Exploring Past Environments with Microbial Biomarkers: A Molecular Approach Using Sedimentary DNA and 3-Hydroxy Fatty Acids

By

Amy Caroline Thorpe

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHHY

School of Geography, Earth and Environmental Sciences

College of Life and Environmental Sciences

University of Birmingham

June 2023

Abstract

Palaeoclimate records are crucial for understanding the causes and rates of past climate change and informing predictions of future climate change. Palaeoenvironmental records of microbial communities are important for understanding the ecological impacts of climatic and environmental change due to their position near the base of many food webs, roles in biogeochemical cycling, and ability to act as indicators of change. Producing detailed and reliable palaeoclimate records and reconstructions of past microbial communities depends on the application of organic molecules preserved in temporally stratified sedimentary archives as biomarkers of change. This thesis explores the use of two promising novel microbial biomarkers: sedimentary DNA (sedDNA) to reconstruct past microbial communities, and bacterial 3-hydroxy fatty acids (3-OH FAs) to reconstruct past climates.

Analysis of sedDNA is becoming more frequently used to reconstruct past microbial communities and can be applied with a relatively high taxonomic resolution. However, the taphonomy of DNA in lakes is poorly understood, and the reliability of sedDNA-based reconstructions of microbial communities is largely unknown. To reconstruct past bacterial and eukaryotic microbial communities, 16S rRNA and 18S rRNA gene amplicon sequencing were carried out on sediment cores collected from Esthwaite Water (English Lake District) spanning approximately 100 years. To assess the reliability of these reconstructions, the sedDNA records were compared against concurrent, long-term microscopy-based monitoring of cyanobacteria and eukaryotic algae in the surface water, revealing comparable broad-scale temporal shifts in community composition in response to eutrophication. However, a decline in the concentration of 16S rRNA genes and proportion of DNA likely originating from the water column down-core suggested that the reliability of reconstructions may be reduced in older sediments. These

results demonstrate that sedDNA is a valuable tool to reconstruct past microbial communities, but further study is needed to better understand the extent and rate of DNA deposition and degradation in lake sediments to enhance its development as a reliable biomarker.

Proxies based on 3-OH FAs show significant potential as palaeoclimate biomarkers that can be applied to various environments. However, the identity of 3-OH FA producers and the influence of bacterial community composition on the performance of 3-OH FA proxies in terrestrial environments have yet to be determined. Through paired bacterial 16S rRNA gene amplicon sequencing and 3-OH FA analysis in soils and lake surface sediments collected across large gradients of mean annual air temperature (MAAT), mean annual precipitation (MAP) and soil and lake sediment pH in the U.S., it was suggested that distinct communities of bacteria specific to each environment may be responsible for the production of 3-OH FAs. Therefore, the bacterial community may be an important driver of 3-OH FA distribution in the environment. Furthermore, the culture-based analysis revealed that bacteria from different taxonomic lineages may produce a distinct suite of 3-OH FA isomers. These findings highlight the importance of considering the bacterial producers of 3-OH FAs in the development of reliable 3-OH FA proxies for palaeoclimate reconstructions.

Acknowledgements

Thank you to my supervisors, Dan Read and James Bendle, for supporting me throughout my PhD. Dan, your continuous guidance, encouragement and feedback have been invaluable. I am incredibly grateful for the sedDNA dataset you shared with me that led to my first, first-author paper. James, thank you for your helpful feedback and for introducing me to a new field of study. It has been a pleasure to work with you both.

I would also like to thank the Molecular Ecology group at the UK Centre for Ecology & Hydrology. Thank you to Tim Goodall for training me in the lab (I learnt from the best!), and always being enthusiastic about my research and molecular ecology, and to Holly Tipper for teaching me qPCR. Thank you to Katharine Moss for sharing your organic geochemistry expertise, assisting with fieldwork, supporting me at conferences and joining me for the holidays afterwards. Your help and encouragement, especially during the writing stage of my PhD, have been invaluable.

I would like to extend my thanks to Steve Thackeray, Stephen Maberly, Ellie Mackay and Soon Gweon, who I had the privilege of co-authoring my first papers with. Your advice and comments greatly improved the quality of my work. A special thanks go to Ellie for showing us around Esthwaite Water and helping us collect our sediment core. Thank you to Joe Taylor and Ellie for providing seasonal water samples from Esthwaite, and to Rob Griffiths for your advice on network analysis.

Thank you to the Organic Geochemistry group at the University of Birmingham for training me in your organic geochemistry methods. I am particularly grateful to Alice Hardman for sharing data from your PhD and accompanying me on fieldwork, and to Grace Duffy for assisting us in the field.

I would also like to thank the Animal Care Team at the Cornish Seal Sanctuary for allowing me to carry out a placement with you. It was a highlight of my postgraduate training, and I hope Limpet and Narwhal are doing well back in the wild.

Finally, I would like to thank Dad, Sam and Catherine for always supporting me, and I will forever be grateful for our family of cats: Cat, Spencer, Jasper, Tigger and Bear, and chickens: Henrietta, Sooty, Dippy, Speckle, Coco, Daisy, Amber, Lucy, Ivy and Pip, who have provided me with immense happiness and motivation.

Author contributions

This thesis is presented in an alternative format, where each experimental chapter is written in the style of a scientific publication. Chapter 2 has already been published, and Chapter 3 is currently under review. I am the first author of both papers. Detailed contributions to each chapter are outlined below. Supervision from Daniel Read and James Bendle applies to all chapters.

Chapter 2 is published in Environmental DNA as: Thorpe, A. C., Anderson, A., Goodall, T., Thackeray, S. J., Maberly, S. C., Bendle, J. A., Gweon, H. S. and Read, D. S. (2022) Sedimentary DNA records long-term changes in a lake bacterial community in response to varying nutrient availability. *Environmental DNA*, 4 (6), 1340-1355. https://doi.org/10.1002/edn3.344.

Fieldwork for **Chapter 2** was carried out by Amy Anderson and Stephen Thackeray, and lab work was performed by Amy Anderson, Soon Gweon and Tim Goodall. Soon Gweon processed the sequencing reads and contributed to the FreshTrain analysis, and Amy Thorpe conducted all remaining data analysis with guidance from Daniel Read and Stephen Thackeray. The original draft was written by Amy Thorpe, and all authors contributed to editing the draft and approved the final version of the manuscript.

Chapter 3 is under review in Molecular Ecology Resources as: Thorpe, A. C., Mackay, E. B., Goodall, T., Bendle, J. A., Thackeray, S. J., Maberly, S. C. and Read, D. S. (2023) Evaluating the use of lake sedimentary DNA in palaeolimnology: A comparison with long-term microscopy-based monitoring of the phytoplankton community. A preprint is available at: https://doi.org/10.22541/au.167819405.56988284/v1.

Fieldwork for **Chapter 3** was carried out by Amy Thorpe, Eleanor Mackay and Alice Hardman, and lab work was performed by Amy Thorpe. The long-term microscopy-based monitoring record was provided by Stephen Maberly and Stephen Thackeray. All data analysis was completed by Amy Thorpe. The original draft was written by Amy Thorpe, and all authors contributed to editing the draft and approved the final version of the manuscript.

Fieldwork for **Chapter 4** was carried out by Amy Thorpe, Eleanor Mackay and Alice Hardman. Seasonal surface water samples were provided by Eleanor Mackay and Joe Taylor. All lab work, data analysis and writing of this chapter were conducted by Amy Thorpe.

Chapter 5 represents a data-sharing collaboration with Alice Hardman, an Earth Sciences PhD student at the University of Birmingham. Fieldwork was carried out by Alice Hardman and James Bendle, and Alice Hardman was responsible for generating the organic geochemistry dataset. DNA sequencing, data analysis and writing of this chapter were conducted by Amy Thorpe.

Soil samples used in **Chapter 6** were collected by Amy Thorpe, Tim Goodall or Catherine Thorpe. Culturing of the bacterial isolates was performed by Amy Thorpe, and lipid extraction was carried out by Amy Thorpe with initial guidance from the University of Birmingham Organic Geochemistry group. All data analysis and writing of this chapter were conducted by Amy Thorpe.

Table of Contents

List of figures

Chapter 5

List of tables

Chapter 6

List of abbreviations

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

rRNA: Ribosomal RNA

sedDNA: Sedimentary DNA

sedRNA: Sedimentary RNA

PCR: Polymerase chain reaction

qPCR: Quantitative PCR

Ct: Cycle threshold

dNTP: Deoxynucleoside triphosphate

dsDNA: Double-stranded DNA

HS: High sensitivity

BR: Broad range

BSA: Bovine serum albumin

DADA2: Divisive amplicon denoising algorithm 2

ASV: Amplicon sequence variant

OTU: Operational taxonomic unit

NMDS: Non-metric multidimensional scaling

ANOSIM: Analysis of similarity

GAM: Generalised additive model

AIC: Akaike information criterion

REML: Restricted maximum likelihood

MST: Microbial source tracking

FEAST: Fast expectation-maximisation source tracking

CRS: Constant rate of ^{210}Pb supply

TP: Total phosphorus

SRP: Soluble reactive phosphorus

Chl *a*: Chlorophyll *a*

MAAT: Mean annual air temperature

MAP: Mean annual precipitation

SST: Sea surface temperature

3-OH FA: 3-hydroxy fatty acid

i: *iso*

a: *anteiso*

n: *normal*

u: Unsaturated

RAN$_{15}$: Ratio of *anteiso* to *normal* C_{15} 3-OH FAs

RAN$_{17}$: Ratio of *anteiso* to *normal* C_{17} 3-OH FAs

RAN$_{13}$: Ratio of *anteiso* to *normal* C_{13} 3-OH FAs

RAN$_S$: Ratio of *i*-C_{12}, *a*-C_{13} and *i*-C_{14} to *i*-C_{12}, *a*-C_{13}, *n*-C_{13}, *i*-C_{14} and *n*-C_{14} 3-OH FAs

RIN: Ratio of *iso* to *normal* 3-OH FAs

RIN$_{17}$: Ratio of *iso* to *normal* C_{17} 3-OH FAs

RIAN: Ratio of *iso* and *anteiso* to normal 3-OH FAs

GDGT: Glycerol dialkyl glycerol tetraether

isoGDGT: Isoprenoid GDGT

brGDGT: Branched GDGT

TEX$_{86}$: Tetraether index of isoGDGTs with 86 carbons

CBT: Cyclisation of brGDGTs

MBT: Methylation of brGDGTs

TLE: Total lipid extract

FAME: Fatty acid methyl ester

OH FAME: Hydroxy FAME

DCM: Dichloromethane

TMSi: Trimethylsilyl ester

BSTFA: *N*, *O*-bis(trimethylsilyl)trifluoroacetamide

GC-MS: Gas chromatography-mass spectrometry

m/z: Mass to charge ratio

PBS: Phosphate-buffered saline

R2A: Reasoner's 2A agar

Chapter 1: General introduction

1.1. Overview

Long-term records of past climate and ecosystem change are crucial for understanding the ecological impacts of change, distinguishing between natural and human-induced changes, and informing predictions of future climate and ecosystem change (Hegerl *et al*., 2019; Willis *et al*., 2010). Long-term records of microbial communities are particularly important as microbes form the base of many food webs, play key roles in biogeochemical cycling, are highly sensitive to the environment, and can act as sensitive indicators of change (Bloem and Breure, 2003; Litchman *et al*., 2015; Newton *et al*., 2011). A range of molecules produced by microbial communities, such as lipids, pigments, microfossils and DNA, are well-preserved in palaeoenvironmental sedimentary archives. These molecules can be used as biomarkers to reconstruct past changes in the occurrence and physiological state of microbes, and infer the environmental and climatic conditions they experienced (Dearing *et al*., 2006; Eglinton and Eglinton, 2008). However, many of these biomarkers are limited in their taxonomic specificity (Capo *et al*., 2021), and existing bacterial palaeoclimate biomarkers are limited in their ability to reliably reconstruct historical change in terrestrial sedimentary archives (Naeher *et al*., 2014; Blaga *et al*., 2009). The development of novel biomarkers that can be applied to the wider microbial community, provide a high taxonomic resolution, and can be used in a range of different environments are essential for producing detailed and reliable palaeoenvironmental reconstructions.

This chapter provides an overview of biomarkers traditionally used to reconstruct past microbial communities and past climates, and discusses their limitations. Two novel biomarkers with significant potential in palaeoenvironmental research are introduced: sedimentary DNA (sedDNA) to study past microbial communities, and bacterial 3-hydroxy fatty acids (3-OH FAs)

to infer past climatic conditions, and areas that require further research to ensure their application as reliable biomarkers are identified.

1.2. Producing reliable palaeoenvironmental reconstructions

1.2.1. Importance of reliable records of past climates and ecosystems

Long-term records of the climate are crucial for understanding the causes and rates of past climate change, and for providing insights into current and future rates of change (Hegerl *et al*., 2019). Historical ecosystem records are important for understanding the resilience and ability of organisms to adapt to climate and environmental change (Willis and Bhagwat, 2010). They are also critical for understanding the ecological consequences of human activities and human-induced climate change, and can be used to inform ecosystem management and conservation strategies to protect vulnerable species and ecosystems (Dearing *et al*., 2006). However, reliable records of the climate are restricted to the instrumental period, which began in the 18th century (Hulme and Jones, 1994), and detailed ecosystem monitoring records rarely exceed 100 years (Willis *et al*., 2010). Furthermore, the majority of long-term monitoring records only cover the last few decades (Rösel *et al*., 2012; Shade *et al*., 2007), and many are biased towards a relatively small number of well-studied taxa that are easily resolved with morphological identification methods (Burlakova *et al*., 2018; Hampton *et al*., 2008). Detailed, long-term records of the climate and ecosystem that precede the onset of human activities are essential for distinguishing between natural and human-induced changes and their impacts (Mills *et al*., 2016; Willis *et al*., 2010).

1.2.2. Sedimentary archives as records of past climates and ecosystems

Material preserved within palaeoenvironmental sedimentary archives can be analysed and used to construct retrospective records of climate and ecosystem change over periods of time much longer than what can be achieved with contemporary monitoring alone (Dearing *et al*., 2006). These sedimentary archives become temporally stratified when deposited material gradually accumulates to form a series of layers which each represent a discrete interval of time. Organic molecules preserved within each layer can be used to infer the state of the climate or ecosystem at the time of deposition (Zolitschka *et al*., 2015). To be used as effective palaeoenvironmental records, the sediment must be annually laminated or amenable to dating methods such as radiometric dating to establish a reliable chronology (Arcusa *et al*., 2022), should experience little to no disturbance or mixing between stratified layers (Glew *et al*., 2002), and environmental conditions should be suitable for the preservation of organic matter, such as a low temperature, low light availability, low oxygenation and minimal microbial activity (Meyers, 1997; Zonneveld *et al*., 2010). Environments that can be sampled to collect temporally stratified samples include marine sediments (Selway *et al*., 2022; Yang *et al*., 2020), lake sediments (Ibrahim *et al*., 2021; Naeher *et al*., 2012), peatlands (Blundell *et al*., 2016; Lamentowicz *et al*., 2015), cave speleothems (Stahlschmidt *et al*., 2019; Wang *et al*., 2018) and permafrost (Ilyashuk *et al*., 2006; Stapel *et al*., 2016).

1.2.3. Properties of effective and reliable biomarkers

Biomarkers are molecules or structures produced by living organisms that can be extracted from the environment and used to provide information about the past or present occurrence or physiological state of the organisms which produced them (Brocks and Grice, 2011; Feng *et al*., 2019). This information can, in turn, be used to infer the environmental or climatic

conditions the producers were responding to (Eglinton and Eglinton, 2008). Biomarkers, often described as molecular fossils, can be extracted from temporally stratified sedimentary archives and possess several qualities that make them useful for reliable reconstructions of past climates and ecosystems (Eglinton and Logan, 1991; Summons and Lincoln, 2012).

The temporal extent of a reconstruction depends on how well biomarkers are preserved in the environment. To produce reliable reconstructions which extend over hundreds, thousands, or even millions of years, biomarkers must be resistant to degradation and should not undergo significant structural modifications during preservation in the environment (Eglinton and Logan, 1991; Meyers, 1997). The processes leading to their preservation in the environment, known as *taphonomy*, should be well understood to ensure any pre- or post-depositional structural modifications that do occur do not compromise the reliability of reconstructions (Maloney *et al*., 2022). Furthermore, the transportation of biomarkers from their source and their residence time in intermediate environments prior to deposition in the sedimentary archive must be well-characterised to ensure that interpretations of their presence are correctly applied to their true environmental source (Eglinton and Eglinton, 2008).

The biological sources of biomarkers must be identifiable to prevent misleading interpretations of their presence in palaeoenvironmental archives (Belt *et al*., 2018). Biomarkers with high taxonomic specificity are important indicators of the occurrence and distribution of certain taxa in past environments, and can be used to reconstruct taxa-specific responses to environmental changes (Moldowan and Jacobson, 2000). Biomarkers produced by organisms that are highly sensitive to the environment and show well-characterised responses to environmental conditions are particularly useful for providing historical insights into environmental and climatic changes (Limaye *et al*., 2010; Meyers, 1997). To enable estimates of past relative abundance, the relative abundance of a biomarker extracted from the environment should be

related to that of the producer (Boschker and Middelburg, 2002). To ensure the reliability of biomarker-based palaeoenvironmental and palaeoclimatic reconstructions, biomarkers must first be calibrated with modern conditions to understand how their presence varies in space and time (Laprida *et al*., 2017). Furthermore, to ensure the inferred trends represent true temporal changes, they must be validated against other biomarker-based reconstructions or instrumental monitoring records of the same time period (Axford *et al*., 2011).

1.3. Biomarkers for the reconstruction of past microbial communities

1.3.1. Importance of long-term records of microbial communities

The activity, abundance and composition of microbial communities are highly sensitive to environmental and climatic changes (Litchman *et al*., 2015; Zingel *et al*., 2018). Microbial communities form the base of many food webs and play important roles in nutrient cycling and primary production across various environments including soils, lakes and oceans (Newton *et al*., 2011; Sarmento, 2012). Additionally, microbes are often the first to respond to environmental changes, making them important indicators of change (Bloem and Breure, 2003). Producing reliable and detailed long-term records of microbial communities is therefore crucial for understanding how the broader ecosystem may respond to long-term drivers of change (Shade *et al*., 2012).

1.3.2. Traditional biomarkers for studying past microbial communities

Microbes produce a wide range of molecules that can act as biomarkers (Fig. 1.1), providing valuable information on their past and present occurrence, abundance, diversity, activity and physiological state (Boschker and Middelburg, 2002; Morgan and Winstanley, 1997).

Microbial biomarkers traditionally used in palaeoenvironmental studies include lipids (Volkman *et al*., 1998), pigments (Leavitt and Hodgson, 2002), and cell structures such as frustules and cysts (Verleyen *et al*., 2017) which are preserved in the environment and are specific to certain microbes and taxonomic groups.

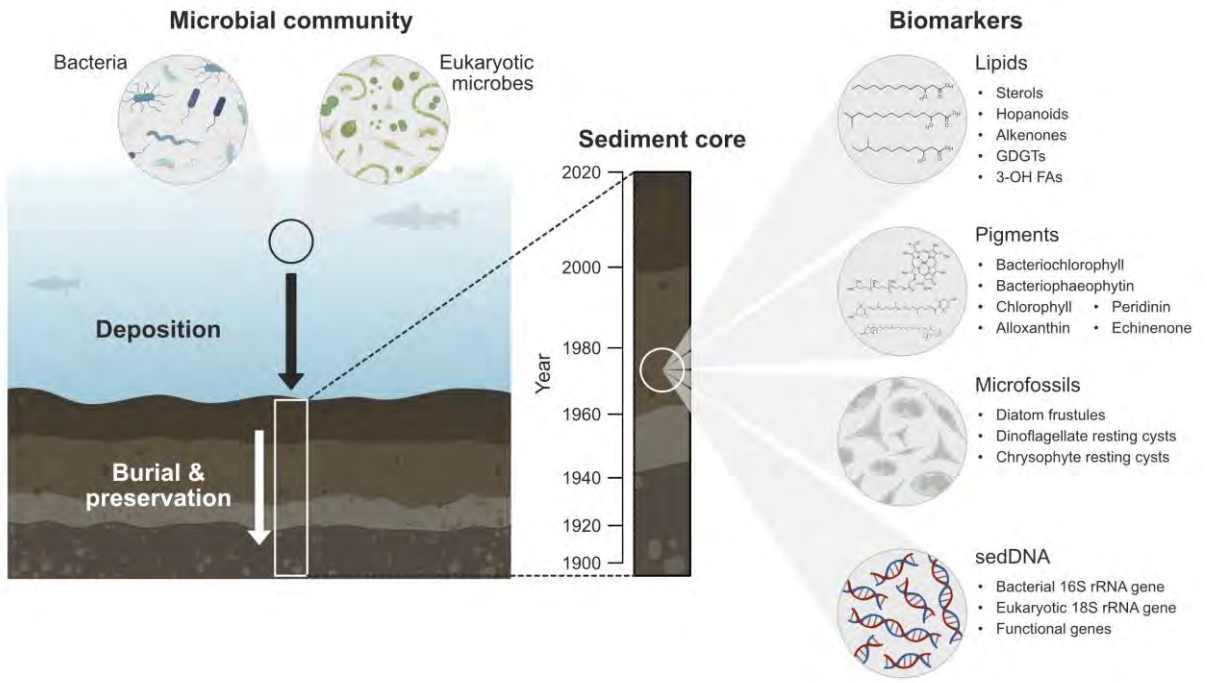

Figure 1.1. Simplified illustration of the deposition and preservation of bacterial and eukaryotic microbial communities and their biomarkers in a dated sedimentary archive. Examples of lipid, pigment, microfossil and sedimentary DNA (sedDNA) biomarkers are given.

Sterols, hopanoids and alkenones are among some of the lipids produced by microbes that have applications as biomarkers (Volkman and Smittenberg, 2017). Sterols, which are predominately produced by eukaryotic algae, and hopanoids, which are predominantly produced by bacteria, are particularly well-preserved in the environment and have been used to trace contribution from algal (Cao *et al*., 2016; Schwark and Empt, 2006) and bacterial (Chen *et al*., 2011; Saito and Suzuki, 2007; Song *et al*., 2013; Xu and Jaffé, 2008) communities to lake and marine sedimentary records over millions of years. Alkenones, which are long-chained unsaturated ketones produced by some haptophytes, are also relatively well-preserved in the environment

and have been used to reconstruct past haptophyte abundance in lake and marine sedimentary records over thousands of years (dos Santos *et al*., 2012; Randlett *et al*., 2014).

Eukaryotic algae and photosynthetic bacteria produce a variety of pigments that can be used as biomarkers (Sanger, 1988). The taxonomic specificity of photosynthetic pigments is variable, as several photosynthetic pigments, such as chlorophyll *a*, diatoxanthin and fucoxanthin, are produced by multiple algal groups. However, some pigments are specific to certain algal phyla, such as alloxanthin, which is produced by cryptophytes, dinoxanthin and peridinin, which are produced by dinoflagellates, and echinenone and zeaxanthin, which are produced by cyanobacteria. Some cyanobacterial pigments are indicative of specific classes or even genera, such as aphanizophyll, which is produced by nitrogen-fixing cyanobacteria, and oscillaxanthin, which is specific to *Oscillatoria* (Leavitt and Hodgson, 2002). These photosynthetic pigments have been extracted from lake and marine sediment cores to reconstruct past cyanobacterial and eukaryotic algal community dynamics over timescales ranging from a few hundred to a million years (Deshpande *et al*., 2014; Florian *et al*., 2015; Funkey *et al*., 2014; Tani *et al*., 2009). Pigments typically produced by purple and green sulfur bacteria, such as bacteriochlorophyll *a*, bacteriophaeophytins, okenone and chlorobactene, have also been used as biomarkers to reconstruct bacterial communities in lake and marine sedimentary archives (Brocks *et al*., 2005; Coolen and Overmann, 1998; Romera-Viana *et al*., 2009; Tani *et al*., 2009).

Microfossil remains of some microbes can be found in sedimentary archives and used as biomarkers (Battarbee, 2000). This includes the silica frustules of diatoms (Verleyen *et al*., 2017) and the resting cysts formed by some dinoflagellates (Mertens *et al*., 2012) and chrysophytes (Smol, 1988) which are well-preserved and have highly specific morphologies (Mertens *et al*., 2012; Verleyen *et al*., 2017). These microfossils have been extracted from lake and marine sedimentary archives and identified to the genus or species level with microscopy

to reconstruct past abundance and community composition over timescales ranging from a few thousand to millions of years (Andrén *et al*., 2000; Drljepan *et al*., 2014; Hembrow *et al*., 2014; Leira, 2005; Paterson *et al*., 2003; Warny *et al*., 2009).

Microbial lipids and photosynthetic pigments are well-preserved in sedimentary archives, but their use as effective biomarkers is limited by a generally low taxonomic resolution that rarely allows identification beyond the class level (Gong *et al*., 2020). Well-preserved and morphologically distinct microfossil remains are only produced by a few microbial groups (Verleyen *et al*., 2017). To study the temporal dynamics of the wider microbial community, including microbes that cannot easily be resolved with traditional biomarkers due to non-specific lipids and pigments or an absence of well-preserved structures, a biomarker that can be applied to the wider microbial community with a relatively high taxonomic resolution is needed.

1.3.3. Use of sedDNA to study past microbial communities

DNA preserved in sedimentary archives can be extracted, sequenced, and used to reconstruct past communities (Capo *et al*., 2021). sedDNA is a relatively recent addition to the suite of biomarkers, and this technique can be used to study past microbial communities in much greater detail than what can be achieved with traditional biomarkers (Capo *et al*., 2021; Domaizon *et al*., 2017). Metabarcoding of sedDNA can target the wider community with broad-range amplicon primers (Tessler *et al*., 2023), or target taxa of interest with specific amplicon primers (Feist and Lance, 2021; Kang *et al*., 2021). sedDNA can be applied to bacteria (Li *et al*., 2019a), eukaryotic microbes (Capo *et al*., 2016), zooplankton (Tsugeki *et al*., 2022) and macrophytes (Stoof-Leichsenring *et al*., 2022), and therefore has the potential to be used to reconstruct interactions within and between different trophic levels (Barouillet *et al*., 2022; Ellegaard *et al*.,

2020). Depending on amplicon primer specificity and reference sequence database coverage, metabarcoding of sedDNA can be used to make genus and sometimes species level assignments (Anslan *et al*., 2020; Tessler *et al*., 2023). Furthermore, sedDNA can also be used to study how microbial activity may have changed in response to environmental and climatic changes with metagenomics and functional gene analysis (Ellegaard *et al*., 2020). For example, genes implicated in metabolic pathways can be targeted to understand the roles of microbes in biogeochemical cycling (Qian *et al*., 2023). Other functional genes that can be targeted include those involved in the biosynthesis of harmful toxins to understand how their production may have been influenced by environmental and climatic conditions in the past, and to inform predictions of the occurrence and impacts of future harmful blooms (Heathcote *et al*., 2023; Tse *et al*., 2018).

Lake and marine sedDNA have been used to reconstruct past eukaryotic algal and cyanobacterial communities and their responses to long-term environmental change such as eutrophication and climate warming (Capo *et al*., 2016; Domaizon *et al*., 2013; Ibrahim *et al*., 2020; Monchamp *et al*., 2016, 2019; Segawa *et al*., 2022), and lake acidification (Mejbel *et al*., 2023) over timescales of up 2200 years. sedDNA reconstructions of bacterial communities are rare in comparison, but lake sedDNA has also been used to reconstruct past bacterial community responses to nutrient enrichment and heavy metal contamination over 150 years (Li *et al*., 2019a). There is evidence that suggests that microbial DNA can be preserved in sedimentary archives for up to 10,000 years (Coolen and Overmann, 1998), and plant and animal DNA has been recovered from 2 million year-old permafrost (Kjær *et al*., 2022). Therefore, under favourable conditions for preservation, sedDNA has significant potential to reconstruct past communities over long timescales.

Similar to traditional microbial biomarkers, the reliability of sedDNA-based reconstructions is dependent upon our understanding of the transport and preservation of organic matter in sedimentary archives (Capo *et al*., 2021; 2022). However, it is not yet known whether the DNA of some bacterial and eukaryotic microbes may deposit more efficiently in sedimentary archives and be more resistant to degradation than others (Capo *et al*., 2021; Mejbel *et al*., 2022). DNA may be degraded by oxidation, photooxidation, hydrolysis or by the activity of enzymes such as DNases, and dissolved DNA may be more vulnerable to these processes compared to intracellular DNA or that which is adsorbed to particles (Mauvisseau *et al*., 2022; Nwosu *et al*., 2021). Furthermore, some sedDNA may originate from modern microbial communities that are active within the sediment and obscure the temporal signal derived from relic DNA (Capo *et al*., 2021). sedDNA-based reconstructions of eukaryotic and cyanobacterial communities are similar to those reconstructed based on sedimentary pigments (Picard *et al*., 2022a; Tse *et al*., 2018) and diatom frustules (Anslan *et al*., 2022). However, biases may also arise in pigment and diatom frustule-based records due to the effects of differential deposition and degradation (Leavitt and Carpenter, 1989; Reuss *et al*., 2005). Therefore, detailed comparisons with long-term monitoring records of microbial communities are needed to validate sedDNA as an effective and reliable biomarker.

1.4. Microbial lipid biomarkers for the reconstruction of past climates

1.4.1. Homeoviscous membrane adaptation

In addition to providing information about the presence and abundance of the source organism, the structure of some microbial lipids can also provide information about past and present climates due to their involvement in homeoviscous membrane adaptation (Ouyang *et al*., 2015).

Homeoviscous adaptation is the mechanism by which microbes adjust their membrane lipid composition to maintain optimal membrane fluidity in response to temperature, pH and other environmental conditions (Ernst *et al*., 2016). A high temperature can cause membranes to become too fluid and compromise their integrity. Long, straight-chained and saturated lipids form an ordered structure with a high density of intermolecular interactions, and incorporating a greater proportion of these lipids can therefore stabilise membranes at a high temperature. Conversely, a low temperature can cause membranes to become too rigid and reduce permeability. Incorporating a greater proportion of short, branched and unsaturated lipids can disrupt the organised structure of the membrane and therefore improve membrane fluidity at a low temperature (De Carvalho and Caramujo, 2018; Ernst *et al*., 2016). Maintaining an optimal intracellular pH is critical for all cellular processes. Microbes in an acidic environment experience a steeper proton gradient, and to prevent an excessive influx of protons, microbes incorporate a higher proportion of long, straight-chained and saturated lipids to reduce membrane permeability (Guan and Liu, 2020). However, while homeoviscous membrane adaptation in response to temperature is highly conserved among many microbes, the adaptive membrane response to pH may differ between acidophilic, neutrophilic and alkaliphilic microbes (Siliakus *et al*., 2017).

1.4.2. Microbial lipids as palaeoclimate biomarkers

Numerous classes of microbial membrane lipids are involved in homeoviscous adaptation (Siliakus *et al*., 2017), some of which have applications as palaeoclimate biomarkers (Volkman and Smittenberg, 2017). Glycerol dialkyl glycerol tetraethers (GDGTs) are membrane-spanning lipids produced by some bacteria and archaea (Weijers *et al*., 2007). GDGTs are some of the most widely applied microbial lipid palaeoclimate biomarkers (Ouyang *et al*., 2015). GDGTs

containing isoprenoid moieties (isoGDGTs) are produced by archaea such as Euryarchaeota and Thaumarchaeota and when grown at higher temperatures, archaea produce isoGDGTs containing more cyclopentane rings than at lower temperatures (Kim *et al*., 2008). The tetraether index based on the degree of cyclisation of isoGDGTs with 86 carbons (TEX_{86}) is positively correlated with temperature and has been used to estimate past sea surface temperature (SST) (Schouten *et al*., 2002) and lake surface temperature (Powers *et al*., 2010).

Branched GDGTs (brGDGTs) are thought to be derived from bacteria, and proxies based on brGDGTs have also been developed for use in terrestrial environments, including the cyclisation of brGDGTs (CBT index) which is negatively correlated with soil pH, and the degree of methyl branching of brGDGTs (MBT index) which is positively correlated with mean annual air temperature (MAAT) (Weijers *et al*., 2007). While TEX_{86} has been applied to marine environments to reliably reconstruct past SST (Huguet *et al*., 2006; O'Brien *et al*., 2017; Robinson *et al*., 2016), the application of TEX_{86}, CBT and MBT in a lake environment may be unreliable, possibly due to inputs from multiple, poorly characterised sources of GDGTs which may include in situ lake production or run-off from the surrounding catchment (Naeher *et al*., 2014; Powers *et al*., 2010; Sinninghe Damsté *et al*., 2012; Tierney and Russell, 2009). The reliability of GDGT-based palaeoclimate reconstructions appears to be strongly influenced by seasonality (Loomis *et al*., 2012), and iso and brGDGT-based temperature proxies have been found to perform poorly in eutrophic lakes compared to less nutrient-rich lakes, although the reasons for this are currently unclear (Blaga *et al*., 2009). The application of brGDGTs as palaeoclimate biomarkers is further limited by uncertainties regarding their bacterial producers. Efforts to constrain their producers have focussed on Acidobacteria (De Jonge *et al*., 2019), but brGDGTs or precursors of brGDGTs have only been identified in a small number of

Acidobacteria strains (Halamka *et al*., 2022; Sinninghe Damsté *et al*., 2018), and production of brGDGTs by other bacterial taxa is yet to be confirmed (Halamka *et al*., 2022).

These limitations of GDGT-based palaeoclimate biomarkers highlight the need for additional lipid biomarkers which are produced by the wider bacterial community and can be applied to both terrestrial and marine environments to reliably reconstruct past climates. 3-OH FAs are thought to be produced by most gram-negative bacteria and are ubiquitous and abundant in terrestrial and marine environments (Wang *et al*., 2016). 3-OH FAs are therefore excellent candidates for development as palaeoclimate biomarkers.

1.4.3. Structure and function of 3-OH FAs

3-OH FAs are a class of fatty acids with a hydroxyl group on the third carbon of the carbon chain. Different structural isomers exist, including *normal* 3-OH FAs with no methyl branching, *iso* 3-OH FAs with a methyl branch on the penultimate carbon and *anteiso* 3-OH FAs with a methyl branch on the third carbon from the terminus (Fig. 1.2). The 3-OH FA carbon chain may be saturated or unsaturated (Christie and Han, 2012).

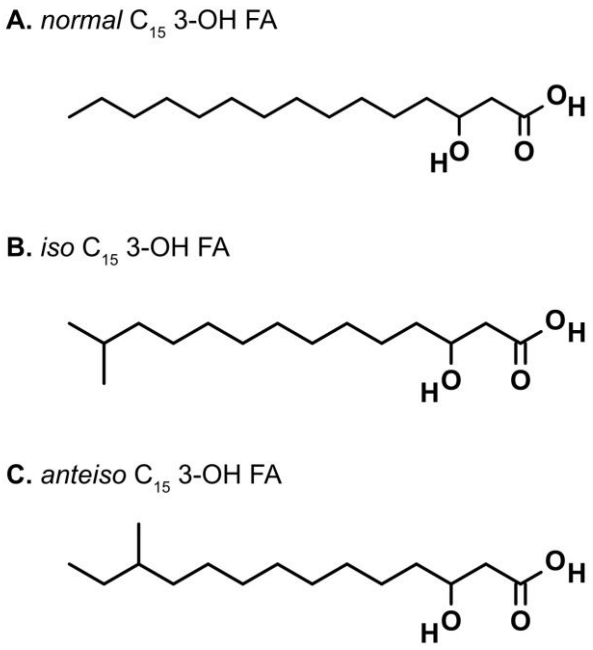

Figure 1.2. Structure of *normal* (A), *iso* (B) and *anteiso* (C) saturated C_{15} 3-OH FAs.

3-OH FAs with carbon chain lengths ranging between C_{10} and C_{18} are predominantly produced by bacteria (Wilkinson, 1988). Other sources of 3-OH FAs include some fungi with 3-OH FA carbon chain lengths between C_3 and C_{20} (Kock *et al*., 2003; Schwarz *et al*., 2004; Van Dyk *et al*., 1994), eukaryotic algae with 3-OH FA carbon chain lengths between C_{10} and C_{30} (Matsumoto and Nagashima, 1984; Volkman *et al*., 1999), and plants with 3-OH FA carbon chain lengths between C_{14} and C_{28} (Dumri *et al*., 2008; Özen *et al*., 2004; Racovita *et al*., 2015), although 3-OH FAs are not considered to be major fatty acids in these organisms.

Bacterial 3-OH FAs are found in the outer gram-negative membrane where they bind via ester or amide bonds to the glucosamine unit of lipid A in lipopolysaccharides (Caroff and Karibian, 2003). Gram-positive bacteria lack an outer membrane and therefore do not have the associated membrane-bound 3-OH FAs (Silhavy *et al*., 2010). However, both gram-negative and gram-positive bacteria can also produce 3-OH FAs as a component of biosurfactants which promote the formation of and adhesion to biofilms (Besson *et al*., 1992; Nakagawa and Matsuyama,

1993; Ron and Rosenberg, 2001; Youssef *et al*., 2005). Other biological functions of bacterial 3-OH FAs include their role as signalling molecules involved in the regulation of gene expression (Fetzner, 2015; Flavier *et al*., 1997), their ability to inhibit the growth of fungi (Pohl *et al*., 2011; Sj gren *et al*., 2003), and similar to many other classes of membrane lipids, 3-OH FAs also have a key role in regulating the fluidity of gram-negative bacterial membranes in homeoviscous adaptation (Hellequin *et al*., 2023; Russell and Fukunaga, 1990; Siliakus *et al*., 2017). Bacteria can incorporate a higher proportion of longer-chained and *normal* 3-OH FAs to prevent the membrane from becoming too fluid at a high temperature and to reduce permeability at an acidic pH. When exposed to low temperatures, a higher proportion of shorter-chained and branched *iso* and *anteiso* 3-OH FAs can restore membrane fluidity. Furthermore, *anteiso* 3-OH FAs have a stronger influence on membrane fluidity compared to *iso* 3-OH FAs (Kaneda, 1991).

1.4.4. Biomarker applications of 3-OH FAs

The detection of 3-OH FAs is considered to be diagnostic for the presence of gram-negative bacteria, and bacterial 3-OH FAs therefore have numerous applications as biomarkers in clinical and environmental settings (Li *et al*., 2010). For example, bacterial lipopolysaccharides, also known as endotoxins, can elicit immune responses in mammalian hosts (Sampath, 2018). The 3-OH FAs associated with lipopolysaccharides can therefore be used as biomarkers in the diagnosis of bacterial infections (Ferrando *et al*., 2005; Kraśnik *et al*., 2006; Szponar *et al*., 2012). 3-OH FAs have also been used as biomarkers of the presence of gram-negative bacteria in a wide range of environmental samples including soils, lake sediments, marine sediments and bioaerosols (Blumenberg *et al*., 2012; Keinänen *et al*., 2003; Koponen *et al*., 2006; Lee *et al*., 2004; Zelles *et al*., 1992). Furthermore, the amount of membrane-bound 3-OH FAs per

bacterial cell is relatively consistent, making 3-OH FAs a useful biomarker of gram-negative bacterial biomass (Keinänen *et al*., 2003; Lee *et al*., 2004). Spatial or temporal changes in 3-OH FA distribution in environmental samples may indicate a shift in bacterial community composition (Zogg *et al*., 1997). However, while 3-OH FA composition may be similar among closely related taxa, their use as a taxonomic biomarker is limited because knowledge of the 3-OH FAs which may be characteristic of certain taxa is restricted to a relatively small proportion of cultured strains (Zelles, 1999). Recently, gram-negative bacterial 3-OH FAs have gained particular interest as novel palaeoclimate biomarkers (Wang *et al*., 2016).

1.4.5. Development of 3-OH FAs as palaeoclimate biomarkers

Wang *et al*. (2016) were the first to generate proxies based on 3-OH FAs which are responsive to MAAT and soil pH and therefore have the potential to be used as palaeoclimate biomarkers. In Chinese soils, the relative abundances of *iso*, *anteiso* and *normal* 3-OH FA isomers with carbon chain lengths between C_{10} and C_{18} were determined and combined in a series of indices. The strength of the correlation between each index and environmental conditions such as MAAT and soil pH was assessed. Indices with the strongest and most significant correlations were investigated further and proposed as potential proxies of MAAT or soil pH. This included the ratio of *anteiso* to *normal* C_{15} and C_{17} 3-OH FAs (RAN_{15} and RAN_{17}, respectively) which were negatively correlated with MAAT, and the negative logarithm of the ratio of *iso* and *anteiso* to *normal* 3-OH FAs (RIAN index) which was negatively correlated with soil pH (Wang *et al*., 2016). The performance of RAN_{15}, RAN_{17} and RIAN indices as MAAT and soil pH proxies were tested in soils from Tanzania, Italy (Huguet *et al*., 2019) and France (Véquaud *et al*., 2021), revealing the existence of region-specific variation between each MAAT and soil pH calibration. The source of this variation is yet to be determined, and has limited the

generation of a reliable global soil calibration of RAN_{15}, RAN_{17} and RIAN (Véquaud *et al*., 2021; Wang *et al*., 2021).

Additional 3-OH FA proxies have been proposed for specific application to lake (Yang *et al*., 2021) and marine (Dong *et al*., 2023; Yang *et al*., 2020) environments. The RAN index based on C_{13} 3-OH FAs (RAN_{13}) and the ratio of *iso* to *normal* C_{17} 3-OH FAs (RIN_{17}) had stronger negative correlations with MAAT in lake sediments compared to RAN_{15} and RAN_{17} developed for use in soils (Yang *et al*., 2021). RAN_{13} was also negatively correlated with SST in marine sediments (Yang *et al*., 2020). The relative abundance of *iso*, *anteiso* and *normal* C_{12}, C_{13} and C_{14} 3-OH FAs in marine sediments was found to be strongly influenced by SST, leading to the proposal of RAN_S, which is based on the ratio of these short-chained isomers and negatively correlated with SST, as a 3-OH FA proxy specific to marine sediments (Dong *et al*., 2023).

1.4.6. Palaeoclimate reconstructions based on 3-OH FAs

The first palaeoclimate reconstruction based on 3-OH FAs was performed on an annually laminated cave stalagmite in central China (Wang *et al*., 2018). The stalagmite record represented an accumulation of material from overlying soil deposited via cave drip water, and the 3-OH FAs extracted from radiometrically dated sections of the stalagmite were therefore assumed to be largely derived from this overlying soil. The RAN_{15}, RAN_{17} and RIAN indices were calculated and used to estimate past MAAT and pH according to the calibrations previously performed in Chinese soils (Wang *et al*., 2016). This analysis revealed a warm period during the early to middle Holocene, followed by a cooler period during the late Holocene. Reconstructed pH was used to predict past hydrological changes, suggesting that two

long wet periods and one drier period may have occurred during the Holocene (Wang *et al*., 2018).

3-OH FAs have also been extracted from sections of a marine sediment core to reconstruct past SST over a 70-year period (Yang *et al*., 2020). The RAN_{13} index which has been shown to perform well in aquatic environments (Yang *et al*., 2020, 2021), and the RAN_S index developed for use in marine sediments (Dong *et al*., 2023) were each calibrated with SST in marine surface sediments and applied down-core to estimate past SST. The RAN_{13} and RAN_S-reconstructed trends in SST were broadly comparable with a long-term instrumental record of SST and performed better than TEX_{86} as a proxy of past SST (Dong *et al*., 2023; Yang *et al*., 2020).

3-OH FAs have been shown to have significant potential as palaeoclimate biomarkers (Wang *et al*., 2018; Yang *et al*., 2020), but their performance as palaeoclimate biomarkers is yet to be assessed in lake sediment cores. Furthermore, the reliability of palaeoclimate reconstructions is dependent on the performance of 3-OH FA calibrations, which have been shown to vary substantially between regions (Véquaud *et al*., 2021; Wang *et al*., 2021) and different environments (Wang *et al*., 2016; Yang *et al*., 2020, 2021). The source of this variation is currently unknown but could be attributed to variation in bacterial community composition. However, the bacterial producers of 3-OH FAs in different environments and how they respond to changes in environmental conditions are poorly understood (Wang *et al*., 2021; Yang *et al*., 2020, 2021). Identification of the main bacterial producers of each 3-OH FA isomer and a more comprehensive understanding of their ecology and physiological responses to environmental conditions are crucial in the development of 3-OH FAs as reliable biomarkers for palaeoclimate reconstruction.

1.5. Thesis outline and aim

The aim of this thesis is to further develop two novel microbial biomarkers, sedDNA for reconstructing past microbial communities, and 3-OH FAs for reconstructing past climates. Chapters 2 and 3 will focus on validating sedDNA as a reliable palaeolimnological record of bacterial and phytoplankton communities, respectively. Chapter 4 will address the source and fate of bacterial and eukaryotic microbial sedDNA in lake sediments. Chapter 5 will explore the influence of the bacterial community on 3-OH FAs in the environment, while Chapter 6 will focus on identifying the bacterial producers in laboratory cultures. Finally, Chapter 7 will bring together the conclusions from each chapter, discuss the implications for the use of sedDNA and 3-OH FAs in palaeolimnology and palaeoclimatology, and provide recommendations for future research to help further their development as effective and reliable biomarkers.

Chapter 2: Sedimentary DNA records long-term changes in a lake bacterial community in response to varying nutrient availability

This chapter has been published in Environmental DNA as: Thorpe, A. C., Anderson, A., Goodall, T., Thackeray, S. J., Maberly, S. C., Bendle, J. A., Gweon, H. S. and Read, D. S. (2022) Sedimentary DNA records long-term changes in a lake bacterial community in response to varying nutrient availability. *Environmental DNA*, 4 (6), 1340-1355. https://doi.org/10.1002/edn3.344.

2.1. Abstract

Microbial communities play important roles in lake ecosystems and are sensitive to environmental change. However, our understanding of their responses to long-term change such as eutrophication is limited as long-term lake monitoring is rare, and traditional palaeolimnological techniques (pigments and microfossils) are restricted to a low taxonomic resolution, or organisms with well-preserved structures. Sedimentary DNA (sedDNA) is a promising technique to reconstruct past microbial communities in sediments, but taphonomic processes and the ability of sedDNA to record bacterial pelagic history accurately are largely unknown. Here, we sequenced the 16S rRNA gene in triplicate sediment cores from Esthwaite Water (English Lake District) which has concurrent long-term monitoring and observational data. The sediment record spanned 113 years and included an episode of increased nutrient availability from the 1970s, followed by a more recent decline. Trends in bacterial community composition were broadly similar among the three sediment cores. Cyanobacterial richness in the sediment cores correlated significantly with that of cyanobacteria in a 65-year microscopy-based monitoring record, and some known pelagic bacterial taxa were detected in the sediment. sedDNA revealed distinct shifts in community composition in response to changing lake physicochemical conditions. The relative abundance of cyanobacteria closely reflected nutrient enrichment, and Proteobacteria, Bacteroidetes and Verrucomicrobia were relatively more abundant in recent sediments, while Chloroflexi, Firmicutes, Acidobacteria, Nitrospirae, Spirochaetes and Planctomycetes declined in more recent sediments. Following lake restoration efforts to reduce nutrient enrichment, the relative abundance of cyanobacteria returned to pre-1970 levels, but the bacterial community did not fully recover from the period of intense eutrophication within the time scale of our study. These results suggest that sedDNA is a valuable approach to reconstruct lake microbial community composition over the 100-year time

scale studied, but an improved understanding of DNA deposition and degradation is required to further the application of sedDNA in palaeolimnology.

2.2. Introduction

Microbial communities form an integral component of lake ecosystems due to their diverse roles in nutrient cycling and position near the base of food webs (Newton *et al*., 2011). However, lake microbial communities are highly sensitive to environmental perturbations (Zingel *et al*., 2018), and many lakes worldwide are experiencing a complex mix of interacting pressures from the effects of human activities, such as nutrient enrichment and climate change (Birk *et al*., 2020). Eutrophication from increased nutrient pollution can change ecosystem dynamics, lead to a change and loss of biodiversity, and increase the frequency and severity of cyanobacterial blooms which can produce toxic compounds that are harmful to wildlife and humans (Battarbee *et al*., 2012). Climate change can exacerbate the effects of eutrophication, which further threatens the diverse biota that lakes support, and can have consequences for biogeochemical cycles (Davidson and Jeppesen, 2013; Richardson *et al*., 2019). Effective management of lakes is therefore crucial to maintain these important ecosystems, but relies on a comprehensive understanding of how lake microbial communities respond to environmental change.

Lake microbial community dynamics have previously been studied on relatively short temporal scales of less than 10 years (Rösel *et al*., 2012; Shade *et al*., 2007), but multi-decadal time series are required to understand how communities respond to, and recover from, long-term drivers such as eutrophication and climate change. Such knowledge can inform predictions of how ecosystems may respond to future change, and provide a baseline for lake restoration (Bennion

et al., 2011; Maberly and Elliott, 2012). However, long-term and high-resolution monitoring records for lakes are rare (Battarbee *et al*., 2012; Dong *et al*., 2012). Sediments can provide a long-term record of past communities, but traditional palaeolimnological techniques are restricted to groups of organisms which leave well-preserved and morphologically distinct remains in sediments (Hobbs *et al*., 2010), or those whose abundance can be inferred from proxies such as pigments (Moorhouse *et al*., 2014).

Molecular-based techniques utilising DNA preserved in lake sediments offer an opportunity to study a wider diversity of organisms with a higher taxonomic resolution compared to traditional palaeolimnological techniques (Domaizon *et al*., 2017). DNA from living organisms is deposited in lake sediment where it is progressively buried and preserved over long periods of time. DNA extracted from different depths within sediment cores can therefore reflect the community present at the time of deposition, and the vertical organisation of sedimentary DNA (sedDNA) can be used to reconstruct temporal patterns in community composition (Capo *et al*., 2021). Lake sedDNA has previously been used to investigate changes in the diversity and community composition of eukaryotic microbes and cyanobacteria in response to trophic status and climate warming over periods ranging between 100 and 2200 years (Capo *et al*., 2016; Domaizon *et al*., 2013; Ibrahim *et al*., 2020; Monchamp *et al*., 2016, 2019). Few sedDNA studies have focused on lake bacterial communities, but Li *et al*. (2019a) demonstrated that sedDNA can be used to explore how different bacterial groups responded to nutrient and heavy metal pollution in a lake in China over a period of 150 years.

Previous sedDNA records for eukaryotic algae and cyanobacteria correlate with sediment core pigment records (Pal *et al*., 2015; Tse *et al*., 2018), and the community detected in the sediment closely reflects the phytoplankton community in the water column according to DNA sequencing (Capo *et al*., 2015) and microscopic analysis of pelagic communities (Monchamp

et al., 2016). However, the extent to which sedimentary records represent the pelagic bacterial community is largely unknown. Due to their small size, bacteria may not deposit in the sediment efficiently, and those that do settle may be mostly particle-associated bacteria whereas more buoyant cells could be flushed from the lake (Thupaki *et al*., 2013; Vuillemin *et al*., 2017). The contribution of the active, in situ sediment microbial community to the sedDNA signal compared to that deposited over time by the pelagic community is also poorly understood (Capo *et al*., 2021; Wurzbacher *et al*., 2017). The taphonomic processes which influence DNA preservation, such as oxidation, hydrolysis, decomposition by heterotrophs and grazing by zooplankton, likely differ in the water column compared to that in the lake sediment (Giguet-Covex *et al*., 2019; Nwosu *et al*., 2021), and as a water-soluble molecule, free, extracellular DNA may be highly unstable in the water column (Mauvisseau *et al*., 2022). DNA originating from different sources may therefore experience different rates of degradation. The extent to which degradation of DNA over time may confound the sedDNA signal, and whether DNA from certain microbial taxa is more susceptible to degradation is relatively unknown (Boere *et al*., 2011). Furthermore, many palaeolimnology studies have used a single sediment core with the assumption that DNA is deposited in a homogenous manner in lakes, but whether different sediment cores show spatial heterogeneity is poorly understood (Capo *et al*., 2021). Lakes with detailed long-term monitoring and observational records offer an ideal opportunity to address these uncertainties and assess the reliability of sedDNA in palaeolimnology.

Our study is focussed on Esthwaite Water, a lake in the Windermere catchment of the English Lake District. It is one of the best studied lakes in the world, as lake conditions have been continually monitored since 1945 and the long history of human activity in the catchment is well-documented (Dong *et al*., 2011). Prior to the 1970s, there was a gradual increase in nutrient concentrations in Esthwaite Water, following inputs from lowland pasture and fertiliser run-

off, and sewage effluent from local villages (Talling and Heaney, 1983). Nutrient enrichment in Esthwaite Water accelerated dramatically since the 1970s, with the construction of wastewater treatment works in 1973, followed by the establishment of a fish farm in the lake in 1981 (Dong *et al*., 2011). Esthwaite Water is one of the most nutrient-enriched lakes in the Lake District, and blooms of cyanobacteria have frequently been observed in recent decades (Dong *et al*., 2012). Efforts have since been made to reduce nutrient enrichment in the lake, with tertiary wastewater treatment initiated in 1986, closure of the fish farm in 2009, and further upgrades to the wastewater treatment works in 2010 (Dong *et al*., 2011, Maberly *et al*., 2011a). However, continued release of phosphorus from sediments may delay recovery of the lake from this intense period of nutrient enrichment (Dong *et al*., 2012). Like other lakes worldwide (Maberly *et al*. 2020), Esthwaite Water is also increasingly influenced by climate change.

Eukaryotic algae and cyanobacteria in Esthwaite Water and the neighbouring lakes in the Lake District have previously been studied using observational records (Feuchtmayr *et al*., 2012; Talling and Heaney, 2015; Thackeray *et al*., 2008). Lake sediment cores from Esthwaite Water have extended these records over longer time periods. Cyanobacterial and algal pigments have shown an overall increasing trend from 1800 to the 2000s (Moorhouse *et al*., 2018), and analysis of diatom frustules revealed distinct shifts in the diatom community over 1200 years (Dong *et al*., 2011). However, previous records have largely been limited to organisms such as diatoms which can be readily identified using traditional palaeolimnological techniques (Bennion *et al*., 2000; Dong *et al*., 2011, 2012). The long-term temporal dynamics of other lake microorganisms, such as bacteria which do not leave well-preserved or morphologically distinct remains, are therefore poorly understood both in the Lake District (Rhodes *et al*., 2012), and globally (Billard *et al*., 2015; Thomas *et al*., 2019).

The present study combines a lake sedDNA record of 113 years with detailed, long-term monitoring records to: (i) assess whether sedDNA is an effective and reliable tool to reconstruct past microbial community composition in lake sediments, (ii) investigate how the bacterial and cyanobacterial community have responded to past human activity and environmental change, and (iii) assess the resilience of lake microbial communities to lake restoration.

2.3. Materials and methods

2.3.1. Study site

Esthwaite Water (54° 21.56' N, 2° 59.15' W) is a relatively small lake in the Windermere catchment of the Lake District National Park, Cumbria, UK (Appendix A, Fig. S1). The lake has a surface area of 0.96 km^2, a volume of 6.7 $\times 10^6$ m^3, mean and maximum depths of 6.9 m and 16 m, respectively, and an average retention time of 91 days. Esthwaite Water has a catchment area of 17 km^2, which is primarily comprised of improved grassland and broadleaf forest (Maberly *et al*., 2011a). There is a well-documented history of human activities and intense eutrophication in Esthwaite Water and its catchment (Dong *et al*., 2011).

2.3.2. Long-term lake monitoring

Physical, chemical and biological conditions in Esthwaite Water have been continually monitored on a weekly or fortnightly basis since 1945 by the Freshwater Biological Association (FBA) until 1989, and then by the UK Centre for Ecology & Hydrology (UKCEH). The dataset used in the present study covered the period from 1945 to 2015 (Maberly *et al*., 2017). Surface water samples and measurements, integrated over 0-5 m depth, were collected from the deepest point of Esthwaite Water, including surface water temperature, surface water pH, alkalinity,

and the concentration of total phosphorus (TP), soluble reactive phosphorus (SRP), nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N), and chlorophyll *a*. Winter SRP was calculated as the mean SRP from December to February because winter SRP represents phosphorus availability before substantial uptake in the lake (Dong *et al*., 2012; Sutcliffe *et al*., 1982). An annual mean was calculated for each variable measured.

2.3.3. Phytoplankton monitoring

Phytoplankton in Esthwaite Water was monitored on a weekly or fortnightly basis, and the record of cyanobacteria from 1945 to 2010 was used in the present study. A sub-sample of each integrated depth water sample used for chemical analysis was preserved in Lugol's iodine and concentrated by sedimentation. Phytoplankton within these concentrated sub-samples were identified to the highest taxonomic level possible under a microscope, and their presence was recorded. The mean morphospecies richness of cyanobacteria was then calculated for each year.

2.3.4. Sediment coring

A HTH 90 mm diameter gravity corer (Pylonex, Sweden) was used to collect three 30 cm sediment cores. Prior to sampling, holes were drilled into the Perspex core tubes at 1 cm intervals and sealed with heavy-duty waterproof tape. All equipment was cleaned with sodium hypochlorite and Decon 90 cleaning agent, and thoroughly rinsed with deionised water before use. The three cores were collected from the deepest point of Esthwaite Water in 2016 and transported to a nearby laboratory managed by the FBA. To prevent contamination, sterile syringes were used to pierce the tape and extract sediment from the centre of each core via the pre-drilled holes, starting from the top of the core and working downwards. The sediment

samples were stored in sterile Eppendorf tubes at -20 °C prior to DNA extraction. A reference core, used for sediment chronology, was collected from the deepest part of Esthwaite Water in 2014 with a gravity corer. This core was sectioned every 0.5 cm with the core extrusion kit, and each section was stored in sterile sample bags and then freeze-dried.

2.3.5. Sediment core chronology

The freeze-dried sediments from the reference core were radiometrically dated at the Environmental Radioactivity Research Centre, University of Liverpool, UK, where they were analysed for the radioactive isotopes, ^{210}Pb, ^{226}Ra, ^{137}Cs and ^{241}Am, by direct gamma assay using Ortec HPGe GWL series well-type coaxial low background intrinsic germanium detectors (Appleby *et al*., 1986, 1992). Radionuclide concentrations with depth for the reference core are given in Appendix A, Table S1. Chronology of the reference sediment core was estimated by the sedimentation rate and ^{210}Pb dates calculated according to the CRS dating model. This model assumes a constant rate of supply of ^{210}Pb to the sediment and allows for a variable sedimentation rate (Appleby and Oldfield, 1978). The chronology, validated using ^{137}Cs dates as reference points which recorded radionuclide fallout incidents (Appleby, 2001), is presented in Appendix A, Table S2. Sample depths for the reference core collected in 2014 were corrected to 2016 using the sedimentation rate to allow for comparison with the sediment cores collected in 2016 (Appendix A, Table S3). The slope and intercept of the age-depth relationship as determined for the reference core (Appendix A, Fig. S2) was then used to estimate the age of each section of the three sediment cores used for DNA sequencing. Sample ages were rounded to the nearest calendar year, and the three sediment cores covered a combined period of 113 years from 1903 to 2016 (Appendix A, Table S4).

2.3.6. DNA extraction, PCR amplification and sequencing

DNA was extracted from the sediment core samples using the Qiagen DNeasy PowerSoil kit according to the manufacturer's protocol (Qiagen, Germany). The concentration and purity of the extracted DNA were checked using the NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, MA, U.S.). The V4 region of the 16S rRNA bacterial gene was amplified with primers, 515F (Forward: GTGYCAGCMGCCGCGGTAA) and 806R (Reverse: GGACTACNVGGGTWTCTAAT) (Walters *et al*., 2016). Each 50 µL PCR reaction contained 0.25 µL of 5 units μL^{-1} Taq DNA polymerase (Sigma-Aldrich, UK), 5 µL of 10x PCR reaction buffer, 0.5 µL of 20 mg mL^{-1} BSA (New England Biolabs, UK), 1 µL of a 10 mM dNTP mix (Bioline, UK), 40.05 µL of molecular grade water, 0.1 µL of the forward and reverse primers, and 3 µL of DNA. The PCR program was set to an initial denaturing temperature of 95 °C for 2 min, followed by 30 cycles of 95 °C for 15 sec, an annealing temperature of 50 °C for 30 sec, an extension temperature of 72 °C for 30 sec, and then a final extension temperature of 72 °C for 10 min. Successful PCR amplification was confirmed with an agarose gel. PCR product was purified with the Zymo DNA clean-up kit according to the manufacturer's protocol (Zymo Research, CA, U.S.).

Second step PCR was performed using a dual-indexing approach (Kozich *et al*., 2013), and 50 µL PCR reactions contained 0.25 µL Taq DNA polymerase, 5 µL PCR reaction buffer, 1 µL dNTPs, 37.75 µL molecular grade water, 5 µL of the indexing primers (Kozich *et al*., 2013), and 1 µL of purified PCR product from the first PCR step. The second step PCR program was set to an initial denaturing temperature of 95 °C for 2 min, followed by 8 cycles of 95 °C for 15 sec, an annealing temperature of 50 °C for 30 sec, an extension temperature of 72 °C for 30 sec, and then a final extension temperature of 72 °C for 10 min. Successful PCR amplification from the second PCR step was confirmed with an agarose gel.

PCR product from the second PCR step was normalised using the Invitrogen SequalPrep normalisation kit according to the manufacturer's protocol (Thermo Fisher Scientific, MA, U.S.). All samples were pooled, and the concentration of pooled DNA was quantified using the Invitrogen Qubit dsDNA HS assay kit with the Qubit 3.0 fluorometer. The pooled amplicon library was denatured with NaOH, and combined with 4% denatured PhiX. The library was then diluted with HT1 buffer (Illumina, CA, U.S.), heat denatured at 96 °C for 5 min and immediately transferred to an ice bath. The denatured library was loaded and sequenced on the Illumina MiSeq Platform (250PE).

2.3.7. Data processing

The sequences were quality filtered and adapters were removed using TrimGalore (https://github.com/FelixKrueger/TrimGalore). The resulting quality-filtered reads were processed with R using the DADA2 pipeline (v1.14.1, Callahan *et al*., 2016) generating an amplicon sequence variant (ASV) abundance table. Each ASV was classified using the naive Bayesian classifier (Wang *et al*., 2007) against the SILVA database (v.132, Quast *et al*., 2013) for kingdom to species assignments. The sequences were rarefied to a uniform sequencing depth of 19,135 reads, and filtered based on taxonomy to include only bacteria and ASVs with more than 10 reads.

2.3.8. Data analysis

Non-metric multidimensional scaling (NMDS) based on a beta diversity Bray-Curtis dissimilarity matrix was performed to visualise variation in community composition in each sediment core with sample age and physicochemical conditions of the lake. The envfit function

in the vegan R package (Oksanen *et al*., 2019) was used to fit vectors for these variables and identify which variables correlated significantly with the beta diversity dissimilarity matrix. Analysis of similarity (ANOSIM) with 999 permutations was used to determine the degree of separation between the three sediment core communities. Alpha diversity in each sample was calculated as the Shannon diversity index, and long-term change in diversity and richness measured at the ASV level throughout the sediment cores was assessed by fitting a linear model of these measures against year.

To assess the ability of sedDNA to record the pelagic cyanobacterial community, ASV richness of cyanobacteria in the sediment was calculated and compared against morphospecies richness of cyanobacteria in the water column. To account for potential inaccuracies in the chronology of the sediment cores, the mean morphospecies richness of cyanobacteria in the water column was calculated for the year before, year of, and year after the corresponding sediment samples, and the mean ASV richness of cyanobacteria in the sediment was calculated for the three sediment core samples within this three-year period (Monchamp *et al*., 2016). As both measures of richness were dependent variables, a model II regression was fitted with the lmodel2 R package (Legendre, 2018).

To determine whether pelagic bacteria could be detected in the sedDNA record, the sedDNA sequences were compared against the FreshTrain database of known pelagic taxa using NCBI BLAST. This database is based on 1,318 16S rRNA reference sequences of heterotrophic bacteria from temperate lake epilimnia (Newton *et al*., 2011; Rohwer *et al*., 2018) and is available at: https://github.com/McMahonLab/TaxAss. Matches were filtered at 97% identity, and taxonomy was assigned based on the Silva SSU v138 database available at: https://www.arb-silva.de/documentation/release-138.

ASV abundances were converted to relative abundances in each sample, and ASVs were grouped at the phylum level. The ten most abundant phyla were chosen for in-depth analysis. Generalised additive models (GAMs) were fitted for each phylum to visualise the main underlying patterns in relative abundance throughout the sediment cores. The GAMs were also used to assess the consistency of these relative abundance trends among cores. To do this, GAMs were fitted with a global smooth combining data for each phylum from all cores (assuming similar patterns of change in all three cores), and then fitted again with individual smooths for each core (allowing core-specific patterns). The Akaike Information Criterion (AIC) was then used to assess which model gave the best fit (Burnham and Anderson, 2002). For all GAMs, restricted maximum likelihood (REML) was used as the smoothness selection method. Relative abundance GAMs were fitted with error distributions from the beta family with a logit link which is suitable for proportion data. GAMs were also used to visualise trends in ASV and morphospecies cyanobacterial richness, but with error distributions from the gamma family with a log link which is suitable for positively skewed, non-negative data (Anderson *et al*., 2010; Simpson, 2018). Spearman's correlation was used to investigate the relationship between the relative abundance of each phylum and sample age and lake physicochemical conditions, and the results were presented as a heat map. All data analysis was performed in R version 4.0.2 (R Core Team, 2020) using vegan (Oksanen *et al*., 2019), phyloseq (McMurdie and Holmes, 2013), microeco (Liu *et al*., 2021) and mgcv (Wood, 2020) packages.

2.4. Results

2.4.1. Long-term lake monitoring record

Over the 70-year period from 1945 to 2015, mean annual surface water temperature varied between 9.4 °C and 12.7 °C, and showed a general cooling trend until 1965, but then a warming trend to 2015 (Fig. 2.1A). Mean annual alkalinity declined from 374.7 µequiv L^{-1} in 1945 to 218.3 µequiv L^{-1} in 1960, and then increased to 436.8 µequiv L^{-1} in 2015 (Fig. 2.1B). Surface water pH was monitored from 1974 to 2015 and was relatively consistent with a geometric mean pH of 7.3 ± 0.21 (SD) over this period (Fig. 2.1C). Measurements of chlorophyll *a* began in 1964, and the mean annual concentration of chlorophyll *a* was highly variable but generally decreased from 26.7 µg L^{-1} in 1988 to 15.2 µg L^{-1} in 2015, with a large peak of 36.4 µg L^{-1} in 1999 (Fig. 2.1D).

