–382.

- 928 Ning, D., Zhang, E., Shulmeister, J., Chang, J., Sun, W., Ni, Z., 2019. Holocene mean annual air temperature
- 929 (MAAT) reconstruction based on branched glycerol dialkyl glycerol tetraethers from Lake Ximenglongtan, 930 southwestern China. Organic Geochemistry 133, 65-76.
- 931 Nürnberg, G.K., 1995. Quantifying anoxia in lakes. Limnology and Oceanography 40, 1100–1111.
- 932 Pearson, E.J., Juggins, S., Talbot, H.M., Weckström, J., Rosén, P., Ryves, D.B., Roberts, S.J., Schmidt, R., 2011. A
- 933 lacustrine GDGT-temperature calibration from the Scandinavian Arctic to Antarctic: Renewed potential for
- 934 the application of GDGT-paleothermometry in lakes. Geochimica et Cosmochimica Acta 75, 6225–6238.
- 935 Pérez-Angel, L.C., Sepúlveda, J., Molnar, P., Montes, C., Rajagopalan, B., Snell, K., Gonzalez-Arango, C., Dildar,
- 936 N., 2020. Soil and Air Temperature Calibrations Using Branched GDGTs for the Tropical Andes of Colombia: 937 Toward a Pan-Tropical Calibration. Geochemistry, Geophysics, Geosystems 21, 1–18.
- 938 Peterse, F., Meer, J. Van Der, Schouten, S., Weijers, J.W.H., Fierer, N., Jackson, R.B., Kim, J., Damste, J.S.S.,
- 939 2012. Revised calibration of the MBT – CBT paleotemperature proxy based on branched tetraether membrane 940 lipids in surface soils 96, 215-229.
- 941 Peterse, F., Vonk, J.E., Holmes, R.M., Giosan, L., Zimov, N., Eglinton, T.I., 2014. Branched glycerol dialkyl
- 942 glycerol tetraethers in Arctic lake sediments: Sources and implications for paleothermometry at high latitudes. 943 Journal of Geophysical Research: Biogeosciences 119, 1738–1754.
- 944 Prahl, F.G., Wakeham, S.G., 1987. Calibration of unsaturation patterns in long-chain ketone compositions for 945 palaeotemperature assessment. Nature 330, 367–369.
- 946 Qian, S., Yang, H., Dong, C., Wang, Y., Wu, J., Pei, H., Dang, X., Lu, J., Zhao, S., Xie, S., 2019. Rapid response of 947 fossil tetraether lipids in lake sediments to seasonal environmental variables in a shallow lake in central China: 948 Implications for the use of tetraether-based proxies. Organic Geochemistry 128, 108-121.
- 949 R Core Team, 2018. R: A language and environment for statistical computing.
- 950 Raynolds, M.K., Comiso, J.C., Walker, D.A., Verbyla, D., 2008. Relationship between satellite-derived land surface 951 temperatures, arctic vegetation types, and NDVI. Remote Sensing of Environment 112, 1884–1894.
- 952 Reizer, J., Grossowicz, N., Barenholz, Y., 1985. The effect of growth temperature on the thermotropic behavior of 953 the membranes of a thermophilic Bacillus. Composition-structure-function relationships. BBA -
- 954 Biomembranes 815, 268-280.
- 955 Russell, J.M., Hopmans, E.C., Loomis, S.E., Liang, J., Sinninghe Damsté, J.S., 2018. Distributions of 5- and 6-956 methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of 957
- temperature, pH, and new lacustrine paleotemperature calibrations. Organic Geochemistry 117, 56-69.
- 958 Schouten, S., Hopmans, E.C., Schefuß, E., Sinninghe Damsté, J.S., 2002. Distributional variations in marine 959 crenarchaeotal membrane lipids: A new tool for reconstructing ancient sea water temperatures? Earth and 960 Planetary Science Letters 204, 265–274.
- 961 Schouten, S., Hopmans, E.C., Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol 962 tetraether lipids: A review. Organic Geochemistry 54, 19-61.
- 963 Schwarz, G., 1978. Estimating the Dimension of a Model. The Annals of Statistics 6, 461–464.
