Lipids at high latitudes: investigation of sources, environmental controls, and new potential applications of brGDGT-based paleoclimate proxies

Jonathan Henrik Råberg

Faculty of Earth Sciences
University of Iceland
2022

Lipids at high latitudes: investigation of sources, environmental controls, and new potential applications of brGDGT-based paleoclimate proxies

Jonathan Henrik Råberg

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Geological Sciences

Thesis Committee

Dr. Áslaug Geirsdóttir
Dr. Gifford H. Miller
Dr. Julio Sepúlveda
Dr. Sebastian Kopf
Dr. Thomas M. Marchitto
Dr. Noah Fierer

Faculty of Earth Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, February 2022

Lipids at high latitudes: investigation of sources, environmental controls, and new potential applications of brGDGT-based paleoclimate proxies
BrGDGT paleoclimate proxies at high latitudes
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Geological Sciences

Faculty of Earth Sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugata 7
101, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Raberg, J. H., 2022, *Lipids at high latitudes: investigation of sources, environmental controls, and new potential applications of brGDGT-based paleoclimate proxies*, PhD dissertation, Faculty of Earth Sciences, University of Iceland, 133 pp.

Author ORCID: 0000-0003-2959-3631

Printing: Háskolaprent
Reykjavik, Iceland, April 2022

Abstract

As high latitude regions continue a decades-long trend of warming at roughly twice the rate of the global average, an understanding of their climatic histories becomes increasing important for predicting their future. Organic molecular proxies preserved in lake sediment archives offer one avenue for reconstructing key elements of such past climates, including their temperature, precipitation, and vegetation regimes. In particular, a class of bacterial membrane-spanning lipids called branched glycerol dialkyl glycerol tetraethers (brGDGTs) form the basis for a paleothermometer that can be applied to reconstruct temperatures as far back as the Cretaceous in sedimentary archives across the globe.

Despite these successes, challenges remain that complicate the development and application of brGDGT-based proxies. First, while they correlate best with temperature and pH, other environmental parameters can influence brGDGT distributions, including seasonality, conductivity, and oxygen availability. Second, it is unknown whether these empirical correlations are the result of a direct physiological response of brGDGT-producing organisms to their environment or an indirect effect resulting from variations in bacterial community composition. Finally, an incomplete understanding of where brGDGTs are produced on the landscape and how they contribute to the sedimentary record hinders our ability to interpret proxies in mixed-source archives.

Herein, I present research addressing each of these three challenges with an emphasis on the Eastern Canadian Arctic and Iceland. First, I develop a technique for grouping brGDGTs based on structural characteristics and show that it can be used to deconvolve the effects of temperature and pH/conductivity. I further find a warm-season bias in brGDGT-derived temperatures and develop calibration equations for temperature and conductivity. Next, I compile >2500 samples from a dozen sample types across the globe and find near-universal trends in the relationships between brGDGTs and temperature, pH, and one another. These commonalities support a physiological basis for observed environmental trends. Finally, by measuring brGDGTs in their intact, polar form, I find that lipid sources in lake catchments can be distinguished and suggest novel applications down core. By advancing our understanding of brGDGTs, my results further our ability to reconstruct key climatic variables from sedimentary archives, especially at high latitudes.

Útdráttur

Lípíð á norðlægum breiddargráðum: rannsókn á uppruna og umhverfisaðstæðum og mögulegri beitingu brGDGT sem nýs veðurvísa í fornloftslagsrannsóknum.

Norðurslóðir hafa á undanförnum áratugum hlýnað um það bil tvöfalt meira en meðaltalið er að jafnaði á heimsvísu, sem gerir rannsóknir og skilning á veðurfarssögu norðurslóða sífellt mikilvægari til að spá fyrir um framtíð þeirra. Lífrænar sameindir sem varðveittar eru í seti stöðuvatna bjóða upp á leið til að endurbyggja eða rekja lykilþætti fornloftslags, þar á meðal hitastig, úrkomu og gróðurfar. Flokkur bakteríumyndandi lípíða, svokölluð „branched glýseról díalkýl glýseról tetraethers“ (brGDGTs) mynda grunn að hitamæli sem hægt er að nota til að endurbyggja hitastig allt aftur til Krítar tímabilsins í setlögum um allan heim.

Þrátt fyrir góðan árangur í rannsóknum á hitanæmi brGDGT skortir enn á almennan skilning á þróun og því notkun á brGDGT sem lífvísi og hitamæli. Í fyrsta lagi, þó að þessi bakteríumyndandi lípíð tengist best hitastigi og pH, geta aðrar umhverfisbreytur haft áhrif á dreifingu þeirra, þar á meðal árstíðarsveifla, leiðni og súrefnisframboð. Í öðru lagi er ekki vitað hvort fylgni brGDGT og þessara umhverfisþátta eru afleiðing af beinni lífeðlisfræðilegri svörun brGDGT-framleiðandi lífvera við umhverfi sitt eða óbein áhrif sem stafa af breytileika í samsetningu bakteríusamfélagsins. Einnig skortir fullnægjandi skilning á því hvar í umhverfinu brGDGTs er framleitt og hver þáttur þess er í upphleðslu setlaga, til að unnt sé að túlka það sem öruggan lífvísi/hitamæli í setlögum sem hlaðist hafa upp við breytileg skilyrði og setmyndunarferli.

Í þessari ritgerð er gerð grein fyrir rannsóknum á hverri þessara þriggja áskorana með áherslu á brGDGT í stöðuvötnum og stöðuvatnaseti á Íslandi og norðausturhluta Kanada. Niðurstöður sýna að unnt er að þróa tækni til að flokka brGDGTs út frá byggingareiginleikum og þannig sýna fram á að hægt er að nota þá flokkun til að skilja á milli áhrifa hitastigs og pH/leiðni. Í öðru lagi er sýnt fram á fylgni brGDGT við hitastig hlýrri árstíða (warm-seasons) og kvörðunarjöfnur þróaðar fyrir hitastig og leiðni. Í þriðja lagi eru >2500 sýni af tylft sýnategunda um allan heim tekin saman og sýnt fram á nær alhliða tengsl brGDGTs og hitastigs, pH og hvors um sig. Þessir sameiginlegu eiginleikar styðja lífeðlisfræðilegar ástæður fyrir umhverfisþróun bakteríanna. Að lokum, með því að mæla brGDGTs í ósnortnu formi, er sýnt fram á að hægt er að greina lípíð uppruna í vatnasviðum, sem styrkir notkun þeirra sem lífvísa/hitamæla aftur í tíma. Heildarniðurstöður sýna því fram á, að með auknum skilningi á uppruna og þróun brGDGTs í setlögum, er unnt að draga fram beinar upplýsingar um helstu loftslagsbreytur aftur í tíma, sérstaklega þegar litið er til norðurslóða.

Table of Contents

List of Figures

List of Tables

Acknowledgements

It is with sincere gratitude that I would like to thank all of those who helped and supported me in my graduate studies. First and foremost, my deepest thanks to my advisors, Gifford Miller, Julio Sepúlveda, and Áslaug Geirsdóttir. It has been an absolute joy to work together and to learn to see Baffin Island, Iceland, the world of lipids, and the earth system at large through your eyes. Thank you sincerely for taking a chance on this outsider. Thank you also to my committee members, Sebastian Kopf, Thomas Marchitto, and Noah Fierer. I am deeply appreciative of the sincerity with which you dedicated time and thought to my research meanderings. The ways in which a word here, a thought there, or a nudge in a certain direction have shaped my PhD journey is not something I will not forget.

I would also like to thank the incredible colleagues I have found in my research community. Thank you to the members of the Miller Group, including Sarah Crump, Simon Pendleton, David Harning, Greg de Wet, Chris Florian, and Darren Larsen. The ways in which you have taken me under your wing are too many to count, and I am endlessly grateful. A special thanks to Sarah – I would not be here today if you hadn't carved out the time to give advice to a stranger over a video call and the countless ways, big and small, in which you've helped to bring me into this wonderful world of science in the outdoors. Thank you also to the members of the Organic Geochemistry Lab, including Sebastian Cantarero, Lina Pérez-Angel, Jamie McFarlin, Garret Boudinot, Katie Rempfert, Nadia Dildar, Aria Blumm, Toby Halamka, Ari Meilijson, Edgart Flores, Brooke Holman, Katie Eaman, and many more. It is hard to imagine a more fun, helpful, and musically talented group of researchers. Thank you to the PACEMAP and ILLUME teams, near and far, including Martha Raynolds, Helga Bueltmann, Shawnee Kasanke, Elizabeth Thomas, Devon Gorbey, Kayla Hollister, Kurt Lindberg, Greg de Wet, Nicolò Ardenghi, and Thor Thordarson. It has been an incredible experience to tackle these projects with such a scientifically diverse and universally talented group of people. A special thanks here to Tako. The patient joy and appreciation with which you approach your work in the field taught me more about what it means to be a natural scientist than any course could have done. Thank you as well to the members of the INSTAAR and Geological Sciences staff and research communities. Being surrounded by such interesting people with the common desire to do good in the world has been an inspiration. None of my research would have been possible without the generous support of a number funding sources. These include the National Science Foundation, the Department of Geological Sciences, the University of Colorado, Boulder, and the University of Iceland. I am eternally grateful for their support. Thank you also to the Qikiqtaani Inuit and the Governments of Nunavut and Nunavik who have supported and facilitated our work on their lands. Thank you to all who have braved the weather to work in the field, including Ian Holmen, Sveinbjörn Steinþórsson, Thor Blöndahl, Lucasie Poisey, and Tony Oqqalak, and to Rick Armstrong, Daniel Taukie, Joshua Akavak, Dennis Lock, and many others for facilitating such trips. I would like to acknowledge the role that my privilege has played in every step of my journey to this PhD and pledge to promote justice, equity, diversity, and inclusion in my work. Finally, I could not have made it to this point without the loving support of my parents, Melissa and Bruno, my sisters, Erika and Liana, my partner Raquel, and my family and friends near and far. Thank you for being there for whatever I might need at every step of the way.

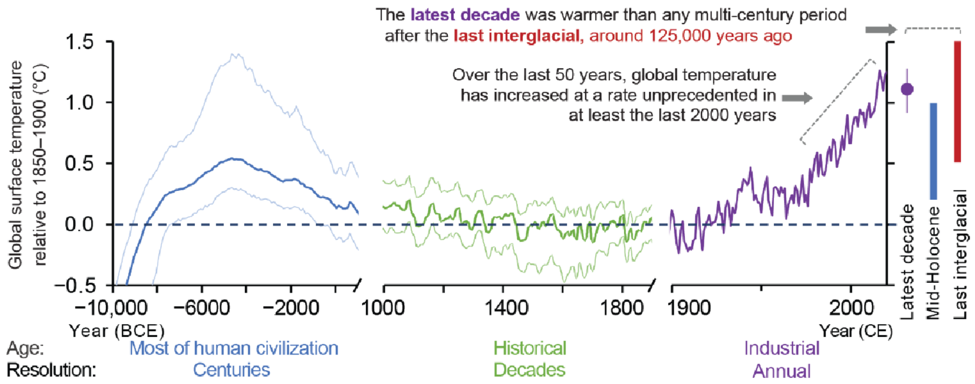
1 CHAPTER I: Introduction

1.1 High Latitudes in a Warming World

Recent decades have shown accelerating manifestations of global-scale climate change. Surface temperatures are increasing at a rate that is unprecedented in at least the last 2,000 years (2 ka), with global temperatures in the last decade rising above those of any multi-century period since the last interglacial (LIG; ~125 ka; Fig. 1-1a; IPCC, 2021). Since 1850-1900, the Earth has warmed globally by ~1°C and is on track to warm an additional ~0.5-3.5°C by century's end (IPCC, 2021). The precise ways in which these rising temperatures manifest in the Earth system has important consequences for human and other life on the planet (IPCC, 2021).

Changes in surface temperature

(a) Recent global temperatures are unprecedented in the era of human civilization



(b) Observed and projected warming are stronger over land than oceans, and strongest in the Arctic

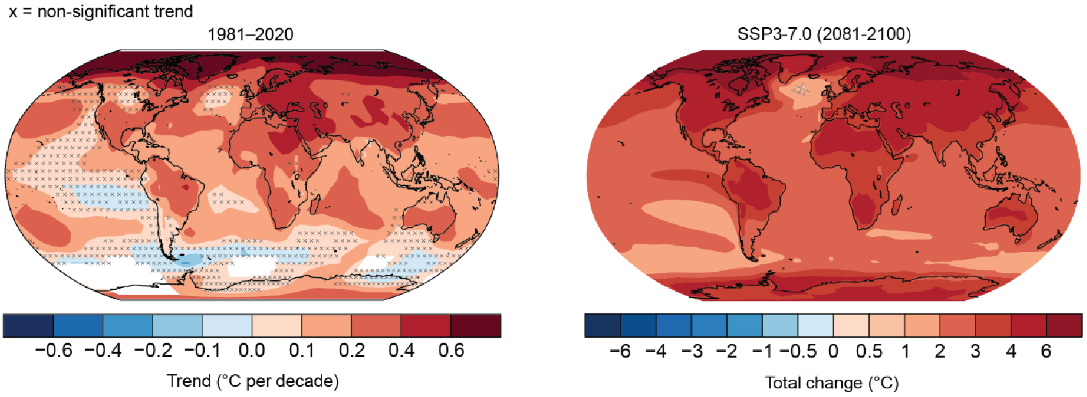


Figure 1-1. Changes in global surface temperature. Figure reproduced as published in the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) with the following caption: "Earth's surface temperature history and future with key findings annotated within each panel. The intent of this figure is to show global surface temperature observed changes from the Holocene to now, and projected changes. (a) Global surface temperature over the Holocene divided into three time scales. (i) 12,000 to 1000 years ago (10,000 BCE to 1000 CE) in 100-year time steps, (ii) 1000 to 1900 CE, 10-year smooth, and (iii) 1900 to 2020 CE (mean of four datasets in panel (c)). Median of the multi-method reconstruction (bold lines), with 5% and 95% percentiles of the ensemble members (thin lines). Vertical bars are 5th to 95th percentile ranges of estimated global surface temperature for the Last Interglacial and mid Holocene (medium confidence) (Section 2.3.1.1). All temperatures relative to 1850–1900. (b) Spatially resolved trends (°C per decade) for ([left hand] map) HadCRUTv5 over 1981– 2020, and ([right hand] map, total change) multi-model mean projected changes from 1995– 2014 to 2081–2010 in the SST3-7.0 scenario. Observed trends have been calculated where data are present in both the first and last decade and for at least 70% of all years within the period using OLS. Significance is assessed with AR(1) correction and denoted by stippling. Hatched areas in the lower map show areas of conflicting model evidence on significance of changes."

Warming has been particularly pronounced at high northern latitudes, where temperature increases outpace the global average by a factor of two or more (Fig. 1-1b; Ballinger et al., 2020; IPCC, 2021). This magnified warming, known as Arctic amplification, is the result of a complex set of physical mechanisms and feedbacks, including reduced albedo through ice loss and vegetation greening and changes in humidity and cloud formation (Serreze and Barry, 2011; Pithan and Mauritsen, 2014). Snow cover, vegetation, marine and terrestrial ice cover, and precipitation regimes in the Arctic have already displayed significant changes as a result (Box et al., 2019 and references therein). Along with their immediate impacts on humans and wildlife, such changes have the potential to accelerate positive warming feedbacks and affect the climate at lower latitudes (Overland et al., 2016; Post et al., 2019; Cohen et al., 2019).

As the globe begins to diverge from historical analogs, our ability to constrain its possible futures becomes increasingly reliant upon the paleoclimate approach, in which climates of the past are reconstructed to inform our understanding of the climates of the future (Miller et al., 2010a; Tierney et al., 2020). In the Arctic, such efforts have revealed multiple periods of past warmth over the Cenozoic (Miller et al., 2010b). Of these periods, the Holocene Thermal Maximum (HTM, ~8-6 ka) and last interglacial (LIG; ~125 ka) may especially serve as important analogs for the coming century due to their temperature ranges (Fig. 1-1) and relative recency. With an emphasis on these time periods and high latitude regions, I therefore aim herein to refine and develop tools for paleoclimate reconstructions to better inform our understanding of our warming world.

1.2 The PACEMAP and ILLUME projects

The work in this thesis represents one component of two collaborative projects aiming to constrain the possible futures of the Arctic by reconstructing their past. The first, Predicting Arctic Change through Ecosystem MoleculAr Proxies (PACEMAP), focuses on the Eastern Canadian Arctic, and especially on Baffin Island. The second, IceLand Landscape reconstrUction with Molecular proxiEs (ILLUME), is centered in Iceland. PACEMAP and ILLUME target warmer-than-present interglacial periods as analogs for future warmth and rely primarily on biomarker proxies preserved in lake sediment archives. Here, I outline the scope of both projects and my role within them.

1.2.1 Study Sites

Two regions that are important for our understanding of high latitude climate change are the Eastern Canadian Arctic (ECA) and Iceland. With low mean annual air temperatures (MAT; Fig. 1-2), substantial late-season snow and ice cover, and tundra ecosystems dipping to some of their southernmost latitudes (Walker et al., 2005), the ECA is a region of cold climate extremes. It is also one of the areas on the planet that is most sensitive to both past (e.g., Brierley et al., 2020; Otto-Bliesner et al., 2021) and modern/future (Fig. 1-1; IPCC, 2021) warming, largely due to the effects of Arctic amplification.

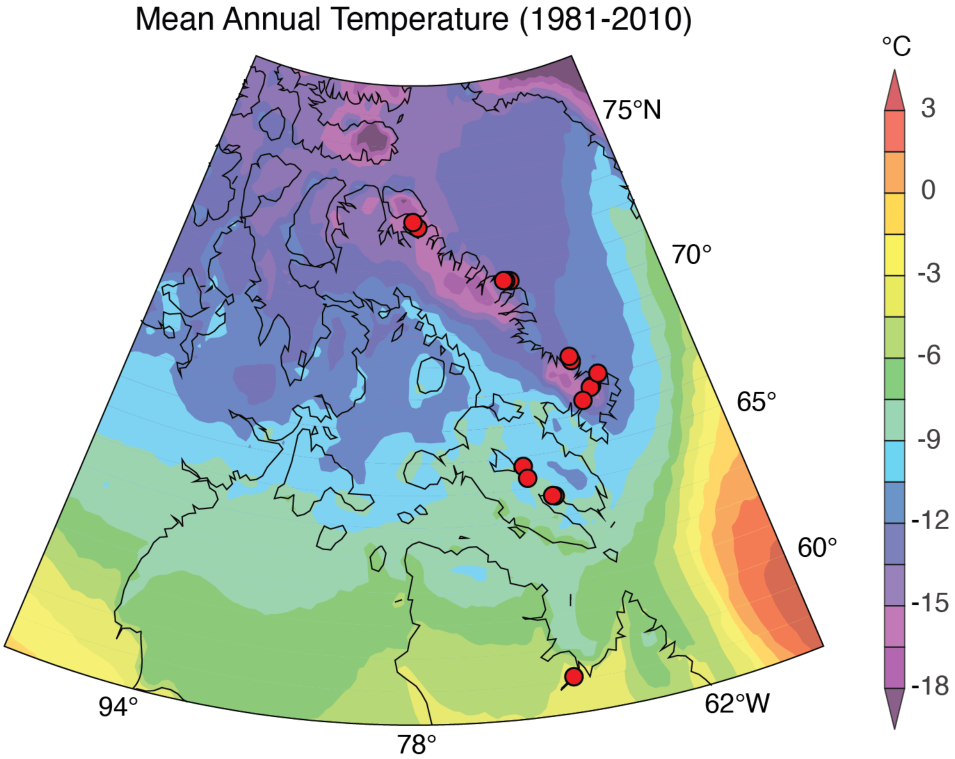


Figure 1-2. Mean annual air temperature (1981-2010) of the Eastern Canadian Arctic. Study sites are shown in red. (Data: NCEP/NCAR North American Regional Reanalysis, NOAA/ESRL Physical Sciences Division).

Though Iceland sits at a comparable latitude, its maritime climate is characterized by warmer temperatures and higher precipitation, in large part due to its position at the northern end of the Atlantic Meridional Overturning Circulation (Wunsch, 1980; Buckley and Marshall, 2016). This major heat advection system buffers winter temperatures and keeps MATs in Iceland above 0°C except at higher altitudes (Fig. 1-3). However, northerly currents bring cold water and occasional sea ice to Iceland's north coast, counteracting this southerly warmth and placing Iceland in a climatically sensitive region that can provide valuable paleoclimate records.

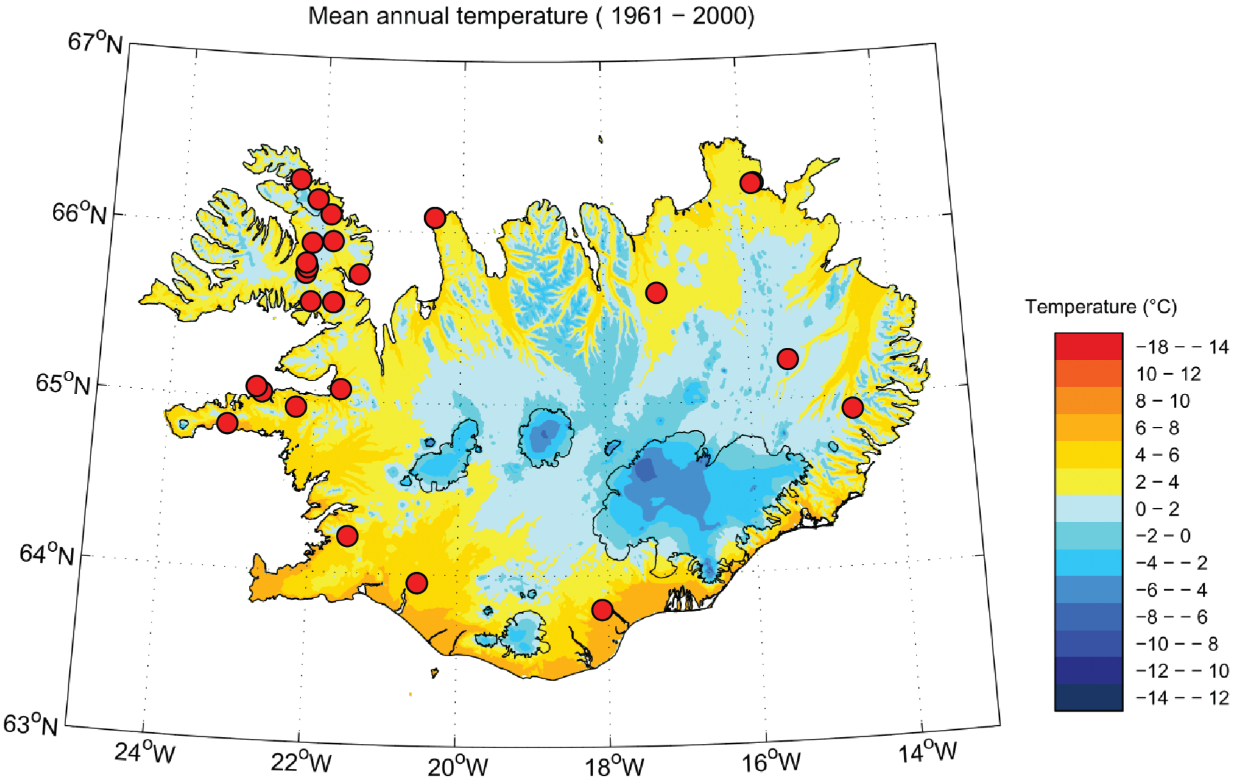


Figure 1-3: Mean annual air temperature (1961-2000) of Iceland. Study sites are shown in red. Figure adapted from Björnsson, (2003).

Recent interglacial periods in these two regions may serve as analogs for future climate and vegetation regimes in the Arctic. The Quaternary temperature histories of both are driven to first order by summer insolation changes in the northern hemisphere associated with periodic fluctuations in the Earth's orbit (c.f., Miller et al., 2010b). Widespread glaciations in the northern hemisphere were punctuated by interglacial periods of relative warmth roughly every 100 kyr. On Baffin Island, some lake sediment archives from the last and penultimate interglacials (LIG and PIG, ~125ka and ~240ka, respectively) have survived the interceding glacial activity (e.g., Briner et al., 2007). Proxy reconstructions from these sediments have suggested LIG summer temperatures of ~+2-7°C compared to 20th century averages and the northward expansion of shrubs (Crump et al., 2021 and references therein).

While cooler than the LIG, the Holocene Thermal Maximum (HTM; ~8-6 ka) of the present interglacial also saw elevated summer temperatures of ~+1-3°C (Miller et al., 2010b) and expanded vegetation ranges (e.g., Crump et al., 2021, 2019; Geirsdóttir et al., 2020). This early warmth was followed by monotonic cooling and neoglaciation (c.f., Geirsdóttir et al., 2019; McKay et al., 2018; Miller et al., 2005), culminating in the Little Ice Age (LIA; c.f., Mann et al., 2009; Miller et al., 2010b, 2012). In Iceland, landscape instability accompanied this decrease in temperature, with widespread soil erosion preceding human settlement and intensifying during the LIA (Geirsdóttir et al., 2020). The extent to which this shift represents the crossing of a threshold in the climate system versus the effect of human activity remains an important question to address (Geirsdóttir et al., 2020).

1.2.2 Research Approach and Proxies

The warmer-than-present temperatures of the geologically recent HTM, LIG, and PIG make them potential analogs for future Arctic conditions. However, temperature reconstructions alone cannot fully characterize the Arctic system. Increases in temperature are likely to drive observed increases in both rainfall and vegetation greening. These variables can in turn feed back to cause further warming through, for example, decreasing surface albedo, changing snowpack characteristics, and increasing relative humidity (Sturm et al., 2005; Miller et al., 2010a; Serreze and Barry, 2011). Reconstructions of the coupled evolution of temperature, vegetation, and precipitation are therefore key to constraining the possible futures of high latitude regions.

Molecular proxies preserved in lake sediment archives offer one avenue for such reconstructions. Ancient environmental DNA preserved in lake sediments can be used to reconstruct local vegetation communities in the Arctic as far back as the LIG (e.g., Crump et al., 2021, 2019). The isotopic compositions of leaf waxes can provide insight into relative humidity and precipitation amounts and seasonality (e.g., Thomas et al., 2016, 2018). Evidence of human activity can be found in the form of polycyclic aromatic hydrocarbons produced from the burning of organic matter (e.g., D'Anjou et al., 2012) and fecal sterols produced by animals such as sheep (e.g., Bull et al., 2002). Finally, the tetraether lipids of bacterial cell membranes form the basis for a quantitative paleothermometer for reconstructing summer warmth (e.g., Crump et al., 2019; Harning et al., 2020).

The PACEMAP and ILLUME projects aim to reconstruct coupled changes in temperature, vegetation, and precipitation using these proxies in lake sediment archives. Both downcore reconstructions and modern proxy development are key to the success of these projects (Fig. 1-4). Detailed vegetation surveys of modern lake catchments will be used to test the preservation of plant DNA in lake sediments. Trends in modern lake water isotopes will inform the influences of seasonality and evaporation on leaf wax isotopes. Finally, temperature calibrations for tetraether lipids will be refined through the use of *in situ* recording of air, soil, and lake water temperatures, repeat measurements of lake water chemistry, and exhaustive sampling of study sites. This final goal is the purpose of this thesis.

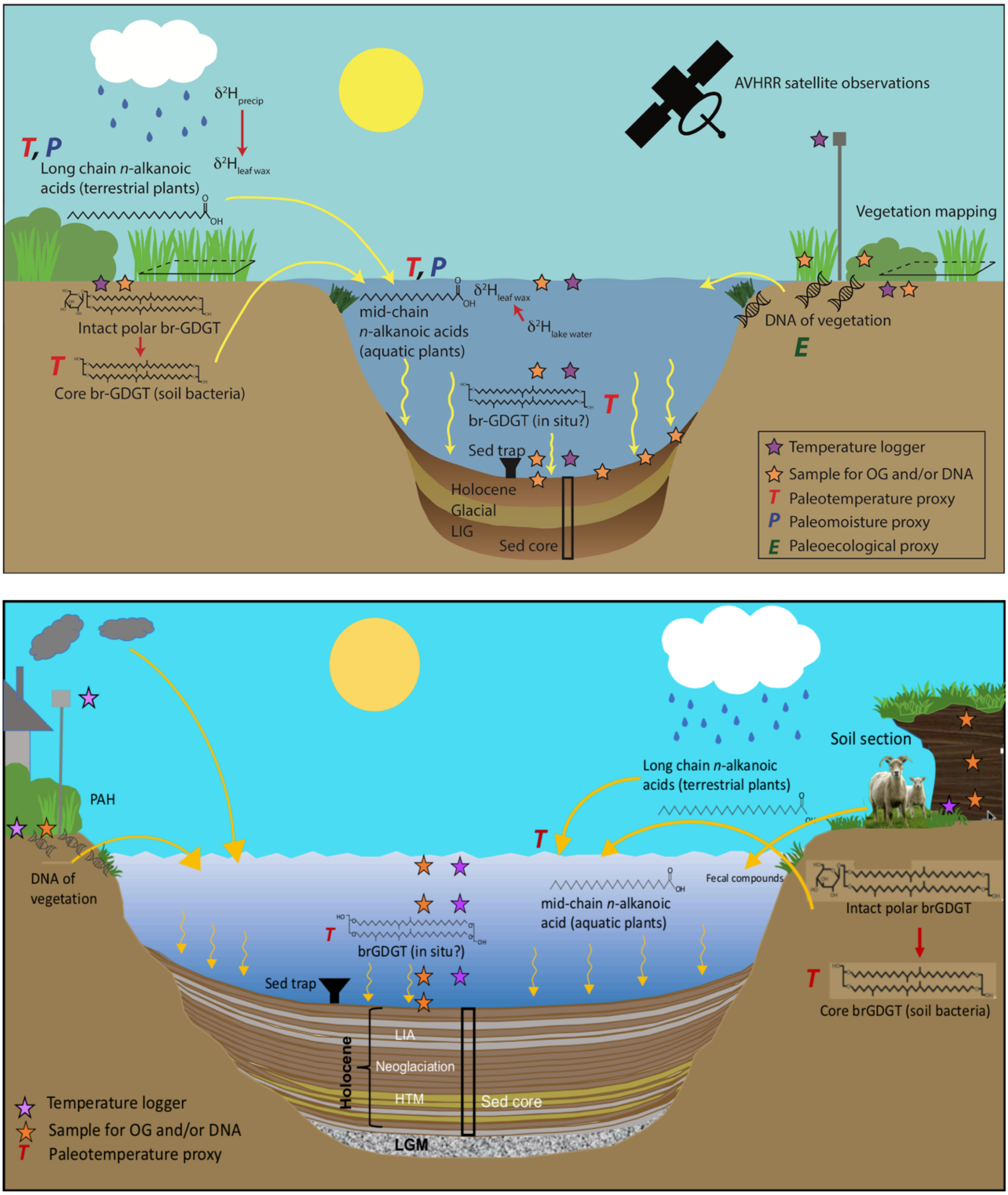


Figure 1-4: Schematic of key proxies and sampling approaches for the PACEMAP (top) and ILLUME (bottom) projects. (Figures courtesy of Sarah Crump and Áslaug Geirsdóttir).

1.3 The brGDGT Paleothermometer – Applications and Challenges

Branched dialkyl glycerol dialkyl tetraethers (brGDGTs; Fig. 1-5) are a class of membrane-spanning lipids whose distributions are empirically related to environmental parameters such as temperature and pH (Weijers et al., 2007b). The lipids' prevalence in natural systems (e.g., soil, peat, lacustrine, and marine; see Chapter 4) coupled with a preservation potential on the order of tens of millions of years (Super et al., 2018) has allowed for their application in numerous sedimentary archives across the globe (e.g., Loomis et al., 2017; Lu et al., 2019; Naafs et al., 2018). In high latitude settings, pioneering works reconstructed temperatures of the Paleocene-Eocene thermal maximum from marine sediment archives in the Arctic Ocean (Weijers et al., 2007c). More recently, the brGDGT paleothermometer has been extended to lacustrine settings, with paleotemperature records being generated in, for example, the ECA (Crump et al., 2019), Iceland (Harning et al., 2020), Greenland (Kusch et al., 2019a), and the Russian Arctic (Lindberg et al., 2021).

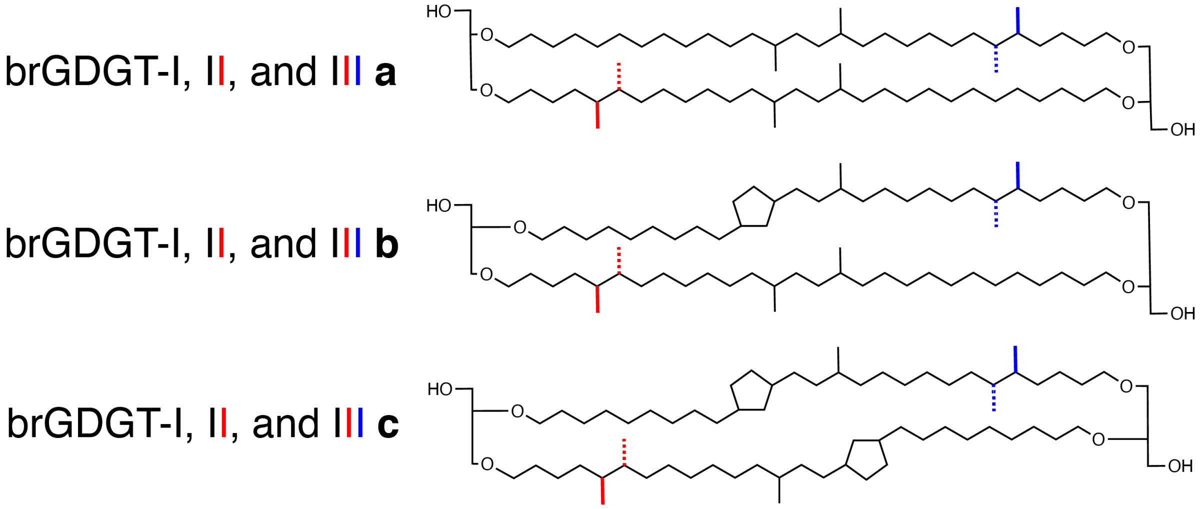


Figure 1-5: Structures of the 15 commonly measured "core" brGDGTs. Top: brGDGT-Ia (black structure only), IIa (with red methyl), and IIIa (with red and blue methyls). Solid and dashed lines indicate the positions of methylations in 5- and 6-methyl isomers, respectively. Middle and lower structures are analogous for mono- and bicyclized compounds.

Despite these successes, however, at least three major gaps exist in our understanding of brGDGTs that complicate their use in paleoclimate research. First, the precise relationships between brGDGT distributions and temperature, pH, and other environmental parameters are still being understood. Early studies found that the number of methylations on brGDGT hydrocarbon backbones was related to temperature, while the number of cyclopentane rings correlated to pH (Fig. 1-5; Weijers et al., 2007a). These relationships have formed the basis for brGDGT-based paleotemperature and paleo-pH proxies to this day. However, other environmental parameters have been shown to interfere with brGDGT-derived temperatures. In lakes, for example, seasonality (e.g., Shanahan et al., 2013), conductivity/salinity (e.g., Raberg et al., 2021; Wang et al., 2021), dissolved oxygen concentration (e.g., Weber et al., 2018), nutrient availability (e.g., Loomis et al., 2014a), and lake water mixing events (e.g., Loomis et al., 2014) have been shown to influence brGDGT distributions. These influences can be difficult to recognize, especially in a paleo record, and few methods exist for applying corrections. For example, some lipid ratios can indicate terrestrial versus marine sources of brGDGTs (e.g., Xiao et al., 2020) and allow for mixed calibrations to be applied (Dearing Crampton-Flood et al., 2018), but such efforts are in their nascent stages.

Next, it is unclear whether the variations in brGDGT distributions with parameters like temperature and pH is a physiological response of source organisms to their environment or an indirect effect resulting from the restructuring of the bacterial community composition. Early works favored the physiological explanation, drawing analogies to similar adjustments observed for other lipid classes (Weijers et al., 2007b). These results were recently supported by molecular dynamics simulations of a brGDGT-based cell membrane (Naafs et al., 2021). However, environmental genomic studies have provided evidence that bacterial community restructuring goes hand-in-hand with changes in brGDGT distributions (e.g., De Jonge et al., 2019; Weber et al., 2018), suggesting that different sets of microbes may produce different sets of brGDGTs. Ultimately, our ability to address this question is hindered by the incomplete identification of brGDGT source organisms, which prevents controlled experimentation in a laboratory setting.

Finally, while the ubiquity of brGDGTs on the landscape is an important advantage, it also poses a challenge as lipids from various sources can become mixed in sedimentary archives. In lake sediments, for example, brGDGTs can be derived from catchment soils (e.g., Peterse et al., 2014), inflowing rivers (e.g., Guo et al., 2020), the lake water column (e.g., Van Bree et al., 2020), and even the sediment itself (e.g., Tierney et al., 2012). As brGDGTs in each of these locations can be expected to record a different set of environmental conditions, deconvoluting their relative contributions becomes an important element of interpreting a brGDGT-based proxy record.

1.4 Scope and goals of this dissertation

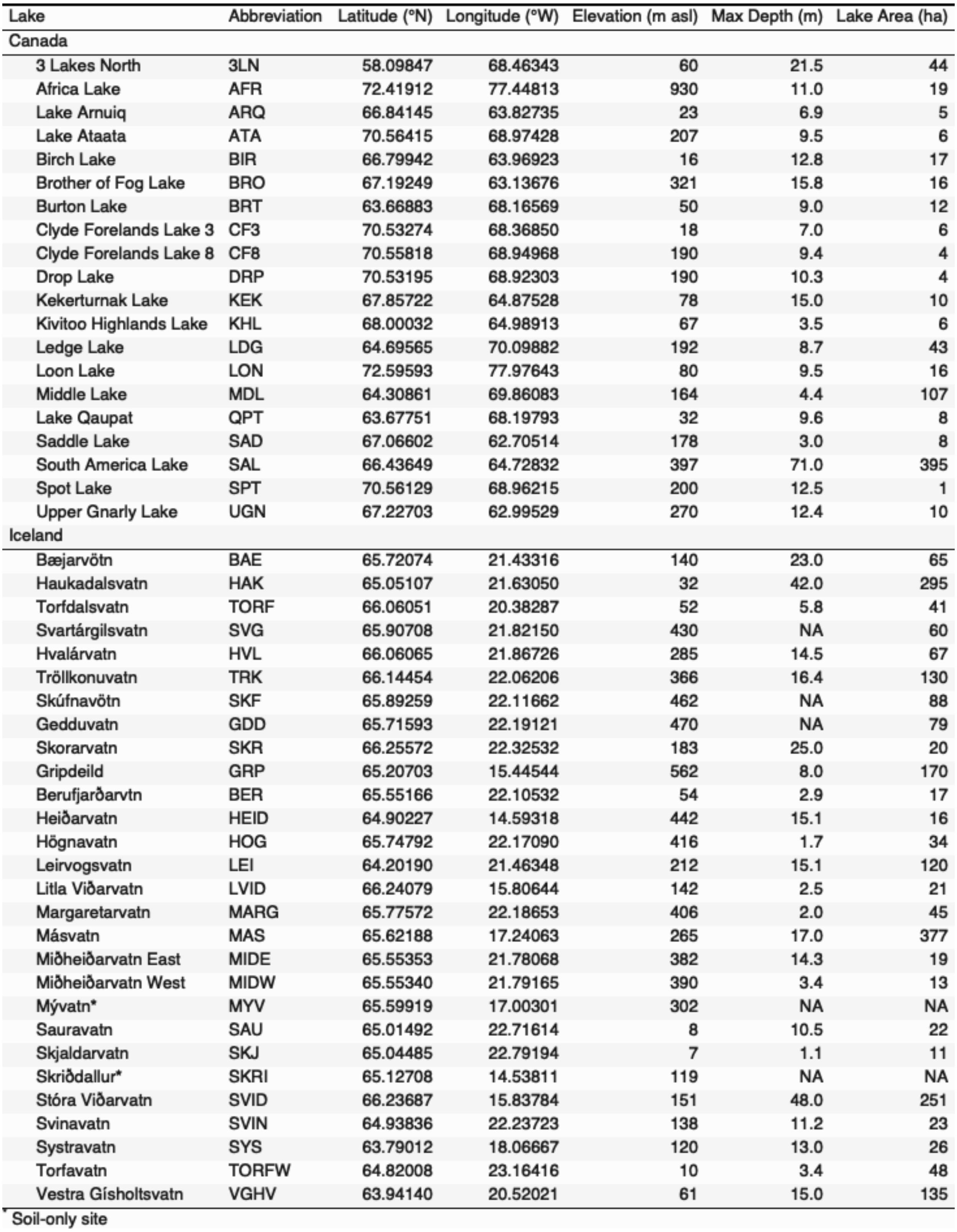
The goal of this dissertation is to develop and refine brGDGT-based paleoclimate proxies for use in high latitude lake sediment archives, specifically those collected as part of the PACEMAP and ILLUME projects. To this end, I present four chapters of study. The first, Chapter 2, aims to characterize the PACEMAP and ILLUME study sites through the use of drone surveys, water chemistry measurements, and *in situ* temperature logging. The next three chapters (Ch. 3-5) each address one of the three major challenges in brGDGT research outlined above. In Chapter 3, I propose a novel technique for deconvoluting the influences of temperature and pH/conductivity on brGDGT distributions. I further examine the effects of seasonality on brGDGTs and generate temperature calibrations for use at high latitudes. In Chapter 4, I compile lipid and environmental data from over a dozen globally distributed sample media. I show that near-universal trends exist across these disparate environments and postulate that an underlying physiological response gives rise to these commonalities. In Chapter 5, I examine a rarely studied form of brGDGTs (intact polar brGDGTs) at Lake Qaupat, Baffin Island and other sites in the ECA. I show that these intact lipids may be specific to different sources in the lake catchment (e.g., soil versus lake water) and propose their use for deconvolving brGDGT signals downcore. Finally, in Chapter 6, I summarize the findings of this thesis, identify important remaining challenges, and discuss future directions.

2 CHAPTER II: Characterization of Study Sites

2.1 Introduction

The PACEMAP and ILLUME projects currently employ 20 Canadian and 28 Iceland study sites, respectively (Table 2-1; Introduction Figs. 1-2 and 1-3). To provide a basic characterization of these sites across the PACEMAP and ILLUME projects, we undertook field campaigns to map the topography, bathymetry, and landscape cover of our lake sites using drone imagery, record in situ air, soil, and lake water temperatures across multiple years, and take repeat measurements of lake water chemistry parameters. Here, I outline our methods, provide basic results, and reference repositories where full datasets are available.

Table 2-1: Study sites of the PACEMAP (Canada) and ILLUME (Iceland) projects.



Lake	Abbreviation	Latitude (°N)	Longitude (°W)	Elevation (m asl)	Max Depth (m)	Lake Area (ha)
Canada						
3 Lakes North	3LN	58.09847	68.46343	60	21.5	44
Africa Lake	AFR	72.41912	77.44813	930	11.0	19
Lake Arnuiq	ARQ	66.84145	63.82735	23	6.9	5
Lake Ataata	ATA	70.56415	68.97428	207	9.5	6
Birch Lake	BIR	66.79942	63.96923	16	12.8	17
Brother of Fog Lake	BRO	67.19249	63.13676	321	15.8	16
Burton Lake	BRT	63.66883	68.16569	50	9.0	12
Clyde Forelands Lake 3	CF3	70.53274	68.36850	18	7.0	6
Clyde Forelands Lake 8	CF8	70.55818	68.94968	190	9.4	4
Drop Lake	DRP	70.53195	68.92303	190	10.3	4
Kekerturnak Lake	KEK	67.85722	64.87528	78	15.0	10
Kivitoo Highlands Lake	KHL	68.00032	64.98913	67	3.5	6
Ledge Lake	LDG	64.69565	70.09882	192	8.7	43
Loon Lake	LON	72.59593	77.97643	80	9.5	16
Middle Lake	MDL	64.30861	69.86083	164	4.4	107
Lake Qaupat	QPT	63.67751	68.19793	32	9.6	8
Saddle Lake	SAD	67.06602	62.70514	178	3.0	8
South America Lake	SAL	66.43649	64.72832	397	71.0	395
Spot Lake	SPT	70.56129	68.96215	200	12.5	1
Upper Gnarly Lake	UGN	67.22703	62.99529	270	12.4	10
Iceland						
Bæjarvötn	BAE	65.72074	21.43316	140	23.0	65
Haukadalsvatn	HAK	65.05107	21.63050	32	42.0	295
Torfdalsvatn	TORF	66.06051	20.38287	52	5.8	41
Svartárgilsvatn	SVG	65.90708	21.82150	430	NA	60
Hvalárvatn	HVL	66.06065	21.86726	285	14.5	67
Tröllkonuvatn	TRK	66.14454	22.06206	366	16.4	130
Skúfnavötn	SKF	65.89259	22.11662	462	NA	88
Gedduvatn	GDD	65.71593	22.19121	470	NA	79
Skorarvatn	SKR	66.25572	22.32532	183	25.0	20
Gripdeild	GRP	65.20703	15.44544	562	8.0	170
Berufjarðarvtn	BER	65.55166	22.10532	54	2.9	17
Heiðarvatn	HEID	64.90227	14.59318	442	15.1	16
Högnavatn	HOG	65.74792	22.17090	416	1.7	34
Leirvogsvatn	LEI	64.20190	21.46348	212	15.1	120
Litla Viðarvatn	LVID	66.24079	15.80644	142	2.5	21
Margaretarvatn	MARG	65.77572	22.18653	406	2.0	45
Másvatn	MAS	65.62188	17.24063	265	17.0	377
Miðheiðarvatn East	MIDE	65.55353	21.78068	382	14.3	19
Miðheiðarvatn West	MIDW	65.55340	21.79165	390	3.4	13
Mývatn*	MYV	65.59919	17.00301	302	NA	NA
Sauravatn	SAU	65.01492	22.71614	8	10.5	22
Skjaldarvatn	SKJ	65.04485	22.79194	7	1.1	11
Skriðdallur*	SKRI	65.12708	14.53811	119	NA	NA
Stóra Viðarvatn	SVID	66.23687	15.83784	151	48.0	251
Svinavatn	SVIN	64.93836	22.23723	138	11.2	23
Systravatn	SYS	63.79012	18.06667	120	13.0	26
Torfavatn	TORFW	64.82008	23.16416	10	3.4	48
Vestra Gísholtsvatn	VGHV	63.94140	20.52021	61	15.0	135

* Soil-only site

2.2 Drone Surveys

We used a Mavic Pro drone (DJI) to map our primary study sites for the PACEMAP project. We used the Maps Made Easy app (Drones Made Easy) to autopilot the drone along a gridded flight path that provided maximum coverage of the immediate lake catchment. Geotagged images taken periodically along this flight path were stitched together using the Maps Made Easy online interface (www.mapsmadeeasy.com). Outputs include high-resolution (~6 cm) true color imagery, digital elevation models (DEMs; absolute elevations referenced at drone takeoff points), and 3-dimensional renderings. True color imagery was further used in combination with manual water depth measurements to generate approximate bathymetric maps (Martha Raynolds, unpublished results). Sample results of the drone mapping campaign for Lake Qaupat, Baffin Island are provided in Figure 2-1. Larger true color and bathymetric maps are shown in the Appendix (Section 8.1.1) for all lakes surveyed (QPT, 3LN, BIR, BRO, CF8, AFR). Additionally, a bathymetric map for Lake ARQ is provided based off satellite imagery (Section 8.1.1). Full resolution versions of all files, along with 3D imagery, are provided in a data repository (Section 8.1.2).