Phosphorus concentrations were measured as SRP (Fig. 2.1E), winter SRP (Fig. 2.1F) and TP (Fig. 2.1G). From 1945 to 1969, SRP and winter SRP were relatively low, with a mean SRP over this period of 1.4 ± 0.3 µg L^{-1} and a mean winter SRP of 2.0 ± 1.3 µg L^{-1}. TP data were not available for this whole period, but the mean TP concentration was 21.8 ± 1.2 µg L^{-1} between 1945 and 1947. From 1970 to 2000, mean annual SRP and winter SRP increased. SRP peaked at 7.7 µg L^{-1} in 1982 and again at 8.5 µg L^{-1} in 1998, while winter SRP reached a peak of 18.4 µg L^{-1} in 2000. The mean annual TP concentration also showed an increasing trend from 1970 and peaked at 39.4 µg L^{-1} in 1998. From 2001 to 2015, mean annual SRP and winter SRP both showed decreasing trends to 2.7 µg L^{-1} and 9.1 µg L^{-1} in 2015, respectively. The mean annual TP concentration also decreased over this period to 12.6 µg L^{-1} in 2013, the lowest concentration observed since monitoring began in 1945.

The mean annual concentration of NO_3-N showed an increasing trend from 85.5 µg L^{-1} in 1945 to a peak of 790.6 µg L^{-1} in 1984, and then declined to 338.4 µg L^{-1} in 2015 (Fig. 2.1H). Monitoring of NH_4-N began in 1961, and there was a period of higher and more variable NH_4-N concentrations between 1986 and 1997, ranging between 45.9 µg L^{-1} and 97.3 µg L^{-1}, followed by a period of relatively lower NH_4-N between 1998 and 2015 when the concentration did not exceed 48.8 µg L^{-1} (Fig. 2.1I).

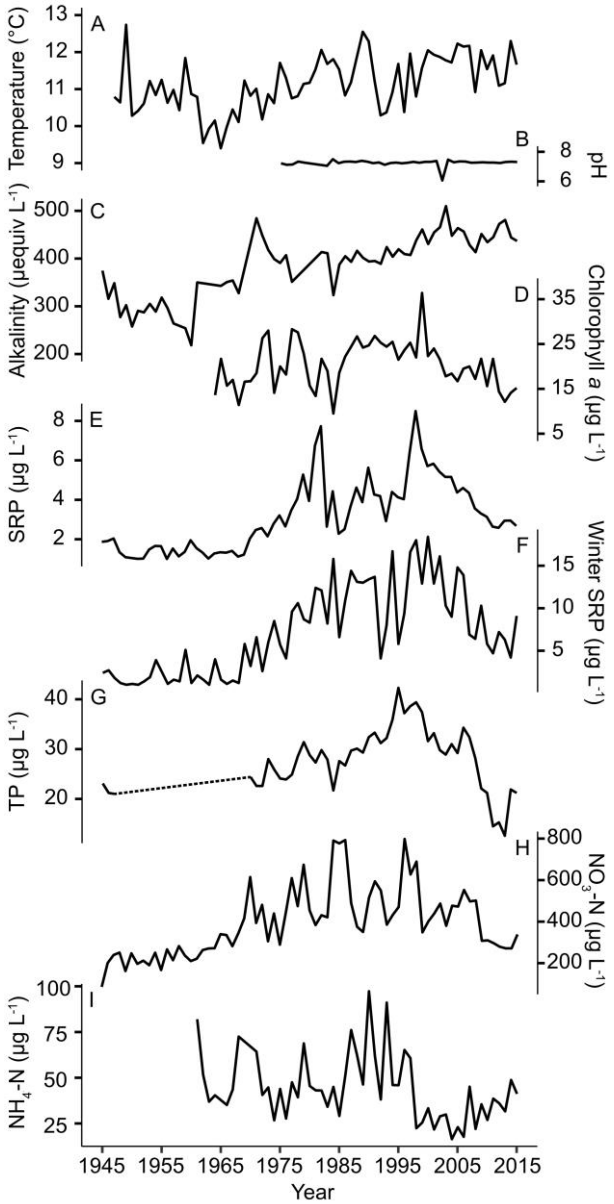

Figure 2.1. Mean annual conditions in Esthwaite Water between 1945 and 2015. (A) surface water temperature, (B) surface water pH, (C) alkalinity, (D) chlorophyll *a*, (E) soluble reactive phosphorus (SRP), (F) winter SRP, (G) total phosphorus (TP), (H) nitrate-nitrogen (NO_3-N), and (I) ammonium-nitrogen (NH_4-N).

2.4.2. Bacterial community composition and diversity

The beta diversity dissimilarity matrix was most strongly and significantly structured by sample age ($R^2 = 0.84$, $p < 0.001$) and alkalinity ($R^2 = 0.47$, $p < 0.001$) according to the permutation test (Fig. 2.2). Significant correlations were also identified with NO_3-N, winter SRP, TP ($p < 0.01$), chlorophyll *a* and SRP ($p < 0.05$) (Appendix A, Table S5). The ANOSIM R value of 0.01 ($p > 0.1$) suggested a high degree of similarity between each of the three sediment core communities.

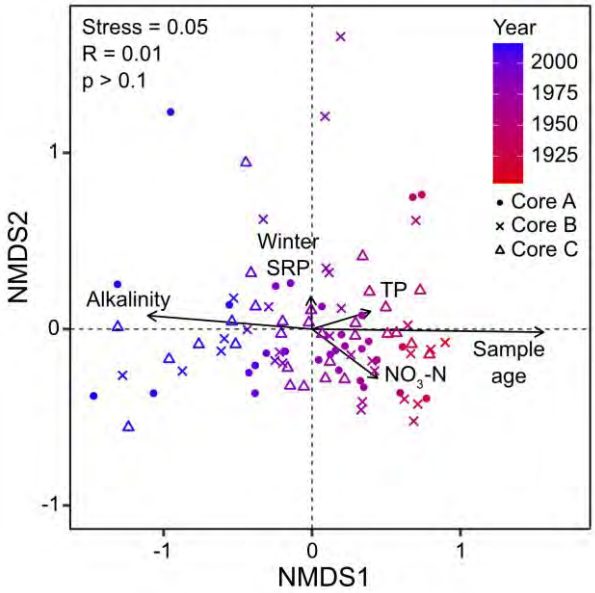

Figure 2.2. Non-metric multidimensional scaling (NMDS) of Bray-Curtis community composition in three sediment cores. For clarity, only the strongest ($R^2 > 0.20$) and most significant ($p < 0.01$) variables are displayed, including sample age, alkalinity, winter soluble reactive phosphorus (winter SRP), total phosphorus (TP) and nitrate-nitrogen (NO_3-N). ANOSIM R and P values are shown. The red to blue gradient indicates older to more recent sediment samples.

Shannon's diversity and richness for the whole community were relatively consistent throughout each sediment core with no significant change based on the linear regression tests ($R^2 = 0.00$, $F_{1,88} = 0.01$, $p > 0.1$ for diversity and richness) (Fig. 2.3A and B). Shannon's diversity ranged between 4.3 and 6.4, and richness ranged between 138 and 1429.

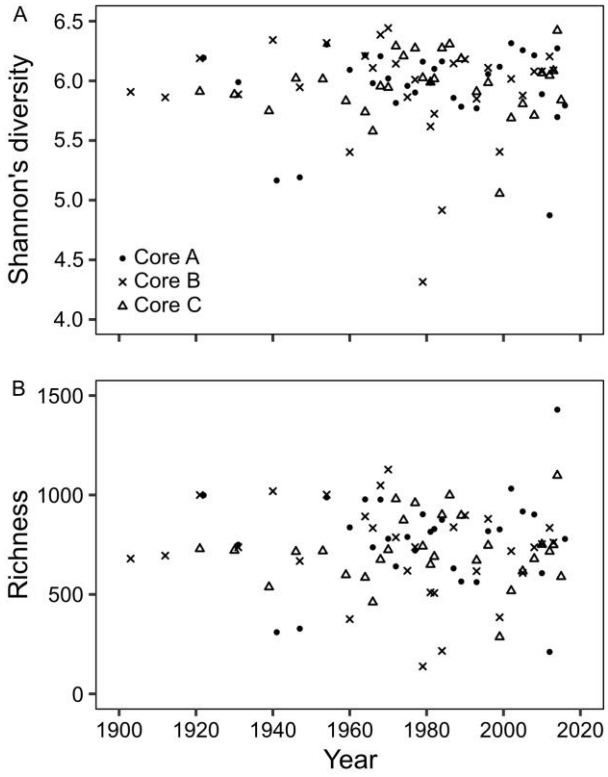

Figure 2.3. Shannon's diversity index (A) and richness (B) for the whole community throughout three sediment cores.

ASV richness of cyanobacteria in the sediment cores and morphospecies richness of cyanobacteria in the phytoplankton record both showed increasing trends from 1945 (Fig. 2.4A), and there was a significant and positive correlation between ASV richness and morphospecies richness ($R^2 = 0.60$, $F_{1,15} = 23.53$, $p < 0.001$) (Fig. 2.4B). The relationship was close to 1:1, but ASV richness was higher compared to morphospecies richness. Mean ASV richness in the sediment cores ranged between 3.0 and 11.3, while mean morphospecies richness in the water column ranged between 1.4 and 6.2.

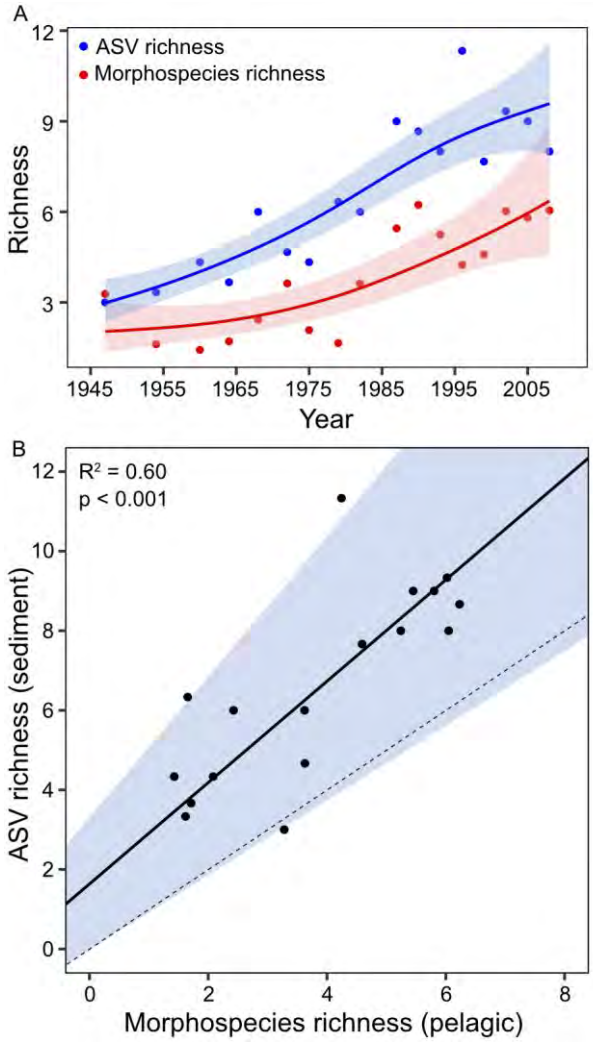

Figure 2.4. Generalised additive models (GAMs) of morphospecies richness of cyanobacteria in the water column and amplicon sequence variant (ASV) richness of cyanobacteria throughout the sediment cores (A), and correlation between morphospecies richness and ASV richness of cyanobacteria (B). Morphospecies richness is the running mean for three years, and ASV richness is the mean for three sediment core samples corresponding to the running mean. Dashed line shows the 1:1 relationship. Shaded area shows the 95% confidence interval.

2.4.3. sedDNA record of known pelagic ASVs

There were 23 sedDNA ASVs that were a match (> 97% identity) with known pelagic bacterial taxa in the FreshTrain database, and 15 of these ASVs were assigned to Proteobacteria, three to Bacteroidetes, three to Verrucomicrobia, and two to Actinobacteria (Appendix A, Table S6). These pelagic ASVs in the sedDNA record demonstrated distinct trends in relative abundance throughout the sediment cores, and selected ASVs which illustrate the range of trends observed

are presented in Fig. 2.5 (refer to Appendix A, Fig. S3 for remaining pelagic ASVs). The relative abundance of some pelagic ASVs increased over time, such as ASV 35 (*Methylobacter tundripaludum*) which increased from the 1930s to 2016 (Fig. 2.5A), and ASV 948 (*Ferruginibacter*) which was not detected until the late 1990s, but increased throughout the 2000s (Fig. 2.5B). Other pelagic ASVs had a distinct peak in relative abundance, such as ASV 290 (*Pseudarthrobacter*) which peaked from the 1960s (Fig. 2.5C), and ASV 264 (*Terrimicrobium*) which increased from 1960 to a peak in 1990, and then declined to 2016 (Fig. 2.5D). ASV 52 (*Crenothrix*) showed a declining trend in relative abundance (Fig. 2.5E), whereas ASV 72 (*Methylocystis*) had a more stable relative abundance over time (Fig. 2.5F).

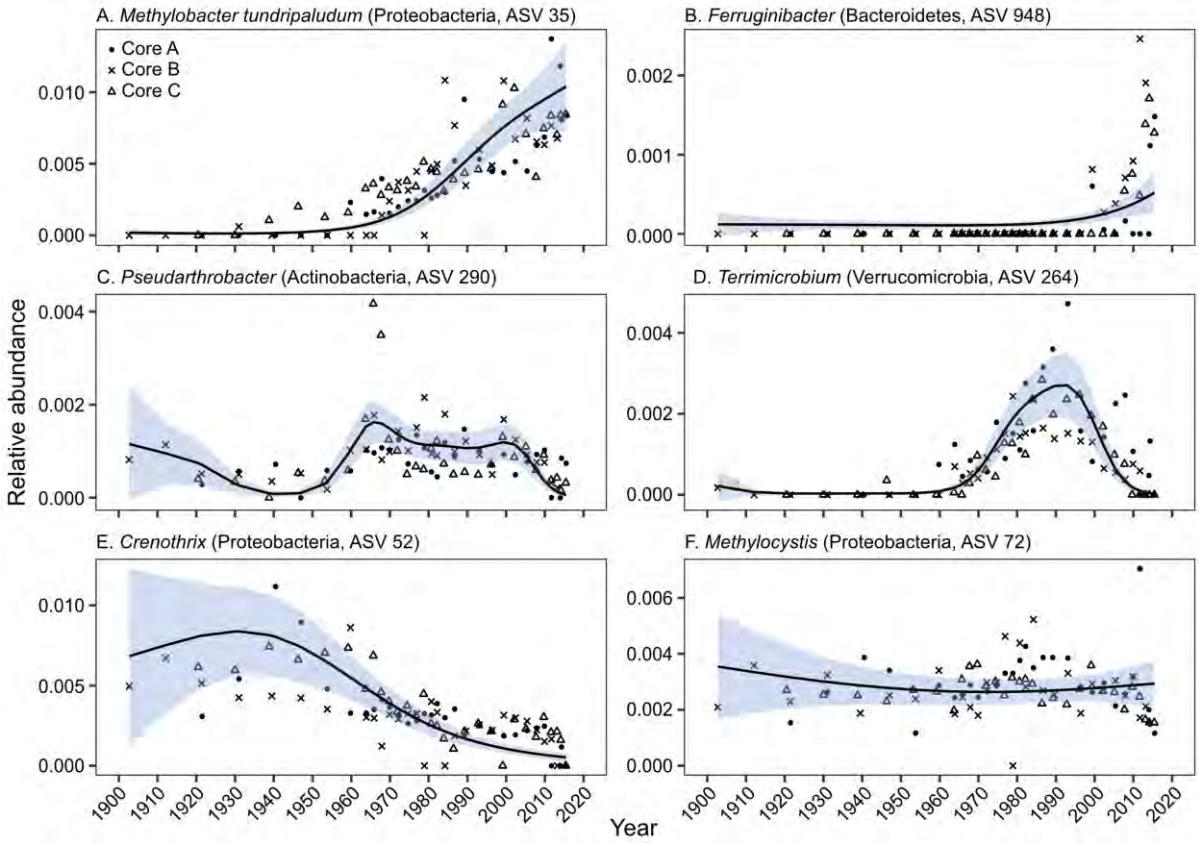

Figure 2.5. Generalised additive models (GAMs) of the relative abundance of selected pelagic ASVs throughout three sediment cores. Shaded area shows the 95% confidence interval. Note the different y-axis scales.

2.4.4. sedDNA record of the lake bacterial community

Bacterial phyla showed distinct trends in relative abundance throughout the sediment cores (Fig. 2.6). Each global (single-smoother) GAM revealed statistically significant down-core trends in relative abundance ($p < 0.001$), and associated statistics are presented in Appendix A, Table S7. However, the AIC comparison suggested that core-specific smooths out-performed the global smooths for all phyla except Cyanobacteria and Verrucomicrobia (Appendix A, Table S8). Core-specific trends in relative abundance were broadly similar, but with some variation in the size and timing of peaks for some phyla, most notably Nitrospirae which displayed a more linear trend in core A, but distinct peaks in cores B and C (Appendix A, Fig. S4-S13). For clarity, GAMs with a global smooth for all cores are presented in Fig. 2.6, and are discussed below.

Proteobacteria comprised more than 25% of the community in each core section. The relative abundance of Proteobacteria was relatively stable between 1903 and 1950, but showed an increasing trend to 2016 (Fig. 2.6A). Chloroflexi made up the second largest proportion of the community from 1903 to 1960, but there was a consistent decline in the relative abundance of Chloroflexi from the 1950s to 2016 (Fig. 2.6B). The relative abundance of Firmicutes exceeded Chloroflexi in the 1960s, but displayed a general decreasing trend to 2016 (Fig. 2.6C). The relative abundance of Bacteroidetes increased from 1970, and this group made up the second largest proportion of the community by 1990. Bacteroidetes then showed a slight decline in relative abundance from 2000 to 2016 (Fig. 2.6D).

Other phyla that were present in the sediment cores at a lower relative abundance include Verrucomicrobia, which increased in relative abundance from 1950, and the rate of increase accelerated after 2000 (Fig. 2.6). Acidobacteria showed a general decreasing trend to 2016 (Fig. 2.6F). Cyanobacteria were at very low relative abundance (< 0.001) in the sediment until the

1960s (Fig. 2.6G). The relative abundance of Cyanobacteria then gradually increased to a peak in the 1990s before decreasing throughout the 2000s to 2016 when the relative abundance returned to a similar level to that in the 1960s. Nitrospirae declined between 1970 and 1990 (Fig. 2.6H), and Spirochaetes and Planctomycetes both showed a general decreasing trend from 1903 to 2016 (Fig. 2.6I and J).

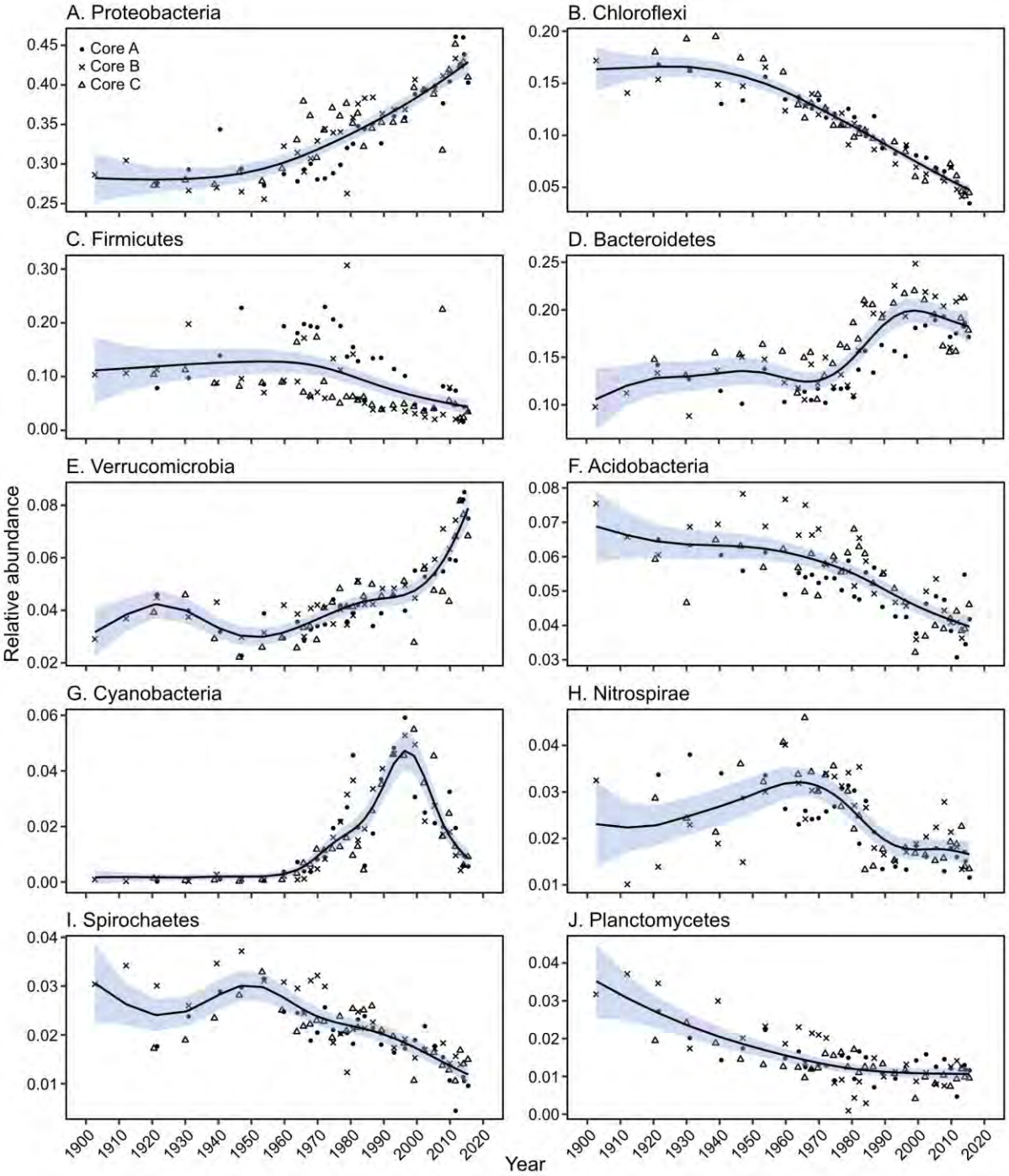

Figure 2.6. Generalised additive models (GAMs) of the relative abundance of bacterial phyla throughout three sediment cores. Shaded area shows the 95% confidence interval. Note the different y-axis scales.

2.4.5. Correlations between bacterial phyla and lake conditions

Spearman's correlation was used to investigate the relationship between the relative abundance of each bacterial phylum and sample age and physicochemical conditions of the lake. The

community could be divided into two groups based on their responses. With increasing sample age, the relative abundance of Proteobacteria, Verrucomicrobia, Bacteroidetes and Cyanobacteria significantly decreased ($p < 0.001$) (Fig. 2.7 and Appendix A, Table S9). Proteobacteria, Verrucomicrobia, Bacteroidetes and Cyanobacteria also had a significant positive correlation with alkalinity, surface water temperature, SRP and winter SRP ($p < 0.05$). The significant positive correlation between Cyanobacteria and SRP and winter SRP was relatively strong ($R_s = 0.69$, $p < 0.001$), and Cyanobacteria were the only phylum which had a significant and positive correlation with TP ($R_s = 0.74$, $p < 0.001$), chlorophyll *a* ($R_s = 0.58$, $p < 0.001$) and NO_3-N ($R_s = 0.54$, $p < 0.001$). With increasing sample age, the relative abundance of Planctomycetes, Nitrospirae, Firmicutes, Acidobacteria, Spirochaetes and Chloroflexi significantly increased ($p < 0.001$), and these phyla had a significant negative correlation with alkalinity, surface water temperature, SRP and winter SRP ($p < 0.05$). However, there was some significant co-correlation among measures of physicochemical conditions (Appendix A, Fig. S14 and Table S10).

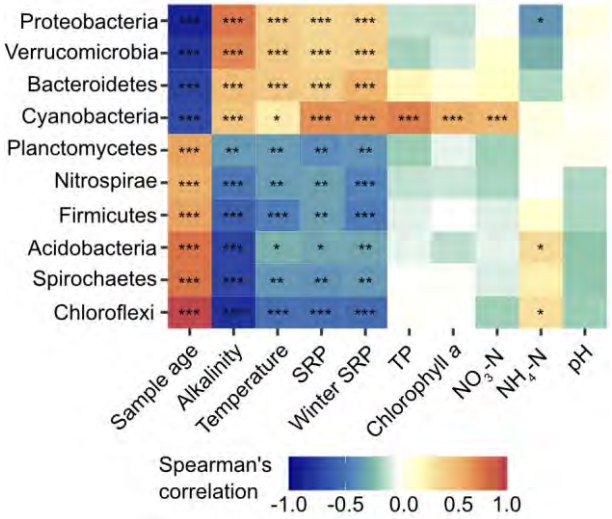

Figure 2.7. Correlations between the bacterial community and sample age and lake physicochemical conditions including alkalinity, surface water temperature, soluble reactive phosphorus (SRP), winter SRP, total phosphorus (TP), chlorophyll *a*, nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N) and surface water pH. Correlations are calculated using Spearman's correlation, where red shades indicate positive correlations, and blue shades indicate negative correlations (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

2.5. Discussion

2.5.1. Eutrophication in Esthwaite Water

The 70-year monitoring record of nutrient concentrations in Esthwaite Water suggests substantial eutrophication which accelerated from the 1970s with the establishment of wastewater treatment works and fish farming in the lake in 1973 and 1981, respectively. Accelerated eutrophication in Esthwaite Water from the 1970s has previously been reported (Bennion *et al*., 2000; Dong *et al*., 2011; Maberly *et al*., 2011a). Nutrient enrichment declined from the late 1990s, indicating that the tertiary wastewater treatment upgrade in 1986 successfully reduced some nutrient pollution, although the response may have been delayed because of continued phosphorus release from sediments and catchment run-off (Dong *et al*., 2011, 2012). The decreasing trend in nutrient enrichment continued following closure of the fish farm in 2009 and further upgrades to the wastewater treatment works in 2010. By 2015, the TP, SRP and NO_3-N concentrations were similar to pre-1970.

2.5.2. Consistency between sediment cores

Three sediment cores were collected from the deepest point of the lake to investigate spatial heterogeneity in the sedDNA signal. Each core recorded broadly similar trends in relative abundance, but slight variations were observed in the core-specific trends for some phyla. This could be a consequence of differing age-depth relationships for each core, or indicate that there was some spatial heterogeneity in the community between replicate cores. Pearman *et al*. (2021) concluded that a single core could adequately record trends in dominant bacteria. Billard *et al*. (2015) also demonstrated that the bacterial community was consistent between three lake sediment cores taken from the same sampling site, although from the much larger Lake Bourget, France. However, it has been reported that the microbial diversity of the littoral zone is often not well-represented in central sediment cores (Anderson, 2014). A single lake sediment core

may therefore be sufficient to capture broad-scale temporal variation in microbial communities, but multiple cores may be valuable when finer-scale spatial and temporal trends are of interest.

2.5.3. Effectiveness of sedDNA in describing pelagic communities

Although we did not have a concurrent DNA-based pelagic community dataset to compare the sedDNA communities against, the 65-year phytoplankton monitoring record for Esthwaite Water offered a rare opportunity to assess the ability of sedDNA to track changes in the pelagic cyanobacterial community over a long period of time. The increase over time in ASV richness of cyanobacteria in the sediment coincided with an increase in morphospecies richness of cyanobacteria in the water column, suggesting that this could largely be a response to changing lake conditions, such as nutrient enrichment, and not only loss of cyanobacterial DNA in older sediments. There was a significant positive correlation between ASV richness and morphospecies richness, supporting the use of sedDNA as a reliable measure of pelagic richness. Monchamp *et al*. (2016) also found a significant positive correlation between cyanobacterial richness in the sediment and water column when comparing sedDNA with a 35-year phytoplankton monitoring record for two temperate lakes. Similar to the present study, richness as determined with sedDNA was slightly higher than that determined with microscopy in the water column, which could be expected because observational methods typically underestimate diversity compared to molecular methods (Monchamp *et al*., 2016; Olsen *et al*., 1986).

The FreshTrain database of known pelagic bacteria based on temperate lake epilimnia was used to determine whether sedDNA recorded known pelagic bacteria. There were 23 pelagic ASVs detected in the sedDNA record, and each ASV displayed a distinct trend in relative abundance,

indicating that sedDNA could track the temporal dynamics of some pelagic bacteria. However, the 23 ASVs that were a match with the FreshTrain database were not diverse as only four different phyla were detected, and the majority were assigned to Proteobacteria. The number of known pelagic bacteria detected was a relatively small proportion of the total sedDNA community. This could suggest that only a proportion of the pelagic bacterial community deposit in the sediment, which could be expected to be mostly larger cells, or those which aggregate with organic material and other cells (Thupaki *et al*., 2013). Cells and extracellular DNA that are quickly adsorbed by particles and buried in sediment are better protected from degradation, and rates of DNA degradation are typically higher in the water column compared to in the sediment (Nwosu *et al*., 2021). Surface water inputs may therefore be underrepresented in the sedDNA record compared to deeper water or sediment inputs. Furthermore, the FreshTrain database is based on a limited number of temperate lakes, most of which are located in North America (Rohwer *et al*., 2018), and it is likely that many more ASVs with pelagic lifestyles from Esthwaite Water may have been recorded by sedDNA but were not present in the FreshTrain database.

Differential degradation of DNA, either in the water column or during the sedimentation process, could limit the potential of sedDNA in palaeolimnology, but the extent of this is relatively unknown for bacterial DNA (Boere *et al*., 2011). Reduced diversity and the disappearance of some taxa in older sediments could indicate that DNA of certain taxa experience greater rates of degradation (Zhang *et al*., 2021a). Diversity and richness for the whole community were relatively consistent throughout the sediment cores, suggesting substantial DNA degradation was not occurring. Many bacterial phyla also displayed contrasting trends in relative abundance that were not consistent with universal degradation of sedDNA through time. However, the relative abundance of cyanobacteria was low prior to

1960. Pigment records for Esthwaite Water suggest that cyanobacteria have increased from the 1800s (Moorhouse *et al*., 2018). Although cyanobacterial richness in the sedDNA record correlated significantly with that in the microscopy-based monitoring record from 1945, the low relative abundance of cyanobacteria in older sediments could be evidence that some post-depositional degradation occurred. Further research is therefore required to determine rates of DNA degradation in lacustrine systems, and to validate cyanobacterial and bacterial sedDNA records over longer periods of time.

The close agreement between cyanobacterial richness measured with sedDNA and that with microscopy, and the successful detection of some pelagic bacteria in the sedDNA record demonstrates that sedDNA has significant potential as a record of pelagic temporal dynamics. However, it must be considered that sedDNA may also record the temporal dynamics of the surface sediment community, and although heterotrophic bacterial activity is typically lower within lake sediments (Haglund *et al*., 2003), deep sediment communities could also contribute to the sedDNA record. These sediment bacterial communities may be less responsive to lake water conditions, and their distribution with depth may partly be a response to sediment conditions. sedDNA gives an indication of temporal bacterial community change, but further research is required to track the incorporation of DNA from different sources into the lake sediment, and to separate ancient DNA from that of metabolically active cells within the sediment.

2.5.4. Community composition change

The sedDNA record revealed distinct shifts in cyanobacterial and bacterial community composition over 113 years, and several phyla underwent marked changes in their relative

abundance during the period of intense eutrophication in Esthwaite Water between 1970 and the early 2000s. The relative abundance of cyanobacteria closely reflected changes in nutrient enrichment in the lake. Frequent cyanobacterial blooms have been observed during the period of intense eutrophication in Esthwaite Water (Dong *et al*., 2012), and photosynthetic pigments extracted from sediment cores in nearby Windermere and Blelham Tarn have also indicated an increase in cyanobacteria during periods of eutrophication from the 1970s (McGowan *et al*., 2012; Moorhouse *et al*., 2014). Eutrophication is a problem for many lakes worldwide, and previous sedDNA records from a large number of lakes have revealed an increase in cyanobacteria in response to past nutrient enrichment (Domaizon *et al*., 2013; Ibrahim *et al*., 2020; Monchamp *et al*., 2019; Zhang *et al*., 2021a).

Proteobacteria made up the largest proportion of the community, and increased in relative abundance from older to more recent sediments. Proteobacteria are typically copiotrophic and often dominate lake bacterial communities (Newton *et al*., 2011). Using sedDNA, Li *et al*. (2019a) also found a substantial increase in the relative abundance of Proteobacteria with increasing nutrient concentrations. However, Proteobacteria continued to increase after nutrient concentrations declined from the 2000s. Proteobacteria had a significant positive correlation with surface water temperature, and the relative abundance of Proteobacteria in bacterioplankton communities has previously been shown to increase with water temperature (He *et al*., 2017). Increasing water temperatures from 1965 onwards may therefore have contributed to the sustained increase in the relative abundance of Proteobacteria, although separation of potential drivers is complex as there were shared long-term trends between the physicochemical conditions.

The increase in the relative abundance of Bacteroidetes from 1970 to 2000 coincided with the period of accelerated eutrophication. Bacteroidetes have frequently been found to increase with

nutrient concentrations, and are a major component of the bacterioplankton community associated with cyanobacterial blooms (De Figueiredo *et al*., 2007; Guedes *et al*., 2018). Many members of Bacteroidetes are chemoorganotrophs capable of metabolising the complex extracellular polysaccharides produced by cyanobacteria, and so Bacteroidetes play an important role in carbon and nitrogen cycling in lakes (Cai *et al*., 2014). Similarly, Verrucomicrobia have previously been found in association with cyanobacterial blooms in eutrophic lakes and can metabolise complex polysaccharides (Cardman *et al*., 2014; Kiersztyn *et al*., 2019). Accelerated nutrient enrichment and a higher abundance of cyanobacteria may therefore have facilitated the increase in the relative abundance of Verrucomicrobia.

Other bacterial phyla declined over the period of accelerated eutrophication, such as Nitrospirae which dropped in relative abundance between 1970 and 1990 when nutrient concentrations were highest. Previous studies have also found that the relative abundance of Nitrospirae declined with increasing phosphorus (Liu *et al*., 2020; Xiong *et al*., 2012) and NH_4-N concentrations (Sun *et al*., 2020). Chloroflexi were relatively abundant from the 1900s but consistently declined over time. Increased turbidity of water during eutrophication and cyanobacterial blooms may reduce light availability in the water column for photosynthetic members of Chloroflexi (Chen *et al*., 2015). The relative abundance of Acidobacteria also declined over time. Acidobacteria have previously been described as oligotrophic taxa (Huang *et al*., 2017), and many species have a preference for acidic environments (Xiong *et al*., 2012). At the deepest point, Esthwaite Water had a stable mean surface water pH of 7.3 ± 0.21. The neutral pH combined with nutrient enrichment may not have been optimal for the growth of Acidobacteria. Firmicutes form resistant endospores (Wunderlin *et al*., 2014), so the DNA of Firmicutes could be expected to be preserved more efficiently in the sediment compared to other bacterial phyla. However, the relative abundance of Firmicutes was not substantially

higher in older sediments compared to other phyla, and this provides further evidence that DNA did not undergo significant differential degradation.

Planctomycetes and Spirochaetes were present at a relatively low abundance, each making up less than 0.4% of the bacterial community, and declined over time. Planctomycetes are usually present in lakes at a low relative abundance, and have slow growth rates and a delayed response to increased nutrient supply (Pollet *et al*., 2014). Previous studies have also found that Spirochaetes decreased with nutrient enrichment (Liu *et al*., 2014; Wan *et al*., 2017). However, many species of Spirochaetes are parasitic, and their abundance may also be related to the population dynamics of other organisms (Sitnikova *et al*., 2002).

Our results clearly demonstrate the potential for using sedDNA to make historical inferences of microbial community change. However, we acknowledge that, due to sediment accumulation rates, this technique is likely limited in its ability to discern intra-annual change, and short-term species turnover in response to environmental forcing. Furthermore, it is challenging to demonstrate causal linkages between environmental drivers and community change, given the inherent multi-dimensionality of lake ecosystem change. Physicochemical drivers, which may act on bacterial community composition, are, at these timescales, highly collinear. Here, we have adopted a correlative approach as a means of tentatively suggesting potential drivers of past change. While it is a commonly used metric, the use of relative abundance in palaeolimnology can obscure temporal trends due to different groups showing reciprocal responses, potentially due to differential DNA degradation. Future studies should therefore aim to quantify absolute abundances of taxa for more accurate reconstructions of community responses.

2.5.5. Recovery from eutrophication

Following the period of accelerated eutrophication from the 1970s, the relative abundance of cyanobacteria declined and returned to pre-1970 levels by 2016. The sedimentary record of cyanobacterial pigments in Esthwaite Water may also indicate some early signs of recovery from the 2000s (Moorhouse *et al*., 2018). Bacteroidetes, which increased over the period of accelerated eutrophication, began to decline from the 2000s, and Nitrospirae, which decreased over this period, showed a slight increase in relative abundance from the 2000s. This could indicate that the bacterial community has begun to gradually recover from intense eutrophication. However, although cyanobacteria declined relatively quickly, the relative abundance of other phyla such as Bacteroidetes remained higher in 2016 compared to pre-1970. Eutrophication and frequent blooms of cyanobacteria may have altered long-term functioning of the ecosystem, and cyanobacterial-derived polysaccharides could remain in the lake which other phyla such as Bacteroidetes and Verrucomicrobia may continue to metabolise. Climate warming could also delay complete recovery, as cyanobacteria and each bacterial phylum had a significant correlation with surface water temperature which has shown an increasing trend since 1965. The abundance of cyanobacteria has previously been found to increase with warming temperatures (Domaizon *et al*., 2013; Monchamp *et al*., 2018), and climate warming combined with eutrophication has been predicted to increase the frequency of cyanobacterial blooms in Esthwaite Water in the future (Elliot, 2010).

2.5.6. Conclusions

Our results have shown that lake sedDNA can be used to infer past changes in cyanobacterial and bacterial community composition. The sedDNA record was validated by long-term monitoring of cyanobacterial richness in the water column, and successfully tracked some

bacterial pelagic history. sedDNA recorded a substantial increase in cyanobacterial relative abundance during the period of intense eutrophication from the 1970s, and this was accompanied by shifts in bacterial community composition in response to changing physicochemical conditions of the lake. Lake restoration efforts appear to have been partially effective, as the relative abundance of cyanobacteria declined with reductions in nutrient enrichment. However, intense eutrophication and the increase in cyanobacteria may have substantially changed long-term ecosystem functioning, and climate warming could delay complete recovery of the lake. These changes in bacterial community composition could have implications for the role of lakes in carbon and nitrogen cycling. We have demonstrated the significant potential of bacterial sedDNA in palaeolimnology, but further studies are required to identify the proportion of pelagic, benthic and sediment inputs, the taphonomic processes each source may be subject to, and the length of time over which sedDNA can act as a reliable record of lake microbial community dynamics.

Chapter 3: Evaluating the use of lake sedimentary DNA in palaeolimnology: A comparison with long-term microscopy-based monitoring of the phytoplankton community

This chapter is under review in Molecular Ecology Resources as: Thorpe, A. C., Mackay, E. B., Goodall, T., Bendle, J. A., Thackeray, S. J., Maberly, S. C. and Read, D. S. (2023) Evaluating the use of lake sedimentary DNA in palaeolimnology: A comparison with long-term microscopy-based monitoring of the phytoplankton community. A pre-print is available at: https://doi.org/10.22541/au.167819405.56988284/v1.

3.1. Abstract

Palaeolimnological records provide valuable information about how phytoplankton respond to long-term drivers of environmental change. Traditional palaeolimnological tools such as microfossils and pigments are restricted to taxa that leave sub-fossil remains, and a method that can be applied to the wider community is required. Sedimentary DNA (sedDNA), extracted from lake sediment cores, shows promise in palaeolimnology, but validation against data from long-term monitoring of lake water is necessary to enable its development as a reliable record of past phytoplankton communities. To address this need, 18S rRNA gene amplicon sequencing was carried out on lake sediments from a core collected from Esthwaite Water (English Lake District) spanning ~105 years. This sedDNA record was compared with concurrent long-term microscopy-based monitoring of phytoplankton in the surface water. Broadly comparable trends were observed between the datasets, with respect to the diversity and relative abundance and occurrence of chlorophytes, dinoflagellates, ochrophytes and bacillariophytes. Up to 20% of genera were successfully captured using both methods, and sedDNA revealed a previously undetected community of phytoplankton. These results suggest that sedDNA can be used as an effective record of past phytoplankton communities, at least over timescales of less than 100 years. However, a substantial proportion of genera identified by microscopy were not detected using sedDNA, highlighting the current limitations of the technique that require further development such as reference database coverage. The taphonomic processes which may affect its reliability, such as the extent and rate of deposition and DNA degradation, also require further research.

3.2. Introduction

Phytoplankton play a vital role in lake ecosystems as primary producers at the base of aquatic food webs. Changes in their community composition in response to environmental change can have extensive ecological and biogeochemical implications (Litchman *et al*., 2015). Many lakes worldwide are experiencing rapid rates of change in response to multiple interacting stressors, but our understanding of how phytoplankton communities respond is limited (Carpenter *et al*., 2011; Heino *et al*., 2009). Multi-decadal records of phytoplankton communities can enable us to understand how they have responded to past environmental change and provide insight into how they may respond in the future (Willis *et al*., 2010).

Detailed long-term monitoring of the phytoplankton community is restricted to a relatively small number of well-studied lakes (Burlakova *et al*., 2018; Hampton *et al*., 2008). Where long-term monitoring records are not available, a range of proxies can be used to produce historical records of the phytoplankton community, such as microfossils and pigments extracted from lake sediment cores (Davidson and Jeppesen, 2013). Microfossil analysis is a widely used technique but is limited to organisms with well-preserved and morphologically distinct remains, such as diatom frustules (Hembrow *et al*., 2014; Leira, 2005) and the resting cysts produced by some dinoflagellates (Drljepan *et al*., 2014). Photosynthetic pigments can provide a record of eukaryotic algal and cyanobacterial community composition, abundance, and primary productivity (Griffiths *et al*., 2022; Kpodonu *et al*., 2016; Makri *et al*., 2019; Watanabe *et al*., 2012), but many pigments are not specific enough to enable taxonomic identification beyond the class level (Gong *et al*., 2020). These limitations of traditional palaeolimnological techniques highlight the need for complementary and improved methods which can be applied to a wider diversity of organisms, such as sedimentary DNA (sedDNA).

sedDNA is a promising palaeolimnological approach which can be used to reconstruct past communities using DNA preserved within lake sediment cores (Edwards, 2020). DNA from living organisms is deposited in the lake sediment, where it is preserved and progressively buried over time. This DNA can then be extracted from layers of a sediment core and sequenced to produce a temporal record of lake communities (Capo *et al*., 2021; Thorpe *et al*., 2022). sedDNA offers many potential benefits compared to traditional palaeolimnological techniques. For example, a relatively high taxonomic resolution can be achieved, and high-throughput amplicon sequencing can process many hundreds or even thousands of samples relatively quickly (Bohmann *et al*., 2022; Gong *et al*., 2020; Mejbel *et al*., 2021). A wider diversity of organisms can be studied using sedDNA, including those previously overlooked with microfossil analysis due to a lack of well-preserved and morphologically distinct remains (Domaizon *et al*., 2017). The applicability of sedDNA to the wider community, including eukaryotic algae (Capo *et al*., 2016), bacteria (Thorpe *et al*., 2022), zooplankton (Tsugeki *et al*., 2022) and macrophytes (Stoof-Leichsenring *et al*., 2022) allows a more complete reconstruction of ecosystem structure which may, in turn, facilitate inferences on past trophic interactions (Barouillet *et al*., 2022; Ellegaard *et al*., 2020).

sedDNA is becoming more widely used in palaeolimnology, but there are currently some uncertainties surrounding the deposition and taphonomy of DNA in lakes (Capo *et al*., 2021; 2022). The extent and rate of DNA degradation may vary among taxa and depend upon the state in which DNA is deposited. For example, intracellular DNA or DNA bound to mineral particles is typically better protected from degradation processes, such as oxidation, hydrolysis, and bacterial degradation than free extracellular DNA (Giguet-Covex *et al*., 2019; Mauvisseau *et al*., 2022). The depositional and degradational processes DNA is subject to could affect the ability of sedDNA to provide a reliable record of past phytoplankton communities. Although

sedDNA has previously been found to be broadly comparable with records from diatom frustules (Anslan *et al*., 2022) and photosynthetic pigments (Picard *et al*., 2022a; Tse *et al*., 2018), these traditional palaeolimnological tools are also subject to pre- and post-depositional losses and subsequent biases. Validation of sedDNA against long-term monitoring of phytoplankton in the water column is therefore needed to further the development of sedDNA as a reliable and robust record of past microbial communities.

To address this need, we analyse and compare sedDNA and water column phytoplankton data from Esthwaite Water, a relatively small lake in the English Lake District which has experienced well-documented changes in human activity and has undergone substantial eutrophication in recent decades (Dong *et al*., 2011; Maberly *et al*., 2011a). Lake physicochemical conditions and the phytoplankton community have been continually monitored since 1945, providing a detailed record against which palaeolimnological records can be compared and validated. Esthwaite Water has been the site of several studies investigating seasonal trends in phytoplankton communities in the water column (Feuchtmayr *et al*., 2012; Talling and Heaney, 2015), and palaeolimnological studies of the bacterial and cyanobacterial community as measured by sedDNA (Thorpe *et al*., 2022), and the microbial eukaryotic community as measured with photosynthetic pigments (Moorhouse *et al*., 2018) and diatom frustules (Bennion *et al*., 2000; Dong *et al*., 2011, 2012). Our study, which combines concurrent microscopy-based monitoring and sedDNA records, is therefore uniquely placed to determine whether sedDNA is a reliable tool for reconstruction of past trends in phytoplankton community composition.

3.3. Materials and methods

3.3.1. Study site

Esthwaite Water (54° 21.56′ N, 2° 59.15′ W) is located within the Lake District National Park, Cumbria, UK (Appendix A, Fig S1), and has a catchment area of 17 km^2, surface area of 0.96 km^2, and mean and maximum depths of 6.9 m and 16 m, respectively (Maberly *et al*., 2011a; Mackay *et al*., 2012). Human activities in Esthwaite Water and its catchment, including construction of wastewater treatment works in 1973 and fish farming between 1983 and 2009, led to this lake becoming one of the most eutrophic lakes in the Lake District (Dong *et al*., 2011; Maberly *et al*., 2011a).

3.3.2. Long-term environmental monitoring record

Physiochemical conditions in Esthwaite Water have been continuously monitored on a weekly to fortnightly basis from 1945 by the Freshwater Biological Association (FBA), and then from 1989 by the UK Centre for Ecology & Hydrology (UKCEH). Measurements and depth integrated surface water samples (0-5 m) were collected from the deepest point of Esthwaite Water, including surface water temperature, pH and alkalinity, and the concentration of total phosphorus (TP), soluble reactive phosphorus (SRP), nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N) and chlorophyll *a* (Chl *a*). Winter SRP was calculated as the mean SRP from December to February. The full dataset is available at: https://doi.org/10.5285/87360d1a-85d9-4a4e-b9ac-e315977a52d3 (Maberly *et al*., 2017), and annual means for these variables have previously been described (Thorpe *et al*., 2022).

3.3.3. Long-term phytoplankton microscopy record

Sub-samples of the surface water samples collected for physiochemical analysis between 1945 and 2010 were used to monitor the phytoplankton community at weekly to fortnightly intervals. These sub-samples were preserved with Lugol's iodine, concentrated by sedimentation, and then placed in a counting chamber under a microscope for identification and enumeration. Cells were counted within a sedimentation chamber until 1994, after which a Lund chamber was used for enumeration. Phytoplankton were identified to species level where possible, and quantified as the number of cells, colonies, or filaments per mL of lake water.

3.3.4. Sediment coring

A sediment core was collected from the deepest point of Esthwaite Water using a HTH 9 cm diameter gravity corer (Pylonex, Sweden) in August 2021. Coring equipment was thoroughly sterilised with ethanol and rinsed with deionised water three times before use. After collection, the 35 cm long sediment core remained intact within the sealed Perspex core tube and was kept upright on ice in a large cool box in the field and during transportation to UKCEH, Wallingford, where it was stored at 4 °C in the dark prior to sectioning.

The sediment core was sectioned in 1 cm intervals using the extruding device (Pylonex, Sweden), beginning with recent sediment at the top and working downwards. Each 1 cm sediment section was pushed out the top of the core tube directly into a sterile petri dish of the same diameter to minimise contact with the air. A broad stainless-steel blade was used to cut between the core tube and the petri dish containing the extruded sediment section, which was then sealed with a lid and secured with parafilm. The blade was sterilised with bleach and ethanol and rinsed with deionised water between each section. Clean lab coats, gloves and

masks were worn when handling the sediment core to minimise contamination risk. Each sediment core section was then sub-sampled in a UV-sterilised laminar flow cabinet. Using a sterilised spatula, a small amount of undisturbed sediment from the centre of each section which did not come into contact with the blade or core tube was transferred to a sterile Eppendorf tube for storage at -20 °C prior to DNA extraction.

3.3.5. Sediment core chronology

The chronology of the sediment core was estimated using the age-depth relationship of a separate reference core collected from the same location within Esthwaite Water in 2014 as described by Thorpe *et al*. (2022). Sample depths for the reference core were corrected to 2021 assuming a constant sedimentation rate (Appendix B, Table S1), and the slope and intercept of the age-depth relationship were then used to estimate the age of each sediment core section (Appendix B, Table S2). The full length of the 35 cm sediment core was estimated to cover 105 years from 1916 to the date of collection in 2021.

To evaluate whether the reference core chronology was accurately aligned with the 2021 core, 16S rRNA gene amplicon sequencing was carried out on sediments from the core and compared with the bacterial sedDNA record obtained from cores collected from the same location within Esthwaite Water in 2016 (Thorpe *et al*., 2022). Temporal trends in the relative abundance of the dominant bacterial groups were closely aligned between the two bacterial sedDNA records, supporting the use of the estimated chronology (refer to Appendix B, Fig. S1 for a detailed comparison).

3.3.6. DNA extraction, PCR amplification and 18S rRNA gene sequencing

DNA was extracted from 0.25 g of each sediment core sample using the Qiagen DNeasy PowerSoil Pro extraction kit (Qiagen, Germany) according to the manufacturer's protocol. DNA extractions were performed in small batches in a random order to minimise bias. A Qiagen PowerBead tube containing no sample material was included in every other batch of extractions as a negative control. The concentration and purity of each DNA sample were checked on the NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, MA, U.S.). Extracted DNA samples were stored at -20 °C.

The V4-V5 region of the 18S rRNA gene was amplified with universal forward and reverse eukaryotic primers, NSF563 (5'-CGCGGTAATTCCAGCTCCA-3') and NSR951 (5'-TTGGYRAATGCTTTCGC-3') (Mangot *et al*., 2012). Each 50 µL PCR mix contained 0.5 µL of 2000 units mL^{-1} Q5 High-Fidelity DNA polymerase, 10 µL of 5x reaction buffer, 10 µL of 5x high GC enhancer (New England Biolabs, UK), 1 µL of a 10 mM dNTP mix (Bioline, UK), 26.3 µL of molecular grade water, 0.1 µL of each 100 µM forward and reverse primer, and 2 µL of DNA. Negative controls containing the PCR mix but no DNA were included. The PCR program was set to an initial denaturing temperature of 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 sec, an annealing temperature of 60 °C for 30 sec, an extension temperature of 72 °C for 30 sec, and then a final extension temperature of 72 °C for 10 min. Successful PCR amplification was confirmed with an agarose gel. PCR product was purified with the Millipore multiscreen PCR filter plate kit according to the manufacturer's protocol (Merck Millipore, MA, U.S.), resulting in purified PCR product eluted in 35 µL of molecular grade water.

Second step PCR was performed using a dual-indexing approach (Kozich *et al*., 2013), and 25 µL reactions contained 0.25 µL of Q5 DNA polymerase, 5 µL of reaction buffer, 5 µL of high GC enhancer, 0.5 µL of dNTPs, 7.25 µL of molecular grade water, 5 µL of the indexing primers

(Kozich *et al*., 2013), and 2 µL of purified PCR product from the first PCR step. The second step PCR program was set to an initial denaturing temperature of 95 °C for 2 min, followed by 8 cycles of 95 °C for 15 sec, an annealing temperature of 50 °C for 30 sec, an extension temperature of 72 °C for 30 sec, and then a final extension temperature of 72 °C for 10 min. Successful PCR amplification from the second PCR step was confirmed with an agarose gel.

PCR product from the second PCR step was normalised using the Invitrogen SequalPrep normalisation kit according to the manufacturer's protocol (Thermo Fisher Scientific, MA, U.S.), resulting in 1-2 ng μL^{-1} of DNA per sample. Samples were pooled, gel-extracted using the Qiagen MinElute gel extraction kit according to the manufacturer's protocol, and the purified DNA concentration was quantified using the Invitrogen Qubit dsDNA high-sensitivity assay kit with the Qubit 3.0 fluorometer. The amplicon library was denatured with NaOH, neutralised with HCl, combined with 10% denatured PhiX, and then diluted with HT1 buffer (Illumina, CA, U.S.). The library was heat denatured at 96 °C for 2 min and immediately transferred to an ice bath prior to sequencing on the Illumina MiSeq Platform with a 500-cycle v2 MiSeq reagent kit.

3.3.7. Sequence data processing

The DADA2 pipeline was implemented to process the sequences (Callahan *et al*., 2016). Primers were trimmed, and reads were truncated where the quality score fell below Q30. The quality-filtered forward and reverse reads were merged, and an amplicon sequence variant (ASV) abundance table was generated. Taxonomy was assigned to each exact ASV using the naive Bayesian classifier (Wang *et al*., 2007) against the PR^2 database v4.14.0 (Guillou *et al*.,

2013). The sequences were rarefied to a uniform sequencing depth of 14,936 reads and two samples that did not meet the rarefaction depth were excluded.

ASV abundance was converted to relative abundance, and ASVs were filtered according to taxonomy to remove those unidentified at the phylum level. Heterotrophic groups that were outside of the scope of the microscopy-based monitoring record were excluded from analysis. Chlorophytes, dinoflagellates, ochrophytes and bacillariophytes were well-represented in both the microscopy and 18S rRNA sedDNA records and were therefore included for in-depth analysis.

3.3.8. Data analysis

Many reference databases use their own taxonomic nomenclature which can lead to conflicting taxonomy assignments when comparing multiple datasets (Canino *et al*., 2021). To allow for comparisons between the microscopy and sedDNA records, taxonomy was homogenised using Phytool v2 (Canino *et al*., 2021) which is based on the taxonomic classifications used in AlgaeBase (Guiry and Guiry, 2022). This ensured that taxa in both records were classified according to the same taxonomic nomenclature and names were updated to the current taxonomically accepted name.

To account for potential inaccuracies in species identification, taxa in both records were grouped at the genus level. As the counting method sometimes varied by size or form (e.g., single cell, colony, or filament), the microscopy-based counts were converted to a binary presence-absence value for each genus on each sampling occasion. The total number of sampling occasions on which each genus was observed was calculated for each year as a

measure of occurrence, and then normalised to the number of sampling occasions per year to account for variable sampling effort.

Non-metric multidimensional scaling (NMDS) was performed based on a beta diversity Bray-Curtis dissimilarity matrix of genus relative abundance as measured using sedDNA and genus occurrence as measured by microscopy from 1945 to 2010. Correlations between each dissimilarity matrix and lake physicochemical conditions were assessed with a permutation test and fitted to the ordination space using the vegan R package v2.6-2 (Oksanen *et al*., 2019). The vegan package was also used to calculate Shannon's alpha diversity at the genus level in both records. Generalised additive models (GAMs) with Gamma error distributions and a log link were fitted to the temporal trend in alpha diversity using the mgcv R package v1.8-40 (Wood, 2020). As there was not a sediment sample corresponding to each year of the microscopy-based monitoring record, annual values of alpha diversity from 1945 to 2010 as measured by sedDNA were estimated using the GAM fitted to the temporal trend. These GAM-estimated annual values were then correlated with GAM-estimated annual values of alpha diversity as measured by microscopy using a model II regression with the lmodel2 R package v1.7-2 (Legendre *et al*., 2018).

GAMs were fitted to the temporal trends in phylum relative abundance as measured by sedDNA using Beta error distributions with a logit link, which is suitable for proportion data. For the trends in phylum occurrence as measured by microscopy, GAMs were fitted using Gamma error distributions with a log link, which is suitable for positively skewed, non-negative data (Anderson *et al*., 2010; Simpson, 2018). Restricted maximum likelihood (REML) was used as the smoothness selection method for all GAMs (Simpson, 2018). Annual values of relative abundance from 1945 to 2010 were estimated using the GAM fitted to the temporal trend and correlated with the GAM-estimated annual values of occurrence using a model II regression.

For each phylum, Venn diagrams were used to illustrate which genera were uniquely detected using sedDNA, which were uniquely detected by microscopy, and which were detected in both records. Venn diagrams were produced with the eulerr R package v7.0.0 (Larsson, 2022), and all data analysis was performed in R v4.2.1 (R Core Team, 2022).

3.4. Results

3.4.1. Beta diversity

The NMDS of the Bray-Curtis dissimilarity matrix based on beta diversity as measured by sedDNA and microscopy both displayed a similar trajectory of community change from older to more recent samples (Fig. 3.1A and B). Pre-1981 sediment core samples were positioned on the left side of the ordination space, and post-1982 samples were on the right. More recent sediment core samples from 1997 to 2010 were closely clustered in the bottom right quadrant (Fig. 3.1A). Water samples collected for microscopic analysis prior to 1978 were positioned on the left half of the ordination space, and those collected after 1979 were positioned on the right with more recent samples from 1994 to 2010 in the bottom quadrant (Fig. 3.1B). The lake physiochemical conditions that correlated significantly with the sedDNA dissimilarity matrix included alkalinity, SRP, and pH ($p < 0.05$) with R^2 values of 0.53, 0.50 and 0.44, respectively. The microscopy dissimilarity matrix also correlated significantly with alkalinity ($p < 0.001$) and SRP ($p < 0.05$), in addition to NH_4-N ($p < 0.01$), NO_3-N, and TP ($p < 0.05$), with R^2 values of 0.61, 0.23, 0.31, 0.19 and 0.25, respectively. Mean annual trends in lake physiochemical conditions are presented in Appendix B, Fig. S2, and statistics for the correlations between these conditions and the dissimilarity matrices are provided in Appendix B, Table S3.

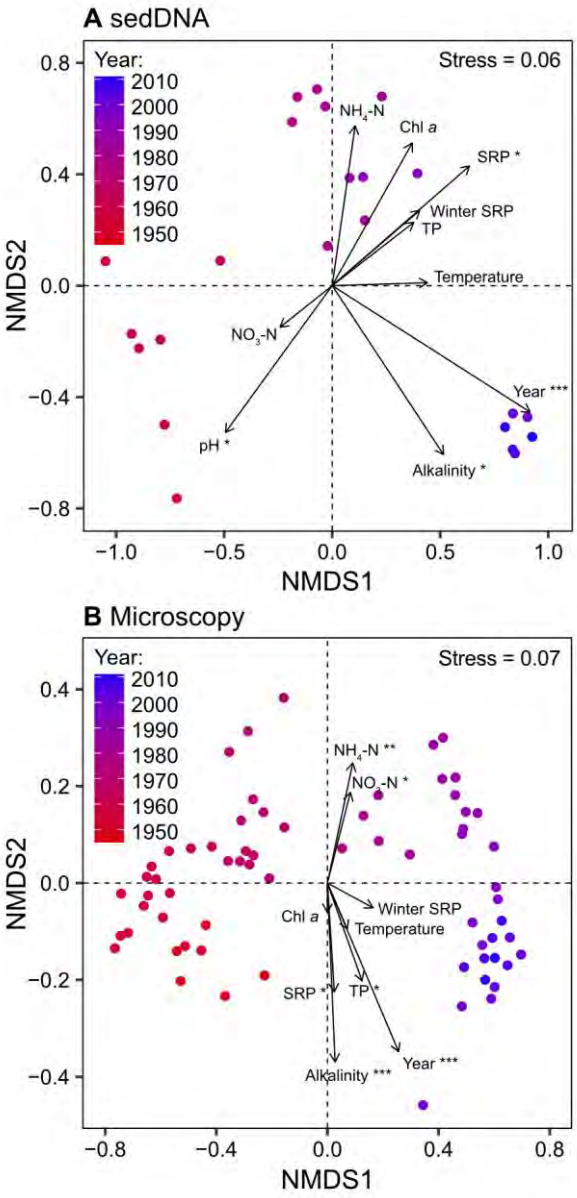

Figure 3.1. NMDS of a Bray-Curtis dissimilarity matrix based on beta diversity as measured by sedDNA (A) and microscopy (B) from 1945 to 2010. The red to blue gradient indicates older to more recent samples. Vectors for sample year and the lake physiochemical conditions that correlated with each dissimilarity matrix are fitted. Vector length is proportional to the strength of the correlation. *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$. NMDS stress values are shown.

3.4.2. Alpha diversity

Shannon's diversity index at the genus level was used as a measure of alpha diversity throughout the sedDNA and microscopy records. Both records displayed a general increasing trend in alpha diversity from the 1970s which began to plateau from the 1990s (Fig. 3.2).

Diversity in the most recent sediment core section was 2.75, which was similar to that in the oldest section with a diversity of 2.50. Alpha diversity as measured by sedDNA ranged between 1.47 and 2.76, and was consistently lower than that measured by microscopy, which ranged between 2.43 and 3.92. There was a significant positive correlation between the sedDNA and microscopy GAM-estimated annual values of alpha diversity between 1945 and 2010 with an r value of 0.75 ($F_{1,64} = 81.45$, standard error = 0.19, $p < 0.001$).

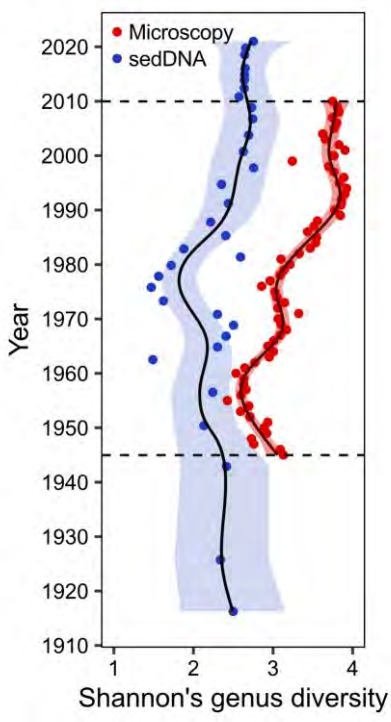

Figure 3.2. GAMs fitted to the trend in Shannon's genus diversity as measured by sedDNA (blue) and microscopy (red). Shaded areas show the 95% confidence intervals.

3.4.3. Temporal trends in community composition

Chlorophytes, dinoflagellates, ochrophytes and bacillariophytes were well-represented in both the sedDNA and microscopy records (Fig. 3.3A-D). Generally, data derived from sedDNA and microscopy showed broadly similar long-term trends for these phyla, but with some differences in the exact timing of the onset of change. Of these dominant phyla, chlorophytes and dinoflagellates made up the largest proportion of the sedDNA community. Chlorophytes were

initially present in the sedDNA record with a relative abundance between 0.03 and 0.10 from 1916 to 1994. Their relative abundance then increased abruptly to between 0.20 and 0.27 in more recent samples from 1997 to 2021. In the microscopy-based monitoring record, chlorophytes had a low occurrence initially, but increased sharply from the 1980s to become the group with the highest occurrence (Fig. 3.3A).

In the sedDNA record, dinoflagellates had a relative abundance of less than 0.01 until 1970. Their relative abundance then increased to two distinct peaks in 1980 when they reached a relative abundance of 0.17, and in 2000 when they reached a relative abundance of 0.25. Dinoflagellates in the microscopy record had three main peaks in 1967, 1986 and 2002 when they reached an occurrence of 2.35, 2.67 and 2.92, respectively (Fig. 3.3B).

The relative abundance of ochrophytes in the sedDNA record was below 0.006 and relatively stable until the 1980s when there was a slight increasing trend to the 2000s. The occurrence of ochrophytes in the microscopy record remained below 1.00 until 1983, but then increased throughout the 1980s and 1990s to their highest occurrence of 4.08 in 2001 (Fig. 3.3C).

Bacillariophytes had the lowest relative abundance of the four phyla analysed in the sedDNA record which was consistently below 0.003. There was a general increasing trend in the relative abundance of bacillariophytes from the 1970s, although there was some scatter around this trend. In the microscopy record, bacillariophytes displayed a slight decreasing trend to the 1980s, and then increased to a period of higher occurrence from the 1990s. Bacillariophytes had the highest occurrence of any phylum in the microscopy record until 1980, after which the only phylum with a higher occurrence were chlorophytes (Fig. 3.3D).

There was a significant positive correlation between the sedDNA and microscopy GAM trends for all four phyla ($p < 0.001$). The correlation between the two records was strongest for

ochrophytes, with an r value of 0.93, followed by 0.76 for chlorophytes and bacillariophytes, and 0.75 for dinoflagellates (Fig. 3.3A-D). All GAM trends for each dataset were significant ($p < 0.01$), and statistics associated with the GAMs are provided in Appendix B, Table S4 and S5.

Charophytes, cryptophytes and haptophytes were also recorded by microscopy. However, charophytes and haptophytes were only detected with a relative abundance greater than 0.001 in three sediment core samples, and cryptophytes were absent from the sedDNA record. The GAM-fitted trends in occurrence as measured by microscopy for charophytes, cryptophytes and haptophytes are presented in Appendix B, Fig. S3.