- 964 Shanahan, T.M., Hughen, K.A., Van Mooy, B.A.S., 2013. Temperature sensitivity of branched and isoprenoid

- 965 GDGTs in Arctic lakes. Organic Geochemistry 64, 119–128.
- 966 Sinninghe Damsté, J.S., 2016. Spatial heterogeneity of sources of branched tetraethers in shelf systems: The
- geochemistry of tetraethers in the Berau River delta (Kalimantan, Indonesia). Geochimica et Cosmochimica
  Acta 186, 13–31.
- Sinninghe Damsté, J.S., Hopmans, E.C., Pancost, R.D., Schouten, S., Geenevasen, J.A.J., 2000. Newly discovered
   non-isoprenoid glycerol dialkyl glycerol tetraether lipids in sediments. Chemical Communications 0, 1683–
   1684.
- 972 Sinninghe Damsté, J.S., Rijpstra, W.I.C., Foesel, B.U., Huber, K.J., Overmann, J., Nakagawa, S., Kim, J.J.,
- Dunfield, P.F., Dedysh, S.N., Villanueva, L., 2018. An overview of the occurrence of ether- and ester-linked
  iso-diabolic acid membrane lipids in microbial cultures of the Acidobacteria: Implications for brGDGT
  paleoproxies for temperature and pH. Organic Geochemistry 124, 63–76.
- Stager, J.C., Ryves, D.B., King, C., Madson, J., Hazzard, M., Neumann, F.H., Maud, R., 2013. Late Holocene
   precipitation variability in the summer rainfall region of South Africa. Quaternary Science Reviews 67, 105–
   120.
- Sun, Q., Chu, G., Liu, M., Xie, M., Li, S., Ling, Y., Wang, X., Shi, L., Jia, G., Lü, H., 2011. Distributions and
  temperature dependence of branched glycerol dialkyl glycerol tetraethers in recent lacustrine sediments from
  China and Nepal. Journal of Geophysical Research: Biogeosciences 116, G01008.
- Super, J.R., Chin, K., Pagani, M., Li, H., Tabor, C., Harwood, D.M., Hull, P.M., 2018. Late Cretaceous climate in
   the Canadian Arctic: Multi-proxy constraints from Devon Island. Palaeogeography, Palaeoclimatology,
   Palaeoecology 504, 1–22.
- Thomas, E.K., Castañeda, I.S., McKay, N.P., Briner, J.P., Salacup, J.M., Nguyen, K.Q., Schweinsberg, A.D., 2018.
  A Wetter Arctic Coincident With Hemispheric Warming 8,000 Years Ago. Geophysical Research Letters 45, 10,637-10,647.
- Tierney, J.E.E., Russell, J.M.M., Eggermont, H., Hopmans, E.C.C., Verschuren, D., Sinninghe Damsté, J.S.S., 2010.
   Environmental controls on branched tetraether lipid distributions in tropical East African lake sediments.
   Geochimica et Cosmochimica Acta 74, 4902–4918.
- Tyler, J.J., Nederbragt, A.J., Jones, V.J., Thurow, J.W., 2010. Assessing past temperature and soil pH estimates from
   bacterial tetraether membrane lipids: Evidence from the recent lake sediments of Lochnagar, Scotland. Journal
   of Geophysical Research 115, G01015.
- van Bree, L., Peterse, F., Baxter, A., De Crop, W., van Grinsven, S., Villanueva, L., Verschuren, D., Sinninghe
   Damsté, J., 2020. Seasonal variability and sources of in situ brGDGT production in a permanently stratified
   African crater lake. Biogeosciences Discussions 1–36.
- 997 Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth. ed. Springer, New York.
- 998 Weber, Y., Damsté, J.S.S., Zopfi, J., De Jonge, C., Gilli, A., Schubert, C.J., Lepori, F., Lehmann, M.F., Niemann,
- 999 H., 2018. Redox-dependent niche differentiation provides evidence for multiple bacterial sources of glycerol
- 1000 tetraether lipids in lakes. Proceedings of the National Academy of Sciences of the United States of America
- 1001 115, 10926–10931.