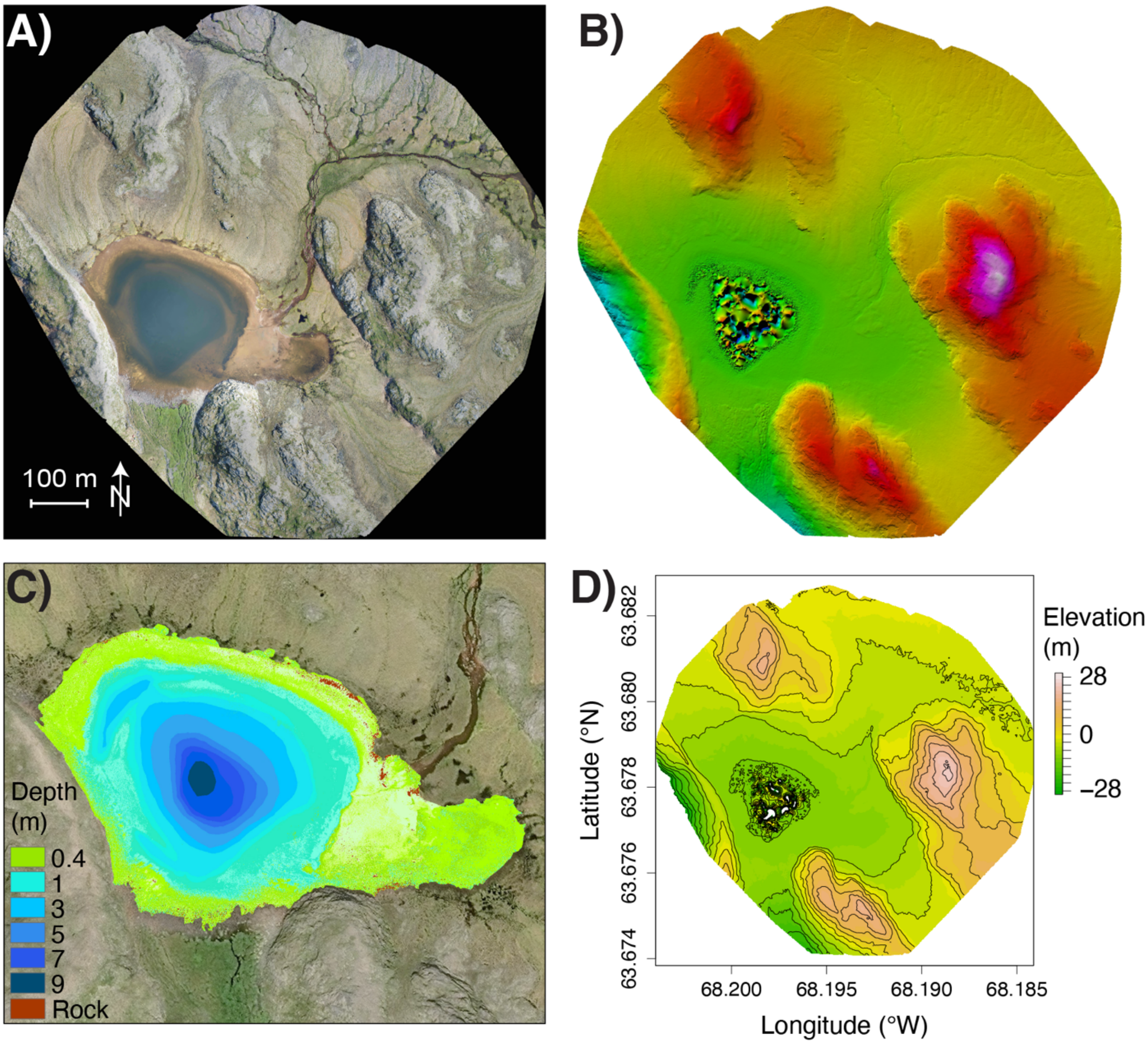


Figure 2-1: Sample results of drone mapping campaign. Survey was conducted at Lake Qaupat, Baffin Island on July 22, 2018. A) True color imagery; B) topography; C) approximate bathymetry; D) digital elevation model. Distortions are present in B) and D) due to light scattered from the surface of the lake.

2.3 Water Chemistry Measurements

Lake water chemistry was measured using two methods. First, we used a multiparameter probe (HydroLab HL4, OTT HydroMet) to measure *in situ* temperature, pH, dissolved oxygen (DO), and electrical conductivity of the water column at 0.5-1 m resolution. These measurements were repeated during each field campaign, resulting in multiple years of summer water chemistry data for many lakes. For some lakes, profiles of the water column were additionally taken through the ice during February (Iceland) or May (Baffin Island) field campaigns. DO measured during the ice-covered season was generally depleted with respect to ice-free conditions. Lakes were on average more acidic in Canada (pH = 5.4 ± 0.1) than Iceland (pH = 6.7 ± 0.1). Sample water chemistry profiles for Lake CF8 are plotted in Figure 2-2 and all data are provided (Section 8.1.2).

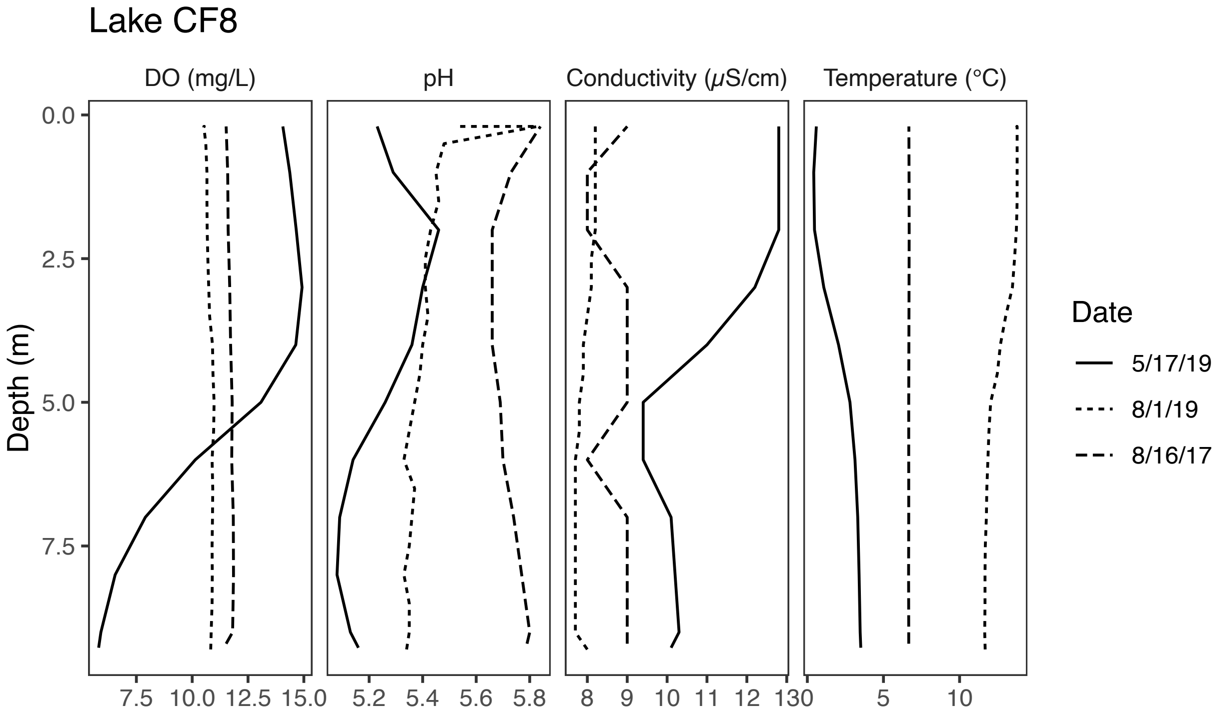


Figure 2-2: Sample water chemistry profiles. Water column dissolved oxygen (DO), pH, raw conductivity, and temperature at Lake Clyde Forelands 8 (CF8) during May 2019 (solid line; ice covered), August 2019 (short dashed; ice-free), and August 2017 (long dashed; ice-free).

In addition to these *in situ* measurements, lake surface (and sometimes bottom or middle) water samples (~0.5 L) were collected for analysis of total phosphorous (TP; acid persulfate digestion), chloride, nitrate, and sulfate (Cl^-, NO_3^{2-}, and SO_4^{2-}; ion chromatography), and total/dissolved organic carbon and total nitrogen (TOC/DOC and TN; combustion using a Shimadzu TOC-L/TNM-L analyzer). Analyses were performed at the Arikaree Laboratory (University of Colorado at Boulder). Canadian and Icelandic lake sites displayed similar average values for most species, with slightly higher concentrations of total phosphorous and sulfate in Iceland on average (Fig. 2-3). Birch Lake in Canada contained sulfate concentrations two orders of magnitude higher than most other lakes (109 mg/L). TOC/DOC, TN, and TP values for Canadian and Icelandic lakes were typical for their climate zone, with values generally falling below the global average (Chen et al., 2015).

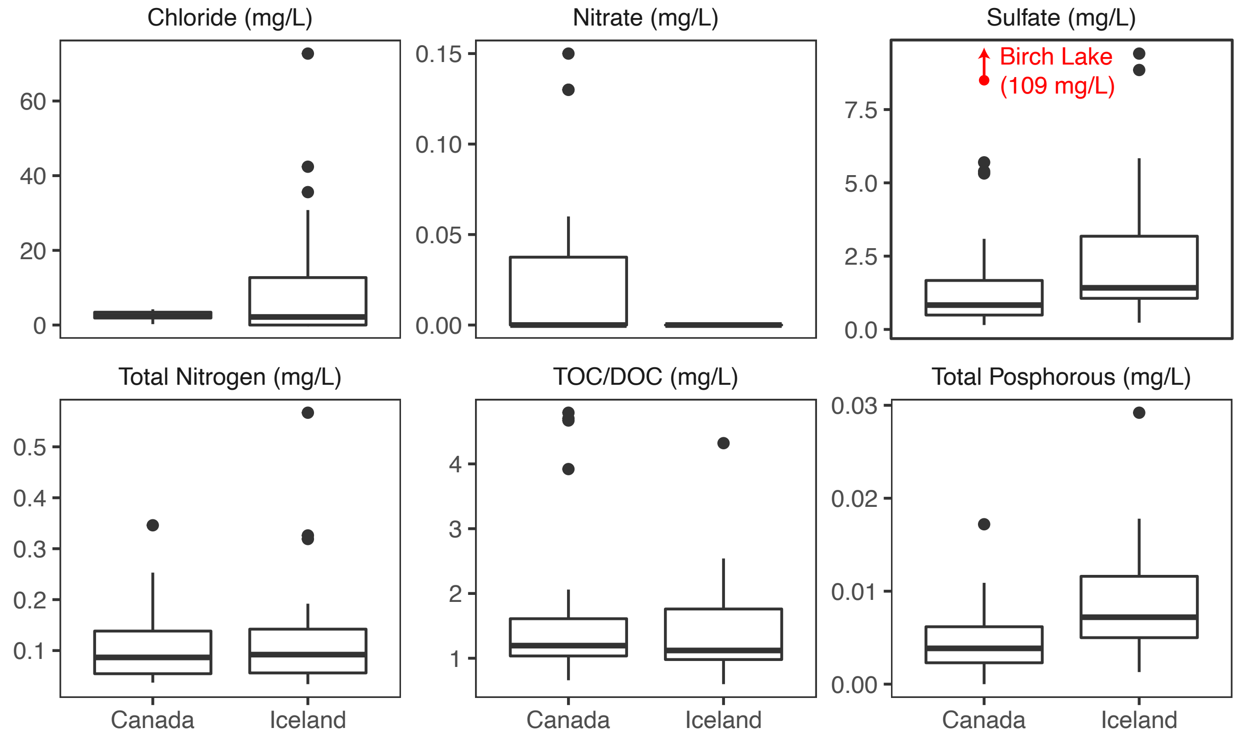


Figure 2-3: Summary of water chemistry measurements. Average lake water chloride, nitrate (NO_3^{2-}), sulfate (SO_4^{2-}), total nitrogen, total/dissolved organic carbon (TOC/DOC), and total phosphorous of Canadian and Icelandic lake sites. Birch Lake in Canada was excluded from the sulfate boxplot due to its anomalously high concentrations (109 mg/L).

2.4 Temperature Logging

Temperature loggers (Thermochron DS1922L or DS1925L ibuttons, Maxim Integrated Products) were used to record air, soil, and lake water temperatures at 1- to 6-hour intervals and 0.0625°C resolution. Air temperature loggers were positioned beneath heat shields and mounted at 2 m above the ground surface. Soil temperature loggers were double wrapped in plastic bags (Whirlpak), placed inside plastic containers (PolyCon), sealed with electrical tape, and buried at 10 cm depth. Water temperature loggers were similarly double wrapped, placed inside 50 mL containers (Nalgene or Falcon), sealed with electrical tape, and suspended at known depths in the lake.

Air, soil, and water column temperatures are plotted for representative Canadian (Lake CF8; Fig. 2-4) and Icelandic (Lake SVID; Fig. 2-5) sites. Air temperatures at Lake CF8 displayed the high seasonality characteristic of Canadian sites. (Air temperatures were not collected on Iceland due to the abundance of existing data from meteorological stations.) Soils displayed a high degree of both spatial and temporal variability, with soils around the same lake sometimes exhibiting substantially different temperature profiles (e.g., Fig. 2-4). These variations are likely the result of differences in snow and vegetation cover, which insulate or shade soils, decoupling them from air temperatures to various degrees (c.f., Myers-Smith et al., 2011). Water temperatures revealed seasonal water column stratification regimes, overturning events, and the onset and duration of ice cover (e.g., Fig. 2-4).

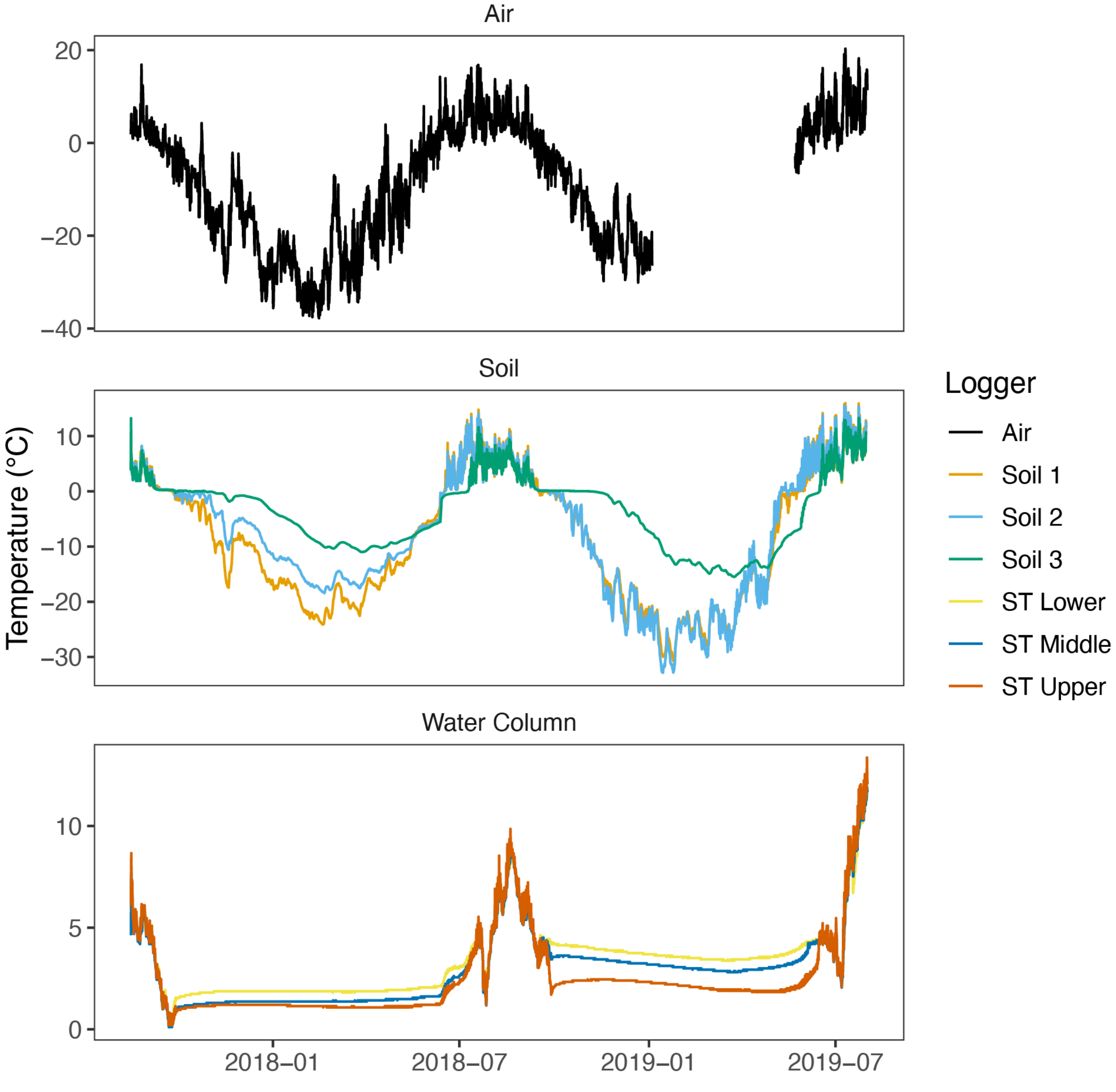


Figure 2-4: Air, soil, and water column temperatures at a representative Canadian lake site (CF8). Air temperature loggers were positioned at a height of 2 m. Soil loggers were buried at 10 cm depth. Water column loggers were attached to sediment traps (ST) at depths of roughly 7 m (Lower), 4.4 m (Middle) and 2.75 m (Upper) in 8 m of water.

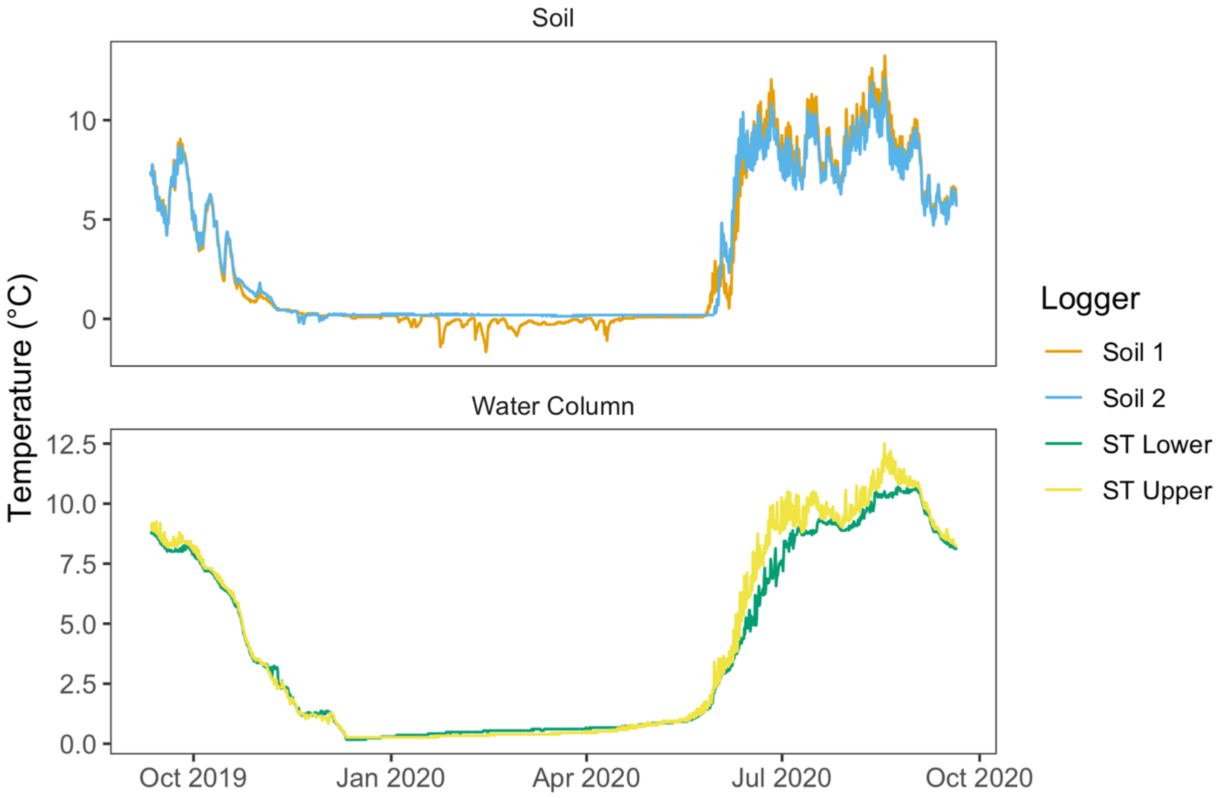


Figure 2-5: Soil and water column temperatures at a representative Icelandic lake site (SVID). Soil loggers were buried at 10 cm depth. Water column loggers were attached to sediment traps (ST) at depths of roughly 18.8 m (Lower) and 2 m (Upper) in 20.8 m of water.

Long-term temperature records from nearby meteorological stations were used to calculate 30-year air temperature normals for all sites. Where air temperature logger data was available (Lakes BIR, BRO, CF8, QPT, and SAL), we estimated local normals by pinning logger data to meteorological records (Fig. 2-6). To achieve this, we first interpolated logger data to hourly resolution. We then calculated the difference between the interpolated and meteorological records (Department of Environment, Government of Canada). Finally, we subtracted the mean monthly differences from the 30-year meteorological station normals (1971-2000) to estimate the site-specific monthly temperature normals. For the remaining Canadian and all Icelandic sites, we generated 30-year climate normals for the roughly same time period (1970-2000) using the WorldClim database (Fick and Hijmans, 2017).

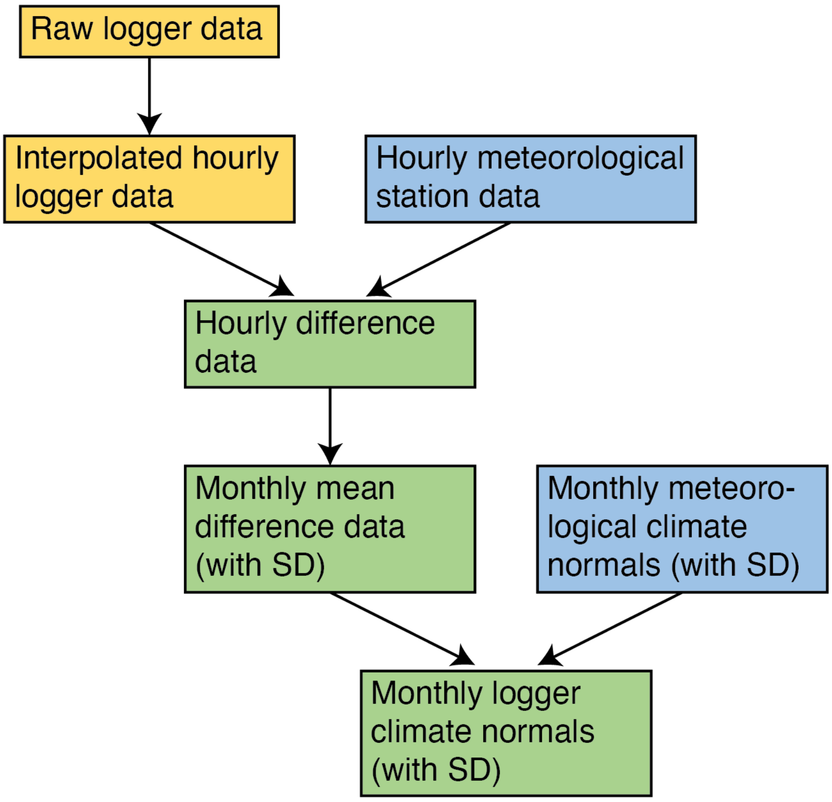


Figure 2-6: Temperature extrapolation methods. Schematic of methods used to extrapolate 1-2 years of air temperature logger data to 30-year normals. "SD" = standard deviation.

The resulting air temperature normals and 1- to 2-year *in situ* soil and water column records highlight some of the defining characteristics of the Canadian and Icelandic study sites. Overall, temperature loggers indicated extreme seasonality in Canada, with winters (December/January/February (DJF)) characterized by their long duration (~9 months of ice cover; e.g., Fig. 2-4) and low air temperatures (-26 ± 3 °C; Fig. 2-7) and brief but substantially warmer summers (June/July/August (JJA); 5 ± 2 °C; Fig. 2-7). Air normals in Iceland displayed only slightly warmer summer temperatures (JJA = 8 ± 1°C) but far milder winters (DJF = -2 ± 2°C; Fig. 2-7). These air temperature trends were for the most part mirrored in soil and water column temperatures in both regions (Fig. 2-7). Notably, however, winter soil temperatures in Canada were highly variable (standard deviation = 6°C; Fig. 2-7), while winter lake water temperatures in all locations were buffered at ~0-4°C by ice formation. All data from temperature loggers is available in Section 8-2-2.

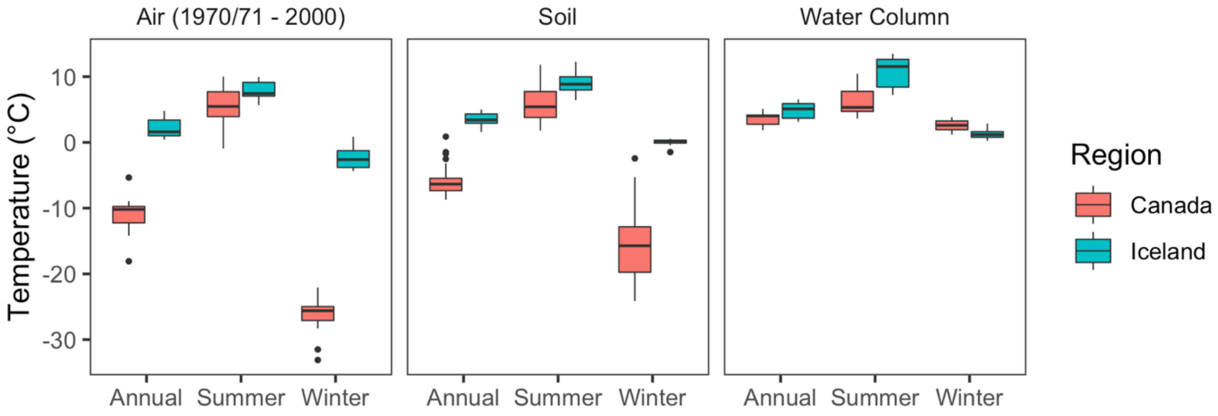


Figure 2-7: Summary of temperature logger results. Average temperatures of air (1971-2000 or 1970-2000 average), soil (10 cm depth, single year between 2017 and 2020), and water column (averaged over depths, single year between 2017 and 2020). Soil and water column temperatures were recorded by temperature loggers in situ. Air temperatures were a combination of in situ and modeled measurements.

2.5 Discussion

The data we generated through the use of drone surveys, water chemistry measurements, and *in situ* temperature logging provided us a with a basic, but informative characterization of the PACEMAP and ILLUME study sites. Drone surveys allowed for high resolution mapping of the landscape topography, vegetation cover, and lake bathymetry. Water chemistry measurements revealed (seasonal) lake stratification regimes and nutrient limitations. Temperature logging provided *in situ* data across highly variable landscapes, markedly reducing the uncertainties associated with extrapolation of temperatures from climate models or meteorological stations. These data improved our understanding of our study sites, aided in interpretation of our results, and guided subsequent field campaigns. They also provide useful baseline measurements for future studies, such as repeat sampling or long-term monitoring campaigns.

3 CHAPTER III: Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments

Abstract. Distributions of branched glycerol dialkyl glycerol tetraethers (brGDGTs) are frequently employed for reconstructing terrestrial paleotemperatures from lake sediment archives. Although brGDGTs are globally ubiquitous, the microbial producers of these membrane lipids remain unknown, precluding a full understanding of the ways in which environmental parameters control their production and distribution. Here, we advance this understanding in three ways. First, we present 43 new high-latitude lake sites characterized by low mean annual air temperatures (MATs) and high seasonality, filling an important gap in the global dataset. Second, we introduce a new approach for analyzing brGDGT data in which compound fractional abundances (FAs) are calculated within structural groups based on methylation number, methylation position, and cyclization number. Finally, we perform linear and nonlinear regressions of the resulting FAs against a suite of environmental parameters in a compiled global lake sediment dataset (n = 182). We find that our approach deconvolves temperature, conductivity, and pH trends in brGDGTs without increasing calibration errors from the standard approach. We also find that it reveals novel patterns in brGDGT distributions and provides a methodology for investigating the biological underpinnings of their structural diversity. Warm-season temperature indices outperformed MAT in our regressions, with Months Above Freezing yielding the highest-performing model (adjusted $R^2 = 0.91$, RMSE = 1.97°C, n = 182). The natural logarithm of conductivity had the second-strongest relationship to brGDGT distributions (adjusted $R^2 = 0.83$, RMSE = 0.66, n = 143), notably outperforming pH in our dataset (adjusted $R^2 = 0.73$, RMSE = 0.57, n = 154) and providing a potential new proxy for paleohydrology applications. We recommend these calibrations for use in lake sediments globally, including at high latitudes, and detail the advantages and disadvantages of each.[1]

3.1 Introduction

Paleotemperature records from lake sediment archives are highly sought after in studies of terrestrial paleoclimate. Bacterial branched glycerol dialkyl glycerol tetraether (brGDGT) lipids have solidified themselves as an important tool in this pursuit (Fig. 8-15; Schouten et al., 2013). First isolated from peat (Sinninghe Damsté et al., 2000), these membrane lipids have since been measured in increasingly diverse settings, from marine, soil, lacustrine, and riverine locations (Hopmans et al., 2004; Weijers et al., 2006; Pearson et al., 2011; De Jonge

---

[1] A version of this chapter is published:
Raberg, J. H., Harning, D. J., Crump, S. E., De Wet, G., Blumm, A., Kopf, S., Geirsdóttir, Á., Miller, G. H. and Sepúlveda, J.: Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments, Biogeosciences, 18, 3579–3603, doi:10.5194/bg-18-3579-2021, 2021

et al., 2014b, respectively) to hot springs, fossil bones, groundwater, deep ocean trenches, and methane seeps (Li et al., 2014; Dillon et al., 2018; Ding et al., 2018; Xiao et al., 2020; Zhang et al., 2020, respectively). Their ubiquity in nature has given them widespread applicability as environmental proxies; brGDGTs have been used to reconstruct temperature in a variety of archives including lake sediments (e.g. de Wet et al., 2016), marine sediments (e.g. Dearing Crampton-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). They have additionally been used to reconstruct lake water pH (e.g. Cao et al., 2017). As the microbial producers of brGDGTs remain elusive (Sinninghe Damsté et al., 2018), these paleoclimate reconstructions currently rely on empirical calibrations at both the regional and global level.

At the heart of brGDGT calibrations is the observation that the degree of alkyl-chain methylation and cyclization are correlated to environmental temperature and pH, respectively. These relationships were first quantified by the Methylation index and Cyclization ratio of Branched Tetraethers (MBT and CBT) in a global soil dataset (Weijers et al., 2007b). The authors proposed physiological explanations for both connections, positing that an increase in methylation number will enhance membrane fluidity, a desirable trait in cold environments, while a greater number of cyclic moieties could improve proton permeability, an advantageous adaptation at high pH. These physiological responses have precedent in other bacterial lipid classes (Reizer et al., 1985; Beales, 2004; Yuk and Marshall, 2004) and appear to function for brGDGTs as well. However, genomic analyses of environmental samples (Weber et al., 2018; De Jonge et al., 2019; Van Bree et al., 2020) have suggested that differences in brGDGT distributions may also stem from shifts in bacterial community composition. Variations in the position of alkyl-chain methylations (Fig. 8-15; De Jonge et al., 2014a) further complicate the picture, with most studies showing 5-methyl brGDGT isomers to correlate better with temperature than their 6-methyl counterparts (e.g. Russell et al., 2018), but others arriving at the opposite result (e.g. Dang et al., 2018). These isomeric variations have additionally been shown to correlate with pH in lake sediments (Dang et al., 2016). These discoveries highlight the multifaceted nature of the empirical relationship between brGDGTs and environmental gradients and the need for further study.

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

$$fx_i = x_i/(x_1 + x_2 + \cdots + x_n) \qquad (3-1)$$

where any x is the absolute abundance of the given compound in the set. For brGDGTs, these FAs are traditionally calculated using all 15 commonly measured compounds (Fig. 8-15),

$$\begin{aligned} fx_i = x_i/(&Ia + Ib + Ic + IIa + IIb + IIc + IIIa + IIIb + IIIc + IIa' + IIb' + IIc' \\ &+ IIIa' + IIIb' \\ &+ IIIc') \end{aligned} \qquad (3-2)$$

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

On the other side of the calibration equations are the environmental variables that are regressed against brGDGT indices and FAs. Mean annual air temperature (MAT) has been the traditional target of brGDGT calibrations in lake sediments (e.g. Tierney et al., 2010; Loomis et al., 2012). It was recognized early on, however, that brGDGT-derived temperatures in cold regions may more accurately reflect warm-season temperatures (Sun et al., 2011; Pearson et al., 2011), an hypothesis that was strongly supported in high-latitude lake sediments (Shanahan et al., 2013; Peterse et al., 2014; Foster et al., 2016). Since the methodological advances that allowed for the separation of 5- and 6-methyl isomers (De Jonge et al., 2014a) and the development of new calibrations, both modern (Hanna et al., 2016; Dang et al., 2018; Cao et al., 2020) and paleo (Super et al., 2018; Thomas et al., 2018; Crump et al., 2019; Harning et al., 2020) studies have continued to support a warm-season bias. Additionally, a recent Bayesian calibration found the mean temperature of Months Above Freezing (MAF) to be the only mode to significantly correlate with brGDGT distributions in a global lake sediment dataset (Martínez-Sosa et al., 2020b). However, as the source organisms of brGDGTs remain incompletely known, other measures of warm-season temperature such as peak warmth or cumulative heat may prove to be more biologically relevant. Furthermore, the nature of the warm-season bias has yet to be tested thoroughly in the regions in which it is most pronounced – namely, those with low MAT and high seasonality. As these are the regions that are currently experiencing the most rapid climate change (Landrum and Holland, 2020), their temperature histories are of high interest (Miller et al., 2010a) and the quantification of the 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 environmental variables are most appropriate to reconstruct. For example, lacustrine brGDGTs were successfully calibrated to lake water temperature at a site in Greenland (Zhao et al., 2021), but such an effort would likely be less successful for a lake with substantial soil input.

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

3.2 Methods

3.2.1 Study sites and sample collection

Surface sediments (0-0.5, 0-1, or 0-2 cm; Ekman box corer or core-top sediment) were collected from 43 lakes (26 from Iceland; 16 from Baffin Island, Arctic Canada; 1 from Northern Quebec; Fig. 3-1 insets) between 2003 and 2019. For 28 of these lakes, water temperature, pH, conductivity, and DO were measured at the time of sampling during the summers of 2017-2020 using a multiparameter probe (HydroLab HL4, OTT HydroMet). These parameters were additionally measured beneath the lake ice in Feb-May of 2018-2020 for 11 lakes. All but one of these lakes experienced depleted bottom water oxygen levels under the ice relative to ice-free conditions. We therefore assume that our ice-free water chemistry profiles do not capture the minimum DO (DO_{min}) levels experienced by our Canadian and Icelandic lakes and exclude them from our analysis of DO_{min}, with the exception of the site from Northern Quebec, which contained a summer oxycline. All other water chemistry parameters were averaged across all depths and seasons before being used in calibrations.

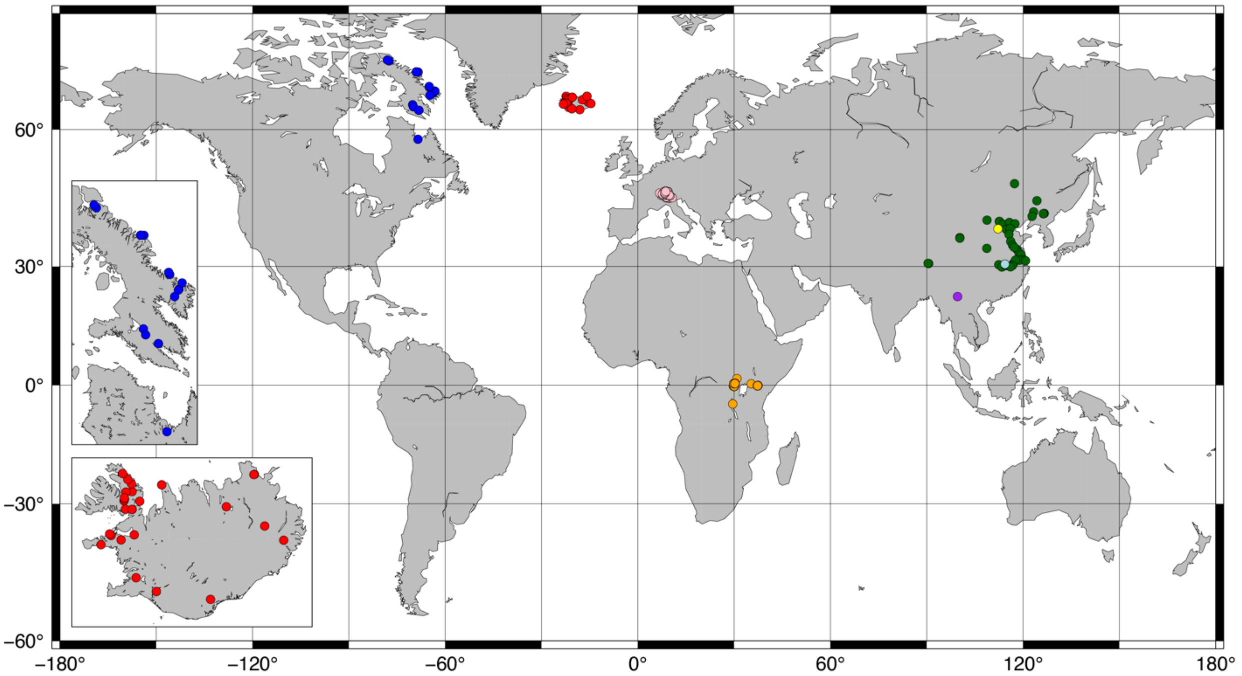


Figure 3-1: Map of sites included in this study. Red and blue: Iceland and Canada (this study; see insets); Orange: Russell et al. (2018); Green: Dang et al. (2018); Pink: Weber et al. (2018); Yellow: Cao et al. (2020); Light blue: Qian et al. (2019); Purple: Ning et al. (2019).

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

3.2.2 Sample extraction and analysis

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

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

3.2.3 Comparison of ASE and BD Extraction Methods

To ensure that brGDGT distributions were agnostic to our extraction method, we extracted three surface sediments, two suspended particulate matter (SPM) samples (2.5 L lake water filtered onto 0.3 µm glass fiber filters), and three soils from Baffin Island using the ASE and BD methods in parallel (Fig. S1). The mean difference in $\mathrm{MBT'_{5Me}}$ (Eq. A3) between the two extraction methods was 0.006 ± 0.004, or 2 ± 2%. This translates to a $\mathrm{MBT'_{5Me}}$-derived temperature difference of 0.2 ± 0.2°C using recent calibrations for soils (Naafs et al., 2017) and lake sediments (Russell et al., 2018), which is well below their respective RMSEs of 5.3°C and 2.14°C (Fig. S2). To test for compound-specific differences, we calculated percent differences in FAs between the two methods. For compounds with FAs > 0.05 in the ASE method, the mean absolute percent difference compared to BD was 4 ± 3%. For the lower abundance compounds (FA ≤ 0.05), this difference was higher (19 ± 17%). No biases in the FA differences were 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 ± 0.6% brGDGTs. These residual brGDGTs had a similar $\mathrm{MBT'_{5Me}}$ to that of original BD extract (mean difference = 0.02 ± 0.01, equivalent to 0.5 ± 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.

3.2.4 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). Monthly temperatures for Central European sites were derived using monthly altitudinal lapse rates constructed from climate normals (1970 to 2013) of 148 meteorological stations (Federal Office of Meteorology and Climatology: MeteoSwiss). Monthly temperature data was not available for the East African lakes. However, the seasonality (standard deviation of monthly temperatures) of these lakes is low (0.5 ± 0.2 °C in the WorldClim database, or $< 2\%$ of range of the dataset). We therefore approximate all monthly temperatures to be equivalent to MAT for these lakes. Monthly temperature data from all other studies were either published or provided by the authors.

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

3.2.5 Statistical Methods

To construct calibrations between brGDGT FAs and environmental variables, we used the following method. Each FA was first regressed alone against the environmental variable being investigated. Compounds with a correlation *p-value* ≥ 0.01 were considered non-significant and removed from further analysis (Pérez-Angel et al., 2020). Fits were then constructed from the remaining compounds using two independent approaches. The first approach was stepwise forward selection/backwards elimination (SFS/SBE) using the MASS package (Venables and Ripley, 2002) in R (Team R Development Core, 2021). This approach finds the best fit by sequentially adding (SFS) or removing (SBE) terms in a generalized linear model and evaluating the resulting fit using the Bayesian Information Criterion (BIC; Schwarz, 1978). The approach is common for constructing brGDGT calibrations (e.g. Dang et al., 2018; Russell et al., 2018), but it is not exhaustive. We therefore additionally used the leaps package (Lumley, 2020) in R to evaluate all possible linear combinations of fitting variables (the "combinatoric" approach, Pérez-Angel et al., 2020). We again used the BIC to select the best fit. For both methods, we imposed the additional criterion that each of the resulting fitting variables must itself be statistically significant ($p < 0.01$). To help avoid overfitting, we additionally used the adjusted R^2 to evaluate calibration performances. Some of our variables (conductivity, depth, surface area to depth, and volume) spanned multiple orders of magnitude. We performed regressions against the natural logarithm of these variables, which greatly improved their normality.

We applied this calibration procedure to the FAs of each brGDGT structural set and subset defined below (Sect. 3.3). In the subset-specific calibrations, it was sometimes possible for a single compound to dominate (FA = 1; see section S1.4). These samples were generally clear outliers resulting from the low natural abundances of all other members of the subset and they were therefore removed from the subset-specific calibration models. We additionally tested for linear regressions against a number of previously defined brGDGT indices. A summary of these indices and their definitions is provided in the Appendix (Section 8.2.1).

All correlations reported in the text and figures were significant ($p < 0.01$) except those marked with an asterisk or with a *p-value* provided. All R^2 values reported in the text and figures are adjusted R^2.

3.3 Partitioning brGDGTs into structural sets for FA calculations

The structure of brGDGTs can vary in three (currently observed) ways: methylation number, methylation position, and cyclization number. These variations result in 15 commonly measured brGDGTs (Fig. 8-15). The standard approach in brGDGT analysis is to calculate fractional abundances using all 15 of these compounds (Eq. 3-2). By mathematically mixing brGDGTs with varying methylation number, methylation position, and cyclization number, however, this approach risks convoluting the influences of disparate environmental variables. Here, we present a method of grouping brGDGTs that highlights one type of structural variation (e.g. methylation number) while holding one or both of the others constant. By calculating FAs within each group, we aim to deconvolve the influences of temperature, pH, and other environmental variables on the structural variations of brGDGTs.

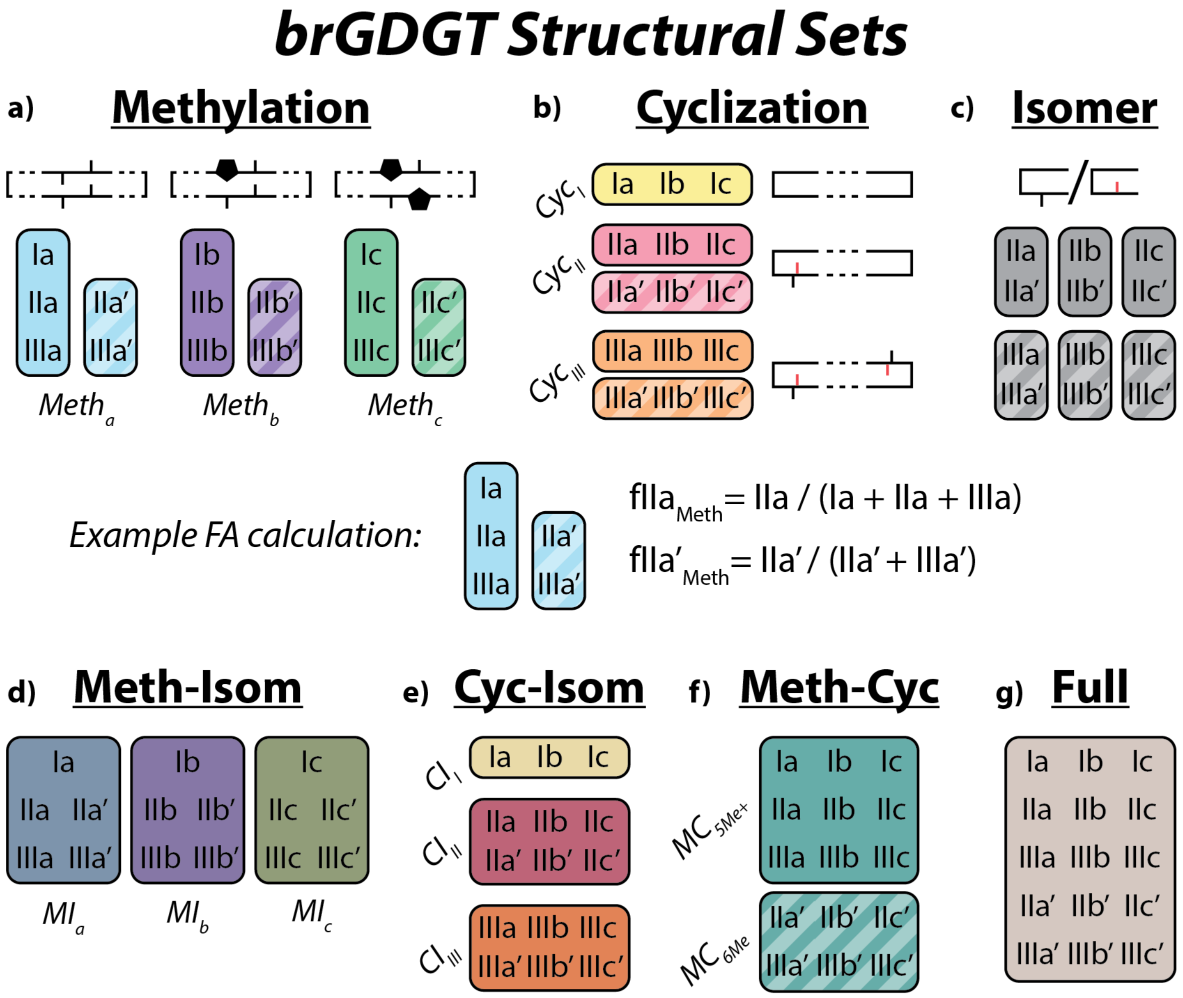


Figure 3-2: Schematic of the basic (a-c) and combined (d-g) brGDGT structural sets. Fractional abundances are calculated within each boxed group independently (Eq. (3-3 to 3-9) and Table 8-1). Schematic structures highlight the defining alkyl-chain moieties, with cyclopentane rings filled in for emphasis and C6 methylations denoted in red. Complete structures are available in Fig. 8-14.

To explore how changes in methylation number alone relate to environmental parameters, we constructed the Methylation (Meth) set (Fig. 3-2a). This set was generated by grouping brGDGTs with the same number of cyclopentane rings and the same methylation positions. Fractional abundances calculated within the Meth set solely reflect changes in methylation number, while ring number and isomer designation are held constant. Within the Meth set, we defined the Meth-5Me and Meth-6Me subsets as those that contained only 5- and 6-methyl lipids, respectively. As the tetramethylated brGDGTs (Ia, Ib, and Ic) are neither 5-methyl nor 6-methyl isomers, we generated versions of these two subsets that excluded (Meth-5Me, Meth-6Me) and included (Meth-5Me+, Meth-6Me+) these compounds (Fig. S3a-d). We found that the correlation between 5-methyl isomers and temperature was significantly improved when the tetramethylated compounds were included in their FA calculations, while the opposite was true for the 6-methyl compounds (see discussion in Sect. 3.4.2.2). We therefore grouped the tetramethylated compounds with the 5-methyl compounds (and not with the 6-methyl compounds) when generating the Meth set. FAs for the Meth set were calculated using Eq. (3-3),

$$fxy_{Meth} = xy\Big/\sum_{n=I}^{III} ny\,;\ fxy'_{Meth} = xy'\Big/\sum_{n=II}^{III} ny' \qquad (3-3)$$

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. 3-2b). 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. (3-4),

$$fxy^{(\prime)}{}_{Cyc} = xy^{(\prime)}\Big/\sum_{m=a}^{c} xm^{(\prime)} \qquad (3-4)$$

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).

Third, we defined the Isomer (Isom) set to isolate changes in the relative abundances of brGDGT isomers (Fig. 3-2c). The Isom set was constructed by grouping brGDGTs with the same number of methylations and cyclizations. Its FAs are solely a measure of the relative abundances of 5- and 6-methyl isomers, without the convoluting influence of ring or methylation number variations. The isomeric diversity of brGDGTs is large, however, and there are structural variations that are not controlled for within this set. For example, hexamethylated brGDGTs have two isomers: one with methylations on different alkyl chains (e.g. at C5 and C5') and another with methylations on the same chain (e.g. C5 and C24; De Jonge et al., 2013). As these compounds coelute, they cannot be treated independently at this time. Additionally, it is unclear whether brGDGT-IIIa" (Weber et al., 2015), which contains both a C5 and a C6' methylation, should be grouped with 5- or 6-methyl brGDGTs, or whether the three recently-identified 7-methyl brGDGTs (Ding et al., 2016) should be included in the Isom series. As these compounds are rarely reported, we excluded them from our analysis here, but suggest the possibility of expanding the Isom set to include them in the future should more data become available. The Isom FAs were calculated using Eq. (3-5),

$$fxy^{(\prime)}{}_{Isom} = xy^{(\prime)}\Big/\sum_{isomers} xy \qquad (3-5)$$

where f*xy* 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. 3-2d). 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. 3-2e) is analogously constructed by holding methylation number constant, while the Meth-Cyc (MC; Fig. 3-2f) set holds methylation position constant (again treating tetramethylated brGDGTs as 5-methyl compounds, Fig. S3g-j). Finally, we defined the Full set (Fig. 3-2g), which takes the standard approach of allowing all three structural characteristics to vary freely by grouping all 15 commonly measured brGDGTs together. The FAs of the combined sets are calculated using Eqs. (3-6 to 3-9),

$$fxy^{(\prime)}{}_{MI} = xy^{(\prime)} \Big/ \sum_{isomers} \sum_{n=I}^{III} ny \qquad (3-6)$$

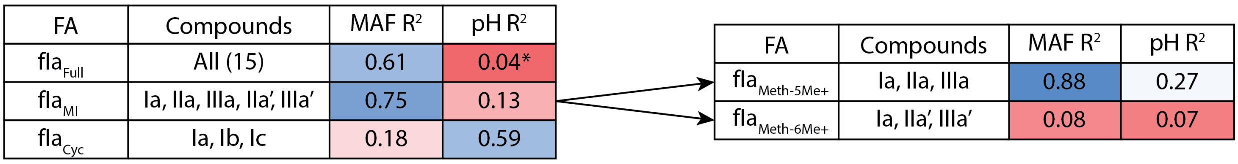
$$fxy^{(\prime)}{}_{CI} = xy^{(\prime)} \Big/ \sum_{isomers} \sum_{m=a}^{c} xm \qquad (3-7)$$

$$fxy_{MC} = xy \Big/ \sum_{n=I}^{III} \sum_{m=a}^{c} nm; \; fxy'_{MC} = xy' \Big/ \sum_{n=II}^{III} \sum_{m=a}^{c} nm' \qquad (3-8)$$

$$fxy^{(\prime)}{}_{Full} = xy^{(\prime)} \Big/ \sum_{isomers} \sum_{n=I}^{III} \sum_{m=a}^{c} nm \qquad (3-9)$$

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 tetramethylated compounds are treated as 5-methyl isomers. An expanded guide to FA calculations is provided in Table 8-1.