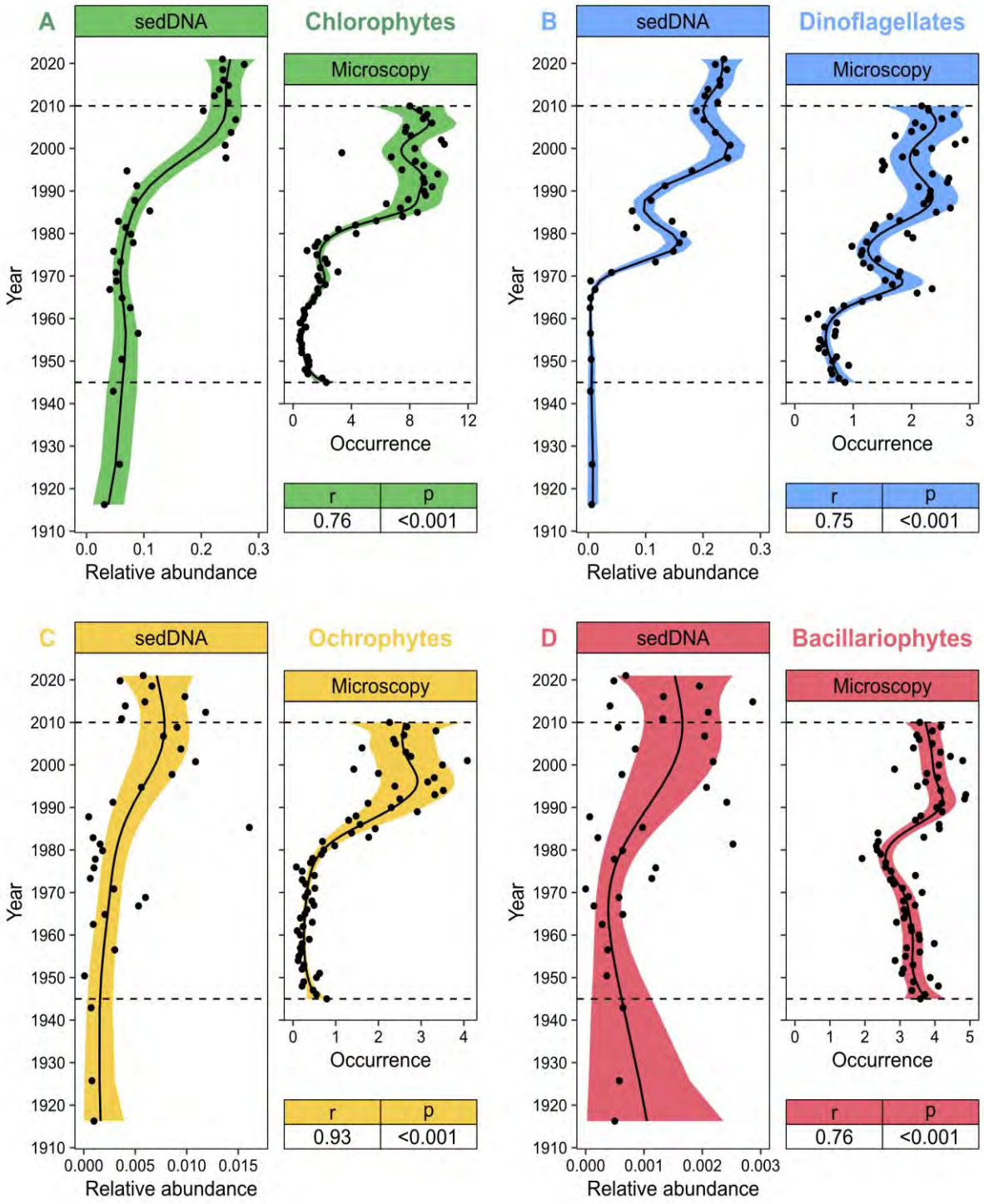

Figure 3.3. GAMs fitted to the trend in relative abundance as measured by sedDNA and occurrence relative to sampling frequency as measured by microscopy for chlorophytes (A), dinoflagellates (B), ochrophytes (C), and bacillariophytes (D). For each phylum, r values and significance levels are shown for the correlation between sedDNA and microscopy GAM-estimated annual values from 1945 to 2010. Shaded areas show the 95% confidence intervals.

3.4.4. Shared and unique genera

Across the four main phyla studied, a total of 215 genera were identified with both sedDNA and microscopy. Of these genera, 113 (52.6%) were uniquely detected by microscopy, 66 (30.7%) were uniquely detected by sedDNA, and 36 (16.7%) were detected in both the sedDNA and microscopy records (Fig. 3.4 A-D). More chlorophyte, ochrophyte and bacillariophyte genera were detected by microscopy compared to sedDNA, whereas more dinoflagellate genera were detected by sedDNA.

The majority of genera detected by each method were chlorophytes. There was a total of 128 chlorophyte genera detected by both methods combined, and almost half of these (65 genera) were only detected by microscopy. sedDNA uniquely detected 38 chlorophyte genera, and 25 genera were detected by both methods (Fig. 3.4A).

Only 4 dinoflagellate genera were uniquely detected by microscopy (*Chimonodinium*, *Glenodiniopsis*, *Glenodinium* and *Gyrodinium*). sedDNA uniquely detected 16 dinoflagellate genera, and a further 4 genera (*Ceratium*, *Gymnodinium* and *Peridinium*) were detected in both the sedDNA and microscopy records (Fig. 3.4B).

Microscopy uniquely detected 21 ochrophyte genera, and 8 genera within this phylum were unique to the sedDNA record. In addition to these, 4 genera (*Dinobryon*, *Mallomonas*, *Ochromonas* and *Uroglena*) were detected with both methods (Fig. 3.4C).

sedDNA detected a total of 8 bacillariophyte genera, 4 of these were uniquely detected by sedDNA (*Staurosira*, *Opephora*, *Planothidium* and *Staurosirella*), and 4 were detected with both methods (*Aulacoseira*, *Diatoma*, *Discostella* and *Stephanodiscus*). In addition to these shared bacillariophyte genera, a further 23 genera were uniquely detected in the microscopy record (Fig. 3.4D).

Outside of these four phyla, microscopy also detected 15 charophyte genera, 5 cryptophyte genera and 1 haptophyte genus. sedDNA only detected 3 charophyte genera, 2 of which were also detected by microscopy.

A total of 669 phytoplankton ASVs were detected in the sedDNA record, and 410 (61%) of these ASVs were grouped into 105 genera. However, 259 phytoplankton ASVs (39%) had no definitive taxonomic assignment at the genus level, including 108 unassigned chlorophyte ASVs, 109 unassigned dinoflagellate ASVs, 35 unassigned ochrophyte ASVs, 4 unassigned bacillariophyte ASVs and 3 unassigned haptophyte ASVs. Within the microscopy-based monitoring record, there were 928 phytoplankton records, and 407 (44%) of these records were grouped into 170 genera. The remaining 521 records (56%) could not be identified to the genus level. This included 16 records of unassigned chlorophytes, 1 record of an unassigned dinoflagellate, 3 records of unassigned ochrophytes, 500 records of unassigned bacillariophytes and 1 record of an unassigned charophyte.

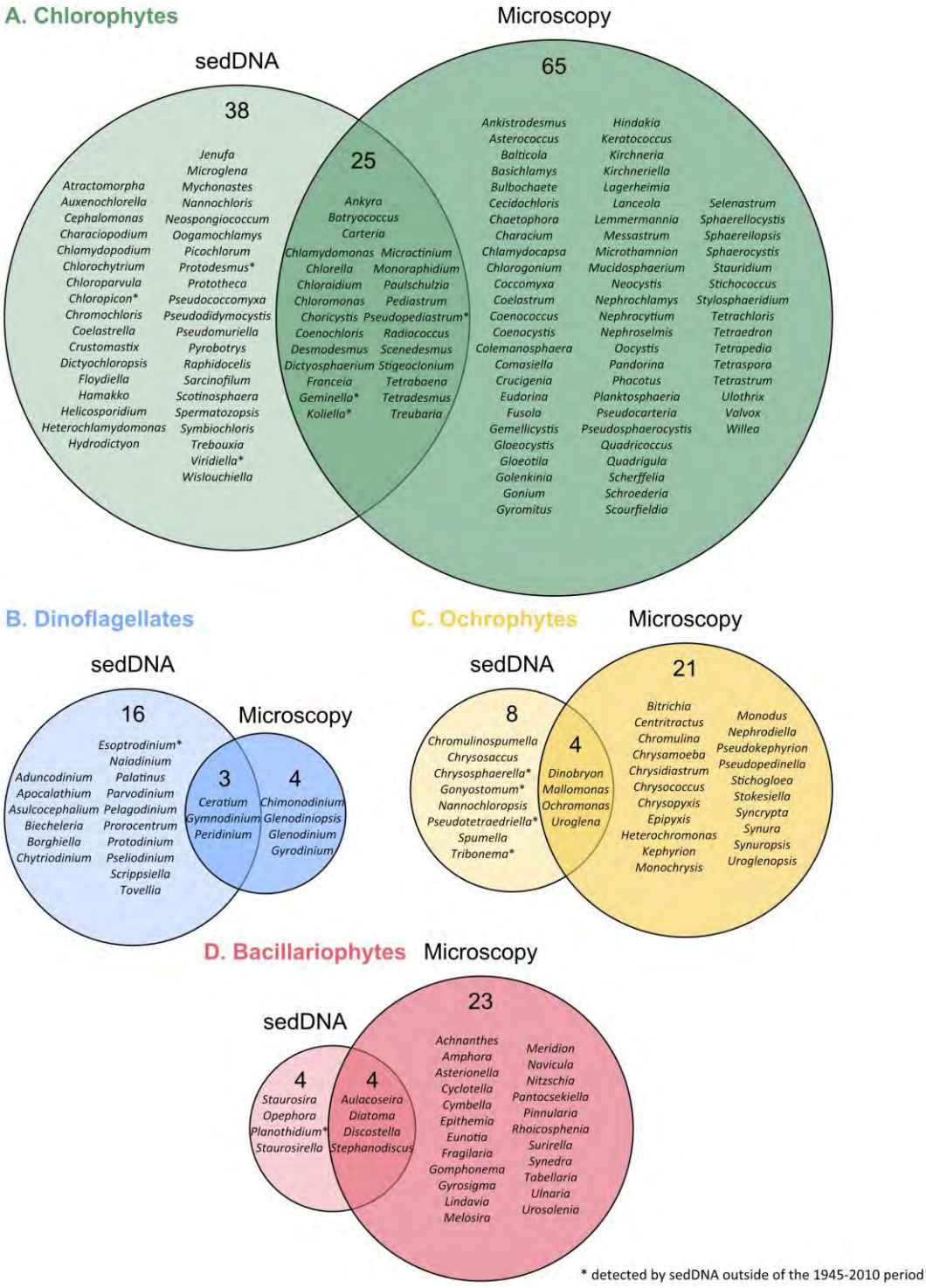

Figure 3.4. Shared and unique chlorophyte (A), dinoflagellate (B), ochrophyte (C), and bacillariophyte (D) genera detected by sedDNA and microscopy. * indicates genera detected by sedDNA outside of the 1945-2010 period covered by the microscopy record.

3.5. Discussion

Here, we have compared a sedDNA record with long-term microscopy-based monitoring to determine whether sedDNA can be used to reliably reconstruct past phytoplankton communities. Temporal trends in diversity and relative abundance and occurrence at the phylum level were broadly comparable between the sedDNA and microscopy records, but with some difference observed in the timing of the onset of community change. Each method detected a distinct composition of genera, with a small proportion of genera detected by both methods.

3.5.1. Data considerations

Differences between the sedDNA and microscopy records may arise from the way the data are collected, and this must be considered when comparing the two temporal records. For example, changes and improvements to the methods used throughout long-term monitoring schemes are to be expected. In the present study, the type of counting chamber used to produce the microscopy record changed from a sedimentation chamber to a Lund chamber in 1994. The way counts were recorded also varied throughout the monitoring scheme as cells were sometimes counted according to size, form or whether they were found in colonies. To alleviate some of the possible biases that may arise from changes to the counting method, the counts were converted to binary presence-absence values as a measure of temporal occurrence. A consequence of converting counts to occurrence is that this measure may not be directly comparable with the relative abundance values used in the sedDNA record, although a positive relationship between species occurrence and abundance has been widely observed (Gaston and He, 2011). Some issues could remain such as the ability of the observer to identify

phytoplankton to genus level by microscopy. This may have varied with the counting method used and the expertise and time investment of the observers, and the counts may have been biased towards more easily identifiable taxa or taxa of particular scientific interest.

There are also methodological factors associated with the sedDNA record which must be considered when making comparisons with the microscopy-based record. For example, the chronology of the sediment core was estimated based on the chronology of a separate sediment core collected in 2014 from the same location within Esthwaite Water. Application of this chronology required the assumption that the sedimentation rate remained constant since 2014, but variation in the sedimentation rate could lead to inaccuracies in the estimated chronology and therefore cause a discrepancy between the sedDNA and microscopy records. Only phytoplankton residing in the surface water were examined in the monitoring scheme, but sedDNA had the potential to record taxa originating from deeper within the water column and at the sediment surface. While contribution from active benthic phytoplankton may be relatively low at the depth the sediment core was collected due to low light availability, benthic taxa originating from littoral areas may have been transported to the sediment in the deeper basin during sediment resuspension and focussing (Mackay *et al*., 2012). The choice of 18S rRNA amplicon primers influences the composition of the community detected, and the accuracy of taxonomic assignment is limited by completeness of the reference database (Francioli *et al*., 2021). The 18S rRNA gene copy number can vary between taxa and lead to overestimations in relative abundance for some groups (Gong and Marchetti, 2019). Despite these data considerations, there were still remarkable similarities between the sedDNA and microscopy records, but possible explanations for the discrepancies between the records are explored further in the following sections.

3.5.2. Temporal trends in diversity

The NMDS of the dissimilarity matrices based on beta diversity as measured by sedDNA and microscopy both displayed comparable trajectories of change from older to more recent samples. The temporal trends in alpha diversity were also similar between the two records, with both showing an increase in diversity from the 1970s which coincided with the intensification of nutrient enrichment. A trend that is driven by the accumulation of DNA degradation with age could be expected to be a monotonic decline in diversity with sediment depth (Dommain *et al*., 2020). However, alpha diversity measured at the core surface was similar to that measured at the bottom of the core, and the temporal trends observed in the sedDNA record were accompanied by similar trends in the microscopy record. This provides evidence that the trends in diversity throughout the sediment core may represent a community response to environmental conditions, and not a trend that is primarily driven by an accumulation of DNA degradation with age. Previous studies have also shown that temporal trends in phytoplankton diversity as measured by sedDNA are consistent with environmental change and not necessarily DNA degradation (Capo *et al*., 2017; Huo *et al*., 2022; Zhang *et al*., 2021a). However, alpha genus diversity as measured by sedDNA was lower compared to that measured by microscopy. A lower diversity could be evidence of at least some DNA degradation, the extent of which may vary with conditions within the sediment (Torti *et al*., 2015).

3.5.3. Temporal trends in community composition

In both records, chlorophytes, dinoflagellates, ochrophytes and bacillariophytes displayed general increasing trends beginning around 1970-1990. These could be responses to nutrient enrichment, which accelerated in Esthwaite Water from the 1970s and remained high until the early 2000s (Appendix B, Fig. S2). A sediment core has previously been collected from

Esthwaite Water for sedimentary pigment analysis. In this record, many algal pigments also displayed increasing trends over time from the 1800s to 2011 with their highest concentrations detected after the 1970s, including chlorophyll *b* and lutein, which are indicative of chlorophytes, and diatoxanthin, which is indicative of bacillariophytes. However, there was a large peak in the concentration of diatoxanthin around the 2000s, and this was not reflected in the sedDNA or microscopy records for bacillariophytes (Moorhouse *et al*., 2018).

Co-occurring patterns in the microscopy record could support sedDNA as a reliable record of past community change. The relative abundance and occurrence of chlorophytes in the sedDNA and microscopy records, respectively, both increased sharply in more recent samples. However, the increase in chlorophyte relative abundance in the sedDNA record occurred over a decade later than the increase in occurrence in the microscopy record. Distinct peaks in the relative abundance and occurrence of dinoflagellates were observed in the sedDNA and microscopy records, but the timing of these peaks was also not aligned. sedDNA and microscopy may have recorded the same trends, but they may have been offset due to uncertainties in the chronology of the sediment core. Taphonomic processes could also have affected the ability of sedDNA to provide a reliable temporal record. For example, it was possible that there was a delay in the time taken for cells in the surface water to deposit in the sediment, particularly for smaller and more buoyant cells. Recently deposited cells and DNA may have become resuspended before complete burial and compaction within the sediment, and DNA may have migrated between sediment layers which could have disrupted the vertical organisation of DNA (Giguet-Covex *et al*., 2019), although it has been suggested that substantial DNA leaching between layers is unlikely to occur in the permanently saturated sediments of lakes (Anderson-Carpenter *et al*., 2011).

Degradation of DNA over time could limit the reliability of sedDNA reconstructions. Prior to 1970, the relative abundance of chlorophytes, dinoflagellates, ochrophytes and bacillariophytes was low and stable in the sedDNA record. Their occurrence in the microscopy record was also relatively low prior to 1970, but there were indications of a slightly higher occurrence in the earlier monitoring records between 1945 and 1950 which were not reflected in the sedDNA record. This could be evidence of some DNA degradation and a reduced ability of sedDNA to detect phytoplankton community change in older sediments. However, separating the effect of DNA degradation from an increase in the relative abundance of phytoplankton with an intensification of nutrient enrichment is complex as both factors could be expected to show a change in the same direction (i.e., an increase from older to more recent sediments). Heterotrophic eukaryotes that may have been active within the sediment such as fungi were also sequenced with the 18S rRNA amplicon primers, and their abundance within the sediment likely contributed to the lower relative abundance of these phytoplankton groups.

Cryptophytes were absent in the sedDNA record but were well-represented in the microscopy-based record, and alloxanthin, the diagnostic pigment of cryptophytes, was detected in the sediment core pigment record from Esthwaite Water (Moorhouse *et al*., 2018). Cryptophytes could therefore be expected to be detected using sedDNA, but similar to the present study, Capo *et al*. (2015) also reported that cryptophytes were poorly represented by sedDNA and suggested that the absence of a cell wall made their DNA vulnerable to degradation, and their high nutritional content made them vulnerable to grazing by zooplankton so that cells did not reach the sediment surface (Capo *et al*., 2015; Capo *et al*., 2021). Haptophytes were also poorly represented by sedDNA, and an underrepresentation of haptophytes in Lake Bourget, France, as measured by sedDNA has previously been reported (Capo *et al*., 2015). However, haptophyte temporal dynamics in an Antarctic lake throughout the Holocene have successfully been

reconstructed using sedDNA (Coolen *et al*., 2004), but the low temperatures in the Antarctic lake may have promoted DNA preservation. Haptophytes were not consistently counted throughout the monitoring scheme, so determining whether this group was underrepresented because they experienced greater rates of DNA degradation, or because they had a low abundance in Esthwaite Water is challenging. The reliability of sedDNA reconstructions depends on the extent of DNA degradation, which may occur at varying rates for different taxa in different environments (Capo *et al*., 2021). Previous efforts have been made to explore DNA degradation patterns in dinoflagellates and bacillariophytes in an Antarctic lake core record (Boere *et al*., 2011), and for cyanobacterial taxa within microcosms (Mejbel *et al*., 2022). However, the extent of DNA degradation that different taxa may be subject to in temperate lake sediments requires further research, particularly for groups that were not well-represented by sedDNA, such as cryptophytes and haptophytes.

3.5.4. Shared and unique genera

More genera were detected by microscopy compared to sedDNA within each phylum except dinoflagellates. For chlorophytes, ochrophytes and bacillariophytes, microscopy may have been more sensitive when distinguishing between genera. Only a small proportion of genera occurred in both records. This was highest for chlorophytes, with 19.5% of chlorophyte genera detected by both methods, but only 16.7% for dinoflagellates, 12.1% for ochrophytes, and 12.9% for bacillariophytes. The majority of genera were uniquely detected by each method, and each method may be capable of recording a different component of the phytoplankton community. Genera unique to the sedDNA record could include taxa that occupied deeper layers of the water column or littoral areas and were therefore beyond the scope of the surface water monitoring scheme, or those that were difficult to identify based on morphology. Depositional

or degradational processes could explain why a large proportion of the phytoplankton community were missed by sedDNA. Previous studies have shown that certain groups of eukaryotic algae (Gauthier *et al*., 2021) and cyanobacteria (Nwosu *et al*., 2021) were differentially represented in surface lake sediments compared to the water column, and this could be because some taxa did not readily deposit. The deposition potential of phytoplankton could be affected by grazing pressure and whether the cells form colonies or aggregate with organic matter which makes them heavier and more likely to deposit and could also protect the DNA from degradation (Gauthier *et al*., 2021; Mauvisseau *et al*., 2022; Nwosu *et al*., 2021).

A larger number of dinoflagellate genera were detected in the sedDNA record compared to the microscopy record, and sedDNA may therefore be a particularly valuable method for studying past dinoflagellate communities. Many dinoflagellates form a robust cyst during the resting stage of their lifecycle which may protect their DNA from grazing by zooplankton and other extracellular degradation processes (Bravo and Figueroa, 2014). Dinoflagellates have previously been shown to be well-represented by sedDNA, but it was possible that they were overrepresented due to their large genomes and high 18S rRNA gene copy number (Gong *et al*., 2020).

Bacillariophyte DNA could also be expected to be preserved in sediments due to the presence of the protective silica frustule (Aguirre *et al*., 2018). However, this group was present at the lowest relative abundance of the four main phyla in the sedDNA record, despite being one of the groups with the highest occurrence in the microscopy record, and a larger number of bacillariophyte genera were detected by microscopy. Another sediment core collected from Esthwaite Water for microfossil analysis found *Asterionella*, *Aulacoseira* and *Fragilaria* to be the most dominant genera between 1945 and 2005 (Dong *et al*., 2012). These genera were also detected by microscopy in the monitoring record from 1945 to 2010, but *Asterionella* and

Fragilaria were absent from the sedDNA record. Targeted primers may be more capable of distinguishing a larger number of bacillariophyte genera compared to the broad range 18S rRNA amplicon primers selected in the present study, such as primers targeting the rbcL gene (Anslan *et al*., 2022; Dulias *et al*., 2017; Kang *et al*., 2021). Although a substantial number of bacillariophytes were missed by sedDNA, a small number of genera were detected which may have been overlooked in the microscopy-based monitoring and microfossil records. This included *Staurosira*, *Opephora*, *Planothidium* and *Staurosirella*. *Planothidium* are typically benthic taxa (Lange-Bertalot *et al*., 2017), and may therefore have been outside of the scope of the surface water monitoring scheme, although *Planothidium* was only detected by sedDNA after 2010. The bacillariophyte community sequenced in lake surface sediments has previously been compared with microscopy-based methods, and also revealed that while microscopy could detect more genera, each method detected a distinct proportion of the community (Dulias *et al*., 2017; Kang *et al*., 2021).

A substantial proportion of ASVs detected by sedDNA (39%) and records in the microscopy-based monitoring scheme (56%) were unidentified at the genus level. In the sedDNA record, the majority of ASVs unassigned at the genus level were chlorophytes and dinoflagellates, while in the microscopy record, a significant number of bacillariophytes were unidentified at the genus level. Taxonomy assignment in the sedDNA record may be limited by reference database coverage (Anslan *et al*., 2022). Taxonomic identification with microscopy may be limited by microscope resolution and the expertise and time investment of the observers, which may vary throughout the monitoring scheme. Separation of the influence of these variables in long-term monitoring schemes from an environmental response is complex (Straile *et al*., 2013). While palaeolimnological tools such as sedDNA typically do not suffer from method changes, they may be subject to other limitations such as DNA degradation. Each method has

its own limitations and biases, and multi-proxy analysis is likely the most reliable approach for reconstructing past phytoplankton communities.

3.5.5. Conclusions and recommendations for the use of sedDNA in palaeolimnology

Validation of sedDNA against concurrent lake monitoring is crucial to further its development and evaluate its performance as a palaeolimnological tool. Our comparison with long-term microscopy-based monitoring of phytoplankton in the lake surface water revealed broadly similar trends in the diversity and relative abundance and occurrence of chlorophytes, dinoflagellates, ochrophytes and bacillariophytes, and up to 20% of genera were detected with both methods. These results support the use of sedDNA as an effective tool for the reconstruction of past phytoplankton communities, at least within the time period investigated in this study. However, DNA degradation may occur in older sediments which could limit the reliability of reconstructions over longer time periods, and a substantial proportion of the phytoplankton community detected by microscopy were missed by sedDNA. Based on these results, we recommend that sedDNA reconstructions over time periods exceeding 100 years or of groups such as cryptophytes that were poorly resolved with sedDNA are treated with caution, and future research should focus on identifying the key determinants of variable DNA degradation and deposition among taxa. Furthermore, due to incomplete reference databases, it is important that future studies consider the fact that phytoplankton sedDNA reconstructions may only represent a subset of the total community in lakes. Continued improvements to reference database coverage, in addition to the combined use of multiple targeted primers may enable the wider phytoplankton community to be captured with sedDNA.

Chapter 4: Source tracking of lake sedimentary DNA and implications for its use in palaeolimnology

4.1. Abstract

Sedimentary DNA (sedDNA) extracted from lake sediment cores is increasingly being used in palaeolimnology, but the reliability of sedDNA reconstructions is limited by unknowns surrounding the transport and preservation of DNA in lakes and lake sediments. Here, we employ a source tracking approach to trace DNA from the water column to the sediment and investigate seasonal and water depth related trends in deposition. 16S rRNA and 18S rRNA gene amplicon sequencing of the bacterial and eukaryotic microbial community, respectively, was carried out on lake water samples collected on a seasonal basis and at varied depths within the water column, and on sediments from a lake sediment core spanning approximately 105 years. Fast expectation-maximisation microbial source tracking (FEAST) was implemented to estimate contribution from the pelagic bacterial and eukaryotic microbial communities to the sediment. Quantitative PCR (qPCR) was performed to quantify total bacterial and cyanobacterial 16S rRNA gene copies down-core. The concentration of 16S rRNA gene copies and contribution from bacterial and eukaryotic microbial communities in the water column to the lake sediment generally declined down-core. These results indicate that DNA degradation may have accumulated over time, and detection of deposited DNA and therefore the reliability of sedDNA reconstructions may be reduced in older sediments. The majority of deposited DNA may originate from pelagic microbial communities in the summer when productivity is higher, and a large proportion of deposited bacterial DNA may originate from communities residing deeper within the water column where rates of DNA degradation may be lower. However, a substantial proportion of the sedDNA community could not be attributed to any of the sampled sources. More comprehensive field surveying across lakes and their catchments is therefore needed to improve source detection of sedDNA, and due to currently unknown rates of interannual community turnover, further research is needed to trace deposition of DNA from

the water column over longer timescales. An improved understanding of the source and fate of lake sedDNA can further its development as a reliable palaeolimnological tool.

4.2. Introduction

Palaeolimnological reconstructions of lake communities are important for understanding how different organisms have responded to environmental change in the past, but these reconstructions rely on the application of reliable palaeolimnological tools (Birks and Birks, 2006; Wingard *et al*., 2017). Lake sedimentary DNA (sedDNA) is a relatively new palaeolimnological tool which can be applied to sedimentary archives to provide temporal records of community change for a wide diversity of lake biota, including bacteria (Thorpe *et al*., 2022), eukaryotic algae (Chapter 3; Capo *et al*., 2016), zooplankton (Tsugeki *et al*., 2022), and macrophytes (Stoof-Leichsenring *et al*., 2022). However, the reliability of sedDNA records depends on the extent and rate of DNA deposition and degradation, and these processes are poorly understood in lakes and lake sediments (Capo *et al*., 2022).

Lake sedDNA consists of DNA derived from sediment-dwelling organisms and DNA that has been deposited to the sediment from the water column (Pawlowski *et al*., 2022; Torti *et al*., 2015). The deposited fraction of sedDNA may originate from communities residing within the lake water column or those which have been transported to the lake from elsewhere within the catchment, such as upstream water and sediment sources or runoff from surrounding soil (Giguet-Covex *et al*., 2019). How much of this DNA is deposited in the lake sediment compared to how much is flushed from the lake may depend on whether it is intracellular or extracellular, whether it is bound to sediment particles, and factors such as cell size, buoyancy, and position within the water column. For example, extracellular DNA or small and buoyant cells may not

readily deposit unless bound to heavier cells or organic matter (Vuillemin *et al*., 2017). Extracellular DNA and cells that reside in the epilimnion during lake stratification may not reach the sediment as mixing between the layers of the water column is restricted, although the ease of deposition may be improved when lake mixing resumes (Harrison *et al*., 2019). Previous studies using sedDNA have found that some pelagic microbes deposit poorly in lake sediments, and this could be related to their buoyancy and position in the water column during stratification (Gauthier *et al*., 2021; Nwosu *et al*., 2021). Differential deposition could result in some taxa becoming over or underrepresented in sedDNA records and may therefore limit the reliability of palaeolimnological reconstructions for pelagic microbes.

Degradation of DNA, which can occur by pre- or post-depositional processes, can also affect the reliability of sedDNA-based reconstructions. DNA degradation may take place in the water column prior to deposition by oxidation, photooxidation, hydrolysis, grazing by zooplankton, and microbial degradation (Brasell *et al*., 2022; Torti *et al*., 2015). The rates of these processes may be affected by lake conditions, and DNA degradation is typically greater in warmer, acidic lakes with higher light and oxygen availability (Torti *et al*., 2015; Zulkefli *et al*., 2019). The rate of degradation can also vary between layers of the water column in stratified lakes, and greater rates of DNA degradation have been observed in the epilimnion compared to the colder hypolimnion (Harrison *et al*., 2019; Matsui *et al*., 2001). Once incorporated within the sediment matrix, sediment particles can physically shield DNA from degradation, and the anoxic environment of sediments may be more favourable towards the preservation of DNA (Torti *et al*., 2015). The extent of DNA degradation may therefore depend on the time taken for deposition, and DNA that remains in the water column for longer may accumulate more damage compared to DNA that is rapidly incorporated within the sediment upon cell death. However, DNA may continue to degrade within the sediment by processes such as microbial degradation,

although this typically declines with sediment depth (Haglund *et al*., 2003). Free, extracellular DNA is particularly vulnerable to degradation, while DNA from organisms with robust cell membranes is more likely to persist as intracellular DNA and is therefore better protected from degradation (Han *et al*., 2022). Extensive DNA degradation could reduce the detection of deposited DNA in the sediment, and if damage accumulates with time, this could limit the time period over which sedDNA can be used reliably.

Previous applications of sedDNA have largely focussed on reconstructing eukaryotic algal and cyanobacterial communities (Capo *et al*., 2016; Domaizon *et al*., 2013; Ibrahim *et al*., 2020; Monchamp *et al*., 2016, 2019), while sedDNA reconstructions of the bacterial community are rare in comparison (Li *et al*., 2019a; Thorpe *et al*., 2022). Reliable sedDNA reconstructions of bacteria are particularly complex because their small cells may not readily deposit in the sediment, and the active community of heterotrophic bacteria within the sediment may obscure the temporal signal from the deposited pelagic community (Thupaki *et al*., 2013; Vuillemin *et al*., 2017). To enable reliable reconstructions of past microbial communities using sedDNA, further research is needed to distinguish between pelagic DNA that has been deposited over time from modern DNA that originated from the sediment community.

This chapter investigates the transport and preservation of DNA in lakes. Microbial community composition in the sediment is compared with that in the water column. A microbial source tracking (MST) approach is then implemented to estimate the degree to which taxa occupying the water column (source community) may be deposited and preserved in the sediment (sink community). Community MST methods include recently developed machine-learning tools which assume that a sink community is comprised of a combination of source communities, and by exploiting the differences in community composition between sinks and sources, these MST tools can quantify contribution from potential sources to a sink (Knights *et al*., 2011;

Shenhav *et al*., 2019). Machine-learning MST has frequently been used to trace transport of microbes in streams, rivers, and lakes in modern samples (Brown *et al*., 2017; Comte *et al*., 2017; Hermans *et al*., 2020; Xu *et al*., 2022; Zhang *et al*., 2019), but to our knowledge, this is the first application of machine-learning MST to a lake sedDNA record in palaeolimnology. This chapter will also use quantitative PCR (qPCR) to measure the concentration of bacterial and cyanobacterial gene copies down-core to understand how the detection and therefore preservation of DNA may vary with sediment depth.

4.3. Materials and methods

4.3.1. Study site

Esthwaite Water (54° 21.56' N, 2° 59.15' W) is a relatively small, eutrophic lake in the Windermere catchment of the Lake District National Park, Cumbria, UK (Appendix A, Fig. S1). Thermal stratification in Esthwaite Water usually begins in April and starts to break down in October (Maberly *et al*., 2011a). Detailed site characteristics and its history of intense eutrophication have previously been described (Dong *et al*., 2011; Maberly *et al*., 2011a; Mackay *et al*., 2012; Thorpe *et al*., 2022).

4.3.2. Water sampling

Water samples of integrated depths were collected from the deepest point of Esthwaite Water using a Ruttner water sampler. Up to 1 L of water was collected and transferred to a sterile Whirl-Pak bag or glass bottle for filtering. In August 2021, a water sample was collected from an integrated depth of 0-5, 5-10 and 10-15 m. Water samples were also collected from this site on a seasonal basis in spring (April 2021), summer (July 2021), autumn (October 2021) and

winter (February 2022) from an integrated depth of 0-5 m. Water sampling depth was measured using a graduated line attached to the water sampler.

Each water sample was filtered on the day of collection. Water samples collected in August were filtered in the field using 0.2 µm Sterivex filter units (Merck Millipore, MA, U.S.). A sterile Luer-lock syringe was filled with 60 ml of water and forced through the filter unit. This was repeated until between 240 and 500 mL of water was filtered per sample depending on turbidity. The filter units were filled with 2 ml DNA shield (Zymo Research, CA, U.S.) to stabilise DNA, and sealed with a Luer-lock cap. Water samples collected in April, July and October 2021 and in February 2022 were transported to the UK Centre for Ecology & Hydrology (UKCEH), Lancaster, and up to 500 mL was filtered using 0.2 µm Supor filter membranes (Pall Corporation, NY, U.S.). All filtered water samples were stored at -20 °C prior to DNA extraction.

4.3.3. Sediment coring and chronology

A sediment core was collected from the deepest point of Esthwaite Water using a HTH gravity corer (Pylonex, Sweden) immediately after water sampling in August 2021. Collection, sectioning, and radiometric dating of the core have been described in detail in Chapter 3. Briefly, this 35 cm long core was sectioned in 1 cm intervals with a blade that was sterilised between sections. Each section was then subsampled to collect a small amount of undisturbed sediment from the centre of each section for DNA extraction. The chronology of the sediment core was estimated using the age-depth relationship of a reference core collected from the same location within Esthwaite Water in 2014 assuming a constant sedimentation rate, as described

in detail in Chapter 2 (Thorpe *et al*., 2022) and Chapter 3. The full length of the 35 cm sediment core was estimated to cover a 105-year time period from 1916 to the date of collection in 2021.

4.3.4. DNA extraction

DNA was extracted from 0.25 g of each sediment core subsample using the Qiagen DNeasy PowerSoil Pro extraction kit (Qiagen, Germany) as previously described in Chapter 3. For water samples collected in August, ethanol-cleaned pipe cutters were used to open Sterivex filter units, and the filter was removed. For all water samples, filters were cut into small pieces using sterile scalpel blades and transferred into lysis tubes for DNA extraction using the Qiagen DNeasy PowerWater extraction kit.

4.3.5. Quantitative PCR

Probe-based qPCR was performed to quantify 16S rRNA gene copies of the total bacterial community and of the cyanobacterial community throughout the sediment core. The bacterial 16S rRNA primers, 1369F (5'-CGGTGAATACGTTCYCGG-3') and 1492R (5'-GGWTACCTTGTTACGACT-3') were used with the TM1389F probe (5'-6-FAM/CTTGTACAC/ZEN/ACCGCCCGTC/IABkFQ-3') (Suzuki *et al*., 2000), and the cyanobacterial-specific 16S rRNA primers, 107F (5'-ACGGGTGAGTAACRCGTRA-3') and 377R (5'-CCATTGCGGAAAATTCCCC-3') were used with the cyanobacterial-specific probe (5'-6-FAM/CTCAGTCCC/ZEN/AGTGTGGCTGNTC/IABkFQ-3') (Picard *et al*., 2022b; Rinta-Kanto *et al*., 2005). PrimeTime probes (IDT, IA, U.S.) were double quenched, and contained a 5' 6-FAM dye, internal ZEN quencher and 3' Iowa Black quencher. gBlock gene fragments (IDT, IA, U.S.) of a target bacterial and cyanobacterial 16S rRNA gene sequence

were quantified using the Invitrogen Qubit dsDNA BR assay kit (Thermo Fisher Scientific, MA, U.S.), serially diluted with qPCR grade water from approximately 10^1 to 10^8 copies (Appendix C, Eq. 1) and used as standards within the qPCR run to produce a standard curve for absolute quantification of gene copy number.

Each 20 µL reaction contained 10 µL of a 2X qPCR master mix (Roche, Switzerland), 4.56 µL of qPCR grade water, 0.2 µL of each 10 µM forward and reverse total bacterial 16S rRNA or cyanobacterial 16S rRNA primer, 0.04 µL of the 2 µM total bacterial or cyanobacterial-specific probe, and 5 µL of 1:20 diluted DNA to overcome inhibition or 5 µL of each serially diluted gBlock standard. Negative controls were included, and all samples and standards were run in duplicate. The qPCR program was set to an initial denaturing temperature of 95 °C for 7 min, followed by 40 cycles of 95 °C for 10 sec, 60 °C for 30 sec and quantification at 72 °C for 1 sec. qPCR was performed on the Roche LightCycler 480 II with the LightCycler 480 software version 1.5. Reactions passed if run efficiency was between 90 and 110%, and if the cycle threshold (Ct) of duplicate technical replicates were within 0.5 cycles of each other, as recommended by Nolan *et al*. (2006). The fit points quantification method (Rasmussen, 2001) was used to estimate gene copy number from the Ct according to the standard curve (Appendix C, Fig. S2), and copies per reaction were converted to copies per g of wet sediment (Appendix C, Eq. 2).

4.3.6. PCR amplification and sequencing

Using DNA extracted from water samples and the sediment core, two-step PCR was performed to amplify the V4 region of the 16S rRNA gene with universal forward and reverse bacterial primers, 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-

GGACTACNVGGGTWTCTAAT-3') (Walters *et al*., 2016), and the V4-V5 region of the 18S rRNA gene with universal forward and reverse microbial eukaryote primers, NSF563 (5'-CGCGGTAATTCCAGCTCCA-3') and NSR951 (5'-TTGGYRAATGCTTTCGC-3') (Mangot *et al*., 2012). First step PCR conditions were as described in Chapter 2 (Thorpe *et al*., 2022) for amplification of the 16S rRNA gene, and in Chapter 3 for amplification of the 18S rRNA gene. First step PCR reagents and all subsequent steps (PCR clean-up, second step PCR conditions and reagents, normalisation, gel-extraction, and quantification) were as described in Chapter 3 for both amplicons. The 16S and 18S rRNA amplicon libraries were pooled and sequenced on the Illumina MiSeq Platform with a 500-cycle v2 MiSeq reagent kit (Illumina, CA, U.S.).

4.3.7. Sequence data processing

The sequence reads were trimmed, quality-filtered and merged using the DADA2 pipeline (Callahan *et al*., 2016) as described in Chapter 3, and an amplicon sequence variant (ASV) abundance table was generated per amplicon. Taxonomy was assigned to each ASV using the naive Bayesian classifier (Wang *et al*., 2007) against the SILVA database v.132 (Quast *et al*., 2013) for the 16S rRNA gene sequences, and against the PR^2 database v4.14.0 (Guillou *et al*., 2013) for the 18S rRNA gene sequences. The 16S and 18S rRNA gene sequences were rarefied to uniform sequencing depths of 6,496 and 14,936 reads, respectively. Three sediment core samples (14, 29 and 32 cm) that did not meet the rarefied 16S rRNA sequencing depth, and two sediment core samples (4 and 33 cm) that did not meet the rarefied 18S rRNA sequencing depth were excluded. Bacterial and eukaryotic microbial ASVs were filtered to remove those with no taxonomic assignment at the phylum level.

4.3.8. Source tracking analysis

Fast expectation-maximisation microbial source tracking (FEAST) was implemented to quantify the contribution of source microbial communities to the sink microbial communities using the FEAST R package with 1000 expectation-maximisation iterations (Shenhav *et al*., 2019). Each water sample was designated as a source, and each section of the sediment core was designated as a sink. The bacterial and eukaryotic microbial communities were analysed in independent FEAST runs using the rarefied ASV abundance tables as the input files.

The source tracking analysis was repeated at the operational taxonomic unit (OTU) level. Using the DECIPHER R package v2.0 (Wright, 2016), rarefied and quality-filtered 16S rRNA and 18S rRNA gene sequences from the DADA2 pipeline were aligned and a distance matrix was generated. The sequences were clustered into OTUs based on 97% sequence similarity. FEAST was implemented as described above, but using the OTU abundance tables as the input files.

4.4. Results

4.4.1. qPCR

qPCR was performed down the sediment core to measure the concentration of 16S rRNA gene copies for the total bacterial community and for cyanobacteria (Fig. 4.1A and B). The concentration of total bacterial and cyanobacterial 16S rRNA gene copies were relatively low and stable from the 1910s to the 1970s and did not exceed 2.64×10^9 total bacterial gene copies g^{-1} wet sediment (Fig. 4.1A) or 2.76×10^7 cyanobacterial gene copies g^{-1} wet sediment (Fig. 4.1B) during this period. The concentration of both genes increased to a small peak from the 1970s to 1980s, when total bacterial 16S rRNA gene copies reached 4.48×10^9 g^{-1} wet sediment before increasing towards the surface of the core, and cyanobacterial 16S rRNA gene copies

reached 8.54 × 10^7 g^{-1} wet sediment. Cyanobacterial 16S rRNA gene copies then displayed a larger peak from the 1980s to the 1990s where the concentration reached 3.11 × 10^8 copies g^{-1} wet sediment. Cyanobacterial 16S rRNA gene copies were between 0.6 and 3.0% of the total bacterial 16S rRNA gene copy concentration except during the larger peak when cyanobacterial copies made up 8.8% of the total bacterial copies.

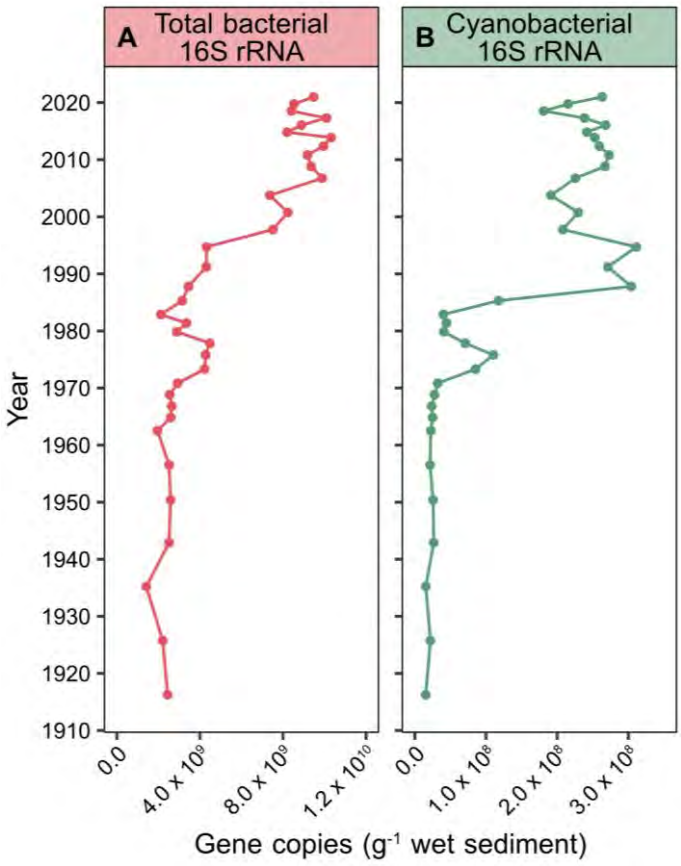

Figure 4.1. Concentration of total bacterial 16S rRNA (A) and cyanobacterial 16S rRNA (B) gene copies per g of wet sediment throughout the sediment core as determined with qPCR.

4.4.2. Community composition

Actinobacteria, Bacteroidetes, Cyanobacteria, Proteobacteria, Verrucomicrobia and Planctomycetes were the most abundant bacterial phyla detected within the water column (Fig. 4.2A and B). Cyanobacteria comprised up to a third of the bacterial community in April and July but were a much smaller component in October, when the community was dominated by Verrucomicrobia, and in February, when the community was dominated by Proteobacteria. Of

each depth sampled within the water column in August, the relative abundance of Cyanobacteria was greater at 5-10 m compared to that at 0-5 or 10-15 m (Fig. 4.2B). The relative abundance of Actinobacteria was consistent between each depth sampled, but the relative abundance of Acidobacteria, Bacteroidetes and Proteobacteria was greater, and the relative abundance of Planctomycetes and Verrucomicrobia was lower at 10-15 m compared to that at 0-5 or 5-10 m.

The sediment bacterial community was dominated by Proteobacteria and Bacteroidetes, which increased in relative abundance from older to more recent sediments, and by Firmicutes and Chloroflexi, which decreased in relative abundance towards the surface of the core. Cyanobacteria peaked in relative abundance in the mid-1990s (Fig. 4.2C).

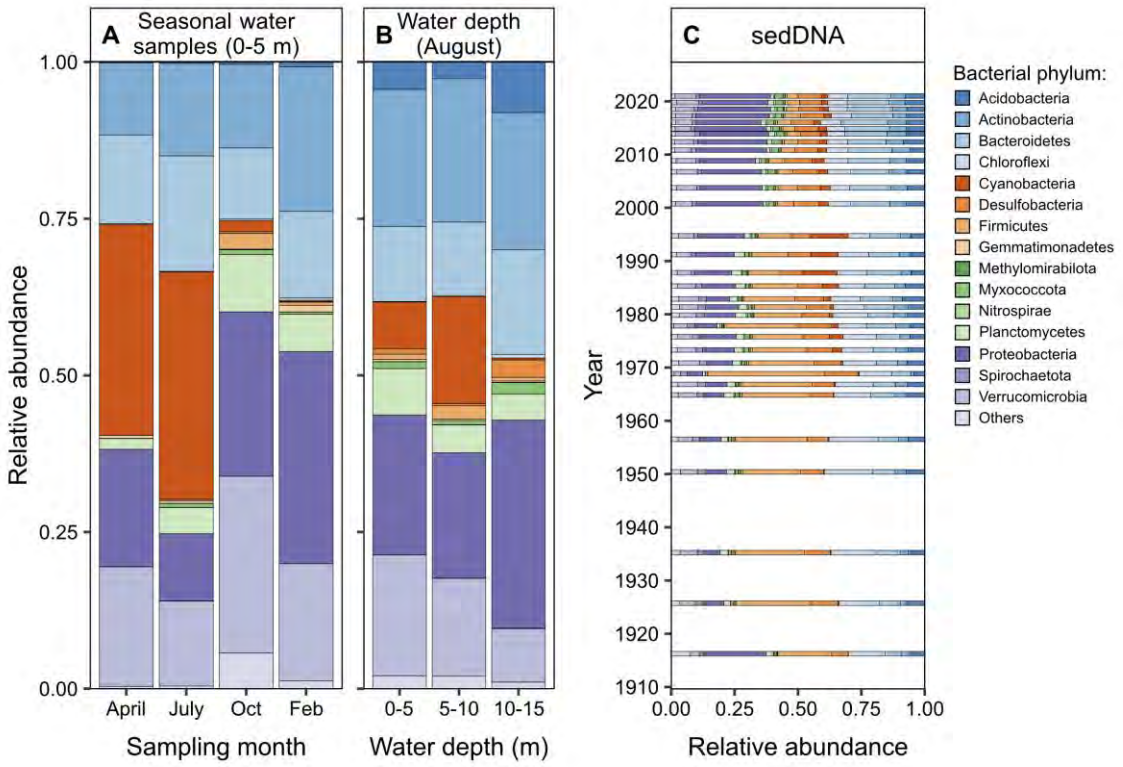

Figure 4.2. Bacterial community composition at the phylum level in water samples collected from an integrated depth of 0-5 m in April 2021, July 2021, October 2021, and February 2022 (A), in water samples collected from an integrated depth of 0-5, 5-10 and 10-15 m in August 2021 (B), and throughout a lake sediment core (C). The top 15 most abundant phyla are shown, and the less abundant phyla are presented as 'others'.

The dominant eukaryotic microbes in the lake water varied seasonally (Fig. 4.3A). In April, the community was dominated by bacillariophytes, but ciliates and fungi were also abundant. In July, ciliates comprised over half of the community, and dinoflagellates were more abundant in July than any other month. Streptophytes dominated the community in October with a relative abundance of 0.86. In February, fungi and ciliates each comprised a third of the community. In August, ciliates and cercozoa dominated the community at each depth (Fig. 4.3B). Bacillariophytes, chlorophytes, choanoflagellates and streptophytes had a higher relative abundance at a depth of 10-15 m, but cryptophytes had a higher relative abundance at 0-5 m.

The eukaryotic microbial community in older sediments was dominated by fungi and ciliates which together comprised over half of the community in the deepest sediments estimated to have been deposited prior to the 1980s (Fig. 4.3C). The relative abundance of fungi declined towards the surface of the core while the relative abundance of dinoflagellates and chlorophytes increased.

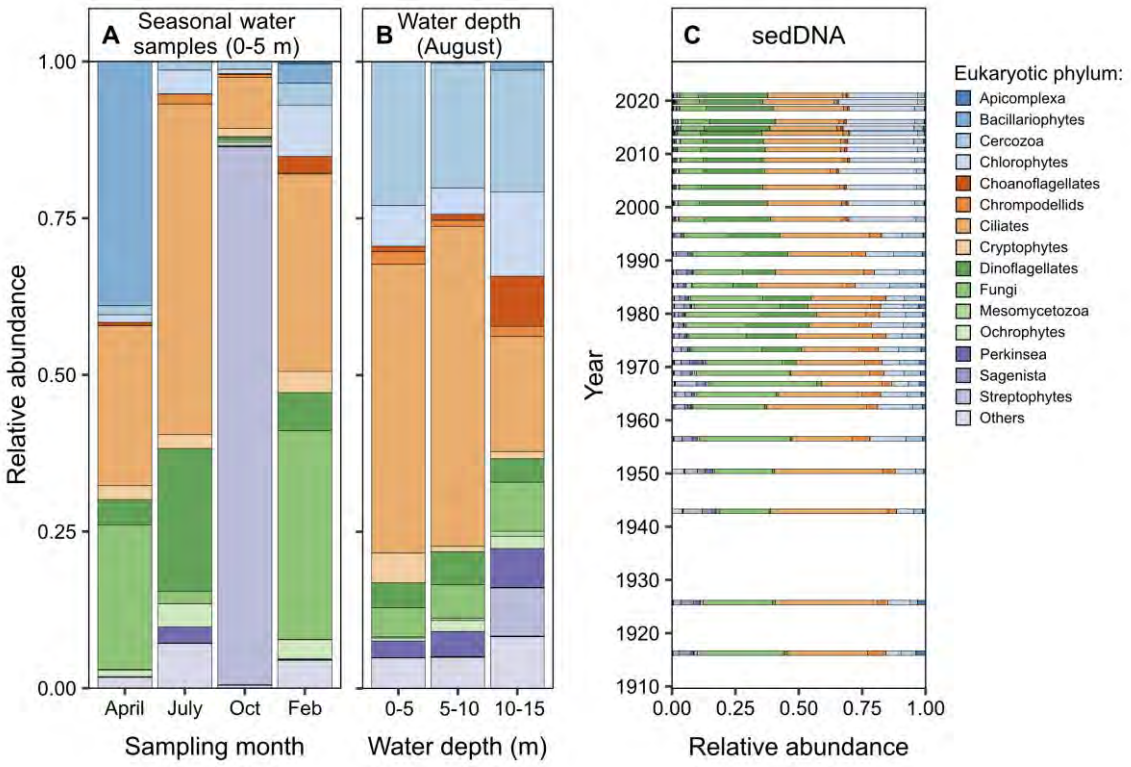

Figure 4.3. Eukaryotic microbial community composition at the phylum level in water samples collected from an integrated depth of 0-5 m in April 2021, July 2021, October 2021, and February 2022 (A), in water samples collected from an integrated depth of 0-5, 5-10 and 10-15 m in August 2021 (B), and throughout a lake sediment core (C). The top 15 most abundant phyla are shown, and the less abundant phyla are presented as 'others'.

4.4.3. Source tracking

The total contribution from the bacterial and eukaryotic microbial communities at a depth of 0-5 m within the water column across all months sampled generally increased towards the surface of the core. For bacteria, total contribution from this depth increased from 0.82% at the bottom of the core to 3.11% at the surface of the core and peaked at 3.76% in 2006 which was largely derived from the community in July (Fig. 4.4A and C). For eukaryotic microbes, total contribution from a depth of 0-5 m within the water column across all months sampled increased from 11.90% at the bottom of the core to 26.17% at the surface of the core and peaked at 33.57% in 2019 (Fig. 4.4B and D). The community in August at a depth of 0-5 m within the

water column was the largest contributor to the eukaryotic microbial sedDNA record, followed by the community in July and February, and in April for sediments deposited in the 1970s and 1980s. Contribution from the eukaryotic community in October was less than 0.13% and only detected in five sediment core sections.

Contribution from the bacterial community at different depths within water column in August to older sediments was relatively low and in sediments deposited prior to 1935, contribution was only detected from a water column depth of 10-15 m (Fig. 4.4C). For sediments deposited from 1973 to 1985, the community at a water depth of 5-10 m was the largest contributor and peaked at 5.75% in 1975. For sediments deposited from 1987 to 2021, contribution from a water depth of 10-15 m increased, peaked at 4.39% in 2014, and was greater than that of all other water column depths sampled in August. Contribution from the community at a water column depth of 0-5 m in August to the bacterial sedDNA record remained relatively low and did not exceed 1.71%. For eukaryotic microbes, contribution from the community at a depth of 0-5 m in August to the sedDNA record was consistently higher than that from deeper within the water column and ranged between 4.21 and 22.45%. Contribution from water column depths of 5-10 and 10-15 m were low in comparison, and peaked at 5.04% and 2.80%, respectively (Fig. 4.4D).

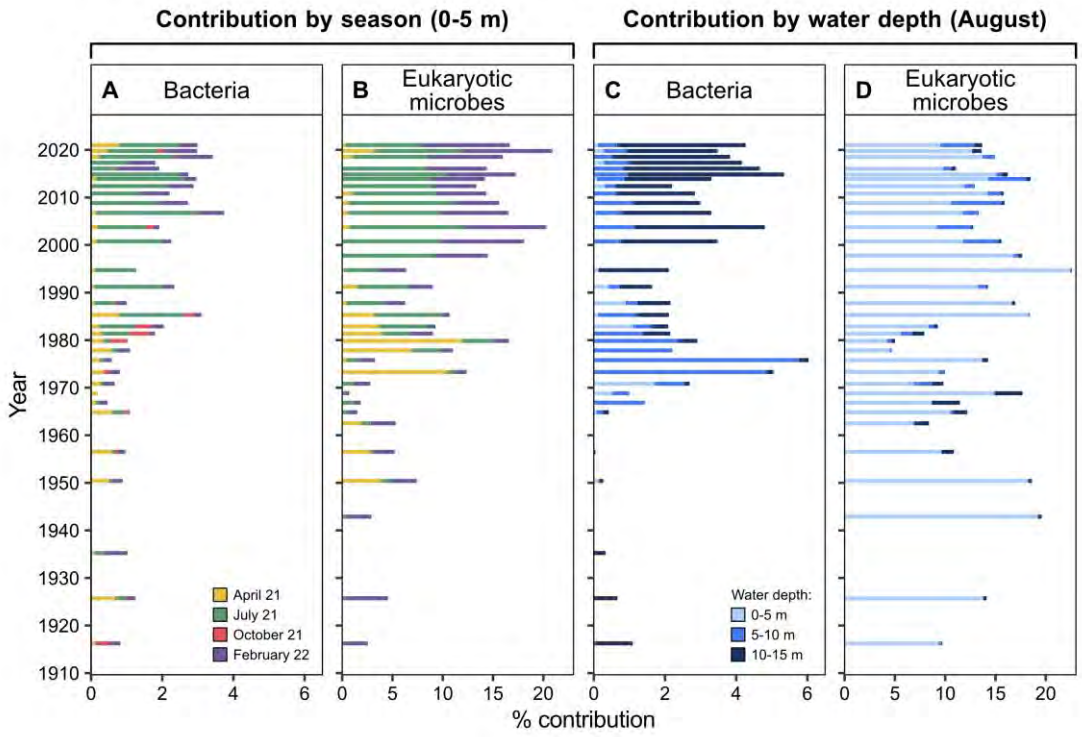

Figure 4.4. FEAST-estimated contribution (%) of the water community at an integrated depth of 0-5 m in April 2021, July 2021, October 2021, and February 2022 to the sedDNA record for bacteria (A) and eukaryotic microbes (B), and contribution of the water community at an integrated depth of 0-5, 5-10 and 10-15 m in August 2021 to the sedDNA record for bacteria (C) and eukaryotic microbes (D).

Source tracking implemented at the OTU level revealed similar temporal trends in source contribution to that performed at the ASV level (Appendix C, Fig. S3). There was an increase of up to 8% and 54% in contribution from bacterial and eukaryotic OTUs in the water column, respectively, to the sedDNA records compared to the ASVs, and this increase was most notable for contribution from the eukaryotic community in July. Unknown contribution to the bacterial and eukaryotic sedDNA record was up to 10% and 46% lower, respectively, compared to that at the ASV level.

4.5. Discussion

4.5.1. Concentration of 16S rRNA gene copies down-core

qPCR was used to quantify 16S rRNA gene copies for the total bacterial community and cyanobacterial community down-core. The concentration of total bacterial 16S rRNA gene copies likely included inputs from the deposited pelagic and sediment-dwelling bacterial communities. As a photosynthetic group, cyanobacteria are not expected to be active below the sediment surface due to limited light availability. Quantifying cyanobacterial gene copies can therefore give an indication of the concentration of deposited DNA down-core without contribution from an active sediment community. The concentration of total bacterial 16S rRNA gene copies was higher in more recent sediments and generally declined down-core. This could suggest that towards the top of the core, recently deposited DNA was less degraded and there was substantial contribution from the in situ sediment bacterial community, while in older sediments, DNA degradation accumulated and input from the sediment bacterial community decreased. A decline in the abundance of live bacterial cells with increasing lake sediment depth has previously been shown (Haglund *et al*., 2003). However, a comparable trend was observed in the concentration of cyanobacterial 16S rRNA gene copies down-core, indicating that this declining trend in 16S rRNA gene copies could largely have been driven by an accumulation of DNA degradation with age, and not only by a smaller sediment community. A decline in the concentration of bacterial gene copies (Pilon *et al*., 2019; Swan *et al*., 2010) and cyanobacterial gene copies (Cao *et al*., 2020; Pal *et al*., 2015; Picard *et al*., 2022a, b) down lake sediment cores have previously been observed, and an accumulation of DNA degradation could limit the ability of sedDNA to reliably detect bacterial community change in some older sediments. However, sedDNA has been used to reconstruct cyanobacterial community dynamics over time periods as long as 1000 years (Picard *et al*., 2022a).

4.5.2. Water and sediment community composition

Proteobacteria and Bacteroidetes were among the most relatively abundant bacterial groups in both water samples and sediments from the sediment core. However, while Firmicutes and Chloroflexi were also found to be dominant in sediments and increase in relative abundance down-core, they were a much smaller component of the bacterial community in the water. Previous studies have also reported a dominance of Firmicutes and Chloroflexi in lake sediments compared to the water column (Newton *et al*., 2011; Ren *et al*., 2019; Ren *et al*., 2022; Zhong *et al*., 2022). A large proportion of sedDNA from Firmicutes and Chloroflexi may have originated from sediment communities as opposed to being deposited from a pelagic community, and the down-core trends observed in their relative abundance may therefore represent spatial responses to sediment conditions instead of temporal trends derived from deposited DNA.

The bacterial water community was dominated by cyanobacteria in April and July, and the eukaryotic water community was dominated by bacillariophytes in April. These seasonal trends correspond with the peak of the summer cyanobacterial bloom and the peak of the spring bacillariophyte bloom observed annually in Esthwaite Water (Maberly *et al*., 2011b). Although bacillariophytes were found to be relatively abundant in the water during the spring, they were only a minor component of the sedDNA community down-core. Furthermore, a large number of bacillariophyte genera identified in a microscopy-based long-term monitoring record of phytoplankton in the surface water were not detected using sedDNA, and this included *Asterionella* and *Fragilaria* which are commonly observed in Esthwaite Water (Chapter 3, Maberly *et al*., 2011b). While the absence of these genera in the sedDNA record could partly be related to limited reference database coverage or primer specificity, these genera were successfully detected in the water using molecular-based methods. Poor deposition of these

genera is unlikely to explain their absence in the sedDNA record because their frustules were identified in a sediment core microfossil record from Esthwaite Water (Dong *et al*., 2012). Instead, DNA degradation may account for the poor detection of some pelagic bacillariophytes in the sedDNA record. A number of bacillariophyte genera were detected in the sedDNA record that were not detected in the water at any depth using molecular (this chapter) or microscopy-based methods (Chapter 3), including *Staurosira*, *Opephora*, *Planothidium* and *Staurosirella*. Phototrophic microbes are unlikely to be active in the sediment at the depth the core was collected from due to limited light availability, and so these genera that were uniquely detected in the sediment may therefore originate from streams or littoral sources.

Cryptophytes were only detected in the sediment with a relative abundance greater than 0.001 in three sediment core sections (Chapter 3). However, cryptophytes were detected in the water across all seasons and depths sampled with a higher relative abundance closer to the lake surface. The successful detection of cryptophytes in the water samples therefore provides further evidence that their poor representation in the sedDNA record may be a consequence of substantial DNA degradation or poor deposition, particularly because their small cells (5-50 μm, Hoef-Emden and Archibald, 2017) may be unable to overcome the barrier of stratification in the water column.

4.5.3. Trends in contribution from the water column community down-core

FEAST-estimated contribution from the bacterial community at a depth of 0-5 m within the water column to the sedDNA record across all months sampled generally declined down-core. Estimated contribution from eukaryotic microbial community at a depth of 0-5 m within the water column in all months sampled except August also declined down-core. These results

suggest that there was an overall loss of DNA likely originating from the pelagic community down-core, and this was most notable for the bacterial community. This could be a consequence of an accumulation of DNA degradation, which may have reduced detection of deposited DNA in older sediments, and for heterotrophic eukaryotes and bacteria, an increase in contribution of modern DNA from the active sediment community relative to the increasingly fragmented deposited DNA with sediment depth. However, a decline in contribution from the water column community could also arise because the water samples collected within one year were not representative of the water column community in the past and may have become more distinct from past communities over time. Previous studies have shown that although some seasonal trends in community composition are observed annually, both bacterial and eukaryotic microbial lake water communities can become more dissimilar over interannual time scales, and community turnover may be particularly great during times of rapid environmental change, such as eutrophication (Korhonen *et al*., 2010; Shade *et al*., 2007; Tammert *et al*., 2015).

4.5.4. Seasonal contribution from the water community to the sedDNA record

Contribution to the sedDNA record from pelagic bacterial and eukaryotic microbial communities near the lake surface could be expected to be lower during the summer stratification period because lake mixing is restricted and the thermocline may act as a physical barrier to deposition (Harrison *et al*., 2019). In August, the depth of the maximum temperature change and therefore an approximation of the thermocline was estimated to be at 6 m (E. Mackay, 2023, pers. comms.). However, contribution from the bacterial and eukaryotic microbial community at a water depth of 0-5 m in August to the sedDNA record was frequently larger compared to contribution from the community in April when stratification was

beginning, in October when stratification had begun to break down, and in February when the lake was fully mixed. Microbial biomass in the water and therefore the amount deposited in the sediment may increase during times of higher productivity (Weisse *et al*., 1990). The formation of cyanobacterial and algal blooms could promote deposition because cells may aggregate in larger biofilms, making them less buoyant and possibly better able to overcome the barrier of stratification and reach the sediment, particularly immediately following the collapse of spring and summer blooms when sinking of organic matter may be greater (Ostrovsky and Yacobi, 2010; Shi *et al*., 2012).

4.5.5. Contribution from the water community with depth to the sedDNA record

Of each depth sampled within the water column in August, the community at a depth of 0-5 m was the largest contributor to the eukaryotic sedDNA record. However, for the bacterial sedDNA record, contribution from the water community at 0-5 m was much lower, and the bacterial community at a depth of 10-15 m within the water column was the largest contributor. Bacterial cells are generally much smaller than many eukaryotic microbes, with typical bacterial cell volumes ranging between 0.4 to 3.0 μm^3 (Levin and Angert *et al*., 2015). Bacteria may not sink as readily as larger eukaryotes, especially during stratification when more energy is required to move between layers of the water column and reach the sediment (Harrison *et al*., 2019). A large proportion of the bacterial community in the water column above the thermocline may therefore not deposit in the sediment, and sedDNA records used to reconstruct past pelagic bacterial communities may be less reliable compared to those of eukaryotic microbes.

Water column DNA may accumulate more DNA damage compared to DNA that is rapidly incorporated within the protective sediment matrix upon cell death (Torti *et al*., 2015). In the oldest sediments, only contribution from the bacterial community at a depth of 10-15 m within the water column was detected. This could be because DNA that originated from deeper within the water column was exposed to fewer DNA degradation processes before it was incorporated within the sediment. The rate of DNA degradation deeper within the water column during stratification may be lower due to anoxia, lower UV radiation and a lower water temperature compared to that closer to the surface (Zulkefli *et al*., 2019). DNA originating from deeper layers of the water column may therefore have been deposited in a relatively undamaged state, allowing it to still be detectable in the oldest sediments.

Although lake sediments were designated as a sink of the water column community, sediment resuspension may also mean that sediments could be considered as a source to the water column. Comte *et al*. (2017) found that resuspension of sediment and the associated sediment community only accounted for a small proportion of the water column bacterial community in two shallow lakes at a depth of 4.2 and 11 m. Sediment resuspension is typically greater in shallow lakes (Evans, 1994), and as the water samples and sediment core were collected from the deepest basin within Esthwaite Water at a depth of 16 m, contamination of water samples due to sediment resuspension could be expected to be relatively small. However, wind-induced currents were found to be sufficient to resuspend small particles at all depths of Esthwaite Water (Mackay *et al*., 2012). Sediment resuspension could therefore make the community detected within the water column, particularly at a depth of 10-15 m which is closer to the water-sediment interface, appear more similar to the sediment community, and account for some of the contribution from the water column to the sediment. Furthermore, resuspension of pelagic taxa that had settled in the sediment could explain why some photosynthetic eukaryotes such

as chlorophytes, bacillariophytes and streptophytes were unexpectedly found to be present with a higher relative abundance in the community at a water column depth of 10-15 m, where light availability is limited, compared to in communities closer to the lake surface.