- 1002 Weber, Y., De Jonge, C., Rijpstra, W.I.C., Hopmans, E.C., Stadnitskaia, A., Schubert, C.J., Lehmann, M.F.,
- Sinninghe Damsté, J.S., Niemann, H., 2015. Identification and carbon isotope composition of a novel
   branched GDGT isomer in lake sediments: Evidence for lacustrine branched GDGT production. Geochimica
   et Cosmochimica Acta 154, 118–129.
- Weijers, J.W.H., Schouten, S., Spaargaren, O.C., Sinninghe Damsté, J.S., 2006. Occurrence and distribution of
   tetraether membrane lipids in soils: Implications for the use of the TEX86 proxy and the BIT index. Organic
   Geochemistry 37, 1680–1693.
- Weijers, J.W.H., Schouten, S., van den Donker, J.C., Hopmans, E.C., Sinninghe Damsté, J.S., 2007. Environmental
  controls on bacterial tetraether membrane lipid distribution in soils. Geochimica et Cosmochimica Acta 71,
  703–713.
- 1012 Wetzel, R.G., 2001. Salinity of Inland Waters, in: Wetzel, R.G. (Ed.), Limnology. Academic Press, pp. 169–186.
- Woltering, M., Werne, J.P., Kish, J.L., Hicks, R., Sinninghe Damsté, J.S., Schouten, S., 2012. Vertical and temporal
  variability in concentration and distribution of thaumarchaeotal tetraether lipids in Lake Superior and the
  implications for the application of the TEX 86 temperature proxy. Geochimica et Cosmochimica Acta 87,
  136–153.
- Wörmer, L., Lipp, J.S., Schröder, J.M., Hinrichs, K.U., 2013. Application of two new LC-ESI-MS methods for
   improved detection of intact polar lipids (IPLs) in environmental samples. Organic Geochemistry 59, 10–21.
- Xiao, W., Wang, Y., Liu, Y., Zhang, X., Shi, L., Xu, Y., 2020. Predominance of hexamethylated 6-methyl branched
   glycerol dialkyl glycerol tetraethers in the Mariana Trench: Source and environmental implication.
   Biogeosciences 17, 2135–2148.
- Yang, X., Wang, S., Kamenik, C., Schmidt, R., Shen, J., Zhu, L., Li, S., 2004. Diatom assemblages and quantitative
  reconstruction for paleosalinity from a sediment core of Chencuo Lake, southern Tibet. Science in China,
  Series D: Earth Sciences 47, 522–528.
- Yao, Y., Zhao, J., Vachula, R.S., Werne, J.P., Wu, J., Song, X., Huang, Y., 2020. Correlation between the ratio of 5 methyl hexamethylated to pentamethylated branched GDGTs (HP5) and water depth reflects redox variations
   in stratified lakes. Organic Geochemistry 147, 104076.
- Yuk, H.G., Marshall, D.L., 2004. Adaptation of Escherichia coli O157:H7 to pH alters membrane lipid composition,
   verotoxin secretion, and resistance to simulated gastric fluid acid. Applied and Environmental Microbiology
   70, 3500–3505.
- Zhang, Z.X., Li, J., Chen, Z., Sun, Z., Yang, H., Fu, M., Peng, X., 2020. The effect of methane seeps on the bacterial
  tetraether lipid distributions at the Okinawa Trough. Marine Chemistry 225, 103845.
- Zhao, B., Castañeda, I.S., Bradley, R.S., Salacup, J.M., de Wet, G.A., Daniels, W.C., Schneider, T., 2021.
  Development of an in situ branched GDGT calibration in Lake 578, southern Greenland. Organic
  Geochemistry 152, 104168.
- Zhao, J., Huang, Y., Yao, Y., An, Z., Zhu, Y., Lu, H., Wang, Z., 2020. Calibrating branched GDGTs in bones to
   temperature and precipitation: Application to Alaska chronological sequences. Quaternary Science Reviews
   240, 106371.

- 1039 Zheng, Y., Pancost, R.D., Liu, X., Wang, Z., Naafs, B.D.A., Xie, X., Liu, Z., Yu, X., Yang, H., 2017. Atmospheric
- 1040 connections with the North Atlantic enhanced the deglacial warming in northeast China. Geology 45, 1031–
- 1041 1034.
- 1042