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



FA	Compounds	MAF R^2	pH R^2
fIa_{Full}	All (15)	0.61	0.04*
fIa_{MI}	Ia, IIa, IIIa, IIa′, IIIa′	0.75	0.13
fIa_{Cyc}	Ia, Ib, Ic	0.18	0.59

FA	Compounds	MAF R^2	pH R^2
$fIa_{Meth\text{-}5Me+}$	Ia, IIa, IIIa	0.88	0.27
$fIa_{Meth\text{-}6Me+}$	Ia, IIa′, IIIa′	0.08	0.07

Figure 3-3: Adjusted R2 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. "Compounds" denote all brGDGTs used in the FA calculation. All values are significant (p < 0.01) unless marked with an asterisk.

3.4 Results and Discussion

3.4.1 Distributions of brGDGTs in Icelandic and Canadian lake sediments

The new Icelandic and Canadian lake sediments bolster the global dataset on the cold end, extending the lowest MAT from -0.2°C to -18°C and containing 28 of the 30 coldest samples by MAT. The low temperatures are reflected in the brGDGT distributions of these sediments (Fig. 3-4), which contain on average a higher $fIIIa_{Full}$ and lower fIa_{Full} than the other samples in this dataset. Additionally, the Canadian and Icelandic datasets provide important end-member samples, including those with the highest $fIIIa_{Full}$, $fIIIb_{Full}$, and $fIIIc_{Full}$ and the lowest or second-lowest fIa_{Full}, fIb_{Full}, and fIc_{Full}. The new samples also contain some of the lowest fractional abundances of 6-methyl isomers, providing low end-member values for $fIIa'_{Full}$, $fIIb'_{Full}$, $fIIc'_{Full}$ and below-average values for the hexamethylated 6-methyl brGDGTs.

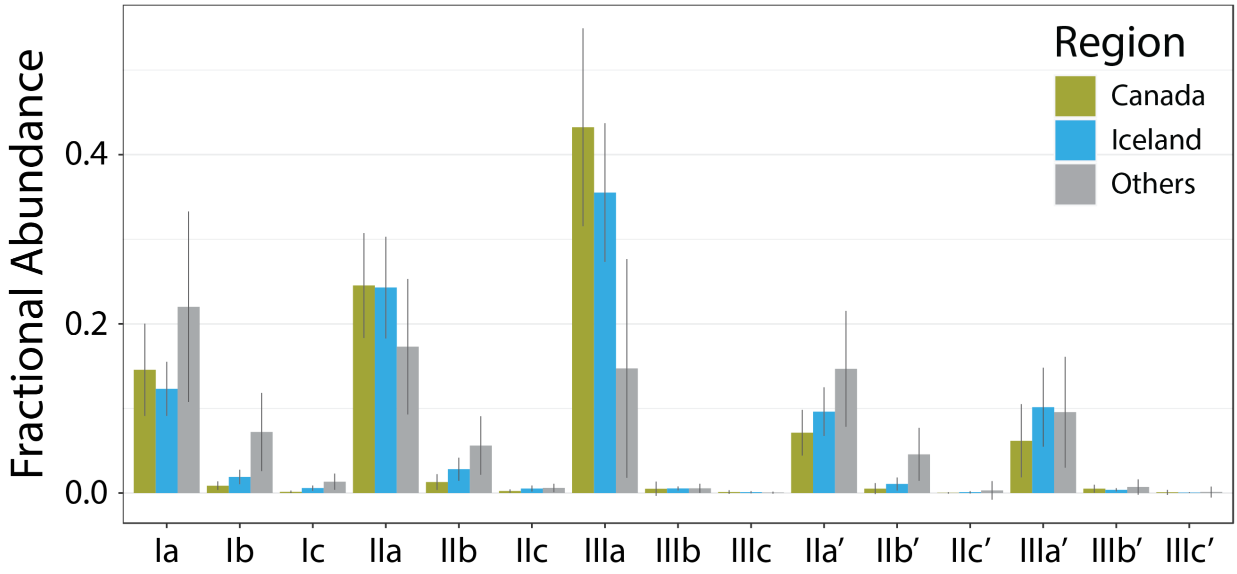


Figure 3-4: Average fractional abundances (calculated within the standard Full set) of new samples from Canada and Iceland as compared to the average of the other samples in this study (Others).

3.4.2 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. 3-5. 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.

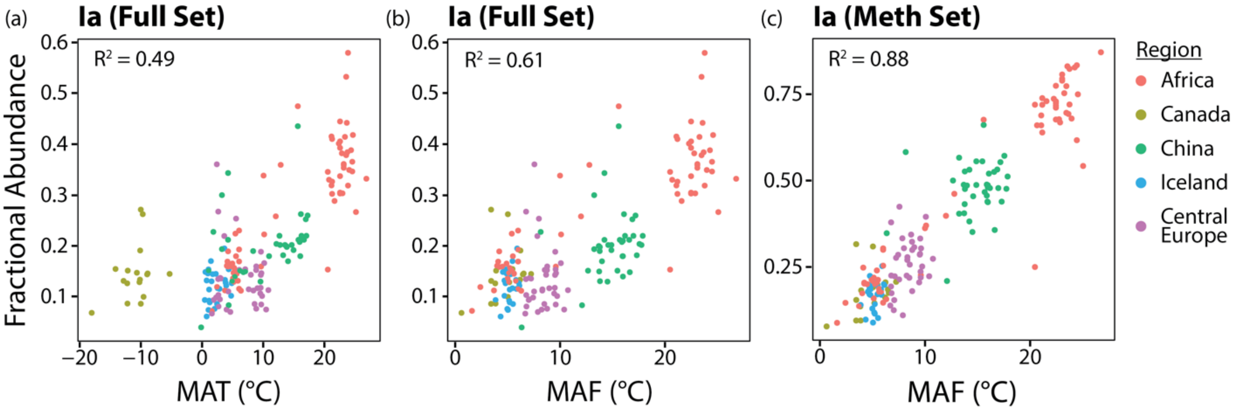


Figure 3-5: From left to right, the effects of substituting the Months Above Freezing (MAF) warm-season index for Mean Annual Temperature (MAT) and fIaMeth for fIaFull on the relationship between the fractional abundance of brGDGT-Ia and temperature.

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. 3.2.4). The correlation between these variables and the FA of one representative compound, fIa_{Meth}, is shown in Fig. 3-6. 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. 3.4.4.1).

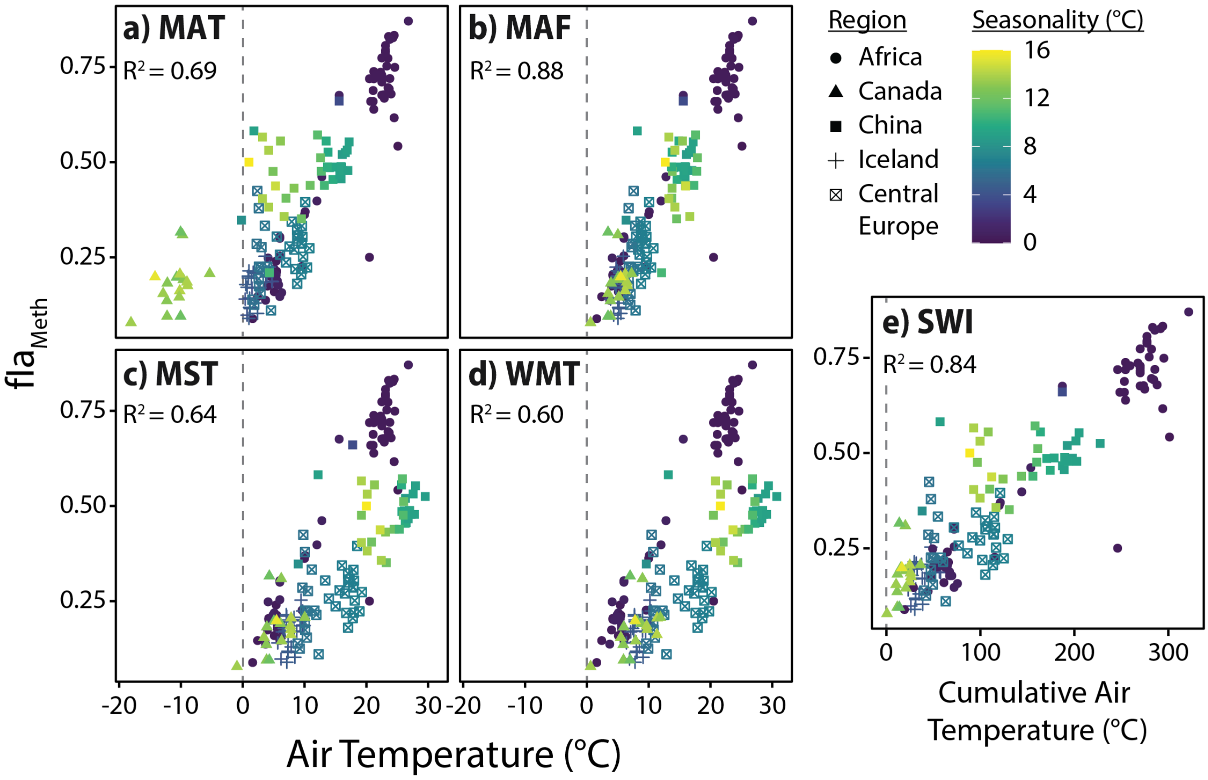


Figure 3-6: Relationships between fla_{Meth} *and the five air temperature variables tested in this study. A) mean annual temperature (MAT), b) months above freezing (MAF), c) mean summer temperature (MST), d) warmest month temperature (WMT), and e) summer warmth index (SWI). Colors indicate air temperature seasonality (standard deviation of mean monthly temperatures). Dashed lines indicate mean or cumulative temperatures of 0°C for reference.* R^2 *values for the relationship between* fla_{Meth} *and each temperature variable are provided in each subplot.*

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

Temperature and the Methylation set

The Methylation set provided the strongest relationships between brGDGT FAs and MAF (Fig. 3-7). This set both strengthened existing correlations (e.g. fIa_{Meth}, Fig. 3-7a) compared to the standard approach (Full set, R^2 values in parentheses in Fig. 3-7) and generated new ones (e.g. $fIIb_{Meth}$, Fig. 3-7e). Furthermore, many relationships between FAs and temperature were revealed to be qualitatively similar regardless of ring number. For example, the FAs of all tetramethylated compounds (fIa_{Meth}, fIb_{Meth}, fIc_{Meth}) had a strong positive linear relationship with temperature (Fig. 3-7a-c), while those of the 5-methyl pentamethylated compounds all had a noisier negative relationship (Fig. 3-7d-f). The hexamethylated FA trends were less clear in part due to their lower abundances, but all were negatively correlated with temperature and nonlinearities were apparent (Fig. 3-7g-l). These analogous trends show that the number of methylations responds similarly to temperature regardless of the number of cyclopentane rings. Within a paleoclimate lens, this observation opens up the possibility of independent temperature calibrations for un-, mono-, and bicyclized brGDGTs (Sect. 3.4.4.1). From a biological standpoint, it could imply that methylation and cyclization play their biological roles independently, as the former appears to vary more or less freely of the latter.

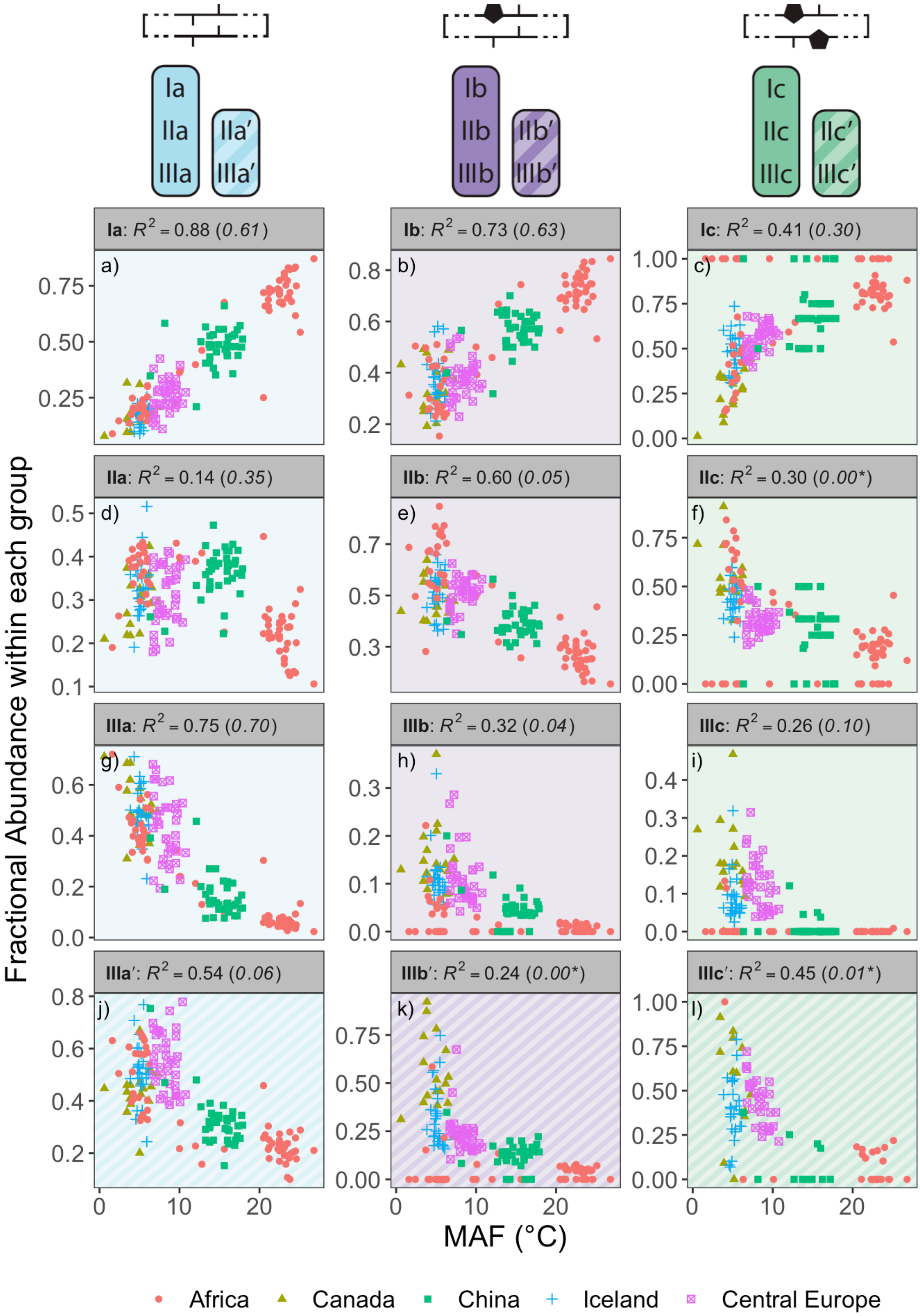

Ia
IIa
IIIa
IIa′
IIIa′
Ib
IIb
IIIb
IIb′
IIIb′
Ic
IIc
IIIc
IIc′
IIIc′
Ia: $R^2 = 0.88$ (0.61)
Ib: $R^2 = 0.73$ (0.63)
Ic: $R^2 = 0.41$ (0.30)
IIa: $R^2 = 0.14$ (0.35)
IIb: $R^2 = 0.60$ (0.05)
IIc: $R^2 = 0.30$ (0.00*)
IIIa: $R^2 = 0.75$ (0.70)
IIIb: $R^2 = 0.32$ (0.04)
IIIc: $R^2 = 0.26$ (0.10)
IIIa′: $R^2 = 0.54$ (0.06)
IIIb′: $R^2 = 0.24$ (0.00*)
IIIc′: $R^2 = 0.45$ (0.01*)
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
Fractional Abundance within each group
MAF (°C)
0
10
20
Africa
Canada
China
Iceland
Central Europe

Figure 3-7: Relationships between the average air temperature of months above freezing (MAF) and brGDGT FAs calculated within the Meth set. R^2 values are provided for each subplot, with R^2 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 IIc' are redundant because they exactly mirror those of IIIa', IIIb', and IIIc', respectively, and are therefore not shown.

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

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

3.4.3 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 et al., 2013; Russell et al., 2018; Dang et al., 2018) and were moderately correlated ($r = 0.75$) in this dataset (Table S2). This correlation is not surprising, as pH will affect conductivity both directly (through the concentration of H_3O^+/OH^-) and indirectly (by altering the solubility of various ions). Furthermore, both pH and conductivity can be affected by the same factors, such as catchment lithology and precipitation chemistry (cf. Wetzel, 2001). BrGDGT FAs in this study showed similar relationships with pH and conductivity, but also displayed potentially important differences. We therefore treat them separately in our analyses but discuss them together in this section.

The Isomer set and conductivity

Conductivity provided the strongest compound-specific correlations with brGDGT FAs after temperature (Figs. 3-8 and S23). Of the basic brGDGT sets (Meth, Cyc, and Isom), the Isom set had the highest statistical performance (Fig. 3-8). The FAs of all 6-methyl brGDGTs showed a positive linear correlation with conductivity in this set, with coefficients of determination as high as $R^2 = 0.70$ (Fig. 3-8d). Furthermore, this positive correlation was broadly independent of both methylation number and cyclization number, indicating that methylation position varies with conductivity irrespective of other structural properties. The 6-methyl brGDGTs were also all positively correlated with pH, but more weakly so ($R^2 \leq 0.54$, Fig. S28). A relationship between brGDGT isomers and pH in lake sediments has been previously observed and quantified by the isomerization of branched tetraethers (IBT, Eq. A14; Ding et al., 2015) and the Isomer Ratio of 6-methyl isomers (IR_{6Me}, Eq. A8; Dang et al., 2016). However, these indices were more 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. These results indicate that isomer abundances are primarily dependent on conductivity, but have some relation to pH as well. Temperature correlations were also present in the Isom subset ($R^2 \leq 0.57$ with MAF, Fig. S14). Further study is needed to elucidate whether temperature acts as a secondary control on isomer abundances or whether these 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 ($R^2 \leq 0.38$, Fig. S35).

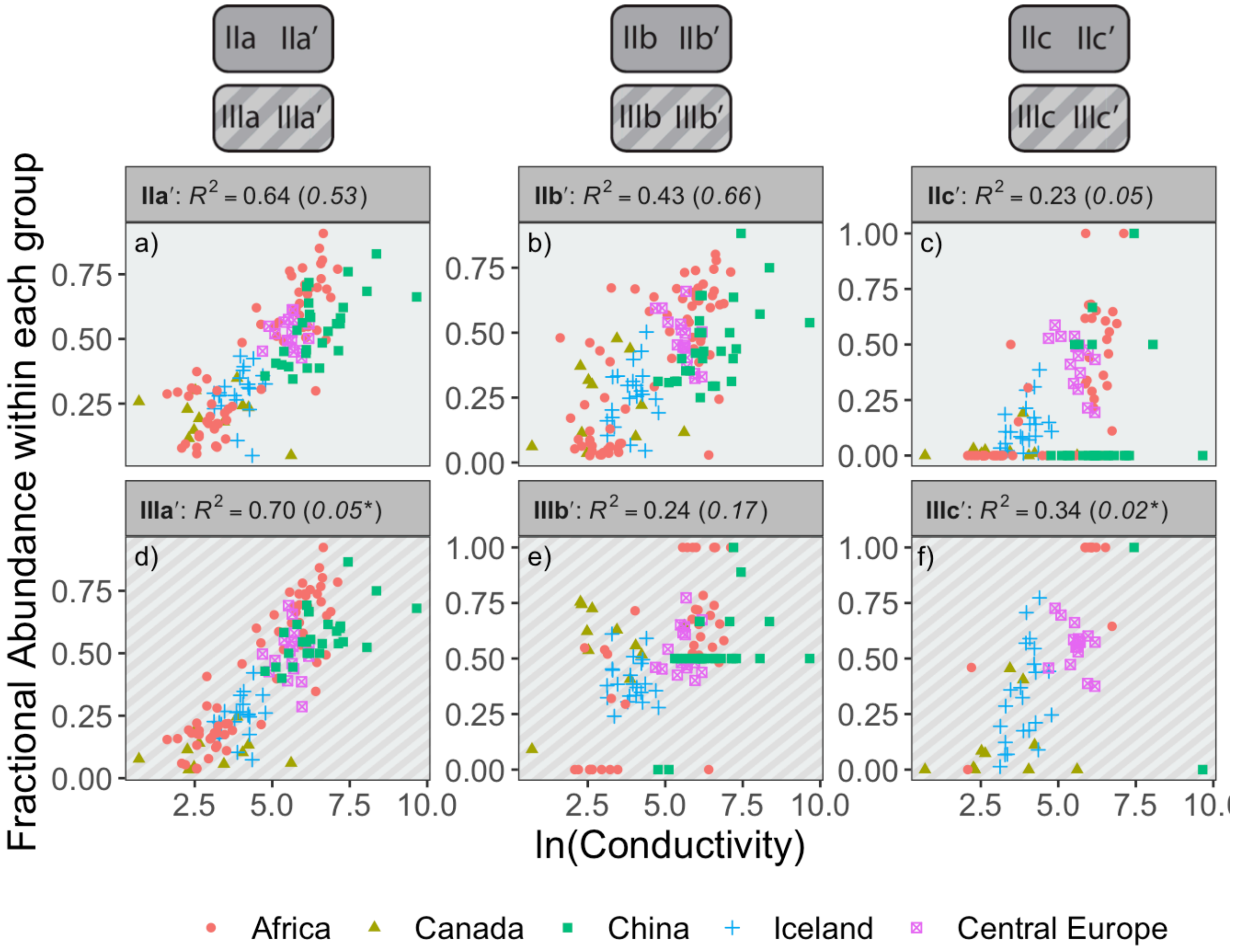


Figure 3-8: Relationships between the natural logarithm of conductivity and brGDGT FAs calculated within the Isom set. R^2 values are provided for each subplot, with R^2 values for the standard Full FAs given in parentheses for comparison. P-values were < 0.01 except where marked with an asterisk. Note 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.

The Cyclization set and pH

The Cyclization set highlighted the relationship between brGDGTs and pH. As pH increased, all cyclized brGDGTs were found in greater relative abundance (Fig. 3-9). This result reinforces previous observations that higher ring number is associated with higher pH, as has been quantified by the CBT (Weijers et al., 2007b), $\#rings_{tetra}$ (Sinninghe Damsté, 2016), Degree of Cyclization (DC; Baxter et al., 2019), and related indices (see Appendix (Section 8.2.1)). However, this pH relationship has thus far been demonstrated primarily in soils; the single-compound FAs of the Cyc subset already provide the strongest pH correlations (R^2 = 0.61, Fig. 3-9b) yet reported in lake sediments. The Cyc set also reveals that compounds with structural similarities exhibit analogous responses to pH. For example, all monocyclized brGDGTs show a nonlinear increase with pH (Fig. 3-9b, e, h, k, n), while uncyclized brGDGTs all exhibit a nonlinear decrease (Fig. 3-9a, d, g, j, m). These trends are apparent regardless of methylation number or methylation position, suggesting that ring number is broadly independent of both. This independence may imply that alkyl-chain cyclization serves its biological function(s) regardless of the number and position of methylations present. It also allows for the construction of independent pH calibrations for tetra-, penta-, and hexamethylated brGDGTs (Sect. 3.4.4.2).

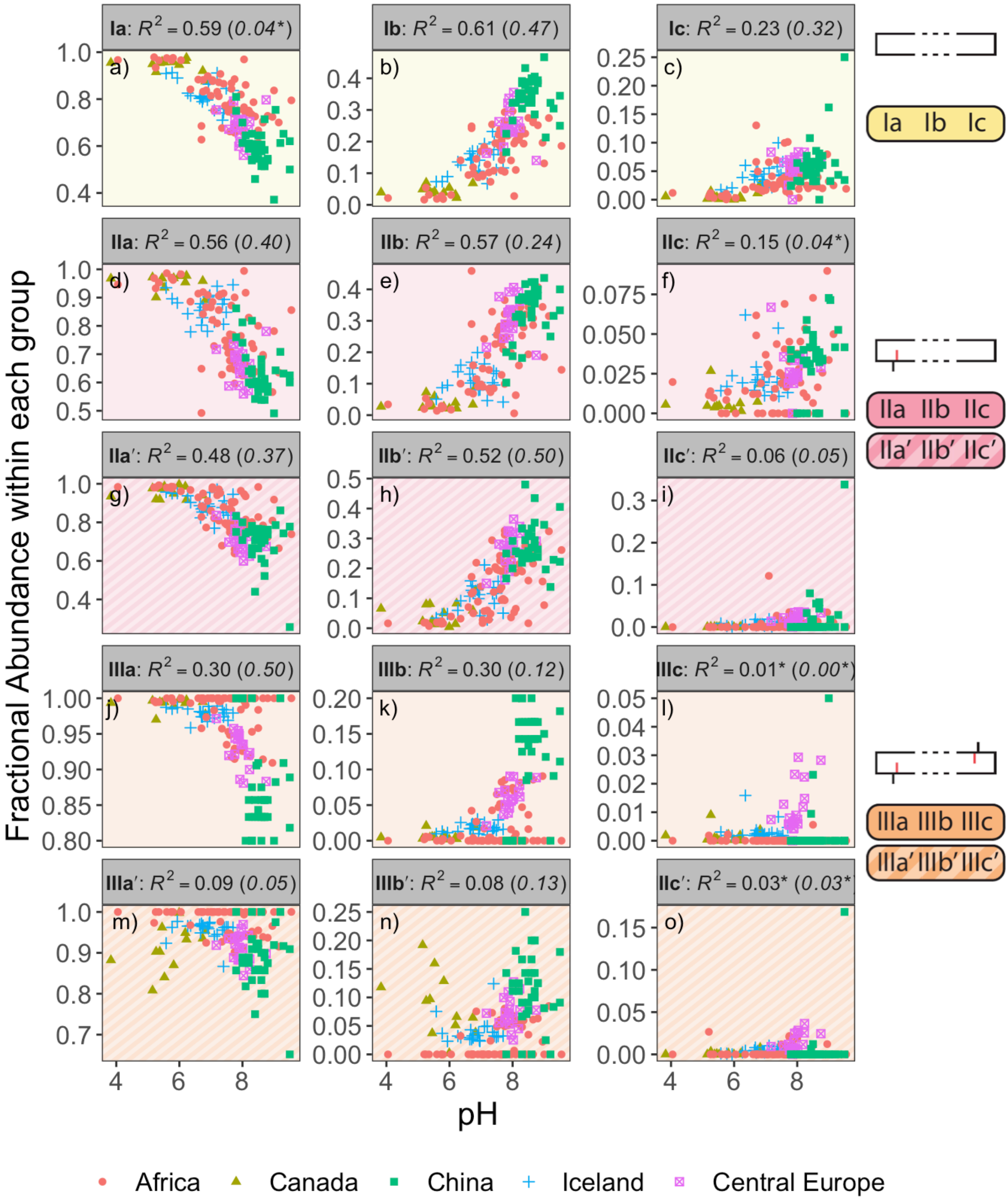


Figure 3-9: Relationships between pH and brGDGT FAs calculated within the Cyc set. R^2 values are provided for each subplot, with R^2 values for the standard Full FAs given in parentheses for comparison. P-values were < 0.01 except where marked with an asterisk.

Though the Cyc set FAs were most strongly correlated with pH ($R^2 \leq 0.61$), they also exhibited robust relationships with conductivity ($R^2 \leq 0.57$; Fig. S20). All cyclized compounds showed positive correlations with conductivity, and this increase was largely independent of methylation number or position. These results indicate that ring number is primarily dependent on pH, but is correlated with conductivity in a similar manner. The brGDGT indices showed an analogous result; all cyclization indices (CBT and related indices, $\#rings_{tetra}$ and related indices, and DC; see Appendix (Section 8.2.1)) correlated most strongly with pH, but also exhibited weaker relationships with conductivity. The Cyc set exhibited little to no correlation with either temperature ($R^2 \leq 0.25$) or DO ($R^2 \leq 0.08$), suggesting that neither of these environmental variables plays an important role in controlling brGDGT cyclization.

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

Though the strongest conductivity trends were displayed by the Isom set ($R^2 \leq 0.70$), correlations were also present in the Cyc FAs ($R^2 \leq 0.57$, Fig. S20). Similarly, the Cyc set contained the highest pH dependencies ($R^2 \leq 0.61$), but notable relationships were visible in the Isom set as well ($R^2 \leq 0.54$, Fig. S28). To take advantage of all of these conductivity and pH relationships, we therefore used the combined Cyc-Isom set, which holds only methylation number constant while allowing both cyclization number and methylation position to vary (Fig. 3-2). Both conductivity and pH trends were strengthened in this combination set ($R^2 \leq 0.73$ and 0.62, Fig. S23 and S30, respectively), especially for the uncyclized compounds. However, temperature correlations were also increased in the CI set ($R^2 \leq 0.60$, Fig. S16), a potentially convoluting influence that may not be desired.

3.4.4 Calibrations

For each set, combined set, and subset defined in Sect. 3.3, we performed linear and quadratic regressions against temperature, conductivity, pH, dissolved oxygen, and lake geometry variables using SFS/SBE and combinatoric fitting methods. We found temperature and conductivity to provide the strongest empirical calibrations with brGDGTs, followed by pH and dissolved oxygen, and discuss our recommended calibrations below.

Temperature Calibrations

For each structural set and subset defined in Section 3.3, we performed regressions of subset-specific brGDGT FAs against five temperature variables to generate multiple global-scale calibrations. Of the temperature indices that we tested, MAF provided the fit with the highest statistical significance ($R^2 = 0.91$), followed closely by MST ($R^2 = 0.90$), SWI ($R^2 = 0.89$), and WMT ($R^2 = 0.88$). MAT provided calibrations with high statistical performance as well ($R^2 \leq 0.87$). However, these fits showed clear seasonality biases in their residuals, resulting in substantial over-estimations of MAT for cold sites (Fig. S40). Furthermore, they often relied heavily on low-abundance compounds as fitting variables (IIc, IIb', IIIc, IIIc', IIIb, and IIIb'). We therefore do not recommend a brGDGT MAT calibration and focus our discussion on the warm-season temperature indices, especially MAF. We note, however, that MAF is equal to MAT in warmer climates or lower latitudes and that MAF calibrations may be interpreted as reconstructing MAT in these cases.

Methylation number was the single most important structural variable for temperature calibrations. The Meth set, which allowed only methylation number to vary, provided a MAF calibration ($R^2 = 0.90$, Fig. 3-10b) that was on par with other recent global and regional lake sediment calibrations ($R^2 = 0.85$ to 0.94; Dang et al., 2018; Russell et al., 2018; Martínez-Sosa et al., 2020b). The MI, MC, and Full sets, which additionally allowed for changes in cyclization number and/or methylation position, added little to the calibration performance ($R^2 = 0.90$ to 0.91, Fig. 3-10a). Furthermore, sets which held methylation number constant – Cyc, Isom, and CI – performed markedly worse ($R^2 = 0.51$, 0.63, and 0.67, respectively, Fig. 3-10a). Analogous results were found using RMSE to compare fit performance (Fig. S41a). These results indicate that effectively all of the temperature dependence of the 15 commonly measured brGDGTs is captured by methylation number alone.

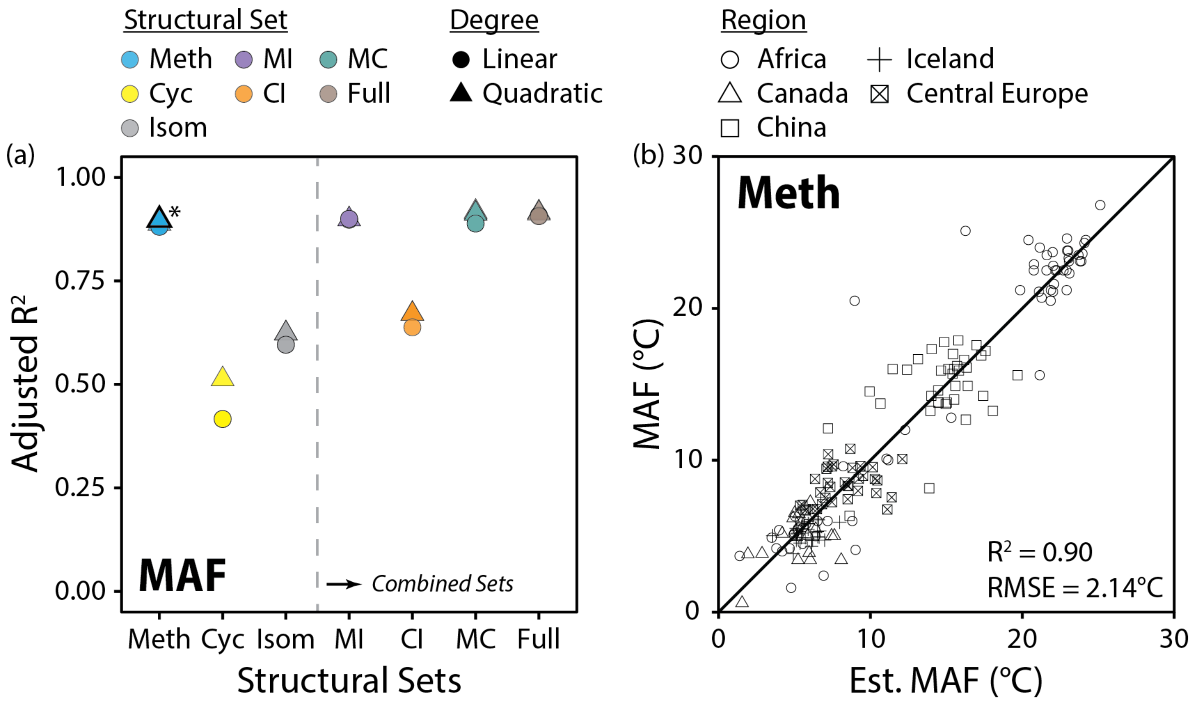


Figure 3-10: Summary of temperature calibrations. A) Performance (adjusted R^2) of all linear and quadratic fits for the mean air temperature of months above freezing (MAF) and brGDGT fractional abundances (FAs) calculated within the basic (Meth, Cyc, Isom; left of dashed line) and combined (Meth-Isom (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. 3-10) is bolded and marked with an asterisk in a) and plotted in b). "Est. MAF" is the MAF temperature estimated using this suggested fit.

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

Of the brGDGT temperature indices that we tested (see Appendix (Section 8.2.1)), MBT'_{5Me} performed best. This index 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 $R^2 = 0.70$; WMT $R^2 = 0.66$). The slope and intercept of the MAF/MBT'_{5Me} calibration (MAF = -0.5 (± 0.4) + 30.4 (± 0.8) * MBT'_{5Me}) were similar to the MAT/MBT'_{5Me} calibration presented by Russell et al. (2018) (MAT = -1.21 + 32.42 * MBT'_{5Me}). This may suggest that MBT'_{5Me}-derived temperatures using the Russell et al. (2018) calibration in cold regions are best considered to reconstruct MAF rather than MAT. Though MBT'_{6Me} was previously found to correlate well with temperature in highly alkaline lakes on a regional scale ($R^2 = 0.75$; Dang et al., 2018), it was not correlated with any temperature variable in our global dataset ($R^2 \leq 0.12$). Finally, we note that the Community Index (De Jonge et al., 2019), which was associated with bacterial community changes in geothermally heated Icelandic soils, is identical to our fIa_{Meth}. This may suggest that the strong connection between fIa_{Meth} and MAF could also be driven at least in part by changes in microbial community composition. However, genomic data is not currently available for the majority of the sites in this study to test this hypothesis.

MAF provided the strongest temperature calibrations at the global level in this study. Unlike for MAT, these calibrations yielded unbiased residuals at the cold end (Fig. 3-10b), encouraging their application in cold regions with high seasonality in particular. However, we note that the range of MAF values in our combined Canadian and Icelandic dataset is small (6.6°C, or 3.2°C without the warm and cold endmembers) compared to that of MAT (22.9°C), especially in relation to the MAF calibration error (RMSE = 2.14°C, Fig. 3-10b). While the global-scale MAF calibrations presented here are valid for application at high latitudes, they may therefore prove unable to resolve smaller temperature variations in paleoclimate archives from those regions. Regional high-latitude calibrations aimed at improving this temperature resolution are the subject of ongoing work.

Conductivity and pH Calibrations

Conductivity outperformed pH in our calibrations ($R^2 = 0.83$ versus 0.74) and was the second most important predictor of brGDGT distributions in our dataset after temperature. Both cyclization number and methylation position were important to the success of these calibrations. The Cyc set, which allowed only cyclization number to change, provided a conductivity fit with $R^2 = 0.73$ (Fig. 3-11a). The Isom set, which isolated trends in isomer abundances, generated a slightly stronger calibration with $R^2 = 0.76$. When both of these structural properties were allowed to vary together in the CI set, the calibration was markedly improved ($R^2 = 0.83$, Fig. 3-11b). In contrast, methylation number was a poorer predictor of conductivity alone (Meth set $R^2 = 0.65$) and did not improve upon the CI correlation in the 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 fit performance (Fig. S41b). The CI_{III} and CI_{II} subsets also provided conductivity calibrations with relatively high statistical performance ($R^2 = 0.75$, 0.73). The CI_I subset performed worse ($R^2 = 0.65$), likely due to the fact that no isomer variations are present in these FAs. These subset-specific fits are the first of their kind, and their success indicates that the relationship between brGDGTs and conductivity is present regardless of methylation number.

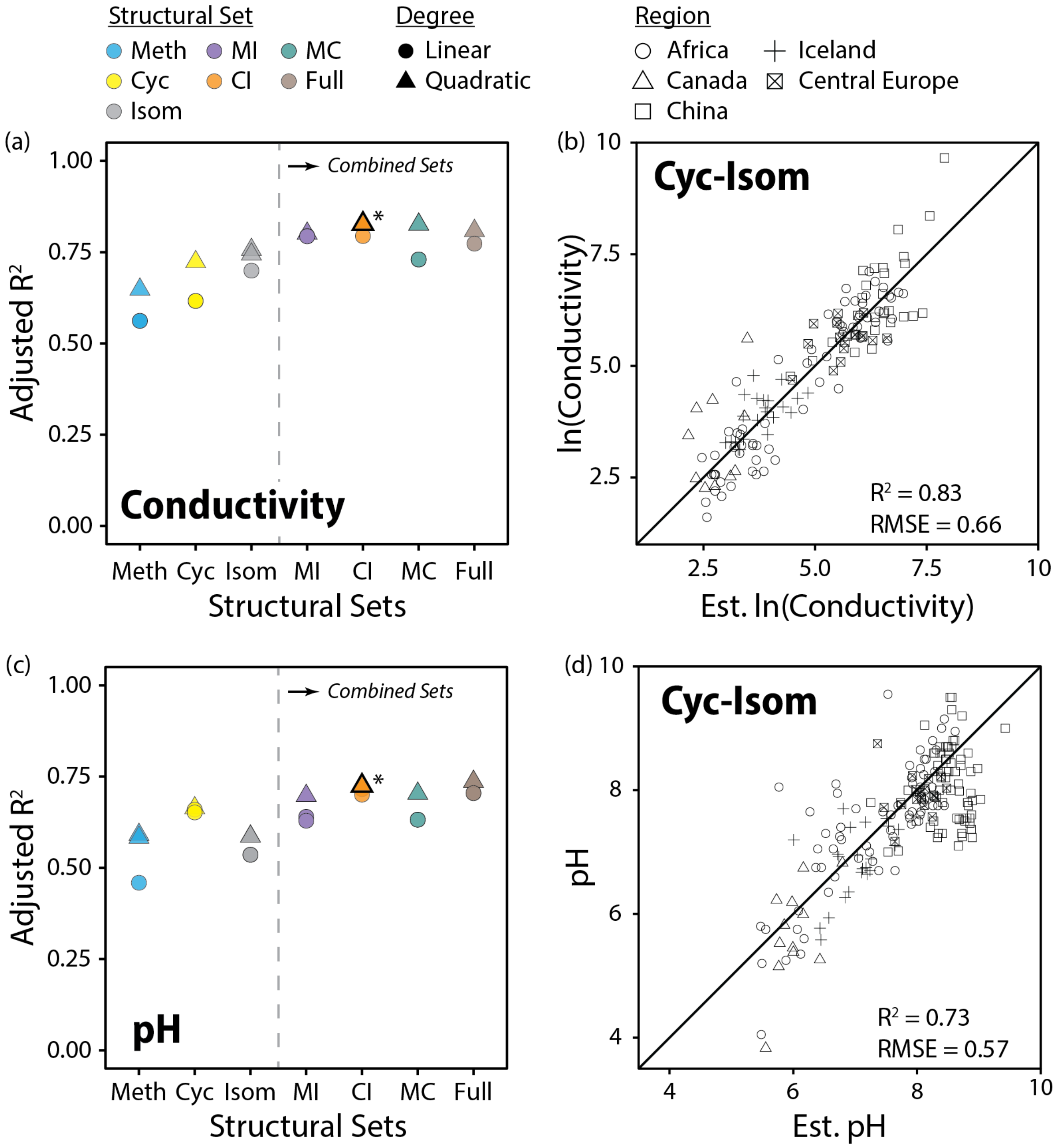


Figure 3-11: Summary of pH and conductivity calibrations. Performance (adjusted R^2) of all linear and quadratic fits for lake water conductivity (a) and pH (c) and brGDGT fractional abundances (FAs) calculated within the basic (Meth, Cyc, Isom; left of dashed lines) and combined (Meth-Isom (MI), Cyc-Isom (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. 3-12 and 3-13) is bolded and marked with an asterisk in a) and c) and plotted in b) and d). "Est. ln(Conductivity)" and "Est. pH" are the natural logarithm of lake water conductivity and pH estimated using these suggested fits.

In addition to conductivity, the CI set also best captured the relationship between brGDGTs and pH ($R^2 = 0.73$; Fig. 3-11c and d). The addition of methylation number variations in the Full set did not substantially improve the calibration ($R^2 = 0.74$), indicating that the majority of the relationship between brGDGTs and pH is captured by cyclization number and methylation position. Of these two structural variables, cyclization number was more important for pH; the Cyc set calibration ($R^2 = 0.67$) outperformed those from the Isom and Meth sets ($R^2 = 0.59$ for both). Analogous results were found using RMSE (Fig. S41c). The CI_I, CI_{II}, and CI_{III} subsets provided weaker, but significant calibrations with comparable performances to one another ($R^2 = 0.67$, 0.68, and 0.62, respectively), indicating again that pH relationships are more or less independent of methylation number.

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_{mean} and DO_{min}). 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).

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

Recommended Calibrations

The Meth, MC, MI, and Full subsets all provided MAF temperature calibrations with comparable R^2 (0.90 to 0.91) and RMSE (1.97 to 2.14°C) values (Fig. 3-10a, Table S3). However, the Meth set provided the FAs with the strongest compound-specific relationships with MAF ($R^2 \leq 0.88$) in the modern dataset and exhibited little influence ($R^2 \leq 0.49$) 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. 3-11a and c). We therefore recommend the highest-performing Meth calibration (Fig. 3-10b; Eq. 3-10; n = 182, $R^2 = 0.90$, RMSE = 2.14°C) for general use in lake sediments:

$$
\begin{aligned}
MAF\ (°C) = {} & 92.9(\pm 15.98) + 63.84(\pm 15.58) \times fIb_{Meth}^{2} - 130.51(\pm 30.73) \times fIb_{Meth} \\
& - 28.77(\pm 5.44) \times fIIa_{Meth}^{2} - 72.28(\pm 17.38) \times fIIb_{Meth}^{2} \\
& - 5.88(\pm 1.36) \times fIIc_{Meth}^{2} + 20.89(\pm 7.69) \times fIIIa_{Meth}^{2} \\
& - 40.54(\pm 5.89) \times fIIIa_{Meth} - 80.47(\pm 19.19) \times fIIIb_{Meth}
\end{aligned}
\tag{3-10}
$$

Importantly, Eq. (3-10) contains terms for $fIIc_{Meth}$ and $fIIIb_{Meth}$, which may be equal to zero in organic-lean samples (e.g. Fig. 3-7f and h) and could thus result in unrealistic values for MAF. In these cases, we recommend applying subset-specific calibrations that exclude bicyclized (Eq. S1; n = 182, $R^2 = 0.89$, RMSE = 2.19°C) or both mono- and bicyclized (Eq. S2; n = 182, $R^2 = 0.88$, RMSE = 2.33°C) compounds. These calibrations offer flexibility in downcore applications with minimal sacrifice of statistical performance. Subset-specific calibrations are also available for mono- and bicyclized compounds alone (Eq. S3-4).

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. (3-11) (n = 182, $R^2 = 0.91$, RMSE = 1.97°C):

$$
\begin{aligned}
MAF\ (°C) = {} & -8.06(\pm 1.56) + 37.52(\pm 2.35) \times fIa_{Full} - 266.83(\pm 98.61) \times fIb_{Full}^{2} \\
& + 133.42(\pm 19.51) \times fIb_{Full} + 100.85(\pm 9.27) \times fIIa'^{2}_{Full} \\
& + 58.15(\pm 10.09) \times fIIIa'^{2}_{Full} + 12.79(\pm 2.89) \times fIIIa_{Full}
\end{aligned}
\tag{3-11}
$$

The calibrations presented in Eqs. (3-10), (3-11), and (S1-4) allow for the quantitative reconstruction of warm-season 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 3.4.3), 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. 3-11b, 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 ($R^2 \leq 0.70$, 0.77, and 0.75, respectively; Figs S18 , S17 and S15). We therefore recommend the top CI fit for use (Eq. 3-12; n = 143, $R^2 = 0.83$, RMSE = 0.66), and provide subset-specific calibrations for $\mathrm{CI_{I}}$, $\mathrm{CI_{II}}$, and $\mathrm{CI_{III}}$ in the Supplement (Eq. S5-7):

$$\begin{aligned} ln(Cond.) = {} & 6.62(\pm 1.01) + 8.87(\pm 1.24) \times fIb_{CI} + 5.12(\pm 1.54) \times fIIa'^{2}_{CI} \\ & + 10.64(\pm 1.88) \times fIIa^{2}_{CI} - 8.59(\pm 2.21) \times fIIa_{CI} \\ & - 4.32(\pm 1.46) \times fIIIa'^{2}_{CI} - 5.31(\pm 0.95) \times fIIIa^{2}_{CI} \\ & - 142.67(\pm 36.08) \times fIIIb^{2}_{CI} \qquad (3-12) \end{aligned}$$

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. (3-12) 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. 3-11c, Table S3). Given the stronger temperature dependencies of the Full set, we recommend the CI calibration (Fig. 3-11d; Eq. 3-13; n = 154, $R^2 = 0.73$, RMSE = 0.57) for use and provide its subset-specific calibrations ($\mathrm{CI_{I}}$, $\mathrm{CI_{II}}$, and $\mathrm{CI_{III}}$) in the Supplement (Eq. S8-10):

$$pH = 8.93(\pm 0.21) - 3.84(\pm 0.25) \times fIa^{2}_{CI} + 2.63(\pm 0.35) \times fIIa'_{CI} \qquad (3-13)$$

The residuals of the fit in Eq. (3-13) have a weak but significant correlation with pH ($R^2 = 0.17$), causing it to overestimate pH for acidic samples and underestimate it for alkaline ones. A similar bias in pH calibrations has been previously observed in soils (De Jonge et al., 2014a). We therefore caution the use of this calibration in acidic ($pH < 5$) or alkaline ($pH > 9$) conditions.

Previous work has demonstrated the value of brGDGT-derived pH records in studies of terrestrial paleoclimate (Tyler et al., 2010; Cao et al., 2017; Fastovich et al., 2020). However, these studies relied on calibrations generated from soils and/or analyses in which the 5- and 6-methyl isomers were not separated. The fit presented in Eq. (3-13) may improve such studies by providing a globally-distributed pH calibration in lake sediments using the latest chromatographic methods which improves upon the error of previously available calibrations (RMSE = 0.80, Russell et al., 2018).

The structural set calibrations for MAF, conductivity, and pH presented above offer potential advantages over traditional calibrations constructed using the standard Full set. The structural set FAs separate the effects of temperature from pH and conductivity in the modern dataset. The application of calibrations such as Eqs. (3-10), (3-12), and (3-13) could therefore provide some protection against the unwanted influences of other environmental variables downcore. Importantly, this advantage is gained without sacrificing the statistical performance of the calibration. Furthermore, the subset-specific calibrations such as Eqs.

(S1) and (S2) use only the most abundant compounds, making it easier to apply brGDGT proxies to organic-lean samples where only some compounds are typically present above the detection limit.

3.5 Conclusions

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

The structural sets also provided insight into the biological underpinnings of brGDGT structural diversity. The Meth, Cyc, and Isom sets gave evidence that methylation number, cyclization number, and methylation position vary more or less independently of one another across environmental gradients. They further revealed that the inclusion of tetramethylated compounds (Ia, Ib, Ic) enhances the temperature dependencies of 5-methyl compounds but erases those of their 6-methyl counterparts. Finally, the results linking methylation number to temperature and cyclization number to pH are consistent with a physiological explanation of brGDGT distributions (Weijers et al., 2007b). 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.