4.5.6. Contribution from unknown sources

A significant proportion of the sedDNA community could not be attributed to any of the sampled water sources. Contribution from unknown sources was between 91.9 and 99.0% for bacteria and 65.5 and 87.8% for eukaryotic microbes. Zhang *et al*. (2019) used a similar machine-learning MST approach to trace a larger proportion of the bacterial community in lake surface sediments to upstream water and sediment sources and found that contribution from unknown sources was only between 23.8% and 61.4%. However, Zhang *et al*. (2019) and many other studies (Brown *et al*., 2017; Comte *et al*., 2017; Hermans *et al*., 2020; Xu *et al*., 2022) applied machine-learning MST at the OTU level instead of at the finer resolution ASV level. In the present study, bacterial and eukaryotic microbial sequences were also clustered into OTUs based on 97% sequence similarity and the FEAST analysis was repeated. Contribution to the bacterial and eukaryotic microbial sedDNA records that could not be attributed to any source was reduced by up to 10% and 46%, respectively, when FEAST was performed at the OTU level compared to the ASV level. The lower specificity of OTUs may lead to overestimations in source contribution, as well as missing ecological dynamics at the strain or ASV level. However, contribution from unknown sources remained relatively high even at the OTU level. Zhang *et al*. (2019) sampled a larger number of possible sources such as upstream rivers, and the small number of water samples analysed in the present study were unlikely to be representative of the microbial community across a full year, of past pelagic communities, or of the community across the whole lake and its catchment. Additional sources to the sediment

community could include host-associated microbiomes from lake biota such as fish, littoral sediments of the lake from which material may be transported towards the deeper basin due to sediment resuspension and focussing (Mackay *et al*., 2012), water and sediment sources upstream of Esthwaite Water, and runoff from the surrounding catchment. Source detection therefore has the potential to be improved with more comprehensive field surveying spanning multiple seasons and years to better capture the spatial and temporal heterogeneity of microbial communities across the lake and its catchment.

4.5.7. Conclusions

The concentration of 16S rRNA gene copies for the total bacterial community and cyanobacterial community, and contribution from pelagic bacterial and eukaryotic microbial communities to the sedDNA record generally declined down-core. These trends could be a consequence of an accumulation of DNA degradation which may reduce the detection of deposited DNA in older sediments. Although some contribution from the pelagic bacterial and eukaryotic microbial community was still detected in the oldest lake sediments, these results indicate that the resolution and reliability of pelagic community reconstructions may decrease with sample age. This could present a more significant issue for reconstructions of past bacterial communities because contribution from the bacterial community near the lake surface to the sedDNA record was relatively low, indicating that a substantial proportion of the bacterial community did not deposit in the sediment, possibly due to their small size. Further validation of sedDNA as a temporal record is needed for microbial groups that may experience substantial DNA degradation or poor deposition, such as smaller microbes, bacillariophytes and cryptophytes, and those that may be active and abundant within the sediment but are much rarer in the water column, such as Firmicutes and Chloroflexi. Interannual turnover of microbial

communities may also explain the decline in contribution from lake water communities down-core, and future research should therefore aim to better characterise temporal variation in lake microbial communities over longer timescales. This could be achieved by combining long-term molecular-based monitoring of pelagic and benthic microbial communities for comparison with sedDNA records, and can help to further our understanding of the transport and preservation of DNA in lakes and lake sediments and contribute to the development of sedDNA as a reliable palaeolimnological tool.

Chapter 5: Investigating the influence of the bacterial community on 3-hydroxy fatty acid palaeoclimate biomarkers in soils and lake sediments

5.1. Abstract

Palaeoclimatic reconstructions are crucial for understanding past climate change. Widely distributed and well-preserved biomarkers that exhibit consistent structural responses to environmental conditions are essential for reliable climate reconstructions. Proxies based on 3-hydroxy fatty acids (3-OH FAs) produced by gram-negative bacteria have shown potential as palaeoclimate biomarkers in soil, lake and marine environments. However, their use as reliable biomarkers is limited by the lack of knowledge regarding the bacterial producers of these biomolecules. A paired 3-OH FA and bacterial DNA sequencing approach was used to identify possible producers of 3-OH FAs. Soils and lake surface sediments were collected from transects across Eastern and Southern U.S. spanning gradients of mean annual air temperature (MAAT), mean annual precipitation (MAP), and pH. 3-OH FAs and DNA were extracted from each sample, and 16S rRNA gene amplicon sequencing of the bacterial community was carried out. The bacterial phyla Proteobacteria, Acidobacteria and Planctomycetes had higher relative abundance in soils, while Desulfobacteria, Chloroflexi and Cyanobacteria had higher relative abundance in lake sediments. Cross-correlation networks between the bacterial community and 3-OH FAs were computed, revealing that 3-OH FAs may be produced by bacteria specific to each environment, and the shorter-chained C_{10}-C_{14} 3-OH FAs may primarily be produced by lake sediment bacteria. The distinct bacterial communities in soils and lake sediments may respond differently to their environments, possibly explaining why proxies developed for use in soils perform poorly in lakes. 3-OH FA proxies specifically developed based on the responses of lake bacteria may be required for reliable palaeoclimate reconstructions using lake sediment cores. Further studies are required to extend the paired 3-OH FA and bacterial DNA dataset with samples spanning wider environmental gradients and different biomes. A culture-based assessment of 3-OH FAs produced by a wide diversity of bacteria is also required to confirm

whether the positive associations identified between 3-OH FAs and bacterial taxa in environmental samples reflect real producer relationships.

5.2. Introduction

Reconstructions of past climates can improve our understanding of the climate system and provide insights for predictions of future climate change (Yang *et al*., 2020). These reconstructions rely on the analysis of preserved biological materials, such as bacterial membrane lipids, as climate biomarkers (Luo *et al*., 2019). 3-hydroxy fatty acids (3-OH FAs) are one of the many classes of membrane lipids produced by gram-negative bacteria and are involved in homeoviscous adaptation to maintain membrane integrity in response to environmental changes (Siliakus *et al*., 2017). Preserved lipids can be extracted from the environment, and based on their structure, inferences can be made regarding the conditions under which bacteria were exposed to. A range of proxies based on the relative abundance of 3-OH FA isomers have been developed, and when calibrated with modern conditions, these proxies can be applied to sedimentary archives to reconstruct palaeoclimates (Wang *et al*., 2016, 2018). However, identification of the main bacterial producers of 3-OH FAs and an assessment of the influence of environmental factors on their community composition are needed to develop 3-OH FAs as reliable palaeoclimate biomarkers (Huguet *et al*., 2019; Wang *et al*., 2016; Yang *et al*., 2020).

3-OH FAs are found in the outer membrane of gram-negative bacteria with typical chain lengths ranging from C_{10} to C_{18} (Wilkinson, 1988). Different 3-OH FA isomers exist, with *normal* 3-OH FAs containing no methyl branch, *iso* 3-OH FAs containing a methyl branch on the penultimate carbon, and *anteiso* 3-OH FAs containing a methyl branch on the third carbon from

the terminus (Christie and Han, 2012). Bacteria can adjust membrane composition in response to environmental conditions by incorporating differing 3-OH FA isomers with differing carbon chain lengths. For example, a higher proportion of longer, straight-chained or saturated isomers relative to shorter, branched or unsaturated isomers can confer greater membrane rigidity, which is beneficial at high temperatures when the membrane may otherwise become too fluid. These modifications can also reduce membrane permeability to protons, allowing bacteria to maintain an optimal intracellular pH when exposed to lower extracellular pH (Russell and Fukunaga, 1990; Siliakus *et al*., 2017).

3-OH FAs have been extracted from a variety of environments, including soil (Huguet *et al*., 2019; Pei *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016), lake surface sediment (Yang *et al*., 2021), marine sediment (Yang *et al*., 2020), seawater (Wakeham *et al*., 2003), a stalagmite (Wang *et al*., 2018), snow (Tyagi *et al*., 2015), and atmospheric aerosols (Cheng *et al*., 2012). Because of their wide-ranging distribution and stability in the environment, 3-OH FAs have been proposed as palaeoclimate biomarkers, and a number of proxies based on the structural responses to environmental conditions such as temperature and pH have been developed (Wang *et al*., 2016). This includes the ratio of *anteiso* to *normal* 3-OH FAs (RAN index). RAN indices based on C_{15} and C_{17} 3-OH FAs (RAN_{15} and RAN_{17}, respectively) were negatively correlated with mean annual air temperature (MAAT) and positively correlated with mean annual precipitation (MAP) in the soil environment (Wang *et al*., 2016). In lake surface sediments, the ratio of *iso* to *normal* C_{17} 3-OH FAs (RIN_{17}) was negatively correlated with and a better determinant of MAAT compared to the RAN_{15} and RAN_{17} indices applied to soils (Yang *et al*., 2021). The RAN index based on C_{13} 3-OH FAs (RAN_{13}) is also negatively correlated with MAAT in lake sediments (Yang *et al*., 2021), in addition to mean annual sea surface temperature (SST) in marine sediments (Dong *et al*., 2023; Yang *et al*., 2020). The RAN

index based on C_{12}, C_{13} and C_{14} *iso*, *anteiso* and *normal* 3-OH FAs (RAN_S) were also negatively correlated with the mean annual SST (Dong *et al*., 2023). The negative logarithm of the ratio of *iso* and *anteiso* to *normal* 3-OH FAs (RIAN index) was negatively correlated with soil pH (Wang *et al*., 2016), however in the lacustrine environment, much weaker correlations were observed between RIAN and lake water pH (Yang *et al*., 2021).

The RAN and RIAN indices have been calibrated with contemporary MAAT and soil pH (Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016). While the overall relationships (i.e., a negative correlation) between RAN_{15} and RAN_{17} and MAAT, and the RIAN index and pH were generally conserved among a number of regional soil calibrations (Huguet *et al*., 2019; Pei *et al*., 2019; Wang *et al*., 2016), some found no correlation between the RAN indices and MAAT in soils (Véquaud *et al*., 2021), and there was variation in the slope and intercept of the relationships between each regional calibration. Developing a soil-based calibration which can be applied globally has proven to be complex because of this region-specific variation in the trends (Véquaud *et al*., 2021; Wang *et al*., 2021). To develop accurate 3-OH FA proxy calibrations for soils and lake sediments, which may be applied regionally or globally, the possible sources of this variation must be addressed.

Bacterial taxa may produce a unique suite of 3-OH FAs isomers with different carbon chain lengths (Ikemoto *et al*., 1978; Miyagawa *et al*., 1979; Oyaizu and Komagata, 1983; Wollenweber *et al*., 1980), resulting in taxa-specific membrane responses to environmental conditions. Bacterial community composition, which is largely driven by environmental conditions, could therefore be a significant source of variation in 3-OH FA calibrations (Huguet *et al*., 2019). Soil pH is frequently reported to be the dominant driver of bacterial diversity and community composition (Fierer and Jackson, 2006; Griffiths *et al*., 2011; Lauber *et al*., 2009). However, communities may also be structured by temperature, nutrient availability, vegetation

cover, soil type and soil moisture (Fierer *et al*., 2007; Liu *et al*., 2010; Zhou *et al*., 2016). Furthermore, the environmental conditions and composition of the bacterial community in lakes and lake sediments differ substantially from those in soils (Chen *et al*., 2016a; Zwart *et al*., 2002), suggesting that soil-based 3-OH FA proxies may not be suitable for application to lake sedimentary archives used to reconstruct palaeoclimates.

Previous research suggests that specific 3-OH FA proxies are required for application to soil vs lake environments, possibly due to the distinct bacterial communities occupying each environment (Yang *et al*., 2020). However, the main bacterial producers of 3-OH FAs and how their communities may differ between soils and lakes are poorly understood (Yang *et al*., 2020). Recent efforts combining the analysis of bacterial communities using DNA metabarcoding with the analysis of 3-OH FAs suggest that 3-OH FA distributions in soils and lake sediments may be influenced by bacterial community composition (Yang *et al*., 2020, 2021). However, bacterial DNA was extracted from a small subset of samples which is unlikely to sufficiently capture the variation in bacterial communities across environmental gradients. Therefore, paired analysis of bacterial communities and 3-OH FAs in a larger number of soil and lake sediment samples spanning large environmental gradients of temperature, precipitation and pH is needed to identify associations between the environmental drivers, bacterial taxa, and 3-OH FAs in each environment, and to further the development and calibration of 3-OH FAs proxies for use in soil and lake environments.

5.3. Materials and methods

5.3.1. Study sites

Soil samples and lake surface sediments were collected from sites spanning two separate transects in the U.S. (Fig. 5.1). Soil samples were collected in July 2018 from 40 sites within a north to south transect extending from Maine to Florida in Eastern U.S. Paired soils and lake surface sediments were collected in August 2018 from 27 lake catchments within an east to west transect across Arizona, New Mexico and Texas in Southern U.S.

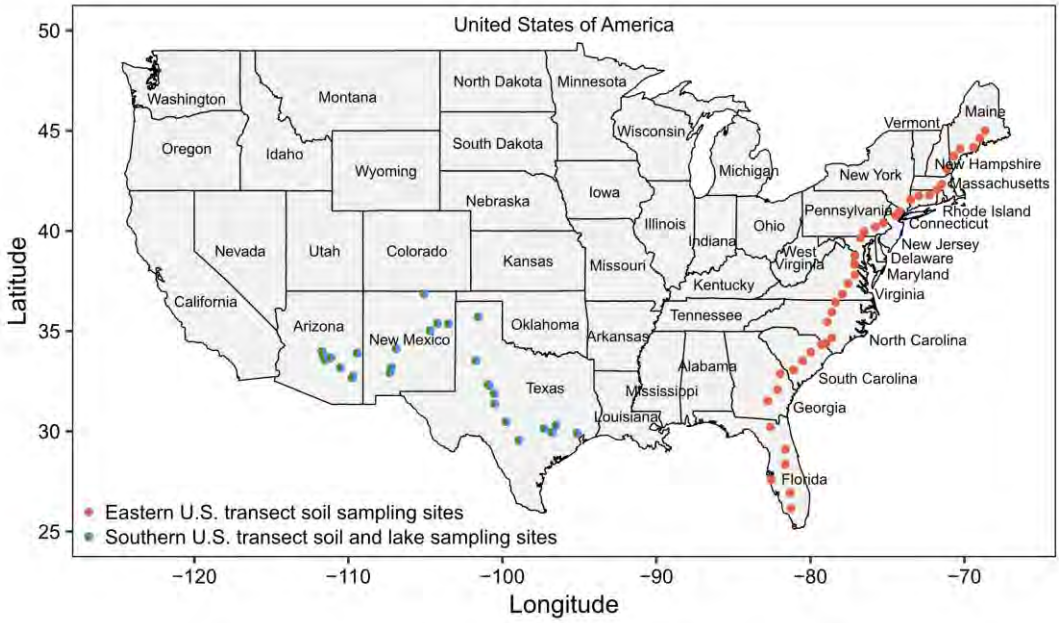

Figure 5.1. Map of the U.S. showing the locations of the soil sampling sites within the Eastern U.S. transect (red) and paired soil and lake surface sediment sampling sites within the Southern U.S. transect (green/blue).

5.3.2. Sample collection

Soil samples were collected from each site within the Eastern and Southern U.S. transects. In the Southern U.S. transect, soil samples were collected in close proximity to the lake shore. Overlying leaf litter was removed, and soil from a mixed depth (0-10 cm) was removed with a trowel and placed on pre-combusted foil. Debris was removed and a spatula was used to transfer a small subsample of well-mixed soil to a Zymo DNA Shield lysis tube (Zymo, CA, U.S.) to

stabilise the DNA. The remaining soil sample was wrapped in foil and stored in a sterile Whirl-Pak bag.

An Ekman grab was used to collect surface sediments from each lake within the Southern U.S. transect. The Ekman grab was deployed from a small boat and the top 15 cm of lake sediment was retrieved. A small subsample of well-mixed sediment was transferred to a lysis tube with a spatula. The remaining sediment sample was wrapped in pre-combusted foil and stored in a sterile Whirl-Pak bag. All equipment used for soil and lake sediment sampling was sterilised with ethanol and thoroughly rinsed with deionised water between uses.

Soil and lake sediment samples were stored in a cool box in the field and during air transport to the UK. Samples for 3-OH FA analysis were homogenised by grinding in a pestle and mortar, sieved and freeze-dried at the University of Birmingham. Subsamples for 16S rRNA gene sequencing of the bacterial community were transported to the UK Centre for Ecology & Hydrology, Wallingford and stored at -20 °C prior to DNA extraction.

5.3.3. Environmental variables

MAAT and MAP at each sample site were obtained from the PRISM database (PRISM Climate Group, 2023), available at: http://www.prism.oregonstate.edu/. MAAT and MAP were averaged over a 30-year period from 1988 to 2018, with a spatial resolution of 4 km^2.

Soil and lake sediment pH were measured using 5 g of each freeze-dried sample. Soil or lake sediment was transferred to a 10 mL Pyrex vial containing 12.5 mL of deionised water to achieve a soil or sediment: water ratio of 1:2.5 (g mL^{-1}). To resuspend the particles, the samples were shaken on a rotary shaker at 150 rpm for 30 min. Before particles settled, a stable pH measurement was taken using the Orion Ross ultra pH/ATC triode and Orion Star pH meter

(Thermo Fisher Scientific, MA, U.S.) calibrated at a pH of 4.01 and 7.01 using buffer solutions (Hanna Instruments, RI, U.S.). The pH probe was thoroughly rinsed with deionised water between measurements. Due to limited sample material, pH was not determined for two Eastern U.S. soil samples.

5.3.4. 3-OH FA extraction

Acid hydrolysis was performed on freeze-dried soil or lake sediment samples. 30 mL of 3 M HCl was added to 10 g of each sample and refluxed at 130 °C for 3 h. The solution was transferred to a furnaced glass centrifuge tube containing 5 mL of DCM, vortexed, and centrifuged at 3000 rpm for 3 min. The supernatant was transferred to a clean furnaced centrifuge tube, 5 mL of DCM was added to the original tube, and the centrifugation step was repeated three times. To obtain the total lipid extract (TLE), DCM was evaporated from the supernatant using a Polyvap (Buchi, UK) at 40 °C, 750 mbar, and 200 rpm. The TLE was cleaned by transferring the sample through a Na_2SO_4 column to remove residual water, and soil or sediment particles. The cleaned TLE was then methylated with 2 mL of 3 M HCl-MeOH and heated at 70 °C for 1.5 h. Hydroxy fatty acid methyl esters (OH FAMEs) were separated from non-OH FAMEs using silica gel column chromatography. Cleaned TLE was adsorbed onto silica gel and first eluted with 1:1 hexane and DCM to yield non-OH FAMEs, and then eluted with 100% ethyl acetate to yield OH FAMEs. The solvent was evaporated from the OH FAME fraction using a N_2 blow down evaporator. OH FAMEs and C_{12}, C_{14} and C_{16} 3-OH FA standards were derivatised with *N*, *O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) at 70 °C for 1.5 h and dried using the N_2 blow down evaporator.

Derivatised OH FAMEs and 3-OH FA standards were diluted in 50 or 400 μL of hexane, depending on OH FAME concentration, and analysed on a gas chromatogram-mass spectrometer (Agilent, CA, U.S.) with a DB-5 fused silica capillary column (length: 60 m, diameter: 0.25 mm, film thickness: 0.25 μm). Helium was used as the carrier gas at a flow rate of 1 mL min^{-1}. The oven temperature was increased from 70 °C to 200 °C at a rate of 10 °C min^{-1}, and then increased at a rate of 3 °C min^{-1} to a 30-min holding temperature of 310 °C. The mass spectrometer ionisation energy was 70 eV.

Gas chromatography-mass spectrometry (GC-MS) data analysis was performed using Agilent ChemStation software v01.11. 3-OH FA trimethylsilyl (TMSi) esters were identified based on their relative retention time and the presence of the diagnostic fragment ion, $[CH_3]_3SiO{=}CHCH_2CO_2CH_3^+$ (175 m/z) resulting from the cleavage between the third and fourth carbon, and the M^+-15 base peak resulting from the cleavage of a methyl group. The mass-to-charge ratio (m/z) of the base peak varies according to the length of the carbon chain. 3-OH FA isomers were identified according to their relative retention time, with isomers eluting in the order: *iso* (*i*), *anteiso* (*a*) and *normal* (*n*) (Wang *et al*., 2021). Peaks corresponding to 3-OH FA TMSi esters were integrated and their relative abundances were determined.

3-OH FA proxies developed by Wang *et al*. (2016), Yang *et al*. (2020, 2021) and Dong *et al*. (2023) were calculated based on the relative abundance of isomers according to the following equations:

$$RAN_{13} = a\text{-}C_{13} / n\text{-}C_{13} \quad \text{(eq. 1)}$$

$$RAN_{15} = a\text{-}C_{15} / n\text{-}C_{15} \quad \text{(eq. 2)}$$

$$RAN_{17} = a\text{-}C_{17} / n\text{-}C_{17} \quad \text{(eq. 3)}$$

$$RAN_S = (i\text{-}C_{12} + a\text{-}C_{13} + i\text{-}C_{14}) / (i\text{-}C_{12} + a\text{-}C_{13} + n\text{-}C_{13} + i\text{-}C_{14} + n\text{-}C_{14}) \quad \text{(eq. 4)}$$

$$RIN_{17} = i\text{-}C_{17} / n\text{-}C_{17} \quad \text{(eq. 5)}$$

$$RIN = I / N \quad \text{(eq. 6)}$$

$$\text{Branched index} = (I + A) / (I + A + N) \quad \text{(eq. 7)}$$

$$\text{Branching ratio} = (I + A) / N \quad \text{(eq. 8)}$$

$$RIAN = -\log(\text{Branching ratio}) \quad \text{(eq. 9)}$$

Where *i*, *a* and *n* represent *iso*, *anteiso* and *normal* 3-OH FA isomers, respectively, and *I*, *A* and *N* represent the sum of all *iso*, *anteiso* and *normal* 3-OH FA isomers, respectively.

5.3.5. DNA extraction, PCR amplification and 16S rRNA gene sequencing

Soil or lake sediment samples contained within the Zymo lysis tubes were defrosted, vortexed, and 500 µL of the soil or lake sediment-lysis buffer solution was transferred into a Qiagen PowerBead tube for DNA extraction using the Qiagen DNeasy PowerSoil kit according to the manufacturer's protocol (Qiagen, Germany). PCR amplification and 16S rRNA gene sequencing have been described in detail in Chapter 2 (Thorpe *et al*., 2022) and Chapter 3. Briefly, the V4 region of the bacterial 16S rRNA gene was amplified using the universal forward and reverse bacterial primers, 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3') (Walters *et al*., 2016). First step PCR conditions were as described in Chapter 2, and reagents were as described in Chapter 3, except that the volumes were halved for 25 µL reactions. The PCR product was purified prior to second step PCR which was performed using a dual-indexing approach (Kozich *et al*., 2013). Samples were normalised, pooled as a single amplicon library and quantified prior to sequencing on the Illumina MiSeq Platform using a 500-cycle v2 MiSeq reagent kit, as described in Chapter 2.

5.3.6. Sequence data processing

Sequence reads were demultiplexed and primers were removed using Cutadapt v4.2 (Martin, 2011). The DADA2 pipeline (Callahan *et al*., 2016) was then implemented in R v4.0.2 (R Core Team, 2020) to truncate reads where the quality score fell below Q30. The quality-filtered forward and reverse reads were merged, and an amplicon sequence variant (ASV) abundance table was generated using the DADA2 error model to identify differences between sequences. Taxonomy was assigned to each ASV using the naive Bayesian classifier (Wang *et al*., 2007) against the SILVA database v.132 (Quast *et al*., 2013). The sequences were rarefied to a uniform sequencing depth of 4,708 reads. ASVs were filtered according to their taxonomy to remove those unidentified at the phylum level.

5.3.7. Data analysis

Pearson's correlations between the 3-OH FA indices and environmental variables for each soil and lake sediment transect were computed and presented as a heat map (Fig. 5.3). No 3-OH FAs were detected in four Eastern U.S. soils, limited 3-OH FAs were detected in two Southern U.S. lake sediments, and two Southern U.S. soil samples and one lake sediment sample displayed broad or tailing chromatogram peaks indicating low quality data. These samples were therefore excluded from the correlations.

ASV abundance was converted to relative abundance and the ASVs were grouped by phylum using the phyloseq R package v1.40.0 (McMurdie and Holmes, 2013). The top 15 most abundant phyla were selected for in-depth analysis of bacterial community composition in soils and lake sediments.

Non-metric multidimensional scaling (NMDS) was performed based on a beta diversity Bray-Curtis dissimilarity matrix of bacterial ASV relative abundance using the vegan R package v2.6-2 (Oksanen *et al*., 2019). A permutation test was used to determine which environmental variables and bacterial phyla were correlated with the dissimilarity matrix. Those that were most strongly ($r > 0.20$) and significantly ($p < 0.001$) correlated, and bacterial phyla that were within the top 15 most abundant groupings were fitted to the ordination space as vectors. NMDS was also performed on a dissimilarity matrix based on 3-OH FA relative abundance and the environmental variables, 3-OH FAs and 3-OH FA indices that correlated with the dissimilarity matrix ($r > 0.20$, $p < 0.001$) were fitted to the ordination space. For both the bacterial and 3-OH FA dissimilarity matrices, analysis of similarity (ANOSIM) with 999 permutations was used to determine the degree of separation between Eastern U.S. soils, Southern U.S. soils and Southern U.S. lake sediments.

Network analysis was performed to identify the associations between bacterial taxa and 3-OH FAs and their indices. Bipartite cross-correlation networks for soils and lake sediments were constructed using the CoNet v1.1.1 (Faust and Raes, 2016) plugin for Cytoscape v3.9.1 (Shannon *et al*., 2003). Pearson's correlations were computed between the relative abundance matrix of ASVs and the relative abundance matrix of 3-OH FAs and their indices, but within-matrix correlations were not computed. A correlation threshold of $r > 0.50$ was applied to the networks, and only positive correlations were displayed. Fisher's Z test was used to compute p values, and only significant ($p < 0.05$) correlations were included. ASVs and 3-OH FAs detected in fewer than five samples were excluded from the networks to reduce the risk of spurious correlations. An attribute circle layout was applied, arranging nodes in clusters according to their strongest edge.

5.4. Results

5.4.1. Environmental variables

MAAT ranged between 6.53 and 23.96 °C within the Eastern U.S. transect, and between 5.19 and 21.22 °C within the Southern U.S. transect. MAP ranged between 1093.45 and 1439.05 mm $year^{-1}$ within the Eastern U.S. transect, and there was a much wider MAP range within the Southern U.S. transect which varied from 245.05 to 1474.04 mm $year^{-1}$. Soil pH ranged between 3.56 and 7.58 within the Eastern U.S. transect, and between 4.25 and 8.77 within the Southern U.S. transect. The pH of lake sediments within the Southern U.S. transect was between 5.76 and 8.80 (Appendix D, Table S1-S3).

5.4.2. 3-OH FA distribution

Soils of the Eastern U.S. transect and soils and lake sediments of the Southern U.S. transect were dominated by the even-chained *normal* 3-OH FAs, *n*-C_{12}, *n*-C_{14}, *n*-C_{16} and *n*-C_{18}, which together comprised more than half of the total 3-OH FAs detected in any soil or lake sediment sample (Fig. 5.2A-C). Soils within the Eastern U.S. transect generally had a higher relative abundance of *n*-C_{10} (up to 0.14) than soils within the Southern U.S. transect (up to 0.03). The relative abundance of *n*-C_{18} was lower in the lake sediments (up to 0.14) than in the soils (up to 0.30). The odd-chained *normal* 3-OH FAs, *n*-C_{11}, *n*-C_{13}, *n*-C_{15} and *n*-C_{17}, were much rarer in soils and lake sediments than the even-chained *normal* isomers, with a combined maximum relative abundance of 0.08 in any sample. *a*-C_{15}, *i*-C_{13}, *i*-C_{15} and *i*-C_{17} were also among the most abundant odd-chained 3-OH FAs found in soils and lake sediments. The relative abundance of *a*-C_{15} was greater in the majority of lake sediment samples (up to 0.12) than in the soils of the Eastern and Southern U.S. transects (up to 0.05). Limited 3-OH FAs were identified in lake

sediments collected from Southern U.S. sites 62 and 63, and included only *n*-C_{10}, *n*-C_{12}, *i*-C_{13}, *n*-C_{14}, *n*-C_{16}, *i*-C_{17}, and *n*-C_{18} (Fig. 5.2C). The relative abundance of unsaturated 3-OH FAs was less than 0.01 in any soil or lake sediment sample (Fig. 5.2A-C).

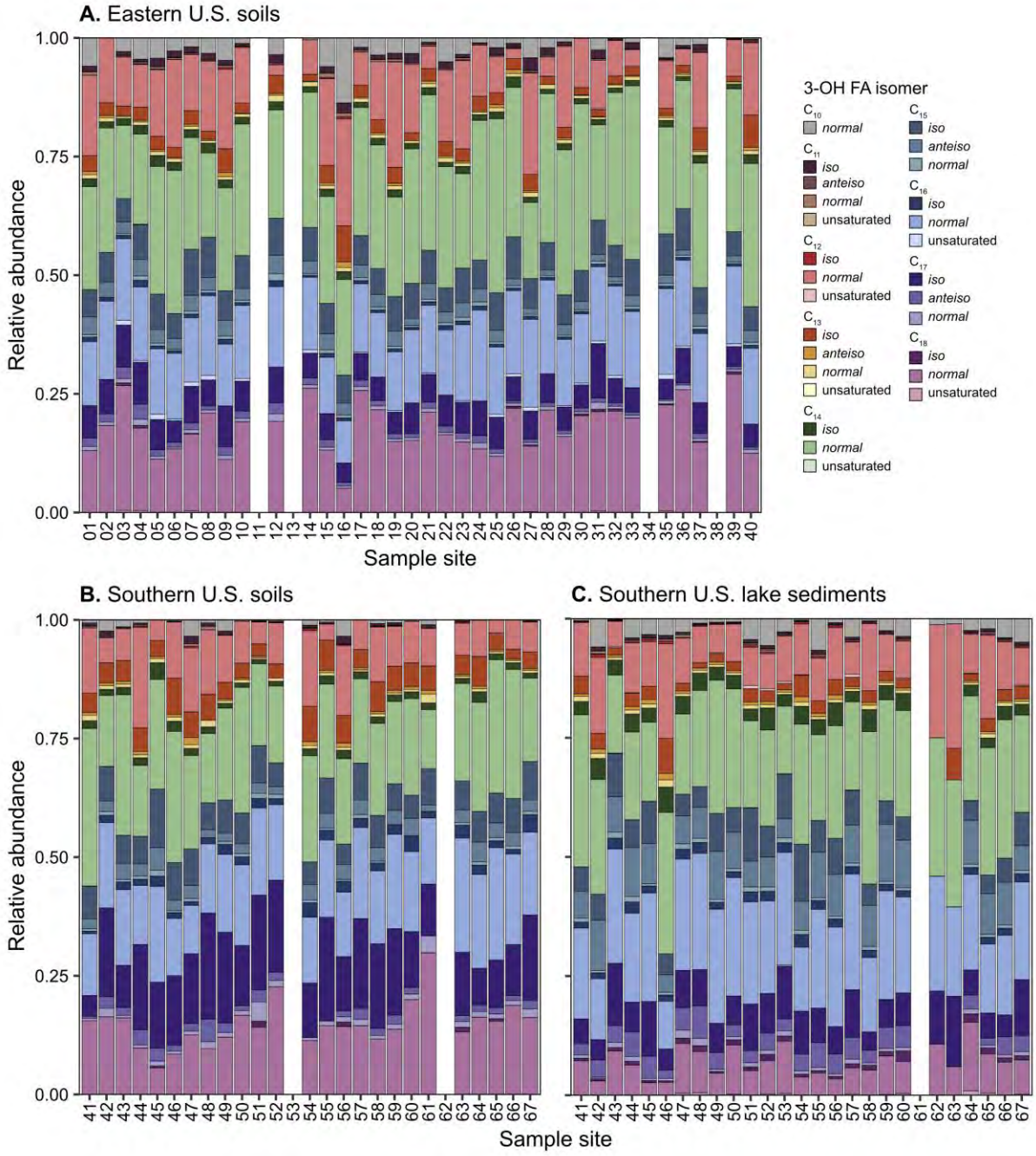

Figure 5.2. Distribution of 3-OH FAs in soils collected within the Eastern U.S. transect (A), and soils (B) and lake surface sediments (C) collected within the Southern U.S. transect. Samples with poor 3-OH FA extraction are not shown.

5.4.3. 3-OH FA index correlations

Of the temperature indices, RAN_{15} and RAN_{17} were negatively correlated with MAAT in soils within the Eastern U.S. ($r = -0.57$ and $r = -0.74$, respectively, $p < 0.001$) and Southern U.S. transects ($r = -0.46$, $p < 0.05$ and $r = -0.58$, $p < 0.01$, respectively) (Fig. 5.3A and B). RAN_S and RIN also had a significant negative correlation with MAAT in Eastern U.S. soils ($r = -0.45$, $p < 0.01$ and $r = -0.35$, $p < 0.05$, respectively), and RIN_{17} had a significant negative correlation with MAAT in Southern U.S. soils ($r = -0.40$, $p < 0.05$). In lake sediments of the Southern U.S. transect, RAN_S, RAN_{13}, and RIN had a significant negative correlation with MAAT ($r = -0.46$, -0.41, -0.51, respectively, $p < 0.01$) and MAP ($r = -0.67$, $p < 0.001$, $r = -0.59$, $p < 0.01$ and $r = -0.48$, $p < 0.01$, respectively) (Fig. 5.3C). RIN, RIN_{17} and RAN_{17} had a significant negative correlation with MAP in Southern U.S. soils ($r = -0.71$, $p < 0.001$, $r = -0.43$, $p < 0.05$ and $r = -0.43$, $p < 0.05$, respectively), but in Eastern U.S. soils, only RAN_{17} had a significant negative correlation with MAP ($r = -0.35$, $p < 0.05$) (Fig. 5.3A and B).

Of the pH indices, the branching ratio and branched index only had a significant positive correlation with pH in the Southern U.S. soils ($r = 0.63$ and 0.67, respectively, $p < 0.001$), but also a significant negative correlation with MAP in soils ($r = -0.70$ and -0.74, respectively, $p < 0.001$) and lake sediments ($r = -0.54$ and $r = -0.57$, respectively, $p < 0.01$) of the Southern U.S. transect (Fig. 5.3B and C). RIAN also had a significant negative correlation with the pH of soils within the Southern U.S. transect ($r = -0.68$, $p < 0.001$) (Fig. 5.3B), but the negative correlation between RIAN and pH was not significant in soils within the Eastern U.S. transect ($r = -0.31$, $p > 0.05$) (Fig. 5.3A) or lake sediments within the Southern U.S. transect ($r = -0.26$, $p > 0.05$) (Fig. 3C). RIAN also had a significant positive correlation with MAP in Southern U.S. soils ($r = 0.76$, $p < 0.001$) (Fig. 5.3B), with MAP and MAAT in lake sediments ($r = 0.59$ and 0.49,

respectively, $p < 0.01$) (Fig. 5.3C), and MAAT in Eastern U.S. soils ($r = 0.50$, $p < 0.01$) (Fig. 5.3A).

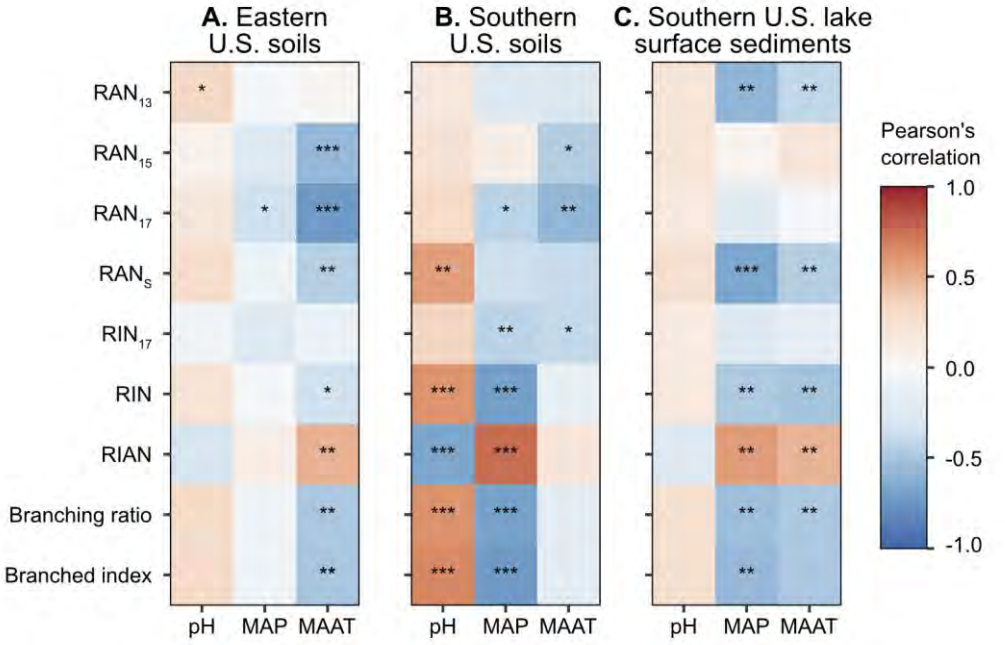

Figure 5.3. Correlations between 3-OH FA indices and environmental variables including soil or lake surface sediment pH, mean annual precipitation (MAP) and mean annual air temperature (MAAT) in Eastern U.S. soils (A), Southern U.S. soils (B) and Southern U.S. lake surface sediments (C). Correlations were calculated using Pearson's correlation, where red shading indicates positive correlations and blue shading indicates negative correlations (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

5.4.4. Bacterial community composition

The bacterial communities in the Eastern and Southern U.S. soils were dominated by the bacterial phyla Proteobacteria and Acidobacteria, which together comprised over half of the community in almost all the samples (Fig. 5.4A and B). The bacterial community in Southern U.S. soils generally had a higher relative abundance of Actinobacteria (up to 0.28) compared to the bacterial community in Eastern U.S. soils where Actinobacteria comprised a smaller proportion (up to 0.13). Other dominant phyla in both soil transects included Bacteroidetes, Planctomycetes and Verrucomicrobia.

Actinobacteria, Bacteroidetes, and Verrucomicrobia were also detected with a comparable relative abundance in lake sediments compared to soils, and the bacterial community could be distinguished from the soil communities with a lower proportion of Proteobacteria, Acidobacteria and Planctomycetes, and a much higher proportion of Desulfobacteria, Chloroflexi and Cyanobacteria (Fig. 5.4C). The relative abundance of Chloroflexi was highly variable between the lake sediment samples, ranging from 0.05 to 0.57. The relative abundance of Cyanobacteria was also variable, ranging from 0.002 to 0.15 for the majority of samples, but the relative abundance of Cyanobacteria in the lake sediment at site 54 was particularly high at 0.34.

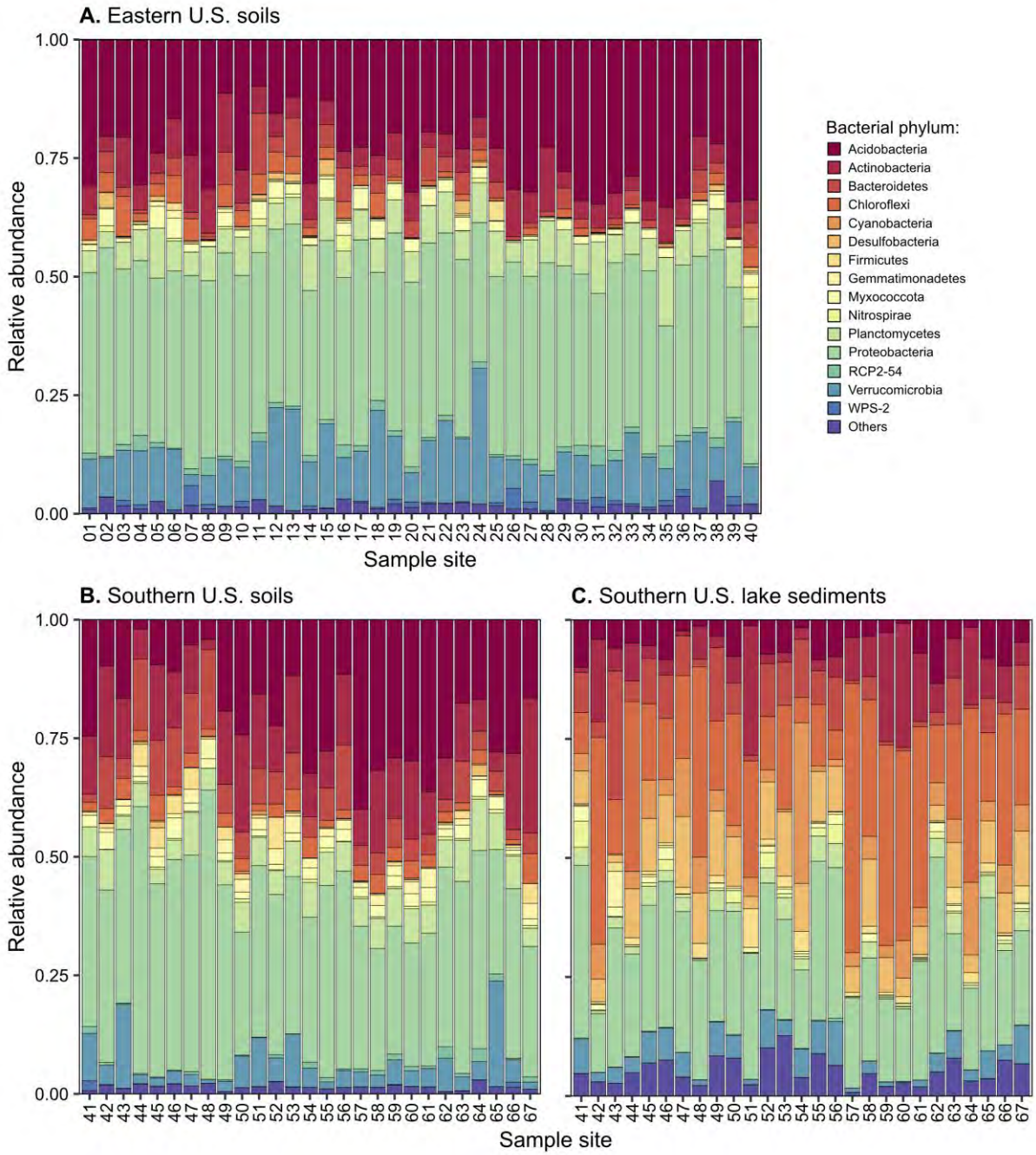

Figure 5.4. Bacterial community composition at the phylum level in soils collected within the Eastern U.S. transect (A), and soils (B) and lake surface sediments (C) collected within the Southern U.S. transect. The top 15 most abundant phyla are shown, and the less abundant phyla are presented as 'others'.

5.4.5. Ordination

The dissimilarity matrices based on bacterial ASVs and 3-OH FAs were most strongly and significantly structured by pH (R^2 = 0.81 and 0.61, respectively, $p < 0.001$) (Fig. 5.5A and B).

Firmicutes, Bacteroidetes and Gemmatimonadetes were positively correlated with a higher pH, whereas Verrucomicrobia, Acidobacteria, RCP2-54 and WPS-2 were positively correlated with a lower pH and higher MAP. Desulfobacteria, Chloroflexi, and Cyanobacteria were associated with lake sediments (Fig. 5.5A).

Unsaturated C_{12} (u-C_{12}), *i*-C_{12}, *a*-C_{13}, *i*-C_{14}, *a*-C_{15}, *a*-C_{17} and *i*-C_{18} were associated with lake sediments, *i*-C_{11}, *n*-C_{14}, u-C_{16} and *n*-C_{18} were associated with Eastern U.S. soils, and *i*-C_{17} and *n*-C_{17} were associated with Southern U.S. soils (Fig. 5.5B). Two lake sediment samples with limited 3-OH FAs were identified as outliers and were therefore removed from the NMDS plot as shown in Appendix D, Fig. S1.

The ANOSIM R value of 0.85 ($p < 0.001$) based on the bacterial ASV dissimilarity matrix indicated a higher degree of separation between the soil and lake sediment transects than that based on the 3-OH FA dissimilarity matrix ($R = 0.63$, $p < 0.001$) (Fig. 5.5A and B). All statistics associated with NMDS are presented in Appendix D, Table S4-S6.

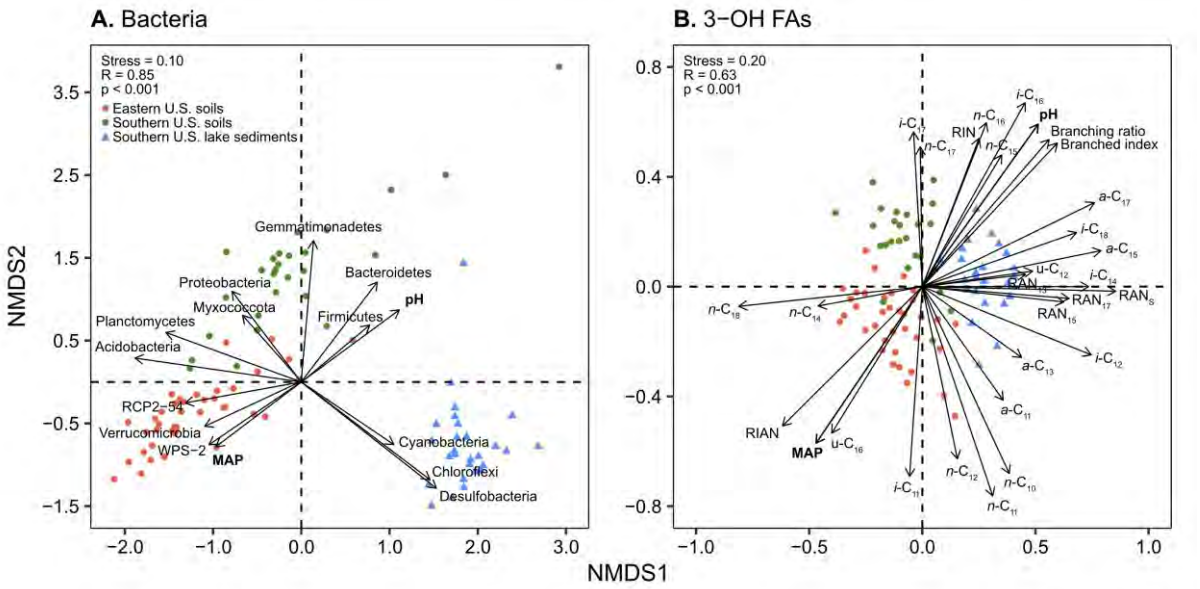

Figure 5.5. NMDS of a dissimilarity matrix based on bacterial ASVs (A) and 3-OH FA distribution (B) in Eastern U.S. soils and Southern U.S. soils and lake surface sediments. Vectors for bacterial phyla that correlated with the bacterial dissimilarity matrix and vectors for 3-OH FAs and their indices that correlated with the 3-OH FA dissimilarity matrix were fitted. The environmental vectors that correlate with each dissimilarity matrix are shown in bold. Only the strongest ($R^2 > 0.20$) and most significant ($p < 0.001$) vectors are displayed. NMDS stress values and ANOSIM R and p values are shown.

5.4.6. Network analysis

The bipartite cross-correlation network of bacterial ASVs and 3-OH FAs and their indices in Eastern and Southern U.S. soils comprised 180 nodes and 205 edges (Fig. 5.6). The network consisted of 11 connected components, the largest of which included *i*-C_{16}, *i*-C_{17}, *n*-C_{17}, *i*-C_{18}, RIN, RIN_{17}, branching ratio and branched index which together connected to 104 ASV nodes via 146 edges. u-C_{16} connected to 13 ASV nodes and u-C_{18} connected to 12 ASV nodes, and together these clusters formed another connected component via ASV-806 (Acidobacteriales, Acidobacteria) and ASV-7369 (Isosphaerales, Planctomycetes) which correlated with both u-C_{16} and u-C_{18}. The smaller components included *a*-C_{11} connected to 15 ASV nodes, *n*-C_{10} connected to eight ASV nodes, RIAN connected to four ASV nodes, and *i*-C_{12} connected to two ASV nodes. Five connected components comprised of only two nodes, including *i*-C_{11} and ASV-2344 (Rhizobiales, Proteobacteria), *i*-C_{13} and ASV-1178 (Pirellulales, Planctomycetes),

i-C_{14} and ASV-6 (Micrococcales, Actinobacteria), RAN_{13} and ASV-571 (Xanthomonadales, Proteobacteria), and RAN_{17} and ASV-450 (Micrococcales, Actinobacteria).

Of the 24,947 bacterial ASVs detected in the soil samples, 162 ASVs were associated with 3-OH FAs in the soil network ($r > 0.50$, $p < 0.05$), of which 36% were assigned to Proteobacteria, 19% to Acidobacteria, 19% to Actinobacteria and 8% to Planctomycetes, while other phyla comprised less than 5% of the ASV nodes (Fig. 5.6). The majority of ASV nodes connected to only one 3-OH FA or 3-OH FA index node. However, a small number of nodes (24 ASVs) were highly connected with between two and five edges. This included ASV-40 (Myxococcales, Proteobacteria), ASV-424 (Thermoanaerobaculales, Acidobacteria), and ASV-4 (Frankiales, Actinobacteria) which correlated with *i*-C_{16}, *i*-C_{17}, RIN, branching ratio and branched index, respectively. 62 ASV nodes correlated with *i*-C_{16} which had the largest degree of association among all the 3-OH FA nodes. Of these, 20 nodes were assigned to Proteobacteria, 14 to Actinobacteria, 12 to Acidobacteria, seven to Chloroflexi, six to Planctomycetes, and one each to Bacteroidetes, Myxococcota and RCP2-54. The strongest correlations in the soil network were observed between ASV-134 (Rhizobiales, Proteobacteria) and *i*-C_{16} ($r = 0.75$), ASV-1222 (unassigned Chloroflexi) and *i*-C_{16} ($r = 0.69$), and ASV-19 (Chthoniobacterales, Verrucomicrobia) and *n*-C_{17} ($r = 0.67$).

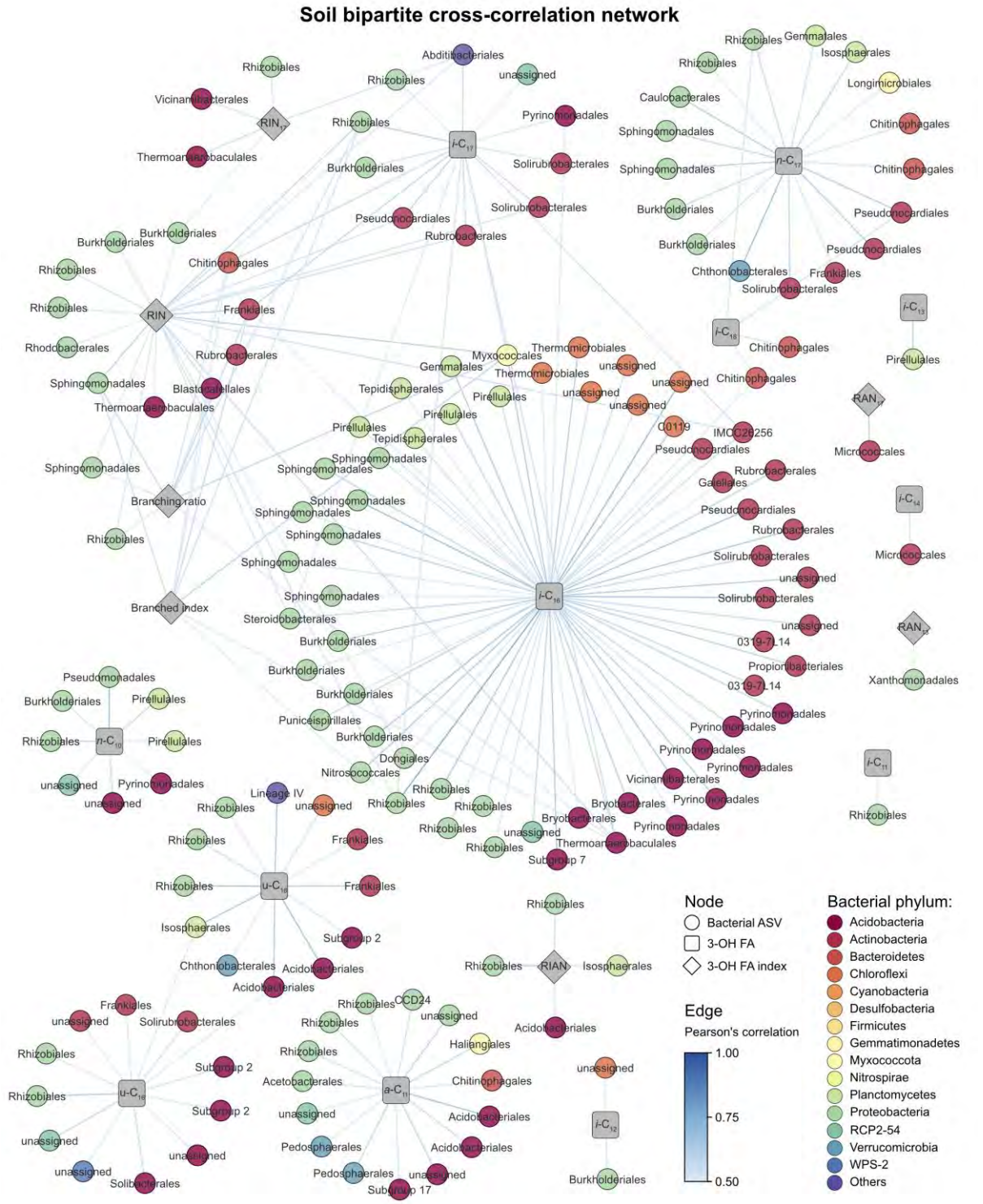

Figure 5.6. Bipartite cross-correlation network of bacterial ASVs, 3-OH FAs and 3-OH FA indices in soils from the Eastern and Southern U.S. transects. Only the strongest ($r > 0.50$) and most significant ($p < 0.05$) Pearson's correlations are shown. Nodes correspond to ASVs (circles), 3-OH FAs (squares) and 3-OH FA indices (diamonds). Edge colour indicates the strength of the correlation. ASV nodes are labelled according to bacterial order and coloured according to bacterial phylum.

The network of bacterial ASVs and 3-OH FAs and their indices in Southern U.S. lake sediments comprised 268 nodes and 389 edges (Fig. 5.7). Nodes and edges were arranged in 4 connected components, the largest of which comprised of 18 3-OH FA nodes (*n*-C_{10}, *n*-C_{11}, *i*-C_{11}, u-C_{11}, *a*-C_{11}, *n*-C_{12}, *i*-C_{12}, u-C_{12}, *n*-C_{13}, *i*-C_{13}, *a*-C_{13}, *i*-C_{14}, u-C_{14}, *i*-C_{15}, *a*-C_{15}, *n*-C_{17}, *i*-C_{17}, *a*-C_{17}), six 3-OH FA index nodes (RAN_{13}, RAN_{15}, RAN_{17}, RAN_S, RIN and RIN_{17}) and 200 ASV nodes connected via 335 edges. Smaller connected components included *i*-C_{18}, u-C_{18} and *n*-C_{18} connected to 33 ASV nodes via 48 edges, RIAN connected to four ASV nodes and *n*-C_{14} connected to two ASV nodes.

Of the 9,278 bacterial ASVs detected in the lake sediment samples, 239 ASVs were associated with 3-OH FAs in the lake sediment network ($r > 0.50$, $p < 0.05$), of which 28% were assigned to Proteobacteria, 18% to Chloroflexi, 13% to Actinobacteria, 9% to Desulfobacteria and 7% to Bacteroidetes, while other phyla comprised less than 5% of the ASV nodes (Fig. 5.7). Only one ASV was found in both soil and lake sediment networks (ASV-124, Rhizobiales, Proteobacteria). The majority of ASV nodes connected to only one 3-OH FA or 3-OH FA index node. 109 ASV nodes were highly connected with between two and five edges. ASV-3080 (Rhizobiales, Proteobacteria) connected to *n*-C_{11}, *i*-C_{11}, *i*-C_{12}, RAN_{15} and RAN_{17}, ASV-5338 (Desulfobacterales, Desulfobacteria) connected to *n*-C_{11}, *i*-C_{11}, *a*-C_{11}, *i*-C_{12} and *a*-C_{13}, ASV-8408 (Rhodobacterales, Proteobacteria) connected to *i*-C_{13}, *a*-C_{13}, RAN_{13}, RIN and RIN_{17}, and ASV-7413 (1013-28-CG33, Proteobacteria) connected to *i*-C_{12}, *i*-C_{13}, *a*-C_{13}, *n*-C_{13} and *i*-C_{14}. RAN_{13} had the highest degree of association and was connected to 60 ASV nodes, 24 of which were assigned to Proteobacteria, seven to Desulfobacteria, seven to Bacteroidetes, four to Verrucomicrobia, four to Nitrospirae, three to Chloroflexi, three to Planctomycetes, two to Acidobacteria and one to each of Actinobacteria, Gemmatimonadetes, Methylomirabilota, Myxococcota, RCP2-54 and Spirochaetes. RIN_{17} was also highly connected and correlated with

40 ASV nodes, 16 of which were assigned to Proteobacteria, five to Desulfobacteria, four to Planctomycetes, three to Chloroflexi, three to Bacteroidetes, three to Verrucomicrobia, two to Acidobacteria and one to each of Actinobacteria, Methylomirabilota, Nitrospirae and RCP2-54. The strongest correlations were observed between ASV-8623 (Phycisphaerales, Planctomycetes) and RIN_{17} ($r = 0.85$), ASV-4276 (Burkholderiales, Proteobacteria) and u-C_{11} ($r = 0.84$), and ASV-1582 (Cellvibrionales, Proteobacteria) and RAN_{13} ($r = 0.82$).

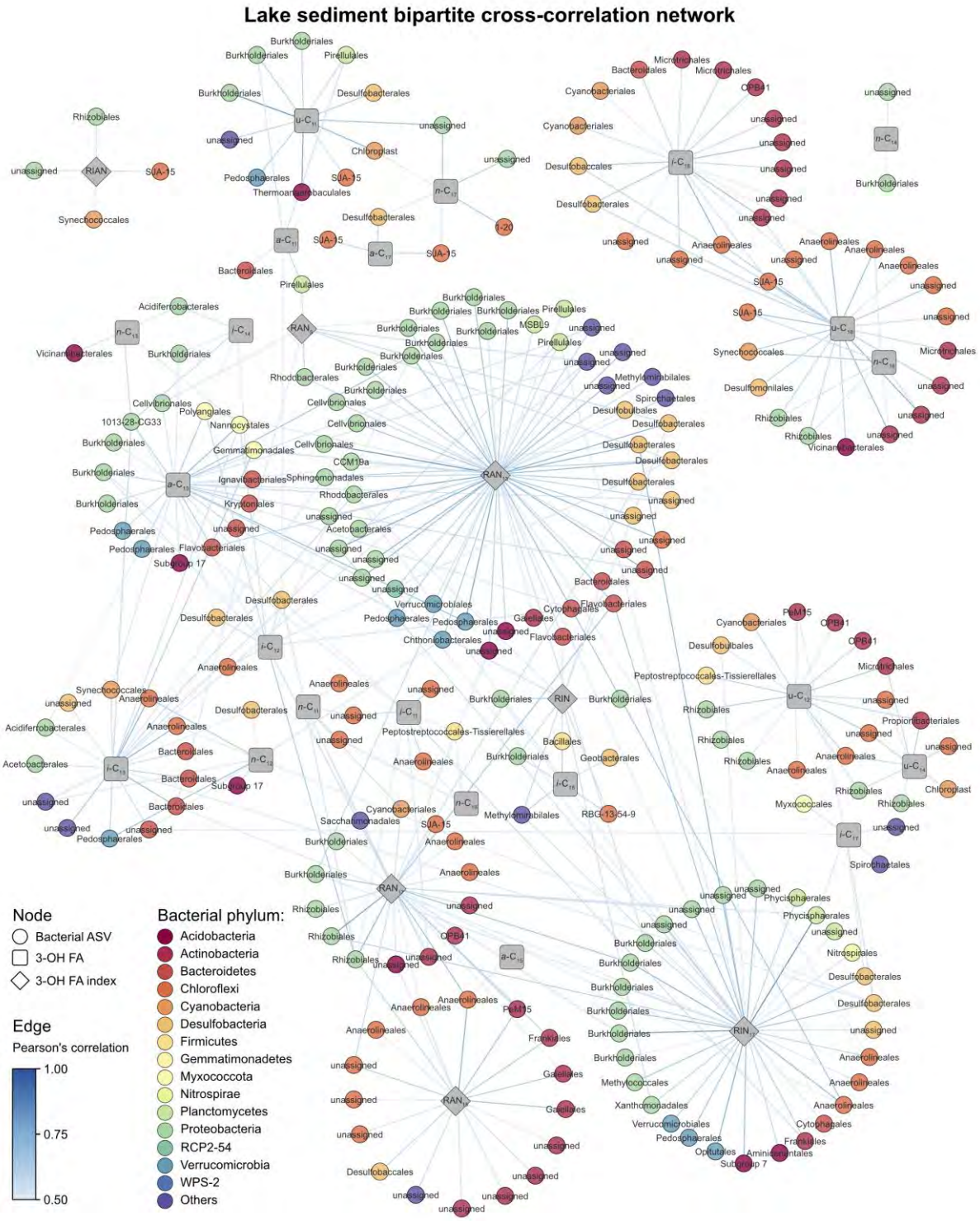

Figure 5.7. Bipartite cross-correlation network of bacterial ASVs, 3-OH FAs and 3-OH FA indices in Southern U.S. lake surface sediments. Only the strongest ($r > 0.50$) and most significant ($p < 0.05$) Pearson's correlations are shown. Nodes correspond to ASVs (circles), 3-OH FAs (squares) and 3-OH FA indices (diamonds). Edge colour indicates the strength of the correlation. ASV nodes are labelled according to bacterial order and coloured according to bacterial phylum.

5.5. Discussion

5.5.1. 3-OH FA distribution and indices

In the soils and lake sediments of the U.S. transects, even-chained *normal* 3-OH FAs were found to be much more abundant than odd-chained 3-OH FAs. This is consistent with previous studies which also reported the dominance of even-chained 3-OH FAs in soils (Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016), lake sediments (Yang *et al*., 2021) and marine sediments (Dong *et al*., 2023; Yang *et al*., 2020). This may suggest that all bacterial producers of 3-OH FAs preferentially produce even-chained 3-OH FAs, or that certain groups of bacteria which may produce even-chained 3-OH FAs are more abundant in soils and sediments than those which may produce more odd-chained 3-OH FAs.

The RAN_{15} and RAN_{17} indices were developed as temperature proxies for use in a soil environment and were found to negatively correlate with MAAT in soils from the Eastern and Southern U.S. transects. This is in agreement with previous studies that covered numerous soil transects in Africa, Asia and Europe (Huguet *et al*., 2019; Wang *et al*., 2016; Wang *et al*., 2021). However, in contrast with these studies, a stronger correlation was observed between MAAT and RAN_{17} ($r = -0.74$, $p < 0.001$ and $r = -0.58$, $p < 0.01$ for the Eastern and Southern U.S. soil transects, respectively) compared to RAN_{15} ($r = -0.57$, $p < 0.001$ and $r = -0.46$, $p < 0.05$ for the Eastern and Southern U.S. soil transects, respectively), and a significant positive correlation was not seen between either RAN index and MAP in the U.S. soil transects. These differences observed in both the 3-OH FA distribution and the index correlations could be a consequence of region-specific variation in the environmental drivers of bacterial activity and community composition between the transects and previous regional studies (Huguet *et al*., 2019; Wang *et al*., 2016; Wang *et al*., 2021).

The RAN_{13} index, originally applied to marine sediments (Yang *et al*., 2020), was found to be negatively correlated with MAAT in Chinese lake sediments (Yang *et al*., 2021) and in Southern U.S. lake sediments of the present study ($r = -0.41$, $p < 0.01$). Similarly, the marine RAN_S index (Dong *et al*., 2023) was also found to be negatively correlated with MAAT in lake sediments in the present study ($r = -0.46$, $p < 0.01$). However, RIN_{17} which was developed for use in lake sediments (Yang *et al*., 2021), was only weakly correlated with MAAT ($r = -0.16$, $p > 0.05$), although the negative correlation between MAAT and RIN was stronger ($r = -0.51$, $p < 0.01$). The negative correlation between MAAT and the marine temperature proxies, RAN_{13} and RAN_S, suggests that they may also be suitable as temperature proxies in lacustrine environments. The applicability of these indices to lake and marine sediments, but not RAN_{15} and RAN_{17}, may indicate that the bacterial producers and their responses to environmental conditions could be distinct between each environment, and perhaps more similar between lake and marine sediments than with soils (Yang *et al*., 2021). RAN_{13}, RAN_S and RIN were also found to negatively correlate with MAP in Southern U.S. lake sediments, and it is possible that bacterial membranes also responded to MAP, although separation of the responses to MAAT and MAP is complex due to co-correlation between these variables.

A negative correlation was observed between RIAN and soil pH in the Eastern and Southern U.S. transects. This agrees with other regional soil calibrations (Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016; Wang *et al*., 2021). However, while the correlation between RIAN and soil pH was relatively strong for soils of the Southern U.S. transect ($r = -0.68$, $p < 0.001$), the correlation was weaker and not significant for soils of the Eastern U.S. transect ($r = -0.31$, $p > 0.05$). This is possibly due to the slightly larger soil pH gradient and much wider MAP gradient within the Southern U.S. transect as precipitation is known to strongly influence soil pH (Slessarev *et al*., 2016). Although not significant, a negative correlation was observed

between RIAN and lake sediment pH ($r = -0.26$, $p > 0.05$). Yang *et al*. (2021) reported a weak negative correlation between lake water pH and RIAN based on 3-OH FAs extracted from lake sediments ($r = -0.09$, $p > 0.05$). The stronger correlation found in the present study between lake sediment pH and RIAN compared to that of lake water pH suggests that the bacterial producers of 3-OH FAs in lake sediments are more responsive to the pH of the sediment than that of the water column.

5.5.2. Bacterial community composition

The bacterial communities in the soils of the Eastern and Southern U.S. transects were dominated by Proteobacteria and Acidobacteria, and to a lesser extent, Actinobacteria, Bacteroidetes, Planctomycetes and Verrucomicrobia. These phyla have been found to be the most abundant bacterial phyla in soils globally (Youssef and Elshahed, 2008). The soil bacterial community was strongly structured by soil pH, which has frequently been found to be the dominant driver of bacterial community composition (Fierer and Jackson, 2006; Griffiths *et al*., 2011; Lauber *et al*., 2009). The bacterial phyla that were negatively correlated with pH included Acidobacteria and Verrucomicrobia, and those that were positively correlated with pH included Bacteroidetes, Firmicutes and Gemmatimonadetes. This is in agreement with previous studies which also reported an abundance of Acidobacteria and Verrucomicrobia in more acidic soils, and an abundance of Bacteroidetes, Firmicutes and Gemmatimonadetes in more alkaline soils (Bartram *et al*., 2014; Jones *et al*., 2009; Lauber *et al*., 2009; Li *et al*., 2021; Wang *et al*., 2019; Yao *et al*., 2017). The composition of the soil community at the phylum level in Eastern and Southern U.S. soils was similar, except that Actinobacteria were generally more abundant in Southern U.S. soils. This could be related to environmental variables such as carbon, nitrogen and phosphorus availability which may have varied between the soil transects and have

previously been shown to strongly influence the relative abundance of Actinobacteria (Dai *et al*., 2018; Liu *et al*., 2017a; Wang *et al*., 2022) and many other bacterial phyla in soils (Fierer *et al*., 2007; Yao *et al*., 2017).