Warm-season temperatures outperformed MAT as the most important predictors of brGDGT distributions, particularly in high-latitude environments. We introduced 43 new lake sediment samples from sites with low MAT and high seasonality. In conjunction with a global dataset, these samples showed a clear warm-season bias in brGDGT temperature relationships, with MAF providing the strongest fits in our dataset. The warm-season bias may suggest a direct or indirect connection to heightened primary productivity in the summer. Alternatively, it may be the result of a more complex relationship with dynamic lake processes such as mixing events. While further study is needed to unravel these complications, the strong empirical calibrations presented here support the use of the brGDGT paleotemperature proxy to quantitatively reconstruct warm-season air temperatures from high-latitude lake sediments.

In summary, the use of brGDGT structural sets and warm-season temperature indices deconvolved relationships between brGDGT structure and environmental gradients, revealed trends with biological implications, and tied brGDGT distributions to warm-season temperatures. Furthermore, they allowed for the construction of temperature and pH calibrations that rely on a smaller number of compounds with clearer modern trends compared to the traditional approach, without sacrificing statistical performance. Finally, we used the new structural set framework to generate the first brGDGT conductivity calibration in a global lake sediment dataset. Our results thus allow 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.

3.6 Acknowledgements

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

4 CHAPTER IV: Near-universal trends in brGDGT distributions in nature

Abstract. Bacterial brGDGT lipids are a prevalent tool in studies of terrestrial paleoclimate. Their distributions correlate empirically with environmental temperature and pH and their ubiquity in terrestrial, freshwater, and marine environments gives them wide applicability. Whether correlations with temperature and pH emerge due to a physiological response of source organisms and/or a shift in bacterial community composition remains an open question with important implications for proxy development and application. We applied a newly described technique for grouping brGDGTs to a globally compiled dataset (n = 2535) consisting of all modern sample media known to host brGDGTs. We found strong resemblances in the relationships between brGDGT fractional abundances and both temperature and pH across nearly all sample types examined. We also found near-universal connections between the brGDGTs themselves. Given the markedly different bacterial communities expected to inhabit these settings, these widespread relationships may suggest physiological and/or biochemical bases for observed brGDGT distributions.[2]

4.1 Introduction

Lipid biomarkers preserved in sedimentary archives are an invaluable tool in studies of Earth's past climates. Of these, bacterial branched glycerol dialkyl glycerol tetraethers (brGDGTs) have quickly become one of the most popular proxies for reconstructing temperature and pH, in part because they are nearly ubiquitous in the natural world. BrGDGTs have been measured in over a dozen different sample media from the High Arctic (Kusch et al., 2019b) to the deepest ocean trenches (Xiao et al., 2020), including in soil (Weijers et al., 2006), peat (Sinninghe Damsté et al., 2000), groundwater (Ding et al., 2018), hot springs (Schouten et al., 2007), aeolian particulate matter (Fietz et al., 2013), marine hydrothermal (Lincoln et al., 2013) and methane cold seep (Zhang et al., 2020) carbonate, fossil bone (Dillon et al., 2018), speleothem (Yang et al., 2011), and sediment and suspended/settling particulate matter (SPM) from lacustrine, riverine, and marine settings (Hopmans et al., 2004; Blaga et al., 2009; Tierney and Russell, 2009; De Jonge et al., 2014b). This ubiquity, along with a preservation window of at least tens of millions of years (Super et al., 2018), makes brGDGTs applicable in a wide variety of sedimentary archives over geologic timescales, even when other proxies may be absent (Weijers et al., 2007a; Dearing Crampton-Flood et al., 2018; Naafs et al., 2018; Crump et al., 2019; Lu et al., 2019; Harning et al., 2020). However, longstanding questions exist in brGDGT research that complicate their use as reliable, quantitative paleoclimate proxies.

The origin of the connection between brGDGT distributions and environmental parameters is one such pertinent question. Drawing analogies to other bacterial membrane lipid classes, early studies posited that the empirical relationships between brGDGTs and environmental parameters were physiological adaptations of source organisms to variations in external

---

[2] A version of this chapter is in review:
Raberg, J. H., Miller, G. H., Geirsdóttir, Á., and Sepúlveda, J.: Near-universal trends in brGDGT distributions in nature, Science Advances, 2021, *in review.*

temperature and pH (Weijers et al., 2007b). Subsequent studies have since shown that changes in brGDGT distributions are also accompanied by shifts in the composition of the associated microbial community, indicating that the empirical correlations between the lipids and environmental parameters may be indirect in nature (Weber et al., 2018; De Jonge et al., 2019; Van Bree et al., 2020; Wu et al., 2021; De Jonge et al., 2021). Whether the variation in a brGDGT paleo record is due to the direct physiological response of microbes to a changing environment or the indirect restructuring of a microbial community has important implications for its interpretation and fidelity (De Jonge et al., 2021). Unfortunately, brGDGT-producing source organisms remain incompletely identified and cultured (Sinninghe Damsté et al., 2018), limiting our ability to test these hypotheses through controlled laboratory experimentation. However, some insight can be gained by other means. Mesocosm studies of both incubated soils (Chen et al., 2018) and lake waters (Martínez-Sosa et al., 2020a) have suggested a bacterial community influence. Molecular dynamics simulations have recently produced a split result, supporting the physiological explanation for the brGDGT paleothermometer but leaving open the possibility of a community-driven pH response (Naafs et al., 2021). Here, we take an alternate approach in addressing this long-standing question by compiling the large and diverse body of existing lipid and environmental data and revisiting the empirical trends that exist in nature.

Early studies of brGDGTs in soils found that the number of methyl groups and cyclopentane rings on the alkyl backbone were empirically correlated with temperature and pH, respectively (Weijers et al., 2007b). Subsequent studies have generally upheld these basic relationships but have also found important differences in their details across sample types and regions. These discrepancies have led to independent modern calibrations for different sample types (e.g. soils and peats (Dearing Crampton-Flood et al., 2020), lake sediments (Martínez-Sosa et al., 2021) or SPM (Zhao et al., 2021), bones (Zhao et al., 2020), speleothems (Blyth and Schouten, 2013), and marine sediments (Dearing Crampton-Flood et al., 2018)) and regions (e.g. East Africa (Russell et al., 2018) and China (Dang et al., 2018)) and have necessitated much effort to disentangle the allochthonous versus autochthonous sources of brGDGTs to sedimentary archives (Buckles et al., 2014a; Loomis et al., 2014b; De Jonge et al., 2015b; Sinninghe Damsté, 2016; Dearing Crampton-Flood et al., 2018; Guo et al., 2020; Xiao et al., 2020; Dearing Crampton-Flood et al., 2021). Further complicating the matter, brGDGT distributions are affected by a wide array of other environmental parameters, including oxygen levels (Xie et al., 2014; Liu et al., 2014; Colcord et al., 2017; Weber et al., 2018; Van Bree et al., 2020; Yao et al., 2020; Dugerdil et al., 2021; Lattaud et al., 2021; Wu et al., 2021), salinity/electrical conductivity (Raberg et al., 2021; Wang et al., 2021), seasonality (Kim et al., 2012; Loomis et al., 2014b; Van Bree et al., 2020; Dearing Crampton-Flood et al., 2020), nutrient availability (Loomis et al., 2014a; Hu et al., 2016), and soil chemistry (De Jonge et al., 2021), which can mask or override relationships with temperature or pH. Finally, an observed bias in brGDGT-derived temperatures towards warmer seasons has proven difficult to quantify, with various studies finding summer air temperature (Shanahan et al., 2013), the mean air temperature of months above freezing (MAF) (Dearing Crampton-Flood et al., 2020; Raberg et al., 2021; Martínez-Sosa et al., 2021), growing degree days above freezing (Naafs et al., 2017), or other temperature indices to provide the strongest correlations in modern training sets. Such discrepancies might be expected as these air temperature indices are themselves only proxies for the *in situ* temperatures that are relevant for brGDGT-producing microbes. A growing body of *in situ* temperature studies (e.g., in lakes (Zhao et al., 2021) and soils (Wang et al., 2020; Pérez-Angel et al., 2020)) show promise, but the vast majority of published works currently rely on the more readily available air temperatures.

Importantly, brGDGT studies have traditionally relied on fractional abundances (FAs) calculated by dividing the compound of interest by the sum of all 15 commonly measured brGDGTs ("Full set FAs"; Fig. 8-16g; Table 8-2). By mathematically mixing these compounds in the denominator, this approach risks convoluting the influences of multiple environmental parameters on lipid distributions and obscuring individual compounds' trends (Raberg et al., 2021). We recently proposed an alternative approach in which brGDGT FAs are calculated within structurally-based brGDGT subgroups (Fig. 8-16; Table 8-2) and showed that these groupings deconvolved relationships between the lipids and temperature, pH, and conductivity in a global lake sediment dataset (Raberg et al., 2021). This study revealed that variations in methylation number, as isolated by the Methylation (Meth) structural set (Fig. 8-16a), captured the entirety of the brGDGT temperature response. Similarly, the pH response was fully captured by variations in cyclization number and isomer abundance, as quantified by the Cyclization-Isomer (Cyc-Isom) set (Fig. 8-16e). Here, we apply this new methodology to a compiled global dataset (n = 2535) of all available modern sample media known to contain brGDGTs to assess the uniformity of the relationships between compound FAs and temperature/pH across sample types and environments. We find near-universal relationships between brGDGT FAs, temperature, and pH across this diverse global dataset and discuss the implications of our findings for developing global brGDGT-based proxies that are independent of sample type, building confidence in downcore applications, and advancing our understanding of the origin of the connection between brGDGT lipid distributions and the environment.

4.2 Results

Recent work showed that the Meth structural set (Fig. 8-16a; Table 8-2) best captured the relationship between brGDGT FAs and temperature in a dataset of globally distributed lake sediments (Raberg et al., 2021). We find this result to be upheld for soils, peats, lacustrine SPM, bones, and marine surface sediments (Fig. 4-1). These sample types do not always share the same temperature variable (e.g., air temperature for soils versus sea surface temperature for marine sediments), due primarily to inherent differences between sample types and data availability constraints (see Methods for details). Despite these differences, all 15 compound-specific temperature relationships of these disparate sample types shared visibly similar functional forms (Fig. 4-1), with linear or quadratic fitting coefficients often within one standard error (Fig. 8-17). Limited datasets of lacustrine mesocosm, low dissolved oxygen (DO) SPM (separated due to the marked influence of oxygen levels on brGDGT distributions), and groundwater samples broadly plot within global trends. Notable deviations are visible for low DO SPM (e.g. Fig. 4-1e), groundwater (e.g. Fig. 4-1g), and marine samples (e.g. Fig. 4-1j). The Meth set FAs unanimously provided stronger correlations than the standard Full set FAs (Fig. 8-16g; Table 8-2) as well as better agreement between sample types (Fig. 8-18).

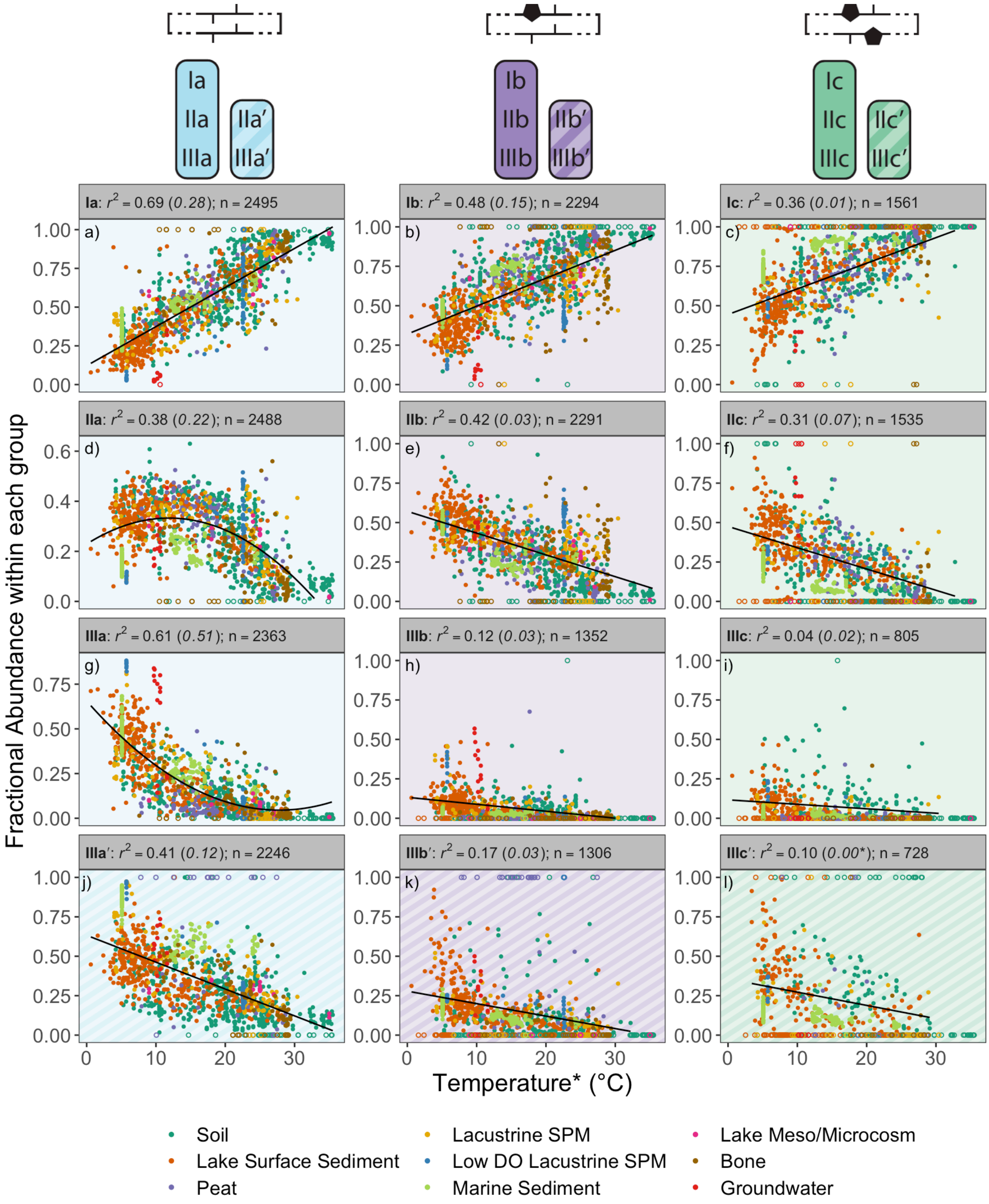

Ia
IIa
IIIa
IIa′
IIIa′
Ib
IIb
IIIb
IIb′
IIIb′
Ic
IIc
IIIc
IIc′
IIIc′
Ia: r^2 = 0.69 (0.28); n = 2495
Ib: r^2 = 0.48 (0.15); n = 2294
Ic: r^2 = 0.36 (0.01); n = 1561
IIa: r^2 = 0.38 (0.22); n = 2488
IIb: r^2 = 0.42 (0.03); n = 2291
IIc: r^2 = 0.31 (0.07); n = 1535
IIIa: r^2 = 0.61 (0.51); n = 2363
IIIb: r^2 = 0.12 (0.03); n = 1352
IIIc: r^2 = 0.04 (0.02); n = 805
IIIa′: r^2 = 0.41 (0.12); n = 2246
IIIb′: r^2 = 0.17 (0.03); n = 1306
IIIc′: r^2 = 0.10 (0.00*); n = 728
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
1.00
0.75
0.50
0.25
0.00
0.6
0.4
0.2
0.0
0
10
20
30
Fractional Abundance within each group
Temperature* (°C)
Soil
Lake Surface Sediment
Peat
Lacustrine SPM
Low DO Lacustrine SPM
Marine Sediment
Lake Meso/Microcosm
Bone
Groundwater

*Figure 4-1: Relationships between temperature and brGDGT fractional abundances (FAs) in the Meth set. The Meth set is formed by grouping brGDGTs by cyclization number and isomer designation (schematics at top); FAs are calculated within each group (Raberg et al., 2021). Plots of IIa', IIb', and IIc' exactly mirror those of IIIa', IIIb', and IIIc', respectively, and are therefore not shown. *Temperatures are (necessarily) different for different sample types (see Methods) and are as follows: warmest month air temperature (Soil, Peat, Bone); mean air temperature of months above freezing (Lake Surface Sediment); sea surface temperature (Marine Sediment); in situ water temperature (SPM, Meso/Microcosm, Groundwater). Best fit curves (quadratic for IIa and IIIa; linear for all other compounds) are plotted for each compound, with R^2 values provided. R^2 values for FAs calculated using the standard approach (grouping all 15 compounds together) are given in parentheses for comparison. P values were < 0.01 except where marked with an asterisk. Open circles denote points with FA = 0 or 1. These points were clear outliers originating from samples with low concentrations of brGDGTs and were excluded from statistical analyses (see Methods).*

The Cyc-Isom set (Fig. 8-16e; Table 8-2) best captured the relationships between brGDGTs and pH for soils, peats, and lacustrine sediments and SPM, with compound-specific relationships again exhibiting similar functional forms across sample types (Fig. 4-2). Linear slopes were often within one standard error (Fig. 8-19). However, unlike for temperature, a consistent offset is visible between the terrestrial (soil and peat) and lacustrine (sediment and SPM) datasets (Fig. 4-2). Sample types with limited datasets, including lacustrine microcosm and low DO SPM, riverine sediments and SPM, and groundwater generally plot within these global trends, with notable exceptions for low DO SPM and groundwater samples (e.g., Fig. 4-2e). Lower-pH microcosm samples also deviated, potentially due to the death of brGDGT-producing microbes in these experiments (Martínez-Sosa et al., 2020a). Again, the Cyc-Isom FAs provided stronger correlations and better agreement between sample types than the standard approach (Fig. 8-20).

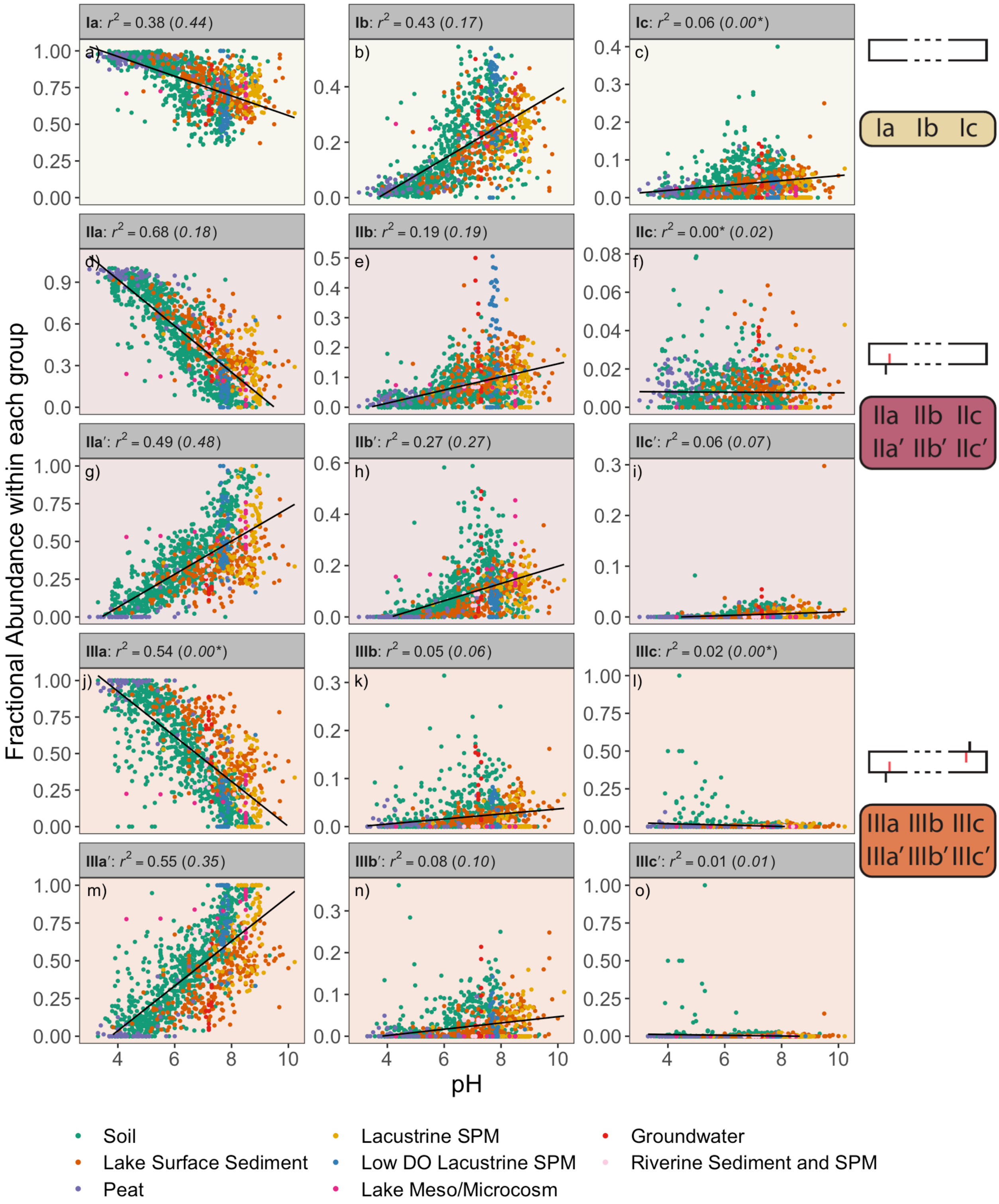


Figure 4-2: Relationships between pH and brGDGT fractional abundances (FAs) calculated within the Cyc-Isom set. The Cyc-Isom set is formed by grouping brGDGTs by methylation number (schematics at right; methylations at C5 (black) or C6 (red)); FAs are calculated within each group (Raberg et al., 2021). Linear fits are plotted for each compound, with R^2 values provided. R^2 values for FAs calculated using the standard approach (grouping all 15 compounds together) are given in parentheses for comparison. P values were < 0.01 except where marked with an asterisk. N = 1295 for all subplots.

We additionally found multiple near-universal relationships between pairs of brGDGT FAs (Fig. 4-3). By partitioning brGDGTs into structural sets, these pairs of FAs are guaranteed to be mathematically independent (Raberg et al., 2021). The FAs of uncyclized and monocyclized tetramethylated brGDGTs ($f\text{Ia}_{\text{MI}}$ and $f\text{Ib}_{\text{MI}}$; Fig. 8-16d; Table 8-2) were tightly correlated in the Meth-Isom set ($R^2 = 0.79$, Fig. 4-3a). Similarly, those of 5-methyl pentamethylated and hexamethylated uncyclized brGDGTs ($f\text{IIa}_{\text{CI}}$ and $f\text{IIIa}_{\text{CI}}$; Fig. 8-16e; Table 8-2) were closely related in the Cyc-Isom set ($R^2 = 0.77$, Fig. 4-3b). Finally, 6-methyl isomers of pentamethylated and hexamethylated uncyclized brGDGTs ($f\text{IIa'}_{\text{Isom}}$ and $f\text{IIIa'}_{\text{Isom}}$; Fig. 8-16c; Table 8-2) are present in proportional abundances in the Isom set ($R^2 = 0.80$, Fig. 4-3c). These relationships were broadly consistent across sample types, with linear slopes and intercepts often within error (Fig. 8-21). A comparable analysis is not possible using traditional Full set FAs as they are not mathematically independent.

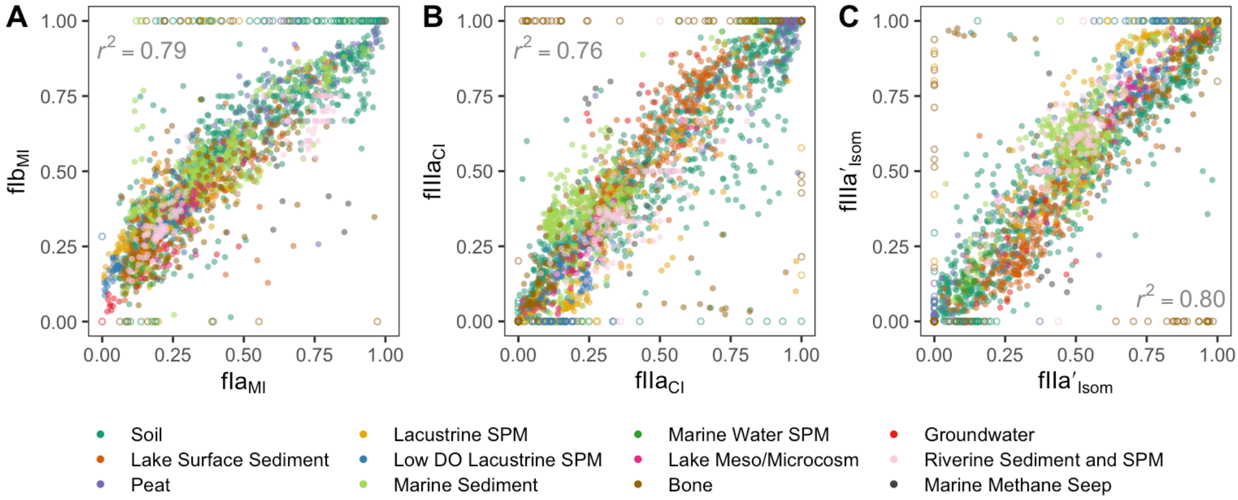


Figure 4-3: Relationships between selected brGDGT fractional abundances (FAs). FAs were calculated in the (A) Meth-Isom (MI), (B) Cyc-Isom (CI), and (C) Isom structural sets (Raberg et al., 2021) (Fig. S1 and Table S1). R^2 values for linear regressions ($p < 0.01$ for all) are provided. Samples with FA = 1 or 0 (open circles) were excluded from statistical analyses (n = 2359, 2203, and 2129 for A, B, and C, respectively). FAs are calculated as follows (Fig. S1 and Table S1): $f\text{Ia}_{MI} = \text{Ia} / (\text{Ia} + \text{IIa} + \text{IIIa} + \text{IIa'} + \text{IIIa'})$ *and* $f\text{Ib}_{MI} = \text{Ib} / (\text{Ib} + \text{IIb} + \text{IIIb} + \text{IIb'} + \text{IIIb'})$; $f\text{IIa}_{CI} = \text{IIa} / (\text{IIa} + \text{IIb} + \text{IIc} + \text{IIa'} + \text{IIb'} + \text{IIc'})$ *and* $f\text{IIIa}_{CI} = \text{IIIa} / (\text{IIIa} + \text{IIIb} + \text{IIIc} + \text{IIIa'} + \text{IIIb'} + \text{IIIc'})$; $f\text{IIa'}_{Isom} = \text{IIa'} / (\text{IIa} + \text{IIa'})$ *and* $f\text{IIIa'}_{Isom} = \text{IIIa'} / (\text{IIIa} + \text{IIIa'})$.

While the basic functional forms of the relationships between brGDGTs, temperature, and pH are in good agreement, differences may be present in their details. An analysis of the fitting coefficients for each sample type revealed that they are often but not always within one standard error. However, challenges, such as a lack of standardized temperature measurements between sample types, limits the utility of this approach at this point. An expanded discussion of these challenges is available in the Supplementary Material (Section 8.3.1).

4.3 Discussion

Microbial community composition can be expected to vary markedly across the diverse media and environments of the compiled dataset. The fact that near-universal temperature and pH trends supersede these community differences is therefore remarkable and suggests that either 1) brGDGTs are produced by a small number of widespread source organisms or 2) brGDGTs play a fundamental physiological role for a wide range of bacterial life. In either case, these near-universal trends have important implications for brGDGT-based proxies. First, they encourage proxy development beyond the well-studied soil/peat and lake sediment datasets. In particular, they support recent work in marine sediments (Dearing Crampton-Flood et al., 2018; Gao et al., 2021; Liu et al., 2021), which exhibited the highest compound-specific temperature correlation in our dataset (fIa_{Meth}; $R^2 = 0.77$, $p < 0.01$), as well as in bones (Zhao et al., 2020) and lacustrine SPM (Zhao et al., 2021). The widespread trends also suggest that it may be possible to construct modern calibrations that join multiple sample types in future studies. Currently, the sample types in the compiled dataset (necessarily) have different associated temperatures (e.g., air versus water; see Methods) and exhibit offsets in pH relationships (Figs. 4-2 and 8-19). While these incongruities preclude the construction of universal calibrations at this time, advances in brGDGT research, such as the growing body of in situ (Wang et al., 2020; Pérez-Angel et al., 2020) and/or more standardizable (Stefanescu et al., 2021) temperature measurements, may allow for such efforts to succeed in the future. Even without these mixed or universal calibrations, the common functional forms exhibited by different sample types lend confidence to qualitative trends in paleoclimate reconstructions from mixed-source archives (provided the relative contribution of each source can be constrained). Furthermore, the commonalities in brGDGT temperature and pH responses support the use of the proxies in paleoclimate reconstructions where modern analogs and details of the depositional setting may be unavailable.

Correlations between pairs of brGDGT FAs (Fig. 4-3) revealed relationships between lipids that spanned nearly every sample type examined. Such connections may provide insight into the biosynthesis and/or biochemical roles of brGDGTs in cells. For example, the correlated Isom set FAs of IIa' and IIIa' (Fig. 4-3c), may indicate that the same enzyme is responsible for methylating both compounds and/or that the two lipids perform related functions. Similarly, the fact that the FA of brGDGT-Ia (with respect to all uncyclized compounds) increases in tandem with the FA of brGDGT-Ib (with respect to all monocyclized compounds) indicates that methylation number can vary independently of cyclization number (Fig. 4-3a). This result may suggest that these two structural variations are able to perform their biological roles independently. New avenues of research based on such observations may provide rare insight into the biochemistry of brGDGTs.

Deviations from the trends described above are nearly as important as the trends themselves. Samples with low associated DO, such as groundwater and some lacustrine SPM, show anomalously high $\text{fIIIa}_{\text{Meth}}$ (Fig. 4-1g), which would lead to underestimates of temperature in most brGDGT calibrations. These samples are also associated with narrow pH ranges yet vary widely in cyclization and isomer abundances (e.g., Fig. 4-2e). Additionally, the origin of the offset between the pH responses of terrestrial and lacustrine samples (e.g., Fig. 4-2j) is unknown. While it is possible that the discrepancy represents important differences between these sample media, it is also possible that bulk pH measurements do not accurately represent the chemical microenvironments experienced by brGDGT producing microbes (e.g., in soils (Baveye et al., 2018)). These or other possibilities warrant further investigation.

Finally, we stress that while clear trends emerge at the global scale, they do not always hold at the regional or site-specific level (e.g., ref. (Van Bree et al., 2020)), indicating that microbial community effects and/or environmental forcing outside of temperature and pH remain important controls.

In conclusion, a comparison of more than 2500 samples spanning a dozen sample types from across the globe revealed near-universal trends between brGDGTs, temperature, and pH. Five distinct relationships displayed R^2 values between 0.68 and 0.80 across 8-12 diverse sample media (Fig. 4-1a, Fig. 4-2d, and Fig. 4-3a-c). Other relationships exhibited weaker correlations but still possessed similar functional forms (e.g., Fig. 4-1d). Recent work has revealed that brGDGT distributions vary in tandem with bacterial communities (Weber et al., 2018; De Jonge et al., 2019; Van Bree et al., 2020; Martínez-Sosa et al., 2020a; Wu et al., 2021; De Jonge et al., 2021), connecting them only indirectly to pertinent variables such as temperature and pH. However, given the wide range of microbial life expected to inhabit the diverse environments of the compiled dataset, the coherent trends presented here suggest at least some physiological basis for the relationships between brGDGTs and their environment. This observation encourages the further development of brGDGT-based proxies, lends confidence to their continued application, and provides a new framework for investigating controls on brGDGT distributions in the future.

4.4 Materials and Methods

4.4.1 Experimental Design

The objective of this study was to compare how brGDGTs in different sample types relate to temperature, pH, and one another. While brGDGT FAs can be quantitatively compared without trouble, environmental parameters pose a challenge. First, temperatures associated with different sample types are often different (e.g., air versus water temperature), with seasonality effects further complicating temperature selection (e.g., summer versus mean annual). Furthermore, many commonly used measurements are themselves only proxies for the true environmental conditions experienced by brGDGT-producing microbes (e.g., air temperature instead of in situ soil temperature). As a result, it is not possible at this time to standardize environmental parameters such that datasets of different sample types can be fully quantitatively merged. However, qualitative and semi-quantitative comparisons of the ways in which brGDGTs relate to temperature and pH can still provide meaningful results with important implications and are worthy of pursuit.

To this end, we compiled the brGDGT FAs and, to the best of our ability, associated temperature and pH values from previously published datasets. We selected temperature parameters that were widely supported in the literature when possible. Where a consensus had yet to be reached (e.g., marine sediments), we selected standardizable and accessible parameters (e.g., sea surface temperatures). These selections are not intended to opine on these areas of research, only to allow for broad comparison with other sample types in this study.

4.4.2 Sample types and associated environmental parameters

Lipid and environmental data were compiled from previously published datasets of modern samples that used the most recent chromatographic methods that separate 5- and 6-methyl isomers (De Jonge et al., 2014a). The compiled dataset (n = 2535) consisted of bone (n = 191), groundwater (n = 14), lake water meso/microcosm (n = 36), lake surface sediment (n = 341), lake water SPM (n = 227, including sediment traps (n = 115) and water filtrates (n = 112)), low DO lake water SPM (n = 135, including sediment traps (n = 29) and water filtrates (n = 106)), authigenic carbonates from a marine methane cold seep (n = 13), marine surface sediment (n = 325, including deep ocean trench sediments (n = 31)), marine SPM (water filtrates, n = 25), peat (n = 96), riverine surface sediments (n = 52) and SPM (water filtrates, n = 85), and soil (n = 995, including permafrost active layer (n = 17)). Data from other sample media, including hot springs, speleothems, and hydrothermal vents, could not be included as these studies did not separate the 5- and 6-methyl isomers. All data and references are available in the dataset provided in the Appendix (Section 8.3.2). Fractional abundances (FAs) were calculated according to Figure 8-16 and Table 8-2 (see ref. (Raberg et al., 2021) for further details).

Temperatures associated with each sample type are as follows: mean air temperature of months above freezing (MAF; lake surface sediments), warmest month temperature (WMT; soils, peats, bones), *in situ* water temperature (lacustrine SPM (integrated summer water column where available for sediment traps, point measurement otherwise), meso/microcosm, and groundwater), sea surface temperature of the warmest month (SST, marine sediments above 72°N), and mean annual SST (all other marine sediments). These temperatures were primarily based on previous calibrations and data availability, with some adjustments as follows.

Lake sediments have recently been shown to correlate with the MAF warm-season air temperature index on a global scale (Raberg et al., 2021; Martínez-Sosa et al., 2021). Soil and peat brGDGTs have also been shown to correlate with MAF (Dearing Crampton-Flood et al., 2020). However, we used another warm-season temperature index, WMT, which provided comparable compound-specific correlations and better overlap with other sample types in this study. To bolster the soil dataset on the cold end, we generated monthly air temperature normals (1970-2000) for permafrost soil active layer samples (Kusch et al., 2019b) using the WorldClim database (Fick and Hijmans, 2017).

Though bones were initially calibrated against mean annual air temperature, a warm-season bias was observed in that study (Zhao et al., 2020). Due to this bias, and to match the soil and peat datasets, we used WMT values generated with the WorldClim database as above. Only bones with substantial *in situ* brGDGT production (Zhao et al., 2020) (isomer ratio > 50%) were included in temperature analyses.

Temperature calibrations for brGDGTs in marine settings are an active area of research. Thus far brGDGT production has been demonstrated in both deep (Liu et al., 2014) and shallow (Zhu et al., 2011) marine water columns and correlations have been found between marine surface sediment brGDGTs and both sea surface (Gao et al., 2021; Liu et al., 2021) and bottom water (Dearing Crampton-Flood et al., 2018) temperatures. Due to these uncertainties and data availability constraints, we used the SST temperature parameter for marine sediments in this study. Where published data were not available, we generated SSTs using the Extended Reconstructed Sea Surface Temperature database (Huang et al., 2017).

We used mean annual SST except for sites above 72°N, which experience substantial sea ice cover, for which we used the SST of the warmest month. To avoid marine sediments with terrestrial influence, we excluded those samples with a $\#\text{rings}_{\text{tetra}}$ index (Sinninghe Damsté, 2016) and (IIIa + IIIa') / (IIa + IIa') < 0.92 from temperature analyses (Xiao et al., 2020; Liu et al., 2021).

A calibration has yet to be generated for lake water filtrates; however, point measurements of water temperature taken at the time of sampling are widely available and were thus used in this study. Integrated summer (June, July, August) water column temperatures were recently shown to correlate well with brGDGTs in sediment traps (Zhao et al., 2021) and were used where available; point measurements of *in situ* water temperature were used otherwise. "Low DO" was defined as DO < 1 mg/L. For the three lakes in this study without a published DO profile, samples below a reported redox transition zone were considered "Low DO" (Yao et al., 2019; Yao et al., 2020).

Where available, *in situ* pH values were used for all sample types except lake surface sediments, for which we used average water column pH.

4.4.3 Statistical Analysis

All statistical analyses were performed in R (Team R Development Core, 2021). We generated quadratic (Figs. 4-1d and g) or linear (all others) regressions for all relationships in the main text and figures. P values were < 0.01 for all regressions or else are denoted by an asterisk (Figs. 4-1 to 4-3, 8-18, and 8-20) or an open point (Figs. 8-17, 8-19, and 8-21). Due primarily to the low abundances of some compounds, some samples in Figures 4-1 and 4-3 displayed an FA = 0 or 1. These points, which are denoted by open circles in the plots, were clear outliers and were not included in statistical analyses (R^2 values, fitting coefficients, or plotted trend lines).

4.5 Acknowledgements

We kindly thank Sebastian Kopf for help with coding and figure-making, the entire PACEMAP and ILLUME teams as well as members of the Organic Geochemistry Lab at CU Boulder for fruitful discussions, and all authors who shared additional data. We additionally thank the following funding sources: National Science Foundation grant OPP-1737712 (GHM, JS), National Science Foundation grant OPP-1836981 (GHM, ÁG, JS), National Science Foundation grant DDRI-1657743 (GHM), a Doctoral Grant from the University of Iceland (ÁG), and a Project Grant from the University of Iceland Research Fund (ÁG).

5 CHAPTER V: Intact polar brGDGTs at Lake Qaupat, Baffin Island: implications for lipid sources and paleoclimate applications

Abstract. Paleotemperature histories derived from lake sediment archives can provide valuable context for modern climate changes and inform future trajectories. BrGDGT lipids have proven to be a valuable tool in such pursuits due to their empirical correlation with temperature and near ubiquity in nature. Despite their widespread application, however, the relative contributions of terrestrial and lacustrine sources of brGDGTs to lake sediment archives is site-dependent and difficult to constrain. Here, we explore the potential for intact polar brGDGTs – the lipids as they exist in living bacteria – to provide insight into the sources of brGDGTs on the landscape and their contributions to the sedimentary record. We examined the distributions of core and intact brGDGTs in soils, surface and downcore sediments, water filtrates, and sediment traps across five lake catchments in the Eastern Canadian Arctic, with an emphasis on Lake Qaupat (QPT), Baffin Island. We found soils to be dominated by brGDGTs with a monoglycosyl (1G) head group, while lacustrine samples displayed a far greater proportion of phosphohexose (PH) brGDGTs, providing evidence for *in situ* brGDGT production in both settings. We further found a heightened relative abundance of core- and PH-brGDGT-IIIa in sediments, implying an additional post-depositional source of in situ brGDGTs. A hierarchical clustering analysis of Lake QPT core and intact brGDGT distributions indicated that fossil core brGDGTs in lake sediments were largely lacustrine in origin, while 1G-brGDGTs were primarily soil derived. Additionally, we found evidence for substantial preservation of intact brGDGTs – especially 1G-brGDGTs – downcore on thousand-year timespans, though in situ production cannot be ruled out. In light of this apparent downcore recalcitrance, we explored the possibility of reconstructing 1G-brGDGT-derived soil temperatures and core-brGDGT derived lake temperatures in tandem from lake sediment archives.[3]

5.1 Introduction

Over the past 50 years, the Arctic has experienced substantial changes in snow cover, vegetation greening, sea and land ice cover, and precipitation (Box et al., 2019 and references within). These changes are linked in large part to a marked rise in surface air temperatures, which have been increasing at more than twice the global average for the past two decades (Ballinger et al., 2020). As high latitude regions begin to diverge from historical analogs, the paleoclimate approach becomes an increasingly important tool for constraining their possible futures (Miller et al., 2010a; Tierney et al., 2020).

---

[3] This chapter is in prep:
Raberg, J. H., Flores, E., Crump, S. E., De Wet, G., Dildar, N., Miller, G. H., Geirsdóttir, Á., and Sepúlveda, J.: Intact polar brGDGTs at Lake Qaupat; Baffin Island: implications for lipid sources and paleoclimate applications, *in prep.*

One tool for reconstructing high latitude paleotemperature records is a class of lipids called branched glycerol dialkyl glycerol tetraether lipids (brGDGTs). Structural variations in the alkyl-chain backbones of these bacterial membrane lipids were found to correlate with temperature and pH in soils (Weijers et al., 2007b). Analogous relationships were soon found in numerous other sample media including lake sediments (Tierney et al., 2010), allowing for the generation of global empirical temperature calibrations (Raberg et al., 2021; Martínez-Sosa et al., 2021) and the reconstruction of temperature histories from lacustrine archives across the high northern latitudes (Thomas et al., 2018; Crump et al., 2019; Harning et al., 2020; Lindberg et al., 2021).

Despite their widespread application, however, the sources of brGDGT lipids to sedimentary archives remain difficult to constrain. The seemingly ubiquitous compounds have been found in terrestrial (e.g., soil), freshwater, and marine environments around the world (see Chapter 4). BrGDGTs in these different environments may record substantially different temperatures (e.g., soil versus water temperature) and can contribute to sedimentary archives to varying degrees. For example, studies that tracked the abundance and distribution of brGDGTs in lakes (Tierney et al., 2012; Loomis et al., 2014b; De Jonge et al., 2015a) and river systems (De Jonge et al., 2014b; Hanna et al., 2016; Guo et al., 2020) have shown significant freshwater production and transport within a catchment. Furthermore, evidence has been found for brGDGT production within the lake sediments themselves (Tierney et al., 2012; Buckles et al., 2014b; Van Bree et al., 2020). As a result, brGDGTs in lake sediment archives can range from largely autochthonous (e.g., Loomis et al., 2014) to primarily soil-derived (e.g., Ning et al., 2019; Peterse et al., 2014). Some tools have been developed to disentangle sources of brGDGTs in marine settings (e.g., Sinninghe Damsté, 2016; Xiao et al., 2020) and even apply mixed calibrations (Dearing Crampton-Flood et al., 2018). However, while these techniques from the marine realm have recently been tested in lacustrine systems (Martin et al., 2019), methods to disentangle the various sources of brGDGTs in lake sediment archives remain in their nascent stages.

A second challenge to the interpretation of brGDGT-based paleotemperature records is the convoluting influence of dissolved oxygen levels on lipid distributions. The potential influence of oxygen depletion on brGDGT distributions was recognized early in the study of lacustrine brGDGTs (Tierney et al., 2012). Its effect has since been demonstrated in numerous lake systems (e.g., Van Bree et al., 2020; Colcord et al., 2017; Weber et al., 2018), and is often associated with an increasing proportion of hexamethylated brGDGTs. As these brGDGTs are also associated with colder temperatures, anoxic conditions in lake waters have the potential to artificially lower reconstructed temperatures. Apart from intensive studies of compound-specific isotopes (Colcord et al., 2017; Weber et al., 2018), tools for circumventing or even indicating an influence of dissolved oxygen on paleotemperature records are virtually non-existent.

The vast majority of studies on brGDGTs have relied on measurements of the recalcitrant "core" skeletons of the brGDGT lipids (c-brGDGTs). However, much additional information is contained in the more diverse pool of intact polar brGDGTs (i-brGDGTs), i.e., the lipids as they exist in living cell membranes. Due to analytical challenges, these intact lipids have been primarily analyzed by hydrolyzing them to their core skeletons and measuring the resulting c-brGDGT distributions. This method has allowed access to a putative "living" pool of brGDGTs. It has provided much information on the production rates and locations of brGDGTs (e.g., Huguet et al., 2017; Zell et al., 2013), but it sacrifices the information contained in the full suite of intact lipids.

In recent years, chromatographic methods have been developed to analyze brGDGTs in their full, intact form (e.g., Wörmer et al., 2013), though their application in the literature is limited (Liu et al., 2010; Peterse et al., 2011). These techniques expand the number of measurable brGDGTs from fifteen to over sixty, allowing for the resolution of more subtle distinctions between brGDGT distributions. Studies employing these methods have provided evidence for enhanced productivity of brGDGTs in anoxic zones of peats (Liu et al., 2010; Peterse et al., 2011). However, the distributions of i-brGDGTs across the terrestrial landscape and in lacustrine systems remain largely unexplored, and to our knowledge, no study has yet examined them in the Arctic.

Here we investigate the potential for i-brGDGTs to distinguish between sources of brGDGTs on the landscape in modern lake catchments from the Eastern Canadian Arctic, with an emphasis on Baffin Island. We measure brGDGTs in both their core and intact forms in soils, surface and downcore lake sediments, and aqueous environments across five lake sites, with a primary focus on Lake Qaupat. We further explore whether our findings can be used to inform c-brGDGT applications downcore and/or develop new i-brGDGT-based proxies.

5.2 Methods

5.2.1 Study sites and sample collection

Soil, water filtrate (WF), sediment trap (ST), and surface sediment (SS) samples were collected from five lake sites across Baffin Island (Clyde Forelands Lake 8 (CF8), Clyde Forelands Lake 3 (CF3), Brother of Fog Lake (BRO), and Lake Qaupat (QPT), informal names) and northern Québec (3 Lakes North (3LN)) during the summers of 2017 and 2018 (Fig. 5-1 inset). Additionally, three samples from sediment core QPT-2A, collected in 2016 and described by Crump et al. (2019) were re-extracted for this study, as well as downcore sediment from a suboxic interval of Early Holocene sediment collected from BRO Lake in 2017. We analyzed 53 samples in total, including 25 samples from our primary study site, Lake QPT (Fig. 5-1).

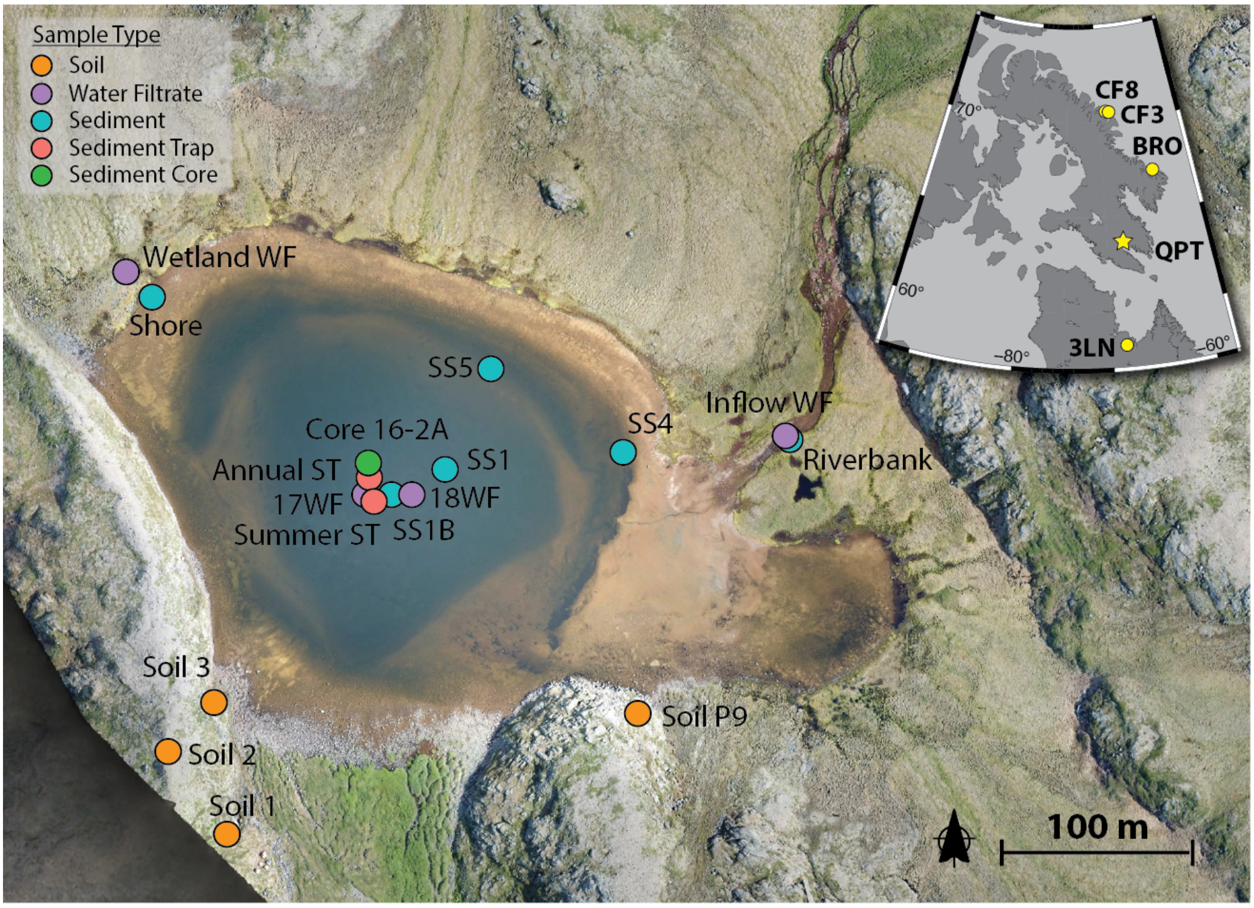


Figure 5-1: Drone imagery (July 2018) and sampling locations for Lake Qaupat (QPT), Baffin Island. Other sites in this study are marked in the inset: Clyde Forelands Lake 8 (CF8); Clyde Forelands Lake 3 (CF3); Brother of Fog Lake (BRO), and 3 Lakes North (3LN). Sample type abbreviations are as follows: surface sediment (SS); sediment trap (ST), water filtrate (WF).

Soils were sampled from beneath the dominant plant communities in the lake catchment (Martha K. Raynolds, personal communication in the field) at a depth interval of 0-10 cm, except for one (QPT P9) which was point sampled at 10 cm. Surface (0 m) and bottom (1-2 m above lakebed) waters were collected from lake centers and filtered through 0.3µm glass fiber filters (Advantec). Sediment traps at deep (2 m above the lakebed) and shallow (2 m below the surface) positions at Lake QPT were deployed from August 2017-July 2018 (ST Upper and ST Lower). After collection, these sediment traps were redeployed for the peak summer month, July 2018-August 2018 (ST Upper Sum and ST Lower Sum). A deep ST was also deployed from May-August 2018 at BRO Lake. Temperature, pH, conductivity, and dissolved oxygen concentration of the lake water columns were measured using a multiparameter sonde (HydroLab HL4, OTT HydroMet) at the time of deployment and collection of STs. Lake water samples were collected at the same time and analyzed for total phosphorous at the Arikaree Lab at the University of Colorado at Boulder. Drone imagery was collected at 6 cm resolution with a Mavic Pro (DJI) drone on July 22, 2018. A composite (Fig. 5-1) was generated using MapsMadeEasy.

5.2.2 Lipid extraction and analysis

After collection, samples were kept as cool as possible in the field and frozen upon returning to Colorado. All samples were freeze-dried and extracted using a modified Bligh & Dyer (BD) extraction method (Bligh and Dyer, 1959; Wörmer et al., 2013; Raberg et al., 2021). We extracted ~1g sediment, ~5g soil, glass fiber filters (2.5-12 L water filtered total), and all available material from sediment traps. Samples were first vortexed and sonicated in solvent Mix A, consisting of dichloromethane (DCM):methanol (MeOH):50 mM phosphate buffer (aq., pH 7.4) [1:2:0.8, v:v:v]. After centrifugation (3000 rpm and 10°C for 10 min), the supernatant was poured or pipetted into a glass separatory funnel. The process was performed twice with Mix A, twice with Mix B (DCM:MeOH:5% trichloroacetic acid buffer (aq., pH 2) [1:2:0.8, v:v:v]), and once with Mix C (DCM:MeOH [1:5, v:v]). Separation of aqueous and organic fractions was induced by adding equal volumes of HPLC-grade water and DCM to the separatory funnel. The organic (bottom) fraction was collected and dried under a nitrogen stream. DCM was then added to the aqueous phase, shaken, and allowed to separate to remove any residual lipids. The resulting organic fraction was then added to the extract and dried. Aliquots of the resulting total lipid extract (TLE) were redissolved in 99:1 (v:v) hexane:isopropanol (c-brGDGTs) or 9:1 (v:v) DCM:MeOH (i-brGDGTs) and filtered (0.45 µm, PTFE) before analysis. Six samples were additionally extracted using a standard Accelerated Solvent Extraction (ASE) method, as described previously (Raberg et al., 2021).

Core and intact brGDGTs were analyzed using a Thermo Scientific UltiMate 3000 high-performance liquid chromatography instrument coupled to a Q Exactive Focus Orbitrap-Quadrupole high-resolution mass spectrometer (HPLC-MS). Core brGDGTs were analyzed using a slightly modified version (Raberg et al., 2021) of the methods of Hopmans et al. (2016). Separation of i-brGDGTs was achieved using the chromatographic conditions described by Wörmer et al. (2013) and Cantarero et al. (2020). Briefly, two mobile phases (Phase A: 0.01% formic acid and 0.01% NH_4OH in acetonitrile:DCM [75:25, v:v]; Phase B; 0.4% formic acid and 0.4% NH_4OH in MeOH:H_2O [50:50, v:v]) were combined in a gradient from 1% B to 40% B. Intact brGDGTs were ionized via electrospray ionization and identified by exact masses (± 5 ppm) and retention times. Due to the lack of a suitable standard, we did not quantify absolute i-brGDGT abundances. Instead, we relied on fractional abundances of integrated peak areas as described below.

5.2.3 Index calculation and terminology

We defined multiple fractional abundances (FAs) for our analysis of both c- and i-brGDGTs. First, we calculated the standard Full Set FAs (Raberg et al., 2021) of c-brGDGTs,

$$\begin{aligned} fx_{Full} = x/(&Ia + Ib + Ic + IIa + IIb + IIc + IIIa + IIIb + IIIc + IIa' + IIb' + IIc' \\ &+ IIIa' + IIIb' \\ &+ IIIc') \end{aligned} \qquad (5-1)$$

where *x* is the integrated peak area of any compound in the denominator. An analogous FA can unfortunately not be calculated for i-brGDGTs as it is currently unknown whether the analytical method can achieve separation of the 5- and 6-methyl isomers. We therefore calculated FAs of brGDGTs for each head group using the 9 compounds that are commonly measured without isomer separation and denote the difference with an asterisk:

$$fx\text{–}HG_{Full*} = x/(HG\text{–}Ia + HG\text{–}Ib + HG\text{–}Ic + HG\text{–}IIa_C + HG\text{–}IIb_C + HG\text{–}IIc_C + HG\text{–}IIIa_C + HG\text{–}IIIb_C + HG\text{–}IIIb_C) \qquad (5-2)$$

For each head group, HG, x is any compound in the denominator and all penta- and hexamethylated compounds represent the combined sum of 5- and 6-methyl isomers, as denoted by the subscript "C". We also calculated the analogous Meth* Set FAs,

$$fx\text{–}HG_{Meth*} = x/(HG\text{–}Ia + HG\text{–}IIa_C + HG\text{–}IIIa_C) \qquad (5-3)$$

and MBT' index (Peterse et al., 2012),

$$MBT'_{HG} = \frac{(HG\text{–}Ia + HG\text{–}Ib + HG\text{–}Ic)}{(HG\text{–}Ia + HG\text{–}Ib + HG\text{–}Ic + HG\text{–}IIa_C + HG\text{–}IIb_C + HG\text{–}IIc_C + HG\text{–}IIIa_C)} \qquad (5-4)$$

again using the sums of penta- and hexamethylated compounds. Finally, we calculated FAs within the entire pool of i-brGDGTs,

$$fx\text{–}HG_{Full\text{-}C\text{-}IPLs} = x/\sum i\text{–}brGDGTs \qquad (5-5)$$

and the FA of each HG,

$$fHG = \sum HG\text{–}brGDGTs/\sum i\text{–}brGDGTs \qquad (5-6)$$

5.2.4 Statistical methods

We used principal component analyses (PCAs) and hierarchical clustering to compare similarities in brGDGT distributions. PCAs were performed using the FactoMineR package in R (Lê et al., 2008; Team R Development Core, 2021). We used Full Set FAs (Eq. 5-1) for c-brGDGTs and Full*-IPLs Set FAs (Eq. 5-5) for i-brGDGTs. We chose the Euclidean distance measure to produce hierarchical clustering of Full* Set FAs (Eq. 5-2). Unweighted Pair Group Method with Arithmetic Mean (UPGMA; average-linkage) agglomerative clustering method was used according to Ramos Emmendorfer and de Paula Canuto (2021). We calculated approximately unbiased p-values (AU) and bootstrap probability values (BP). The bootstrap supports were based on 10,000 replicates and clusters with AU ≥ 95% confidence were highlighted for this study (Suzuki and Shimodaira, 2006).

5.3 Results

5.3.1 Source-dependent i-brGDGT distributions

BrGDGTs were detected both as core (HG = H) and intact polar lipids with monoglycosyl (1G) or phosphohexose (PH) head groups (Fig. 5-2). We used a chromatographic method capable of separating the 5- and 6-methyl brGDGT isomers (solid and dashed colored methylations in Fig. 5-2a, respectively) for c-brGDGTs (see Methods). The multifaceted peaks present for many i-brGDGTs (e.g., PH-IIa; Fig. 5-2c) suggest the existence of isomers for these compounds as well. However, as the method used for i-brGDGTs separates compounds based on their polar head groups rather than their alkyl-chain moieties, it is unlikely that these peaks represent 5- and 6-methyl i-brGDGTs. We did not attempt to integrate these isomers separately in this study; however, they may reflect meaningful differences in head group position and their existence warrants further investigation.

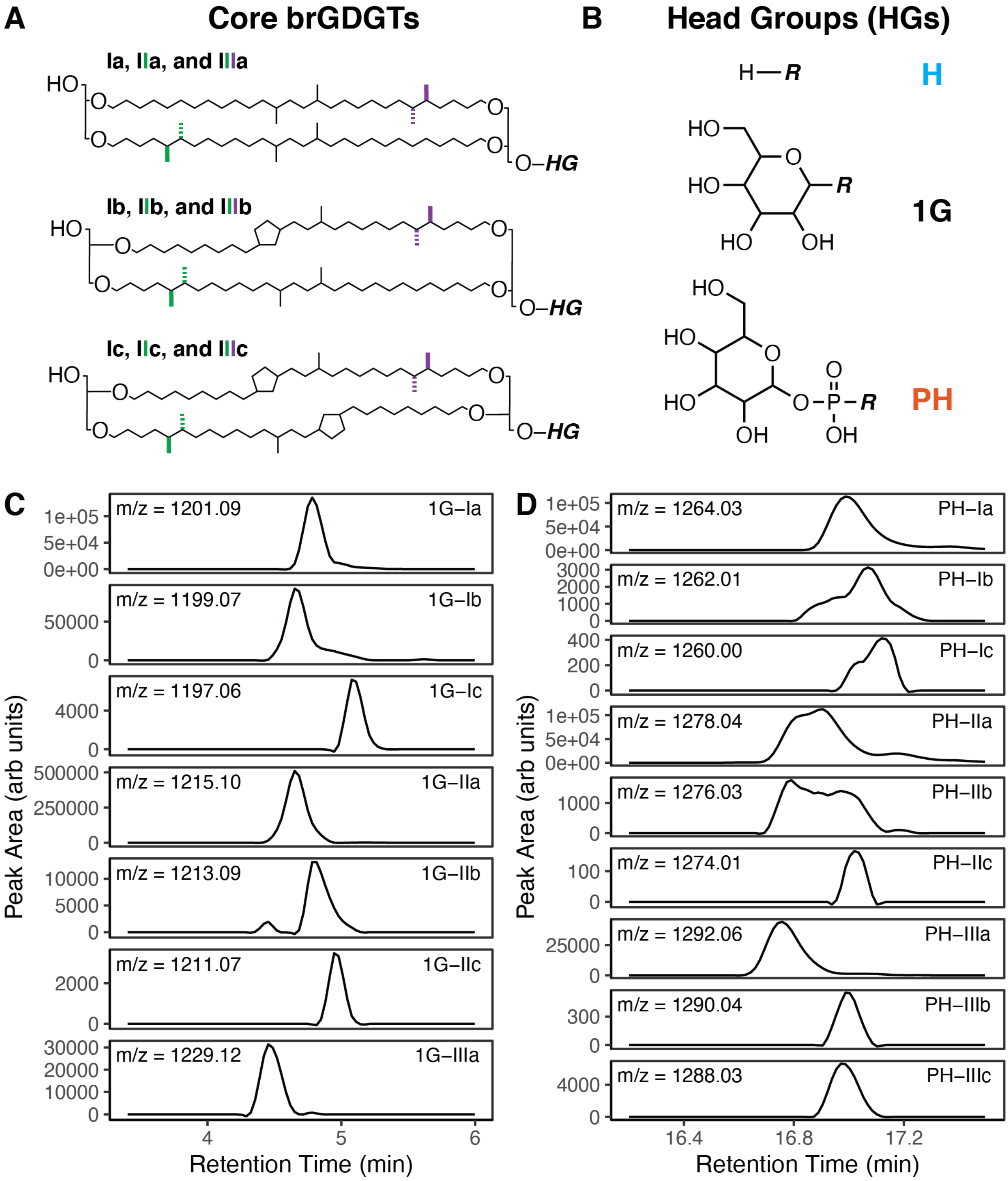


Figure 5-2: Structures of the A) core brGDGTs and B) head groups (HGs) measured in this study. Solid and dashed lines represent methylations for the 5- and 6-methyl isomers of penta- (green) and hexa- (green + purple) methylated compounds, respectively. "R" denotes any core brGDGT; "HG" denotes any head group. Structures shown do not encapsulate all possible isomers. Chromatograms (smoothed, 10-point moving average) of the C) 1G- and D) PH-brGDGTs are shown for a representative sample (3LN Soil 2; 1G-IIIb and 1G-IIIc not detected).

Both 1G- and PH-brGDGTs were detected in most samples (Fig. 5-3). Clear differences in the relative proportions of these two HGs existed between sample types. Soils were dominated by 1G-brGDGTs at all sites, with a mean fractional abundance of 1G versus PH (f1G; Eq. 5-6) of 0.89 ± 0.10 (Figs. 5-3 and 5-4). Surface sediments had a consistently lower proportion of 1G-brGDGTs (f1G = 0.34 ± 0.21) but ranged substantially across lakes (f1G = 0.20-0.68; Figs. 5-3 and 5-4). Water column samples (water filtrates and sediment traps) displayed both the largest range and within-site variance (Figs. 5-3 and 5-4), likely due at least in part to low lipid concentrations in the water filtrates that may have left many compounds below the detection limit.

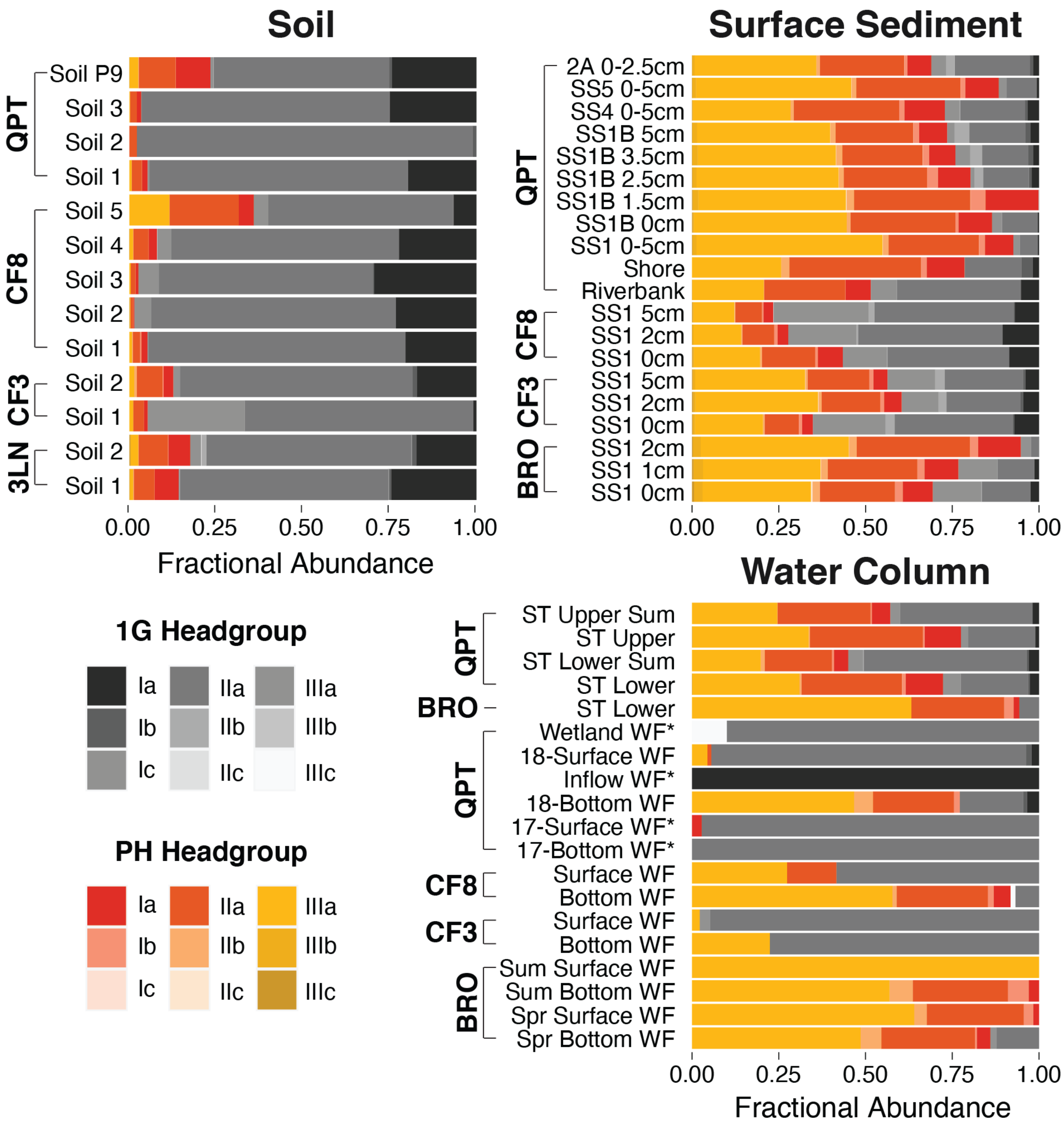


Figure 5-3: Distributions of i-brGDGTs. Fractional abundances (calculated using Eq. 5 of the main text) of 1G- and PH-brGDGTs in Canadian soil, surface sediment (SS), sediment trap (ST), and water filtrate (WF) samples. QPT samples marked with an asterisk were excluded from principal component and clustering analyses due to low compound abundances.

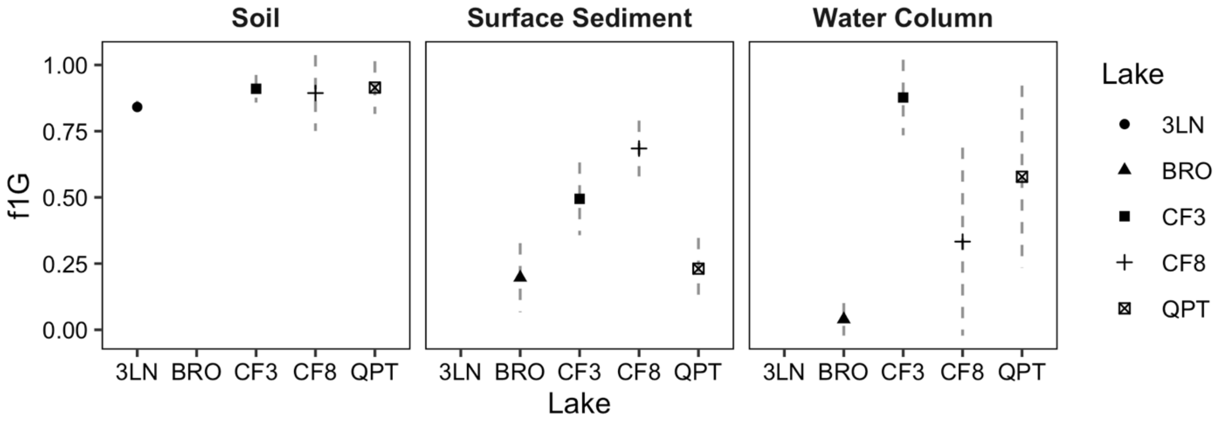


Figure 5-4: Average fraction of 1G head group by sample type. "Water Column" samples consist of sediment traps and lacustrine water filtrates.

Finally, we examined the relationship between lake water total phosphorous (TP) and the fractional abundance of PH-brGDGTs (fPH; Eq. 5-6) in surface sediments. While we found a weak positive correlation (adjusted $R^2 = 0.39$), it was not significant ($p > 0.2$; Fig. 8-22).

5.3.2 Connections between c- and i-brGDGT distributions

To elucidate connections between c- and i-brGDGTs, we compared the distributions of brGDGTs within each head group (H (i.e., c-brGDGTs), 1G, and PH) at Lake QPT. We performed principal component analyses (PCAs) on the resulting c- and i-brGDGT FAs (Figs. 5-5a and 5-5b, respectively). In the c-brGDGT PCA, uncyclized compounds (Ia, IIa/IIa', and IIIa) were the primary drivers of principal component 1 (PC1), which explained 77.2% of the variance and separated soils from lacustrine samples (Fig. 5-5a). Uncyclized i-brGDGTs (especially PH-IIIa, 1G-IIa, PH-IIa, and PH-Ia) drove a similar separation of these sample types (Fig. 5-5b). In the c-brGDGT PCA, lake surface sediments were separated from water column samples by a heightened proportion of brGDGT-IIIa (Fig. 5-5a). Surface sediments were separated in a similar manner from sediment traps, but not water filtrates, by PH-IIIa in the i-brGDGT PCA (Fig. 5-5b). As suggested by the c- and i-brGDGT PCAs, the Meth* set FAs of brGDGTs IIIa and PH-IIIa displayed a strong one-to-one correlation (adjusted $R^2 = 0.80$, $p < 2.2e-16$) across all sample types, including soils (Fig. 5-6a). In contrast, the analogous FAs of IIIa and 1G-IIIa were only weakly related (Fig. 5-6b, adjusted $R^2 = 0.27$, $p = 0.0003$).

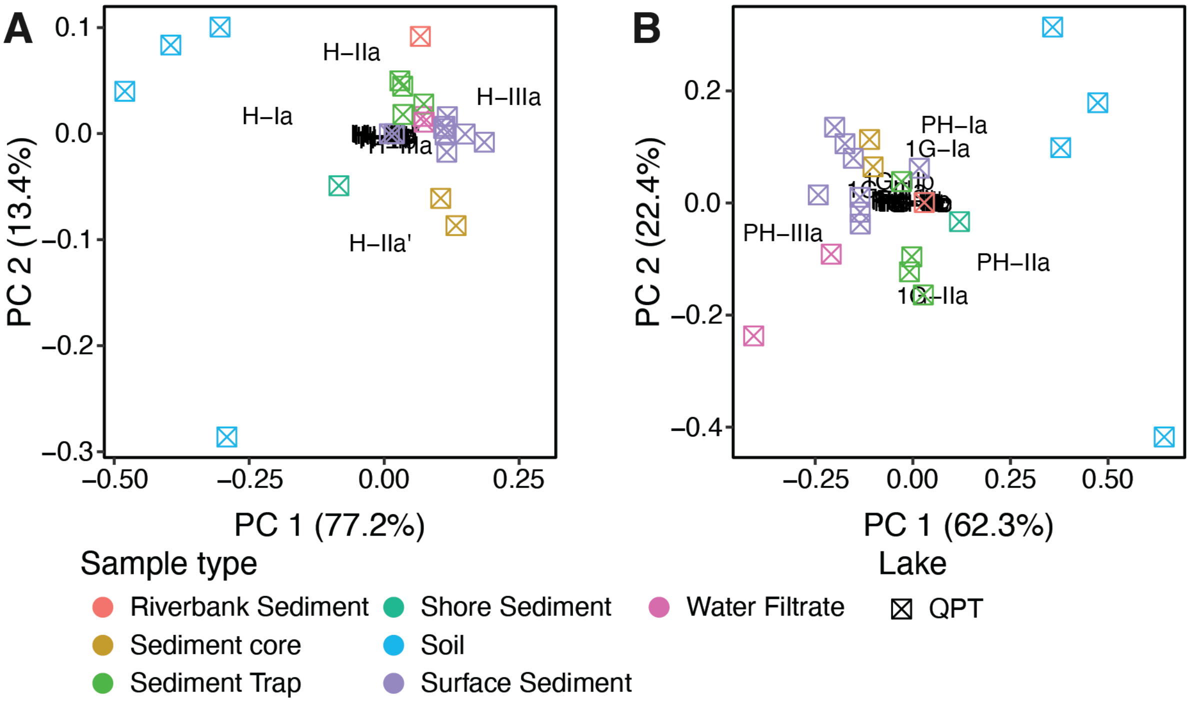


Figure 5-5: Principal component analyses (PCAs) of fractional abundances (FAs) of Lake QPT A) core and B) intact brGDGTs. FAs of c-brGDGTs were calculated according to Eq. 1 (i.e., with 5- and 6-methyl isomers separated); those of i-brGDGTs were calculated with Eq. 5-5. Compound loadings are plotted in black text (scaled by 1.8 for clarity).

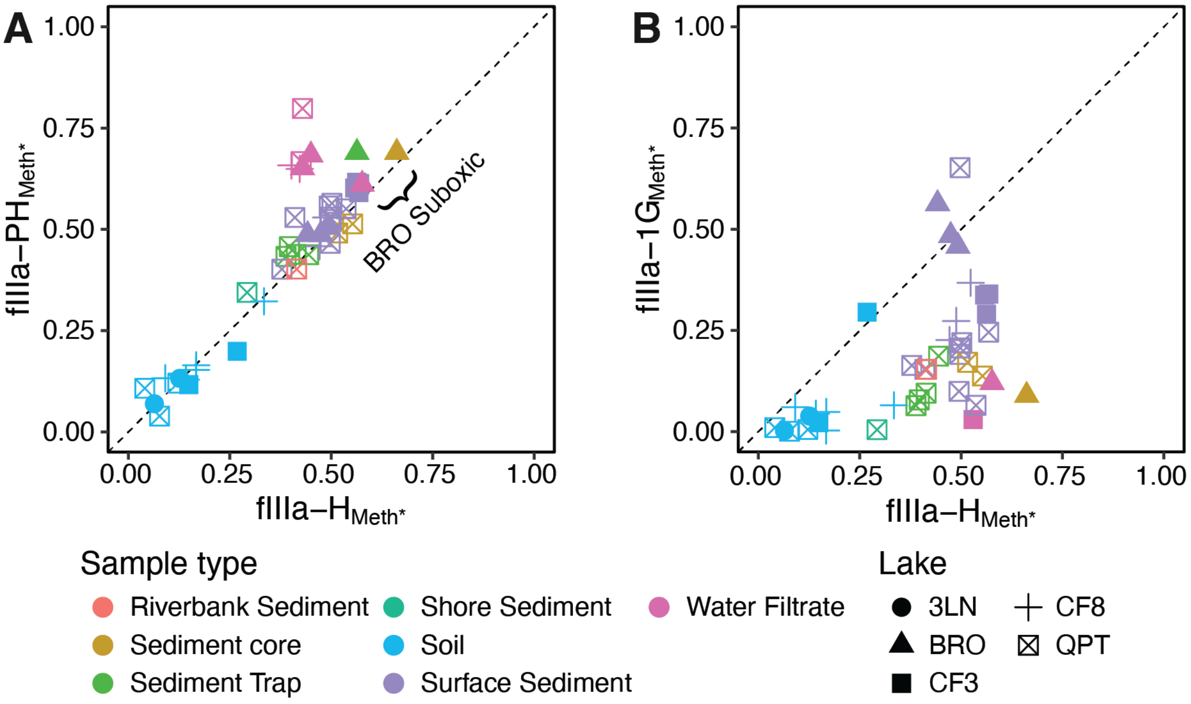


Figure 5-6: Fractional abundances of core and intact brGDGT-IIIa. Relationship between the fractional abundance (FA) of IIIa in the Meth set (Eq. 5-3) for the core and A) PH- and B) 1G-brGDGTs. Samples with FA = 0 or 1 were excluded. A one-to-one relationship is plotted with a dashed line. The three samples influenced by suboxic conditions (BRO Spr. Bottom WF, Lower ST, and Early Holocene downcore sediment) are marked in A).*

To further compare the similarities of c- and i-brGDGT distributions, we performed a hierarchical cluster analysis, again using FAs calculated within each HG independently (Eq. 5-2). The analysis produced three statistically significant groupings (red boxes in Fig. 5-7; see Methods). Group 1 clustered all lacustrine core (H, light blue) and PH (orange) brGDGTs. Group 2 contained all 1G-brGDGT distributions, including soils and lacustrine samples. Group 3 was comprised of soil core and PH distributions. Groups 2 and 3 were more similar to each other than to Group 1.

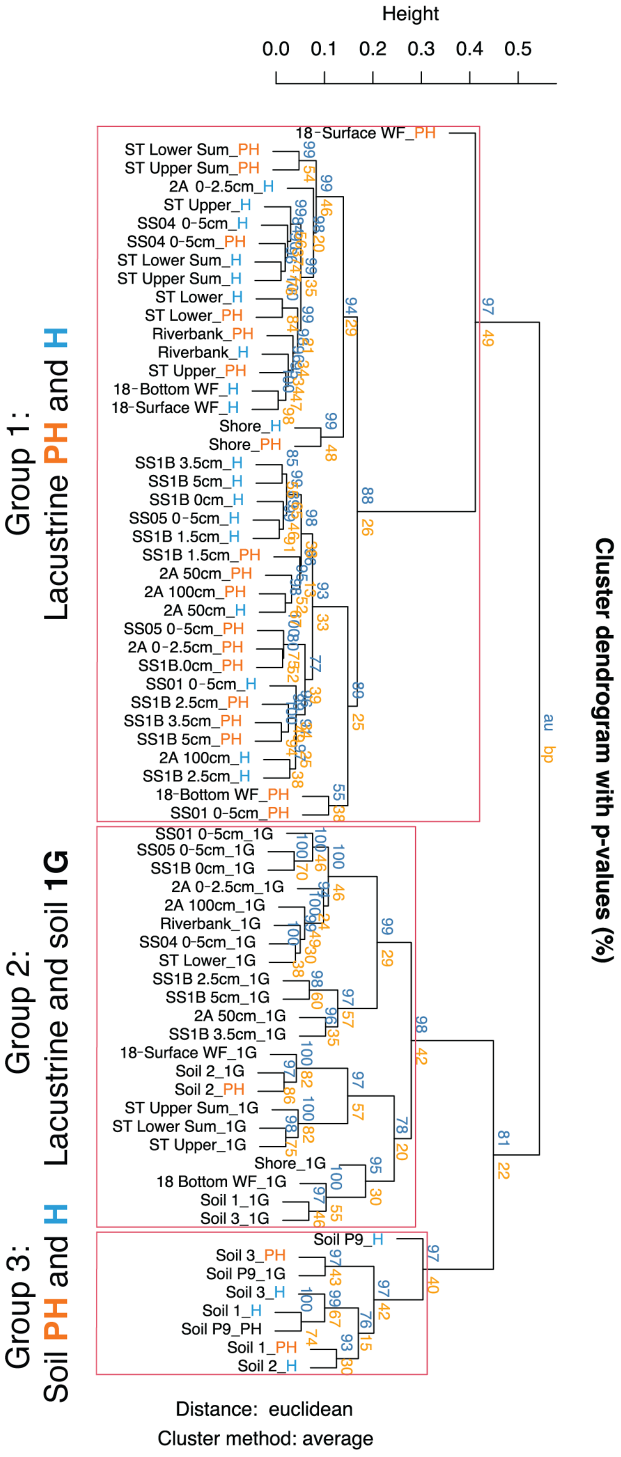


Figure 5-7: Clustering dendrogram of all samples at Lake QPT. Fractional abundances of core lipid backbones were calculated independently for head group (HG = H (core), PH, and 1G) accord to Eq. 5. Approximately unbiased p-values (au; blue) and bootstrap probability values (bp; orange) are provided. Clusters with au ≥ 95% confidence are highlighted in red boxes.

5.3.3 Potential for downcore applications

In addition to using i-brGDGTs to distinguish lipid sources in modern lake catchments, we examined their potential for applications in sedimentary archives. To test the preservation potential of i-brGDGTs, we extracted three downcore samples from Lake QPT (0, 50, and 100 cm; ~0, 4.3, and 6.3 ka (Crump et al., 2019)) and one from BRO Lake (~50-60 cm, Early Holocene in age) using the BD extraction method. Both 1G- and PH-brGDGTs were detected in all four samples (Fig. 5-8a). However, downcore sediments are generally extracted using the ASE rather than the BD method. We therefore extracted six samples in parallel with the ASE and BD methods. We found both 1G- and PH- head groups in the BD-extracted samples, but only 1G after extracting with ASE (Fig. 5-8b).

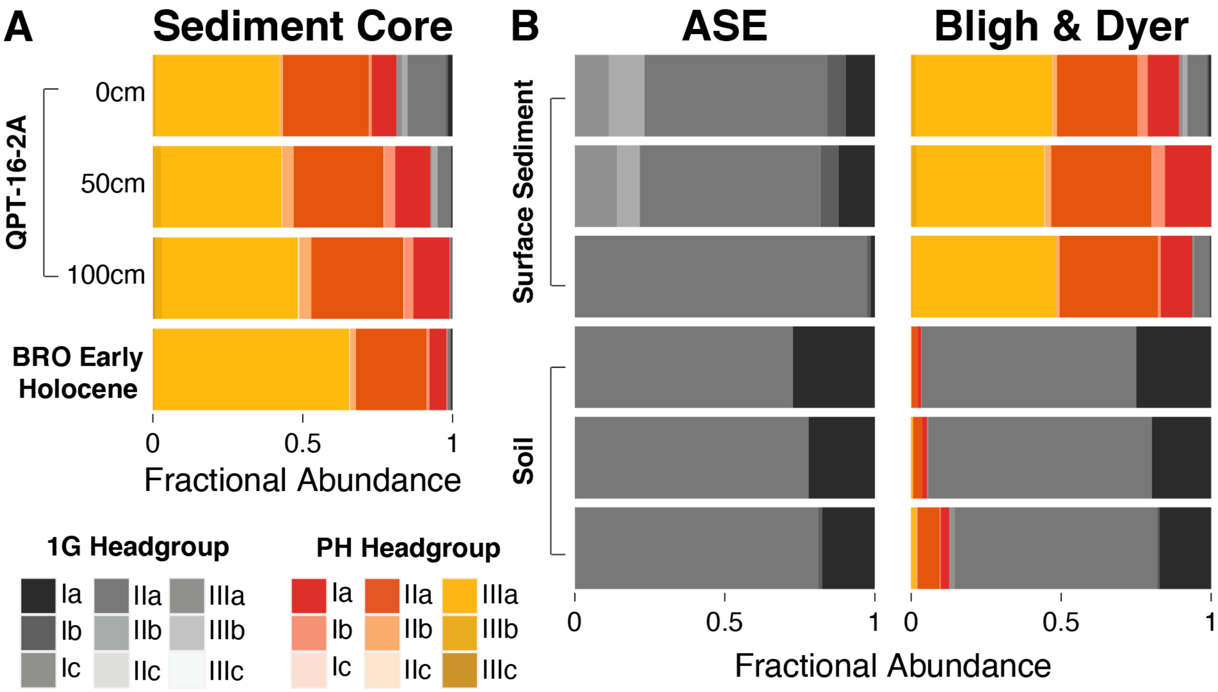


Figure 5-8: Potential for downcore applications of intact brGDGTs. A) 1G- and PH-brGDGTs present in downcore sediments extracted using the Bligh and Dyer (BD) method. B) Samples extracted in parallel using the ASE and BD methods. From top to bottom, surface sediments in B) are QPT-SS1B 2.5cm, 1.5cm, and 0cm and soils are QPT Soil 3, QPT Soil 1, and CF3 Soil 2 – see also Fig. 1.

To test whether i-brGDGTs could serve as proxies for environmental conditions in the same manner as c-brGDGTs, we calculated MBT', $fIIIa_{Meth*}$, and mean annual air temperature (MAAT) on c- and i-brGDGTs at Lake QPT. The MBT' index was similar for core, PH-, and 1G-brGDGTs for lacustrine samples, but varied widely for soils (Fig. 5-9a). The FA of IIIa in the Meth* set (Eq. 5-3) was in good agreement across HGs for soils (Fig. 5-9b). In contrast, the values of fIIIa-PH_{Meth*} and $fIIIa_{Meth*}$ were similar for lacustrine samples, but fIIIa-$1G_{Meth*}$ displayed substantially lower values that were in closer agreement with those of the soils (Fig. 5-9b). Finally, we applied a transfer function (Eq. 12 of Loomis et al. (2012)) to calculate MAAT from core, PH-, and 1G-brGDGTs. Again, soils showed good agreement of MAAT values across head groups (Fig. 5-9c). For lacustrine samples, MAATs calculated from 1G-brGDGTs broadly agreed with soil MAATs, while those from core and PH-brGDGTs were substantially lower (Fig. 5-9c). Temperatures reconstructed using two other published transfer functions (Pearson et al., 2011; Foster et al., 2016) did not show the same agreements (not shown).

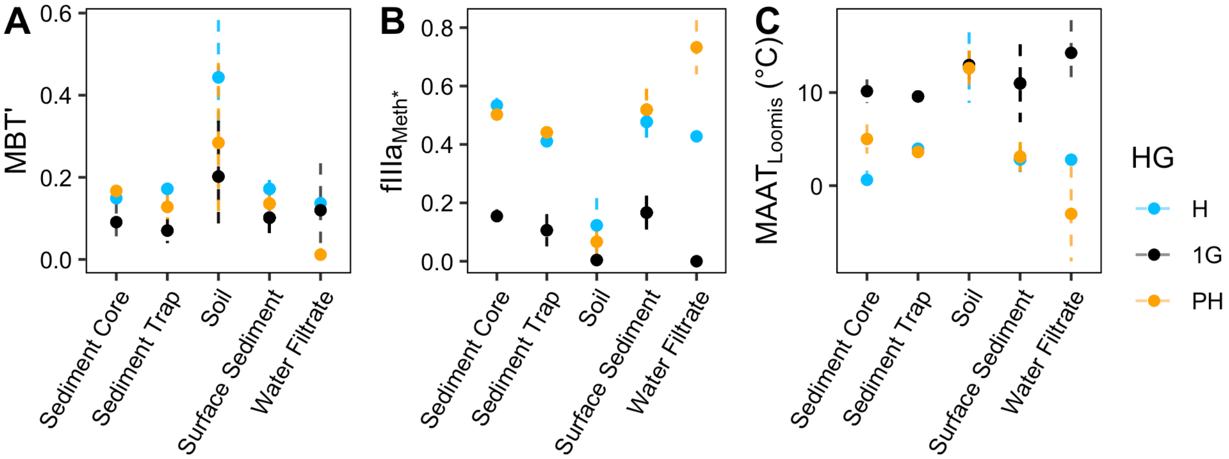


Figure 5-9: Comparison of core and intact brGDGT indices. Mean (± one standard deviation) of a) the MBT' index, b) the fractional abundance of IIIa in the Meth structural set (Eq. 5-3), and c) mean annual air temperature (MAAT) calculated using the Eq. 12 of Loomis et al. (2012) for each sample type at Lake QPT. "Sediment Core" consists of the QPT16-2A 50cm and 100cm samples.*

5.4 Discussion

5.4.1 Intact brGDGTs as a tool for distinguishing lipid sources

An analysis of c- and i-brGDGT distributions revealed in situ production of brGDGTs in the soil, water column, and surface sediment at Lake QPT. Soils were distinguished from lacustrine samples by a stark difference in the relative abundances of 1G-brGDGTs (Figs. 5-3 and 5-4). Within the lake, surface sediments were further distinguished from water column samples by a heightened relative abundance of hexamethylated brGDGTs, particularly IIIa and PH-IIIa (Figs. 5-5 and 5-6). These results support the conclusions of previous work using core lipids, which suggested that the production of brGDGTs is commonplace in both terrestrial and lacustrine environments (e.g., Buckles et al., 2014; Guo et al., 2020; Peterse et al., 2014; Tierney et al., 2012).

Widespread production poses a challenge for downcore applications, where a mixture of brGDGTs derived from multiple sources can complicate interpretations (c.f., Blaga et al., 2010). Promisingly, our results from Lake QPT suggest that 1G-brGDGTs in lacustrine samples may be primarily soil-derived and may therefore serve as a tool for disentangling lipid sources at this site. Soils at Lake QPT (and across the Canadian Arctic) were characterized by high f1G values (Fig. 5-3 and 5-4). Furthermore, 1G-brGDGTs in lacustrine samples were found to have similar distributions to those in the soils, grouping together in a cluster analysis (Fig. 5-7). These patterns suggest either that lacustrine 1G-brGDGTs are largely soil derived or that 1G-brGDGTs are produced in similar distributions regardless of environmental setting. The fact that the cluster of 1G-brGDGTs (Group 2) is more similar to soil (Group 3) than lacustrine (Group 1) core and PH-brGDGTs provides evidence for the former (Fig. 5-7), though the latter cannot be ruled out. Finally, it is important to note that while the 1G-brGDGT distributions group together in the Lake QPT cluster analysis, different results may emerge with a larger sample set or at other lake sites. For example, surface sediments at Lakes CF3, CF8, and BRO displayed higher relative abundances of 1G-IIIa than almost all soils in this study (Figs. 5-3 and 5-6b). While it is possible that soils with distributions similar to CF3 Soil 1 (Figs. 5-3 and 5-6b) contribute an outsized proportion of 1G-brGDGTs to these lakes, it seems likely that at least some of the 1G-brGDGTs were produced in the lacustrine environment at these sites, perhaps driven by lower available phosphorous (Fig. 8-22), which has been shown to lead to HG remodeling (Van Mooy et al., 2009). Despite these unknowns in other lake systems, our detailed study at Lake QPT suggests that 1G-brGDGTs throughout the Lake QPT study site are primarily soil-derived.

Despite the presence of soil-derived 1G-brGDGTs in lacustrine samples, our results suggest that the fossil lipid pool is primarily derived from PH-brGDGTs produced in situ. A cluster analysis grouped lacustrine core and PH-brGDGTs separately from soil and 1G-brGDGTs (Fig. 5-7). Furthermore, the MBT' index, $\mathrm{fIIIa}_{\mathrm{Meth*}}$, and $\mathrm{MAAT}_{\mathrm{Loomis}}$ of core brGDGTs agreed better with those derived from PH-brGDGTs than 1G-brGDGTs (Fig. 5-9). Finally, the strong correlation between the Meth* set FAs of IIIa and PH-IIIa (Fig. 5-6) suggest that the phospho- rather than glycolipid is the primary precursor of the core brGDGT-IIIa. This last connection is of particular interest for two reasons. First, a heightened relative abundance of core brGDGT-IIIa was the distinguishing marker of in situ surface sediment production at Lake QPT. Second, this same heightened abundance has been observed in multiple locations (e.g., Peterse et al., 2014; Tierney et al., 2012) and has been identified as the cause of an artificial "core-top cooling" signal in paleotemperature reconstructions (e.g., Miller et al., 2018; Zhao et al., 2021). The tight correlation between IIIa and PH-IIIa in our study suggests that this signal may be driven by the production of hexamethylated phosphohexose i-brGDGTs in the lake surface sediment.

Interestingly, the connection between core and PH-brGDGTs was apparent even for soils, despite their high f1G values. One possible explanation for the stronger relationship between PH- and c-brGDGTs is the higher lability of the phosphohexose head group (Harvey et al., 1986), which could lead to a greater turnover rate and a disproportionate contribution to the fossil lipid pool. Under this hypothesis, distributions of PH-brGDGTs may respond more rapidly to changes in environmental conditions and could prove useful in studies of shorter-term (e.g., seasonal) trends in brGDGTs.

5.4.2 Potential applications of i-brGDGTs in sedimentary archives

Both PH- and 1G-brGDGTs were detected in sediments up to 1 m below the lake floor and Early Holocene in age (Fig. 5-8a). While these compounds represent the first i-brGDGTs measured in their intact forms downcore to our knowledge, IPL-derived c-brGDGTs have been studied in lake sediment archives (e.g., Tierney et al., 2012) and marine sediments >140 ka old (Lengger et al., 2013). Whether these downcore i-brGDGTs are produced in situ, preserved over geologic timescales, or a combination of both is not clear, however. Some studies have found evidence for in situ production of intact branched (Lengger et al., 2013) and isoprenoidal (Liu et al., 2011) GDGTs in marine sediments. However, these IPLs have also demonstrated a remarkable recalcitrance. Tetraethers with phosphate-based head groups have been shown to survive over millennial timescales in marine sediments (Lengger et al., 2014). Glycosidic GDGTs are even more recalcitrant, with a degradation rate on the order of millions of years (Lengger et al., 2014). Furthermore, both br- (Fig. 5-8b; this study) and iso- (Lengger et al., 2012) GDGTs with 1G head groups have been shown to survive the high temperatures and pressures of ASE extractions. Finally, soil-derived 1G-brGDGTs appear to survive transport, deposition, and burial without significant alteration to their distributions at Lake QPT (Fig. 5-7). Therefore, while further study is needed to determine whether a significant in situ contribution exists, it is likely that the pool of downcore i-brGDGTs – and especially 1G-brGDGTs – at Lake QPT is ancient in origin.

The preservation of uniquely soil-derived 1G-brGDGTs into the geologic record could open the door for the development of qualitatively new proxies. BrGDGTs in lacustrine systems have been increasingly shown to be influenced by lake water chemistry, especially dissolved oxygen (e.g., Loomis et al., 2014a; Raberg et al., 2021; Weber et al., 2018). This influence affects brGDGT-derived temperatures and poses a major challenge for obtaining accurate paleotemperature reconstructions. Temperatures calculated from soil-derived compounds would circumvent this issue. Furthermore, as 1G-brGDGTs survive the ASE extraction process, which is commonly used for downcore sediments, samples may not need to be (re-)extracted with the more time-consuming BD method to apply 1G-brGDGT-based proxies.

To test whether i-brGDGTs display similar temperatures relationships to c-brGDGTs, we calculated and compared temperatures and temperature-related indices derived from PH, 1G, and core brGDGTs. The MBT' index has a positive correlation with temperature (Peterse et al., 2012). At Lake QPT, we find a general agreement in the MBT' values of lacustrine samples from PH-, 1G-, and c-brGDGTs and higher and more variable values in soils. However, the MBT' index relies on several compounds (e.g., IIc) that were rare or absent from our i-brGDGT distributions, perhaps due to lower compound abundances overall, and may therefore not be appropriate for use here. A better index for this study is $fIIIa_{Meth*}$. This fractional abundance relies only on the three most abundant brGDGTs (Ia, IIa_C, and $IIIa_C$) and can be expected to have a robust negative relationship with temperature (Raberg et al., 2021). As shown in Figure 5-9b, all three values of $fIIIa_{Meth*}$ agree well in the soils, indicating that 1G- and PH-brGDGTs have similar temperature sensitivities and that the core pool may be primarily derived from these IPLs. In contrast, lacustrine samples show good agreement between core- and PH-derived values, but a strong warm bias in those recorded by the 1G-brGDGTs (Fig. 5-9). Furthermore, these 1G-derived $fIIIa_{Meth*}$ values generally agree with those recorded in the soils. This discrepancy can be explained if the PH-brGDGTs in the lake are primarily autochthonous, recording lake water temperatures, while the 1G-brGDGTs are largely allochthonous and record the temperatures of catchment soils. Under this hypothesis, a single lake sediment sample can record both soil and lake water temperatures independently. Considering that both sets of IPLs can be measured downcore, these results present the potential for independent soil and lake temperature reconstructions from the same sample extracts. Indeed, temperatures calculated using the Loomis et al. (2012) calibration appear to do just this (Fig. 5-9c). However, other calibrations (Foster et al., 2016; Pearson et al., 2011; not shown) produce qualitatively different results, indicating that IPL-specific calibrations are likely to be necessary for downcore applications.

5.5 Conclusions

Intact brGDGTs at Lake Qaupat and across the Eastern Canadian Arctic provided new insight into the sources of brGDGTs in terrestrial and lacustrine environments. We found a stark distinction between the i-brGDGT composition of soils and lacustrine samples, with 1G-brGDGTs dominating in the soils and PH-brGDGTs in the lakes. We further found heightened relative abundances of IIIa and PH-IIIa in the lake sediments, suggesting post-depositional in situ production. 1G-brGDGTs in lacustrine samples at Lake QPT were likely soil derived, showed a high preservation potential, and survived the ASE extraction procedure, opening the door for the development of soil-specific proxies that can be measured in lake sediment archives. In contrast, core brGDGTs in the lacustrine environment were primarily derived from PH-brGDGTs produced in situ at Lake QPT. A single downcore sediment sample extracted using standard ASE methods can therefore theoretically be used to reconstruct both soil and lake water temperatures at this site. However, further work is needed to calibrate, test, and refine 1G-brGDGT-based proxies and to expand our understanding of i-brGDGT distributions to other lake sites.

6 CHAPTER VI: Conclusions

6.1 Key results of the thesis

In this thesis, I have presented research focusing on three of the major challenges facing the development and application of brGDGT-based proxies in high latitude regions. My first study (Chapter 3) examined the effects of spurious environmental parameters such as seasonality, conductivity, and dissolved oxygen on brGDGT temperature and pH calibrations. My second study (Chapter 4) addressed the issue of a physiological versus community-based origin for the relationships between brGDGTs and the environment. My third study (Chapter 5) aimed to fingerprint sources of brGDGTs in terrestrial and lacustrine settings and determine their relative contributions to the sedimentary pool. Characterization of the study sites (Chapter 2) provided valuable information on lake catchment morphology, lake water chemistry, and *in situ* temperatures which informed each of these three studies. Key takeaways include:

- Drone surveys, water chemistry measurements, and *in situ* temperature monitoring provide valuable context and baseline data for a variety of studies.