The bacterial community detected in lake sediments was distinct from that in soils, with a lower relative abundance of Proteobacteria, Acidobacteria, and Planctomycetes, and a higher relative abundance of Desulfobacteria, Chloroflexi and Cyanobacteria. Desulfobacteria, Chloroflexi and Cyanobacteria have previously been found to be characteristic of lake sediment bacterial communities compared with soil bacterial communities (Biessy *et al*., 2022; Kuang *et al*., 2022; Li *et al*., 2022; Zhang *et al*., 2020). The relative abundance of Chloroflexi was highly variable between lakes, which could be a response to factors such as nitrogen availability which may have varied between lakes and has previously been shown to strongly influence the relative abundance of Chloroflexi and other phyla in lake sediments (Chen *et al*., 2015; Zhang *et al*., 2015). It is therefore important to note that many other environmental variables which were not determined in the present study may also drive bacterial community composition and consequently 3-OH FA distributions in soils, such as nutrient availability, particle size, soil type and vegetation structure (Hu *et al*., 2014; Zak *et al*., 2003) and in lake sediments, such as trophic status, lake depth and overlying water temperature, pH and dissolved oxygen concentration (Li *et al*., 2019b; Song *et al*., 2022). Furthermore, bacterial taxa may exhibit differential responses to environmental drivers within phyla (Jones *et al*., 2009). The 3-OH FA distribution was also found to be strongly pH dependent. The variations observed in bacterial community composition within and between soil and lake sediment transects in response to conditions such as pH may in turn drive variation in the 3-OH FA distribution.

5.5.3. Associations between possible bacterial producers and 3-OH FAs

Although all saturated C_{10}-C_{18} 3-OH FA isomers detected in lake sediments were also detected in soils, the vast majority of ASVs were specific to each network, with only one ASV found in both networks. The near-distinct communities of bacteria associated with 3-OH FAs in soils and lake sediments indicate that 3-OH FAs may be produced by a community of bacteria specific to each environment. These distinct communities may respond differently to environmental conditions in soils than in lake sediments. For example, it is possible that the producers of C_{15} and C_{17} 3-OH FAs in soils may have stronger responses to temperature than the producers of these 3-OH FAs in lake sediments, leading to a stronger correlation between the RAN_{15} and RAN_{17} indices and MAAT in soils. This could explain why proxies originally developed based on 3-OH FA distributions in soils (RAN_{15} and RAN_{17}) performed poorly in lacustrine environments (Yang *et al*., 2021 and the present study).

RAN_{13} has been found to be more appropriate for lake sediments than RAN_{15} and RAN_{17} in the present study and others (Yang *et al*., 2021). In the soil network, RAN_{13} was only correlated with one ASV, but in the lake sediment network, RAN_{13} had the highest degree of association with all 3-OH FA indices, and a number of ASVs associated with RAN_{13} were assigned to Desulfobacteria and Chloroflexi which were more abundant in lake sediments than in soils and were also associated with *i*-C13 and *a*-C13. This could indicate that C_{13} 3-OH FAs may primarily be produced by a community of bacteria active and responsive to temperature within the lake sediment, and possibly explain why RAN_{13} has proven to be a more applicable temperature proxy in lacustrine environments compared to soils.

In the soil network, C_{16}, C_{17} and C_{18} 3-OH FA isomers were relatively well connected with bacterial ASVs; however, in the lake sediment network, C_{10}, C_{11}, C_{12}, C_{13} and C_{14} 3-OH FA isomers were also well connected with bacterial ASVs, including those assigned to

Desulfobacteria, Chloroflexi and Cyanobacteria. This suggests that more bacterial taxa found in the lake sediment may be capable of producing shorter-chained 3-OH FAs than in the soil community. The RAN_S index originally developed for use in marine sediments is based on the relative abundance of short-chained 3-OH FAs, C_{12}, C_{13} and C_{14} which displayed stronger responses to temperature than longer-chained 3-OH FAs (Dong *et al*., 2023). The associations between these shorter-chained 3-OH FAs and potential bacterial producers in the lake sediment may therefore explain why the RAN_S index performed well as a temperature proxy in Southern U.S. lake sediments compared to catchment soils.

RIAN only correlated with four bacterial ASVs in both networks. This was unexpected because RIAN is calculated from the relative abundance of 3-OH FAs with chain lengths between C_{10} to C_{18}, so a large proportion of the bacterial community could be expected to produce the 3-OH FAs included in the RIAN equation (Wang *et al*., 2021). However, a strong correlation between RIAN and pH was only observed in Southern U.S. soils, and there could be region-specific variation in the responses of the bacterial producers to pH between Eastern and Southern U.S. soils. Future culture-based analyses are needed to better understand bacterial 3-OH FA production in response to pH.

If closely related taxa were responsible for the production of particular 3-OH FA isomers, then associations with an isomer could be expected to be largely dominated by ASVs of the same taxonomic lineage. However, clustering between 3-OH FAs and ASVs according to taxonomy was not apparent, and bacterial ASVs with different taxonomic lineages were correlated with each isomer. This suggests that the 3-OH FAs detected in each environment were produced by a variety of different taxa across multiple taxonomic groups, as opposed to individual taxonomic groups responsible for the production of a specific 3-OH FA isomer. However, bacterial ASVs and 3-OH FAs that vary substantially in their relative abundance across

environmental gradients may be more likely to show strong and significant correlations, many of which may be a consequence of co-correlation due to similar responses, and not necessarily production. While the producers in each environment are likely to be different, as shown by the unique bacterial communities in soils compared to lake sediments, the influence of co-correlation makes identifying positive correlations which represent real producer relationships complex and could mask any taxonomic pattern in 3-OH FA production.

The issue of co-correlation is highlighted by the fact that correlations were detected between 3-OH FAs and bacterial taxa that are not expected to produce 3-OH FAs, such as *Sphingomonas* (Asker *et al*., 2007) which contributed three nodes to the soil network, and gram-positive bacteria including many members of Actinobacteria and Firmicutes. Firmicutes only contributed three nodes to the lake sediment network, whereas Actinobacteria contributed 30 nodes to each of the soil and lake sediment networks. Therefore, these correlations are unlikely to represent producer relationships and may instead be a consequence of co-correlation. However, although studies addressing 3-OH FA production in gram-positive bacteria are limited, some research suggests that production of C_{10}-C_{18} 3-OH FAs may not be exclusive to gram-negative bacteria. For example, C_{10}-C_{18} 3-OH FAs have been identified in biosurfactants produced by several strains of *Bacillus* (Firmicutes) (Besson *et al*., 1992; Youssef *et al*., 2005). Microbial sources of C_{10}-C_{18} 3-OH FAs other than bacteria also need to be considered, as previous studies have revealed that eukaryotic algae may be capable of producing C_{10}-C_{18} 3-OH FAs (Matsumoto and Nagashima, 1984) and fungi may be capable of producing C_{16} and C_{18} 3-OH FAs (Van Dyk *et al*., 1994).

Identifying the bacterial producers of 3-OH FAs in soils and lake sediments is further complicated by the unknown sources of bacteria. For example, 3-OH FAs detected in the lake sediment may have been produced by the bacterial community active within the sediment, but

may have also deposited in the sediment from the water column. Therefore, analysis of the bacterial community and 3-OH FAs in the water column in addition to sediments is important for determining the sources of 3-OH FAs in lakes. Furthermore, it must also be considered that bacteria and their 3-OH FAs may be transported from soils into the lake, making it difficult to identify the in situ producers of 3-OH FAs in lakes.

5.5.4. Conclusions and implications for the use of 3-OH FAs as palaeoclimate biomarkers

More associations were identified between C_{10}-C_{14} 3-OH FAs and the lake sediment bacterial community than with the soil bacterial community. This suggests that these short-chained 3-OH FAs are primarily produced by bacteria active in the lake sediment, and possibly explain why the RAN_{13} and RAN_S indices which are based on the shorter-chained 3-OH FAs performed well in lake sediments compared to RAN_{15} and RAN_{17} which appear to be more appropriate as temperature proxies in a soil environment. Based on these results, soil-based proxies are therefore not recommended for use in lacustrine environments without further validation in a larger number of lakes. This has implications for application of these proxies to lake sediment cores that could be used to reconstruct palaeoclimates.

No clear taxonomic patterns in 3-OH FA production were identified using the cross-correlation network approach. This suggests that diverse bacteria may be capable of producing various 3-OH FA isomers. However, the bacterial producers may not be the same between soils and lake sediments as the 3-OH FAs detected in each environment were found to be associated with highly distinct bacterial communities. Although isolation of soil and lake sediment bacteria for culture-based assessments of 3-OH FA production is needed to confirm this, this suggests that the variation observed in 3-OH FA distribution and in the performance of each 3-OH FA index

in soils and lake sediments may arise due to variations in the bacterial community and their responses to the environment. Therefore, 3-OH FA distributions may be specific to the bacterial community present in a particular environment, and due to the immense diversity of bacteria and strong biogeographic patterns in their community composition (Chu *et al*., 2020; Fierer and Jackson, 2006), this provides an ecological explanation for why regional calibrations were found to be more appropriate compared to a global calibration (Véquaud *et al*., 2021; Wang *et al*., 2021).

While correlation-based analysis is likely limited in its ability to definitively identify bacterial taxa as producers of specific 3-OH FAs, particularly due to the effects of co-correlation, it can be used to identify groups of bacteria which may be associated and therefore be possible producers of certain 3-OH FAs in different environments. In turn, this information can be used to target bacterial taxa for culture-based analysis to directly measure the 3-OH FAs produced. Bacterial groups found to be differentially represented in soils compared to lake sediments, such as Proteobacteria, Acidobacteria and Planctomycetes which were more abundant in soils, and Desulfobacteria, Chloroflexi and Cyanobacteria which were more abundant in lake sediments, may produce different 3-OH FAs and possibly explain the different distribution of 3-OH FAs and performance of proxies in each environment. Therefore, these phyla may be promising targets for culture-based assessments of 3-OH FA production.

The power of this correlation-based approach has the potential to be improved with future additions to the paired 3-OH FA and bacterial DNA metabarcoding datasets, which may extend this analysis to cover wider gradients of environmental conditions and span multiple regions and biomes. Here, we demonstrate the value of a paired 3-OH FA and DNA sequencing approach to explore possible associations between bacterial biomarkers and their producers, which also has significant potential for application to a wide range of palaeoclimate biomarkers,

many of which have uncertain microbial producers, such as the widely used lipid biomarker, glycerol dialkyl glycerol tetraether lipids (GDGTs) (Weijers *et al*., 2007).

Chapter 6: Characterising bacterial 3-hydroxy fatty acid producers for improved palaeoclimatic proxy calibration

6.1. Abstract

3-hydroxy fatty acids (3-OH FAs) produced by gram-negative bacteria have shown promising potential as biomarkers for palaeoclimate reconstruction. Numerous proxies based on the relative abundance of 3-OH FA isomers extracted from soils, lake sediments and marine sediments have been developed to study past temperature and pH. However, their application in palaeoclimatology is currently limited by a poor understanding of the bacterial producers of 3-OH FAs. To identify possible producers, 3-OH FAs were extracted from pure cultures of bacteria isolated from soil and analysed with gas-chromatography mass-spectrometry (GC-MS). 16S rRNA gene sequencing revealed that the isolated bacterial strains were most closely related to *Flavobacterium* (Bacteroidetes) and *Bradyrhizobium*, *Burkholderia*, *Variovax*, *Collimonas*, *Massilia* and *Pseudomonas* (Proteobacteria). These results suggest that members of Bacteroidetes may predominantly produce C_{15}, C_{16} and C_{17} 3-OH FA isomers, and members of Proteobacteria may be among the main producers of even-chained 3-OH FAs. However, these results also showed that 3-OH FA composition can vary within and between taxonomic lineages. Therefore, bacterial community composition may be an important source of variation in 3-OH FA distribution and proxy performance in environmental samples. Even-chained 3-OH FAs are particularly abundant in the environment, appear to be widely produced by bacteria, and are therefore promising candidates for proxy development. Based on these results, we recommend that the possible bacterial sources of 3-OH FAs should be considered in future proxy calibrations and used to inform the development of reliable 3-OH FA proxies for palaeoclimatic reconstruction. However, our understanding of 3-OH FA production is currently limited to readily-cultured bacterial isolates, and further studies are needed to assess the 3-OH FA composition of a wider diversity of bacteria.

6.2. Introduction

Bacterial membrane integrity is highly sensitive to changes in environmental conditions such as temperature and pH. By incorporating lipids with differing carbon chain lengths, degree of branching and degree of unsaturation, bacteria can maintain optimal membrane fluidity in response to environmental change (Russell and Fukunaga, 1990). This response is known as homeoviscous adaptation and can involve many different classes of membrane lipids (Siliakus *et al*., 2017). Among these are 3-hydroxy fatty acids (3-OH FAs) which are found in the outer membrane of gram-negative bacteria with chain lengths ranging between C_{10} and C_{18} (Wilkinson, 1988). These lipids are widely distributed in the environment (Cheng *et al*., 2012; Huguet *et al*., 2019; Tyagi *et al*., 2015; Yang *et al*., 2020, 2021), and as a consequence of their role in homeoviscous adaption, the structure of 3-OH FAs preserved in environmental samples can provide information about the conditions the bacteria that produced them responded to. These characteristics make 3-OH FAs ideal candidates for use as palaeoclimate biomarkers, but their development is currently limited by a poor understanding of the bacterial producers of 3-OH FAs (Huguet *et al*., 2019; Wang *et al*., 2016).

Temperature proxies based on 3-OH FAs include the ratio of *anteiso* to *normal* C_{13}, C_{15} or C_{17} 3-OH FAs (RAN_{13}, RAN_{15} and RAN_{17} indices, respectively) (Wang *et al*., 2016; Yang *et al*., 2020), the RAN index based on *iso*, *anteiso* and *normal* C_{12}, C_{13} and C_{14} 3-OH FAs (RAN_S index) (Dong *et al*., 2023), and the ratio of *iso* to *normal* C_{17} 3-OH FAs (RIN_{17} index) (Yang *et al*., 2021). The negative logarithm of the total *iso* and *anteiso* 3-OH FAs relative to *normal* 3-OH FAs (RIAN index) was developed as a pH proxy (Wang *et al*., 2016). These 3-OH FA indices have been shown to correlate with environmental temperature and pH (Dong *et al*., 2023; Wang *et al*., 2016; Yang *et al*., 2020, 2021). However, some indices appear to be specific to particular environments, as those developed in a soil environment performed poorly in lakes

(Chapter 5; Yang *et al*., 2021), and the slope and intercept of the correlations between these indices and temperature or pH appear to be region-specific (Véquaud *et al*., 2021; Wang *et al*., 2021). Previous studies combining the analysis of 3-OH FAs and 16S rRNA gene amplicon sequencing of the bacterial community have shown that bacterial community composition may drive variation in 3-OH FA distribution across environmental gradients of temperature, precipitation and pH, and 3-OH FAs extracted from soils, lake sediments and marine sediments may be produced by bacteria specific to each environment (Chapter 5; Yang *et al*., 2020, 2021). Therefore, variations in bacterial community composition have been suggested to contribute to the discrepancies observed in 3-OH FA proxy calibrations between different environments and regions (Chapter 5; Wang *et al*., 2021; Yang *et al*., 2020, 2021).

It is possible that bacterial 3-OH FA composition and therefore membrane responses to the environment are taxa-specific (Raetz *et al*., 2007; Simpson and Trend, 2019). However, while the bacterial community is likely to be an important factor influencing 3-OH FA distribution in the environment, a direct examination of the performance of 3-OH FA proxies in pure bacterial cultures has, to our knowledge, only been attempted once. Recently, Hellequin *et al*. (2023) revealed that the relationship between growth temperature and indices based on 3-OH FAs extracted from three isolates of Bacteroidetes was consistent with environmental calibrations, confirming that the correlations observed in the environment are likely to be a consequence of bacterial homeoviscous adaptation. However, this study was based on three closely related isolates, and possible taxonomic differences in 3-OH FA production for a wide diversity of bacteria and the implications for the application of 3-OH FAs as palaeoclimate biomarkers are yet to be shown.

To further the development of 3-OH FAs in palaeoclimatology, a need for the identification of the bacterial producers of 3-OH FAs has been highlighted (Dong *et al*., 2023; Huguet *et al*.,

2019; Wang *et al*., 2016, 2021; Yang *et al*., 2020, 2021). An understanding of how 3-OH FA composition may vary between soil bacteria in particular is needed to explore whether taxa-specific 3-OH FA production has the potential to contribute to the variation observed between soil-based proxy calibrations (Véquaud *et al*., 2021; Wang *et al*., 2021). The aim of this chapter was to isolate and culture bacteria from soil to analyse and compare their 3-OH FA composition. This may help to reveal taxonomic associations in 3-OH FA production, improve our understanding of how the bacterial community may lead to variation in 3-OH FA distribution in environmental samples, and therefore contribute to the development of 3-OH FAs as reliable palaeoclimate biomarkers.

6.3. Materials and methods

6.3.1. Isolation of soil bacteria

Soil was collected range a range of locations (Table 6.1), and 1 g of each soil sample was suspended in 9 ml of autoclaved phosphate-buffered saline (PBS). Serial dilutions up to 10^{-8} were prepared, and 20 µL aliquots of the 10^{-6} and 10^{-8} dilutions were spread onto R2A plates. R2A media consisted of 0.5 g L^{-1} of each of the following components: casein acid hydrolysate, yeast extract, proteose peptone, dextrose and starch, 0.3 g L^{-1} of dipotassium phosphate, 0.3 g L^{-1} of sodium pyruvate, 0.024 g L^{-1} of magnesium sulfate and 15 g L^{-1} of agar. The R2A media had a pH of 7.2. The plates were incubated at room temperature in the dark and monitored periodically for growth. Morphologically distinct colonies were selected from the plates with a sterile 1 µL inoculation loop and streaked onto fresh R2A plates until pure cultures were achieved. To preserve each pure culture, a 1 µL inoculation loop was used to transfer a single

colony from the plate to a sterile cryovial containing a 1:1 glycerol and PBS solution. Glycerol stock cultures were stored at -80 °C until further use.

6.3.2. 16S rRNA gene sequencing of bacterial isolates

For each pure culture, colony PCR (Woodman, 2008) was performed to amplify the full length of the 16S rRNA gene using the forward primer, 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and reverse primer, 1429R (5'-GGTTACCTTGTTACGACTT-3') (Chen *et al*., 2015). A sterile 10 µL pipette tip was used to collect cells from a single colony grown on an R2A plate which were then added to a 50 µL PCR mix. Each reaction mix contained 0.25 µL of 5 units μL^{-1} Taq DNA polymerase (Sigma-Aldrich, UK), 5 µL of 10x PCR buffer (Sigma-Aldrich, UK), 0.5 µL of 20 mg mL^{-1} BSA (New England Biolabs, UK), 1 µL of a 10 mM dNTP mix (Bioline, UK), 43.05 µL of molecular grade water, 0.1 µL of the forward primer and 0.1 µL of the reverse primer. Negative controls with no cells added were included. The PCR program was set to an initial bacterial lysis and denaturing temperature of 95 °C for 5 min, followed by 30 cycles of 95 °C for 15 sec, an annealing temperature of 50 °C for 30 sec, an extension temperature of 72 °C for 1 min, and then a final extension temperature of 72 °C for 10 min. Successful PCR amplification was confirmed with SYBR Safe-stained (Life Technologies Corporation, CA, U.S.) agarose gel electrophoresis. PCR product was purified with the Zymo DNA clean-up kit according to the manufacturer's protocol (Zymo Research, CA, U.S.), resulting in 35 µL ultra-pure DNA eluted in molecular grade water. The DNA concentration was quantified using an Invitrogen Qubit dsDNA HS assay kit with the Qubit 3.0 fluorometer (Thermo Fisher Scientific, MA, U.S).

Purified PCR product was combined with 3.2 pmol of the 27F and 1429R primers and diluted to a DNA concentration of 1-4 ng μL^{-1} with molecular grade water to achieve a final volume of 10 µL in a DNA LoBind Eppendorf tube. Prepared samples were sent to the University of Birmingham Functional Genomics, Proteomics and Metabolomics Facility, where 10 µL terminator dye was added to the PCR product for Sanger sequencing on the Applied Biosystems 3730xl DNA analyser (Applied Biosystems, MA, U.S).

The Seq Scanner v2.0 software (Applied Biosystems, MA, U.S) was used to check the quality score of each forward and reverse read. Each read was then imported into Geneious Prime v2021.0.1 (available at: https://www.geneious.com) to remove primers and trim ends where the error probability exceeded 5%. The forward and reverse reads were aligned and merged and each consensus sequence was imported into NCBI BLAST for taxonomic classification against the BLAST nucleotide collection (BLASTn) database (Altschul *et al*., 1990). The most closely related sequence in the database to each consensus sequence was determined based on percent nucleotide identity.

6.3.3. Cultivation of bacterial isolates for 3-OH FA extraction

Glycerol stocks of bacterial isolates were used to inoculate R2A plates, and colony morphology was examined to ensure a pure culture was retrieved. To obtain sufficient biomass for 3-OH-FA extraction, isolates were grown in liquid culture. A 1 µL loop was used to select a single colony from a plate and inoculate 50 ml R2 broth in 250 ml acid-washed, autoclaved Erlenmeyer flasks sealed with parafilm. Liquid cultures were incubated at 20 °C and 200 rpm in the dark in an Infors HT Multitron incubator (Infors AG, Switzerland). To ensure cultures became acclimated to these conditions and exponential growth was maintained, two sub-

culturing cycles were performed by transferring 500 µL of culture to 50 ml of fresh media every 48 h. With each sub-culturing cycle, an R2A plate was inoculated, and colony morphology was examined to check for contamination. All cultures were handled in a UV-sterilised laminar flow cabinet to minimise contamination risk.

To collect cells for 3-OH-FA extraction, 50 ml of each culture was transferred to a sterile centrifuge tube and centrifuged at 5000 rpm for 10 min. Liquid media and exudates were removed, and the pelleted cells were resuspended in 1 ml PBS in a sterile Eppendorf tube. The cells were washed three times by centrifuging at 15,000 rpm for 2 min, removing PBS and resuspending in fresh PBS. After the final wash, all remaining PBS was removed, taking care not to disturb the pellet. Washed, pelleted cells were stored at -80 °C prior to 3-OH-FA extraction.

6.3.4. 3-OH FA extraction

Pelleted cells of each bacterial isolate were resuspended in 500 µL of sterile deionised water and transferred to a furnaced glass centrifuge tube containing 5 ml 3 M HCl-MeOH. The solution was vortexed prior to heating on a Techne heating block (Cole-Parmer, IL, U.S.) at 70 °C for 14 h for acid hydrolysis and methylation. Total lipid extract (TLE) was then extracted with DCM as described in Chapter 5, but solutions were sonicated in an Elmasonic sonicating bath (Fisherbrand, UK) for 20 min prior to each centrifugation step to aid lysis of cell membranes. TLE was cleaned by transferring the sample through a Na_2SO_4 column. Hydroxy FAMES (OH FAMES) were then separated from non-OH FAMES with silica gel column chromatography and derivatised with *N*, *O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) prior to analysis on a gas chromatogram-mass spectrometer (GC-MS) (Agilent, CA, U.S.) as

described in Chapter 5. An aliquot of R2 broth (500 µL) was subjected to the same extraction protocol and included in the GC-MS run as a negative control.

GC-MS data analysis was performed using the Agilent ChemStation software v01.11. 3-OH FA trimethylsilyl (TMSi) esters were identified as described in detail in Chapter 5. Briefly, *iso* (*i*), *anteiso* (*a*) and *normal* (*n*) isomers of chain lengths between C_{10} and C_{18} were identified based on their relative retention time, the presence of the diagnostic 175 m/z peak and the mass-to-charge ratio (m/z) of the M^+-15 base peak (Wang *et al*., 2021). Due to variability in the instrument's measurement error, the diagnostic 175 m/z peak drifted to 177 m/z for the first of two GC-MS runs. Peaks corresponding to 3-OH FA TMSi esters were integrated and their relative abundances were determined.

6.4. Results

6.4.1. Taxonomic classification of bacterial isolates

According to 16S rRNA gene sequence similarity with bacterial strains in the NCBI BLASTn database (Altschul *et al*., 1990), isolate B1 was most closely related to a strain of *Flavobacterium* sp. (Bacteroidetes) with a nucleotide identity of 99.4% and isolate E2 was most closely related to a strain of *Flavobacterium pectinovorum* with a nucleotide identity of 99.1% (Table 6.1).

Isolate A6 was most closely related to a strain of *Bradyrhizobium* sp. (Proteobacteria, Alphaproteobacteria) with a nucleotide identity of 97.9%. Isolates D7, C9, D9 and F1 were each found to be most closely related to strains belonging to Betaproteobacteria. Isolate D7 was most closely related to a strain of *Burkholderia* sp. with a nucleotide identity of 97.4%, isolate

C9 to a strain of *Variovorax boronicumulans* (96.6%), isolate D9 to a strain of *Collimonas fungivorans* (96.9%), and isolate F1 to a strain of *Massilia* sp. (99.7%).

Isolates D4, G1, B2 and A2 were each found to be most closely related to strains of *Pseudomonas* (Gammaproteobacteria). Isolate D4 was most closely related to *Pseudomonas* sp. strain MBQS34 with a nucleotide identity of 98.3%, isolate G1 to a strain of *Pseudomonas chlororaphis* subsp. *aurantiaca* (99.6%), isolate B2 to *Pseudomonas* sp. strain F-14 (99.9%) and isolate A2 to *Pseudomonas* sp. strain MSM-10-5 (94.5%).

97% sequence similarity is currently accepted as the threshold for species delineation (Stackebrandt and Goebel, 1994). The lower percent nucleotide identity between isolates C9, D9 and A2 and their top matches in the NCBI BLASTn database indicated that these bacterial isolates may be novel strains or species.

Table 6.1. Sources of soil bacterial isolates and their top hit in the NCBI BLASTn database based on percent nucleotide identity.

Bacterial isolate	Source	Top hit	Top hit accession no.	Nucleotide identity (%)
B1	Tenerife, Spain	*Flavobacterium* sp. strain CC8A2	KM187332.1	99.4
E2	Tenerife, Spain	*Flavobacterium pectinovorum* strain E27CS2	MK474994.1	99.1
A6	Oxfordshire, England	*Bradyrhizobium* sp. strain MG-2011-117-FT	FR872447.1	97.9
D7	Anglesey, Wales	*Burkholderia* sp. strain BL23 I1R1	KR154603.1	97.4
C9	Argyll & Bute, Scotland	*Variovorax boronicumulans* strain Port 1	MT825595.1	96.6
D9	Argyll & Bute, Scotland	*Collimonas fungivorans* strain GM300	AB740927.1	96.9
F1	Warwickshire, England	*Massilia* sp. strain HY1-41	OR083862.1	99.7
D4	Warwickshire, England	*Pseudomonas* sp. strain MBWS34. (14)	OP990590.1	98.3
G1	Warwickshire, England	*Pseudomonas chlororaphis* subsp. *aurantiaca* strain KR2-8	MN752851.1	99.6
B2	Warwickshire, England	*Pseudomonas* sp. strain F-14	MG266302.1	99.9
A2	Warwickshire, England	*Pseudomonas* sp. strain MSM-10-5	KY907020.1	94.5

6.4.2. 3-OH FA composition of bacterial isolates

Isolate B1 was found to produce the widest range of 3-OH FA isomers of all bacterial isolates studied (Table 6.2). This included a dominance of *i*-C_{17} with a relative abundance of 0.390 and *i*-C_{15} with a relative abundance of 0.271. Isolate B1 was the only isolate found to produce *iso*, *anteiso* and *normal* isomers of C_{15} and C_{17} 3-OH FAs, although a trace (relative abundance < 0.001) of *i*-C_{17} was detected in isolate G1. *n*-C_{14}, *n*-C_{16} and *a*-C_{18} were the dominant 3-OH FA isomers detected in isolate E2, with a relative abundance of 0.239, 0.454 and 0.180, respectively.

Isolate A6 had the highest relative abundance of *n*-C_{10} and *n*-C_{14} 3-OH FAs with a relative abundance of 0.107 and 0.842, respectively, and was the only isolate found to produce *a*-C_{14} of all isolates studied. Isolate D7 produced a dominance of *n*-C_{14} and *n*-C_{16} 3-OH FAs, with a relative abundance of 0.434 and 0.549, respectively. Isolate C9 also produced a dominance of *n*-C_{16} with a relative abundance of 0.873. *n*-C_{12} was found to be the most relatively abundant 3-OH FA isomer detected in isolates D9, F1, D4, G1, B2 and A2 and comprised more than 50% of all 3-OH FAs detected in each of these isolates except G1. *n*-C_{16} was also detected with a high relative abundance in isolates F1, D4 and G1.

The even-chained *normal* 3-OH FA isomers, *n*-C_{10}, *n*-C_{12} and *n*-C_{18}, were detected in most bacterial isolates studied, and *n*-C_{14} and *n*-C_{16} were detected in all isolates. However, the odd-chained *normal* 3-OH FA isomers, *n*-C_{11}, *n*-C_{13}, *n*-C_{15} and *n*-C_{17}, were detected in a smaller proportion of the isolates with a lower relative abundance (< 0.023) than the even-chained *normal* isomers. *iso* and *anteiso* isomers of C_{10}, C_{11} and C_{12} and *i*-C_{18} 3-OH FAs were not detected in any isolate studied. Chromatograms for each bacterial isolate with peaks corresponding to 3-OH FA isomers identified are presented in Appendix E, Fig S1-11, and representative mass spectra of each isomer detected are presented in Appendix E, Fig. S12-31.

Table 6.2. Relative abundance of 3-OH FA isomers detected in bacterial isolates.

3-OH FA	Bacterial isolate										
	B1	E2	A6	D7	C9	D9	F1	D4	G1	B2	A2
n-C_{10}	-	0.015	0.107	-	0.009	-	0.001	0.008	0.001	0.041	0.001
n-C_{11}	-	-	-	-	-	-	-	0.001	0.001	-	-
n-C_{12}	0.012	0.029	-	0.011	-	0.879	0.554	0.541	0.453	0.861	0.651
i-C_{13}	0.008	-	-	-	-	-	-	-	T	-	-
a-C_{13}	0.003	-	-	-	-	-	-	0.002	-	-	-
n-C_{13}	0.002	-	0.001	0.001	-	0.001	0.003	T	T	-	-
i-C_{14}	0.005	-	-	-	-	-	-	-	-	-	-
a-C_{14}	-	-	0.012	-	-	-	-	-	-	-	-
n-C_{14}	0.015	0.239	0.842	0.434	0.003	0.029	0.006	0.036	0.017	0.006	0.004
i-C_{15}	0.271	-	-	-	-	-	-	-	-	-	-
a-C_{15}	0.042	-	-	-	-	-	-	-	-	-	-
n-C_{15}	0.013	-	-	-	-	-	-	-	-	-	-
i-C_{16}	0.067	-	-	-	0.057	-	-	-	T	-	0.113
a-C_{16}	-	0.044	-	-	-	-	0.040	-	0.001	0.016	0.090
n-C_{16}	0.089	0.452	0.025	0.549	0.873	0.088	0.341	0.400	0.202	0.061	0.059
i-C_{17}	0.390	-	-	-	-	-	-	-	T	-	-
a-C_{17}	0.062	-	-	-	-	-	-	-	-	-	-
n-C_{17}	0.023	-	-	-	-	-	-	-	-	-	-
a-C_{18}	-	0.180	0.013	0.002	0.045	0.001	0.054	0.008	0.319	0.015	0.063
n-C_{18}	-	0.040	T	0.002	0.014	0.002	-	0.003	0.006	-	0.018

i = iso, a = anteiso, n = normal 3-OH FA. T (trace) < 0.001, - = not detected.

6.5. Discussion

6.5.1. Bacterial 3-OH FA producers

Isolates B1 and E2 were each found to be closely related to a different strains of *Flavobacterium*. However, these two bacterial isolates produced a distinct combination of 3-OH FAs. A wide range of 3-OH FA isomers were detected in isolate B1 including odd and even *iso*, *anteiso* and *normal* isomers between C_{12} and C_{17}, and a particularly high relative abundance of *i*-C_{15} and *i*-C_{17}. However, C_{15} and C_{17} isomers were not detected in isolate E2 which was instead characterised by a dominance of *n*-C_{14} and *n*-C_{16} 3-OH FAs. Previous studies have

reported that *i*-C_{15}, *n*-C_{16} and *i*-C_{17} 3-OH FAs are major fatty acids in a number of different species of *Flavobacterium* isolated from soils, freshwater sediment and clinical samples (Lipski *et al*., 1992; Moss and Dees, 1978). *n*-C_{14} 3-OH FAs have also been identified in some species of *Flavobacterium* but it does not appear to be as widely produced compared to C_{15}, C_{16} and C_{17} isomers (Yano *et al*., 1976). Production of C_{14}-C_{17} 3-OH FAs appears to be common among many other members of Bacteroidetes (Bernardet *et al*., 1996; Chen *et al*., 2016b; Fautz *et al*., 1979; Johne *et al*., 1988; Lee *et al*., 2007; Lim *et al*., 2009; Miyagawa *et al*., 1979). However, the detection of C_{10}, C_{12}, C_{13} and C_{18} 3-OH FAs which were detected in one or both isolates of *Flavobacterium* in the present study appear to be rare within Bacteroidetes (Liu *et al*., 2017b; Miyagawa *et al*., 1979).

The 3-OH FA composition of isolates assigned to Alpha, Beta and Gammaproteobacteria were similar with a dominance of even-chained *normal* 3-OH FAs and an absence or low relative abundance (< 0.003) of odd-chained 3-OH FA isomers. However, the most abundant even-chained 3-OH FA isomer differed between isolates. Isolate A6 was most closely related to a strain of *Bradyrhizobium* sp. (Alphaproteobacteria) and produced a dominance of *n*-C_{14}. Isolates D7 and C9 were most closely related to strains of *Burkholderia* sp. and *V. boronicumulans*, respectively (Betaproteobacteria) and produced a dominance of *n*-C_{16}. Isolates D9 and F1 which were most closely related to strains of *C. fungivorans* and *Massilia* sp., respectively (Betaproteobacteria), and isolates D4, G1, B2 and A2 which were each found to be most closely related to different strains of *Pseudomonas* (Gammaproteobacteria) all contained a dominance of *n*-C_{12}. These results are consistent with previous studies which have also reported a dominance of even-chained *normal* 3-OH FAs, particularly C_{12}, C_{14} and C_{16}, in Alphaproteobacteria (Jarvis *et al*., 1996; Lipski *et al*., 1992; Rietschel, 1976; Skerratt *et al*., 1992; Welch, 1991), Betaproteobacteria (Katayama-Fujimura *et al*., 1982; Lipski *et al*., 1992;

Rietschel, 1976; Welch, 1991; Yabuuchi *et al*., 2000) and Gammaproteobacteria (Ikemoto *et al*., 1978; Lipski *et al*., 1992; Parker *et al*., 1982; Rietschel, 1976; Skerratt *et al*., 1992; Welch, 1991). However, there is evidence that some members of Alphaproteobacteria may also be capable of producing odd-chained 3-OH FA isomers including C_{15} and C_{17} (Jarvis *et al*., 1996; Skerratt *et al*., 1992). The presence of *n*-C_{12} 3-OH FAs is remarkably consistent between many different species and strains of *Pseudomonas* isolated from a wide range of environmental and clinical samples (Ikemoto *et al*., 1978; Lipski *et al*., 1992; Rietschel, 1976; Welch *et al*., 1991; Wilkinson *et al*., 1973). However, even-chained 3-OH FAs other than C_{12} may be dominant in other members of Gammaproteobacteria, such as C_{10} which is consistently detected in many species of *Xanthomonas* (Rietschel, 1976), and C_{14} which appears to be characteristic of many genera belonging to Enterobacteriaceae including *Escherichia*, *Salmonella*, *Shigella* and *Proteus* (Rietschel *et al*., 1975).

6.5.2. Implications for 3-OH FAs as palaeoclimate biomarkers

Even-chained *normal* 3-OH FA isomers, particularly *n*-C_{12}, *n*-C_{14} and *n*-C_{16}, were detected with a higher relative abundance than the odd-chained *normal* isomers in the majority of the bacterial strains isolated in the present study. Previous studies have also reported a dominance of these even-chained 3-OH FAs in a number of bacterial groups, including Proteobacteria (Lipski *et al*., 1992; Rietschel, 1976; Welch, 1991), Fusobacteria (Jantzen and Hofstad, 1981; Miyagawa *et al*., 1979; Rietschel, 1976) and Myxococcota (Fautz *et al*., 1979; Rosenfelder *et al*., 1974). These results suggest that many bacterial taxa may predominantly produce even-chained *normal* 3-OH FAs and this could explain why even-chained *normal* 3-OH FAs have been found to dominate soil (Chapter 5; Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016), lake sediment (Chapter 5; Yang *et al*., 2021) and marine sediment samples (Dong *et al*., 2023; Yang

et al., 2020). While a number of different bacterial groups may produce even-chained 3-OH FAs, their abundance in the environment could be largely driven by Proteobacteria which were found to produce mostly even-chained isomers in the present study and previous studies (Ikemoto *et al*., 1978; Lipski *et al*., 1992; Rietschel, 1976; Welch *et al*., 1991; Wilkinson *et al*., 1973), and are also frequently found to be the most abundant phylum in soil (Chapter 5, Fierer *et al*., 2007; Janssen, 2006), lake sediment (Wan *et al*., 2017; Zhang *et al*., 2014) and marine sediment (Li *et al*., 2009; Vipindas *et al*., 2020) bacterial communities.

Isomers of C_{15} and C_{17} 3-OH FAs form the basis of the temperature proxies, RAN_{15}, RAN_{17} (Wang *et al*., 2016) and RIN_{17} (Yang *et al*., 2021). C_{15} and C_{17} isomers were only detected with a relative abundance greater than 0.001 in one isolate in the present study (B1, *Flavobacterium* sp.), and although they were not detected in the other isolate of *Flavobacterium* (E2, *F. pectinovorum*), previous studies have shown that many members of the phylum Bacteroidetes including *Flavobacterium* produce a dominance of C_{15} and C_{17} 3-OH FA isomers (Bernardet *et al*., 1996; Lipski *et al*., 1992; Miyagawa *et al*., 1979; Moss and Dees, 1978; Yano *et al*., 1976). *Flavobacterium* and other Bacteroidetes genera may therefore be among the main producers of C_{15} and C_{17} 3-OH FAs. Bacteroidetes are a dominant bacterial phylum in soils (Chapter 5, Fierer *et al*., 2007), and their relative abundance has been found to be strongly influenced by soil temperature (Oliverio *et al*., 2017). Furthermore, the 3-OH FA composition of Bacteroidetes has been shown to be highly responsive to temperature in laboratory cultures (Hellequin *et al*., 2023). Bacteroidetes could therefore be important drivers of the relationship observed between the RAN_{15}, RAN_{17} and RIN_{17} indices and temperature.

Isomers of C_{13} 3-OH FAs form the basis of RAN_{13} index which is another important temperature proxy with applications in lake and marine surface sediments (Chapter 5; Yang *et al*., 2020, 2021). While C_{13} 3-OH FAs were detected in a number of bacterial isolates including

B1, A6, D7, D9, F1, D4 and G1, the relative abundance of C_{13} isomers was low. Previous studies have shown that C_{13} 3-OH FA isomers may be among the major fatty acids produced by several members of Acidobacteria (Dedysh *et al*., 2012; Eichorst *et al*., 2007; Losey *et al*., 2013; Pankratov *et al*., 2012). Acidobacteria are one of the dominant bacterial phyla in soils (Chapter 5; Janssen, 2006; Kalam *et al*., 2020), lake sediments (Huang *et al*., 2019) and marine sediments (Du *et al*., 2011), and could therefore be one of the main contributors to the relationship observed between RAN_{13} and temperature. However, production of C_{13} 3-OH FAs does not appear to be consistent among all Acidobacteria isolates (Dedysh *et al*., 2012; Pankratov *et al*., 2012). Acidobacteria are notoriously difficult to culture in the laboratory (Kielak *et al*., 2016), but further study of the 3-OH FA composition of a wider diversity of Acidobacteria is needed to elucidate the role of these bacteria in environmental 3-OH FA signals.

The importance of the bacterial community in determining 3-OH FA distribution and proxy performance in environmental samples has been demonstrated (Chapter 5; Yang *et al*., 2020, 2021). Here we show that 3-OH FA composition can vary among bacterial isolates, and that the bacterial sources of 3-OH FA isomers should therefore be taken into consideration when developing reliable 3-OH FA proxies. The ubiquitous nature of biomarkers is important to allow comparisons between proxy calibrations performed in different environments and regions to be made. Even-chained 3-OH FAs are much more abundant in the environment than odd-chained isomers (Chapter 5; Dong *et al*., 2023; Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016; Yang *et al*., 2020, 2021) and may be produced by a wide range of bacterial groups that dominate bacterial communities. The ratio of even-chained *normal*, *iso* and *anteiso* 3-OH FA isomers may therefore be promising candidates for development as proxies which can be applied to a wide variety of environments and regions. However, C_{15} and C_{17} 3-OH FAs may

be predominantly produced by Bacteroidetes, and the more restricted bacterial source compared to many of the even-chained 3-OH FAs may lead to more consistent relationships between environmental conditions and proxies based on these isomers, such as RAN_{15}, RAN_{17} and RIN_{17} indices.

6.5.3. Conclusions

These results and those of previous studies indicate that different bacterial taxonomic lineages may produce a characteristic suite of 3-OH FAs, with members of Bacteroidetes predominantly producing C_{15}, C_{16} and C_{17} 3-OH FA isomers and members of Proteobacteria predominantly producing even-chained *normal* 3-OH FAs. However, each 3-OH FA isomer is unlikely to be unique to a single taxonomic lineage as many 3-OH FAs such as *n*-C_{14} and *n*-C_{16} were detected in isolates of Bacteroidetes and multiple classes of Proteobacteria. This variation in 3-OH FA production within and between taxonomic lineages may explain the trends in 3-OH FA distribution and discrepancies in proxy performance observed in different environments and regions. Therefore, it is recommended that bacterial community composition should be accounted for in future 3-OH FA proxy calibrations and should also be considered when developing reliable 3-OH FA proxies.

Culture-based analysis of bacterial 3-OH FA production is biased towards readily cultured or well-studied isolates, and our understanding of bacterial 3-OH FA production is therefore largely based on members of Bacteroidetes and Proteobacteria which are readily cultured in the laboratory (Lewis *et al*., 2020). Future efforts should focus on isolating a wide diversity of bacteria including those from underrepresented and difficult to culture groups to determine their 3-OH FA composition such as Acidobacteria. Numerous studies focusing on characterising

novel bacterial strains have determined their membrane lipid composition, including their ability to produce 3-OH FAs (Chen *et al*., 2015; Králová *et al*., 2018; Lee *et al*., 2007; Lim *et al*., 2009; Liu *et al*., 2017b; Yabuuchi *et al*., 2000). A comprehensive and systematic review of these studies may be particularly useful for identifying taxonomic patterns in 3-OH FA production. Furthermore, genomic studies of 3-OH FA biosynthesis pathways may allow the 3-OH FA composition of unculturable bacteria to be predicted based on metagenomic data. Further research is also needed to understand how 3-OH FA production by a wider diversity of bacteria varies in response to conditions such as temperature and pH, as previously shown by Hellequin *et al*. (2023) for three closely related isolates of Bacteroidetes, in addition to conditions such as nutrient concentrations which may also influence 3-OH FA production. 3-OH FA proxy calibrations obtained from laboratory cultures could be compared with environmental calibrations to improve our understanding of which bacterial taxa are the main drivers responsible for the relationship observed between these proxies and environmental conditions.

Chapter 7: General Discussion

7.1. Is sedDNA an effective approach to reconstruct microbial communities?

Lake sedDNA (sedimentary DNA) is a promising approach to reconstruct past microbial communities, but its application in palaeolimnology is limited by a number of uncertainties concerning the reliability of community reconstructions, and the source and fate of DNA in lakes and lake sediments (Capo *et al*., 2021; 2022). The ability of sedDNA to provide a palaeolimnological record of past bacterial and phytoplankton communities was assessed in Chapters 2 and 3, respectively, and deposition of bacterial and eukaryotic microbial DNA to the sediment and possible evidence of DNA degradation were explored in Chapter 4. The main findings of each chapter are discussed below, in addition to the outstanding questions that require further study to enable the development of microbial sedDNA as a reliable palaeolimnological tool.

7.1.1. sedDNA as a record of past bacterial communities

Bacterial sedDNA has significant potential in palaeolimnology because bacteria are often neglected in microscopy-based monitoring schemes and by traditional palaeolimnological tools due to difficulties in identification by morphology, their small size, variable deposition, and an absence of well-preserved and taxonomically distinct remains (Capo *et al*., 2021). However, sedDNA reconstructions of the lake bacterial community are uncertain due to the possibility of overprinting of the temporal signal by active heterotrophic sediment communities (Thupaki *et al*., 2013; Vuillemin *et al*., 2017), and previous attempts to reconstruct past bacterial communities using sedDNA are rare (Li *et al*., 2019a).

In Chapter 2 (Thorpe *et al*., 2022), we reconstructed over 100 years of lake bacterial community change, showing that bacterial phyla displayed distinct trends in their relative abundance,

possibly in response to eutrophication. The cyanobacterial temporal trend was validated by comparable trends observed in the long-term microscopy-based monitoring record of cyanobacterial richness in the surface water, and known pelagic bacteria were successfully detected in the sediment. Temporal trends in the relative abundance of the dominant bacterial phyla reconstructed using sedDNA were also shown to be similar between replicate sediment cores collected from Esthwaite Water on the same date (Chapter 2, Thorpe *et al*., 2022) and several years later (Appendix B, Fig S1 and Chapter 4), demonstrating the robustness of the sedDNA approach to reconstruct past bacterial community change. These results indicated that sedDNA can be a useful tool to reconstruct past bacterial community change. However, bacterial sedDNA records may be confounded by contribution from in situ sediment bacterial communities. To ensure the reliability of sedDNA, clear distinctions must therefore be made to separate relic DNA of past bacterial communities from modern DNA of the active bacterial community at depth within the sediment.

7.1.2. sedDNA as a record of past phytoplankton communities

sedDNA reconstructions of past phytoplankton communities are more common in palaeolimnology compared to bacterial community reconstructions (Capo *et al*., 2016; Huo *et al*., 2022; Ibrahim *et al*., 2020; Zhang *et al*., 2021b). When studying past communities of photosynthetic microbes in deep lakes, there is more certainty that the fraction of sedDNA originating from active sediment communities is minor in comparison to that which may have been deposited from the water column over time. However, differential DNA degradation and deposition may influence the reliability of the reconstructed temporal trends, and validation of the sedDNA record with concurrent monitoring is therefore needed.

In Chapter 3, it was found that the phytoplankton sedDNA record was generally supported by long-term microscopy-based monitoring of phytoplankton in the surface water, with broadly comparable trends observed in the relative abundance and occurrence of the dominant phytoplankton phyla over the 65-year period shared between the sedDNA and microscopy-based records. Up to 20% of genera were successfully captured using both methods, and sedDNA also proved to be effective at capturing phytoplankton genera previously overlooked using microscopy. However, our comparisons between sedDNA and microscopy-based monitoring revealed an off-set in the timing of the onset of community change, possibly due to taphonomic processes or dating inaccuracies of the sedDNA record, and identified a number of phytoplankton taxa which were poorly represented with sedDNA, including cryptophytes and a substantial number of bacillariophyte genera. Before sedDNA can reliably be used to reconstruct past microbial communities, further research is needed to investigate the possible reasons why certain taxa were underrepresented, which may be related to DNA degradation, poor deposition, limited coverage in reference databases or poor amplification with the chosen broad-range amplicon primers.

7.1.3. Tracing microbial sedDNA from the water column to the sediment

The results of Chapter 2 (Thorpe *et al*., 2022) and Chapter 3 revealed that sedDNA does have significant potential to be used as a record of past bacterial and phytoplankton community change but highlighted that the transport and preservation of DNA in lakes and lake sediments and the influence of these processes on the reliability of the sedDNA record require further study. To address this need in Chapter 4, a microbial source tracking (MST) approach was applied to the bacterial and eukaryotic microbial sedDNA records to trace transport of DNA likely originating from the water column to the sediment on a seasonal scale. The MST

approach revealed that a substantial proportion of deposited bacterial and eukaryotic microbial sedDNA may derive from the water community in the summer, possibly because productivity and community turnover are high, and aggregations of cells and organic matter during algal and cyanobacterial blooms may promote sinking of cells and DNA. sedDNA reconstructions may therefore be biased towards past microbial communities that were present in the water column during periods of greater productivity. Contribution from the bacterial near-surface water community to the sedDNA record was much lower compared to that from deeper within the water column and from the eukaryotic water community, indicating that due to their small size, bacteria may not easily overcome the barrier of stratification. Furthermore, a number of dominant bacterial phyla detected in the sediment were found to be relatively rare in the water column. The fraction of bacterial sedDNA that was deposited from the water column may therefore be relatively small, and bacterial sedDNA reconstructions may only represent temporal responses of a subset of the community.

Quantitative PCR (qPCR) was performed to understand how the detection and therefore preservation of DNA may change down-core. The concentration of bacterial and cyanobacterial 16S rRNA gene copies was found to decline with sediment depth. The results of qPCR and MST suggest that the detectability of deposited DNA may diminish over time, which could be a consequence of DNA degradation. The relatively low and stable relative abundance of cyanobacteria (Chapter 2) and phytoplankton (Chapter 3) observed in older sediments could also be interpreted as further evidence of DNA degradation. Taken together, these results indicate that at least some DNA degradation is likely to have occurred down-core, and further research is needed to determine the extent of this DNA degradation for different taxa. However, the MST approach was limited by the relatively small number of sources that were sampled within one year. Due to interannual turnover of microbial communities, the water samples were

unlikely to be representative of past water column communities. In older sediments, DNA that was deposited from past water column communities may therefore not have been detected as originating from the water column and could partly explain the decline in estimated contribution down-core. Concurrent analyses of water column and deposited sediment microbial communities over time periods spanning multiple years are needed to further investigate how DNA is transported and preserved in the sediment.

7.1.4. Summary

These results suggest that sedDNA can be used to reconstruct past bacterial and eukaryotic microbial communities over a period of up to 100 years, and it was shown that:

(i) sedDNA is capable of capturing some pelagic history of bacterial and phytoplankton communities, possibly in response to environmental drivers such as eutrophication,

(ii) these reconstructions are broadly comparable with the trends observed in the long-term microscopy-based monitoring records of cyanobacteria and eukaryotic phytoplankton,

(iii) deposition of DNA to the sediment may vary seasonally, according to depth within the water column, and between taxa,

(iv) and the detection of deposited DNA in older sediments may be reduced, possibly due to an accumulation of DNA degradation.

To improve the reliability of sedDNA reconstructions of past microbial communities, our results highlight that there is a need to better characterise the depositional and taphonomic processes that may act differently on certain taxa and lead to biases in palaeolimnological

reconstructions, and to further distinguish deposited, relic DNA from that which originates from active communities within the sediment.

7.1.5. Recommendations for further study of sedDNA in palaeolimnology

7.1.5.1. Source and fate of sedDNA

To better characterise the depositional processes sedDNA is subject to, detailed studies of the microbial communities residing in lake water and sediments and how their communities change over longer periods of time are needed. This could improve our understanding of the connectivity between water and sediment communities and how cells and DNA may be transported between them. These studies may benefit from the use of sediment traps which allow collection of deposited DNA separate from that which originates from in situ sediment communities (Gauthier *et al*., 2021; Nwosu *et al*., 2021). Furthermore, seasonally laminated sediment cores are rare, but seasonal analysis of the material deposited within the sediment traps may allow the detection of more detailed intraannual trends in deposition (Apolinarska *et al*., 2020).

Microcosm experiments have proven to be useful to investigate the rate of DNA degradation. The concentration of DNA can be monitored over time, and the influence of different environmental conditions on the rate of DNA degradation can be investigated by manipulating incubation conditions (Mejbel *et al*., 2022). Previous studies have focused on the degradation of cyanobacterial DNA (Mejbel *et al*., 2022), but further studies investigating the extent of DNA degradation among different microbial taxa are needed.

7.1.5.2. Separating relic and modern sedDNA

To further distinguish between the fraction of sedDNA that was deposited from the water column over time and that which may have originated from active sediment communities, comparisons of extracellular and intracellular DNA or DNA and RNA down-core may be useful. Extracellular DNA may largely be derived from lysed dead cells while intracellular DNA may represent intact live cells (Torti *et al*., 2015). Extracellular DNA could be separated from intracellular DNA by avoiding cell lysis during DNA extraction (Gauthier *et al*., 2022) or with the use of DNA binding dyes that cannot permeate intact cells and so inhibit PCR amplification of extracellular DNA only, such as propidium monoazide (Ramírez *et al*., 2018). Previous studies have demonstrated that reconstructed trends of the eukaryotic microbial community based on intracellular and extracellular sedDNA differ, and it has been suggested that they may represent modern and relic sedDNA, respectively (Gauthier *et al*., 2022). While the total pool of sedDNA may include both modern and relic DNA, sedimentary RNA (sedRNA) could be assumed to originate only from the actively metabolising community due to its instability in the environment (Pearman *et al*., 2021). DNA and RNA can be independently extracted from sediments by following different extraction protocols, and taxa poorly represented with sedRNA but well represented with sedDNA may indicate that they are not active within the sediment (Pearman *et al*., 2021). Analyses of sedRNA or intracellular DNA can therefore give an indication of contribution from active communities to the sedimentary record. However, intracellular DNA may also originate from dormant cells or robust dead cells and extracellular DNA may be actively released by live cells, and relic RNA may exist if stabilised within the sediment matrix (Torti *et al*., 2015). These techniques have not been widely applied to lake sediment cores, and their ability to separate relic and active microbial communities therefore needs further assessment.

7.1.5.3. Potential of metabarcoding in long-term monitoring schemes

Validation of the sedDNA record in Esthwaite Water was made possible by the ongoing lake monitoring scheme run by the UK Centre for Ecology & Hydrology. Our comparisons between the sedDNA and microscopy-based records revealed that while a proportion of the phytoplankton community was captured using both methods, sedDNA and microscopy were each capable of detecting a unique community of phytoplankton not detected by the other method. Molecular methods are typically considered to be more sensitive and can distinguish microbial taxa at the sequence level (MacKeigan *et al*., 2022), and reference sequence database coverage is continually improving (Glöckner *et al*., 2017). Combining metabarcoding with microscopy in long-term monitoring schemes therefore has significant potential to broaden their scope to study the seasonal and interannual temporal dynamics of the wider lake community beyond phytoplankton, which have been a focus of monitoring schemes in the English Lake District and globally (Burlakova *et al*., 2018; Hampton *et al*., 2008), to also include lake biota such as bacteria and improve the detection of taxa that may be difficult to identify with microscopy alone. In turn, the use of metabarcoding in long-term monitoring schemes may enable further validation of sedDNA records, particularly of bacterial sedDNA records which remain uncertain due to contribution from active sediment communities.

7.2. Exploring the bacterial producers of 3-OH FAs

Proxies based on bacterial 3-hydroxy fatty acids (3-OH FAs) have been proposed as novel palaeoclimate biomarkers. However, the main bacterial producers of 3-OH FAs in different environments are yet to be confirmed, and this information is needed to further their development as biomarkers for reliable palaeoclimate reconstructions. To address this need,

Chapter 5 explored how bacterial community composition may influence variation in 3-OH FA distribution and the performance of proxy calibrations in soil and lake surface sediments. Chapter 6 focussed on identifying possible bacterial producers of 3-OH FA isomers in pure culture. The main findings of these chapters are summarised below, and the implications and recommendations for the use of 3-OH FAs as palaeoclimate biomarkers are discussed.

7.2.1. The relationship between the bacterial community and 3-OH FAs

Comparisons of 3-OH FA proxy calibrations performed in different locations have highlighted the existence of region-specific variation in the relationship between 3-OH FA proxies and environmental temperature and pH (Véquaud *et al*., 2021; Wang *et al*., 2021), and some proxies appear to be applicable only to soil, lake or marine environments (Yang *et al*., 2020, 2021). It has been suggested that bacterial community composition may drive this variation, but previous studies combining analyses of 3-OH FAs and the bacterial community are restricted to a small subset of samples (Yang *et al*., 2020, 2021). Associations between 3-OH FAs and the bacterial community across large environmental gradients of temperature, precipitation and pH in soils and lake surface sediments were therefore explored with a correlation-based approach in Chapter 5. Soil and lake sediment bacterial communities were found to be highly distinct, and 3-OH FAs extracted from soils and lake sediments may therefore be produced by bacteria specific to each environment. If producers in soils and lake sediments respond differently to their environment, this may explain the poor performance of soil-based proxies in lake sediments. Associations between each 3-OH FA isomer and a particular taxonomic lineage were not apparent, suggesting that 3-OH FAs may be produced by a variety of bacterial taxa across multiple lineages. However, this correlation-based approach was limited in its ability to

distinguish producer relationships from co-occurrence and emphasised the need for culture-based analysis to confirm the producers of each 3-OH FA isomer.

7.2.2. Bacterial producers of 3-OH FAs

To identify the bacterial producers of 3-OH FAs and explore how 3-OH FA production may vary between taxonomic lineages, strains of bacteria isolated from soil were grown in pure culture and their 3-OH FAs were extracted and analysed in Chapter 6. These results suggested that 3-OH FA composition may be broadly similar within taxonomic lineages. Bacteroidetes may predominantly produce C_{16}, C_{16} and C_{17} 3-OH FA isomers and could therefore be a major driver of the relationships observed between RAN_{15}, RAN_{17} and RIN_{17} indices and temperature (Wang *et al*., 2016; Yang *et al*., 2021). Proteobacteria appeared to produce mostly even-chained 3-OH FAs and as a dominant phylum in a variety of different environments (Chapter 5; Fierer *et al*., 2007; Janssen, 2006; Li *et al*., 2009; Vipindas *et al*., 2020; Wan *et al*., 2017; Zhang *et al*., 2014), this may explain the abundance of even-chained isomers observed in environmental samples (Chapter 5; Dong *et al*., 2023; Huguet *et al*., 2019; Véquaud *et al*., 2021; Wang *et al*., 2016; Yang *et al*., 2020, 2021). However, these results also revealed that some 3-OH FA isomers can be produced by multiple taxonomic lineages and 3-OH FA composition may vary between taxa within the same phylum, class or even genera. This finding supports the conclusions of Chapter 5 and provides evidence that while some taxonomic patterns in 3-OH FA composition exist, each 3-OH FA isomer may be produced by a variety of bacterial taxa across multiple taxonomic lineages. To ensure the development of 3-OH FAs as reliable palaeoclimate biomarkers, future proxy calibrations should assess bacterial community composition to account for possible sources of variation. The bacterial producers of 3-OH FAs should also be taken into consideration to inform the development of novel 3-OH FA proxies

which can be applied to a wide range of environments and regions and show strong and consistent responses to environmental conditions.

7.2.3. Summary

The results of Chapters 5 and 6 demonstrated that variation in bacterial community composition may lead to discrepancies in 3-OH FA distribution and proxy performance in environmental samples, possibly due to bacterial taxa producing a distinct suite of 3-OH FAs, and it was shown that:

(i) temperature and pH proxies based on 3-OH FAs have differential success in soils compared to lake sediments,

(ii) 3-OH FAs extracted from soils and lake sediments may be produced by distinct communities of bacteria specific to each environment,

(iii) and although there may be broad-scale taxonomic patterns in 3-OH FA composition, each 3-OH FA isomer may be produced by a range of bacterial taxa across multiple taxonomic lineages.

However, further study is needed to analyse and compare the composition of 3-OH FAs produced by a wider diversity of bacteria and determine how 3-OH FA production may vary in response to conditions including growth temperature and pH. This information may further the development and allow refinement of 3-OH FA palaeoclimate proxies that can be applied to a range of different environments and locations.

7.2.4. Recommendations to further elucidate the producers of 3-OH FAs

7.2.4.1. Production of 3-OH FAs by a wide diversity of bacteria

The results of Chapter 5 suggested that the bacterial producers of 3-OH FAs may differ between soils and lake sediments, and while Chapter 6 revealed that different bacterial lineages may produce different 3-OH FAs, these results were restricted to a small number of bacterial strains isolated from soil. Further study is therefore needed to isolate bacteria from both soils and lake sediments to investigate how 3-OH FA production may vary between bacteria isolated from different environments. An assessment of the 3-OH FAs produced by a much wider diversity of bacteria is also needed, particularly those with few cultured representatives such as Acidobacteria (Kielak *et al*., 2016). Some bacterial taxa may be difficult to obtain in pure culture because they may have slow growth rates, require specific nutrients or are syntrophic and require co-culture with other taxa (Chaudhary *et al*., 2019; Pham and Kim, 2012). Longer incubation times, lower nutrient concentrations and growth media that mimics the natural environment and contains specific growth factors may improve the recovery of previously uncultured bacteria from environmental samples (Bollmann *et al*., 2007; Janssen *et al*., 2002; Pham and Kim, 2012).

Growth of bacterial isolates over a range of different temperatures and pH levels and analysis of their 3-OH FAs may improve our understanding of the role of 3-OH FAs in homeoviscous membrane adaptation, as demonstrated by Hellequin *et al*. (2023) for three Bacteroidetes strains. If repeated for a wide diversity of bacterial isolates, this may reveal whether 3-OH FA production in response to growth conditions differs within or between taxonomic lineages. Furthermore, 3-OH FA indices can be calculated based on the 3-OH FA isomers extracted from these pure bacterial cultures and correlated with growth temperature or pH for a comparison with environmental proxy calibrations. This may allow the dominant bacterial taxa contributing

to the relationships observed between each 3-OH FA index and temperature or pH to be identified. However, responses from bacterial cultures may not be representative of those in the natural environment due to possible acclimation to laboratory conditions and strain domestication (Eydallin *et al*., 2014). Experimental sites such as the Rothamsted UK soil experiment (Jenkinson, 1991) where conditions such as soil pH and nutrient concentrations have been manipulated over long time periods may therefore be particularly useful for studying bacterial responses in a controlled environmental setting (Bull *et al*., 2000; Pietri and Brookes, 2008). For example, with paired 3-OH FA and bacterial community analyses in experimental soils covering a large pH gradient, the community response and influence on 3-OH FA distribution can be studied. Bacterial 3-OH FA responses to pH are poorly understood compared to responses to temperature (Hellequin *et al*., 2023), and Chapter 5 highlighted the difficulty associated with identifying bacterial community responses to co-correlating environmental conditions. This approach may allow separation of the bacterial membrane response to conditions such as pH from other environmental variables such as soil type, MAAT, MAP or geographic location which remain relatively consistent, and possibly allow a 3-OH FA proxy based on a specific response to pH to be refined.

7.2.4.2. Identification of 3-OH FA genes

A number of genes involved in the addition of 3-OH carbon chains to lipid A have been identified, including *lpxA*, *lpxD*, *lpxM* and *lpxL* (Raetz *et al*., 2008). Genomic and transcriptomic studies of a wide diversity of bacterial isolates found to produce different 3-OH FA isomers may reveal whether these key acyltransferase genes are differentially expressed in different taxonomic lineages and in taxa exposed to different growth conditions. Knowledge of the genes responsible for 3-OH FA production in different bacterial taxa may allow their 3-OH

FA composition to be predicted based on metagenomic data, which is particularly useful for taxa that are difficult to obtain in pure culture.

A comprehensive understanding of 3-OH FA producers may allow the development of reliable 3-OH FA proxies based on the responses of bacterial membranes to temperature or pH independently, or on the responses of specific members of the bacterial community which may be widely distributed and produce a subset of 3-OH FA isomers which show strong, well-characterised responses to environmental conditions that are consistent between sample types and regions. The development and refinement of bacterial 3-OH FA proxies that can be applied widely to sedimentary archives are crucial for producing reliable palaeoclimate reconstructions.

7.3. Thesis conclusions

The results of this thesis have shown that sedDNA can be used to reconstruct past bacterial and eukaryotic microbial communities over timescales of up to 100 years. sedDNA can be applied to the wider microbial community and may provide a higher taxonomic resolution compared to traditional pigment, lipid or microfossil biomarkers. With improved reference sequence database coverage, sedDNA also has the potential to provide a higher taxonomic resolution compared to microscopy-based monitoring records. However, the reliability of sedDNA-based reconstructions is likely dependent on the extent of differential deposition and degradation and contribution from active sediment communities, which require further study.

This thesis has also revealed that bacteria across different taxonomic lineages may produce a distinct suite of 3-OH FAs, and the bacterial community may therefore be an important source of variation in regional 3-OH FA calibrations performed in different environments. It is recommended that the possible bacterial sources of 3-OH FAs are considered when developing

3-OH FA palaeoclimate proxies for use in different environments. 3-OH FAs have previously been shown to be effective palaeoclimate biomarkers in both terrestrial and marine environments (Dong *et al*., 2023; Wang *et al*., 2018; Yang *et al*., 2020), and their production by a range of different bacterial taxa suggests that 3-OH FAs have the potential to be applied more widely compared to other bacterial lipid palaeoclimate biomarkers, which may not be applicable to terrestrial environments, and whose bacterial sources are uncertain.

Although all microbial biomarkers are associated with a number of biases which must be considered when producing palaeoenvironmental reconstructions, each biomarker can provide complementary information. Where possible, the application of multiple microbial biomarkers to sedimentary archives is likely to be the most effective approach for producing comprehensive, detailed and reliable reconstructions of past ecosystems and climates.

References

Aguirre, L.E., Ouyang, L., Elfwing, A., Hedblom, M., Wulff, A. and Inganäs, O. (2018) Diatom frustules protect DNA from ultraviolet light. *Scientific Reports*, 8 (1): 1-6. https://doi.org/10.1038/s41598-018-21810-2.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3): 403-410.

Anderson, C. J., Verkuilen, J. and Johnson, T. (2010) Applied generalized linear mixed models: Continuous and discrete data. New York: Springer.

Anderson, N.J. (2014) Landscape disturbance and lake response: temporal and spatial perspectives. *Freshwater Reviews*, 7 (2): 77-120. https://doi.org/10.1608/FRJ-7.2.811.

Anderson-Carpenter, L.L., McLachlan, J.S., Jackson, S.T., Kuch, M., Lumibao, C.Y. and Poinar, H.N. (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology*, 11 (1): 1-15. https://doi.org/10.1186/1471-2148-11-30.

Andrén, E., Andrén, T. and Sohlenius, G. (2000) The Holocene history of the southwestern Baltic Sea as reflected in a sediment core from the Bornholm Basin. *Boreas*, 29 (3): 233-250. https://doi.org/10.1111/j.1502-3885.2000.tb00981.x.

Anslan, S., Kang, W., Dulias, K., Wünnemann, B., Börner, N., Schwarz, A., Echeverría-Galindo, P., Liu, Y., Liu, K., Künzel, S. and Kisand, V. (2022) Compatibility of diatom valves records with sedimentary ancient DNA amplicon data: a case study in a brackish, alkaline Tibetan Lake. *Frontiers in Earth Science*, 181. https://doi.org/10.3389/feart.2022.824656.

Apolinarska, K., Pleskot, K., Pełechata, A., Migdałek, M., Siepak, M. and Pełechaty, M. (2020) The recent deposition of laminated sediments in highly eutrophic Lake Kierskie, western Poland: 1 year pilot study of limnological monitoring and sediment traps. *Journal of Paleolimnology*, 63 (4): 283-304. https://doi.org/10.1007/s10933-020-00116-2.

Appleby, P.G. (2001) Chronostratigraphic techniques in recent sediments. In: Last, W.M. and Smol, J.P. (eds.), *Tracking Environmental Change Using Lake Sediments Volume 1: Basin Analysis, Coring, and Chronological Techniques* (pp 171-203). Dordrecht: Kluwer Academic.

Appleby, P.G. and Oldfield, F. (1978) The calculation of ^{210}Pb dates assuming a constant rate of supply of unsupported ^{210}Pb to the sediment. *Catena*, 5:1-8. https://doi.org/10.1016/S0341-8162(78)80002-2.

Appleby, P.G., Nolan, P.J., Gifford, D.W., Godfrey, M.J., Oldfield, F. Anderson, J.N. and Battarbee, R.W. (1986) ^{210}Pb dating by low background gamma counting. *Hydrobiologia*, 141: 21-27.

Appleby, P.G., Richardson, N. and Nolan, P.J. (1992) Self-absorption corrections for well-type germanium detectors. *Nuclear Instruments and Methods*, 71: 228-233. https://doi.org/10.1016/0168-583X(92)95328-O.

Arcusa, S.H., McKay, N.P., Wiman, C., Patterson, S., Munoz, S.E. and Aquino-López, M.A. (2022) A Bayesian approach to integrating radiometric dating and varve measurements in intermittently indistinct sediment. *Geochronology*, 4 (1): 409-433. https://doi.org/10.5194/gchron-4-409-2022.

Asker, D., Beppu, T. and Ueda, K. (2007) *Sphingomonas astaxanthinifaciens* sp. nov., a novel astaxanthin-producing bacterium of the family Sphingomonadaceae isolated from Misasa, Tottori, Japan. *FEMS Microbiology Letters*, 273 (2): 140-148. https://doi.org/10.1111/j.1574-6968.2007.00760.x.

Axford, Y., Andresen, C.S., Andrews, J.T., Belt, S.T., Geirsdóttir, Á., Massé, G., Miller, G.H., Ólafsdóttir, S. and Vare, L.L. (2011) Do paleoclimate proxies agree? A test comparing 19 late Holocene climate and sea-ice reconstructions from Icelandic marine and lake sediments. *Journal of Quaternary Science*, 26 (6): 645-656. https://doi.org/10.1002/jqs.1487.

Barouillet, C., Monchamp, M.E., Bertilsson, S., Brasell, K., Domaizon, I., Epp, L.S., Ibrahim, A., Mejbel, H., Nwosu, E.C., Pearman, J.K., Picard, M., Thomson-Laing, G., Tsugeki, N., Von Eggers, J., Gregory-Eaves, I., Pick, F., Wood, S. and Capo, E. (2022) Investigating the effects of anthropogenic stressors on lake biota using sedimentary DNA. *Freshwater Biology*. https://doi.org/10.1111/fwb.14027.

Bartram, A.K., Jiang, X., Lynch, M.D., Masella, A.P., Nicol, G.W., Dushoff, J. and Neufeld, J.D. (2014) Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiology Ecology*, 87 (2): 403-415. https://doi.org/10.1111/1574-6941.12231.

Battarbee, R.W. (2000) Palaeolimnological approaches to climate change, with special regard to the biological record. *Quaternary Science Reviews*, 19 (1-5): 107-124. https://doi.org/10.1016/S0277-3791(99)00057-8.

Battarbee, R.W., Anderson, N.J., Bennion, H. and Simpson, G.L. (2012) Combining limnological and palaeolimnological data to disentangle the effects of nutrient pollution and climate change on lake ecosystems: problems and potential. *Freshwater Biology*, 57 (10): 2091-2106. https://doi.org/10.1111/j.1365-2427.2012.02860.x.

Belt, S.T. (2018) Source-specific biomarkers as proxies for Arctic and Antarctic sea ice. *Organic geochemistry*, 125: 277-298. https://doi.org/10.1016/j.orggeochem.2018.10.002.

Bennion, H., Battarbee, R.W., Sayer, C.D., Simpson, G.L. and Davidson, T.A. (2011) Defining reference conditions and restoration targets for lake ecosystems using palaeolimnology: a synthesis. *Journal of Paleolimnology*, 45 (4): 533-544. https://doi.org/10.1007/s10933-010-9419-3.

Bennion, H., Monteith, D. and Appleby, P. (2000) Temporal and geographical variation in lake trophic status in the English Lake District: evidence from (sub) fossil diatoms and aquatic macrophytes. *Freshwater Biology*, 45 (4): 394-412. https://doi.org/10.1046/j.1365-2427.2000.00626.x.

Bernardet, J.F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K. and Vandamme, P. (1996) Cutting a Gordian knot: emended classification and description of the genus *Flavobacterium*, emended description of the family Flavobacteriaceae, and proposal of *Flavobacterium hydatis* nom. nov. (basonym, *Cytophaga aquatilis* Strohl and Tait 1978). *International Journal of Systematic Bacteriology*, 46 (1): 128-148. https://doi.org/10.1099/00207713-46-1-128

Besson, F., Tenoux, I., Hourdou, M.L. and Michel, G. (1992) Synthesis of β-hydroxy fatty acids and β-amino fatty acids by the strains of *Bacillus subtilis* producing iturinic antibiotics. *Biochimica et Biophysica Acta – Lipids and Lipid Metabolism*, 1123 (1): 51-58. 10.1016/0005-2760(92)90170-z.

Biessy, L., Pearman, J.K., Waters, S., Vandergoes, M.J. and Wood, S.A. (2022) Metagenomic insights to the functional potential of sediment microbial communities in freshwater lakes. *Metabarcoding and Metagenomics*, 6: e79265. https://doi.org/10.3897/mbmg.6.79265.

Billard, E., Domaizon, I., Tissot, N., Arnaud, F. and Lyautey, E. (2015) Multi-scale phylogenetic heterogeneity of archaea, bacteria, methanogens and methanotrophs in lake sediments. *Hydrobiologia*, 751 (1): 159-173. https://doi.org/10.1007/s10750-015-2184-6.

Birk, S., Chapman, D., Carvalho, L., Spears, B.M., Andersen, H.E., Argillier, C., Auer, S., Baattrup-Pedersen, A., Banin, L., Beklioğlu, M. and Bondar-Kunze, E. (2020) Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. *Nature Ecology and Evolution*, 4 (8): 1060-1068. https://doi.org/10.1038/ s41559-020-1216-4.

Birks, H.H. and Birks, H.J.B. (2006) Multi-proxy studies in palaeolimnology. *Vegetation History and Archaeobotany*, 15: 235-251. https://doi.org/10.1007/s00334-006-0066-6.

Blaga, C.I., Reichart, G.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. *Journal of Paleolimnology*, 41: 523-540. https://doi.org/10.1007/s10933-008-9242-2.

Bloem, J. and Breure, A.M. (2003) Microbial indicators. In: Markert, B.A., Breure, A.M. and Zechmeister, H.G. (eds), *Trace Metals and Other Contaminants in the Environment*. Amsterdam: Elsevier. https://doi.org/10.1016/S0927-5215(03)80138-8.

Blumenberg, M., Seifert, R., Buschmann, B., Kiel, S. and Thiel, V. (2012) Biomarkers reveal diverse microbial communities in black smoker sulfides from turtle pits (Mid-Atlantic Ridge, Recent) and Yaman Kasy (Russia, Silurian). *Geomicrobiology Journal*, 29 (1): 66-75. https://doi.org/10.1080/01490451.2010.523445.

Blundell, A., Holden, J. and Turner, T.E. (2016) Generating multi-proxy Holocene palaeoenvironmental records from blanket peatlands. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 443: 216-229. https://doi.org/10.1016/j.palaeo.2015.11.048.

Boere, A.C., Sinninghe Damsté, J.S., Rijpstra, W.I.C., Volkman, J.K. and Coolen, M.J. (2011) Source-specific variability in post-depositional DNA preservation with potential implications for DNA based paleoecological records. *Organic Geochemistry*, 42 (10): 1216-1225. https://doi.org/10.1016/j.orggeochem.2011.08.005.

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., Yu, D.W., Seymour, M., Dumbrell, A.J. and Creer, S. (2022) Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*, 22 (4): 1231-1246. https://doi.org/10.1111/1755-0998.13512.

Bollmann, A., Lewis, K. and Epstein, S.S. (2007) Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. *Applied and Environmental Microbiology*, 73 (20): 6386-6390. https://doi.org/10.1128/AEM.01309-07.

Boschker, H.T.S. and Middelburg, J.J. (2002) Stable isotopes and biomarkers in microbial ecology. *FEMS Microbiology Ecology*, 40 (2): 85-95. https://doi.org/10.1111/j.1574-6941.2002.tb00940.x.

Brasell, K.A., Pochon, X., Howarth, J., Pearman, J.K., Zaiko, A., Thompson, L., Vandergoes, M.J., Simon, K.S. and Wood, S.A. (2022) Shifts in DNA yield and biological community composition in stored sediment: implications for paleogenomic studies. *Metabarcoding and Metagenomics*, 6: 78128. https//doi.org/.10.3897/mbmg.6.78128.

Bravo, I. and Figueroa, R.I. (2014) Towards an ecological understanding of dinoflagellate cyst functions. *Microorganisms*, 2 (1): 11-32. https://doi.org/10.3390/microorganisms2010011.

Brocks, J.J., Grice, K. (2011) Biomarkers (molecular fossils). In: Reitner, J., Thiel, V. (eds), *Encyclopedia of Geobiology*. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9212-1_30.

Brocks, J.J., Love, G.D., Summons, R.E., Knoll, A.H., Logan, G.A. and Bowden, S.A. (2005) Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature*, 437 (7060): 866-870. https://doi.org/10.1038/nature04068.

Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L. and Sadowsky, M.J. (2017) A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary. *Environmental Science and Technology*, 51 (15): 8263-8271. https://doi.org/10.1021/acs.est.7b01353.

Bull, I.D., van Bergen, P.F., Nott, C.J., Poulton, P.R. and Evershed, R.P. (2000) Organic geochemical studies of soils from the Rothamsted classical experiments - V. The fate of lipids in different long-term experiments. *Organic geochemistry*, 31 (5): 389-408. https://doi.org/10.1016/S0146-6380(00)00008-5.

Burlakova, L.E., Hinchey, E.K., Karatayev, A.Y. and Rudstam, L.G. (2018) US EPA Great Lakes National Program Office monitoring of the Laurentian Great Lakes: insights from 40 years of data collection. *Journal of Great Lakes Research*, 44 (4): 535-538. https://doi.org/10.1016/j.jglr.2018.05.017.

Burnham K. P. and Anderson D. R. (2002) Model selection and multimodel inference: A practical information-theoretic approach. New York: Springer. https://doi.org/10.1007/b97636.

Cai, H.Y., Jiang, H.L., Krumholz, L.R. and Yang, Z. (2014) Bacterial community composition of size-fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater lake. *PloS One*, 9 (8): e102879. https://doi.org/10.1371/journal.pone.0102879.

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A. and Holmes S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13: 581-583. https://doi.org/10.1038/nmeth.3869.

Canino, A., Bouchez, A., Laplace-Treyture, C., Domaizon, I., Rimet, F. (2021) Phytool, a ShinyApp to homogenise taxonomy of freshwater microalgae from DNA barcodes and microscopic observations. *Metabarcoding and Metagenomics* 5: e74096. https://doi.org/10.3897/mbmg.5.74096.

Cao, H., Hu, J., Xi, D., Tang, Y., Lei, Y. and Shilling, A. (2016) Paleoenvironmental reconstruction of the late Santonian Songliao Paleo-lake. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 457: 290-303. https://doi.org/10.1016/j.palaeo.2016.05.027.

Cao, X., Xu, X., Bian, R., Wang, Y., Yu, H., Xu, Y., Duan, G., Bi, L., Chen, P., Gao, S. and Wang, J. (2020) Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities. *Water Research*, 183: 116077. https://doi.org/10.1016/j.watres.2020.116077.

Capo, E., Debroas, D., Arnaud, F. and Domaizon, I. (2015) Is planktonic diversity well recorded in sedimentary DNA? Toward the reconstruction of past protistan diversity. *Microbial Ecology*, 70 (4): 865-875. https://doi.org/10.1007/s00248-015-0627-2.

Capo, E., Debroas, D., Arnaud, F., Guillemot, T., Bichet, V., Millet, L., Gauthier, E., Massa, C., Develle, A.L., Pignol, C. and Lejzerowicz, F. (2016) Long-term dynamics in microbial eukaryotes communities: A palaeolimnological view based on sedimentary DNA. *Molecular Ecology*, 25 (23): 5925-5943. https://doi.org/10.1111/mec.13893.

Capo, E., Debroas, D., Arnaud, F., Perga, M.E., Chardon, C. and Domaizon, I. (2017) Tracking a century of changes in microbial eukaryotic diversity in lakes driven by nutrient enrichment and climate warming. *Environmental Microbiology*, 19 (7): 2873-2892. https://doi.org/10.1111/1462-2920.13815.

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A.G., Clarke, C., Crump, S.E., Debroas, D., Englund, G., Ficetola, G.F., Garner, R., Gauthier, J., Gregory-Eaves, I., Heinecke, L., Herzschuh, U., Ibrahim, A., Kisand, V., Kjær, K.H., Lammers, Y., Littlefair, J., Messager, E., Monchamp, M., Olajos, F., Orsi, W., Pedersen, M.W., Rijal, D.P., Rydberg, J., Spanbauer, T., Stoof-Leichsenring, K.R., Taberlet, P., Talas, L., Thomas, C., Walsh, D., Wang, Y., Willerslev, E., van Woerkom, A., Zimmermann, H.H., Coolen, M.J.L., Epp, L.S., Domaizon, I., Alsos, I.G. and Parducci, L. (2021). Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: Overview and recommendations. *Quaternary*, 3. https://doi.org/10.3390/quat4010006.

Capo, E., Monchamp, M.E., Coolen, M.J., Domaizon, I., Armbrecht, L. and Bertilsson, S. (2022) Environmental paleomicrobiology: using DNA preserved in aquatic sediments to its full potential. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.15913.

Cardman, Z., Arnosti, C., Durbin, A., Ziervogel, K., Cox, C., Steen, A.D. and Teske, A. (2014) Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. *Applied and Environmental Microbiology*, 80 (12): 3749-3756. https://doi.org/10.1128/AEM.00899-14.

Carlson, R.E. (1977) A trophic state index for lakes. *Limnology and Oceanography*. 22 (2): 361-369. https://doi.org/10.4319/lo.1977.22.2.0361.

Caroff, M. and Karibian, D. (2003) Structure of bacterial lipopolysaccharides. *Carbohydrate Research*, 338 (23): 2431-2447. https://doi.org/10.1016/j.carres.2003.07.010.

Carpenter, S.R., Stanley, E.H. and Vander Zanden, M.J. (2011) State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annual Review of Environment and Resources*, 36: 75-99. https://doi.org/10.1146/annurev-environ-021810-094524.

Chaudhary, D.K., Khulan, A. and Kim, J. (2019) Development of a novel cultivation technique for uncultured soil bacteria. *Scientific Reports*, 9 (1): 1-11. https://doi.org/10.1038/s41598-019-43182-x.

Chen, L., Wang, Y., Xie, S., Kershaw, S., Dong, M., Yang, H., Liu, H. and Algeo, T.J. (2011) Molecular records of microbialites following the end-Permian mass extinction in Chongyang, Hubei Province, South China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 308 (1-2): 151-159. https://doi.org/10.1016/j.palaeo.2010.09.010.

Chen, N., Yang, J.S., Qu, J.H., Li, H.F., Liu, W.J., Li, B.Z., Wang, E.T. and Yuan, H.L. (2015) Sediment prokaryote communities in different sites of eutrophic Lake Taihu and their interactions with environmental factors. *World Journal of Microbiology and Biotechnology*, 31 (6): 883-896. https://doi.org/10.1007/s11274-015-1842-1.

Chen, W.M., Chen, Z.H., Young, C.C. and Sheu, S.Y. (2016b) *Hymenobacter paludis* sp. nov., isolated from a marsh. *International Journal of Systematic and Evolutionary Microbiology*, 66 (3): 1546-1553. https://doi.org/10.1099/ijsem.0.000915.

Chen, Y., Dai, Y., Wang, Y., Wu, Z., Xie, S. and Liu, Y. (2016a) Distribution of bacterial communities across plateau freshwater lake and upslope soils. *Journal of Environmental Sciences*, 43: 61-69. https://doi.org/10.1016/j.jes.2015.08.012.

Chen, Y.L., Lee, C.C., Lin, Y.L., Yin, K.M., Ho, C.L. and Liu, T. (2015) Obtaining long 16S rDNA sequences using multiple primers and its application on dioxin-containing samples. *BMC Bioinformatics*, 16 (S13). https://doi.org/10.1186/1471-2105-16-S18-S13.

Cheng, J.Y., Hui, E.L. and Lau, A.P. (2012) Bioactive and total endotoxins in atmospheric aerosols in the Pearl River Delta region, China. *Atmospheric Environment*, 47: 3-11. https://doi.org/10.1016/j.atmosenv.2011.11.055.

Christie, W.W. and Han, X. (2010) Gas chromatographic analysis of fatty acid derivatives. In: *Lipid Analysis: Isolation, Separation, Identification and Lipidomic Analysis*. Woodhead Publishing: Cambridge, UK.

Chu, H., Gao, G.F., Ma, Y., Fan, K. and Delgado-Baquerizo, M. (2020) Soil microbial biogeography in a changing world: recent advances and future perspectives. *MSystems*, 5 (2): e00803-19. https://doi.org/10.1128/mSystems.00803-19.

Comte, J., Berga, M., Severin, I., Logue, J.B. and Lindström, E.S. (2017) Contribution of different bacterial dispersal sources to lakes: population and community effects in different seasons. *Environmental Microbiology*, 19 (6): 2391-2404. https://doi.org/10.1111/1462-2920.13749.

Cooke, M.P., Talbot, H.M. and Farrimond, P. (2008) Bacterial populations recorded in bacteriohopanepolyol distributions in soils from Northern England. *Organic Geochemistry*, 39 (9): 1347-1358. https://doi.org/10.1016/j.orggeochem.2008.05.003.

Coolen, M.J. and Overmann, J. (1998) Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment. *Applied and Environmental Microbiology*, 64 (11): 4513-4521. https://doi.org/10.1128/AEM.64.11.4513-4521.1998.

Coolen, M.J., Muyzer, G., Rijpstra, W.I.C., Schouten, S., Volkman, J.K. and Sinninghe Damsté, J.S. (2004) Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. *Earth and Planetary Science Letters*, 223 (1-2): 225-239. https://doi.org/10.1016/j.epsl.2004.04.014.

Dai, Z., Su, W., Chen, H., Barberán, A., Zhao, H., Yu, M., Yu, L., Brookes, P.C., Schadt, C.W., Chang, S.X. and Xu, J. (2018) Long-term nitrogen fertilization decreases bacterial diversity and favors the growth of Actinobacteria and Proteobacteria in agro-ecosystems across the globe. *Global Change Biology*, 24 (8): 3452-3461. https://doi.org/10.1111/gcb.14163.

Davidson, T.A. and Jeppesen, E. (2013) The role of palaeolimnology in assessing eutrophication and its impact on lakes. *Journal of Paleolimnology*, 49 (3): 391-410. https://doi.org/10.1007/s10933-012-9651-0.

De Carvalho, C.C. and Caramujo, M.J. (2018) The various roles of fatty acids. *Molecules*, 23 (10): 2583. https://doi.org/10.3390/molecules23102583.

De Figueiredo, D.R., Pereira, M.J., Moura, A., Silva, L., Barrios, S., Fonseca, F., Henriques, I. and Correia, A. (2007) Bacterial community composition over a dry winter in meso-and eutrophic Portuguese water bodies. *FEMS Microbiology Ecology*, 59 (3): 638-650. https://doi.org/10.1111/j.1574-6941.2006.00241.x.

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. *Organic Geochemistry*, 137: 103897. https://doi.org/10.1016/j.orggeochem.2019.07.006.

Dearing, J.A., Battarbee, R.W., Dikau, R., Larocque, I. and Oldfield, F. (2006) Human-environment interactions: learning from the past. *Regional Environmental Change*, 6: 1-16. https://doi.org/10.1007/s10113-005-0011-8.

Dedysh, S.N., Kulichevskaya, I.S., Serkebaeva, Y.M., Mityaeva, M.A., Sorokin, V.V., Suzina, N.E., Rijpstra, W.I.C. and Sinninghe Damsté, J.S. (2012) *Bryocella elongata* gen. nov., sp. nov., a member of subdivision 1 of the Acidobacteria isolated from a methanotrophic enrichment culture, and emended description of *Edaphobacter aggregans* Koch *et al*. 2008. *International Journal of Systematic and Evolutionary Microbiology*, 62 (3): 654-664. https://doi.org/10.1099/ijs.0.031898-0.

Deshpande, B.N., Tremblay, R., Pienitz, R. and Vincent, W.F. (2014) Sedimentary pigments as indicators of cyanobacterial dynamics in a hypereutrophic lake. *Journal of Paleolimnology*, 52: 171-184. https://doi.org/10.1007/s10933-014-9785-3.

Domaizon, I., Savichtcheva, O., Debroas, D., Arnaud, F., Villar, C., Pignol, C., Alric, B. and Perga, M.E. (2013) DNA from lake sediments reveals the long-term dynamics and diversity of Synechococcus assemblages. *Biogeosciences*, 10 (6): 3817-3838. https://doi.org/10.5194/bg-10-3817-2013.

Domaizon, I., Winegardner, A., Capo, E., Gauthier, J. and Gregory-Eaves, I. (2017) DNA-based methods in paleolimnology: new opportunities for investigating long-term dynamics of lacustrine biodiversity. *Journal of Paleolimnology*, 58 (1): 1-21. https://doi.org/10.1007/s10933-017-9958-y.

Dommain, R., Andama, M., McDonough, M.M., Prado, N.A., Goldhammer, T., Potts, R., Maldonado, J.E., Nkurunungi, J.B. and Campana, M.G. (2020) The challenges of reconstructing tropical biodiversity with sedimentary ancient DNA: A 2200-year-long metagenomic record from Bwindi impenetrable forest, Uganda. *Frontiers in Ecology and Evolution*, 8: 218. https://doi.org/10.3389/fevo.2020.00218.

Dong, X., Bennion, H., Battarbee, R.W. and Sayer, C.D. (2011) A multiproxy palaeolimnological study of climate and nutrient impacts on Esthwaite Water, England over the past 1200 years. *The Holocene*, 22 (1): 107-118. https://doi.org/10.1177/0959683611409780.

Dong, X., Bennion, H., Maberly, S.C., Sayer, C.D., Simpson, G.L. and Battarbee, R.W. (2011) Nutrients exert a stronger control than climate on recent diatom communities in Esthwaite Water: evidence from monitoring and palaeolimnological records. *Freshwater Biology*, 57 (10): 2044-2056. https://doi.org/10.1111/j.1365-2427.2011.02670.x.

Dong, Z., Yang, Y., Wang, C., Bendle J.A., Ruan, X., Lü, X. and Xie S. (2023) Development of a novel sea surface temperature proxy based on bacterial 3-hydroxy fatty acids. *Frontiers in Marine Science*, 9: 1050269. https://doi.org/10.3389/fmars.2022.1050269.

dos Santos, R.A.L., Wilkins, D., De Deckker, P. and Schouten, S. (2012) Late Quaternary productivity changes from offshore Southeastern Australia: A biomarker approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 363: 48-56. https://doi.org/10.1016/j.palaeo.2012.08.013.

Drljepan, M., McCarthy, F.M. and Hubeny, J.B. (2014) Natural and cultural eutrophication of Sluice Pond, Massachusetts, USA, recorded by algal and protozoan microfossils. *The Holocene*, 24 (12): 1731-1742. https://doi.org/10.1177/0959683614551227.

Du, J., Xiao, K., Huang, Y., Li, H., Tan, H., Cao, L., Lu, Y. and Zhou, S. (2011) Seasonal and spatial diversity of microbial communities in marine sediments of the South China Sea. *Antonie Van Leeuwenhoek*, 100: 317-331. https://doi.org/10.1007/s10482-011-9587-9.

Dulias, K., Stoof-Leichsenring, K.R., Pestryakova, L.A. and Herzschuh, U. (2017) Sedimentary DNA versus morphology in the analysis of diatom-environment relationships. *Journal of Paleolimnology*, 57: 51-66. https://doi.org/10.1007/s10933-016-9926-y.

Dumri, K., Seipold, L., Schmidt, J., Gerlach, G., Dötterl, S., Ellis, A.G. and Wessjohann, L.A. (2008) Non-volatile floral oils of *Diascia* spp. (Scrophulariaceae). *Phytochemistry*, 69 (6): 1372-1383. https://doi.org/10.1016/j.phytochem.2007.12.012.

Edwards, M.E. (2020) The maturing relationship between Quaternary paleoecology and ancient sedimentary DNA. *Quaternary Research*, 96: 39-47. https://doi:10.1017/qua.2020.52.

Eglinton, G. and Logan, G.A. (1991) Molecular preservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 333 (1268): 315-328. https://doi.org/10.1098/rstb.1991.0081.

Eglinton, T.I. and Eglinton, G. (2008) Molecular proxies for paleoclimatology. *Earth and Planetary Science Letters*, 275 (1-2): 1-16. https://doi.org/10.1016/j.epsl.2008.07.012.

Eichorst, S.A., Breznak, J.A. and Schmidt, T.M. (2007) Isolation and characterization of soil bacteria that define *Terriglobus* gen. nov., in the phylum Acidobacteria. *Applied and Environmental Microbiology*, 73 (8): 2708-2717. https://doi.org/10.1128/AEM.02140-06.

Ellegaard, M., Clokie, M.R., Czypionka, T., Frisch, D., Godhe, A., Kremp, A., Letarov, A., McGenity, T.J., Ribeiro, S. and Anderson, J.N. (2020) Dead or alive: sediment DNA archives as tools for tracking aquatic evolution and adaptation. *Communications Biology*, 3 (1): 1-11. https://doi.org/10.1038/s42003-020-0899-z.

Elliott, J.A., Irish, A.E. and Reynolds, C.S. (2010) Modelling phytoplankton dynamics in fresh waters: affirmation of the PROTECH approach to simulation. *Freshwater Reviews*, 3 (1): 75-96. https://doi.org/10.1608/FRJ-3.1.4.

Ernst, R., Ejsing, C.S. and Antonny, B. (2016) Homeoviscous adaptation and the regulation of membrane lipids. *Journal of Molecular Biology*, 428 (24): 4776-4791. https://doi.org/10.1016/j.jmb.2016.08.013.

Evans, R.D. (1994) Empirical evidence of the importance of sediment resuspension in lakes. *Hydrobiologia*, 284: 5-12. https://doi.org/10.1007/BF00005727.

Eydallin, G., Ryall, B., Maharjan, R. and Ferenci, T. (2014) The nature of laboratory domestication changes in freshly isolated *Escherichia coli* strains. *Environmental Microbiology*, 16 (3): 813-828. https://doi.org/10.1111/1462-2920.12208.

Faust, K. and Raes, J. (2016) CoNet app: inference of biological association networks using Cytoscape. F1000 Research, 5: 1519. https://doi.org/10.12688/f1000research.9050.2.

Fautz, E., Rosenfelder, G. and Grotjahn, L. (1979) Iso-branched 2- and 3-hydroxy fatty acids as characteristic lipid constituents of some gliding bacteria. *Journal of Bacteriology*, 140 (3): 852-858. https://doi.org/10.1128/jb.140.3.852-858.1979.

Feist, S.M. and Lance, R.F. (2021) Genetic detection of freshwater harmful algal blooms: A review focused on the use of environmental DNA (eDNA) in *Microcystis aeruginosa* and *Prymnesium parvum*. *Harmful Algae*, 110: 102124. https://doi.org/10.1016/j.hal.2021.102124.

Feng, X., Wang, Y., Liu, T., Jia, J., Ma, T. and Liu, Z. (2019) Biomarkers and their applications in ecosystem research. *Chinese Journal of Plant Ecology*, 44 (4): 384-394. https://doi.org/ 10.17521/cjpe.2019.0139.

Ferrando, R., Szponar, B., Sánchez, A., Larsson, L. and Valero-Guillén, P.L. (2005) 3-Hydroxy fatty acids in saliva as diagnostic markers in chronic periodontitis. *Journal of Microbiological Methods*, 62 (3): 285-291. https://doi.org/10.1016/j.mimet.2005.04.014.

Fetzner, S. (2015) Quorum quenching enzymes. *Journal of Biotechnology*, 201: 2-14. https://doi.org/10.1016/j.jbiotec.2014.09.001.

Feuchtmayr, H., Thackeray, S.J., Jones, I.D., De Ville, M., Fletcher, J., James, B.E.N. and Kelly, J. (2012) Spring phytoplankton phenology - are patterns and drivers of change consistent among lakes in the same climatological region? *Freshwater Biology*, 57 (2): 331-344. https://doi.org/10.1111/j.1365-2427.2011.02671.x.

Fierer, N. and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103 (3): 626-631. https://doi.org/10.1073/pnas.0507535103.

Fierer, N., Bradford, M.A. and Jackson, R.B. (2007) Toward an ecological classification of soil bacteria. *Ecology*, 88 (6): 1354-1364. https://doi.org/10.1890/05-1839.

Flavier, A.B., Clough, S.J., Schell, M.A. and Denny, T.P. (1997) Identification of 3-hydroxypalmitic acid methyl ester as a novel autoregulator controlling virulence in *Ralstonia solanacearum*. *Molecular Microbiology*, 26 (2): 251-259. https://doi.org/10.1046/j.1365-2958.1997.5661945.x.

Florian, C.R., Miller, G.H., Fogel, M.L., Wolfe, A.P., Vinebrooke, R.D. and Geirsdóttir, Á. (2015) Algal pigments in Arctic lake sediments record biogeochemical changes due to Holocene climate variability and anthropogenic global change. *Journal of Paleolimnology*, 54: 53-69. https://doi.org/10.1007/s10933-015-9835-5.

Francioli, D., Lentendu, G., Lewin, S. and Kolb, S. (2021) DNA metabarcoding for the characterization of terrestrial microbiota – pitfalls and solutions. *Microorganisms*, 9 (2): 361. https://doi.org/10.3390/microorganisms9020361.

Funkey, C.P., Conley, D.J., Reuss, N.S., Humborg, C., Jilbert, T. and Slomp, C.P. (2014) Hypoxia sustains cyanobacteria blooms in the Baltic Sea. *Environmental Science and Technology*, 48 (5): 2598-2602. https://doi.org/10.1021/es404395a.

Gaston, K.J. and He, F. (2011) Species occurrence and occupancy. In: Magurran, A.E. and McGill, B.J. (eds.), *Biological diversity: frontiers in measurement and assessment* (141-151). Oxford: Oxford University Press.

Gauthier, J., Walsh, D., Selbie, D.T., Bourgeois, A., Griffiths, K., Domaizon, I. and Gregory-Eaves, I. (2021) Evaluating the congruence between DNA-based and morphological taxonomic approaches in water and sediment trap samples: Analyses of a 36-month time series from a temperate monomictic lake. *Limnology and Oceanography*, 66 (8): 3020-3039. https://doi.org/10.1002/lno.11856.

Gauthier, J., Walsh, D., Selbie, D.T., Domaizon, I. and Gregory-Eaves, I. (2022) Sedimentary DNA of a human-impacted lake in Western Canada (Cultus Lake) reveals changes in micro-eukaryotic diversity over the past ~200 years. *Environmental DNA*, 4 (5): 1106-1119. https://doi.org/10.1002/edn3.310.

Giguet-Covex, C., Ficetola, G.F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A.L. and David, F. (2019) New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. *Scientific Reports*, 9 (1): 1-21. Https://doi.org/10.1038/s41598-019-50339-1.

Glew, J.R., Smol, J.P., Last, W.M. (2002) Sediment core collection and extrusion. In: Last, W.M., Smol, J.P. (eds), *Tracking Environmental Change Using Lake Sediments*. Dordrecht: Springer. https://doi.org/10.1007/0-306-47669-X_5.

Glöckner, F.O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R. and Ludwig, W. (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, 261: 169-176. https://doi.org/10.1016/j.jbiotec.2017.06.1198.

Gong, W. and Marchetti, A. (2019) Estimation of 18S gene copy number in marine eukaryotic plankton using a next-generation sequencing approach. *Frontiers in Marine Science*, 6: 219. https://doi.org/10.3389/fmars.2019.00219.

Gong, W., Hall, N., Paerl, H. and Marchetti, A. (2020) Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. *Environmental Microbiology*, 22 (11): 4718-4731. https://doi.org/10.1111/1462-2920.15221.

Griffiths, K., Jeziorski, A., Antoniades, D., Beaulieu, M., Smol, J.P. and Gregory-Eaves, I. (2022) Pervasive changes in algal indicators since pre-industrial times: A paleolimnological study of changes in primary production and diatom assemblages from ~200 Canadian lakes. *Science of The Total Environment*, p.155938. https://doi.org/10.1016/j.scitotenv.2022.155938.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M. and Whiteley, A.S. (2011) The bacterial biogeography of British soils. *Environmental Microbiology*, 13 (6): 1642-1654. https://doi.org/10.1111/j.1462-2920.2011.02480.x.

Guan, N. and Liu, L. (2020) Microbial response to acid stress: mechanisms and applications. *Applied Microbiology and Biotechnology*, 104 (1): 51-65. https://doi.org/10.1007/s00253-019-10226-1.

Guedes, I.A., Rachid, C.T., Rangel, L.M., Silva, L.H., Bisch, P.M., Azevedo, S.M. and Pacheco, A.B. (2018) Close link between harmful cyanobacterial dominance and associated bacterioplankton in a tropical eutrophic reservoir. *Frontiers in Microbiology*, 9: 424. https://doi.org/10.3389/fmicb.2018.00424.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. and Christen, R. (2012) The Protist Ribosomal Reference database (PR^2): a catalogue of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41 (D1): D597-D604. https://doi.org/10.1093/nar/gks1160.

Guiry, M.D. and Guiry, G.M. (2022) *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org; searched on 08 August 2022.

Haglund, A.L., Lantz, P., Törnblom, E. and Tranvik, L. (2003) Depth distribution of active bacteria and bacterial activity in lake sediment. *FEMS Microbiology Ecology*, 46 (1): 31-38. https://doi.org/10.1016/S0168-6496(03)00190-9.

Halamka, T.A., Raberg, J.H., McFarlin, J.M., Younkin, A.D., Mulligan, C., Liu, X.L. and Kopf, S.H. (2022) Production of diverse brGDGTs by Acidobacterium *Solibacter usitatus* in response to temperature, pH, and O2 provides a culturing perspective on brGDGT proxies and biosynthesis. *Geobiology*, 21 (1): 102-118. https://doi.org/10.1111/gbi.12525.

Hampton, S.E., Izmest'eva, L.R., Moore, M.V., Katz, S.L., Dennis, B. and Silow, E.A. (2008) Sixty years of environmental change in the world's largest freshwater lake – Lake Baikal, Siberia. *Global Change Biology*, 14 (8): 1947-1958. https://doi.org/10.1111/j.1365-2486.2008.01616.x.

Han, X., Tolu, J., Deng, L., Fiskal, A., Schubert, C.J., Winkel, L.H. and Lever, M.A. (2022) Long-term preservation of biomolecules in lake sediments: potential importance of physical shielding by recalcitrant cell walls. *PNAS Nexus*, 1 (3): pgac076. https://doi.org/10.1093/pnasnexus/pgac076.

Harrison, J.B., Sunday, J.M. and Rogers, S.M. (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B*, 286 (1915): 20191409. https://doi.org/10.1098/rspb.2019.1409.

He, Y., Sen, B., Zhou, S., Xie, N., Zhang, Y., Zhang, J. and Wang, G. (2017) Distinct seasonal patterns of bacterioplankton abundance and dominance of phyla α-Proteobacteria and cyanobacteria in Qinhuangdao coastal waters off the Bohai sea. *Frontiers in Microbiology*, 8: 1579. https://doi.org/10.3389/fmicb.2017.01579.

Heathcote, A.J., Taranu, Z.E., Tromas, N., MacIntyre-Newell, M., Leavitt, P.R. and Pick, F.R. (2023) Sedimentary DNA and pigments show increasing abundance and toxicity of cyanoHABs during the Anthropocene. *Freshwater Biology*. https://doi.org/10.1111/fwb.14069.

Hegerl, G.C., Brönnimann, S., Cowan, T., Friedman, A.R., Hawkins, E., Iles, C., Müller, W., Schurer, A. and Undorf, S. (2019) Causes of climate change over the historical record. *Environmental Research Letters*, 14 (12): 123006. https://doi.org/10.1088/1748-9326/ab4557.

Heino, J., Virkkala, R. and Toivonen, H. (2009) Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. *Biological Reviews*, 84 (1): 39-54. https://doi.org/10.1111/j.1469-185X.2008.00060.x.

Hellequin, E., Collin, S., Seder-Colomina, M., Véquaud, P., Anquetil, C., Kish, A. and Huguet, A. (2023) Membrane lipid adaptation of soil Bacteroidetes isolates to temperature and pH. *Frontiers in Microbiology*, 14. https://doi.org/10.3389/fmicb.2023.1032032.

Hembrow, S.C., Taffs, K.H., Atahan, P., Parr, J., Zawadzki, A. and Heijnis, H. (2014) Diatom community response to climate variability over the past 37,000 years in the sub-tropics of the Southern Hemisphere. *Science of the Total Environment*, 468: 774-784. https://doi.org/10.1016/j.scitotenv.2013.09.003.

Hermans, S.M., Buckley, H.L., Case, B.S. and Lear, G. (2020) Connecting through space and time: catchment-scale distributions of bacteria in soil, stream water and sediment. *Environmental Microbiology*, 22: 1000-1010. https://doi.org/10.1111/1462-2920.1479.

Hobbs, W.O., Telford, R.J., Birks, H.J.B., Saros, J.E., Hazewinkel, R.R., Perren, B.B., Saulnier-Talbot, E. and Wolfe, A.P. (2010) Quantifying recent ecological changes in remote lakes of North America and Greenland using sediment diatom assemblages. *PloS One*, 5 (4): e10026. https://doi.org/10.1371/journal.pone.0010026.

Hoef-Emden, K., Archibald, J.M. (2017) Cryptophyta (Cryptomonads). In: Archibald, J., Simpson, A., Slamovits, C. (eds.), *Handbook of the Protists*. Cham: Springer. https://doi.org/10.1007/978-3-319-28149-0_35.

Hu, Y., Xiang, D., Veresoglou, S.D., Chen, F., Chen, Y., Hao, Z., Zhang, X. and Chen, B. (2014) Soil organic carbon and soil structure are driving microbial abundance and community composition across the arid and semi-arid grasslands in northern China. *Soil Biology and Biochemistry*, 77: 51-57. https://doi.org/10.1016/j.soilbio.2014.06.014.

Huang, W., Chen, X., Jiang, X. and Zheng, B. (2017) Characterization of sediment bacterial communities in plain lakes with different trophic statuses. *Microbiology Open*, 6 (5): e00503. https://doi.org/10.1002/mbo3.503.

Huang, W., Chen, X., Wang, K., Chen, J., Zheng, B. and Jiang, X. (2019) Comparison among the microbial communities in the lake, lake wetland, and estuary sediments of a plain river network. *Microbiology Open*, 8 (2): e00644. https://doi.org/10.1002/mbo3.644.

Huguet, A., Coffinet, S., Roussel, A., Gayraud, F., Anquetil, C., Bergonzini, L., Bonanomi, G., Williamson, D., Majule, A. and Derenne, S. (2019) Evaluation of 3-hydroxy fatty acids as a pH and temperature proxy in soils from temperate and tropical altitudinal gradients. *Organic Geochemistry*, 129: 1-13. https://doi.org/10.1016/j.orggeochem.2019.01.002.

Huguet, C., Kim, J.H., Sinninghe Damsté, J.S. and Schouten, S. (2006) Reconstruction of sea surface temperature variations in the Arabian Sea over the last 23 kyr using organic proxies (TEX_{86} and $U_{37}^{K'}$). *Paleoceanography*, 21 (3). https://doi.org/10.1029/2005PA001215.

Hulme, M. and Jones, P.D. (1994) Global climate change in the instrumental period. *Environmental Pollution*, 83 (1-2): 23-36. https://doi.org/10.1016/0269-7491(94)90019-1.

Huo, S., Zhang, H., Monchamp, M.E., Wang, R., Weng, N., Zhang, J., Zhang, H. and Wu, F. (2022) Century-long homogenization of algal communities is accelerated by nutrient enrichment and climate warming in lakes and reservoirs of the North Temperate Zone. *Environmental Science and Technology*, 56 (6): 3780-3790. https://doi.org/10.1021/acs.est.1c06958.

Ibrahim, A., Capo, E., Wessels, M., Martin, I., Meyer, A., Schleheck, D. and Epp, L.S. (2021) Anthropogenic impact on the historical phytoplankton community of Lake Constance reconstructed by multimarker analysis of sediment-core environmental DNA. *Molecular Ecology*, 30 (13): 3040-3056. https://doi.org/10.1111/mec.15696.

Ikemoto, S., Kuraishi, H., Komagata, K., Azuma, R., Suto, T. and Murooka, H. (1978) Cellular fatty acid composition in *Pseudomonas* species. *The Journal of General and Applied Microbiology*, 24 (4): 199-213. https://doi.org/10.2323/jgam.24.199.

Ilyashuk, B.P., Andreev, A.A., Bobrov, A.A., Tumskoy, V.E. and Ilyashuk, E.A. (2006) Interglacial history of a palaeo-lake and regional environment: a multi-proxy study of a permafrost deposit from Bol'shoy Lyakhovsky Island, Arctic Siberia. *Journal of Paleolimnology*, 35: 855-872. https://doi.org/10.1007/s10933-005-5859-6.

Janssen, P.H. (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, 72 (3): 1719-1728. https://doi.org/10.1128/AEM.72.3.1719-1728.2006.

Janssen, P.H., Yates, P.S., Grinton, B.E., Taylor, P.M. and Sait, M. (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied and Environmental Microbiology*, 68 (5): 2391-2396. https://doi.org/10.1128/AEM.68.5.2391-2396.2002.

Jantzen, E. and Hofstad, T. (1981) Fatty acids of *Fusobacterium* species: taxonomic implications. *Microbiology*, 123 (1): 163-171. https://doi.org/10.1099/00221287-123-1-163.

Jarvis, B.D.W., Sivakumaran, S., Tighe, S.W. and Gillis, M. (1996) Identification of *Agrobacterium* and *Rhizobium* species based on cellular fatty acid composition. *Plant and Soil*, 184: 143-158. https://doi.org/10.1007/BF00029284.

Jenkinson, D.S. (1991) The Rothamsted long-term experiments: Are they still of use? *Agronomy Journal*, 83 (1): 2-10. https://doi.org/10.2134/agronj1991.00021962008300010008x.

Johne, B., Olson, I. and Bryn, K. (1988) Fatty acids and sugars in lipopolysaccharides from *Bacteroides intermedius*, *Bacteroides gingivalis* and *Bacteroides loescheii*. *Oral Microbiology and Immunology*, 3 (1): 22-27. https://doi.org/10.1111/j.1399-302X.1988.tb00600.x.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) A comprehensive survey of soil Acidobacterial diversity using pyrosequencing and clone library analyses. *The ISME journal*, 3 (4): 442-453. https://doi.org/10.1038/ismej.2008.127.

Kalam, S., Basu, A., Ahmad, I., Sayyed, R.Z., El-Enshasy, H.A., Dailin, D.J. and Suriani, N.L. (2020) Recent understanding of soil Acidobacteria and their ecological significance: a critical review. *Frontiers in Microbiology*, 11: 580024. https://doi.org/10.3389/fmicb.2020.580024.

Kaneda, T. (1991) Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiological Reviews*, 55 (2): 288-302. https://doi.org/10.1128/mr.55.2.288-302.1991.

Kang, W., Anslan, S., Börner, N., Schwarz, A., Schmidt, R., Künzel, S., Rioual, P., Echeverría-Galindo, P., Vences, M., Wang, J. and Schwalb, A. (2021) Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: The effect of sample-size and bioinformatics on the identified communities. *Ecological Indicators*, 121: 107070. https://doi.org/10.1016/j.ecolind.2020.107070.

Katayama-Fujimura, Y., Tsuzaki, N. and Kuraishi, H. (1982) Ubiquinone, fatty acid and DNA base composition determination as a guide to the taxonomy of the genus *Thiobacillus*. *Microbiology*, 128 (7): 1599-1611. https://doi.org/10.1099/00221287-128-7-1599.

Keinänen, M.M., Korhonen, L.K., Martikainen, P.J., Vartiainen, T., Miettinen, I.T., Lehtola, M.J., Nenonen, K., Pajunen, H. and Kontro, M.H. (2003) Gas chromatographic–mass spectrometric detection of 2-and 3-hydroxy fatty acids as methyl esters from soil, sediment and biofilm. *Journal of Chromatography B*, 783 (2): 443-451. https://doi.org/10.1016/S1570-0232(02)00713-4.

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., Van Veen, J.A. and Kuramae, E.E. (2016) The ecology of Acidobacteria: moving beyond genes and genomes. *Frontiers in Microbiology*, 7: 744. https://doi.org/10.3389/fmicb.2016.00744.

Kiersztyn, B., Chróst, R., Kaliński, T., Siuda, W., Bukowska, A., Kowalczyk, G. and Grabowska, K. (2019) Structural and functional microbial diversity along a eutrophication gradient of interconnected lakes undergoing anthropopressure. *Scientific Reports*, 9 (1): 1-14. https://doi.org/10.1038/s41598-019-47577-8.

Kim, J.H., Schouten, S., Hopmans, E.C., Donner, B. and Sinninghe Damsté, J.S. (2008) Global sediment core-top calibration of the TEX_{86} paleothermometer in the ocean. *Geochimica et Cosmochimica Acta*, 72 (4): 1154-1173. https://doi.org/10.1016/j.gca.2007.12.010.

Kjær, K.H., Pedersen, M.W., De Sanctis, B., De Cahsan, B., Korneliussen, T.S., Michelsen, C.S., Sand, K.K., Jelavić, S., Ruter, A.H., Schmidt, A.M. and Kjeldsen, K.K. (2022) A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, 612 (7939): 283-291. https://doi.org/10.1038/s41586-022-05453-y.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R. and Kelley, S.T. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, 8 (9): 761-763. https://doi.org/10.1038/nmeth.1650.

Kock, J.L., Strauss, C.J., Pohl, C.H. and Nigam, S. (2003) The distribution of 3-hydroxy oxylipins in fungi. *Prostaglandins and other Lipid Mediators*, 71 (3-4): 85-96. https://doi.org/10.1016/S1098-8823(03)00046-7.

Koponen, H.T., Jaakkola, T., Keinänen-Toivola, M.M., Kaipainen, S., Tuomainen, J., Servomaa, K. and Martikainen, P.J. (2006) Microbial communities, biomass, and activities in soils as affected by freeze thaw cycles. *Soil Biology and Biochemistry*, 38 (7): 1861-1871. https://doi.org/10.1016/j.soilbio.2005.12.010.

Korhonen, J.J., Soininen, J. and Hillebrand, H. (2010) A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. *Ecology*, 91 (2): 508-517. https://doi.org/10.1890/09-0392.1.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, 79 (17): 5112-5120. https://doi.org/10.1128/AEM.01043-13.

Kpodonu, A.T.N., Hamilton, D.P., Hartland, A., Laughlin, D.C. and Lusk, C.H. (2016) Coupled use of sediment phosphorus speciation and pigment composition to infer phytoplankton phenology over 700 years in a deep oligotrophic lake. *Biogeochemistry*, 129 (1): 181-196. https://doi.org/10.1007/s10533-016-0227-3.

Králová, S., Švec, P., Busse, H.J., Staňková, E., Váczi, P. and Sedláček, I. (2018) *Flavobacterium chryseum* sp. nov. and *Flavobacterium psychroterrae* sp. nov., novel environmental bacteria isolated from Antarctica. *International Journal of Systematic and Evolutionary Microbiology*, 68 (10): 3132-3139. https://doi.org/10.1099/ijsem.0.002952.

Kraśnik, L., Szponar, B., Walczak, M., Larsson, L. and Gamian, A. (2006) Routine clinical laboratory tests correspond to increased serum levels of 3-hydroxy fatty acids, markers of endotoxins, in cardiosurgery patients. *Archivum Immunologiae et Therapiae Experimentalis*, 54: 55-60. https://doi.org/10.1007/s00005-006-0006-2.

Kuang, B., Xiao, R., Wang, C., Zhang, L., Wei, Z., Bai, J., Zhang, K., Campos, M. and Jorquera, M.A. (2022) Bacterial community assembly in surface sediments of a eutrophic shallow lake in northern China. *Ecohydrology and Hydrobiology*. https://doi.org/10.1016/j.ecohyd.2022.01.005.

Lamentowicz, M., Gałka, M., Obremska, M., Kühl, N., Lücke, A. and Jassey, V.E.J. (2015) Reconstructing climate change and ombrotrophic bog development during the last 4000 years in northern Poland using biotic proxies, stable isotopes and trait-based approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 418: 261-277. https://doi.org/10.1016/j.palaeo.2014.11.015.

Lange-Bertalot, H., Hofmann, G., Werum, M., Cantonati, M. (2017) Freshwater benthic diatoms of Central Europe: over 800 common species used in ecological assessment. Lange-Bertalot, H., Werum, M., Cantonati, M, Kelly, M. (eds.). Czechia: Koeltz Botanical Books.

Laprida, C., García Chapori, N.L., Violante, R.A. (2017). Principles of Paleoceanographic Reconstruction. In: *The Argentina Continental Margin*. Cham: Springer. https://doi.org/10.1007/978-3-319-04196-4_6.

Larsson, J. (2022) eulerr: area-proportional Euler and Venn diagrams with ellipses. https://cran.r-project.org/package=eulerr.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied *Environmental Microbiology*, 75 (15): 5111-5120. https://doi.org/10.1128/AEM.00335-09.

Leavitt, P.R. and Carpenter, S.R. (1989) Effects of sediment mixing and benthic algal production on fossil pigment stratigraphies. *Journal of Paleolimnology*, 2: 147-158. https://doi.org/10.1007/BF00177044.

Leavitt, P.R., Hodgson, D.A. (2002) Sedimentary pigments. In: Smol, J.P., Birks, H.J.B., Last, W.M., Bradley, R.S., Alverson, K. (eds.), *Tracking Environmental Change Using Lake Sediments. Developments in Paleoenvironmental Research*. Dordrecht: Springer. https://doi.org/10.1007/0-306-47668-1_15.

Lee, A.K., Chan, C.K., Fang, M. and Lau, A.P. (2004) The 3-hydroxy fatty acids as biomarkers for quantification and characterization of endotoxins and gram-negative bacteria in atmospheric aerosols in Hong Kong. *Atmospheric Environment*, 38 (37): 6307-6317. https://doi.org/10.1016/j.atmosenv.2004.08.013.

Lee, H.G., An, D.S., Im, W.T., Liu, Q.M., Na, J.R., Cho, D.H., Jin, C.W., Lee, S.T. and Yang, D.C. (2007) *Chitinophaga ginsengisegetis* sp. nov. and *Chitinophaga ginsengisoli* sp. nov., isolated from soil of a ginseng field in South Korea. *International Journal of Systematic and Evolutionary Microbiology*, 57 (7): 396-1401. https://doi.org/10.1099/ijs.0.64688-0.

Legendre, P. (2018) lmodel2: Model II regression. https://cran.rproject.org/package=lmodel2.

Leira, M. (2005) Diatom responses to Holocene environmental changes in a small lake in northwest Spain. *Quaternary International*, 140: 90-102. https://doi.org/10.1016/j.quaint.2005.05.005.

Levin, P.A. and Angert, E.R. (2015) Small but mighty: cell size and bacteria. *Cold Spring Harbor Perspectives in Biology*, 7 (7): a019216. https://doi.org/10.1101/cshperspect.a019216.

Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z. and Ettema, T.J. (2021) Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, 19 (4): 225-240. https://doi.org/10.1038/s41579-020-00458-8.

Li, F., Zhang, X., Xie, Y. and Wang, J. (2019a) Sedimentary DNA reveals over 150 years of ecosystem change by human activities in Lake Chao, China. *Environment International*, 133: 105214. https://doi.org/10.1016/j.envint.2019.105214.

Li, H., Yu, Y., Luo, W., Zeng, Y. and Chen, B. (2009) Bacterial diversity in surface sediments from the Pacific Arctic Ocean. *Extremophiles*, 13: 233-246. https://doi.org/10.1007/s00792-009-0225-7.

Li, H.Q., Shen, Y.J., Wang, W.L., Wang, H.T., Li, H. and Su, J.Q. (2021) Soil pH has a stronger effect than arsenic content on shaping plastisphere bacterial communities in soil. *Environmental Pollution*, 287: 117339. https://doi.org/10.1016/j.envpol.2021.117339.

Li, S., Fang, J., Zhu, X., Spencer, R.G., Álvarez-Salgado, X.A., Deng, Y., Huang, T., Yang, H. and Huang, C. (2022) Properties of sediment dissolved organic matter respond to eutrophication and interact with bacterial communities in a plateau lake. *Environmental Pollution*, 301: 118996. https://doi.org/10.1016/j.envpol.2022.118996.

Li, Y., Wu, S., Wang, L., Li, Y., Shi, F. and Wang, X. (2010) Differentiation of bacteria using fatty acid profiles from gas chromatography–tandem mass spectrometry. *Journal of the Science of Food and Agriculture*, 90 (8): 1380-1383. https://doi.org/10.1002/jsfa.3931.

Li, Y., Zhang, J., Zhang, J., Xu, W. and Mou, Z. (2019b) Microbial community structure in the sediments and its relation to environmental factors in eutrophicated Sancha Lake. *International Journal of Environmental Research and Public Health*, 16 (11): 1931. https://doi.org/10.3390/ijerph16111931.

Lim, J.H., Baek, S.H. and Lee, S.T. (2009) *Ferruginibacter alkalilentus* gen. nov., sp. nov. and *Ferruginibacter lapsinanis* sp. nov., novel members of the family 'Chitinophagaceae' in the phylum Bacteroidetes, isolated from freshwater sediment. *International Journal of Systematic and Evolutionary Microbiology*, 59 (10): 2394-2399. https://doi.org/10.1099/ijs.0.009480-0.

Limaye, R.B., Kumaran, K.P.N., Nair, K.M. and Padmalal, D. (2010) Cyanobacteria as potential biomarkers of hydrological changes in the Late Quaternary sediments of South Kerala Sedimentary Basin, India. *Quaternary International*, 213 (1-2): 79-90. https://doi.org/10.1016/j.quaint.2009.09.016.

Lipski, A., Klatte, S., Bendinger, B. and Altendorf, K. (1992) Differentiation of gram-negative, nonfermentative bacteria isolated from biofilters on the basis of fatty acid composition, quinone system, and physiological reaction profiles. *Applied and Environmental Microbiology*, 58 (6): 2053-2065. https://doi.org/10.1128/aem.58.6.2053-2065.1992.

Litchman, E., de Tezanos Pinto, P., Edwards, K.F., Klausmeier, C.A., Kremer, C.T. and Thomas, M.K. (2015) Global biogeochemical impacts of phytoplankton: a trait-based perspective. *Journal of Ecology*, 103 (6): 1384-1396. https://doi.org/10.1111/1365-2745.12438.

Liu, C., Cui, Y., Li, X., Yao, M. (2021) microeco: An R package for data mining in microbial community ecology, *FEMS Microbiology Ecology*, 97 (2): fiaa255. https://doi.org/10.1093/femsec/fiaa255.

Liu, Q., Siddiqi, M.Z., Kim, M.S., Kim, S.Y. and Im, W.T. (2017b) *Mucilaginibacter hankyongensis* sp. nov., isolated from soil of ginseng field Baekdu Mountain. *Journal of Microbiology*, 55: 525-530. https://doi.org/10.1007/s12275-017-7180-2.

Liu, S., Wang, H., Chen, L., Wang, J., Zheng, M., Liu, S., Chen, Q. and Ni, J. (2020) Comammox *Nitrospira* within the Yangtze River continuum: community, biogeography, and ecological drivers. *The ISME Journal*, 14 (10): 2488-2504. https://doi.org/10.1038/s41396-020-0701-8.

Liu, X., Cong, J., Lu, H., Xue, Y., Wang, X., Li, D. and Zhang, Y. (2017a) Community structure and elevational distribution pattern of soil Actinobacteria in alpine grasslands. *Acta Ecologica Sinica*, 37 (4): 213-218. https://doi.org/10.1016/j.chnaes.2017.02.010.

Liu, Y., Zhang, J., Zhao, L., Zhang, X. and Xie, S. (2014) Spatial distribution of bacterial communities in high-altitude freshwater wetland sediment. *Limnology*, 15 (3): 249-256. https://doi.org/10.1007/s10201-014-0429-0.

Liu, Z., Fu, B., Zheng, X. and Liu, G. (2010) Plant biomass, soil water content and soil N: P ratio regulating soil microbial functional diversity in a temperate steppe: a regional scale study. *Soil Biology and Biochemistry*, 42 (3): 445-450. https://doi.org/10.1016/j.soilbio.2009.11.027.

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 and Planetary Science Letters*, 357: 277-288. https://doi.org/10.1016/j.epsl.2012.09.031.

Losey, N.A., Stevenson, B.S., Busse, H.J., Damste, J.S.S., Rijpstra, W.I.C., Rudd, S. and Lawson, P.A. (2013) *Thermoanaerobaculum aquaticum* gen. nov., sp. nov., the first cultivated member of Acidobacteria subdivision 23, isolated from a hot spring. *International Journal of Systematic and Evolutionary Microbiology,* 63 (11): 4149-4157. https://doi.org/10.1099/ijs.0.051425-0.

Luo, G., Yang, H., Algeo, T.J., Hallmann, C. and Xie, S. (2019) Lipid biomarkers for the reconstruction of deep-time environmental conditions. *Earth-Science Reviews*, 189: 99-124. https://doi.org/10.1016/j.earscirev.2018.03.005.

Maberly, S.C. and Elliott, J.A. (2012) Insights from long-term studies in the Windermere catchment: external stressors, internal interactions and the structure and function of lake ecosystems. *Freshwater Biology*, 57 (2): 233-243. https://doi.org/10.1111/j.1365-2427.2011.02718.x.

Maberly, S.C., Brierley, B., Carter, H.T., Clarke, M.A., De Ville, M.M., Fletcher, J.M., James, J.B., Keenan, P., Kelly, J.L., Mackay, E.B., Parker, J.E., Patel, M., Pereira, M.G., Rhodes, G., Tanna, B., Thackeray, S.J., Vincent, C., Feuchtmayr, H. (2017) Surface temperature, surface oxygen, water clarity, water chemistry and phytoplankton chlorophyll a data from Esthwaite Water, 1945 to 2013. *NERC Environmental Information Data Centre*. https://doi.org/10.5285/87360d1a-85d9-4a4e-b9ac-e315977a52d3.

Maberly, S.C., De Ville, M.M., Feuchtmayr, H., Jones, I.D., Mackay, E.B., May, L., Thackeray, S.J. and Winfield, I.J. (2011b) The limnology of Esthwaite Water: historical change and its causes, current state and prospects for the future. *Report to Natural England*. Centre for Ecology and Hydrology, Lancaster.

Maberly, S.C., De Ville, M.M., Kelly, J. and Thackeray, S.J. (2011a) The state of Esthwaite Water in 2010. *Report to Natural England*. Centre for Ecology and Hydrology, Lancaster.

Maberly, S.C., O'Donnell, R.A., Woolway, R.I., Cutler, M.E., Gong, M., Jones, I.D., Merchant, C.J., Miller, C.A., Politi, E., Scott, E.M. and Thackeray, S.J. (2020) Global lake thermal regions shift under climate change. *Nature Communications*, 11 (1): 1-9. https://doi.org/10.1038/s41467-020-15108-z.

Mackay, E.B., Jones, I.D., Folkard, A.M. and Barker, P. (2012) Contribution of sediment focussing to heterogeneity of organic carbon and phosphorus burial in small lakes. *Freshwater Biology*, 57 (2): 290-304. https://doi.org/10.1111/j.1365-2427.2011.02616.x.

MacKeigan, P.W., Garner, R.E., Monchamp, M.È., Walsh, D.A., Onana, V.E., Kraemer, S.A., Pick, F.R., Beisner, B.E., Agbeti, M.D., da Costa, N.B. and Shapiro, B.J. (2022) Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. *Harmful Algae*, 113: 102187. https://doi.org/10.1016/j.hal.2022.102187.

Makri, S., Lami, A., Lods-Crozet, B. and Loizeau, J.L. (2019) Reconstruction of trophic state shifts over the past 90 years in a eutrophicated lake in western Switzerland, inferred from the sedimentary record of photosynthetic pigments. *Journal of Paleolimnology*, 61 (2): 129-145. https://doi.org/10.1007/s10933-018-0049-5.

Maloney, K.M., Schiffbauer, J.D., Halverson, G.P., Xiao, S. and Laflamme, M. (2022) Preservation of early Tonian macroalgal fossils from the Dolores Creek Formation, Yukon. *Scientific Reports*, 12 (1): 6222. https://doi.org/10.1038/s41598-022-10223-x.

Mangot, J.F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G. and Debroas, D. (2012) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental Microbiology*, 15 (6): 1745-1758. https://doi.org/10.1111/1462-2920.12065.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17 (1): 10-12. https://doi.org/10.14806/ej.17.1.200.

Matsui, K., Honjo, M. and Kawabata, Z. (2001) Estimation of the fate of dissolved DNA in thermally stratified lake water from the stability of exogenous plasmid DNA. *Aquatic Microbial Ecology*, 26 (1): 95-102. https://doi.org/10.3354/ame026095.

Matsumoto, G.I. and Nagashima, H. (1984) Occurrence of 3-hydroxy acids in microalgae and cyanobacteria and their geochemical significance. *Geochimica et Cosmochimica Acta*, 48 (8): 1683-1687. https://doi.org/10.1016/0016-7037(84)90337-5.

Mauvisseau, Q., Harper, L.R., Sander, M., Hanner, R.H., Kleyer, H. and Deiner, K. (2022) The multiple states of environmental DNA and what is known about their persistence in aquatic environments. *Environmental Science and Technology*, 56 (9): 5322-5333. https://doi.org/ 10.1021/acs.est.1c07638.

McGowan, S., Barker, P., Haworth, E.Y., Leavitt, P.R., Maberly, S.C. and Pates, J. (2012) Humans and climate as drivers of algal community change in Windermere since 1850. *Freshwater Biology*, 57 (2): 260-277. https://doi.org/10.1111/j.1365-2427.2011.02689.x.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PloS ONE*, 8 (4): e61217. https://doi.org/10.1371/journal.pone.0061217.

Mejbel, H.S., Dodsworth, W. and Pick, F.R. (2022) Effects of temperature and oxygen on cyanobacterial DNA preservation in sediments: A comparison study of major taxa. *Environmental DNA*, 4 (4): 717-731. https://doi.org/10.1002/edn3.289.

Mejbel, H.S., Dodsworth, W., Baud, A., Gregory-Eaves, I. and Pick, F.R. (2021) Comparing quantitative methods for analyzing sediment DNA records of cyanobacteria in experimental and reference lakes. *Frontiers in Microbiology*, 12: 1532. https://doi.org/10.3389/fmicb.2021.669910.

Mejbel, H.S., Irwin, C.L., Dodsworth, W., Higgins, S.N., Paterson, M.J. and Pick, F.R. (2023) Long-term cyanobacterial dynamics from lake sediment DNA in relation to experimental eutrophication, acidification and climate change. *Freshwater Biology*. https://doi.org/10.1111/fwb.14074.

Mertens, K.N., Rengefors, K., Moestrup, Ø. and Ellegaard, M. (2012) A review of recent freshwater dinoflagellate cysts: taxonomy, phylogeny, ecology and palaeocology. *Phycologia*, 51 (6): 612-619. https://doi.org/10.2216/11-89.1.

Meyers, P.A. (1997) Organic geochemical proxies of paleoceanographic, paleolimnologic, and paleoclimatic processes. *Organic geochemistry*, 27 (5-6): 213-250. https://doi.org/10.1016/S0146-6380(97)00049-1.

Mills, K., Schillereff, D., Saulnier-Talbot, É., Gell, P., Anderson, N.J., Arnaud, F., Dong, X., Jones, M., McGowan, S., Massaferro, J. and Moorhouse, H. (2017) Deciphering long-term records of natural variability and human impact as recorded in lake sediments: a palaeolimnological puzzle. *Wiley Interdisciplinary Reviews: Water*, 4 (2): e1195. https://doi.org/10.1002/wat2.1195.

Miyagawa, E., Azuma, R. and Suto, T. (1979) Cellular fatty acid composition in gram-negative obligately anaerobic rods. *The Journal of General and Applied Microbiology*, 25 (1): 41-51. https://doi.org/10.2323/jgam.25.41.

Moldowan, J.M. and Jacobson, S.R. (2000) Chemical signals for early evolution of major taxa: biosignatures and taxon-specific biomarkers. *International Geology Review*, 42 (9): 805-812. https://doi.org/10.1080/00206810009465112.

Monchamp, M.E., Spaak, P. and Pomati, F. (2019) Long term diversity and distribution of non-photosynthetic cyanobacteria in peri-alpine lakes. *Frontiers in Microbiology*, 9: 3344. https://doi.org/10.3389/fmicb.2018.03344.

Monchamp, M.E., Walser, J.C., Pomati, F. and Spaak, P. (2016) Sedimentary DNA reveals cyanobacterial community diversity over 200 years in two perialpine lakes. Applied and *Environmental Microbiology*, 82 (21): 6472-6482. https://doi.org/10.1128/AEM.02174-16.

Moorhouse, H.L., McGowan, S., Jones, M.D., Barker, P., Leavitt, P.R., Brayshaw, S.A. and Haworth, E.Y. (2014) Contrasting effects of nutrients and climate on algal communities in two lakes in the Windermere catchment since the late 19th century. *Freshwater Biology*, 59 (12): 2605-2620. https://doi.org/10.1111/fwb.12457.