- Structurally based groupings of brGDGTs can deconvolve the influences of environmental parameters on brGDGT distributions.

- Through methylation number alone, brGDGTs in lake sediments record the mean temperature of the months above freezing.

- Cyclization numbers and isomer abundances of lake sediment brGDGTs are related to lake water pH and conductivity.

- Near-universal trends across a dozen disparate sample types suggest a physiological origin for the relationships between brGDGTs and the environment.

- Intact polar brGDGTs can be source specific.

- The preservation of i-brGDGTs downcore – especially the soil-derived 1G-brGDGTs – offers a potential new avenue for soil- and lake-specific temperature reconstructions from the same sample extracts.

The studies in this thesis were undertaken with the intent of furthering paleoclimate efforts focused on Quaternary lake sediment archives in high latitude regions. Additional key takeaways in this regard include:

- The brGDGT paleothermometer will reconstruct warm-season temperatures.

- Core brGDGTs in lacustrine settings (including sediments) in the Eastern Canadian Arctic are primarily produced *in situ*.

- Samples associated with low dissolved oxygen deviated from global trends (Chapter 4). Considering that lake water dissolved oxygen generally becomes depleted under ice in the ECA and Iceland (Chapter 2), current calibrations may not produce accurate temperature from anoxic lake sediments in these locations.

- The development of soil-specific 1G-brGDGT calibrations may allow for soil temperature reconstructions from lake sediments, avoiding the complications of lake water anoxia.

6.2 Important remaining challenges

While the results of this thesis have made progress in our understanding of brGDGTs, important challenges remain. First, the influence of oxygen availability on brGDGT distributions must be constrained. While grouping brGDGTs into structural sets deconvolved temperature and pH/conductivity effects, it was unable to characterize oxygen effects. This was likely due to the sparse dataset of oxygen measurements, which often consisted of a single point measurement taken at the time of sampling. A long-term and large-scale study including continuous monitoring of dissolved oxygen levels may be necessary for determining its role in natural settings. Next, though the results of Chapter 4 supported a physiological basis for the environmental relationships of brGDGTs at the global scale, bacterial community composition may still have an important influence on the regional or site-specific level. Such an influence would have important implications for the fidelity of the brGDGT paleothermometer. Finally, the decoupling of lake water and soil temperatures from air temperatures in our high latitude study sites (Chapter 2) highlights the importance of *in situ* temperature monitoring for calibrations studies.

6.3 Future directions

The studies I presented in this thesis have inspired ongoing and future research, some of which I will highlight here. First, the extensive datasets generated as part of this thesis can continue to lend themselves to studies exploring modern relationships between lipid distributions and environmental parameters. For example, *in situ* soil and water temperatures can be used to further improve brGDGT calibrations. Similarly, water (and in the future, soil) chemistry measurements (e.g., total phosphorous) may reveal other parameters that are important to brGDGT distributions. This extensive set of modern environmental data can also be used in conjunction with other lipid classes, such as i-brGDGTs or iso-GDGTs, while the compiled dataset of Chapter 4 may prove useful for examining additional global trends. Finally, in addition to these research questions in the modern environment, I also aim to investigate the paleoclimate questions central to the PACEMAP and ILLUME goals by generating brGDGT-based proxy reconstructions from Icelandic and Canadian sedimentary archives.

7 References

References

Ballinger T. J., Overland J. E., Wang M., Bhatt U. S., Hanna E., Hanssen-Bauer I., Kim S. J.-J., Thoman R. L. and Walsh J. E. (2020) Surface Air Temperature. *NOAA Arct. Rep. Card 2020*, 1–7. Available at: https://repository.library.noaa.gov/view/noaa/27901 [Accessed October 5, 2021].

Baveye P. C., Otten W., Kravchenko A., Balseiro-Romero M., Beckers É., Chalhoub M., Darnault C., Eickhorst T., Garnier P., Hapca S., Kiranyaz S., Monga O., Mueller C. W., Nunan N., Pot V., Schlüter S., Schmidt H. and Vogel H.-J. (2018) Emergent Properties of Microbial Activity in Heterogeneous Soil Microenvironments: Different Research Approaches Are Slowly Converging, Yet Major Challenges Remain. *Front. Microbiol.* **0**, 1929.

Baxter A. J., Hopmans E. C., Russell J. M. and Sinninghe Damsté J. S. (2019) Bacterial GMGTs in East African lake sediments: Their potential as palaeotemperature indicators. *Geochim. Cosmochim. Acta* **259**, 155–169. Available at: https://www.sciencedirect.com/science/article/pii/S0016703719303230?dgcid=raven_sd_via_email [Accessed July 10, 2019].

Beales N. (2004) Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review. *Compr. Rev. Food Sci. Food Saf.* **3**, 1–20. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1541-4337.2004.tb00057.x [Accessed November 22, 2020].

Björnsson H. (2003) The annual cycle of temperature in Iceland. *Tech. Rep. 03037*, 45.

Blaga C. I., Reichart G.-J. J., Heiri O. and Sinninghe Damsté J. S. (2009) Tetraether membrane lipid distributions in water-column particulate matter and sediments: A study of 47 European lakes along a north-south transect. *J. Paleolimnol.* **41**, 523–540. Available at: https://link.springer.com/article/10.1007/s10933-008-9242-2 [Accessed November 8, 2018].

Blaga C. I., Reichart G. J., Schouten S., Lotter A. F., Werne J. P., Kosten S., Mazzeo N., Lacerot G. and Sinninghe Damsté J. S. (2010) Branched glycerol dialkyl glycerol tetraethers in lake sediments: Can they be used as temperature and pH proxies? *Org. Geochem.* **41**, 1225–1234. Available at: https://www.sciencedirect.com/science/article/pii/S0146638010001841 [Accessed November 8, 2018].

Bligh E. G. and Dyer W. J. (1959) A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917. Available at: www.nrcresearchpress.com [Accessed October 7, 2021].

Blyth A. J. and Schouten S. (2013) Calibrating the glycerol dialkyl glycerol tetraether temperature signal in speleothems. *Geochim. Cosmochim. Acta* **109**, 312–328.

Box J. E., Colgan W. T., Christensen T. R., Schmidt N. M., Lund M., Parmentier F.-J. W., Brown R., Bhatt U. S., Euskirchen E. S., Romanovsky V. E., Walsh J. E., Overland J. E., Wang M., Corell R. W., Meier W. N., Wouters B., Mernild S., Mård J., Pawlak J. and Olsen M. S. (2019) Key indicators of Arctic climate change: 1971–2017. *Environ. Res. Lett.* **14**, 045010. Available at: https://iopscience.iop.org/article/10.1088/1748-9326/aafc1b [Accessed October 5, 2021].

Van Bree L. G. J., Peterse F., Baxter A. J., De Crop W., Van Grinsven S., Villanueva L., Verschuren D. and Sinninghe Damsté J. S. (2020) Seasonal variability and sources of in situ brGDGT production in a permanently stratified African crater lake. *Biogeosciences Discuss.* **17**, 1–36. Available at: https://doi.org/10.5194/bg-17-5443-2020 [Accessed July 28, 2020].

Brierley C. M., Zhao A., Harrison S. P., Braconnot P., Williams C. J. R., Thornalley D. J. R., Shi X., Peterschmitt J. Y., Ohgaito R., Kaufman D. S., Kageyama M., Hargreaves J. C., Erb M. P., Emile-Geay J., D'Agostino R., Chandan D., Carré M., Bartlein P. J., Zheng W., Zhang Z., Zhang Q., Yang H., Volodin E. M., Tomas R. A., Routson C., Richard Peltier W., Otto-Bliesner B., Morozova P. A., McKay N. P., Lohmann G., Legrande A. N., Guo C., Cao J., Brady E., Annan J. D. and Abe-Ouchi A. (2020) Large-scale features and evaluation of the PMIP4-CMIP6 midHolocene simulations. *Clim. Past* **16**, 1847–1872.

Briner J. P., Axford Y., Forman S. L., Miller G. H. and Wolfe A. P. (2007) Multiple generations of interglacial lake sediment preserved beneath the Laurentide Ice Sheet. *Geology* **35**, 887–890. Available at: https://pubs.geoscienceworld.org/geology/article/35/10/887-890/129672 [Accessed April 24, 2018].

Buckles L. K., Weijers J. W. H., Tran X.-M., Waldron S. and Sinninghe Damsté J. S. (2014a) Provenance of tetraether membrane lipids in a large temperate lake (Loch Lomond, UK): implications for glycerol dialkyl glycerol tetraether (GDGT)-based palaeothermometry. *Biogeosciences* **11**, 5539–5563. Available at: www.biogeosciences.net/11/5539/2014/ [Accessed July 12, 2020].

Buckles L. K., Weijers J. W. H., Verschuren D. and Sinninghe Damsté J. S. (2014b) Sources of core and intact branched tetraether membrane lipids in the lacustrine environment: Anatomy of Lake Challa and its catchment, equatorial East Africa. *Geochim. Cosmochim. Acta* **140**, 106–126.

Buckley M. W. and Marshall J. (2016) Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review. *Rev. Geophys.* **54**, 5–63. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/2015RG000493 [Accessed October 21, 2021].

Bull I. D., Lockheart M. J., Elhmmali M. M., Roberts D. J. and Evershed R. P. (2002) The origin of faeces by means of biomarker detection. *Environ. Int.* **27**, 647–654. Available at: www.elsevier.com/locate/envint [Accessed October 24, 2021].

Cantarero S. I., Henríquez-Castillo C., Dildar N., Vargas C. A., von

Dassow P., Cornejo-D'Ottone M. and Sepúlveda J. (2020) Size-Fractionated Contribution of Microbial Biomass to Suspended Organic Matter in the Eastern Tropical South Pacific Oxygen Minimum Zone. *Front. Mar. Sci.* **7**, 745. Available at: www.frontiersin.org [Accessed October 14, 2020].

Cao J., Rao Z., Jia G., Xu Q. and Chen F. (2017) A 15 ka pH record from an alpine lake in north China derived from the cyclization ratio index of aquatic brGDGTs and its paleoclimatic significance. *Org. Geochem.* **109**, 31–46.

Cao J., Rao Z., Shi F. and Jia G. (2020) Ice formation on lake surfaces in winter causes warm-season bias of lacustrine brGDGT temperature estimates. *Biogeosciences* **17**, 2521–2536.

Chen M., Zeng G., Zhang J., Xu P., Chen A. and Lu L. (2015) Global Landscape of Total Organic Carbon, Nitrogen and Phosphorus in Lake Water. *Sci. Reports 2015 51* **5**, 1–7. Available at: https://www.nature.com/articles/srep15043 [Accessed October 18, 2021].

Chen Y., Zheng F., Chen S., Liu H., Phelps T. J. and Zhang C. (2018) Branched GDGT production at elevated temperatures in anaerobic soil microcosm incubations. *Org. Geochem.* **117**, 12–21.

Cohen J., Zhang X., Francis J., Jung T., Kwok R., Overland J., Ballinger T. J., Bhatt U. S., Chen H. W., Coumou D., Feldstein S., Gu H., Handorf D., Henderson G., Ionita M., Kretschmer M., Laliberte F., Lee S., Linderholm H. W., Maslowski W., Peings Y., Pfeiffer K., Rigor I., Semmler T., Stroeve J., Taylor P. C., Vavrus S., Vihma T., Wang S., Wendisch M., Wu Y. and Yoon J. (2019) Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nat. Clim. Chang. 2019 101* **10**, 20–29. Available at: https://www.nature.com/articles/s41558-019-0662-y [Accessed October 15, 2021].

Colcord D. E., Pearson A. and Brassell S. C. (2017) Carbon isotopic composition of intact branched GDGT core lipids in Greenland lake sediments and soils. *Org. Geochem.* **110**, 25–32. Available at: https://www.sciencedirect.com/science/article/pii/S014663801730044X [Accessed June 14, 2018].

Crump S. E., Fréchette B., Power M., Cutler S., de Wet G., Raynolds M. K., Raberg J. H., Briner J. P., Thomas E. K., Sepúlveda J., Shapiro B., Bunce M. and Miller G. H. (2021) Ancient plant DNA reveals High Arctic greening during the Last Interglacial. *Proc. Natl. Acad. Sci. U. S. A.* **118**.

Crump S. E., Miller G. H., Power M., Sepúlveda J., Dildar N., Coghlan M. and Bunce M. (2019) Arctic shrub colonization lagged peak postglacial warmth: Molecular evidence in lake sediment from Arctic Canada. *Glob. Chang. Biol.* **25**, 1–13. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.14836 [Accessed April 18, 2021].

D'Anjou R. M., Bradley R. S., Balascio N. L. and Finkelstein D. B. (2012) Climate impacts on human settlement and agricultural activities in northern Norway revealed through sediment biogeochemistry. *Proc. Natl. Acad. Sci.* **109**, 20332–20337. Available at:

https://www.pnas.org/content/109/50/20332 [Accessed October 24, 2021].

Dang X., Ding W., Yang H., Pancost R. D., Naafs B. D. A., Xue J., Lin X., Lu J. and Xie S. (2018) Different temperature dependence of the bacterial brGDGT isomers in 35 Chinese lake sediments compared to that in soils. *Org. Geochem.* **119**, 72–79. Available at: https://www.sciencedirect.com/science/article/pii/S0146638018300275 [Accessed April 17, 2019].

Dang X. Y., Xue J. T., Yang H. and Xie S. C. (2016) Environmental impacts on the distribution of microbial tetraether lipids in Chinese lakes with contrasting pH: Implications for lacustrine paleoenvironmental reconstructions. *Sci. China Earth Sci.* **59**, 939–950. Available at: http://cdc.cma.gov.cn/ [Accessed July 20, 2020].

Dearing Crampton-Flood E., Peterse F., Munsterman D. and Sinninghe Damsté J. S. (2018) Using tetraether lipids archived in North Sea Basin sediments to extract North Western European Pliocene continental air temperatures. *Earth Planet. Sci. Lett.* **490**, 193–205. Available at: https://doi.org/10.1016/j.epsl.2018.03.030 [Accessed September 23, 2021].

Dearing Crampton-Flood E., Tierney J. E., Peterse F., Kirkels F. M. S. A. and Sinninghe Damsté J. S. (2020) BayMBT: A Bayesian calibration model for branched glycerol dialkyl glycerol tetraethers in soils and peats. *Geochim. Cosmochim. Acta* **268**, 142–159.

Dearing Crampton-Flood E., van der Weijst C. M. H., van der Molen G., Bouquet M., Yedema Y., Donders T. H., Sangiorgi F., Sluijs A., Sinninghe Damsté J. S. and Peterse F. (2021) Identifying marine and freshwater overprints on soil-derived branched GDGT temperature signals in Pliocene Mississippi and Amazon River fan sediments. *Org. Geochem.* **154**, 104200.

Dillon J. T., Lash S., Zhao J., Smith K. P., van Dommelen P., Scherer A. K. and Huang Y. (2018) Bacterial tetraether lipids in ancient bones record past climate conditions at the time of disposal. *J. Archaeol. Sci.* **96**, 45–56.

Ding S., Kohlhepp B., Trumbore S., Küsel K., Totsche K. U., Pohnert G., Gleixner G. and Schwab V. F. (2018) In situ production of core and intact bacterial and archaeal tetraether lipids in groundwater. *Org. Geochem.* **126**, 1–12. Available at: https://www.sciencedirect.com/science/article/pii/S0146638018302304 [Accessed March 1, 2019].

Ding S., Schwab V. F., Ueberschaar N., Roth V.-N., Lange M., Xu Y., Gleixner G. and Pohnert G. (2016) Identification of novel 7-methyl and cyclopentanyl branched glycerol dialkyl glycerol tetraethers in lake sediments. *Org. Geochem.* **102**, 52–58. Available at: https://www.sciencedirect.com/science/article/pii/S0146638016302248 [Accessed July 21, 2019].

Ding S., Xu Y., Wang Y., He Y., Hou J., Chen L. and 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. Available at: www.biogeosciences.net/12/3141/2015/ [Accessed July 21, 2019].

Dugerdil L., Joannin S., Peyron O., Jouffroy-Bapicot I., Vannière B., Boldgiv B., Unkelbach J., Behling H. and Ménot G. (2021) Climate reconstructions based on GDGT and pollen surface datasets from Mongolia and Baikal area: calibrations and applicability to extremely cold-dry environments over the Late Holocene. **17**, 1199–1226. Available at: https://doi.org/10.5194/cp-17-1199-2021 [Accessed June 22, 2021].

Fastovich D., Russell J. M., Jackson S. T. and Williams J. W. (2020) Deglacial temperature controls on no-analog community establishment in the Great Lakes Region. *Quat. Sci. Rev.* **234**, 106245.

Fick S. E. and Hijmans R. J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315.

Fietz S., Prahl F. G., Moraleda N. and Rosell-Melé A. (2013) Eolian transport of glycerol dialkyl glycerol tetraethers (GDGTs) off northwest Africa. *Org. Geochem.* **64**, 112–118.

Foster L. C., Pearson E. J., Juggins S., Hodgson D. A., Saunders K. M., Verleyen E. and Roberts S. J. (2016) Development of a regional glycerol dialkyl glycerol tetraether (GDGT)-temperature calibration for Antarctic and sub-Antarctic lakes. *Earth Planet. Sci. Lett.* **433**, 370–379. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X15007220 [Accessed June 26, 2018].

Gao C., Yang Y., Yang H., Zhang Y. G., Lü X., Wang H., Yu X. and Ruan X. (2021) Different temperature dependence of marine-derived brGDGT isomers in a sediment core from the Chukchi Sea shelf. *Org. Geochem.* **152**, 104169.

Geirsdóttir Á., Harning D. J., Miller G. H., Andrews J. T., Zhong Y. and Caseldine C. (2020) Holocene history of landscape instability in Iceland: Can we deconvolve the impacts of climate, volcanism and human activity? *Quat. Sci. Rev.* **249**, 106633. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0277379120305953 [Accessed October 14, 2020].

Geirsdóttir Á., Miller G. H., Andrews J. T., Harning D. J., Anderson L. S., Florian C., Larsen D. J. and Thordarson T. (2019) The onset of neoglaciation in Iceland and the 4.2 ka event. *Clim. Past* **15**, 25–40. Available at: https://www.clim-past.net/15/25/2019/ [Accessed March 6, 2020].

Guo J., Glendell M., Meersmans J., Kirkels F., Middelburg J. J. and 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. Available at: https://doi.org/10.5194/bg-17-3183-2020 [Accessed June 24, 2020].

Hanna A. J. M., Shanahan T. M. and 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. *J. Geophys. Res. Biogeosciences* **121**, 1762–1780. Available at: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015JG003266 [Accessed April 8, 2019].

Harning D. J., Andrews J. T., Belt S. T., Cabedo-Sanz P., Geirsdóttir Á., Dildar N., Miller G. H. and 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. *Paleoceanogr. Paleoclimatology* **34**, 1006–1021.
Harning D. J., Curtin L., Geirsdóttir Á., D'Andrea W. J., Miller G. H. and Sepúlveda J. (2020) Lipid Biomarkers Quantify Holocene Summer Temperature and Ice Cap Sensitivity in Icelandic Lakes. *Geophys. Res. Lett.* **47**, 1–11. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085728 [Accessed August 16, 2020].
Harvey H. R., Fallon R. D. and Patton J. S. (1986) The effect of organic matter and oxygen on the degradation of bacterial membrane lipids in marine sediments. *Geochim. Cosmochim. Acta* **50**, 795–804. Available at: https://www.sciencedirect.com/science/article/pii/0016703786903558 [Accessed April 14, 2019].
Hopmans E. C., Schouten S. and Sinninghe J. S. (2016) The effect of improved chromatography on GDGT-based palaeoproxies. *Org. Geochem.* **93**, 1–6. Available at: http://dx.doi.org/10.1016/j.orggeochem.2015.12.006.
Hopmans E. C., Weijers J. W. ., Schefuß E., Herfort L., Sinninghe Damsté J. S. and Schouten S. (2004) A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. *Earth Planet. Sci. Lett.* **224**, 107–116. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X04003176 [Accessed November 8, 2018].
Hu J., Zhou H., Peng P. and 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. *Chem. Geol.* **420**, 200–212.
Huang B., Thorne P. W., Banzon V. F., Boyer T., Chepurin G., Lawrimore J. H., Menne M. J., Smith T. M., Vose R. S. and Zhange H.-M. (2017) NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5.
Huguet A., Meador T. B., Laggoun-Défarge F., Könneke M., Wu W., Derenne S. and Hinrichs K.-U. U. (2017) Production rates of bacterial tetraether lipids and fatty acids in peatland under varying oxygen concentrations. *Geochim. Cosmochim. Acta* **203**, 103–116. Available at: https://www.sciencedirect.com/science/article/pii/S0016703717300169#f0010 [Accessed September 21, 2018].
Huguet C., Hopmans E. C., Febo-Ayala W., Thompson D. H., Sinninghe Damsté J. S. and Schouten S. (2006) An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids. *Org. Geochem.* **37**, 1036–1041. Available at: https://www.sciencedirect.com/science/article/pii/S0146638006001094 [Accessed November 1, 2018].
Hutchinson G. E. (1938) On the Relation between the Oxygen Deficit and

the productivity and Typology of Lakes. *Int. Rev. der gesamten Hydrobiol. und Hydrogr.* **36**, 336–355. Available at: http://doi.wiley.com/10.1002/iroh.19380360205 [Accessed November 27, 2020].

IPCC (2021) *IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* eds. V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Available at: https://www.ipcc.ch/report/ar6/wg1/.

De Jonge C., Hopmans E. C., Stadnitskaia A., Rijpstra W. I. C., Hofland R., Tegelaar E. and 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. *Org. Geochem.* **54**, 78–82. Available at: https://www.sciencedirect.com/science/article/pii/S0146638012002185 [Accessed November 18, 2018].

De Jonge C., Hopmans E. C., Zell C. I., Kim J. H., Schouten S. and Sinninghe Damsté J. S. (2014a) Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction. *Geochim. Cosmochim. Acta* **141**, 97–112. Available at: https://ac.els-cdn.com/S0016703714004141/1-s2.0-S0016703714004141-main.pdf?_tid=1fada79d-5d3e-4847-9eec-7ad21a0aeb41&acdnat=1525465009_c8674b56137e687aef4788cae99fda32 [Accessed May 4, 2018].

De Jonge C., Kuramae E. E., Radujković D., Weedon J. T., Janssens I. A. and Peterse F. (2021) The influence of soil chemistry on branched tetraether lipids in mid- and high latitude soils: implications for brGDGT-based paleothermometry. *Geochim. Cosmochim. Acta* **310**, 95–112. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0016703721003951 [Accessed July 12, 2021].

De Jonge C., Radujković D., Sigurdsson B. D., Weedon J. T., Janssens I. and Peterse F. (2019) Lipid biomarker temperature proxy responds to abrupt shift in the bacterial community composition in geothermally heated soils. *Org. Geochem.* **137**, 103897.

De Jonge C., Stadnitskaia A., Fedotov A. and Sinninghe Damsté J. S. (2015a) Impact of riverine suspended particulate matter on the branched glycerol dialkyl glycerol tetraether composition of lakes: The outflow of the Selenga River in Lake Baikal (Russia). *Org. Geochem.* **83–84**, 241–252. Available at: http://dx.doi.org/10.1016/j.orggeochem.2015.04.004 [Accessed September 21, 2018].

De Jonge C., Stadnitskaia A., Hopmans E. C., Cherkashov G., Fedotov A. and Sinninghe Damsté J. S. (2014b) In situ produced branched glycerol dialkyl glycerol tetraethers in suspended particulate matter from the Yenisei River, Eastern Siberia. *Geochim. Cosmochim. Acta* **125**, 476–491.

Available at: http://dx.doi.org/10.1016/j.gca.2013.10.031 [Accessed September 20, 2018].

De Jonge C., Stadnitskaia A., Hopmans E. C., Cherkashov G., Fedotov A., Streletskaya I. D., Vasiliev A. A. and Sinninghe Damsté J. S. (2015b) Drastic changes in the distribution of branched tetraether lipids in suspended matter and sediments from the Yenisei River and Kara Sea (Siberia): Implications for the use of brGDGT-based proxies in coastal marine sediments. *Geochim. Cosmochim. Acta* **165**, 200–225. Available at: www.elsevier.com/locate/gcaAvailableonlineatwww.sciencedirect.com [Accessed September 20, 2018].

Kim J.-H., Zell C., Moreira-Turcq P., Pérez M. A. P., Abril G., Mortillaro J.-M., Weijers J. W. H., Meziane T. and Sinninghe Damsté J. S. (2012) Tracing soil organic carbon in the lower Amazon River and its tributaries using GDGT distributions and bulk organic matter properties. *Geochim. Cosmochim. Acta* **90**, 163–180. Available at: https://www.sciencedirect.com/science/article/pii/S0016703712003018 [Accessed November 8, 2018].

Kusch S., Bennike O., Wagner B., Lenz M., Steffen I. and Rethemeyer J. (2019a) Holocene environmental history in high-Arctic North Greenland revealed by a combined biomarker and macrofossil approach. *Boreas* **48**, 273–286. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/bor.12377 [Accessed April 8, 2019].

Kusch S., Winterfeld M., Mollenhauer G., Höfle S. T., Schirrmeister L., Schwamborn G. and Rethemeyer J. (2019b) Glycerol dialkyl glycerol tetraethers (GDGTs) in high latitude Siberian permafrost: Diversity, environmental controls, and implications for proxy applications. *Org. Geochem.* **136**, 103888.

Landrum L. and Holland M. M. (2020) Extremes become routine in an emerging new Arctic. *Nat. Clim. Chang.* **10**, 1108–1115. Available at: http://dx.doi.org/10.1038/s41558-020-0892-z.

Lattaud J., De Jonge C., Pearson A., Elling F. J. and Eglinton T. I. (2021) Microbial lipid signatures in Arctic deltaic sediments - insights into methane cycling and climate variability. *Org. Geochem.* **157**, 104242.

Lê S., Josse J. and Husson F. (2008) FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **25**. Available at: http://www.jstatsoft.org/ [Accessed April 16, 2019].

Lengger S. K., Hopmans E. C., Sinninghe Damsté J. S. and Schouten S. (2012) Comparison of extraction and work up techniques for analysis of core and intact polar tetraether lipids from sedimentary environments. *Org. Geochem.* **47**, 34–40. Available at: https://www.sciencedirect.com/science/article/pii/S0146638012000381 [Accessed October 19, 2018].

Lengger S. K., Hopmans E. C., Sinninghe Damsté J. S. and Schouten S. (2014) Fossilization and degradation of archaeal intact polar tetraether lipids in deeply buried marine sediments (Peru Margin). *Geobiology* **12**, 212–220. Available at: http://doi.wiley.com/10.1111/gbi.12081 [Accessed

September 21, 2018].
Lengger S. K., Kraaij M., Tjallingii R., Baas M., Stuut J. B., Hopmans E. C., Sinninghe Damsté J. S. and Schouten S. (2013) Differential degradation of intact polar and core glycerol dialkyl glycerol tetraether lipids upon post-depositional oxidation. *Org. Geochem.* **65**, 83–93. Available at: http://dx.doi.org/10.1016/j.orggeochem.2013.10.004 [Accessed September 20, 2018].
Li F., Zhang C. L., Wang S., Chen Y., Sun C., Dong H., Li W., Klotz M. G. and Hedlund B. P. (2014) Production of branched tetraether lipids in Tibetan hot springs: A possible linkage to nitrite reduction by thermotolerant or thermophilic bacteria? *Chem. Geol.* **386**, 209–217.
Lincoln S. A., Bradley A. S., Newman S. A. and Summons R. E. (2013) Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in chimneys of the Lost City Hydrothermal Field. *Org. Geochem.* **60**, 45–53.
Lindberg K. R., Daniels W. C., Castañeda I. S. and Brigham-grette J. (2021) Biomarker Proxy Records of Arctic Climate Change During the Mid- Pleistocene Transition from Lake El ' gygytgyn (Far East Russia). , 1–26. Available at: https://doi.org/10.5194/cp-2021-66 [Accessed October 5, 2021].
Liu X.-L., Leider A., Gillespie A., Gröger J., Versteegh G. J. M. and Hinrichs K.-U. (2010) Identification of polar lipid precursors of the ubiquitous branched GDGT orphan lipids in a peat bog in Northern Germany. *Org. Geochem.* **41**, 653–660. Available at: https://ac.els-cdn.com/S0146638010001002/1-s2.0-S0146638010001002-main.pdf?_tid=5bd78f22-6450-4d67-90cd-13eafbf42f34&acdnat=1537831822_cc69f2df4171f65db103c097f79848a5 [Accessed September 24, 2018].
Liu X. L., Zhu C., Wakeham S. G. and Hinrichs K. U. (2014) In situ production of branched glycerol dialkyl glycerol tetraethers in anoxic marine water columns. *Mar. Chem.* **166**, 1–8.
Liu X., Lipp J. S. and Hinrichs K. U. (2011) Distribution of intact and core GDGTs in marine sediments. *Org. Geochem.* **42**, 368–375. Available at: https://ac.els-cdn.com/S0146638011000337/1-s2.0-S0146638011000337-main.pdf?_tid=7aed4a16-1296-4ffb-916f-2b5aeb9a34a6&acdnat=1537383601_22ee9fb4183302a96addd6c68848b6ff [Accessed September 19, 2018].
Liu Y., Xiao W., Wu J., Han L., Zhang H. and Xu Y. (2021) Source, composition, and distributional pattern of branched tetraethers in sediments of northern Chinese marginal seas. *Org. Geochem.* **157**, 104244.
Loomis S. E., Russell J. M., Eggermont H., Verschuren D. and Sinninghe Damsté J. S. (2014a) Effects of temperature, pH and nutrient concentration on branched GDGT distributions in East African lakes: Implications for paleoenvironmental reconstruction. *Org. Geochem.* **66**, 25–37. Available at: http://dx.doi.org/10.1016/j.orggeochem.2013.10.012 [Accessed September 21, 2018].
Loomis S. E., Russell J. M., Heureux A. M., D'Andrea W. J. and Sinninghe Damsté J. S. (2014b) Seasonal variability of branched glycerol dialkyl

glycerol tetraethers (brGDGTs) in a temperate lake system. *Geochim. Cosmochim. Acta* **144**, 173–187.

Loomis S. E., Russell J. M., Ladd B., Street-Perrott F. A. and Sinninghe Damsté J. S. (2012) Calibration and application of the branched GDGT temperature proxy on East African lake sediments. *Earth Planet. Sci. Lett.* **357–358**, 277–288. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X12005274 [Accessed June 26, 2018].

Loomis S. E., Russell J. M., Verschuren D., Morrill C., De Cort G., Sinninghe Damsté J. S., Olago D., Eggermont H., Alayne Street-Perrott F. and Kelly M. A. (2017) The tropical lapse rate steepened during the Last Glacial Maximum. *Sci. Adv.* **3**, e1600815. Available at: http://advances.sciencemag.org/ [Accessed May 3, 2021].

Lu H., Liu W., Yang H., Wang H., Liu Z., Leng Q., Sun Y., Zhou W. and An Z. (2019) 800-kyr land temperature variations modulated by vegetation changes on Chinese Loess Plateau. *Nat. Commun.* **10**, 1–10. Available at: https://doi.org/10.1038/s41467-019-09978-1 [Accessed April 18, 2021].

Lumley T. (2020) leaps: Regression Subset Selection.

Mann M. E., Zhang Z., Rutherford S., Bradley R. S., Hughes M. K., Shindell D., Ammann C., Faluvegi G. and Ni F. (2009) Global signatures and dynamical origins of the little ice age and medieval climate anomaly. *Science (80-.).* **326**, 1256–1260. Available at: https://www.science.org/doi/abs/10.1126/science.1177303 [Accessed October 23, 2021].

Martin C., Ménot G., Thouveny N., Davtian N., Andrieu-Ponel V., Reille M. and Bard E. (2019) Impact of human activities and vegetation changes on the tetraether sources in Lake St Front (Massif Central, France). *Org. Geochem.* **135**, 38–52.

Martínez-Sosa P., Tierney J. E. and Meredith L. K. (2020a) Controlled lacustrine microcosms show a brGDGT response to environmental perturbations. *Org. Geochem.* **145**, 104041.

Martínez-Sosa P., Tierney J. E., Stefanescu I. C., Dearing Crampton-Flood E., Shuman B. N. and Routson C. (2020b) A global Bayesian temperature calibration for lacustrine brGDGTs. *EarthArXiv [preprint].* Available at: https://eartharxiv.org/repository/view/1719/.

Martínez-Sosa P., Tierney J. E., Stefanescu I. C., Dearing Crampton-Flood E., Shuman B. N. and Routson C. (2021) A global Bayesian temperature calibration for lacustrine brGDGTs. *Geochim. Cosmochim. Acta* **305**, 87–105.

McKay N. P., Kaufman D. S., Routson C. C., Erb M. and Zander P. D. (2018) The onset and rate of Holocene Neoglacial cooling in the Arctic. *Geophys. Res. Lett.* Available at: http://doi.wiley.com/10.1029/2018GL079773.

Michelutti N., Wolfe A. P., Briner J. P. and Miller G. H. (2007) Climatically controlled chemical and biological development in Arctic lakes. *J. Geophys. Res. Biogeosciences* **112**, n/a-n/a. Available at: http://doi.wiley.com/10.1029/2006JG000396 [Accessed April 19, 2018].

Miller D. R., Habicht M. H., Keisling B. A., Castañeda I. S. and 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. Available at: https://doi.org/10.5194/cp-14-1653-2018 [Accessed February 12, 2020].

Miller G. H., Alley R. B., Brigham-Grette J., Fitzpatrick J. J., Polyak L., Serreze M. C. and White J. W. C. (2010a) Arctic amplification: Can the past constrain the future? *Quat. Sci. Rev.* **29**, 1779–1790. Available at: https://ac.els-cdn.com/S0277379110000405/1-s2.0-S0277379110000405-main.pdf?_tid=7e152b9e-d3a4-11e7-aa02-00000aacb35d&acdnat=1511809090_c2d8c34b9d08f5fdfa97082be4df5e48 [Accessed November 27, 2017].

Miller G. H., Brigham-Grette J., Alley R. B., Anderson L., Bauch H. A., Douglas M. S. V., Edwards M. E., Elias S. A., Finney B. P., Fitzpatrick J. J., Funder S. V., Herbert T. D., Hinzman L. D., Kaufman D. S., MacDonald G. M., Polyak L., Robock A., Serreze M. C., Smol J. P., Spielhagen R., White J. W. C., Wolfe A. P. and Wolff E. W. (2010b) Temperature and precipitation history of the Arctic. *Quat. Sci. Rev.* **29**, 1679–1715. Available at: https://www.sciencedirect.com/science/article/pii/S0277379110000673 [Accessed April 19, 2019].

Miller G. H., Geirsdóttir Á., Zhong Y., Larsen D. J., Otto-Bliesner B. L., Holland M. M., Bailey D. A., Refsnider K. A., Lehman S. J., Southon J. R., Anderson C., Björnsson H. and Thordarson T. (2012) Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophys. Res. Lett.* **39**, 1–5. Available at: http://doi.wiley.com/10.1029/2011GL050168 [Accessed April 24, 2018].

Miller G. H., Wolfe A. P., Briner J. P., Sauer P. E. and Nesje A. (2005) Holocene glaciation and climate evolution of Baffin Island, Arctic Canada. *Quat. Sci. Rev.* **24**, 1703–1721. Available at: https://www.sciencedirect.com/science/article/pii/S0277379105000752 [Accessed April 19, 2019].

Van Mooy B. A. S., Fredricks H. F., Pedler B. E., Dyhrman S. T., Karl D. M., Koblížek M., Lomas M. W., Mincer T. J., Moore L. R., Moutin T., Rappé M. S. and Webb E. A. (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**, 69–72.

Myers-Smith I. H., Forbes B. C., Wilmking M., Hallinger M., Lantz T., Blok D., Tape K. D., MacIas-Fauria M., Sass-Klaassen U., Lévesque E., Boudreau S., Ropars P., Hermanutz L., Trant A., Collier L. S., Weijers S., Rozema J., Rayback S. A., Schmidt N. M., Schaepman-Strub G., Wipf S., Rixen C., Ménard C. B., Venn S., Goetz S., Andreu-Hayles L., Elmendorf S., Ravolainen V., Welker J., Grogan P., Epstein H. E. and Hik D. S. (2011) Shrub expansion in tundra ecosystems: Dynamics, impacts and research priorities. *Environ. Res. Lett.* **6**, 045509. Available at: http://stacks.iop.org/1748-9326/6/i=4/a=045509?key=crossref.3e2b3e3172ca6f9fc39a607c137b880f [Accessed April 16, 2019].

Naafs B. D. A., Gallego-Sala A. V., Inglis G. N. and Pancost R. D. (2017) Refining the global branched glycerol dialkyl glycerol tetraether (brGDGT) soil temperature calibration. *Org. Geochem.* **106**, 48–56.

Naafs B. D. A., Oliveira A. S. F. and Mulholland A. J. (2021) Molecular dynamics simulations support the hypothesis that the brGDGT paleothermometer is based on homeoviscous adaptation. *Geochim. Cosmochim. Acta.* Available at: https://doi.org/10.1016/j.gca.2021.07.034.

Naafs B. D. A., Rohrssen M., Inglis G. N., Lähteenoja O., Feakins S. J., Collinson M. E., Kennedy E. M., Singh P. K., Singh M. P., Lunt D. J. and Pancost R. D. (2018) High temperatures in the terrestrial mid-latitudes during the early Palaeogene. *Nat. Geosci.* **11**, 766–771. Available at: https://doi.org/10.1038/s41561-018-0199-0 [Accessed March 24, 2021].

Naeher S., Peterse F., Smittenberg R. H., Niemann H., Zigah P. K. and 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. *Org. Geochem.* **66**, 164–173. Available at: https://www.sciencedirect.com/science/article/pii/S0146638013002441 [Accessed April 8, 2019].

Ng S. L. and 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. *J. Paleolimnol.* **22**, 367–382.

Ning D., Zhang E., Shulmeister J., Chang J., Sun W. and Ni Z. (2019) Holocene mean annual air temperature (MAAT) reconstruction based on branched glycerol dialkyl glycerol tetraethers from Lake Ximenglongtan, southwestern China. *Org. Geochem.* **133**, 65–76. Available at: https://www.sciencedirect.com/science/article/pii/S0146638019300877?dgcid=raven_sd_via_email [Accessed May 28, 2019].

Nürnberg G. K. (1995) Quantifying anoxia in lakes. *Limnol. Oceanogr.* **40**, 1100–1111. Available at: http://doi.wiley.com/10.4319/lo.1995.40.6.1100 [Accessed November 27, 2020].

Otto-Bliesner B. L., Brady E. C., Zhao A., Brierley C. M., Axford Y., Capron E., Govin A., Hoffman J. S., Isaacs E., Kageyama M., Scussolini P., Tzedakis P. C., Williams C. J. R., Wolff E., Abe-Ouchi A., Braconnot P., Ramos Buarque S., Cao J., De Vernal A., Vittoria Guarino M., Guo C., Legrande A. N., Lohmann G., Meissner K. J., Menviel L., Morozova P. A., Nisancioglu K. H., O'Ishi R., Mélia D. S. Y., Shi X., Sicard M., Sime L., Stepanek C., Tomas R., Volodin E., Yeung N. K. H., Zhang Q., Zhang Z. and Zheng W. (2021) Large-scale features of Last Interglacial climate: Results from evaluating the lig127k simulations for the Coupled Model Intercomparison Project (CMIP6)-Paleoclimate Modeling Intercomparison Project (PMIP4). *Clim. Past* **17**, 63–94.

Overland J. E., Dethloff K., Francis J. A., Hall R. J., Hanna E., Kim S. J., Screen J. A., Shepherd T. G. and Vihma T. (2016) Nonlinear response of mid-latitude weather to the changing Arctic. *Nat. Clim. Chang.* **6**, 992–

999. Available at: http://www.nature.com/articles/nclimate3121 [Accessed April 10, 2019].

Pearson E. J., Juggins S., Talbot H. M., Weckström J., Rosén P., Ryves D. B., Roberts S. J. and Schmidt R. (2011) A lacustrine GDGT-temperature calibration from the Scandinavian Arctic to Antarctic: Renewed potential for the application of GDGT-paleothermometry in lakes. *Geochim. Cosmochim. Acta* **75**, 6225–6238. Available at: https://www.sciencedirect.com/science/article/pii/S0016703711004406 [Accessed April 19, 2018].

Pérez-Angel L. C., Sepúlveda J., Molnar P., Montes C., Rajagopalan B., Snell K., Gonzalez-Arango C. and Dildar N. (2020) Soil and Air Temperature Calibrations Using Branched GDGTs for the Tropical Andes of Colombia: Toward a Pan-Tropical Calibration. *Geochemistry, Geophys. Geosystems* **21**, 1–18. Available at: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020GC008941 [Accessed September 4, 2020].

Peterse F., Hopmans E. C., Schouten S., Mets A., Rijpstra W. I. C. and Sinninghe Damsté J. S. (2011) Identification and distribution of intact polar branched tetraether lipids in peat and soil. *Org. Geochem.* **42**, 1007–1015. Available at: https://ac.els-cdn.com/S0146638011001859/1-s2.0-S0146638011001859-main.pdf?_tid=508d3f59-9d0a-4272-ae17-de787024095c&acdnat=1537484724_77ef72a87fbe289f1ca9b3179f31b98b [Accessed September 20, 2018].

Peterse F., Meer J. Van Der, Schouten S., Weijers J. W. H. H., Fierer N., Jackson R. B., Kim J. H. and Damste J. S. S. (2012) Revised calibration of the MBT – CBT paleotemperature proxy based on branched tetraether membrane lipids in surface soils. *Geochim. Cosmochim. Acta* **96**, 215–229.

Peterse F., Vonk J. E., Holmes R. M., Giosan L., Zimov N. and Eglinton T. I. (2014) Branched glycerol dialkyl glycerol tetraethers in Arctic lake sediments: Sources and implications for paleothermometry at high latitudes. *J. Geophys. Res. Biogeosciences* **119**, 1738–1754. Available at: http://doi.wiley.com/10.1002/2014JG002639 [Accessed June 26, 2018].

Pithan F. and Mauritsen T. (2014) Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nat. Geosci.* **7**, 181–184. Available at: http://www.nature.com/articles/ngeo2071 [Accessed April 16, 2019].

Post E., Alley R. B., Christensen T. R., Macias-Fauria M., Forbes B. C., Gooseff M. N., Iler A., Kerby J. T., Laidre K. L., Mann M. E., Olofsson J., Stroeve J. C., Ulmer F., Virginia R. A. and Wang M. (2019) The polar regions in a 2°C warmer world. *Sci. Adv.* **5**. Available at: https://www.science.org/doi/abs/10.1126/sciadv.aaw9883 [Accessed October 15, 2021].

Prahl F. G. and Wakeham S. G. (1987) Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. *Nature* **330**, 367–369.

Qian S., Yang H., Dong C., Wang Y., Wu J., Pei H., Dang X., Lu J., Zhao S. and Xie S. (2019) Rapid response of fossil tetraether lipids in lake sediments to seasonal environmental variables in a shallow lake in central China: Implications for the use of tetraether-based proxies. *Org. Geochem.* **128**, 108–121.

Raberg J. H., Harning D. J., Crump S. E., De Wet G., Blumm A., Kopf S., Geirsdóttir Á., Miller G. H. and Sepúlveda J. (2021) Revised fractional abundances and warm-season temperatures substantially improve brGDGT calibrations in lake sediments. *Biogeosciences* **18**, 3579–3603. Available at: https://doi.org/10.5194/bg-18-3579-2021 [Accessed June 23, 2021].

Ramos Emmendorfer L. and de Paula Canuto A. M. (2021) A generalized average linkage criterion for Hierarchical Agglomerative Clustering. *Appl. Soft Comput.* **100**, 106990.

Raynolds M. K., Comiso J. C., Walker D. A. and Verbyla D. (2008) Relationship between satellite-derived land surface temperatures, arctic vegetation types, and NDVI. *Remote Sens. Environ.* **112**, 1884–1894. Available at: https://ac.els-cdn.com/S0034425707004294/1-s2.0-S0034425707004294-main.pdf?_tid=0a3d1de6-0125-44f7-8cf5-877022f4a562&acdnat=1525461946_92289516480b428530709c423dff8558 [Accessed May 4, 2018].

Reizer J., Grossowicz N. and Barenholz Y. (1985) The effect of growth temperature on the thermotropic behavior of the membranes of a thermophilic Bacillus. Composition-structure-function relationships. *BBA - Biomembr.* **815**, 268–280.

Russell J. M., Hopmans E. C., Loomis S. E., Liang J. and Sinninghe Damsté J. S. (2018) Distributions of 5- and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of temperature, pH, and new lacustrine paleotemperature calibrations. *Org. Geochem.* **117**, 56–69. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0146638017304394 [Accessed September 17, 2018].

Schouten S., Hopmans E. C., Schefuß E. and Sinninghe Damsté J. S. (2002) Distributional variations in marine crenarchaeotal membrane lipids: A new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X02009792 [Accessed October 11, 2018].

Schouten S., Hopmans E. C. and Sinninghe Damsté J. S. (2013) The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Org. Geochem.* **54**, 19–61. Available at: http://dx.doi.org/10.1016/j.orggeochem.2012.09.006 [Accessed April 18, 2021].

Schouten S., Van Der Meer M. T. J., Hopmans E. C., Rijpstra W. I. C., Reysenbach A. L., Ward D. M. and Damsté J. S. S. (2007) Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in hot springs of Yellowstone National Park. *Appl. Environ. Microbiol.* **73**, 6181–6191. Available at: https://journals.asm.org/journal/aem [Accessed July 28, 2021].

Schwarz G. (1978) Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464. Available at: https://projecteuclid.org/euclid.aos/1176344136 [Accessed January 3, 2021].

Serreze M. C. and Barry R. G. (2011) Processes and impacts of Arctic amplification: A research synthesis. *Glob. Planet. Change* **77**, 85–96. Available at: https://www.colorado.edu/geography/class_homepages/geog_4271_f13/readings/reading_9_serreze_barry_arctic_amp.pdf [Accessed November 26, 2017].

Shanahan T. M., Hughen K. A. and Van Mooy B. A. S. (2013) Temperature sensitivity of branched and isoprenoid GDGTs in Arctic lakes. *Org. Geochem.* **64**, 119–128. Available at: http://dx.doi.org/10.1016/j.orggeochem.2013.09.010 [Accessed April 3, 2019].

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). *Geochim. Cosmochim. Acta* **186**, 13–31. Available at: https://www.sciencedirect.com/science/article/pii/S0016703716302022 [Accessed March 11, 2019].

Sinninghe Damsté J. S., Hopmans E. C., Pancost R. D., Schouten S. and Geenevasen J. A. J. (2000) Newly discovered non-isoprenoid glycerol dialkyl glycerol tetraether lipids in sediments. *Chem. Commun.* **0**, 1683–1684. Available at: http://xlink.rsc.org/?DOI=b004517i [Accessed November 15, 2018].

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. and 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. *Org. Geochem.* **124**, 63–76. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0146638018301578 [Accessed September 19, 2018].

Stager J. C., Ryves D. B., King C., Madson J., Hazzard M., Neumann F. H. and Maud R. (2013) Late Holocene precipitation variability in the summer rainfall region of South Africa. *Quat. Sci. Rev.* **67**, 105–120.

Stefanescu I. C., Shuman B. N. and Tierney J. E. (2021) Temperature and water depth effects on brGDGT distributions in sub-alpine lakes of mid-latitude North America. *Org. Geochem.* **152**, 104174. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0146638020302096 [Accessed February 1, 2021].

Sturm M., Douglas T., Racine C. and Liston G. E. (2005) Changing snow and shrub conditions affect albedo with global implications. *J. Geophys. Res.* **110**, G01004. Available at: http://doi.wiley.com/10.1029/2005JG000013 [Accessed April 12, 2018].

Sun Q., Chu G., Liu M., Xie M., Li S., Ling Y., Wang X., Shi L., Jia G. and Lü H. (2011) Distributions and temperature dependence of branched glycerol dialkyl glycerol tetraethers in recent lacustrine sediments from China and Nepal. *J. Geophys. Res. Biogeosciences* **116**, G01008. Available at: http://doi.wiley.com/10.1029/2010JG001365 [Accessed November 8, 2018].

Super J. R., Chin K., Pagani M., Li H., Tabor C., Harwood D. M. and Hull P. M. (2018) Late Cretaceous climate in the Canadian Arctic: Multi-proxy constraints from Devon Island. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **504**, 1–22. Available at: https://www.sciencedirect.com/science/article/pii/S0031018217307666 [Accessed April 8, 2019].

Suzuki R. and Shimodaira H. (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542.

Team R Development Core (2021) A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **2**, https://www.R-project.org. Available at: http://www.r-project.org/.

Thomas E. K., Briner J. P., Ryan-Henry J. J. and Huang Y. (2016) A major increase in winter snowfall during the middle Holocene on western Greenland caused by reduced sea ice in Baffin Bay and the Labrador Sea. *Geophys. Res. Lett.* **43**, 5302–5308. Available at: http://doi.wiley.com/10.1002/2016GL068513 [Accessed December 22, 2017].

Thomas E. K., Castañeda I. S., McKay N. P., Briner J. P., Salacup J. M., Nguyen K. Q. and Schweinsberg A. D. (2018) A Wetter Arctic Coincident With Hemispheric Warming 8,000 Years Ago. *Geophys. Res. Lett.* **45**, 10,637-10,647. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL079517 [Accessed April 8, 2019].

Tierney J. E. E., Russell J. M. M., Eggermont H., Hopmans E. C. C., Verschuren D. and Sinninghe Damsté J. S. S. (2010) Environmental controls on branched tetraether lipid distributions in tropical East African lake sediments. *Geochim. Cosmochim. Acta* **74**, 4902–4918. Available at: https://www.sciencedirect.com/science/article/pii/S0016703710003236 [Accessed September 17, 2018].

Tierney J. E., Poulsen C. J., Montañez I. P., Bhattacharya T., Feng R., Ford H. L., Hönisch B., Inglis G. N., Petersen S. V., Sagoo N., Tabor C. R., Thirumalai K., Zhu J., Burls N. J., Foster G. L., Goddéris Y., Huber B. T., Ivany L. C., Turner S. K., Lunt D. J., McElwain J. C., Mills B. J. W., Otto-Bliesner B. L., Ridgwell A. and Zhang Y. G. (2020) Past climates inform our future. *Science (80-.).* **370**. Available at: https://science.sciencemag.org/content/370/6517/eaay3701 [Accessed August 11, 2021].

Tierney J. E. and Russell J. M. (2009) Distributions of branched GDGTs in a tropical lake system: Implications for lacustrine application of the MBT/CBT paleoproxy. *Org. Geochem.* **40**, 1032–1036. Available at: https://www.sciencedirect.com/science/article/pii/S0146638009000989 [Accessed April 11, 2019].

Tierney J. E., Schouten S., Pitcher A., Hopmans E. C. and Sinninghe

Damsté J. S. (2012) Core and intact polar glycerol dialkyl glycerol tetraethers (GDGTs) in Sand Pond, Warwick, Rhode Island (USA): Insights into the origin of lacustrine GDGTs. *Geochim. Cosmochim. Acta* **77**, 561–581. Available at: www.elsevier.com/locate/gca [Accessed September 20, 2018].

Tyler J. J., Nederbragt A. J., Jones V. J. and 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. *J. Geophys. Res.* **115**, G01015. Available at: http://doi.wiley.com/10.1029/2009JG001109 [Accessed October 14, 2020].

Venables W. N. and Ripley B. D. (2002) *Modern Applied Statistics with S.* Fourth., Springer, New York.

Walker D. A., Raynolds M. K., Daniëls F. J. A., Einarsson E., Elvebakk A., Gould W. A., Katenin A. E., Kholod S. S., Markon C. J., Melnikov E. S., Moskalenko N. G., Talbot S. S., Yurtsev B. A. (†. and The other members of the CAVM Team (2005) The Circumpolar Arctic vegetation map. *J. Veg. Sci.* **16**, 267–282. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1654-1103.2005.tb02365.x [Accessed April 20, 2021].

Wang H., An Z., Lu H., Zhao Z. and Liu W. (2020) Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence. *Quat. Sci. Rev.* **231**, 106172.

Wang H., Liu W., He Y., Zhou A., Zhao H., Liu H., Cao Y., Hu J., Meng B., Jiang J., Kolpakova M., Krivonogov S. and Liu Z. (2021) Salinity-controlled isomerization of lacustrine brGDGTs impacts the associated MBT5ME' terrestrial temperature index. *Geochim. Cosmochim. Acta* **305**, 33–48.

Weber Y., Damsté J. S. S., Zopfi J., De Jonge C., Gilli A., Schubert C. J., Lepori F., Lehmann M. F. and Niemann H. (2018) Redox-dependent niche differentiation provides evidence for multiple bacterial sources of glycerol tetraether lipids in lakes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10926–10931. Available at: https://www.pnas.org/content/115/43/10926 [Accessed January 29, 2020].

Weber Y., De Jonge C., Rijpstra W. I. C., Hopmans E. C., Stadnitskaia A., Schubert C. J., Lehmann M. F., Sinninghe Damsté J. S. and Niemann H. (2015) Identification and carbon isotope composition of a novel branched GDGT isomer in lake sediments: Evidence for lacustrine branched GDGT production. *Geochim. Cosmochim. Acta* **154**, 118–129.

Weijers J. W. H., Schefuß E., Schouten S. and Damsté J. S. S. S. (2007a) Coupled thermal and hydrological evolution of tropical Africa over the last deglaciation. *Science (80-.).* **315**, 1701–1704. Available at: https://science.sciencemag.org/content/315/5819/1701 [Accessed August 4, 2021].

Weijers J. W. H., Schouten S., Donker J. C. Van Den, Hopmans E. C., van den Donker J. C., Hopmans E. C. and Sinninghe Damsté J. S. (2007b) Environmental controls on bacterial tetraether membrane lipid distribution in soils. *Geochim. Cosmochim. Acta* **71**, 703–713. Available at:

https://ac.els-cdn.com/S0016703706021090/1-s2.0-S0016703706021090-main.pdf?_tid=5e63c70e-dc48-11e7-90d3-00000aab0f01&acdnat=1512759084_77c5f06f28c2e88c1cf1d71c3fb8d2ac [Accessed December 8, 2017].

Weijers J. W. H., Schouten S., Sluijs A., Brinkhuis H. and Sinninghe Damsté J. S. (2007c) Warm arctic continents during the Palaeocene-Eocene thermal maximum. *Earth Planet. Sci. Lett.* **261**, 230–238. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X07004190 [Accessed November 11, 2018].

Weijers J. W. H., Schouten S., Spaargaren O. C. and 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. *Org. Geochem.* **37**, 1680–1693. Available at: https://www.sciencedirect.com/science/article/pii/S0146638006001835 [Accessed April 10, 2019].

de Wet G. A., Castañeda I. S., DeConto R. M. and 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 Planet. Sci. Lett.* **436**, 56–63. Available at: https://www.sciencedirect.com/science/article/pii/S0012821X15007840?via%3Dihub [Accessed April 8, 2019].

Wetzel R. G. (2001) Salinity of Inland Waters. In *Limnology* (ed. R. G. Wetzel). Academic Press, San Diego. pp. 169–186.

Woltering M., Werne J. P., Kish J. L., Hicks R., Sinninghe Damsté J. S. and 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. *Geochim. Cosmochim. Acta* **87**, 136–153.

Wörmer L., Lipp J. S., Schröder J. M. and Hinrichs K. U. (2013) Application of two new LC-ESI-MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. *Org. Geochem.* **59**, 10–21. Available at: https://www.sciencedirect.com/science/article/pii/S0146638013000557 [Accessed June 13, 2018].

Wu J., Yang H., Pancost R. D., Naafs B. D. A., Qian S., Dang X., Sun H., Pei H., Wang R., Zhao S. and Xie S. (2021) Variations in dissolved O2 in a Chinese lake drive changes in microbial communities and impact sedimentary GDGT distributions. *Chem. Geol.* **579**, 120348. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0009254121002928 [Accessed June 1, 2021].

Wunsch C. (1980) Meridional heat flux of the North Atlantic Ocean. *Proc. Natl. Acad. Sci.* **77**, 5043–5047.

Xiao W., Wang Y., Liu Y., Zhang X., Shi L. and 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. Available at:

https://www.biogeosciences.net/17/2135/2020/ [Accessed July 19, 2020].

Xie S., Liu X. L., Schubotz F., Wakeham S. G. and Hinrichs K. U. (2014) Distribution of glycerol ether lipids in the oxygen minimum zone of the Eastern Tropical North Pacific Ocean. *Org. Geochem.* **71**, 60–71.

Yang H., Ding W., Zhang C. L., Wu X., Ma X., He G., Huang J. and Xie S. (2011) Occurrence of tetraether lipids in stalagmites: Implications for sources and GDGT-based proxies. *Org. Geochem.* **42**, 108–115.

Yang X., Wang S., Kamenik C., Schmidt R., Shen J., Zhu L. and Li S. (2004) Diatom assemblages and quantitative reconstruction for paleosalinity from a sediment core of Chencuo Lake, southern Tibet. *Sci. China, Ser. D Earth Sci.* **47**, 522–528. Available at: https://link.springer.com/article/10.1360/02yd0142 [Accessed December 28, 2020].

Yao Y., Zhao J., Bauersachs T. and Huang Y. (2019) Effect of water depth on the TEX 86 proxy in volcanic lakes of northeastern China. *Org. Geochem.* **129**, 88–98.

Yao Y., Zhao J., Vachula R. S., Werne J. P., Wu J., Song X. and 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. *Org. Geochem.* **147**, 104076. Available at: https://doi.org/10.1016/j.orggeochem.2020.104076 [Accessed August 24, 2020].

Yuk H. G. and 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. *Appl. Environ. Microbiol.* **70**, 3500–3505. Available at: http://aem.asm.org/ [Accessed November 22, 2020].

Zell C., Kim J.-H., Moreira-Turcq P., Abril G., Hopmans E. C., Bonnet M.-P., Sobrinho R. L. and Damsté J. S. S. (2013) Disentangling the origins of branched tetraether lipids and crenarchaeol in the lower Amazon River: Implications for GDGT-based proxies. *Limnol. Oceanogr.* **58**, 343–353. Available at: https://onlinelibrary.wiley.com/doi/full/10.4319/lo.2013.58.1.0343 [Accessed September 23, 2021].

Zhang Z. X., Li J., Chen Z., Sun Z., Yang H., Fu M. and Peng X. (2020) The effect of methane seeps on the bacterial tetraether lipid distributions at the Okinawa Trough. *Mar. Chem.* **225**, 103845.

Zhao B., Castañeda I. S., Bradley R. S., Salacup J. M., de Wet G. A., Daniels W. C. and Schneider T. (2021) Development of an in situ branched GDGT calibration in Lake 578, southern Greenland. *Org. Geochem.* **152**, 104168.

Zhao J., Huang Y., Yao Y., An Z., Zhu Y., Lu H. and Wang Z. (2020) Calibrating branched GDGTs in bones to temperature and precipitation: Application to Alaska chronological sequences. *Quat. Sci. Rev.* **240**, 106371.

Zheng Y., Pancost R. D., Liu X., Wang Z., Naafs B. D. A., Xie X., Liu Z., Yu X. and Yang H. (2017) Atmospheric connections with the North Atlantic enhanced the deglacial warming in northeast China. *Geology* **45**, 1031–1034. Available at: http://pubs.geoscienceworld.org/gsa/geology/article-pdf/45/11/1031/3948872/1031.pdf [Accessed September 27, 2020].

Zhu C., Weijers J. W. H., Wagner T., Pan J. M., Chen J. F. and Pancost R. D. (2011) Sources and distributions of tetraether lipids in surface sediments across a large river-dominated continental margin. *Org. Geochem.* **42**, 376–386. Available at: https://www.sciencedirect.com/science/article/pii/S0146638011000325 [Accessed April 16, 2019].

8 Appendices

8.1 Appendix 1: Chapter 2 Supplemental Information

8.1.1 True color and bathymetric maps

True color and bathymetric maps are provided below in alphabetical order for all lakes surveyed. A drone survey was not performed for Lake ARQ. However, an approximate bathymetric map based off satellite imagery is provided. Geospatial support for the ARQ imagery was provided by the Polar Geospatial Center under NSF-OPP awards 1043681 and 1559691. Full resolution files from drone surveys are available in an online data repository (Section 8.1.2).

Figure 8-1: Lake 3LN drone imagery. Imagery collected August 23, 2019.

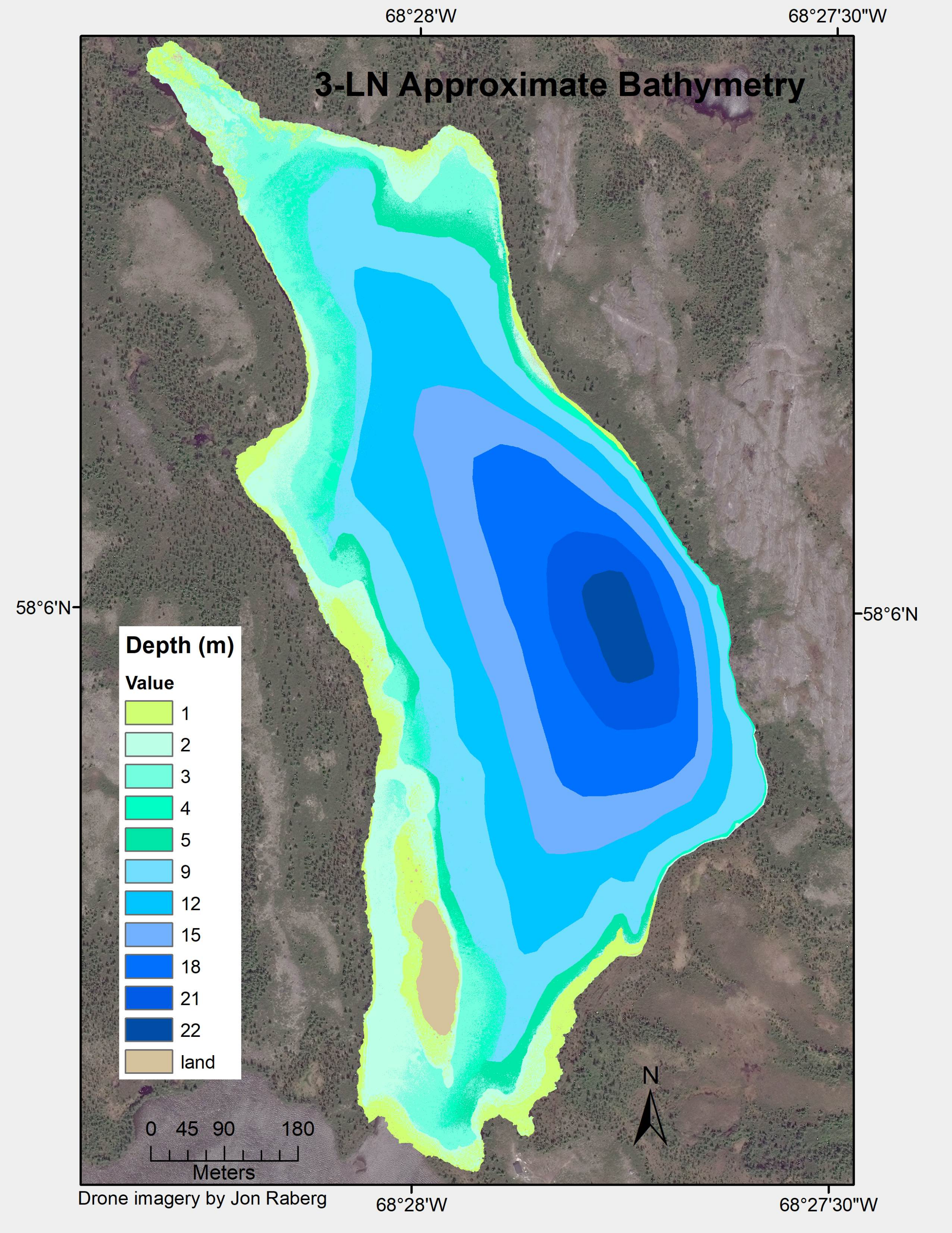


Figure 8-2: Lake 3LN approximate bathymetry. Figure courtesy of Martha Raynolds.

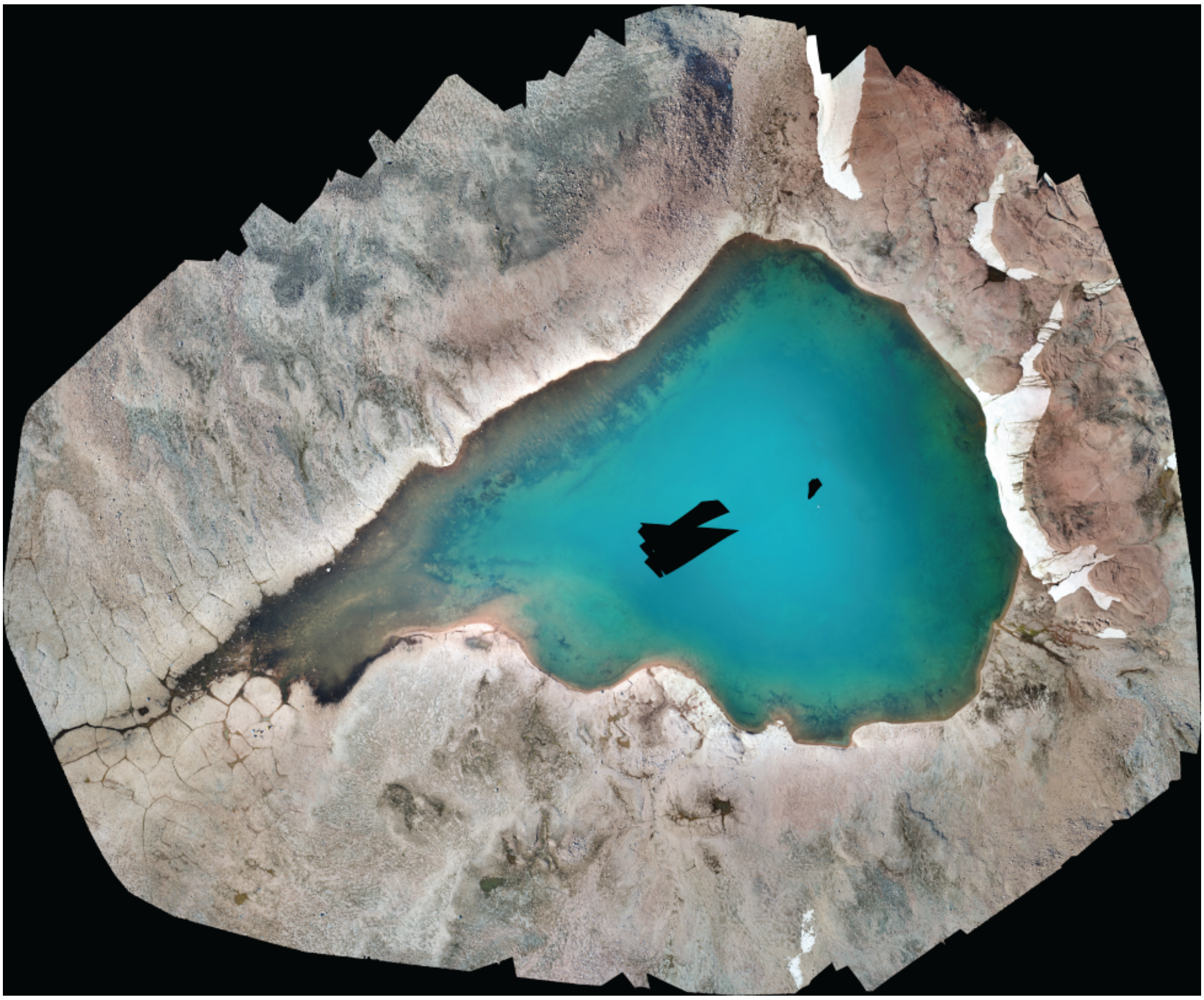


Figure 8-3: Lake AFR drone imagery. Imagery collected on August 14, 2019.

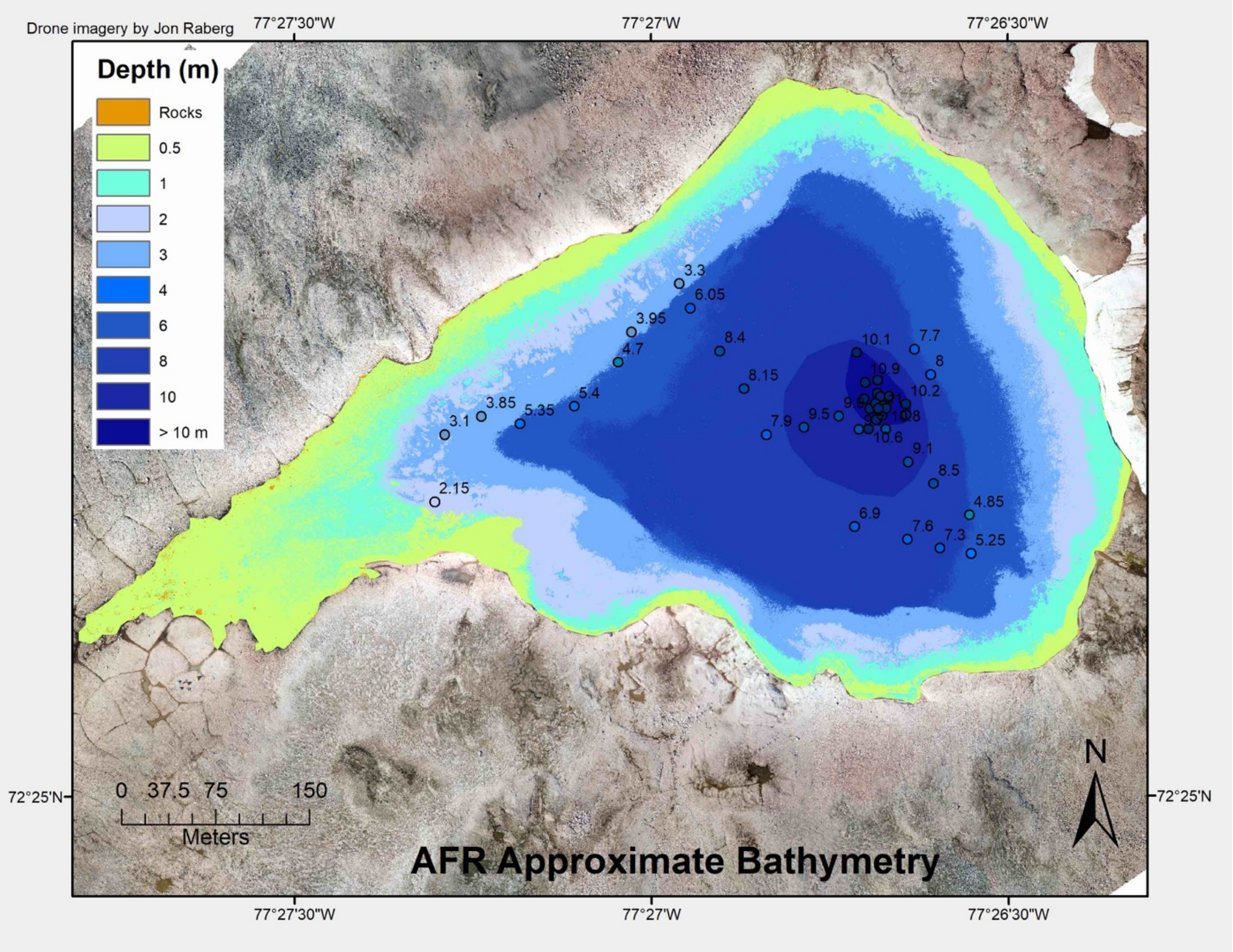


Figure 8-4: Lake AFR approximate bathymetry. Points show manually measured depths in meters. Figure courtesy of Martha Raynolds.

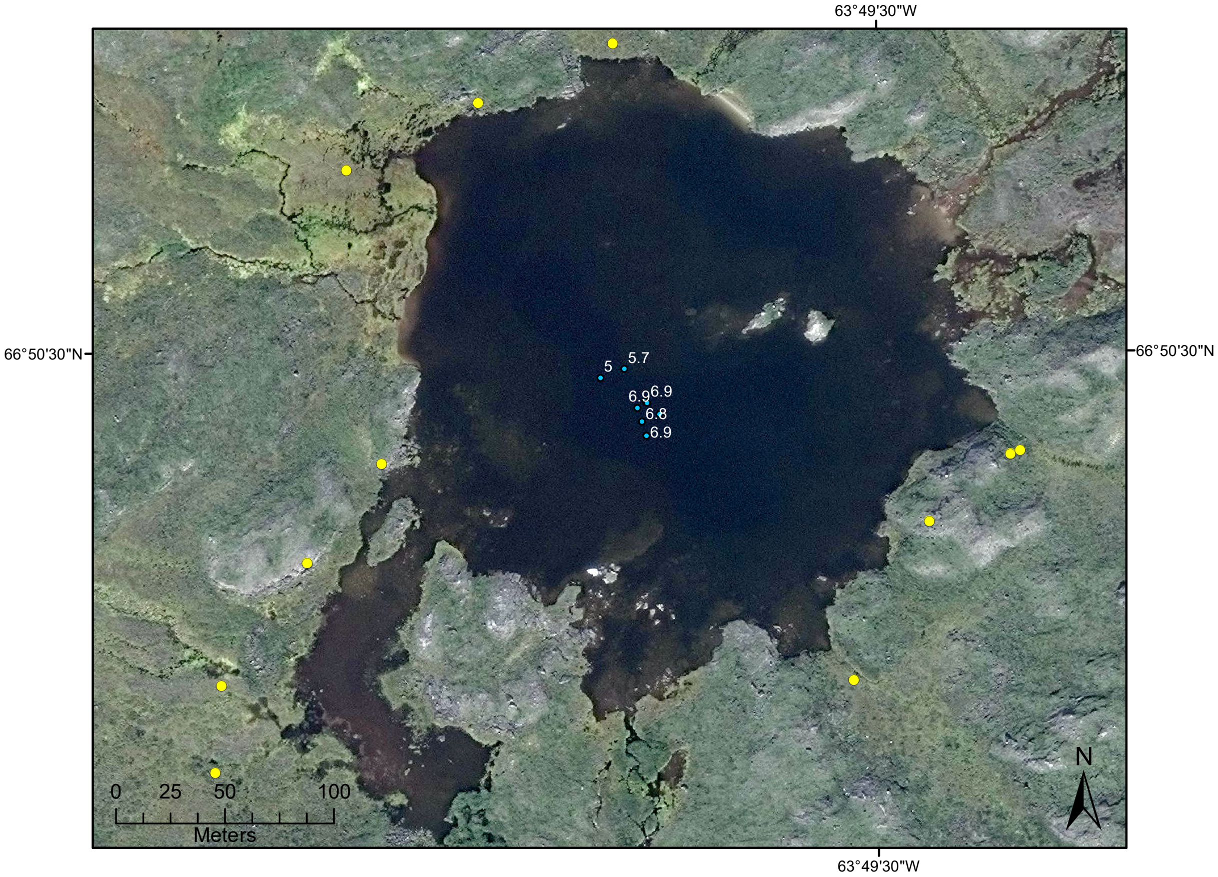


Figure 8-5: Lake ARQ satellite imagery. Points in show manually measured depths in meters. (Yellow points represent vegetation plots.) Figure courtesy of Martha Raynolds. Imagery (summer of 2020) was provided by the Polar Geospatial Center.

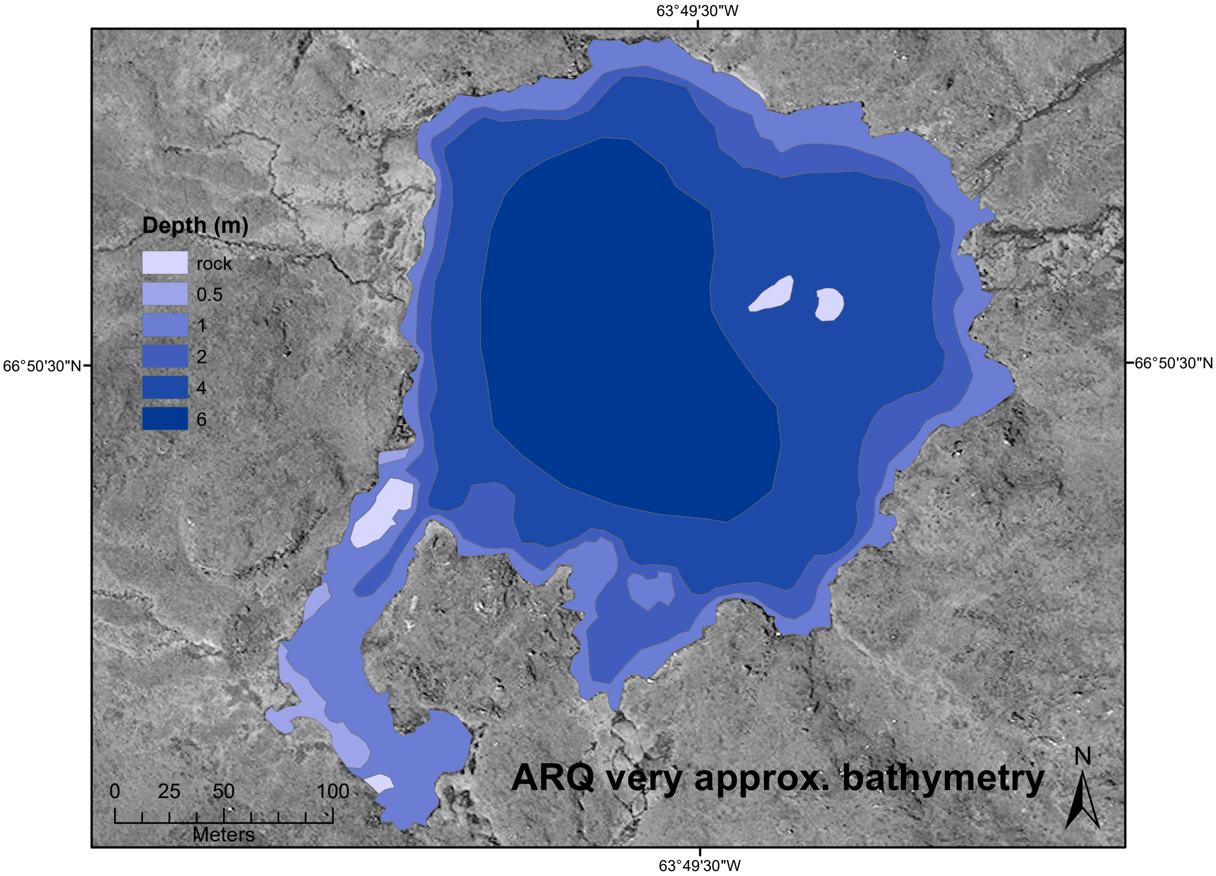


Figure 8-6: Lake ARQ very approximate bathymetry. Figure courtesy of Martha Raynolds.

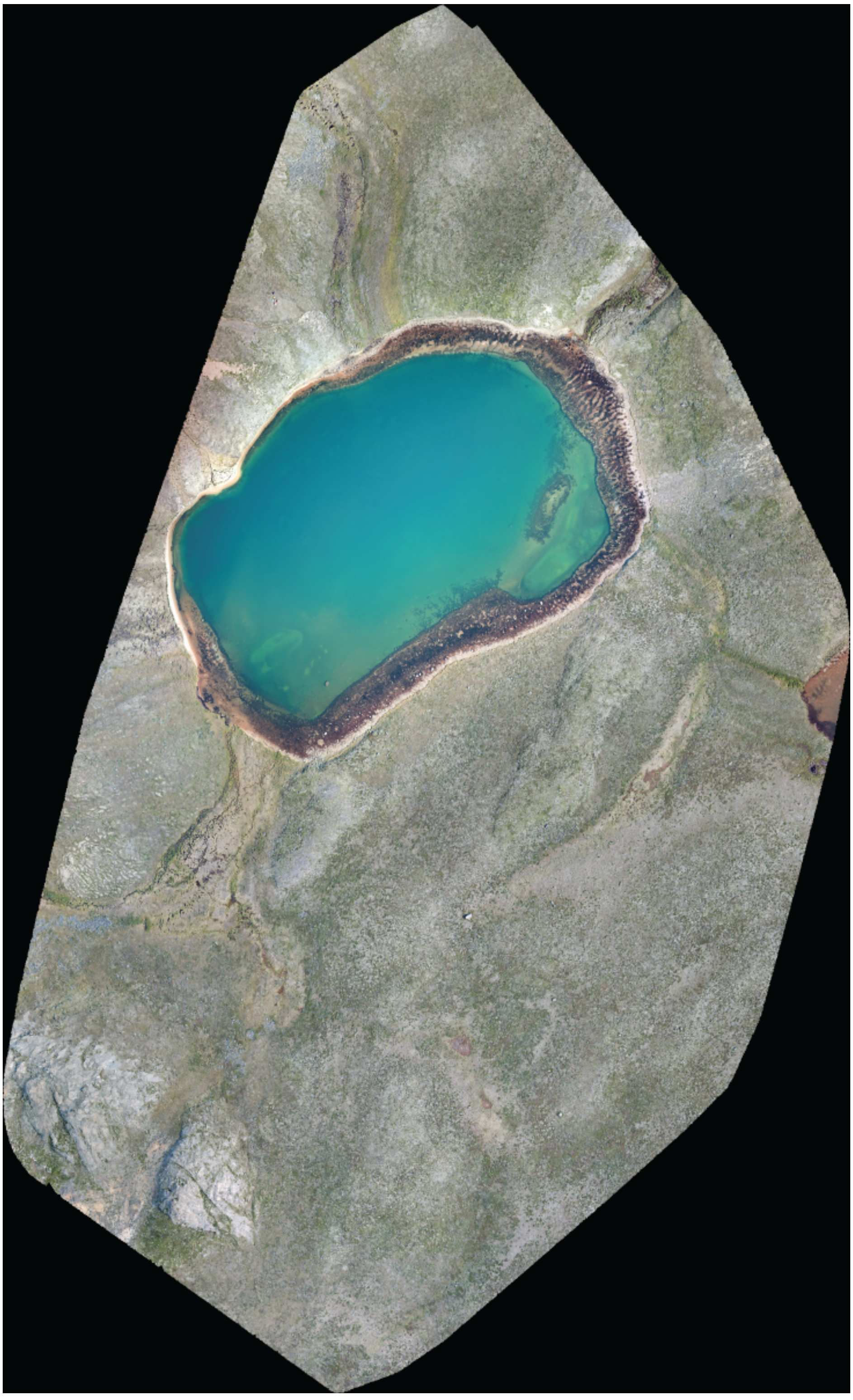


Figure 8-7: Lake BIR drone imagery. Imagery collected on July 30, 2018.

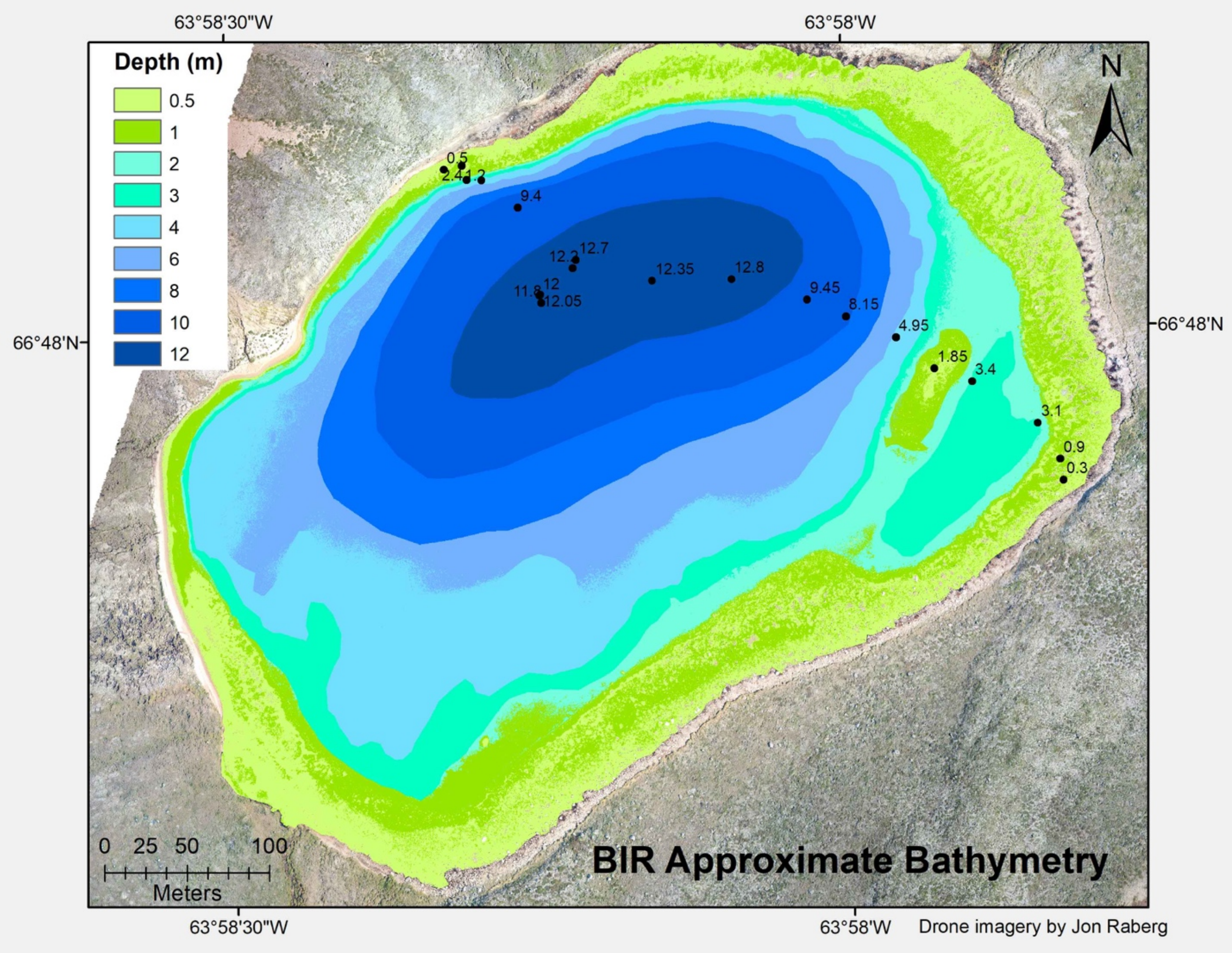


Figure 8-8: Lake BIR approximate bathymetry. Points show manually measured depths in meters. Figure courtesy of Martha Raynolds.

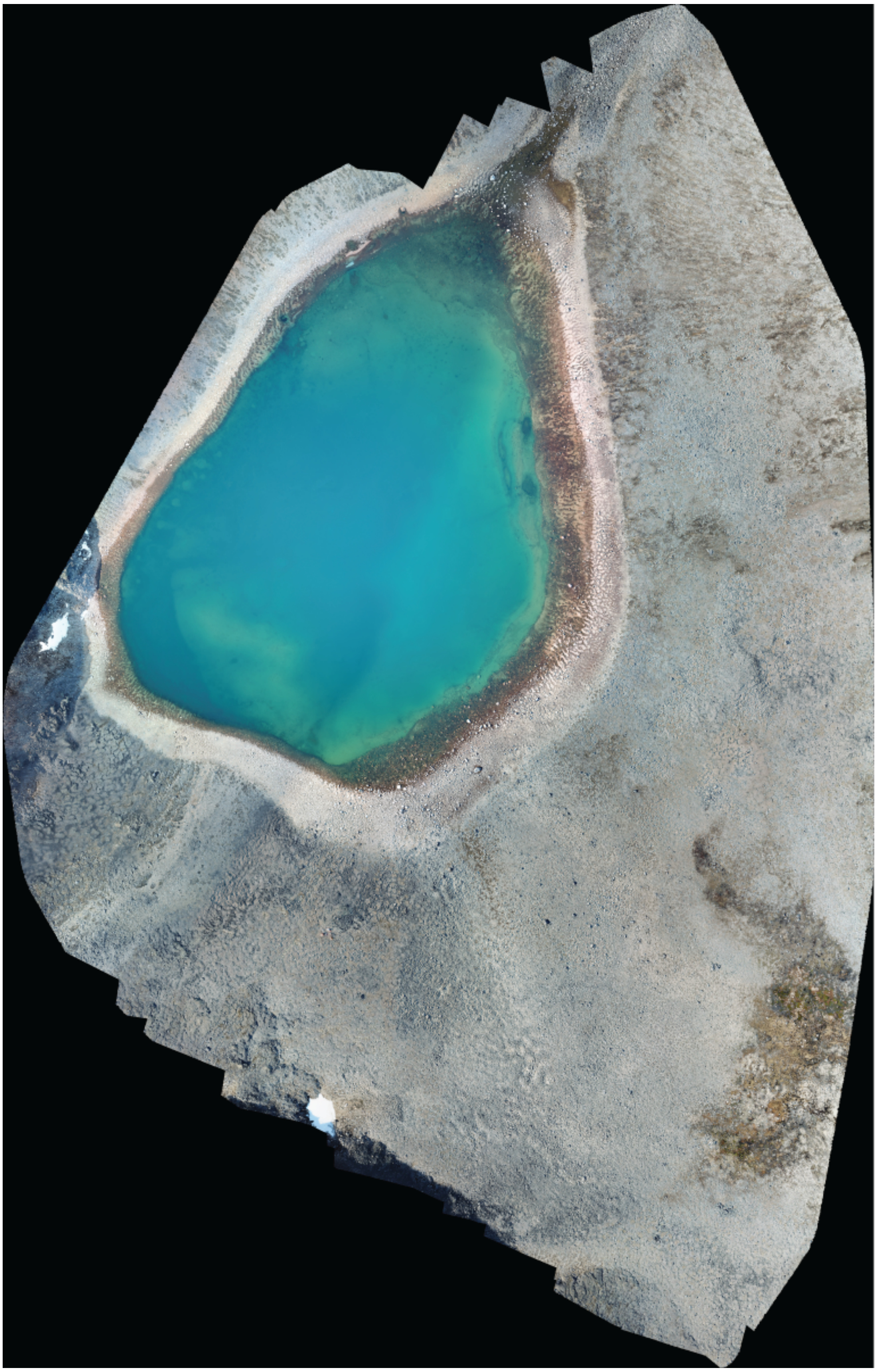


Figure 8-9: Lake BRO drone imagery. Imagery collected on August 16, 2018.

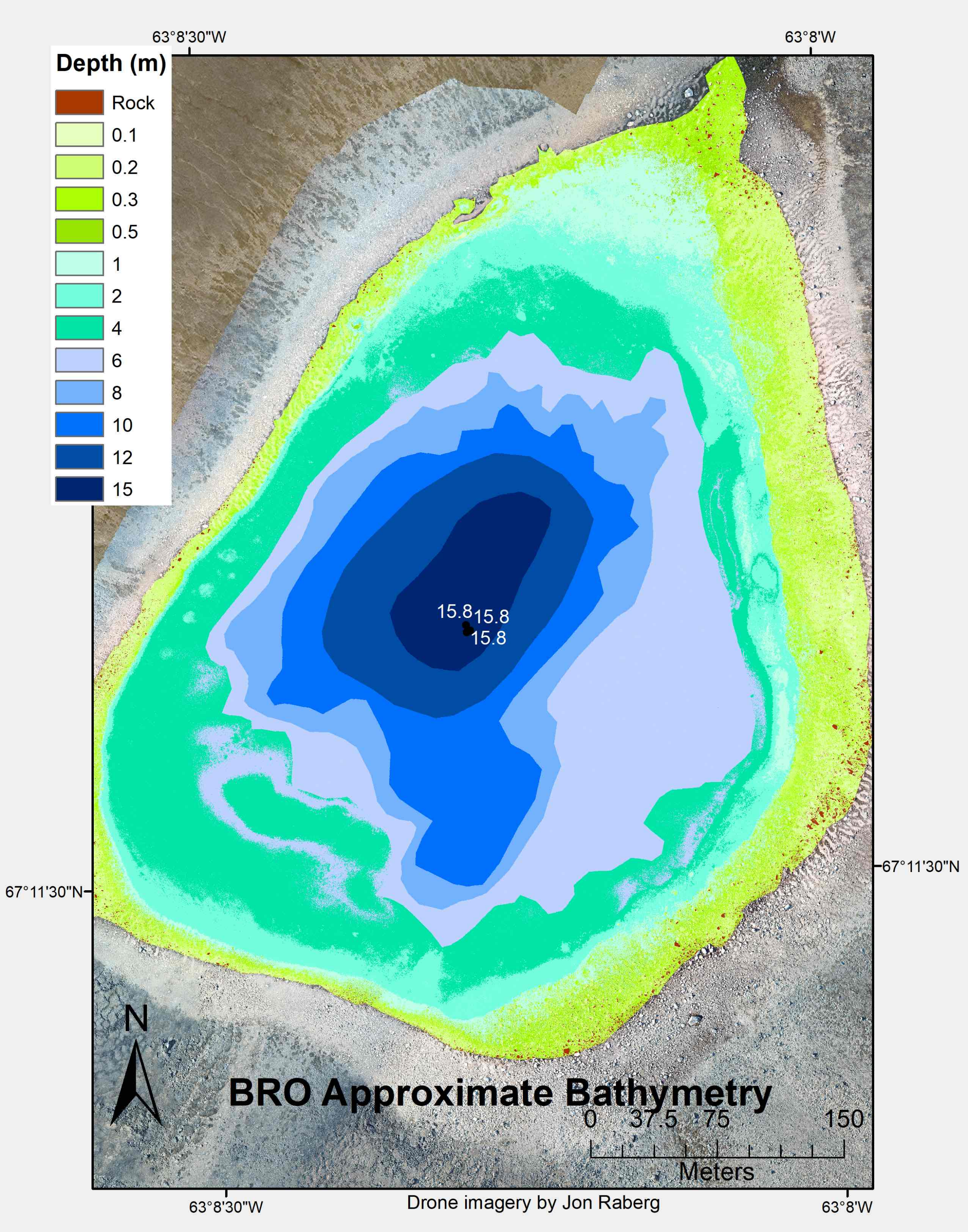


Figure 8-10: Lake BRO approximate bathymetry. Points show manually measured depths in meters. Figure courtesy of Martha Raynolds.

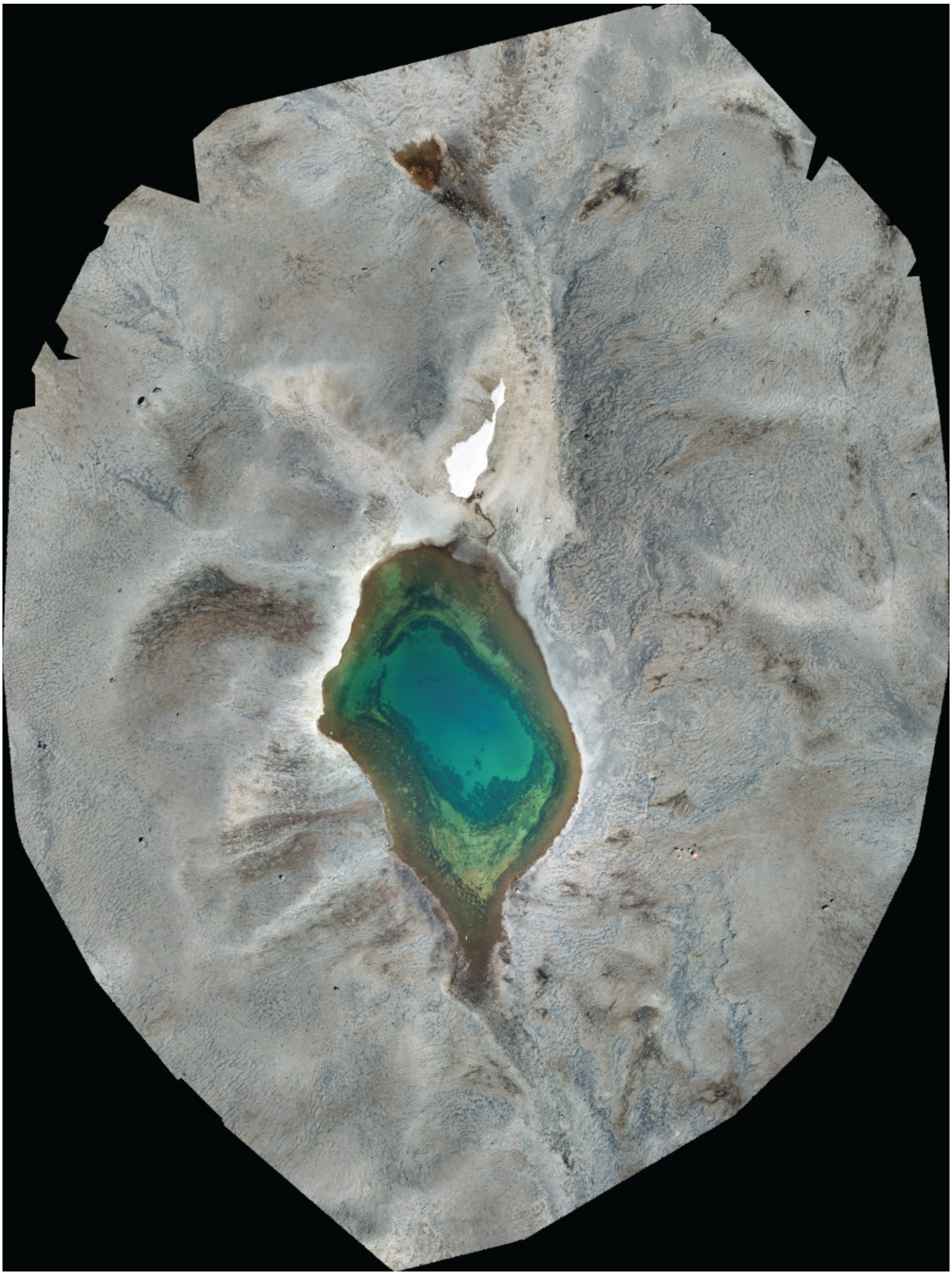
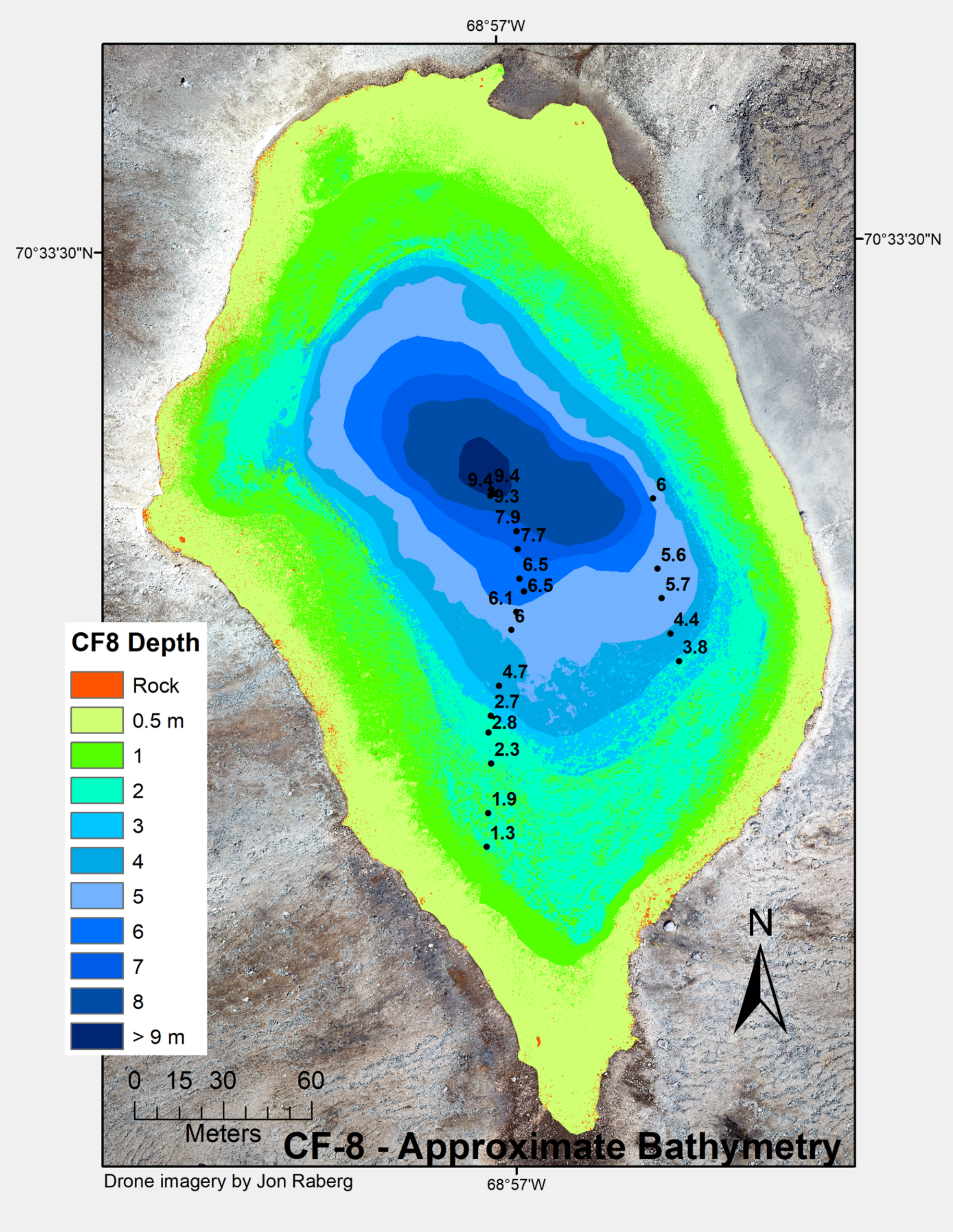


Figure 8-11: Lake CF8 drone imagery. Imagery collected on August 1, 2019.