Moorhouse, H.L., McGowan, S., Taranu, Z.E., Gregory-Eaves, I., Leavitt, P.R., Jones, M.D., Barker, P. and Brayshaw, S.A. (2018) Regional versus local drivers of water quality in the Windermere catchment, Lake District, United Kingdom: The dominant influence of wastewater pollution over the past 200 years. *Global Change Biology*, 24 (9): 4009-4022. https://doi.org/10.1111/gcb.14299.

Morgan, J.A.W., Winstanley, C. (1997) Microbial biomarkers. In: van Elsas, J.D., Trevors, J.T., Wellington, E.M.H. (eds.), *Modern Soil Microbiology*. New York: Marcel Dekker.

Moss, C.W. and Dees, S.B. (1978) Cellular fatty acids of *Flavobacterium meningosepticum* and *Flavobacterium* species group IIb. *Journal of Clinical Microbiology*, 8 (6): 772-774. https://doi.org/10.1099/ijs.0.64119-0.

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. *Organic Geochemistry*, 66: 164-173. https://doi.org/10.1016/j.orggeochem.2013.10.017.

Naeher, S., Smittenberg, R.H., Gilli, A., Kirilova, E.P., Lotter, A.F. and Schubert, C.J. (2012) Impact of recent lake eutrophication on microbial community changes as revealed by high resolution lipid biomarkers in Rotsee (Switzerland). *Organic Geochemistry*, 49: 86-95. https://doi.org/10.1016/j.orggeochem.2012.05.014.

Nakagawa, Y. and Matsuyama, T. (1993) Chromatographic determination of optical configuration of 3-hydroxy fatty acids composing microbial surfactants. *FEMS Microbiology Letters*, 108 (1): 99-102. https://doi.org/10.1111/j.1574-6968.1993.tb06080.x.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D. and Bertilsson, S. (2011) A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*, 75 (1): 14-49. https://doi.org/10.1128/MMBR.00028-10.

Nolan, T., Hands, R.E. and Bustin, S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, 1 (3): 1559-1582. https://doi.org/10.1038/nprot.2006.236.

Nwosu, E.C., Roeser, P., Yang, S., Ganzert, L., Dellwig, O., Pinkerneil, S., Brauer, A., Dittmann, E., Wagner, D. and Liebner, S. (2021) From water into sediment – tracing freshwater cyanobacteria via DNA analyses. *Microorganisms*, 9 (8): 1778. https://doi.org/10.3390/microorganisms9081778.

O'Brien, C.L., Robinson, S.A., Pancost, R.D., Sinninghe Damsté, J.S., Schouten, S., Lunt, D.J., Alsenz, H., Bornemann, A., Bottini, C., Brassell, S.C. and Farnsworth, A. (2017) Cretaceous sea-surface temperature evolution: Constraints from TEX_{86} and planktonic foraminiferal oxygen isotopes. *Earth-Science Reviews*, 172: 224-247. https://doi.org/10.1016/j.earscirev.2017.07.012.

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. and Wagner, H. (2019) vegan: Community ecology packages. https://CRAN.R-project.org/package=vegan.

Oliverio, A.M., Bradford, M.A. and Fierer, N. (2017) Identifying the microbial taxa that consistently respond to soil warming across time and space. *Global Change Biology*, 23 (5): 2117-2129. https://doi.org/10.1111/gcb.13557.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R. and Stahl, D.A. (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology*, 40: 337-365. https://doi.org/10.1146/annurev.mi.40.100186.002005.

Ostrovsky, I. and Yacobi, Y.Z. (2010) Sedimentation flux in a large subtropical lake: spatiotemporal variations and relation to primary productivity. *Limnology and Oceanography*, 55 (5): 1918-1931. https://doi.org/10.4319/lo.2010.55.5.1918.

Ouyang, X., Guo, F. and Bu, H. (2015) Lipid biomarkers and pertinent indices from aquatic environment record paleoclimate and paleoenvironment changes. *Quaternary Science Reviews*, 123: 180-192. https://doi.org/10.1016/j.quascirev.2015.06.029.

Oyaizu, H. and Komagata, K. (1983) Grouping of *Pseudomonas* species on the basis of cellular fatty acid composition and the quinone system with special reference to the existence of 3-hydroxy fatty acids. *The Journal of General and Applied Microbiology*, 29 (1): 17-40. https://doi.org/10.2323/jgam.29.17.

Özen, H.Ç., Başshan, M., Keskin, C. and Toker, Z. (2004) Fatty acid and 3-hydroxy fatty acid composition of two *Hypericum* species from Turkey. *European Journal of Lipid Science and Technology*, 106 (1): 68-70. https://doi.org/10.1002/ejlt.200300828.

Pal, S., Gregory-Eaves, I. and Pick, F.R. (2015) Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. *Journal of Paleolimnology*, 54 (1): 87-101. https://doi.org/10.1007/s10933-015-9839-1.

Pankratov, T.A., Kirsanova, L.A., Kaparullina, E.N., Kevbrin, V.V. and Dedysh, S.N. (2012) *Telmatobacter bradus* gen. nov., sp. nov., a cellulolytic facultative anaerobe from subdivision 1 of the Acidobacteria, and emended description of *Acidobacterium capsulatum* Kishimoto *et al*. 1991. *International Journal of Systematic and Evolutionary Microbiology*, 62 (2): 430-437. https://doi.org/10.1099/ijs.0.029629-0.

Parker, J.H., Smith, G.A., Fredrickson, H.L., Vestal, J.R. and White, D.C. (1982) Sensitive assay, based on hydroxy fatty acids from lipopolysaccharide lipid A, for gram-negative bacteria in sediments. *Applied and Environmental Microbiology*, 44 (5): 1170-1177. https://doi.org/10.1128/aem.44.5.1170-1177.1982.

Paterson, A.M., Betts-Piper, A.A., Smol, J.P. and Zeeb, B.A. (2003) Diatom and chrysophyte algal response to long-term PCB contamination from a point-source in northern Labrador, Canada. *Water, Air, and Soil Pollution*, 145: 377-393. https://doi.org/10.1023/A:1023654105342.

Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F.I., Amalfitano, S., Apothéloz-Perret-Gentil, L., Baussant, T., Bouchez, A., Carugati, L., Cermakova, K. and Cordier, T. (2022) Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Science of the Total Environment*, 181: 151783. https://doi.org/10.1016/j.scitotenv.2021.151783.

Pearman, J.K., Biessy, L., Howarth, J.D., Vandergoes, M.J., Rees, A. and Wood, S.A. (2022) Deciphering the molecular signal from past and alive bacterial communities in aquatic sedimentary archives. *Molecular Ecology Resources*, 22 (3): 877-890. https://doi.org/10.1111/1755-0998.13515.

Pearman, J.K., Thomson-Laing, G., Howarth, J.D., Vandergoes, M.J., Thompson, L., Rees, A. and Wood, S.A. (2021) Investigating variability in microbial community composition in replicate environmental DNA samples down lake sediment cores. *PloS One*, 16 (5): e0250783. https://doi.org/10.1371/journal.pone.0250783.

Pei, H., Wang, C., Wang, Y., Yang, H. and Xie, S. (2019) Distribution of microbial lipids at an acid mine drainage site in China: Insights into microbial adaptation to extremely low pH conditions. *Organic Geochemistry*, 134: 77-91. https://doi.org/10.1016/j.orggeochem.2019.05.008.

Pham, V.H. and Kim, J. (2012) Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, 30 (9): 475-484. https://doi.org/10.1016/j.tibtech.2012.05.007.

Picard, M., Pochon, X., Atalah, J., Pearman, J.K., Rees, A., Howarth, J.D., Moy, C.M., Vandergoes, M.J., Hawes, I., Khan, S. and Wood, S.A. (2022a) Using metabarcoding and droplet digital PCR to investigate drivers of historical shifts in cyanobacteria from six contrasting lakes. *Scientific Reports*, 12 (1): 1-17. https://doi.org/10.1038/s41598-022-14216-8.

Picard, M., Wood, S.A., Pochon, X., Vandergoes, M.J., Reyes, L., Howarth, J.D., Hawes, I. and Puddick, J. (2022b) Molecular and pigment analyses provide comparative results when reconstructing historic cyanobacterial abundances from lake sediment cores. *Microorganisms*, 10 (2): 279. https://doi.org/10.3390/microorganisms10020279.

Pietri, J.A. and Brookes, P.C. (2008) Relationships between soil pH and microbial properties in a UK arable soil. *Soil Biology and Biochemistry*, 40 (7): 1856-1861. https://doi.org/10.1016/j.soilbio.2008.03.020.

Pilon, S., Zastepa, A., Taranu, Z.E., Gregory-Eaves, I., Racine, M., Blais, J.M., Poulain, A.J. and Pick, F.R. (2019) Contrasting histories of microcystin-producing cyanobacteria in two temperate lakes as inferred from quantitative sediment DNA analyses. *Lake and Reservoir Management*, 35 (1): 102-117. https://doi.org/10.1080/10402381.2018.1549625.

Pohl, C.H., Kock, J.L. and Thibane, V.S. (2011) Antifungal free fatty acids: a review. *Science Against Microbial Pathogens: Communicating Current Research and Technological Advances*, 3: 61-71.

Pollet, T., Humbert, J.F. and Tadonléké, R.D. (2014) Planctomycetes in lakes: poor or strong competitors for phosphorus? *Applied and Environmental Microbiology*, 80 (3): 819-828. https://doi.org/10.1128/AEM.02824-13.

Powers, L., Werne, J.P., Vanderwoude, A.J., Sinninghe Damsté, J.S., Hopmans, E.C. and Schouten, S. (2010) Applicability and calibration of the TEX_{86} paleothermometer in lakes. *Organic Geochemistry*, 41 (4): 404-413. https://doi.org/10.1016/j.orggeochem.2009.11.009.

PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu, data created 4 February 2014, accessed 02 March 2023.

Qian, L., Yu, X., Gu, H., Liu, F., Fan, Y., Wang, C., He, Q., Tian, Y., Peng, Y., Shu, L. and Wang, S. (2023) Vertically stratified methane, nitrogen and sulphur cycling and coupling mechanisms in mangrove sediment microbiomes. *Microbiome*, 11 (1): 1-19. https://doi.org/10.1186/s40168-023-01501-5.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41 (D1): D590-D596. https://doi.org/10.1093/nar/gks1219.

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.r.project.org.

Racovita, R.C., Peng, C., Awakawa, T., Abe, I. and Jetter, R. (2015) Very-long-chain 3-hydroxy fatty acids, 3-hydroxy fatty acid methyl esters and 2-alkanols from cuticular waxes of *Aloe arborescens* leaves. *Phytochemistry*, 113: 183-194. https://doi.org/10.1016/j.phytochem.2014.08.005.

Raetz, C.R., Reynolds, C.M., Trent, M.S. and Bishop, R.E. (2007) Lipid A modification systems in gram-negative bacteria. *Annual Review of Biochemistry*, 76: 295-329. https://doi.org/10.1146/annurev.biochem.76.010307.145803.

Ramírez, G.A., Jørgensen, S.L., Zhao, R. and D'Hondt, S. (2018) Minimal influence of extracellular DNA on molecular surveys of marine sedimentary communities. *Frontiers in Microbiology*, 9: 2969. https://doi.org/10.3389/fmicb.2018.02969.

Randlett, M.È., Coolen, M.J., Stockhecke, M., Pickarski, N., Litt, T., Balkema, C., Kwiecien, O., Tomonaga, Y., Wehrli, B. and Schubert, C.J. (2014) Alkenone distribution in Lake Van sediment over the last 270 ka: influence of temperature and haptophyte species composition. *Quaternary Science Reviews*, 104: 53-62. https://doi.org/10.1016/j.quascirev.2014.07.009.

Rasmussen, R. (2001) Quantification on the LightCycler instrument. In: Meuer, S., Wittwer, C. and Nakagawara, K. (eds.), *Rapid cycle real-time PCR: Methods and applications* (21-34). Heidelberg: Springer Press.

Ren, Z., Qu, X., Peng, W., Yu, Y. and Zhang, M. (2019) Nutrients drive the structures of bacterial communities in sediments and surface waters in the river-lake system of Poyang Lake. *Water*, 11 (5): 930. https://doi.org/10.3390/w11050930.

Ren, Z., Zhang, C., Li, X., Ma, K. and Cui, B. (2022) Abundant and rare bacterial taxa structuring differently in sediment and water in thermokarst lakes in the Yellow River Source area, Qinghai-Tibet Plateau. *Frontiers in Microbiology*, 13. https://doi.org/10.3389/fmicb.2022.774514.

Reuss, N., Conley, D.J. and Bianchi, T.S. (2005) Preservation conditions and the use of sediment pigments as a tool for recent ecological reconstruction in four Northern European estuaries. *Marine Chemistry*, 95 (3-4): 283-302. https://doi.org/10.1016/j.marchem.2004.10.002.

Rhodes, G., Porter, J. and Pickup, R.W. (2012) The bacteriology of Windermere and its catchment: insights from 70 years of study. *Freshwater Biology*, 57 (2): 305-320. https://doi.org/10.1111/j.1365-2427.2011.02700.x.

Richardson, J., Feuchtmayr, H., Miller, C., Hunter, P.D., Maberly, S.C. and Carvalho, L. (2019) Response of cyanobacteria and phytoplankton abundance to warming, extreme rainfall events and nutrient enrichment. *Global Change Biology*, 25 (10): 3365-3380.

Rietschel, E.T. (1976) Absolute configuration of 3-hydroxy fatty acids present in lipopolysaccharides from various bacterial groups. *European Journal of Biochemistry*, 64 (2): 423-428. https://doi.org/10.1111/j.1432-1033.1976.tb10318.x.

Rietschel, E.T., Lüderitz, O. and Volk, W.A. (1975) Nature, type of linkage, and absolute configuration of (hydroxy) fatty acids in lipopolysaccharides from *Xanthomonas sinensis* and related strains. *Journal of Bacteriology*, 122 (3): 1180-1188. https://doi.org/10.1128/jb.122.3.1180-1188.1975.

Rinta-Kanto, J.M., Ouellette, A.J.A., Boyer, G.L., Twiss, M.R., Bridgeman, T.B. and Wilhelm, S.W. (2005) Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. *Environmental Science and Technology*, 39 (11): 4198-4205. https://doi.org/10.1021/es048249u.

Robinson, S.A., Ruhl, M., Astley, D.L., Naafs, B.D.A., Farnsworth, A.J., Bown, P.R., Jenkyns, H.C., Lunt, D.J., O'Brien, C., Pancost, R.D. and Markwick, P.J. (2017) Early Jurassic North Atlantic sea-surface temperatures from TEX_{86} palaeothermometry. *Sedimentology*, 64 (1): 215-230. https://doi.org/10.1111/sed.12321.

Rohwer, R.R., Hamilton, J.J., Newton, R.J. and McMahon, K.D. (2018) TaxAss: leveraging a custom freshwater database achieves fine-scale taxonomic resolution. *mSphere*, 3 (5): e00327-18. https://doi.org/10.1101/214288.

Romero-Viana, L., Keely, B.J., Camacho, A., Vicente, E. and Miracle, M.R. (2010) Primary production in Lake La Cruz (Spain) over the last four centuries: reconstruction based on sedimentary signal of photosynthetic pigments. *Journal of Paleolimnology*, 43: 771-786. https://doi.org/10.1007/s10933-009-9367-y.

Ron, E.Z. and Rosenberg, E. (2001) Natural roles of biosurfactants: Minireview. *Environmental Microbiology*, 3 (4): 229-236. https://doi.org/10.1046/j.1462-2920.2001.00190.x.

Rösel, S., Allgaier, M. and Grossart, H.P. (2012) Long-term characterization of free-living and particle-associated bacterial communities in Lake Tiefwaren reveals distinct seasonal patterns. *Microbial Ecology*, 64 (3): 571-583. https://doi.org/10.1007/s00248-012-0049-3.

Rosenfelder, G., Lüderetz, O. and Westphal, O. (1974) Composition of Lipopolysaccharides from *Myxococcus fulvus* and Other Fruiting and Non-Fruiting Myxobacteria. *European Journal of Biochemistry*, 44 (2): 411-420. https://doi.org/10.1111/j.1432-1033.1974.tb03499.x.

Russell, N.J. and Fukunaga, N. (1990) A comparison of thermal adaptation of membrane lipids in psychrophilic and thermophilic bacteria. *FEMS Microbiology Reviews*, 6 (2-3): 171-182. https://doi.org/10.1111/j.1574-6968.1990.tb04093.x.

Ruzicka, S., Edgerton, D., Norman, M. and Hill, T. (2000) The utility of ergosterol as a bioindicator of fungi in temperate soils. *Soil Biology and Biochemistry*, 32 (7): 989-1005. https://doi.org/10.1016/S0038-0717(00)00009-2.

Saito, H. and Suzuki, N. (2007) Distributions and sources of hopanes, hopanoic acids and hopanols in Miocene to recent sediments from ODP Leg 190, Nankai Trough. *Organic Geochemistry*, 38 (10): 1715-1728. https://doi.org/10.1016/j.orggeochem.2007.05.012.

Sampath, V. (2018) Bacterial endotoxin-lipopolysaccharide; structure, function and its role in immunity in vertebrates and invertebrates. *Agriculture and Natural Resources*, 52 (2): 115-120. https://doi.org/10.1016/j.anres.2018.08.002.

Sanger, J.E. (1988) Fossil pigments in paleoecology and paleolimnology. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 62 (1-4): 343-359. https://doi.org/10.1016/0031-0182(88)90061-2.

Sarmento, H. (2012) New paradigms in tropical limnology: the importance of the microbial food web. *Hydrobiologia*, 686 (1): 1-14. https://doi.org/10.1007/s10750-012-1011-6.

Schouten, S., Hopmans, E.C., Schefuß, E. and Damste, J.S.S. (2002) Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth and Planetary Science Letters*, 204 (1-2): 265-274. https://doi.org/10.1016/S0012-821X(02)00979-2.

Schwark, L. and Empt, P. (2006) Sterane biomarkers as indicators of Palaeozoic algal evolution and extinction events. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 240 (1-2): 225-236. https://doi.org/10.1016/j.palaeo.2006.03.050.

Schwarz, M., Köpcke, B., Weber, R.W., Sterner, O. and Anke, H. (2004) 3-Hydroxypropionic acid as a nematicidal principle in endophytic fungi. *Phytochemistry*, 65 (15): 2239-2245. https://doi.org/10.1016/j.phytochem.2004.06.035.

Segawa, Y., Yamamoto, M., Kuwae, M., Moriya, K., Suzuki, H. and Suzuki, K. (2022) Reconstruction of the eukaryotic communities in Beppu Bay over the past 50 years based on sedimentary DNA barcoding. *Journal of Geophysical Research: Biogeosciences*, 127 (6): e2022JG006825. https://doi.org/10.1029/2022JG006825.

Selway, C.A., Armbrecht, L. and Thornalley, D. (2022) An outlook for the acquisition of marine sedimentary ancient DNA (sedaDNA) from north Atlantic Ocean archive material. *Paleoceanography and Paleoclimatology*, 37 (5): e2021PA004372. https://doi.org/10.1029/2021PA004372.

Shade, A., Kent, A.D., Jones, S.E., Newton, R.J., Triplett, E.W. and McMahon, K.D. (2007) Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnology and Oceanography*, 52 (2): 487-494. https://doi.org/10.4319/lo.2007.52.2.0487.

Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Bürgmann, H., Huber, D.H., Langenheder, S., Lennon, J.T., Martiny, J.B. and Matulich, K.L. (2012) Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, 3: 417. https://doi.org/10.3389/fmicb.2012.00417.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13 (11): 2498-504. https://doi.org/10.1101/gr.1239303.

Shenhav, L., Thompson, M., Joseph, T.A., Briscoe, L., Furman, O., Bogumil, D., Mizrahi, I., Pe'er, I. and Halperin, E. (2019) FEAST: fast expectation-maximization for microbial source tracking. *Nature Methods*, 16 (7): 627-632. https://doi.org/10.1038/s41592-019-0431-x.

Shi, L., Cai, Y., Kong, F. and Yu, Y. (2012). Specific association between bacteria and buoyant Microcystis colonies compared with other bulk bacterial communities in the eutrophic Lake Taihu, China. *Environmental Microbiology Reports*, 4 (6): 669-678. https://doi.org/10.1111/1758-2229.12001.

Silhavy, T.J., Kahne, D. and Walker, S. (2010) The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*, 2 (5): a000414. https://doi.org/10.1101/cshperspect.a000414.

Siliakus, M.F., van der Oost, J. and Kengen, S.W. (2017) Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles*, 21: 651-670. https://doi.org/10.1007/s00792-017-0939-x.

Simpson, B.W. and Trent, M.S. (2019) Pushing the envelope: LPS modifications and their consequences. *Nature Reviews Microbiology*, 17 (7): 403-416. https://doi.org/10.1038/s41579-019-0201-x.

Simpson, G.L. (2018) Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*, 6: 149. https://doi.org/10.3389/fevo.2018.00149.

Sinninghe Damsté, J.S., Ossebaar, J., Schouten, S. and Verschuren, D. (2012) Distribution of tetraether lipids in the 25-ka sedimentary record of Lake Challa: extracting reliable TEX_{86} and MBT/CBT palaeotemperatures from an equatorial African lake. *Quaternary Science Reviews*, 50: 43-54. https://doi.org/10.1016/j.quascirev.2012.07.001.

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. *Organic Geochemistry*, 124: 63-76. https://doi.org/10.1016/j.orggeochem.2018.07.006.

Sitnikova, T., Michel, E., Tulupova, Y., Khanaev, I., Parfenova, V. and Prozorova, L., 2012. Spirochetes in gastropods from Lake Baikal and North American freshwaters: New multi-family, multi-habitat host records. *Symbiosis*, 56 (3): 103-110. https://doi.org/10.1007/s13199-012-0167-1.

Sj gren, J., Magnusson, J., Broberg, A., Schnürer, J. and Kenne, L. (2003) Antifungal 3-hydroxy fatty acids from *Lactobacillus plantarum* MiLAB 14. *Applied and Environmental Microbiology*, 69 (12): 7554-7557. https://doi.org/10.1128/AEM.69.12.7554-7557.2003.

Skerratt, J.H., Nichols, P.D., Bowman, J.P. and Sly, L.I. (1992) Occurrence and significance of long-chain (ω-1)-hydroxy fatty acids in methane-utilizing bacteria. *Organic Geochemistry*, 18 (2): 189-194. https://doi.org/10.1016/0146-6380(92)90129-L.

Slessarev, E.W., Lin, Y., Bingham, N.L., Johnson, J.E., Dai, Y., Schimel, J.P. and Chadwick, O.A. (2016) Water balance creates a threshold in soil pH at the global scale. *Nature*, 540 (7634): 567-569. https://doi.org/10.1038/nature20139.

Smol, J.P. (1988) Chrysophycean microfossils in paleolimnological studies. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 62 (1-4): 287-297. https://doi.org/10.1016/0031-0182(88)90058-2.

Song, D., Huo, T., Zhang, Z., Cheng, L., Wang, L., Ming, K., Liu, H., Li, M. and Du, X. (2022) Metagenomic analysis reveals the response of microbial communities and their functions in lake sediment to environmental factors. *International Journal of Environmental Research and Public Health*, 19 (24): 16870. https://doi.org/10.3390/ijerph192416870.

Song, Z., Qin, Y., George, S.C., Wang, L., Guo, J. and Feng, Z. (2013) A biomarker study of depositional paleoenvironments and source inputs for the massive formation of Upper Cretaceous lacustrine source rocks in the Songliao Basin, China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 385: 137-151. https://doi.org/10.1016/j.palaeo.2012.12.007.

Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44 (4): 846-849. https://doi.org/10.1099/00207713-44-4-846.

Stahlschmidt, M.C., Collin, T.C., Fernandes, D.M., Bar-Oz, G., Belfer-Cohen, A., Gao, Z., Jakeli, N., Matskevich, Z., Meshveliani, T., Pritchard, J.K. and McDermott, F. (2019) Ancient mammalian and plant DNA from Late Quaternary stalagmite layers at Solkota Cave, Georgia. *Scientific Reports*, 9 (1): 6628. https://doi.org/10.1038/s41598-019-43147-0.

Stapel, J.G., Schirrmeister, L., Overduin, P.P., Wetterich, S., Strauss, J., Horsfield, B. and Mangelsdorf, K. (2016) Microbial lipid signatures and substrate potential of organic matter in permafrost deposits: Implications for future greenhouse gas production. *Journal of Geophysical Research: Biogeosciences*, 121 (10): 2652-2666. https://doi.org/10.1002/2016JG003483.

Stoof-Leichsenring, K.R., Huang, S., Liu, S., Jia, W., Li, K., Liu, X., Pestryakova, L.A. and Herzschuh, U. (2022) Sedimentary DNA identifies modern and past macrophyte diversity and its environmental drivers in high-latitude and high-elevation lakes in Siberia and China. *Limnology and Oceanography*. 67 (5): 1126-1141. https://doi.org/10.1002/lno.12061.

Straile, D., Jochimsen, M.C. and Kümmerlin, R. (2013) The use of long-term monitoring data for studies of planktonic diversity: a cautionary tale from two Swiss lakes. *Freshwater Biology*, 58 (6): 1292-1301. https://doi.org/10.1111/fwb.12118.

Summons, R.E. and Lincoln, S.A. (2012) Biomarkers: Informative molecules for studies in geobiology. In: Knoll, A.H., Canfield, D.E. and Konhauser, K.O. (eds), *Fundamentals of Geobiology*, Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118280874.ch15.

Sun, R., Tu, Z., Fan, L., Qiao, Z., Liu, X., Hu, S., Zheng, G., Wu, Y., Wang, R. and Mi, X. (2020) The correlation analyses of bacterial community composition and spatial factors between freshwater and sediment in Poyang Lake wetland by using artificial neural network (ANN) modelling. *Brazilian Journal of Microbiology*, 51: 1191-1207. https://doi.org/10.1007/s42770-020-00285-2.

Sutcliffe D.W., Carrick T.R., Heron J., Rigg E., Talling J.F., Woof C. and Lund, J.W.G. (1982) Long-term and seasonal changes in the chemical composition of precipitation and surface waters of lakes and tarns in the English Lake District. *Freshwater Biology*, 12: 451–506. https://doi.org/10.1111/j.1365-2427.1982.tb00640.x.

Suzuki, M.T., Taylor, L.T. and DeLong, E.F. (2000) Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5′-nuclease assays. *Applied and Environmental Microbiology*, 66 (11): 4605-4614. https://doi.org/10.1128/aem.66.11.4605-4614.2000.

Swan, B.K., Ehrhardt, C.J., Reifel, K.M., Moreno, L.I. and Valentine, D.L. (2010) Archaeal and bacterial communities respond differently to environmental gradients in anoxic sediments of a California hypersaline lake, the Salton Sea. *Applied and Environmental Microbiology*, 76 (3): 757-768. https://doi.org/10.1128/AEM.02409-09.

Szponar, B., Larsson, L. and Domagała-Kulawik, J. (2012) Endotoxin markers in bronchoalveolar lavage fluid of patients with interstitial lung diseases. *Multidisciplinary Respiratory Medicine*, 7: 1-6. https://doi.org/10.1186/2049-6958-7-54.

Talling, J.F. and Heaney, S.I. (1983) Evaluation of historical data on the nutrient status of Esthwaite Water, Cumbria. *Report to Nature Conservancy Council*. Freshwater Biological Association, Ambleside.

Talling, J.F. and Heaney, S.I. (2015) Novel tests of regular seasonality, types of variability, and modes of succession in lake phytoplankton. *Inland Waters*, 5 (4): 331-338. https://doi.org/10.5268/IW-5.4.768.

Tammert, H., Tšertova, N., Kiprovskaja, J., Baty, F., Nõges, T. and Kisand, V. (2015) Contrasting seasonal and interannual environmental drivers in bacterial communities within a large shallow lake: evidence from a seven year survey. *Aquatic Microbial Ecology*, 75 (1): 43-54. https://doi.org/10.3354/ame01744.

Tani, Y., Matsumoto, G.I., Soma, M., Soma, Y., Hashimoto, S. and Kawai, T. (2009) Photosynthetic pigments in sediment core HDP-04 from Lake Hovsgol, Mongolia, and their implication for changes in algal productivity and lake environment for the last 1 Ma. *Quaternary International*, 205 (1-2): 74-83. https://doi.org/10.1016/j.quaint.2009.02.007.

Tessler, M., Cunningham, S.W., Ingala, M.R., Warring, S.D. and Brugler, M.R. (2023) An environmental DNA primer for microbial and restoration ecology. *Microbial Ecology*, 1-13. https://doi.org/10.1007/s00248-022-02168-5.

Thackeray, S.J., Jones, I.D. and Maberly, S.C. (2008) Long-term change in the phenology of spring phytoplankton: Species-specific responses to nutrient enrichment and climatic change. *Journal of Ecology*, 96 (3): 523-535. https://doi.org/10.1111/j.1365-2745.2008.01355.x.

Thomas, S.P., Shanmuganathan, B., Jaiswal, M.K., Kumaresan, A. and Sadasivam, S.K. (2019) Legacy of a Pleistocene bacterial community: Patterns in community dynamics through changing ecosystems. *Microbiological Research*, 226: 65-73. https://doi.org/10.1016/j.micres.2019.06.001.

Thorpe, A. C., Anderson, A., Goodall, T., Thackeray, S. J., Maberly, S. C., Bendle, J. A., Gweon, H. S. and Read, D. S. (2022) Sedimentary DNA records long-term changes in a lake bacterial community in response to varying nutrient availability. *Environmental DNA*, 4 (6), 1340-1355. https://doi.org/10.1002/edn3.344.

Thupaki, P., Phanikumar, M.S., Schwab, D.J., Nevers, M.B. and Whitman, R.L. (2013) Evaluating the role of sediment-bacteria interactions on *Escherichia coli* concentrations at beaches in southern Lake Michigan. *Journal of Geophysical Research*: Oceans, 118 (12): 7049-7065. https://doi.org/10.1002/2013JC008919.

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. *Organic Geochemistry*, 40 (9): 1032-1036. https://doi.org/10.1016/j.orggeochem.2009.04.014.

Torti, A., Lever, M.A. and Jørgensen, B.B. (2015) Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, 24: 185-196. https://doi.org/10.1016/j.margen.2015.08.007.

Tse, T.J., Doig, L.E., Tang, S., Zhang, X., Sun, W., Wiseman, S.B., Feng, C.X., Liu, H., Giesy, J.P., Hecker, M. and Jones, P.D. (2018) Combining high-throughput sequencing of sedaDNA and traditional paleolimnological techniques to infer historical trends in cyanobacterial communities. *Environmental Science and Technology*, 52 (12): 6842-6853. https://doi.org/10.1021/acs.est.7b06386.

Tsugeki, N., Nakane, K., Ochi, N. and Kuwae, M. (2022) Reconstruction of 100-year dynamics in Daphnia spawning activity revealed by sedimentary DNA. *Scientific Reports*, 12 (1): 1-11. https://doi.org/10.1038/s41598-021-03899-0.

Tyagi, P., Yamamoto, S. and Kawamura, K. (2015) Hydroxy fatty acids in fresh snow samples from northern Japan: long-range atmospheric transport of gram-negative bacteria by Asian winter monsoon. *Biogeosciences*, 12 (23): 7071-7080. https://doi.org/10.5194/bg-12-7071-2015.

Van Dyk, M.S., Kock, J.L.F. and Botha, A. (1994) Hydroxy long-chain fatty acids in fungi. *World Journal of Microbiology and Biotechnology*, 10: 495-504. https://doi.org/10.1007/BF00367653.

Véquaud, P., Derenne, S., Thibault, A., Anquetil, C., Bonanomi, G., Collin, S., Contreras, S., Nottingham, A.T., Sabatier, P., Salinas, N. and Scott, W.P. (2021) Development of global temperature and pH calibrations based on bacterial 3-hydroxy fatty acids in soils. *Biogeosciences*, 18 (12): 3937-3959. https://doi.org/10.5194/bg-18-3937-2021.

Verleyen, E., Sabbe, K., Vynerman, W., Nicosia, C. (2017) Siliceous microfossils from single-celled organisms: Diatoms and Chrysophycean stomatocysts, In: Nicosia, C. and Stoops, G. (eds.), *Archaeological Soil and Sediment Micromorphology*, New Jersey: John Wiley and Sons. https://doi.org/10.1002/9781118941065.ch19.

Vipindas, P.V., Mujeeb, R.K.M., Jabir, T., Thasneem, T.R. and Hatha, A.M. (2020) Diversity of sediment bacterial communities in the South Eastern Arabian Sea. *Regional Studies in Marine Science*, 35: 101153. https://doi.org/10.1016/j.rsma.2020.101153.

Volkman, J.K., Barrett, S.M. and Blackburn, S.I. (1999) Fatty acids and hydroxy fatty acids in three species of freshwater eustigmatophytes. *Journal of Phycology*, 35 (5): 1005-1012. https://doi.org/10.1046/j.1529-8817.1999.3551005.x.

Volkman, J.K., Barrett, S.M., Blackburn, S.I., Mansour, M.P., Sikes, E.L. and Gelin, F. (1998) Microalgal biomarkers: a review of recent research developments. *Organic Geochemistry*, 29 (5-7): 1163-1179. https://doi.org/10.1016/S0146-6380(98)00062-X.

Volkman, J.K., Smittenberg, R.H. (2017) Lipid biomarkers as organic geochemical proxies for the paleoenvironmental reconstruction of estuarine environments. In: Weckström, K., Saunders, K., Gell, P., Skilbeck, C. (eds), *Applications of Paleoenvironmental Techniques in Estuarine Studies*. Dordrecht: Springer. https://doi.org/10.1007/978-94-024-0990-1_8.

Vuillemin, A., Horn, F., Alawi, M., Henny, C., Wagner, D., Crowe, S.A. and Kallmeyer, J. (2017) Preservation and significance of extracellular DNA in ferruginous sediments from Lake Towuti, Indonesia. *Frontiers in Microbiology*, 8: 1440. https://doi.org/10.3389/fmicb.2017.01440.

Vuorio, K., Mäki, A., Salmi, P., Aalto, S.L. and Tiirola, M. (2020) Consistency of targeted metatranscriptomics and morphological characterization of phytoplankton communities. *Frontiers in Microbiology*, 11: 96. https://doi.org/10.3389/fmicb.2020.00096.

Wakeham, S.G., Pease, T.K. and Benner, R. (2003) Hydroxy fatty acids in marine dissolved organic matter as indicators of bacterial membrane material. *Organic Geochemistry*, 34 (6): 857-868. https://doi.org/10.1016/S0146-6380(02)00189-4.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J.A., Jansson, J.K., Caporaso, J.G., Fuhrman, J.A. and Apprill, A. (2016) Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems*, 1 (1): e00009-15. https://doi.org/10.1128/mSystems.00009-15.

Wan, Y., Ruan, X., Zhang, Y. and Li, R. (2017) Illumina sequencing-based analysis of sediment bacteria community in different trophic status freshwater lakes. *Microbiology Open*, 6 (4): e00450. https://doi.org/10.1002/mbo3.450.

Wang, C., Bendle, J., Yang, Y., Yang, H., Sun, H., Huang, J. and Xie, S. (2016) Impacts of pH and temperature on soil bacterial 3-hydroxy fatty acids: Development of novel terrestrial proxies. *Organic Geochemistry*, 94: 21-31. https://doi.org/10.1016/j.orggeochem.2016.01.010.

Wang, C., Bendle, J.A., Yang, H., Yang, Y., Hardman, A., Yamoah, A., Thorpe, A., Mandel, I., Greene, S.E., Huang, J. and Xie, S. (2021) Global calibration of novel 3-hydroxy fatty acid based temperature and pH proxies. *Geochimica et Cosmochimica Acta*, 302: 101-119. https://doi.org/10.1016/j.gca.2021.03.010.

Wang, C., Bendle, J.A., Zhang, H., Yang, Y., Liu, D., Huang, J., Cui, J. and Xie, S. (2018) Holocene temperature and hydrological changes reconstructed by bacterial 3-hydroxy fatty acids in a stalagmite from central China. *Quaternary Science Reviews*, 192: 97-105. https://doi.org/10.1016/j.quascirev.2018.05.030.

Wang, C.Y., Zhou, X., Guo, D., Zhao, J.H., Yan, L., Feng, G.Z., Gao, Q., Yu, H. and Zhao, L.P. (2019) Soil pH is the primary factor driving the distribution and function of microorganisms in farmland soils in north-eastern China. *Annals of Microbiology*, 69 (13): 1461-1473. https://doi.org/10.1007/s13213-019-01529-9.

Wang, P., Li, Q., Ge, F., Li, F., Liu, Y., Deng, S., Zhang, D. and Tian, J. (2022) Correlation of bacterial community with phosphorus fraction drives discovery of Actinobacteria involved soil phosphorus transformation during the trichlorfon degradation. *Environmental Pollution*, 302: 119043. https://doi.org/10.1016/j.envpol.2022.119043.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73 (16): 5261-5267. https://doi.org/10.1128/AEM.00062-07.

Warny, S., Askin, R.A., Hannah, M.J., Mohr, B.A., Raine, J.I., Harwood, D.M., Florindo, F. and SMS Science Team (2009) Palynomorphs from a sediment core reveal a sudden remarkably warm Antarctica during the middle Miocene. *Geology*, 37 (10): 955-958. https://doi.org/10.1130/G30139A.1.

Watanabe, K., Park, H.D. and Kumon, F. (2012) Historical change of phytoplankton in a eutrophic lake in Japan as determined by analysis of photosynthetic pigments in a lakebed sediment core. *Environmental Earth Sciences*, 66 (8): 2293-2300. https://doi.org/10.1007/s12665-011-1452-7.

Weijers, J.W., Schouten, S., van den Donker, J.C., Hopmans, E.C. and Sinninghe Damsté, J.S. (2007) Environmental controls on bacterial tetraether membrane lipid distribution in soils. *Geochimica et Cosmochimica Acta*, 71 (3): 703-713. https://doi.org/10.1016/j.gca.2006.10.003.

Weisse, T., Müller, H., Pinto-Coelho, R.M., Schweizer, A., Springmann, D. and Baldringer, G. (1990) Response of the microbial loop to the phytoplankton spring bloom in a large prealpine lake. *Limnology and Oceanography*, 35 (4): 781-794. https://doi.org/10.4319/lo.1990.35.4.0781.

Welch, D.F. (1991) Applications of cellular fatty acid analysis. *Clinical Microbiology Reviews*, 4 (4): 422-438. https://doi.org/10.1128/CMR.4.4.422.

Wilkinson, S. G. (1988) Gram-negative bacteria. In: Ratledge, C., and Wilkinson, S. G. (eds.), *Microbial Lipids* (199-488). New York: Academic Press.

Wilkinson, S.G., Galbraith, L. and Lightfoot, G.A. (1973) Cell walls, lipids, and lipopolysaccharides of *Pseudomonas* species. *European Journal of Biochemistry*, 33 (1): 158-174. https://doi.org/10.1111/j.1432-1033.1973.tb02666.x.

Willis, K.J. and Bhagwat, S.A. (2010) Questions of importance to the conservation of biological diversity: answers from the past. *Climate of the Past*, 6 (6): 759-769. https://doi.org/10.5194/cp-6-759-2010.

Willis, K.J., Bailey, R.M., Bhagwat, S.A. and Birks, H.J.B. (2010) Biodiversity baselines, thresholds and resilience: testing predictions and assumptions using palaeoecological data. *Trends in Ecology and Evolution*, 25 (10): 583-591. https://doi.org/10.1016/j.tree.2010.07.006.

Wingard, G.L., Bernhardt, C.E. and Wachnicka, A.H. (2017) The role of paleoecology in restoration and resource management – the past as a guide to future decision-making: review and example from the Greater Everglades Ecosystem, USA. *Frontiers in Ecology and Evolution*, 5: 11. https://doi.org/10.3389/fevo.2017.00011.

Wollenweber, H.W., Rietschel, E.T., Hofstad, T., Weintraub, A., and Lindberg, A.A. (1980) Nature, type of linkage, quantity, and absolute configuration of (3-hydroxy) fatty acids in lipopolysaccharides from *Bacteroides fragilis* NCTC 9343 and related strains. *Journal of Bacteriology*, 144 (3): 898-903. https://doi.org/10.1128/jb.144.3.898-903.1980.

Wood, S.N. (2020) mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. https://cran. rproject.org/web/packages/mgcv/index.html.

Woodman, M. E. (2008) Direct PCR of intact bacteria (colony PCR). *Current Protocols in Microbiology*, 9 (1): A.3D.1–A.3D.6. https://doi.org/10.1002/9780471729259.mca03ds9.

Wright, E.S. (2016) Using DECIPHER v2.0 to analyse big biological sequence data in R. The R Journal, 8 (1): 352-459.

Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N. and Junier, P. (2014) Endospore-enriched sequencing approach reveals unprecedented diversity of Firmicutes in sediments. *Environmental Microbiology Reports*, 6 (6): 631-639. https://doi.org/10.1111/1758-2229.12179.

Wurzbacher, C., Fuchs, A., Attermeyer, K., Frindte, K., Grossart, H.P., Hupfer, M., Casper, P. and Monaghan, M.T. (2017) Shifts among eukaryota, bacteria, and archaea define the vertical organization of a lake sediment. *Microbiome*, 5 (1): 1-16. https://doi.org/10.1186/s40168-017-0255-9.

Xiao, X., Sogge, H., Lagesen, K., Tooming-Klunderud, A., Jakobsen, K.S. and Rohrlack, T. (2014) Use of high throughput sequencing and light microscopy show contrasting results in a study of phytoplankton occurrence in a freshwater environment. *PloS One*, 9 (8): e106510. https://doi.org/10.1371/journal.pone.0106510.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., Yang, Y., Yao, T., Knight, R. and Chu, H. (2012) Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environmental Microbiology*, 14 (9): 2457-2466. https://doi.org/10.1111/j.1462-2920.2012.02799.x.

Xu, Y. and Jaffé, R. (2008) Biomarker-based paleo-record of environmental change for a eutrophic, tropical freshwater lake, Lake Valencia, Venezuela. *Journal of Paleolimnology*, 40: 179-194. https://doi.org/10.1007/s10933-007-9150-x.

Xu, Y., Han, G., Zhang, H., Yu, Z. and Liu, R. (2022) Application of fast expectation-maximization microbial source tracking to discern fecal contamination in rivers exposed to low fecal inputs. *Journal of Microbiology*, 60 (6): 594-601. https://doi.org/10.1007/s12275-022-1651-9.

Yabuuchi, E., Kawamura, Y., Ezaki, T., Ikedo, M., Dejsirilert, S., Fujiwara, N., Naka, T. and Kobayashi, K. (2000) *Burkholderia uboniae* sp. nov., L-arabinose-assimilating but different from *Burkholderia thailandensis* and *Burkholderia vietnamiensis*. *Microbiology and Immunology*, 44 (4): 307-317.

Yang, Y., Wang, C., Bendle, J.A., Luo, Z., Dang, X., Xue, J., Xiang, X. and Xie, S. (2021) Appraisal of paleoclimate indices based on bacterial 3-hydroxy fatty acids in 20 Chinese alkaline lakes. *Organic Geochemistry*, 160: 104277. https://doi.org/10.1016/j.orggeochem.2021.104277.

Yang, Y., Wang, C., Bendle, J.A., Yu, X., Gao, C., Lü, X., Ruan, X., Wang, R. and Xie, S. (2020) A new sea surface temperature proxy based on bacterial 3-hydroxy fatty acids. *Organic Geochemistry*, 141: 103975. https://doi.org/10.1016/j.orggeochem.2020.103975.

Yano, I., Ohno, Y., Masui, M., Kato, K., Yabuuchi, E. and Ohyama, A. (1976) Occurrence of 2-and 3-hydroxy fatty acids in high concentrations in the extractable and bound lipids of *Flavobacterium meningosepticum* and *Flavobacterium* IIb. *Lipids*, 11: 685-688. https://doi.org/10.1007/BF02532887.

Yao, M., Rui, J., Niu, H., Heděnec, P., Li, J., He, Z., Wang, J., Cao, W. and Li, X. (2017) The differentiation of soil bacterial communities along a precipitation and temperature gradient in the eastern Inner Mongolia steppe. *Catena*, 152: 47-56. https://doi.org/10.1016/j.catena.2017.01.007.

Youssef, N.H. and Elshahed, M.S. (2009) Diversity rankings among bacterial lineages in soil. *The ISME journal*, 3 (3): 305-313. https://doi.org/10.1038/ismej.2008.106.

Youssef, N.H., Duncan, K.E. and McInerney, M.J. (2005) Importance of 3-hydroxy fatty acid composition of lipopeptides for biosurfactant activity. *Applied and Environmental Microbiology*, 71 (12): 7690-7695. https://doi.org/10.1128/AEM.71.12.7690-7695.2005.

Zak, D.R., Holmes, W.E., White, D.C., Peacock, A.D. and Tilman, D., 2003. Plant diversity, soil microbial communities, and ecosystem function: are there any links? *Ecology*, 84 (8): 2042-2050. https://doi.org/10.1890/02-0433.

Zelles, L. (1999) Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: a review. *Biology and Fertility of Soils*, 29: 111-129. https://doi.org/10.1007/s003740050533.

Zelles, L., Bai, Q.Y., Beck, T. and Beese, F. (1992) Signature fatty acids in phospholipids and lipopolysaccharides as indicators of microbial biomass and community structure in agricultural soils. *Soil Biology and Biochemistry*, 24 (4): 317-323. https://doi.org/10.1016/0038-0717(92)90191-Y.

Zhang, H., Huo, S., Wang, R., Xiao, Z., Li, X. and Wu, F. (2021b) Hydrologic and nutrient-driven regime shifts of cyanobacterial and eukaryotic algal communities in a large shallow lake: Evidence from empirical state indicator and ecological network analyses. *Science of The Total Environment*, 783: 147059. https://doi.org/10.1016/j.scitotenv.2021.147059.

Zhang, H., Huo, S., Xiao, Z., He, Z., Yang, J., Yeager, K.M., Li, X. and Wu, F. (2021a) Climate and nutrient-driven regime shifts of cyanobacterial communities in low-latitude plateau lakes. *Environmental Science and Technology*, 55 (5): 3408-3418. https://doi.org/10.1021/acs.est.0c05234.

Zhang, J., Yang, Y., Zhao, L., Li, Y., Xie, S. and Liu, Y. (2015) Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Applied Microbiology and Biotechnology*, 99: 3291-3302. https://doi.org/10.1007/s00253-014-6262-x.

Zhang, L., Shen, T., Cheng, Y., Zhao, T., Li, L. and Qi, P. (2020) Temporal and spatial variations in the bacterial community composition in Lake Bosten, a large, brackish lake in China. *Scientific Reports*, 10 (1): 1-10. https://doi.org/10.1038/s41598-019-57238-5.

Zhang, W., Gu, J., Li, Y., Lin, L., Wang, P., Wang, C., Qian, B., Wang, H., Niu, L., Wang, L. and Zhang, H. (2019) New insights into sediment transport in interconnected river–lake systems through tracing microorganisms. *Environmental Science and Technology*, 53 (8): 4099-4108. https://doi.org/10.1021/acs.est.8b07334.

Zhong, M., Capo, E., Zhang, H., Hu, H., Wang, Z., Tian, W., Huang, T. and Bertilsson, S. (2022) Homogenisation of water and sediment bacterial communities in a shallow lake (lake Balihe, China). *Freshwater Biology*, 68 (1): 155-171. https://doi.org/10.1111/fwb.14016.

Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., Qin, Y., Xue, K., Wu, L., He, Z. and Voordeckers, J.W. (2016) Temperature mediates continental-scale diversity of microbes in forest soils. *Nature Communications*, 7: 12083. https://doi.org/10.1038/ncomms12083.

Zingel, P., Cremona, F., Nõges, T., Cao, Y., Neif, É.M., Coppens, J., Işkın, U., Lauridsen, T.L., Davidson, T.A., Søndergaard, M. and Beklioglu, M. (2018) Effects of warming and nutrients on the microbial food web in shallow lake mesocosms. *European Journal of Protistology*, 64: 1-12. https://doi.org/10.1016/j.ejop.2018.03.001.

Zogg, G.P., Zak, D.R., Ringelberg, D.B., White, D.C., MacDonald, N.W. and Pregitzer, K.S. (1997) Compositional and functional shifts in microbial communities due to soil warming. *Soil Science Society of America Journal*, 61 (2): 475-481. https://doi.org/10.2136/sssaj1997.03615995006100020015x.

Zolitschka, B., Francus, P., Ojala, A.E. and Schimmelmann, A. (2015) Varves in lake sediments – a review. *Quaternary Science Reviews*, 117: 1-41. https://doi.org/10.1016/j.quascirev.2015.03.019.

Zonneveld, K.A., Versteegh, G.J., Kasten, S., Eglinton, T.I., Emeis, K.C., Huguet, C., Koch, B.P., de Lange, G.J., de Leeuw, J.W., Middelburg, J.J. and Mollenhauer, G. (2010) Selective preservation of organic matter in marine environments; processes and impact on the sedimentary record. *Biogeosciences*, 7 (2): 483-511. https://doi.org/10.5194/bg-7-483-2010.

Zulkefli, N.S., Kim, K.H. and Hwang, S.J. (2019) Effects of microbial activity and environmental parameters on the degradation of extracellular environmental DNA from a eutrophic lake. *International Journal of Environmental Research and Public Health*, 16 (18): 3339. https://doi.org/10.3390/ijerph16183339.

Zwart, G., Crump, B.C., Kamst-van Agterveld, M.P., Hagen, F. and Han, S.K. (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology*, 28 (2): 141-155. https://doi.org/10.3354/AME028141.

Appendix A: Supplementary information to Chapter 2

Figure S1. Map of Esthwaite Water in the Lake District, UK, showing the location of the coring and water sampling site (red circle).

Table S1. Fallout radionuclide concentrations in the 2014 reference sediment core.

Depth		Total ^{210}Pb		Unsupported ^{210}Pb		Supported ^{210}Pb		^{137}Cs		^{241}Am	
cm	g cm^{-2}	Bq kg^{-1}	SD	Bq kg^{-1}	SD	Bq kg^{-1}	SD	Bq kg^{-1}	SD	Bq kg^{-1}	SD
0.25	0.01	418.3	15.8	357.0	16.1	61.4	3.1	46.0	2.4	0.0	0.0
2.25	0.16	512.5	16.5	443.6	16.8	68.9	3.1	52.1	2.6	0.0	0.0
4.25	0.38	513.4	16.2	431.5	16.5	81.9	3.4	64.1	2.8	0.0	0.0
6.25	0.60	501.1	19.9	436.5	20.2	64.6	3.5	61.5	2.9	0.0	0.0
8.25	0.81	482.4	21.1	403.0	21.5	79.4	4.4	72.2	3.7	0.0	0.0
10.25	1.03	351.3	16.6	265.8	17.0	85.4	3.6	128.0	3.7	0.0	0.0
12.25	1.30	305.9	18.0	216.2	18.4	89.7	3.9	137.2	3.9	0.0	0.0
14.25	1.57	296.1	19.5	226.4	20.0	69.7	4.2	106.4	4.1	0.0	0.0
16.25	1.85	273.8	12.9	197.8	13.2	76.0	3.0	108.4	2.8	0.0	0.0
18.25	2.13	233.6	14.8	156.2	15.3	77.3	3.7	121.5	3.7	0.0	0.0
20.25	2.42	185.0	10.6	113.4	10.9	71.6	2.8	129.1	2.9	3.6	1.0
22.25	2.73	213.5	17.9	138.4	18.4	75.1	4.2	150.2	4.4	5.9	1.7
24.25	3.05	127.0	8.7	59.5	9.0	67.5	2.3	41.8	1.8	0.0	0.0
26.25	3.45	133.4	12.6	67.5	13.0	65.9	3.2	28.3	2.5	0.0	0.0
28.25	3.84	102.1	8.5	29.1	8.9	72.9	2.4	12.5	1.4	0.0	0.0
29.75	4.17	76.8	7.3	19.8	7.5	57.0	2.0	7.8	1.1	0.0	0.0

Table S2. ^{210}Pb chronology of the 2014 reference sediment core.

Depth		Chronology			Sedimentation rate		
cm	g cm^{-2}	Year	Age (years from 2014)	SD	g cm^{-2} y^{-1}	cm y^{-1}	SD (%)
0.00	0.00	2014	0	0			
0.25	0.01	2014	0	0	0.058	0.81	5.2
2.25	0.16	2011	3	1	0.051	0.55	4.6
4.25	0.38	2007	7	2	0.043	0.40	4.8
6.25	0.60	2001	13	2	0.037	0.35	5.7
8.25	0.81	1995	19	2	0.034	0.31	6.4
10.25	1.03	1988	26	3	0.042	0.34	7.6
12.25	1.30	1983	31	3	0.064	0.48	9.6
14.25	1.57	1980	34	4	0.071	0.51	10.1
16.25	1.85	1976	38	4	0.064	0.46	9.0
18.25	2.13	1971	43	5	0.070	0.49	12.1
20.25	2.42	1967	47	5	0.073	0.48	13.1
22.25	2.73	1963	51	6	0.038	0.25	17.1
24.25	3.05	1951	63	6	0.024	0.15	17.9
26.25	3.45	1936	78	7	0.024	0.12	17.9
28.25	3.84	1917	97	10	0.024	0.11	17.9
29.75	4.17	1903	111	12	0.024	0.11	17.9

Table S3. Chronology of the 2014 reference sediment core with sample depth corrected to 2016 using the sedimentation rate to allow for comparison with sediment cores collected in 2016.

Corrected depth (cm)	Age (years from 2016)	SD
0.00	0	0
1.62	2	0
1.87	2	0
3.87	5	1
5.87	9	2
7.87	15	2
9.87	21	2
11.87	28	3
13.87	33	3
15.87	36	4
17.87	40	4
19.87	45	5
21.87	49	5
23.87	53	6
25.87	65	6
27.87	80	7
29.87	99	10
31.37	113	12

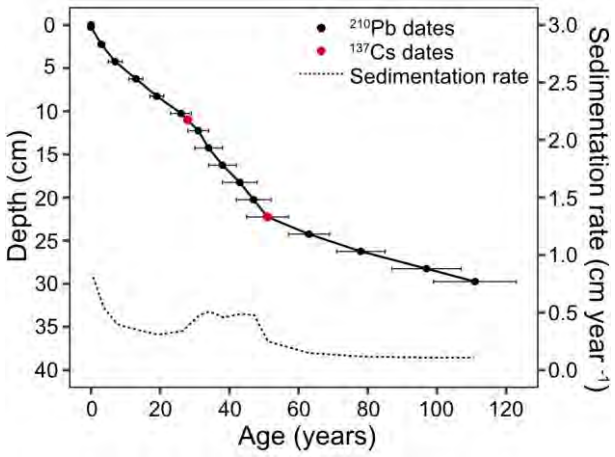

Figure S2. Chronology of the 2014 reference sediment core showing the 1986 and 1963 depths suggested by the ^{137}Cs record, ^{210}Pb dates and sedimentation rate calculated using the ^{137}Cs dates as reference points.

Table S4. Chronology of three sediment cores collected in 2016 estimated using the age-depth model based on the reference sediment core.

Core A			Core B			Core C		
Depth (cm)	Year	SD	Depth (cm)	Year	SD	Depth (cm)	Year	SD
0.4	2016	0	2.4	2013	0	0.5	2015	0
1.4	2014	0	3.4	2012	1	1.5	2014	0
2.4	2014	0	4.4	2010	1	2.5	2013	0
3.4	2012	1	5.4	2008	2	3.5	2012	1
4.4	2010	1	6.4	2005	2	4.5	2010	1
5.4	2008	2	7.4	2002	2	5.5	2008	2
6.4	2005	2	8.4	1999	2	6.5	2005	2
7.4	2002	2	9.4	1996	2	7.5	2002	2
8.4	1999	2	10.4	1993	2	8.5	1999	2
9.4	1996	2	11.4	1990	3	9.5	1996	2
10.4	1993	2	12.4	1987	3	10.5	1993	2
11.4	1989	3	13.4	1984	3	11.5	1989	3
12.4	1987	3	14.4	1982	3	12.5	1986	3
13.4	1984	3	15.4	1981	4	13.5	1984	3
14.4	1982	3	16.4	1979	4	14.5	1982	3
15.4	1981	4	17.4	1977	4	15.5	1981	4
16.4	1979	4	18.4	1975	4	16.5	1979	4
17.4	1977	4	19.4	1972	5	17.5	1977	4
18.4	1975	4	20.4	1970	5	18.5	1974	4
19.4	1972	5	21.4	1968	5	19.5	1972	5
20.4	1970	5	22.4	1966	5	20.5	1970	5
21.4	1968	5	23.4	1964	6	21.5	1968	5
22.4	1966	5	24.4	1960	6	22.5	1966	5
23.4	1964	6	25.4	1954	6	23.5	1964	6
24.4	1960	6	26.4	1947	6	24.5	1959	6
25.4	1954	6	27.4	1940	7	25.5	1953	6
26.4	1947	6	28.4	1931	8	26.5	1946	6
27.4	1941	6	29.4	1921	9	27.5	1939	7
28.4	1931	8	30.4	1912	11	28.5	1930	8
29.4	1922	9	31.4	1903	12	29.5	1921	9

Table S5. Results of the permutation test between the beta diversity Bray-Curtis dissimilarity matrix and sample age and lake physicochemical conditions including alkalinity, nitrate-nitrogen (NO_3-N), winter soluble reactive phosphorus (SRP), total phosphorus (TP), chlorophyll *a*, SRP, ammonium-nitrogen (NH_4-N), surface water temperature and surface water pH.

Variable	R^2	p
Sample age	0.84	***
Alkalinity	0.47	***
NO_3-N	0.28	***
Winter SRP	0.23	**
TP	0.22	**
Chlorophyll *a*	0.18	*
SRP	0.17	*
NH_4-N	0.07	
Temperature	0.02	
pH	0.01	

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S6. BLAST results for sedDNA ASV matches in the FreshTrain database.

sedDNA ASV	FreshTrain match	Percent identity	Phylum	Class	Order	Family	Genus	Species
ASV 19	UncMeth6	97.6	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Crenothrix*	Unassigned
ASV 35	UncMeth6	100	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Methylobacter*	*tundripaludum*
ASV 52	UncMethy	98	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Crenothrix*	Unassigned
ASV 72	JbaZZZZ8	98.8	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	*Methylocystis*	Unassigned
ASV 93	MsFpbF02	97.6	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Rhizobacter*	Unassigned
ASV 102	UncMethy	98	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Methylobacter*	Unassigned
ASV 141	UncuBet8	100	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Methylophilaceae	*Methylotenera*	Unassigned
ASV 152	Betau125	100	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Rhodoferax*	Unassigned
ASV 169	UncMeth6	98.4	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Crenothrix*	Unassigned
ASV 256	CbbYyy08	100	Verrucomicrobia	Verrucomicrobiae	Chthoniobacterales	Terrimicrobiaceae	*FukuN18_freshwater_group*	Unassigned
ASV 264	Bctrm544	100	Verrucomicrobia	Verrucomicrobiae	Chthoniobacterales	Terrimicrobiaceae	*Terrimicrobium*	Unassigned
ASV 283	MspbBpb2	99.2	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	*Flavobacterium*	Unassigned
ASV 290	AbmAc208	100	Actinobacteria	Actinobacteria	Micrococcales	Micrococcaceae	*Pseudarthrobacter*	Unassigned
ASV 292	UncMeth6	97.6	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Crenothrix*	Unassigned
ASV 348	BetauB97	100	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Acidovorax*	Unassigned
ASV 408	JbgZZZZ3	98	Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	*Mycobacterium*	*madagascariense*
ASV 426	MpppcnpE	99.6	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Rhodobacter*	Unassigned
ASV 457	MbGubG02	99.2	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiales_Incertae_Sedis	*unassigned*	Unassigned
ASV 702	ClonLD29	98.8	Verrucomicrobia	Verrucomicrobiae	Chthoniobacterales	Chthoniobacteraceae	*LD29*	Unassigned
ASV 718	UncMeth6	97.2	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonaceae	*Crenothrix*	Unassigned
ASV 784	BcrBa621	97.2	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	*Dinghuibacter*	Unassigned
ASV 930	BctGKS39	97.2	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	*Rhodoblastus*	Unassigned
ASV 948	BcrBac55	97.2	Bacteroidetes	Bacteroidia	Chitinophagales	Chitinophagaceae	*Ferruginibacter*	Unassigned

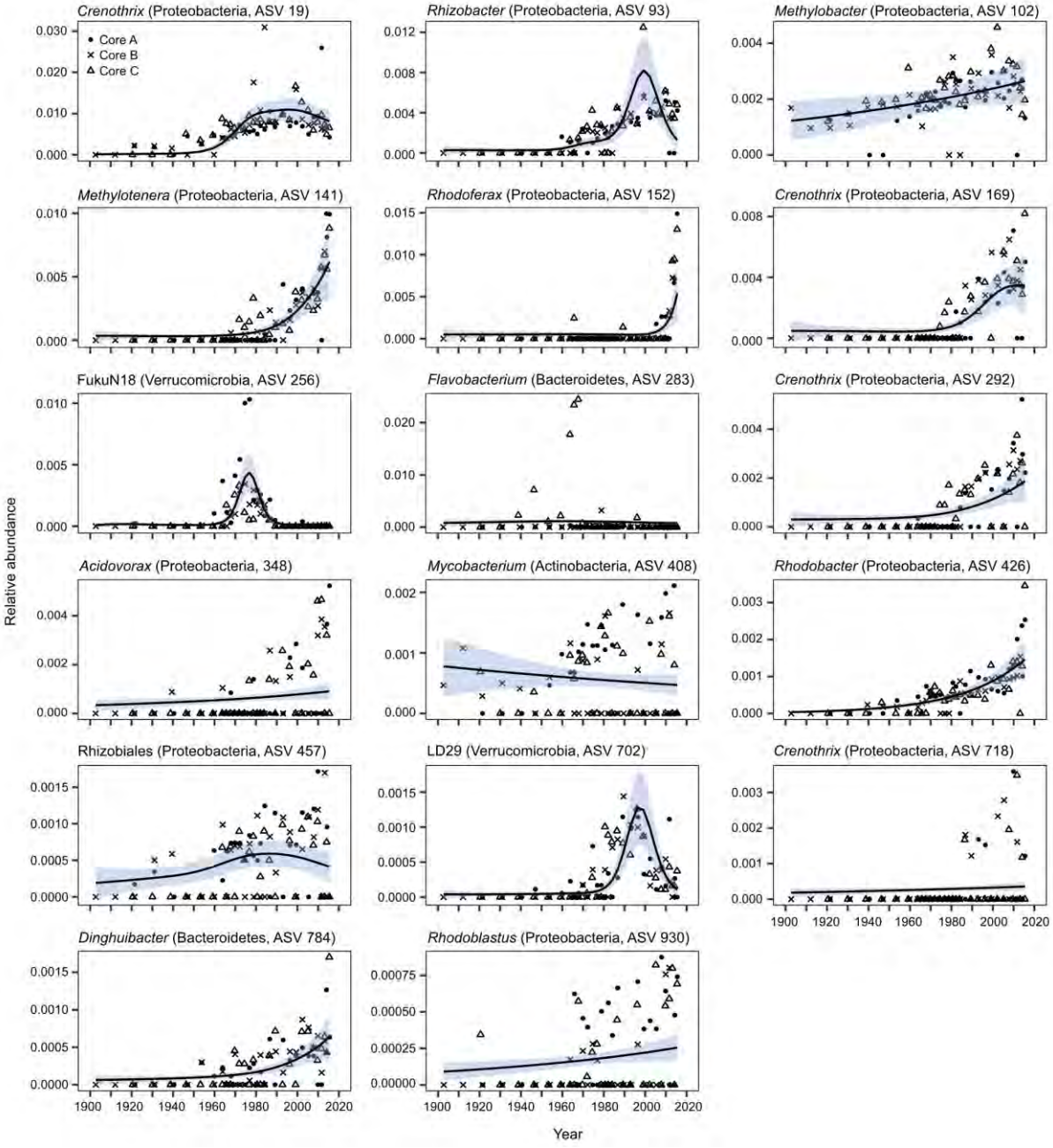

Figure S3. Generalised additive models (GAMs) of the relative abundance of pelagic ASVs throughout three sediment cores. Shaded area shows the 95% confidence interval.

Table S7. Statistics associated with the global (i.e. single smoother) generalised additive models (GAMs) fitted to the relative abundance of each phylum throughout three sediment cores.

Phylum	edf	R^2	Deviance explained (%)	p
Proteobacteria	3.22	0.76	77.0	***
Chloroflexi	4.56	0.92	93.1	***
Firmicutes	3.00	0.26	39.2	***
Bacteroidetes	6.03	0.62	64.3	***
Verrucomicrobia	6.54	0.80	79.1	***
Acidobacteria	3.45	0.61	63.4	***
Cyanobacteria	7.00	0.79	88.8	***
Nitrospirae	5.45	0.51	54.7	***
Spirochaetes	5.55	0.67	68.6	***
Planctomycetes	2.66	0.59	47.3	***

(edf = effective degrees of freedom, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S8. Akaike Information Criterion (AIC) of generalised additive models (GAMs) with a global smooth and with a separate smooth for each sediment core. Delta AIC is the difference between the global smooth AIC and the core-specific smooth AIC.