CF-8 - Approximate Bathymetry

Drone imagery by Jon Raberg

Figure 8-12: Lake CF8 approximate bathymetry. Points show manually measured depths in meters. Figure courtesy of Martha Raynolds.

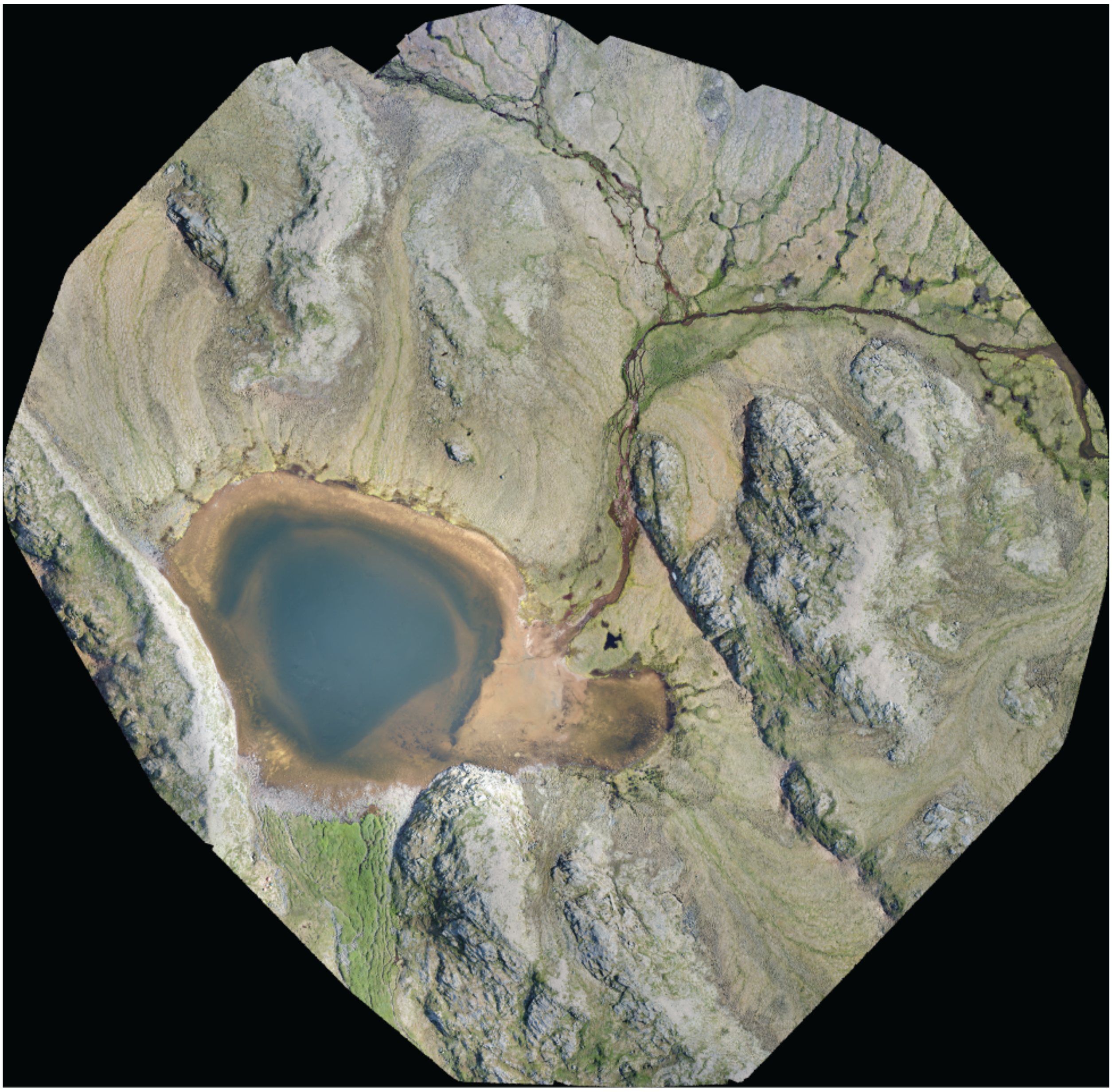


Figure 8-13: Lake QPT drone imagery. Imagery collected on July 22, 2018.

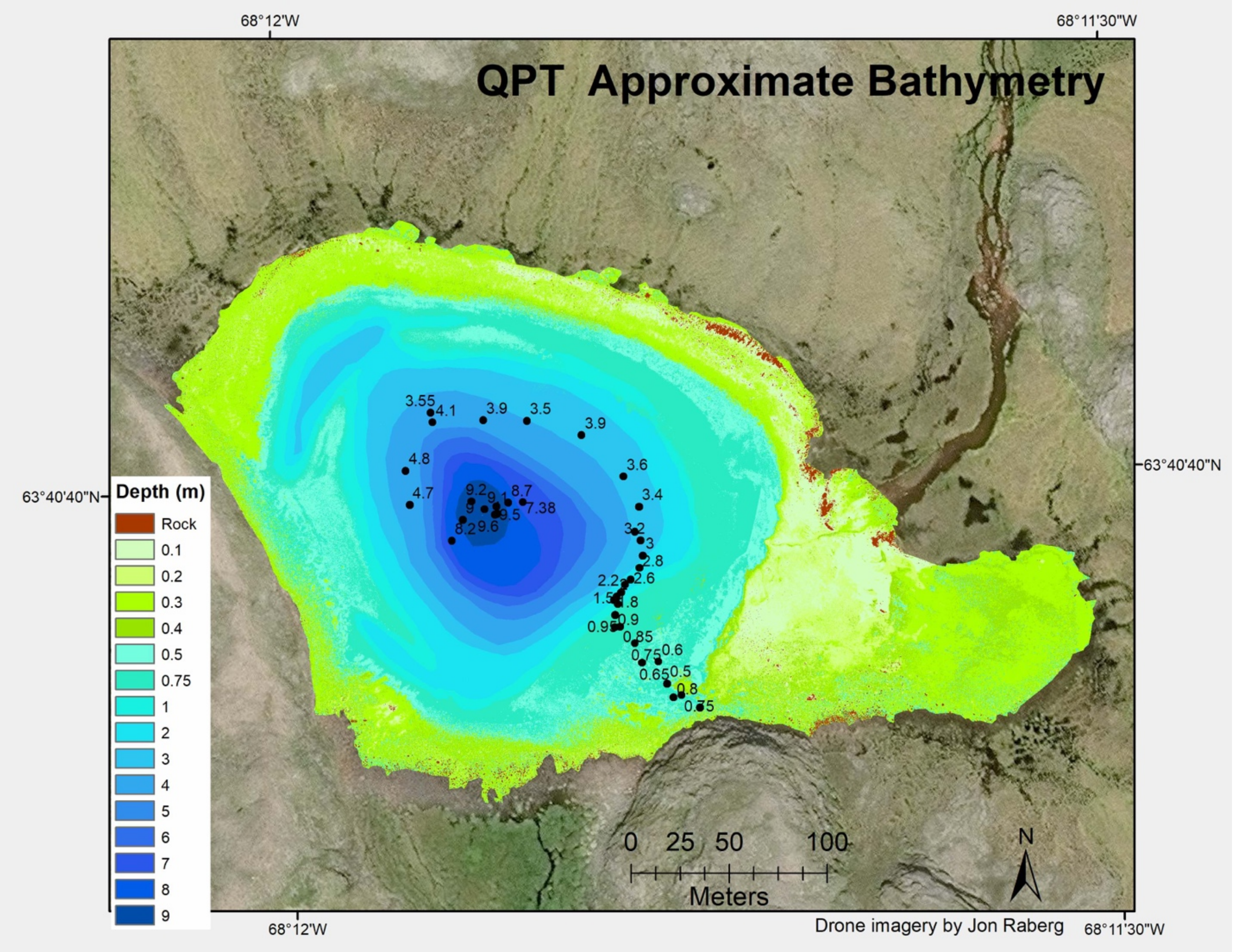


Figure 8-14: Lake QPT approximate bathymetry. Points show manually measured depths in meters. Figure courtesy of Martha Raynolds.

8.1.2 Supplementary data

All drone imagery is available at https://doi.org/10.18739/A2FN10T2V. Water chemistry measurements are available at https://doi.org/10.18739/A29Z90D0D (PACEMAP project), https://doi.org/10.18739/A26688K7F (ILLUME project). Temperature logger data is available at https://doi.org/10.18739/A22F7JS2T (PACEMAP project) and https://doi.org/10.18739/A2XP6V46R (ILLUME project), which will be updated periodically as new data is collected.

8.2 Appendix 2: Chapter 3 Supplemental Information

8.2.1 Appendix to the main text

Full and schematic structures of the 15 commonly measured brGDGTs are provided in Fig. 8-15. Table 8-1 details equations for calculating FAs within the structural sets.

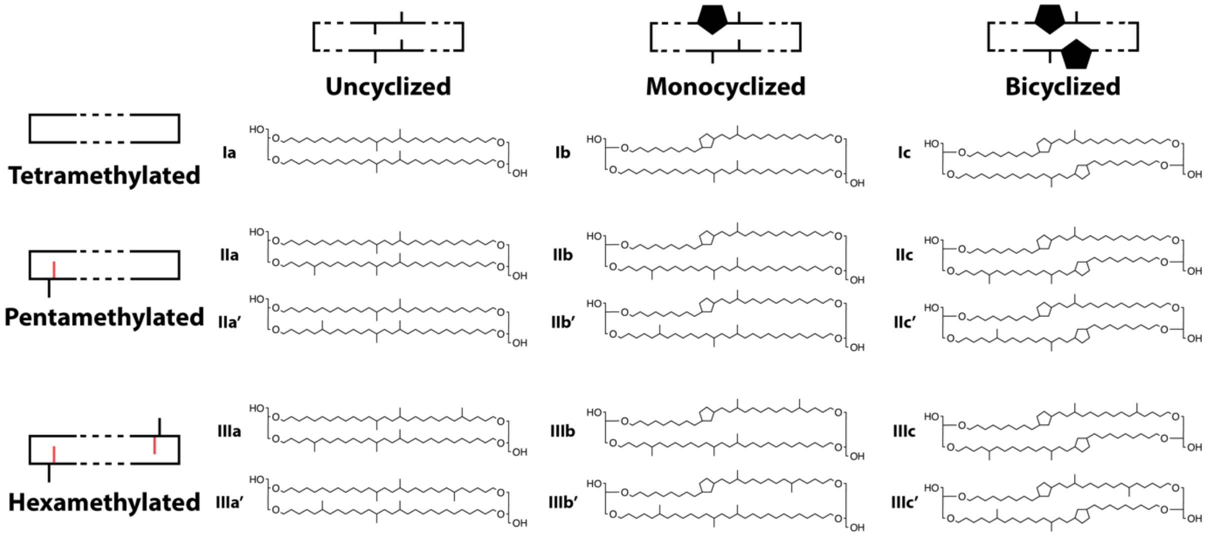


Figure 8-15: Structures of the 15 commonly measured brGDGTs. Full structures and schematic representations are provided, with C6 methylations denoted in red in the schematics.

Table 8-1: Equations for calculating FAs within brGDGT subsets. The variables 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).

(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’}	$\text{f}xy_{[(sub)set\ name]} = xy/\sum S$
MC-5Me+	S = {Ia, Ib, Ic, IIa, IIb, IIc, IIIa, IIIb, IIIc}	
MC-6Me+	S = {Ia, Ib, Ic, IIa’, IIb’, IIc’, IIIa’, IIIb’, IIIc’}	
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’}	$\text{f}xy_{[(sub)set\ name]} = xy/\sum S_y$
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’} S_c = {Ic, IIc’, IIIc’}	
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	*Use Meth-5Me+ and Meth-6Me FAs*	
CI	S_{I} = {Ia, Ib, Ic} S_{II} = {IIa, IIb, IIc, IIa’, IIb’, IIc’} S_{III} = {IIIa, IIIb, IIIc, IIIa’, IIIb’, IIIc’}	$\text{f}xy_{[(sub)set\ name]} = xy/\sum S_x$
Cyc-5Me	S_{I} = {Ia, Ib, Ic} S_{II} = {IIa, IIb, IIc} S_{III} = {IIIa, IIIb, IIIc}	
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’} S_{IIb} = {IIb, IIb’}; S_{IIIb} = {IIIb, IIIb’} S_{IIc} = {IIc, IIc’}; S_{IIIc} = {IIIc, IIIc’}	$\text{f}xy_{[(sub)set\ name]} = xy/\sum S_{xy}$

We used the following previously-defined brGDGT indices in this study: CBT (Weijers et al., 2007b); MBT’ (Peterse et al., 2012); MBT’$_{5Me}$, MBT’$_{6Me}$, CBT$_{5Me}$, CBT’, and Index1 (De Jonge et al., 2014a); IR$_{6Me}$ (Dang et al., 2016); #rings$_{tetra}$, #rings$_{penta\ 5Me}$, and #rings$_{penta}$

$_{6Me}$ (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., 2015), community index (CI; De Jonge et al., 2019). Their equations are given below:

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

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

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

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

$$CBT_{5Me} = -log\left(\frac{Ib + IIb}{Ia + IIa}\right) \tag{A5}$$

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

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

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

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

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

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

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

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

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

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

8.2.2 Supplementary information

For all supplemental text, figures, and data, which have the prefix “S” in the main text (e.g., “Fig. S1”), the reader is referred to the online version of the published article (Raberg et al., 2021).

8.3 Appendix 3: Chapter 4 Supplemental Information

8.3.1 Comparison of fitting coefficients

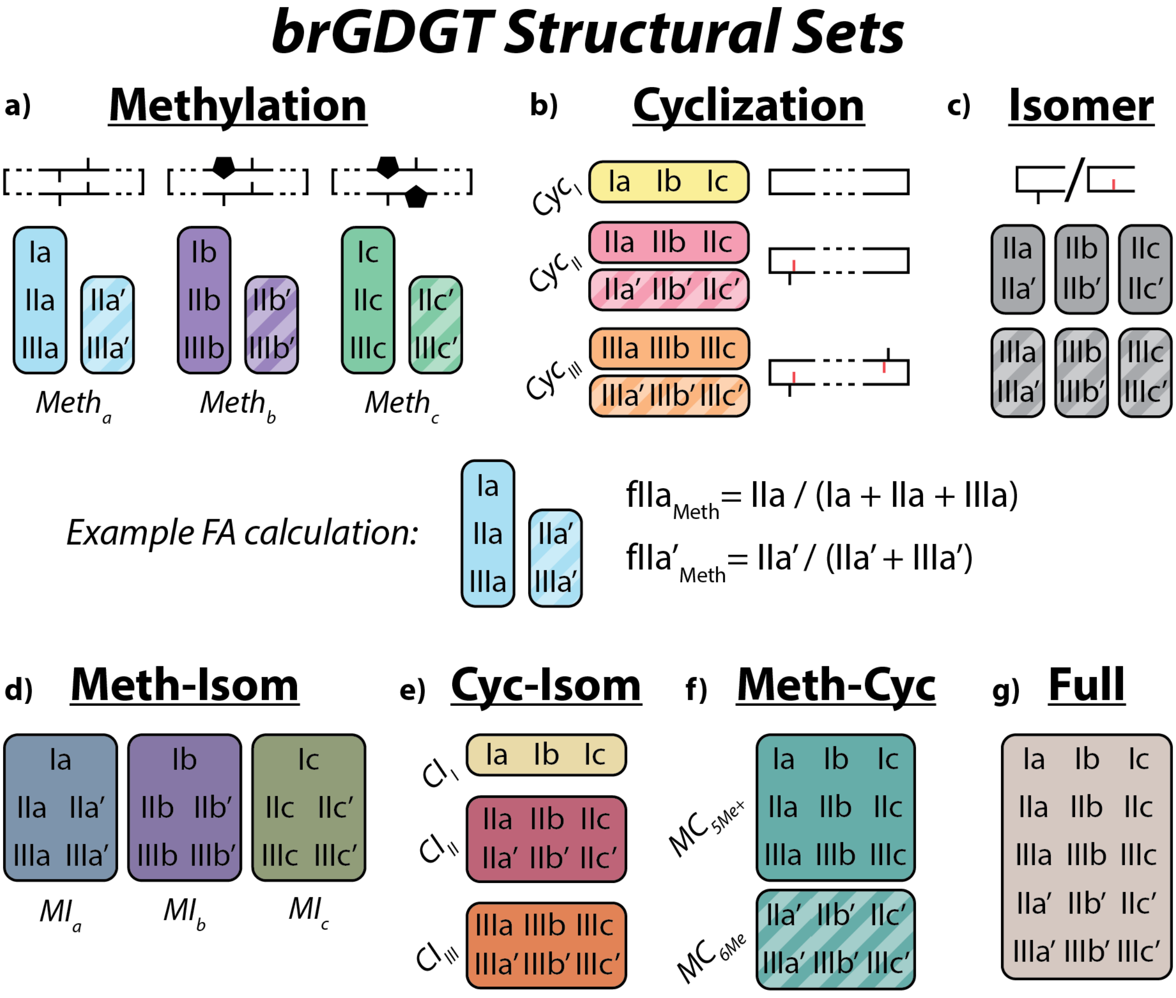


Figure 8-16: Schematic of the basic (a-c) and combined (d-g) brGDGT structural sets. Fractional abundances are calculated within each boxed group independently (Table S1). Schematic structures highlight the defining alkyl-chain moieties, with cyclopentane rings filled in for emphasis and C6 methylations denoted in red. (Reproduced from ref. (Raberg et al., 2021).)

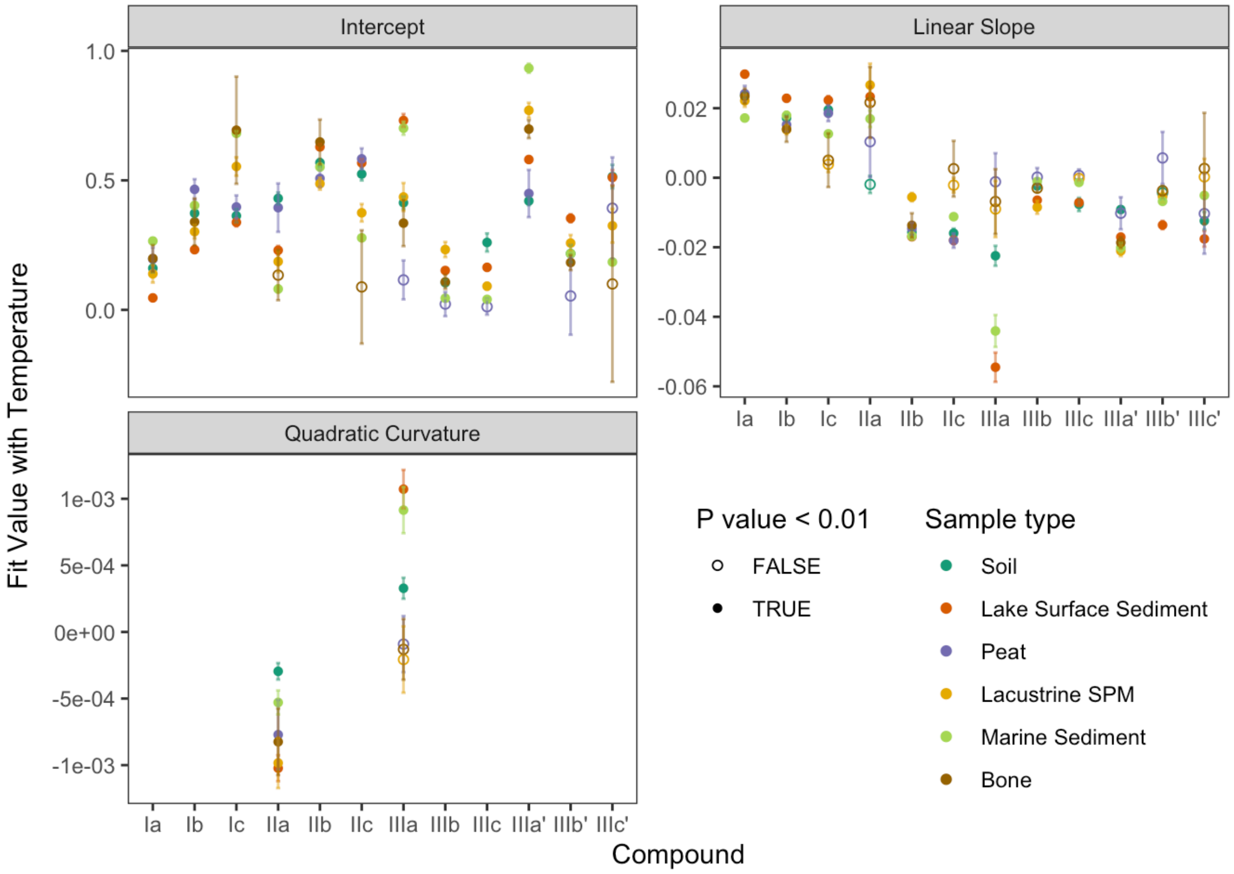


Figure 8-17: Coefficients of temperature fits. Fitting coefficients for quadratic (IIa, IIIa) and linear (all other compounds) regressions between brGDGT Meth Set FAs and Temperature, as plotted in Figure 1 of the main text, are provided. Error bars represent one standard error. Coefficients with p values ≥ 0.01 are plotted as open circles.

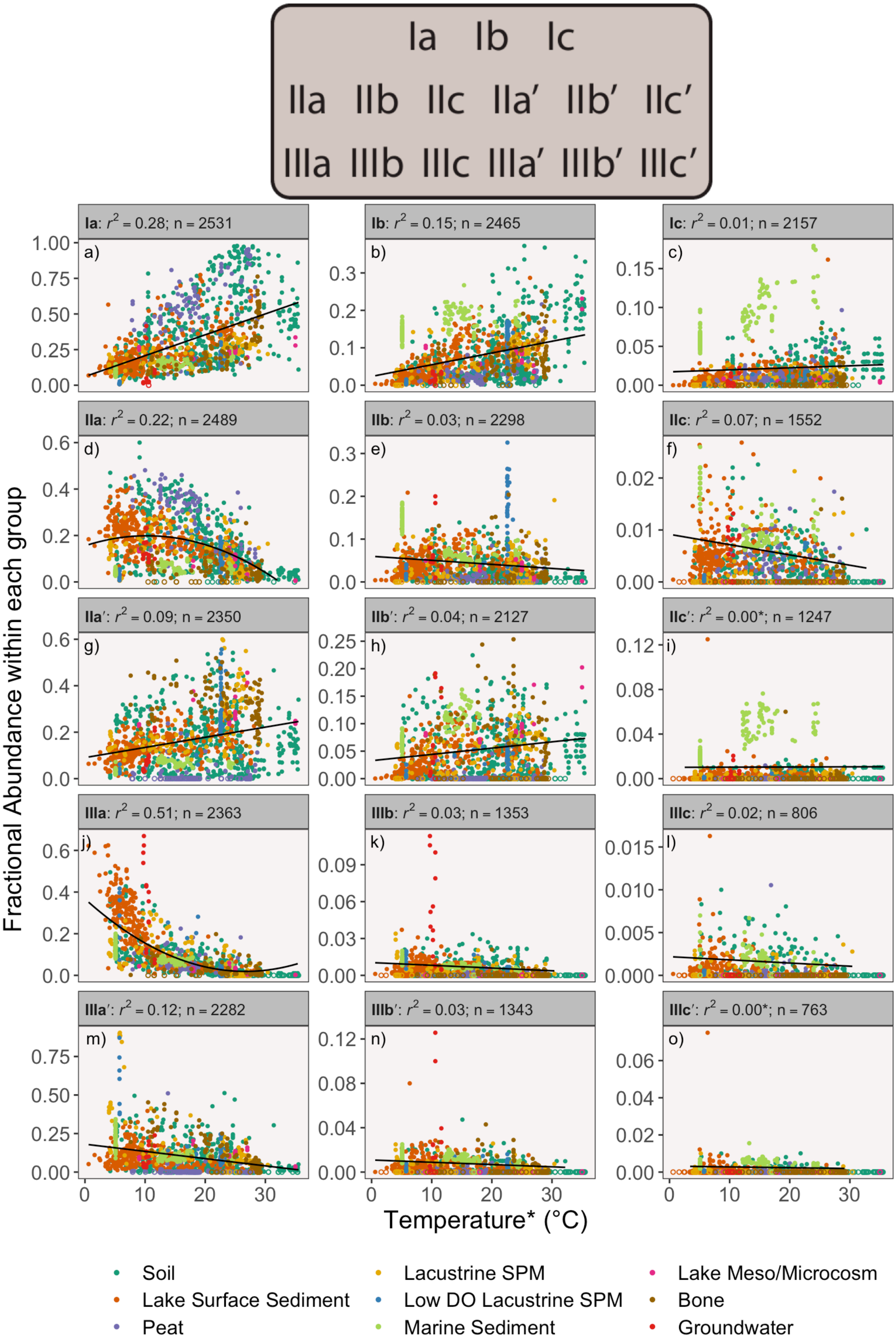

Ia Ib Ic
IIa IIb IIc IIa′ IIb′ IIc′
IIIa IIIb IIIc IIIa′ IIIb′ IIIc′
Ia: r² = 0.28; n = 2531
Ib: r² = 0.15; n = 2465
Ic: r² = 0.01; n = 2157
IIa: r² = 0.22; n = 2489
IIb: r² = 0.03; n = 2298
IIc: r² = 0.07; n = 1552
IIa′: r² = 0.09; n = 2350
IIb′: r² = 0.04; n = 2127
IIc′: r² = 0.00*; n = 1247
IIIa: r² = 0.51; n = 2363
IIIb: r² = 0.03; n = 1353
IIIc: r² = 0.02; n = 806
IIIa′: r² = 0.12; n = 2282
IIIb′: r² = 0.03; n = 1343
IIIc′: r² = 0.00*; n = 763
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
Fractional Abundance within each group
Temperature* (°C)
0
10
20
30
Soil
Lake Surface Sediment
Peat
Lacustrine SPM
Low DO Lacustrine SPM
Marine Sediment
Lake Meso/Microcosm
Bone
Groundwater

*Figure 8-18: Relationships between temperature and brGDGT fractional abundances (FAs) calculated within the Full set. FAs were calculated using all compounds (Fig. 8-16 and Table 8-2). *Temperatures are the same as in Figure 1. Best fit curves (quadratic for IIa and IIIa; linear for all other compounds) and are plotted for each compound, with R^2 values provided. P values were < 0.01 except where marked with an asterisk. Open circles denote points with FA = 0 or 1. These points were excluded from statistical analyses to match the analysis done with the Meth set (Fig. 4-1).*

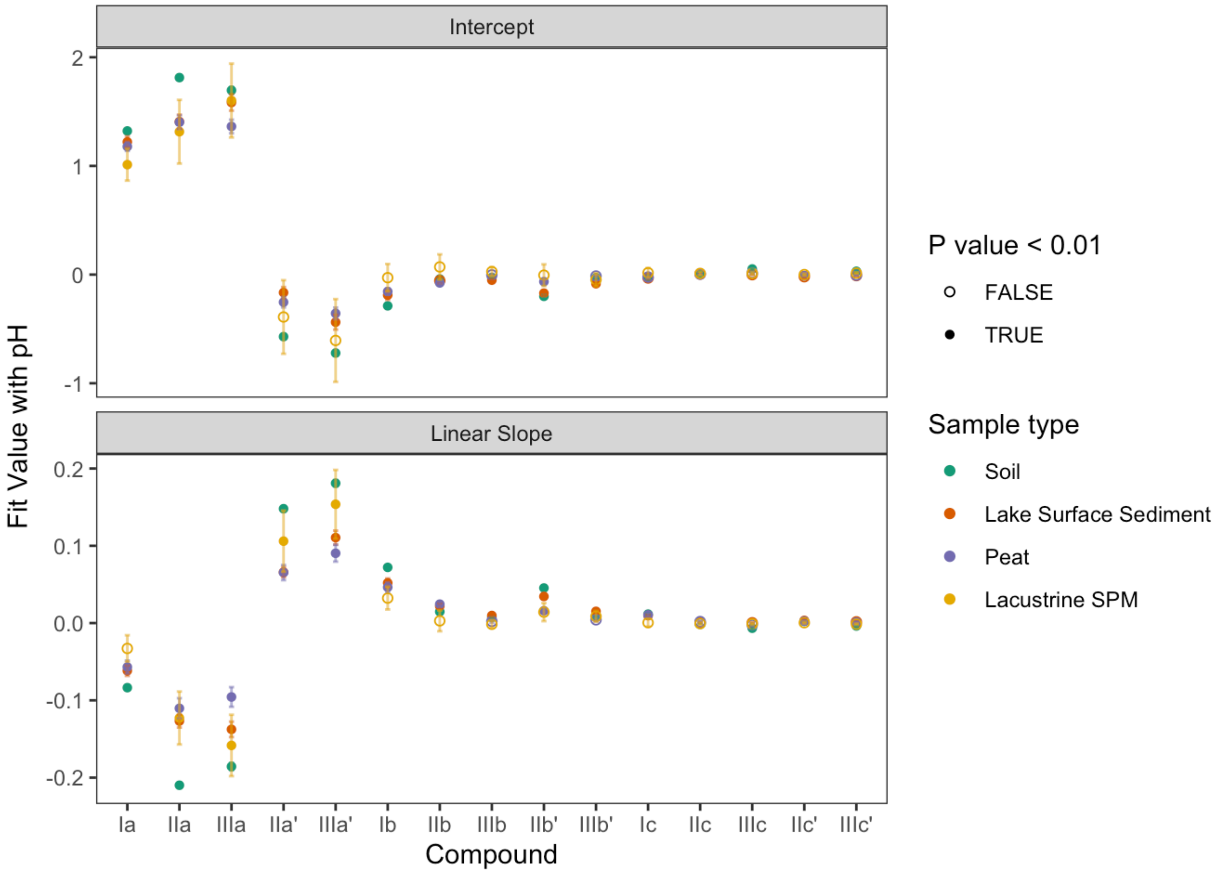


Figure 8-19: Fitting coefficients for linear regressions between brGDGT Cyc-Isom Set FAs and pH. Error bars represent one standard error. Coefficients with p values ≥ 0.01 are plotted as open circles.

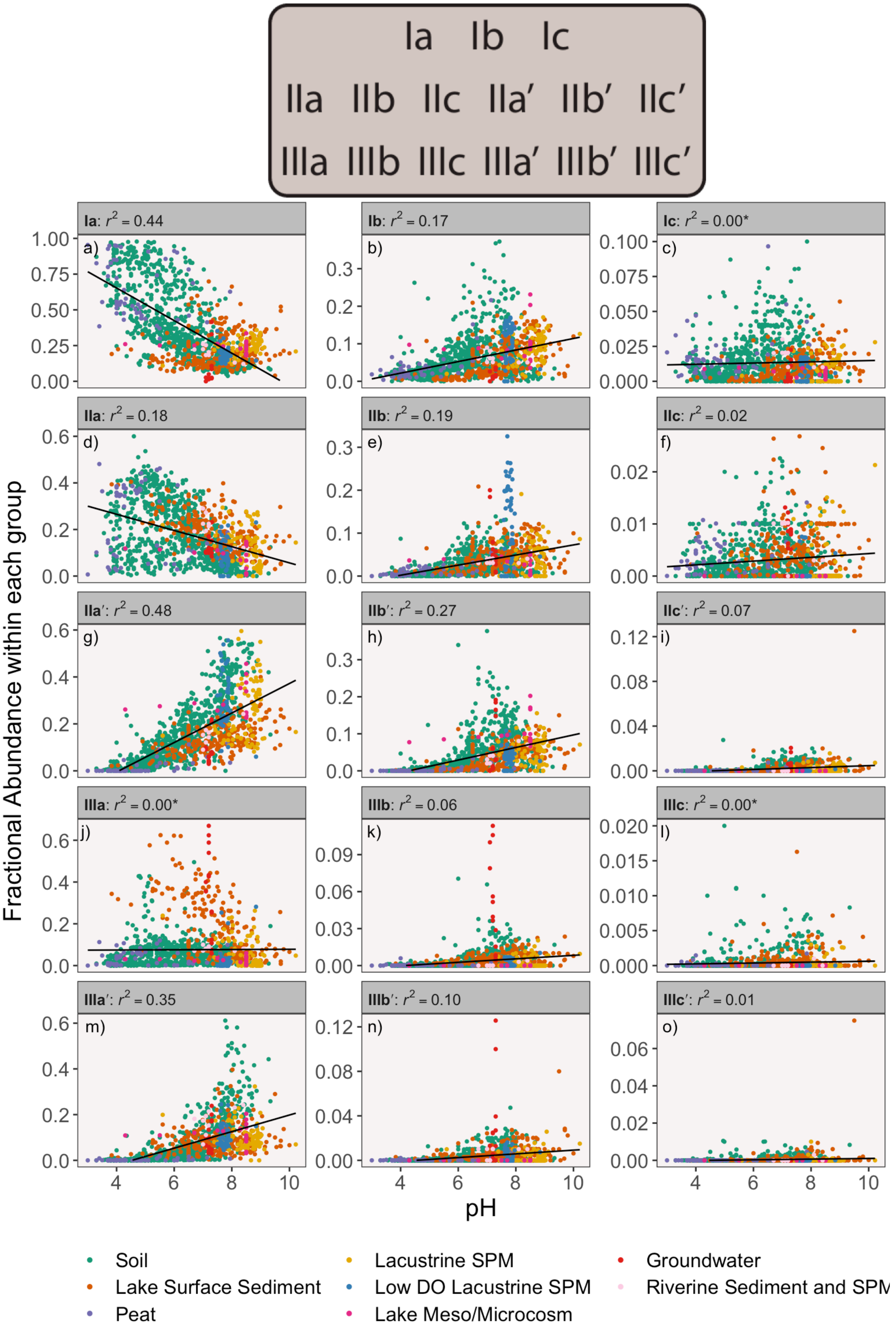
Ia Ib Ic
IIa IIb IIc IIa′ IIb′ IIc′
IIIa IIIb IIIc IIIa′ IIIb′ IIIc′
Ia: $r^2 = 0.44$
Ib: $r^2 = 0.17$
Ic: $r^2 = 0.00$*
IIa: $r^2 = 0.18$
IIb: $r^2 = 0.19$
IIc: $r^2 = 0.02$
IIa′: $r^2 = 0.48$
IIb′: $r^2 = 0.27$
IIc′: $r^2 = 0.07$
IIIa: $r^2 = 0.00$*
IIIb: $r^2 = 0.06$
IIIc: $r^2 = 0.00$*
IIIa′: $r^2 = 0.35$
IIIb′: $r^2 = 0.10$
IIIc′: $r^2 = 0.01$
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
Fractional Abundance within each group
pH
4
6
8
10
Soil
Lake Surface Sediment
Peat
Lacustrine SPM
Low DO Lacustrine SPM
Lake Meso/Microcosm
Groundwater
Riverine Sediment and SPM

Figure 8-20: Relationships between pH and brGDGT fractional abundances (FAs) calculated within the Full set. FAs were calculated using all compounds (Fig. 8-16 and Table 8-2). Linear fits are plotted for each compound, with R^2 values provided. P values were < 0.01 except where marked with an asterisk. N = 1295 for all subplots.

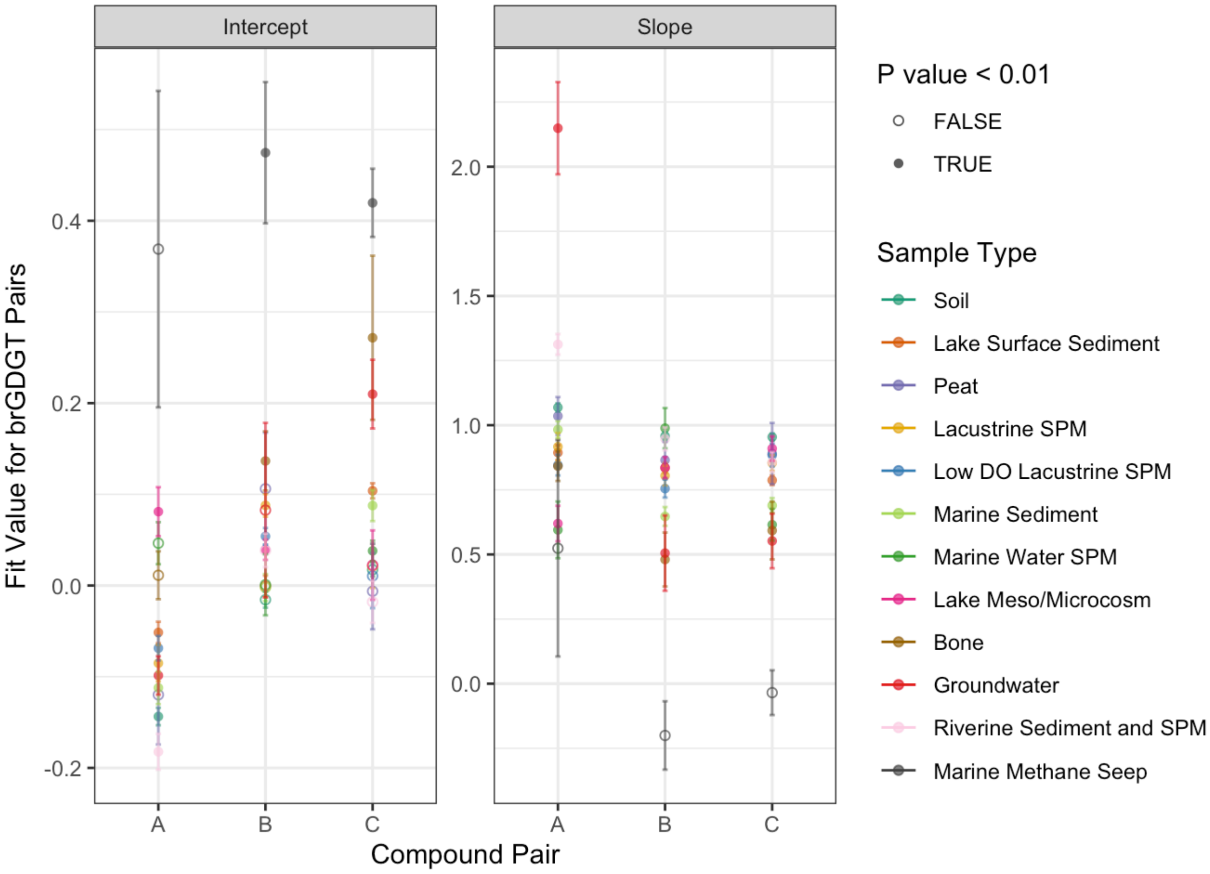


Figure 8-21: Fitting coefficients for linear regressions between the brGDGT compound pairs. Pairs are as plotted in Figure 4-3 of the main text. (Compound Pair A refers to Figure 4-3A, for example.) Error bars represent one standard error. Coefficients with p values ≥ 0.01 are plotted as open circles.

Table 8-2: Equations for calculating FAs within brGDGT structural sets used in this study. The terms 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).

Structural set name	Compound Groupings	Fractional abundance equation
Full	S = {Ia, Ib, Ic, IIa, IIb, IIc, IIIa, IIIb, IIIc, IIa', IIb', IIc', IIIa', IIIb', IIIc'}	$\mathrm{f}xy_{Full} = xy/\sum S$
Meth	S_a = {Ia, IIa, IIIa}; $S_{a'}$ = {IIa', IIIa'} S_b = {Ib, IIb, IIIb}; $S_{b'}$ = {IIb', IIIb'} S_c = {Ic, IIc, IIIc}; $S_{c'}$ = {IIc', IIIc'}	$\mathrm{f}xy_{Meth} = xy/\sum S_y$
Meth-Isom	S_a = {Ia, IIa, IIIa, IIa', IIIa'} S_b = {Ib, IIb, IIIb, IIb', IIIb'} S_c = {Ic, IIc, IIIc, IIc', IIIc'}	$\mathrm{f}xy_{MI} = xy/\sum S_y$
Cyc-Isom	S_I = {Ia, Ib, Ic} S_{II} = {IIa, IIb, IIc, IIa', IIb', IIc'} S_{III} = {IIIa, IIIb, IIIc, IIIa', IIIb', IIIc'}	$\mathrm{f}xy_{CI} = xy/\sum S_x$
Isom	S_{IIa} = {IIa, IIa'}; S_{IIIa} = {IIIa, IIIa'} S_{IIb} = {IIb, IIb'}; S_{IIIb} = {IIIb, IIIb'} S_{IIc} = {IIc, IIc'}; S_{IIIc} = {IIIc, IIIc'}	$\mathrm{f}xy_{Isom} = xy/\sum S_{xy}$

The qualitative similarities in Figures 4-1 to 4-3 can to some extent be quantified by comparing the fitting coefficients for individual sample types. Such an analysis can be informative but has important limitations that we list below.

A comparison of fitting coefficients reveals sample-type-specific slopes and intercepts that are often within error (Figs. 8-17, 8-19, and 8-21). For soils and peats, this overlap may be expected as these sample types were shown to have statistically indistinguishable relationships between MBT'$_{5Me}$ and temperature (Dearing Crampton-Flood et al., 2020). However, no other datasets have been successfully merged in this manner. It is therefore remarkable that even such dissimilar sample media as soil and marine sediment (e.g., IIb in Fig. 8-17), peat and lake sediment (e.g., Ia in Fig. 8-19), or groundwater and bone (e.g., pair C in Fig. 8-21), display compound-specific relationships with environmental parameters that are within error of one another.

In several cases, however, fitting coefficients for different sample types do not overlap. One notable example is that of soils and lake sediments. While these sample types display undeniably similar relationships with temperature (e.g., Fig. 4-1e) and pH (e.g., Fig. 4-2a), offsets consistently appear in their slopes and intercepts (Fig. 8-17 and 8-19). This disparity agrees with previous work showing that the relationships between MBT'$_{5Me}$ and temperature were distinct for lake sediments versus soils and peats (Martínez-Sosa et al., 2021). Even for sample types that might be more likely to overlap, such as lacustrine sediments and SPM, offsets in slope and intercept can appear (e.g., Fig. 8-17).

The disagreements in fitting coefficients could be evidence of important differences between sample types. However, several uncertainties exist that make such a conclusion beyond the resolution and scope of our approach with the existing data at this time. First, while we have excluded samples with multiple sources of brGDGTs as much as possible (e.g., marine sediments with clear soil input), it is likely that some still exist in our dataset. This issue could be alleviated in future studies as further methods are developed to fingerprint mixed sources or by making use of intact polar brGDGTs, which are more likely to be produced in situ. Second, while grouping brGDGTs into the Meth and Cyc-Isom Sets aided in isolating the effects of temperature and pH for lake sediments (Raberg et al., 2021), other environmental parameters can still influence brGDGT FAs (e.g., low DO lacustrine SPM in Fig. 4-1). Third, samples from a single sample type sometimes do not span the full range of values in the dataset (e.g., marine sediments in Fig. 4-3c) or are biased to one end (e.g., lacustrine SPM in Fig. 4-2), making it difficult to infer fitting coefficients across the full dataset range. Finally, and perhaps most simply, the environmental variables associated with each sample type are often (necessarily) different (see Methods). For example, air temperatures were used for the lake sediment dataset, while water temperature was used for lacustrine SPM. While these variables can be standardized to a certain extent (e.g., by using WMT for all terrestrial sample types), most of the available temperature and pH measurements are themselves imperfect proxies for the true physical or chemical conditions that brGDGT-producing organisms may experience (e.g., air temperature instead of *in situ* soil or water temperature). When *in situ* temperature values are available, likely biases in biological activity towards more favorable growing conditions (e.g., warmer temperatures (Cao et al., 2020) or higher nutrient availability (Loomis et al., 2014a)) can further complicate the selection of an appropriate temperature index. Even measurements of pH are imperfect to some extent as bulk properties may not accurately represent the chemical microenvironments in which microbes operate (e.g. in soils (Baveye et al., 2018)).

As a result of these challenges, we are unable to determine at this time whether the discrepancies in the precise relationships between brGDGTs, temperature, and pH are due to true differences between sample types or not. Advances in brGDGT research, such as the growing body of in situ temperature measurements (Wang et al., 2020; Pérez-Angel et al., 2020) or more advanced ways of approximating them (Stefanescu et al., 2021), may allow for this question to be addressed with more confidence in the future. However, despite these lingering uncertainties, we can still conclude that remarkable similarities exist in the relationships between brGDGTs, temperature, and pH across nearly all sample types examined and explore the implications of these broader agreements.

8.3.2 Supplementary Data

Supplementary data for Chapter 4 will be available with the published article or upon request.

8.4 Appendix 4: Chapter 5 Supplemental Information

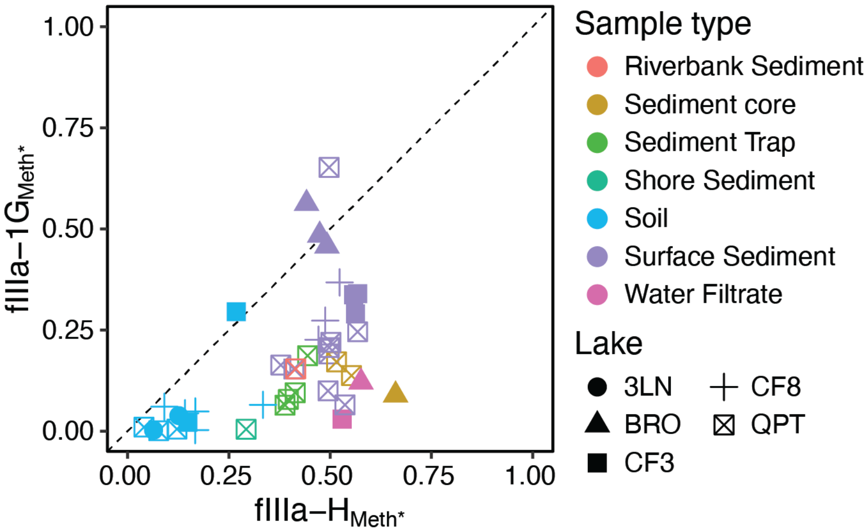


Figure 8-22: Average surface sediment fPH (Eq. 6) versus total phosphorous (µg/L) of lake water. Uncertainties (one standard deviation of repeat measurements; dashed gray lines) are shown where available. Data for Lake CF3 and some data for Lake CF8 were provided by Neil (unpublished; Michelutti et al., 2007).

8.5 Appendix 5: Glossary of terms

1G: monoglycosyl head group

brGDGT: branched glycerol dialkyl glycerol tetraether

c-brGDGT: core brGDGT

DO: dissolved oxygen

ECA: Eastern Canadian Arctic

FA: fractional abundance

H: hydrogen (the “head group” for a core brGDGT)

HG: head group

i-brGDGT: intact brGDGT

MAF: mean temperature of months above freezing

MAT/MAAT: mean annual (air) temperature

PH: phosphohexose head group