Phylum	Global smooth		Individual core smooths		Delta AIC
	df	AIC	df	AIC	
Proteobacteria	5.58	-394.07	12.77	-401.02	6.95
Chloroflexi	6.81	-551.51	17.31	-559.05	7.53
Firmicutes	5.61	-310.81	11.88	-331.24	20.43
Bacteroidetes	8.61	-411.56	21.60	-433.51	21.94
Verrucomicrobia	9.04	-643.45	19.59	-634.71	-8.74
Acidobacteria	6.21	-638.30	16.75	-658.10	19.80
Cyanobacteria	9.88	-703.78	21.78	-680.87	-22.91
Nitrospirae	8.59	-667.57	21.71	-701.77	34.20
Spirochaetes	8.64	-733.69	18.36	-739.99	6.30
Planctomycetes	5.32	-710.88	14.96	-729.48	18.61

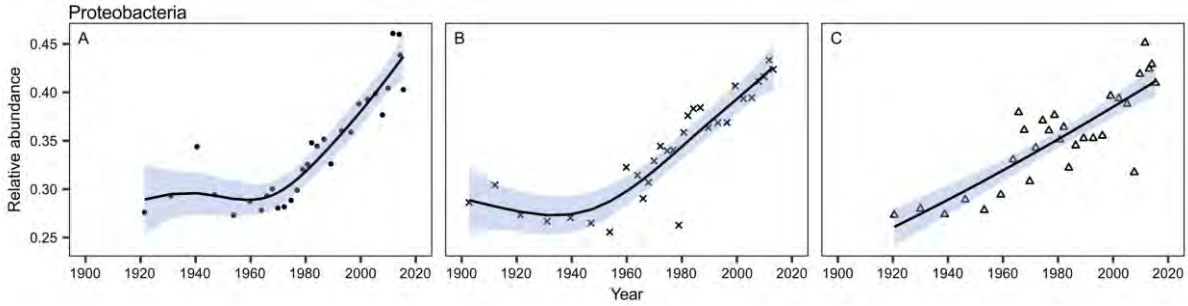

Figure S4. Generalised additive models (GAMs) of the relative abundance of Proteobacteria throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

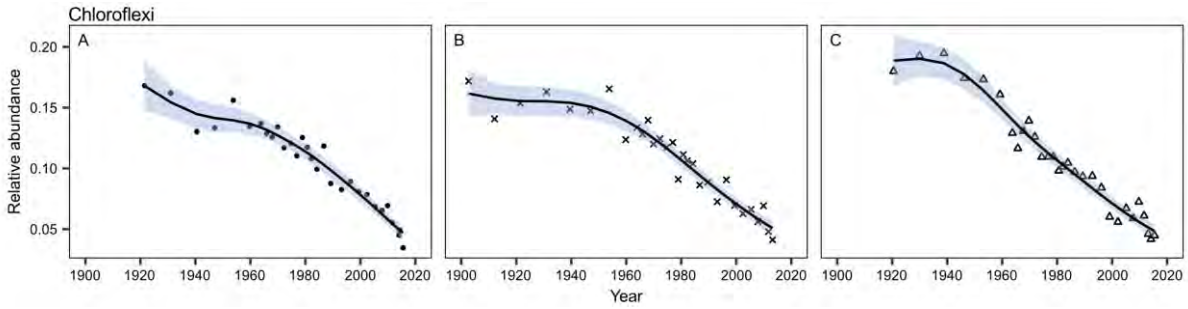

Figure S5. Generalised additive models (GAMs) of the relative abundance of Chloroflexi throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

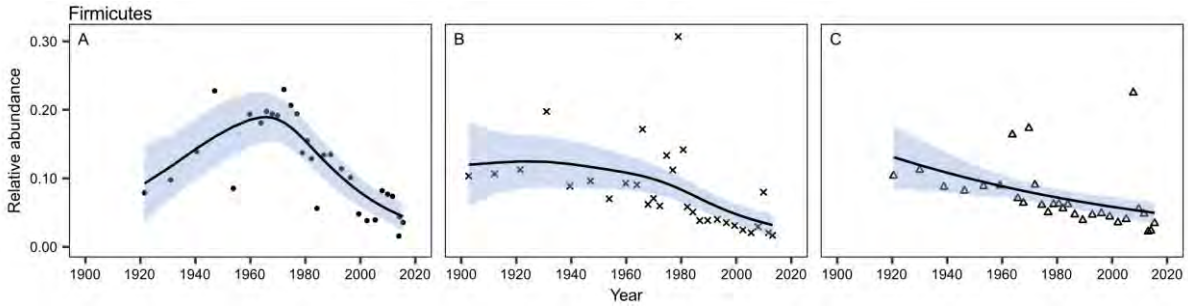

Figure S6. Generalised additive models (GAMs) of the relative abundance of Firmicutes throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

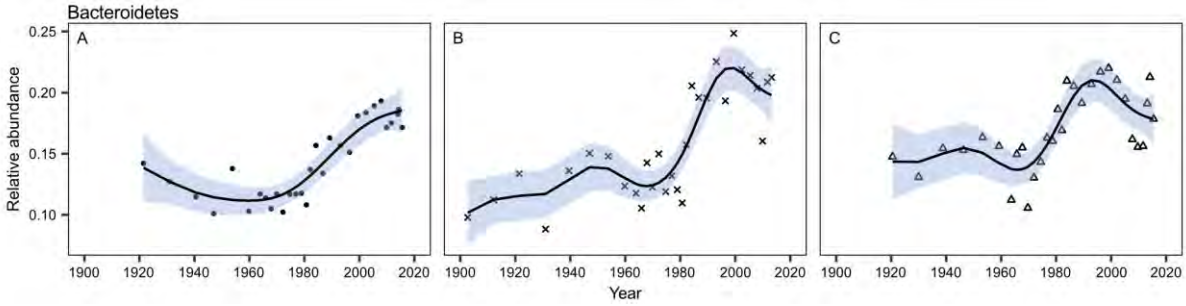

Figure S7. Generalised additive models (GAMs) of the relative abundance of Bacteroidetes throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

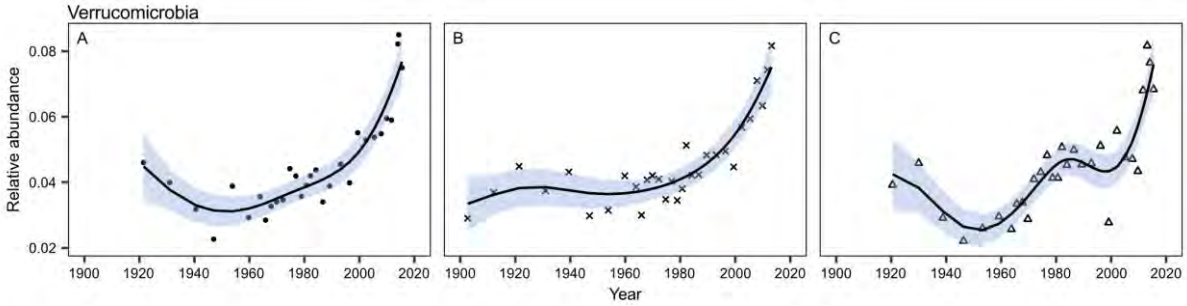

Figure S8. Generalised additive models (GAMs) of the relative abundance of Verrucomicrobia throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

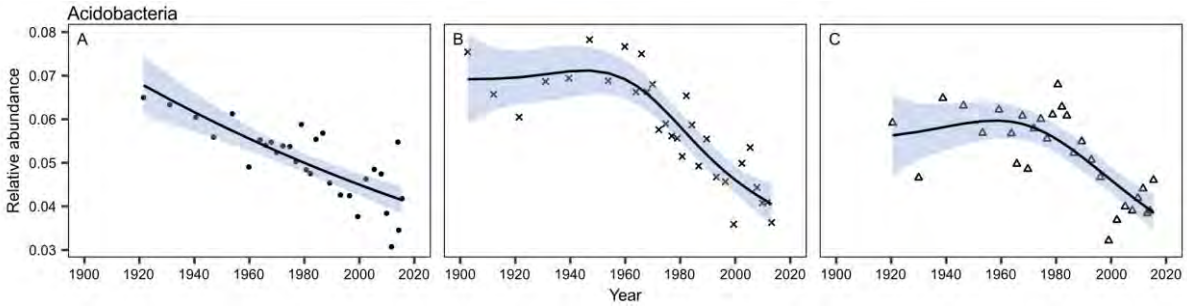

Figure S9. Generalised additive models (GAMs) of the relative abundance of Acidobacteria throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

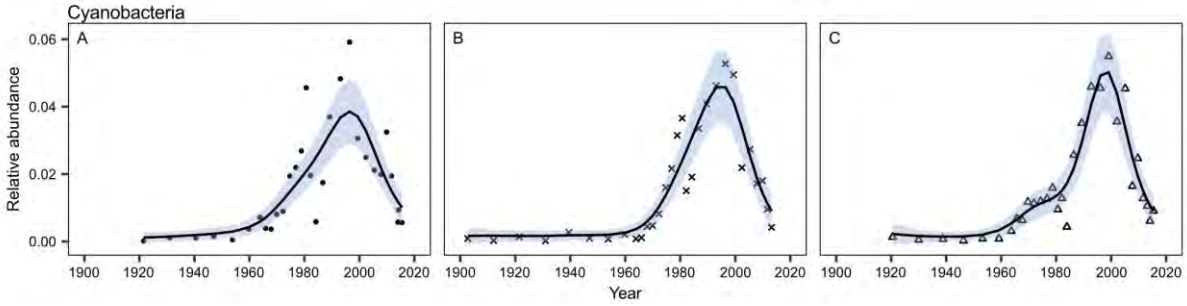

Figure S10. Generalised additive models (GAMs) of the relative abundance of Cyanobacteria throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

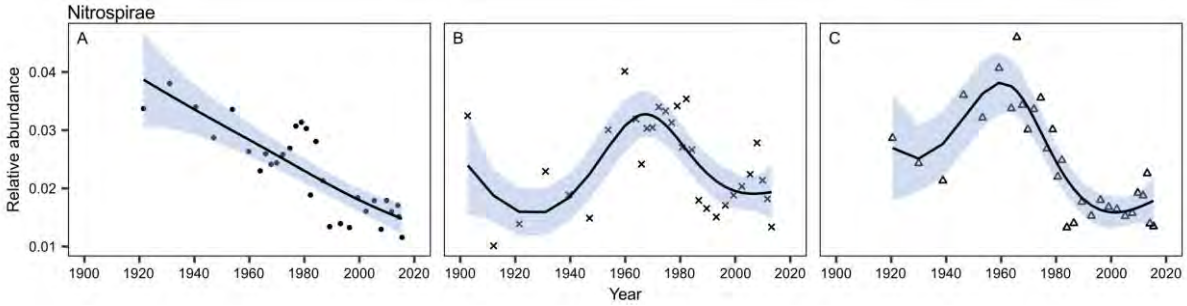

Figure S11. Generalised additive models (GAMs) of the relative abundance of Nitrospirae throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

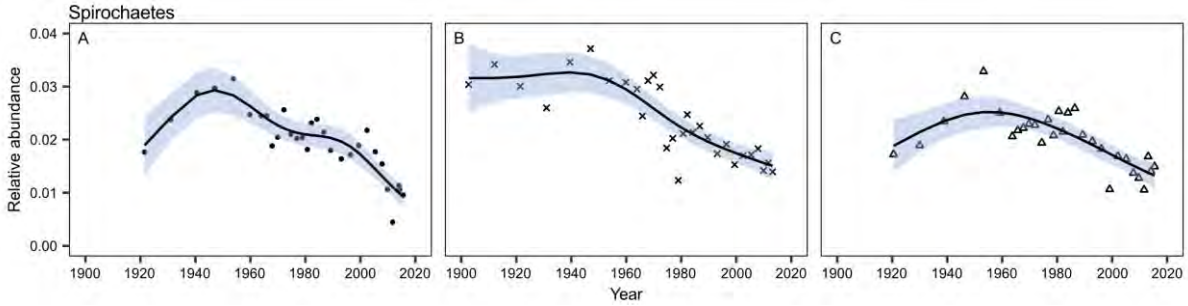

Figure S12. Generalised Additive Models (GAMs) of the relative abundance of Spirochaetes throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

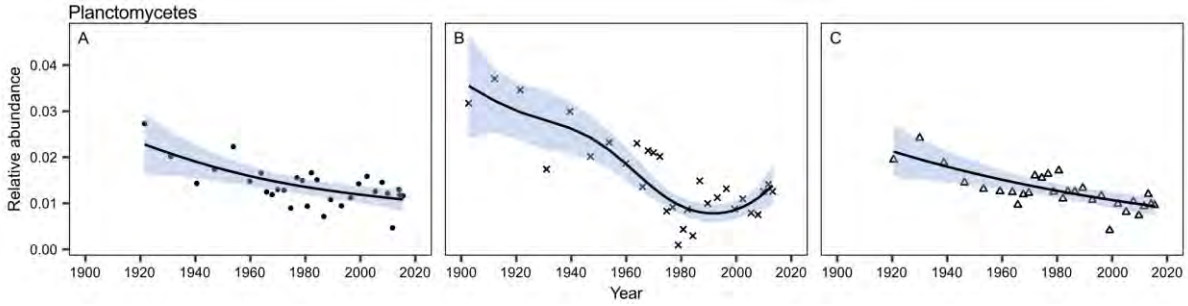

Figure S13. Generalised Additive Models (GAMs) of the relative abundance of Planctomycetes throughout sediment core A, B and C. Shaded area shows the 95% confidence interval.

Table S9. Spearman's correlation between the relative abundance of each bacterial phylum and sample age and lake physicochemical conditions including alkalinity, surface water temperature, soluble reactive phosphorus (SRP), winter SRP, total phosphorus (TP), chlorophyll *a*, nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N) and surface water pH.

	Sample age		Alkalinity		Temperature		SRP		Winter SRP		TP		Chlorophyll *a*		NO_3-N		NH_4-N		pH	
	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p
Proteobacteria	-0.87	***	0.76	***	0.44	***	0.40	***	0.42	***	-0.10		-0.10		0.04		-0.35	*	0.09	
Verrucomicrobia	-0.77	***	0.71	***	0.46	***	0.40	***	0.44	***	-0.14		-0.08		0.14		-0.25		0.04	
Bacteroidetes	-0.71	***	0.52	***	0.47	***	0.44	***	0.56	***	0.23		0.12		0.18		-0.13		0.07	
Cyanobacteria	-0.67	***	0.48	***	0.29	*	0.69	***	0.69	***	0.74	***	0.58	***	0.54	***	0.09		0.03	
Planctomycetes	0.62	***	-0.34	***	-0.33	**	-0.38	**	-0.36	**	-0.16		-0.04		-0.16		0.05		0.05	
Nitrospirae	0.59	***	-0.47	***	-0.33	**	-0.31	**	-0.41	***	-0.10		-0.08		-0.14		0.01		-0.13	
Firmicutes	0.63	***	-0.57	***	-0.43	***	-0.31	**	-0.44	***	-0.05		0.01		-0.05		0.15		-0.13	
Acidobacteria	0.78	***	-0.71	***	-0.26	*	-0.29	*	-0.30	**	-0.03		-0.10		-0.04		0.37	*	-0.18	
Spirochaetes	0.78	***	-0.71	***	-0.34	**	-0.36	**	-0.33	**	0.01		0.02		-0.06		0.27		-0.18	
Chloroflexi	0.97	***	-0.80	***	-0.50	***	-0.51	***	-0.52	***	0.02		0.00		-0.16		0.34	*	-0.16	

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

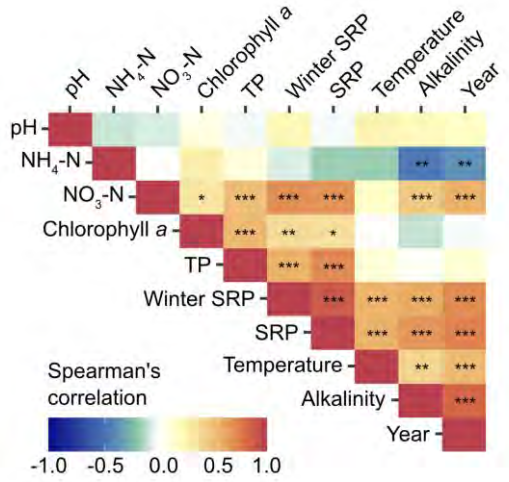

Figure S14. Correlations between year and lake physicochemical conditions including alkalinity, surface water temperature, soluble reactive phosphorus (SRP), winter SRP, total phosphorus (TP), chlorophyll *a*, nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N) and surface water pH. Correlations are calculated using Spearman's correlation, where red shades indicate positive correlations, and blue shades indicate negative correlations (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S10. Spearman's correlation between year and lake physicochemical conditions including alkalinity, surface water temperature, soluble reactive phosphorus (SRP), winter SRP, total phosphorus (TP), chlorophyll *a*, nitrate-nitrogen (NO_3-N), ammonium-nitrogen (NH_4-N) and surface water pH.

	Year		Alkalinity		Temperature		SRP		Winter SRP		TP		Chlorophyll *a*		NO_3-N		NH_4-N		pH	
	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p	R_s	p
Year	1.00	-	0.83	***	0.53	***	0.73	***	0.70	***	0.14		-0.01		0.58	***	-0.36	**	0.18	
Alkalinity	0.83	***	1.00	-	0.41	**	0.67	***	0.60	***	0.02		-0.09		0.46	***	-0.45	**	0.23	
Temperature	0.53	***	0.41	**	1.00	-	0.58	***	0.57	***	0.13		0.01		0.18		-0.16		0.25	
SRP	0.73	***	0.67	***	0.58	***	1.00	-	0.88	***	0.71	***	0.35	*	0.67	***	-0.16		-0.01	
Winter SRP	0.70	***	0.60	***	0.57	***	0.88	***	1.00	-	0.58	***	0.37	**	0.70	***	-0.06		0.24	
TP	0.14		0.02		0.13		0.71	***	0.58	***	1.00	-	0.54	***	0.51	***	0.11		-0.02	
Chlorophyll *a*	-0.01		-0.09		0.01		0.35	*	0.37	**	0.54	***	1.00	-	0.33	*	0.28		0.09	
NO_3-N	0.58	***	0.46	***	0.18		0.67	***	0.70	***	0.51	***	0.33	*	1.00	-	0.01		-0.06	
NH_4-N	-0.36	**	-0.45	**	-0.16		-0.16		-0.06		0.11		0.28		0.01		1.00	-	-0.08	
pH	0.18		0.23		0.25		-0.01		0.24		-0.02		0.09		-0.06		-0.08		1.00	-

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Appendix B: Supplementary information to Chapter 3

Table S1. Chronology of the 2014 reference sediment core (Thorpe *et al*., 2022) with sample depths corrected to 2021 assuming a constant sedimentation rate.

Corrected depth (cm)	Age (years from 2021)
0.00	0
5.67	7
5.92	7
7.92	10
9.92	14
11.92	20
13.92	26
15.92	33
17.92	38
19.92	41
21.92	45
23.92	50
25.92	54
27.92	58
29.92	70
31.92	85
33.92	104
35.42	118

Table S2. Chronology of the sediment core collected in 2021 estimated using the age-depth model based on the reference sediment core (Thorpe *et al*., 2022).

Depth (cm)	Year
0	2021
1	2020
2	2019
3	2017
4	2016
5	2015
6	2014
7	2012
8	2011
9	2009
10	2007
11	2004
12	2001
13	1998
14	1995
15	1991
16	1988
17	1985
18	1983
19	1981
20	1980
21	1978
22	1976
23	1973
24	1971
25	1969
26	1967
27	1965
28	1963
29	1957
30	1950
31	1943
32	1935
33	1926
34	1916

Comparison of two bacterial sedDNA records to support chronology estimation

A dated reference core collected from the deepest point of Esthwaite Water in 2014 was used to date cores collected in 2016 (Thorpe *et al*., 2022), and was also applied to the 2021 core assuming a constant sedimentation rate. However, as there was a longer period from 2014 to 2021, and possibly more variation in the sedimentation rate over time, the bacterial sedDNA records obtained from the 2016 and 2021 cores were compared to evaluate whether the chronology was accurately aligned with the 2021 core.

Sequencing of the bacterial community in sediments of the core was performed using the same 16S rRNA amplicon primers, and PCR amplification, sequencing and data processing steps followed the same general protocol described by Thorpe *et al*. (2022). Generalised additive models (GAMs) were fitted to the temporal trends in the relative abundance of ten abundant bacterial phyla. GAM-estimated annual trends were estimated for the period shared between the cores (1916-2016), and the correlation between the GAM-estimated annual trends for each bacterial phylum in the two sedDNA records was assessed with a model II regression.

The temporal trends of the dominant bacterial phyla (maximum relative abundance > 0.1) were well aligned between the 2016 (Thorpe *et al*., 2022) and 2021 cores (the present study) when the chronology of the reference core was applied. Proteobacteria showed comparable increasing trends (Fig. S2 A), and Chloroflexi (Fig. S2 B) and Firmicutes (Fig. S2 C) showed comparable decreasing trends. The increase in the relative abundance of Bacteroidetes was estimated to occur in 1970 in both records (Fig. S2 D). The large peak in the relative abundance of Cyanobacteria in the 2021 core was estimated to occur less than five years earlier than in the 2016 core (Fig. S2 G). Correlations between GAM-estimated trends were significant for all phyla ($p < 0.001$) and relatively strong for the dominant phyla (Fig. S2). While there may still

be some uncertainty associated with radiometric dating and with the application of the reference core chronology to a more recent core, the close alignment between the bacterial sedDNA records supported the use of the estimated chronology and demonstrated the robustness of the bacterial sedDNA record with repeated sediment coring.

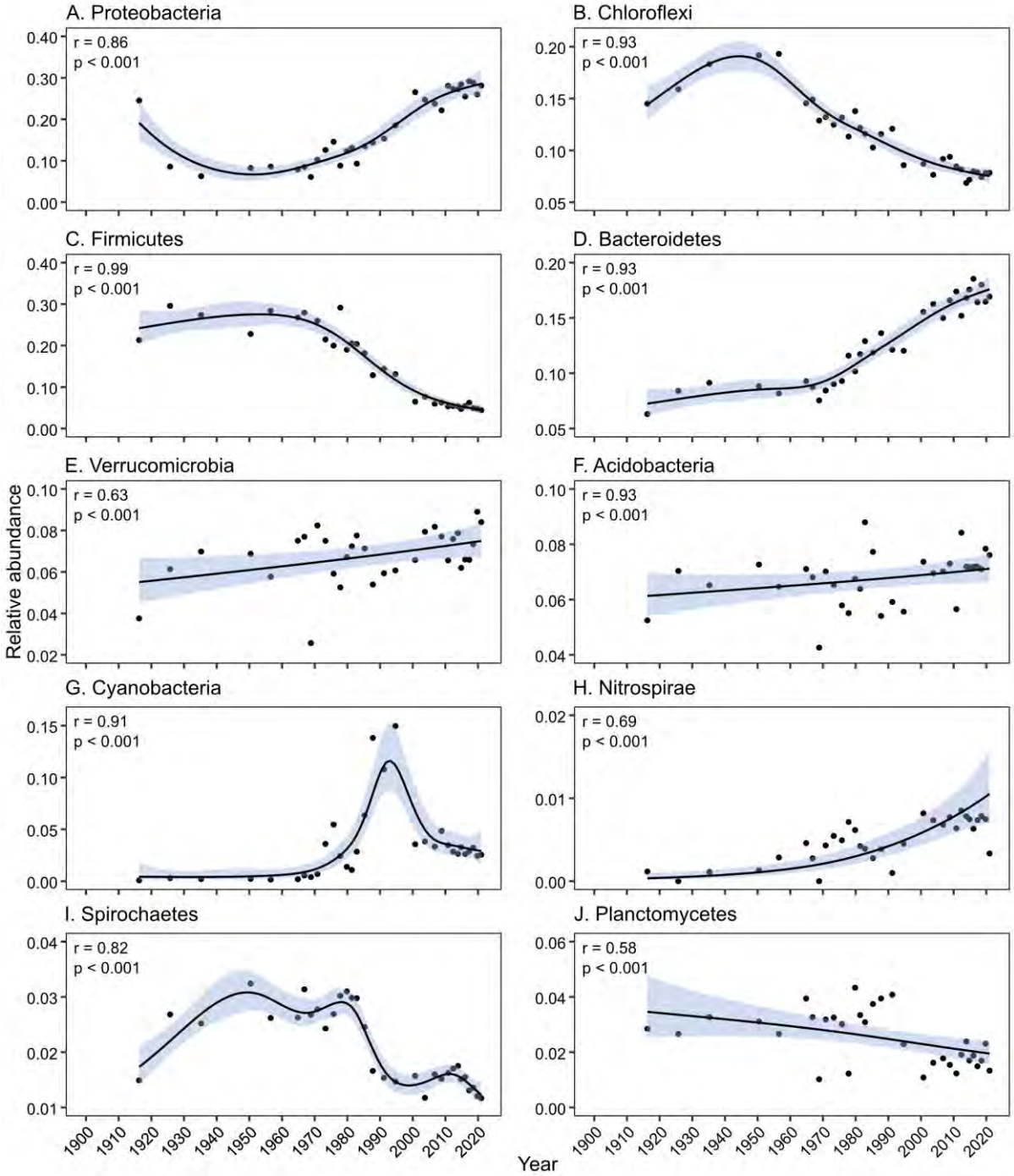

Figure S1. GAMs fitted to the trend in relative abundance of bacterial phyla as measured by sedDNA for comparison with a previous Esthwaite Water bacterial sedDNA record (Thorpe *et al*., 2022). r values and significance levels are shown for the correlation between the GAM-estimated annual trends in the relative abundance of each phylum in sediments from the 2016 and 2021 cores. Shaded areas show the 95% confidence intervals.

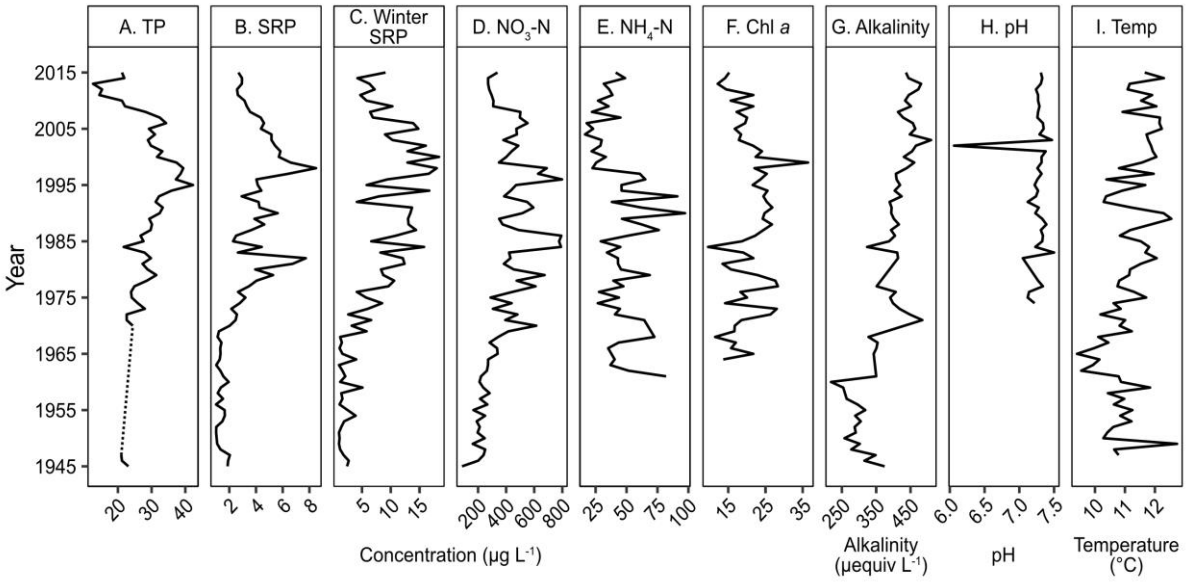

Figure S2. Mean annual conditions in Esthwaite Water between 1945 and 2015, including the concentration of (A) total phosphorus (TP), (B) soluble reactive phosphorus (SRP), (C) winter SRP from December to February, (D) nitrate-nitrogen (NO_3-N), (E) ammonium-nitrogen (NO_4-N), (F) chlorophyll *a* (Chl *a*), and (G) alkalinity, (H) pH and (I) temperature (temp).

Table S3. Results of the permutation test between the beta diversity Bray-Curtis dissimilarity matrix and sample year and lake physicochemical conditions including alkalinity, soluble reactive phosphorus (SRP), pH, chlorophyll *a*, winter SRP, ammonium-nitrogen (NH_4-N), total phosphorus (TP), temperature and nitrate-nitrogen (NO_3-N) for the sedDNA and microscopy-based records between 1945 and 2010.

	sedDNA		Microscopy	
Variable	R^2	p	R^2	p
Year	0.88	***	0.84	***
Alkalinity	0.53	*	0.61	***
SRP	0.50	*	0.23	*
pH	0.44	*	0.00	
Chlorophyll *a*	0.34		0.02	
Winter SRP	0.20		0.13	
NH_4-N	0.29		0.31	**
TP	0.16		0.25	*
Temperature	0.16		0.06	
NO_3-N	0.07		0.19	*

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S4. Statistics associated with the GAMs fitted to the relative abundance of each phylum as measured by sedDNA.

Phylum	edf	R^2	Deviance explained (%)	p
Chlorophytes	6.19	0.91	92.6	***
Dinoflagellates	9.99	0.97	98.7	***
Ochrophytes	3.70	0.31	54.8	***
Bacillariophytes	4.48	0.06	47.2	**

(edf = effective degrees of freedom, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S5. Statistics associated with the GAMs fitted to the occurrence of each phylum as measured by microscopy in the long-term monitoring record.

Phylum	edf	R^2	Deviance explained (%)	p
Chlorophytes	9.95	0.92	95.2	***
Dinoflagellates	9.25	0.81	84.1	***
Ochrophytes	6.61	0.85	86.0	***
Bacillariophytes	7.87	0.56	61.8	***
Charophytes	8.26	0.82	80.0	***
Cryptophytes	10.46	0.90	91.0	***
Haptophytes	1.00	0.68	43.5	***

(edf = effective degrees of freedom, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

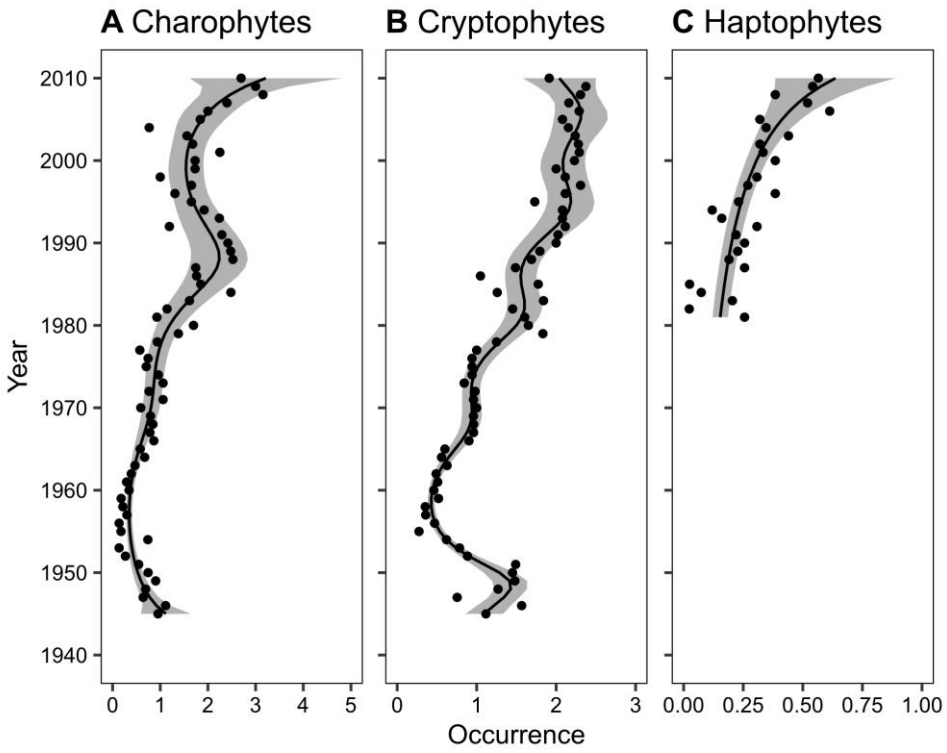

Figure S3. GAMs fitted to the trend in occurrence relative to sampling frequency as measured by microscopy for charophytes (A), cryptophytes (B), and haptophytes (C). Shaded areas show the 95% confidence intervals.

Appendix C: Supplementary information to Chapter 4

Equation 1. Calculation of the number of copies of 16S rRNA gene fragments used in qPCR as standards.

$$\text{Copies} = \frac{\text{X} \times N_{\text{A}}}{(\text{L} \times 660) \times \text{dilution}}$$

Where:

Copies = number of gene copies (molecules)

X = amount of gene fragment (ng),

L = length of gene fragment (bp),

660 = average mass of 1 bp of dsDNA (g mole^{-1}),

N_{A} = Avogadro's constant (6.0221×10^{23} molecules mole^{-1}),

dilution = starting dilution (1×10^{9}).

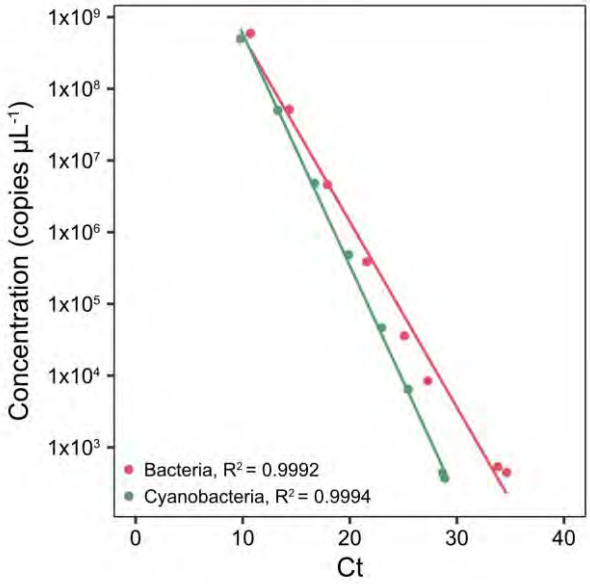

Figure S2. qPCR standard curves of the cycle threshold (Ct) and copy concentration of a serially diluted bacterial (red) and cyanobacterial (green) 16S rRNA gene fragment.

Equation 2. Conversion of gene copy concentration per qPCR reaction to gene copy concentration per mass of sample.

$$\text{Copies } (\text{g}^{-1} \text{ wet sediment}) = \frac{\text{copies per reaction} \times \left(\frac{1}{(\text{DF} \times \text{TV})}\right) \times \text{EV}}{\text{Mass of sample}}$$

Where:

Copies = gene copy concentration (g^{-1} wet sediment),

Copies per reaction = gene copy concentration per 20 µL reaction (average of duplicates),

DF = dilution factor of template DNA (1:20),

TV = volume of template DNA added (5 µL),

EV = elution volume of extracted DNA (100 µL),

Mass of sample = mass of wet sediment DNA was extracted from (0.25 g).

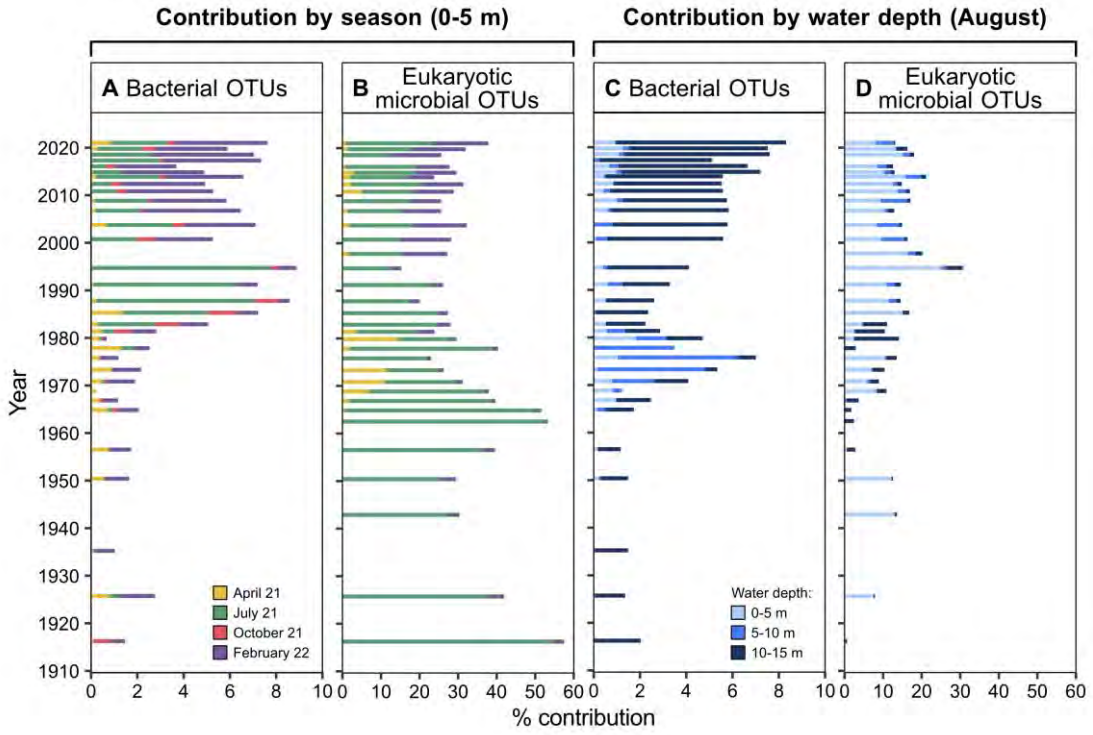

Figure S3. OTU-level FEAST-estimated contribution (%) of the water community at an integrated depth of 0-5 m in April 2021, July 2021, October 2021, and February 2022 to the sedDNA record for bacteria (A) and eukaryotic microbes (B), and contribution of the water column community at an integrated depth of 0-5, 5-10 and 10-15 m in August 2021 to the sedDNA record for bacteria (C) and eukaryotic microbes (D).

Appendix D: Supplementary information to Chapter 5

Table S1. Soil pH, mean annual air temperature (MAAT) and mean annual precipitation (MAP) of Eastern U.S. soil sampling sites.

Site	State	Latitude	Longitude	Soil pH	MAAT (°C)	MAP (mm year^{-1})
1	Maine	44.17371	-69.4111	5.20	7.67	1239.74
2	Maine	44.61978	-69.0118	5.04	6.87	1165.13
3	Maine	44.99938	-68.6504	4.28	6.53	1093.45
4	Maine	44.09081	-70.2817	5.06	7.50	1189.59
5	Maine	43.73451	-70.6940		6.84	1254.09
6	New Hampshire	43.08329	-71.1523	4.11	8.84	1178.97
7	Massachusetts	42.33098	-71.4753	3.84	9.64	1211.80
8	Massachusetts	42.05655	-71.7861	3.56	9.23	1243.26
9	Connecticut	41.78419	-72.2798	6.12	9.66	1258.09
10	Connecticut	41.74604	-72.9565	4.13	9.02	1359.34
11	Connecticut	41.56823	-73.4966		9.42	1270.47
12	New Jersey	40.97923	-74.2053	5.34	10.59	1292.84
13	New Jersey	40.76775	-74.4520	5.79	11.21	1277.22
14	Pennsylvania	40.39370	-75.2773	4.33	11.15	1263.30
15	Pennsylvania	40.19304	-75.7995	5.36	11.04	1225.49
16	Pennsylvania	39.97460	-76.5104	6.47	11.88	1137.64
17	Pennsylvania	39.64721	-76.7619	4.61	11.90	1212.79
18	Virginia	38.76014	-77.1195	4.76	14.18	1136.36
19	Virginia	38.33434	-77.1447	4.90	14.16	1132.07
20	Virginia	37.80631	-77.1208	4.14	14.94	1148.95
21	Virginia	37.37310	-77.5534	5.05	14.53	1155.51
22	Virginia	36.84687	-77.9296	5.38	14.64	1152.50
23	North Carolina	36.43981	-78.3661	5.08	14.80	1125.85
24	North Carolina	35.94941	-78.6088	5.99	15.31	1175.59
25	North Carolina	35.46396	-78.9130	4.36	16.04	1198.09
26	North Carolina	34.66763	-78.6047	3.70	17.01	1264.48
27	North Carolina	34.38813	-79.0020	4.07	16.97	1231.78
28	South Carolina	34.32800	-79.2793	4.34	17.13	1211.65
29	South Carolina	33.94747	-79.9784	4.99	17.51	1255.55
30	South Carolina	33.52045	-80.4987	4.76	18.00	1235.19
31	South Carolina	33.06203	-81.0900	4.01	18.42	1162.03
32	Georgia	32.87283	-81.9609	4.14	18.34	1157.67
33	Georgia	32.08360	-82.1353	4.92	19.13	1203.50
34	Georgia	31.51843	-82.7657	4.20	19.31	1206.13
35	Florida	30.21353	-82.5942	3.70	20.35	1344.70
36	Florida	29.10675	-81.6171	4.09	21.36	1352.13
37	Florida	28.35394	-81.6362	6.01	22.71	1295.60
38	Florida	26.15982	-81.2665	7.58	23.96	1439.05
39	Florida	26.94133	-81.3063	4.49	23.06	1263.98
40	Florida	27.60095	-82.5451	7.09	23.04	1273.24

Table S2. Soil pH, mean annual air temperature (MAAT) and mean annual precipitation (MAP) of Southern U.S. soil sampling sites.

Site	State	Lake catchment	Latitude	Longitude	Soil pH	MAAT (°C)	MAP (mm $year^{-1}$)
41	Texas	Lake Houston	29.90043	-95.1475	5.17	21.07	1474.04
42	Texas	Lake Meredith	35.70550	-101.557	8.00	15.28	474.04
43	New Mexico	Mary's Lake	36.85505	-105.096	6.38	6.90	505.60
44	New Mexico	Lake Escondida	34.12197	-106.891	8.09	14.40	245.05
45	New Mexico	Elephant Butte Lake	33.18170	-107.208	8.50	16.83	251.19
46	New Mexico	Lake Caballo	32.91138	-107.307	8.36	16.70	251.19
47	Arizona	Dankworth Pond	32.72065	-109.706	7.63	18.02	276.19
48	Arizona	Roper Lake	32.75513	-109.704	7.76	18.09	253.14
49	Arizona	Theodore Roosevelt Reservoir	33.67180	-111.133	7.29	20.49	245.85
50	Arizona	Lake Saguaro	33.57670	-111.536	6.56	21.92	368.58
51	Arizona	Horseshoe Lake	33.98670	-111.723	5.94	20.77	319.09
52	Arizona	Bartlett Lake	33.83801	-111.642	7.36	21.71	331.56
53	Arizona	Crescent Lake	33.90830	-109.424	5.83	6.12	334.78
54	New Mexico	Santa Rosa Lake	35.02990	-104.684	8.00	14.30	629.30
55	New Mexico	Conchas Lake	35.37940	-104.204	7.91	15.54	404.86
56	New Mexico	Ute Reservoir	35.36550	-103.493	7.90	15.34	409.71
57	Texas	Buffalo Springs Lake	33.53335	-101.720	8.77	16.15	413.60
58	Texas	Lake Colorado	32.33940	-100.929	8.06	17.84	458.90
59	Texas	Champion Creek Reservoir	32.27869	-100.848	7.88	17.80	545.88
60	Texas	EV Spence Reservoir	31.88980	-100.533	7.89	18.20	553.70
61	Texas	Lake Nasworthy	31.37194	-100.493	7.85	18.63	578.95
62	Texas	Lake Junction	30.48560	-99.7585	7.60	18.73	559.10
63	Texas	Medina Lake	29.54561	-98.9253	7.67	20.24	572.51
64	Texas	Lake Bastrop	30.14010	-97.2847	7.55	20.29	785.49
65	Texas	Lake Fayette	29.94670	-96.7497	6.70	20.69	915.63
66	Texas	Lake Sommerville	30.30690	-96.5171	4.25	20.18	998.34
67	Arizona	San Carlos Reservoir	33.18359	-110.519	8.10	19.35	1022.18

Table S3. Lake surface sediment pH, mean annual air temperature (MAAT) and mean annual precipitation (MAP) of Southern U.S. lake surface sediment sampling sites.

Site	State	Lake	Latitude	Longitude	Sediment pH	MAAT (°C)	MAP (mm $year^{-1}$)
41	Texas	Lake Houston	29.92703	-95.1518	5.76	21.09	1474.04
42	Texas	Lake Meredith	35.71152	-101.556	7.86	15.30	474.04
43	New Mexico	Mary's Lake	36.85165	-105.097	8.80	7.02	509.97
44	New Mexico	Lake Escondida	34.12158	-106.890	7.62	14.41	245.05
45	New Mexico	Elephant Butte Lake	33.18322	-107.193	7.37	16.84	251.19
46	New Mexico	Lake Caballo	32.91125	-107.306	8.23	16.71	276.19
47	Arizona	Dankworth Pond	32.72069	-109.705	7.12	18.03	253.14
48	Arizona	Roper Lake	32.75560	-109.705	7.23	18.10	245.85
49	Arizona	Theodore Roosevelt Reservoir	33.68630	-111.123	7.34	20.51	368.58
50	Arizona	Lake Saguaro	33.57276	-111.532	7.67	21.94	319.09
51	Arizona	Horseshoe Lake	33.98826	-111.718	7.51	20.79	331.56
52	Arizona	Bartlett Lake	33.82470	-111.627	7.52	21.73	334.78
53	Arizona	Crescent Lake	33.91050	-109.421	5.90	6.13	629.30
54	New Mexico	Santa Rosa Lake	35.03385	-104.685	7.48	14.32	404.86
55	New Mexico	Conchas Lake	35.38940	-104.194	7.52	15.55	409.71
56	New Mexico	Ute Reservoir	35.35727	-103.489	8.14	15.35	413.60
57	Texas	Buffalo Springs Lake	33.52910	-101.712	7.76	16.17	458.90
58	Texas	Lake Colorado	32.34075	-100.928	8.23	17.85	545.88
59	Texas	Champion Creek Reservoir	32.28290	-100.848	7.40	17.82	553.70
60	Texas	EV Spence Reservoir	31.89422	-100.532	7.57	18.22	578.95
61	Texas	Lake Nasworthy	31.37950	-100.490	7.82	18.64	559.10
62	Texas	Lake Junction	30.48720	-99.7621	8.01	18.74	572.51
63	Texas	Medina Lake	29.54460	-98.9318	7.40	20.26	785.49
64	Texas	Lake Bastrop	30.15176	-97.2865	7.25	20.23	909.23
65	Texas	Lake Fayette	29.94678	-96.7457	8.01	20.71	998.34
66	Texas	Lake Sommerville	30.31486	-96.5289	6.76	20.05	1013.96
67	Arizona	San Carlos Reservoir	33.17792	-110.516	7.45	19.37	355.72

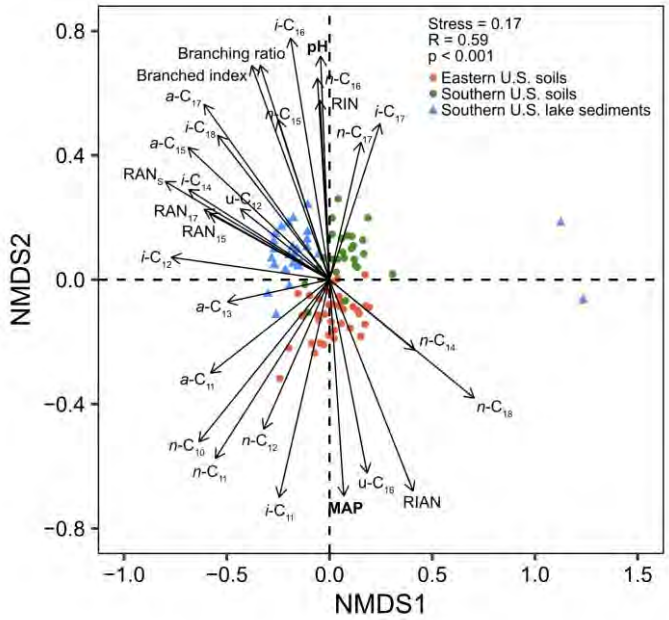

Figure S1. NMDS of a dissimilarity matrix based on 3-OH FA distribution in Eastern U.S. soils and Southern U.S. soils and lake surface sediments including two lake surface sediment outliers with limited 3-OH FAs. Vectors for environmental vectors (bold) and 3-OH FAs and their indices that correlated with the dissimilarity matrix were fitted. Only the strongest ($R^2 > 0.20$) and most significant ($p < 0.001$) vectors are displayed. NMDS stress values and ANOSIM R and p values are shown.

Table S4. Results of the permutation test between the dissimilarity matrices based on bacterial beta diversity and 3-OH FA distribution and environmental variables including soil or lake surface sediment pH, mean annual precipitation (MAP) and mean annual air temperature (MAAT).

Variable	Bacterial dissimilarity matrix		3-OH FA dissimilarity matrix	
	R^2	p	R^2	p
pH	0.81	***	0.61	***
MAP	0.63	***	0.54	***
MAAT	0.06		0.08	*

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S5. Results of the permutation test between the bacterial beta diversity dissimilarity matrix and the top 15 most abundant bacterial phyla.

Phylum	R^2	p
Desulfobacteria	0.79	***
Acidobacteria	0.71	***
Chloroflexi	0.69	***
Gemmatimonadetes	0.58	***
Planctomycetes	0.53	***
Bacteroidetes	0.43	***
Proteobacteria	0.36	***
RCP2-54	0.35	***
Cyanobacteria	0.33	***
WPS-2	0.33	***
Verrucomicrobia	0.29	***
Myxococcota	0.22	***
Firmicutes	0.21	***
Nitrospirae	0.09	*
Actinobacteria	0.07	*

(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Table S6. Results of the permutation test between the dissimilarity matrix based on 3-OH FA distribution and 3-OH FAs and 3-OH FA indices.

Phylum	R^2	p
RAN_S	0.73	***
n-C_{11}	0.68	***
a-C_{17}	0.67	***
n-C_{18}	0.66	***
i-C_{16}	0.66	***
a-C_{15}	0.64	***
RIAN	0.64	***
Branched index	0.63	***
i-C_{12}	0.62	***
n-C_{10}	0.61	***
Branching ratio	0.60	***
i-C_{14}	0.54	***
i-C_{18}	0.50	***
i-C_{11}	0.49	***
u-C_{16}	0.44	***
n-C_{16}	0.44	***
RAN_{17}	0.42	***
n-C_{12}	0.42	***
RAN_{15}	0.39	***
RIN	0.35	***
n-C_{15}	0.35	***
i-C_{17}	0.32	***
a-C_{11}	0.30	***
a-C_{13}	0.26	***
n-C_{17}	0.26	***
u-C_{12}	0.24	***
n-C_{14}	0.22	***
RAN_{13}	0.21	***
u-C_{18}	0.19	**
u-C_{14}	0.14	**
u-C_{11}	0.14	***
n-C_{13}	0.12	**
i-C_{15}	0.07	
u-C_{13}	0.07	*
RIN_{17}	0.01	
i-C_{13}	0.00	

(*i = iso, a = anteiso, n = normal,* u = unsaturated, *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$).

Appendix E: Supplementary information to Chapter 6

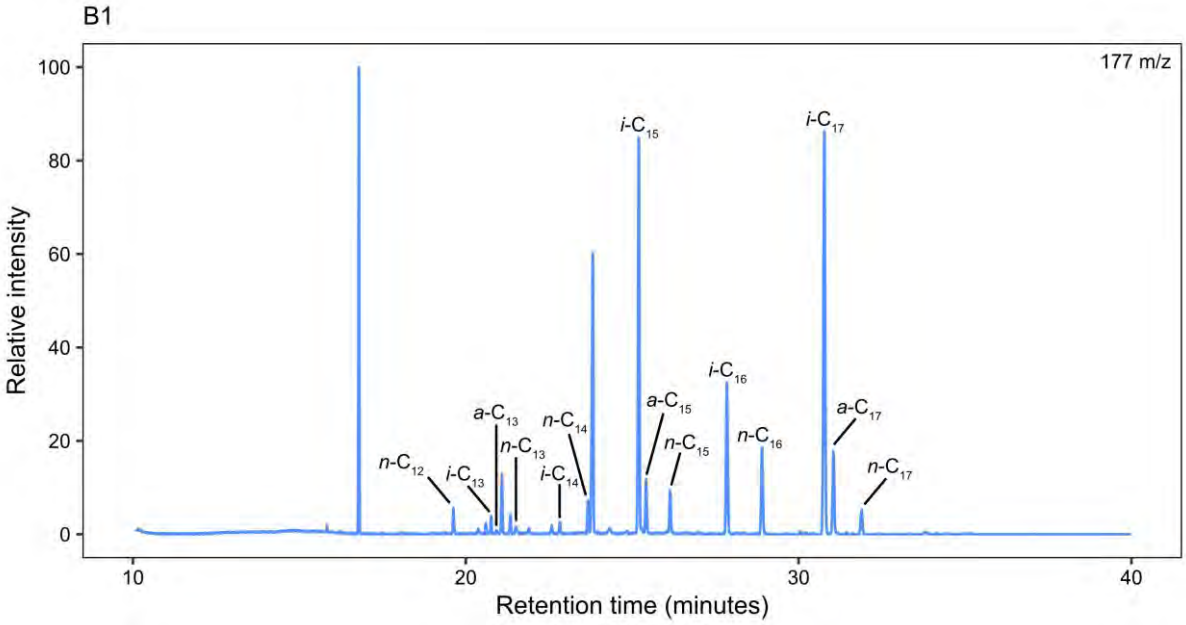

Figure S1. Chromatogram of 3-OH FA isomers extracted from bacterial isolate B1. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 177.

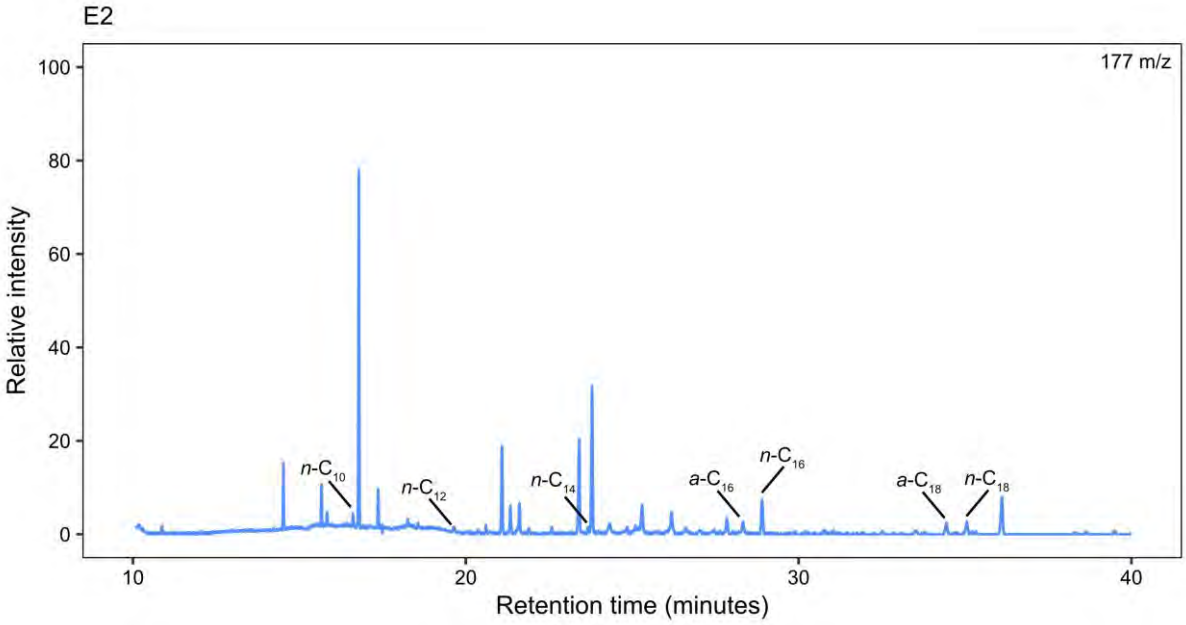

Figure S2. Chromatogram of 3-OH FA isomers extracted from bacterial isolate E2. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 177.

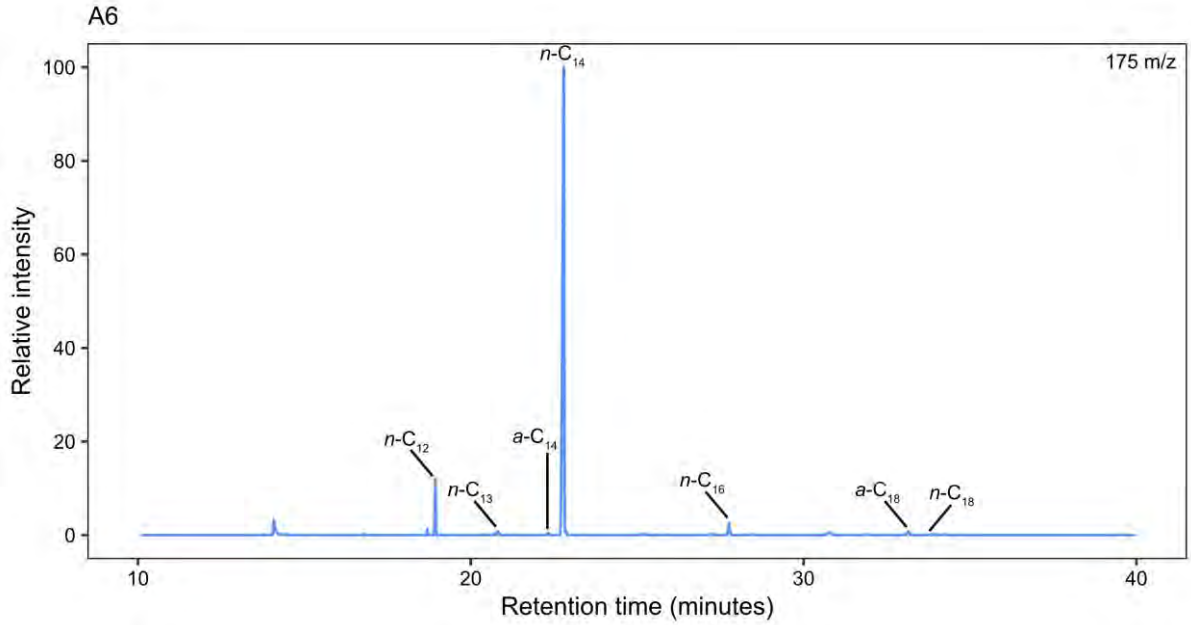

Figure S3. Chromatogram of 3-OH FA isomers extracted from bacterial isolate A6. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

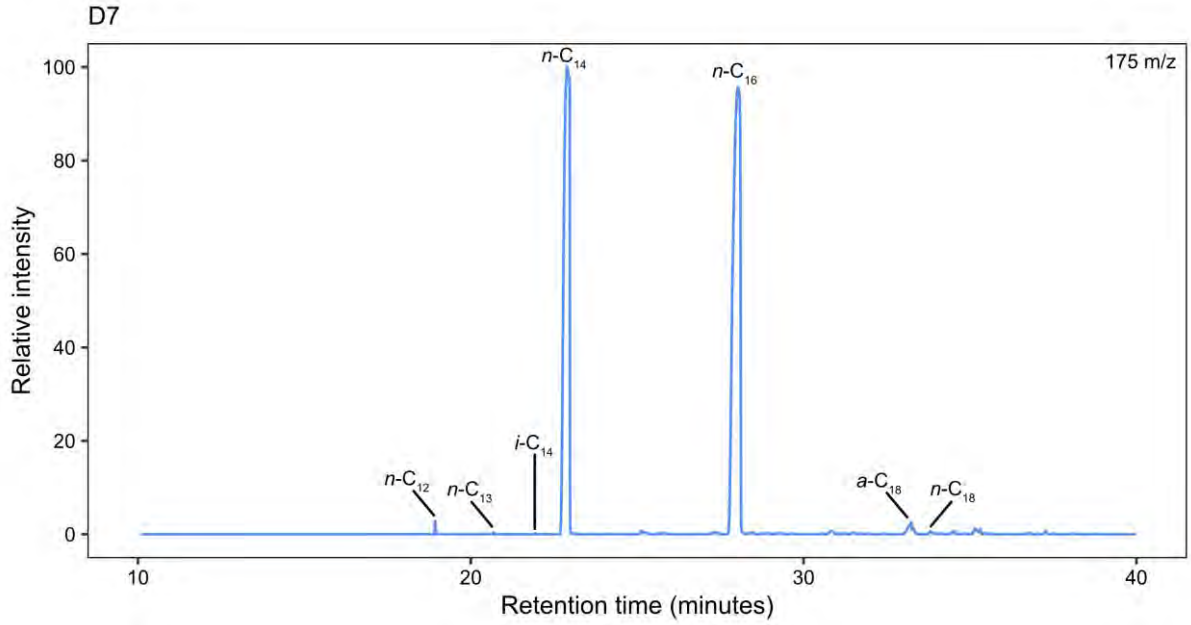

Figure S4. Chromatogram of 3-OH FA isomers extracted from bacterial isolate D7. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

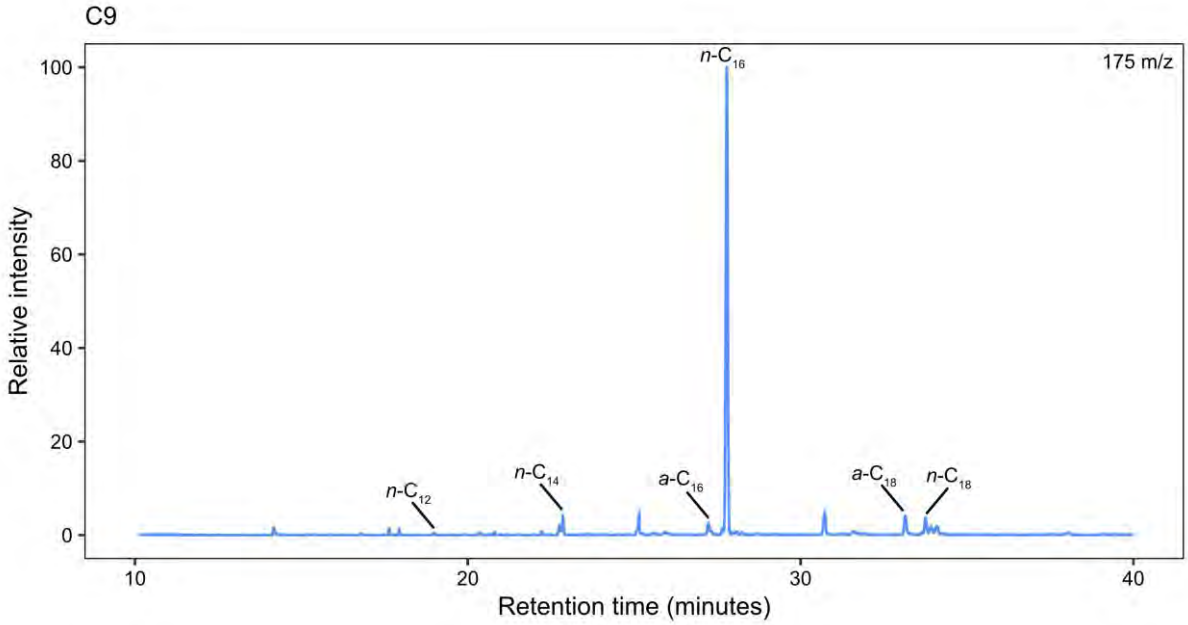

Figure S5. Chromatogram of 3-OH FA isomers extracted from bacterial isolate C9. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

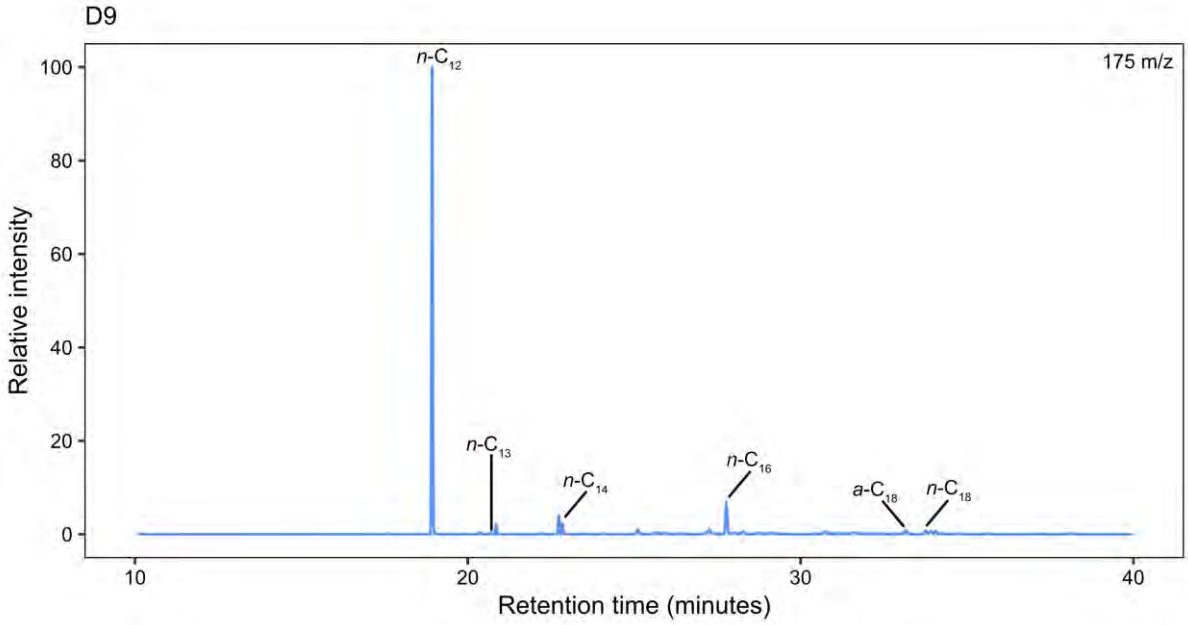

Figure S6. Chromatogram of 3-OH FA isomers extracted from bacterial isolate D9. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

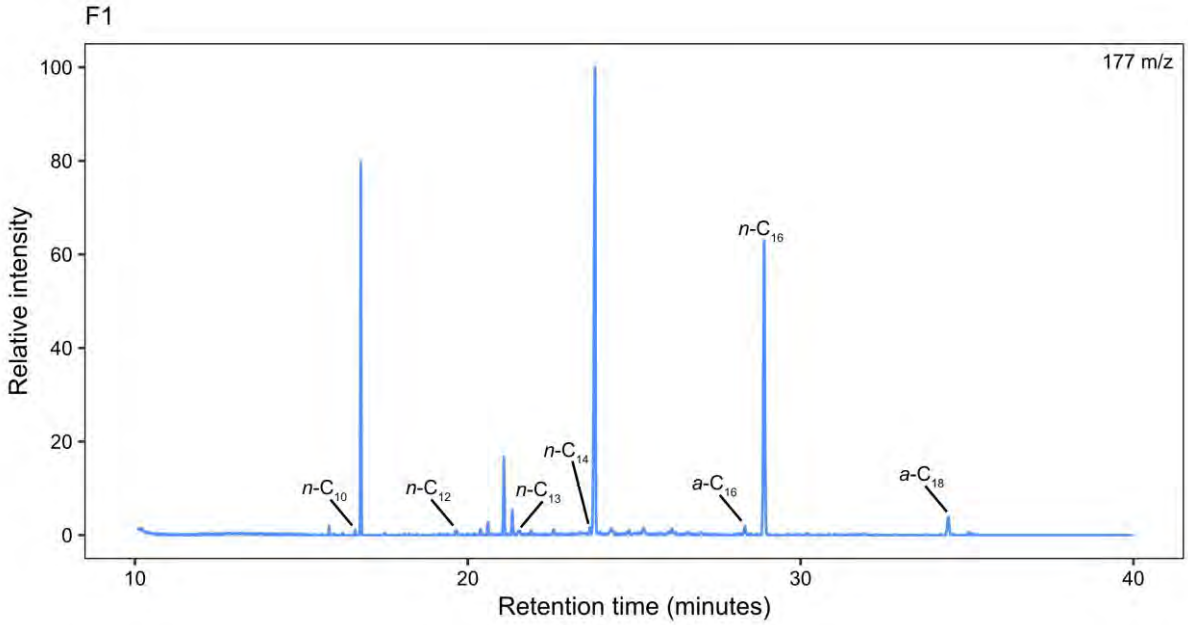

Figure S7. Chromatogram of 3-OH FA isomers extracted from bacterial isolate F1. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 177.

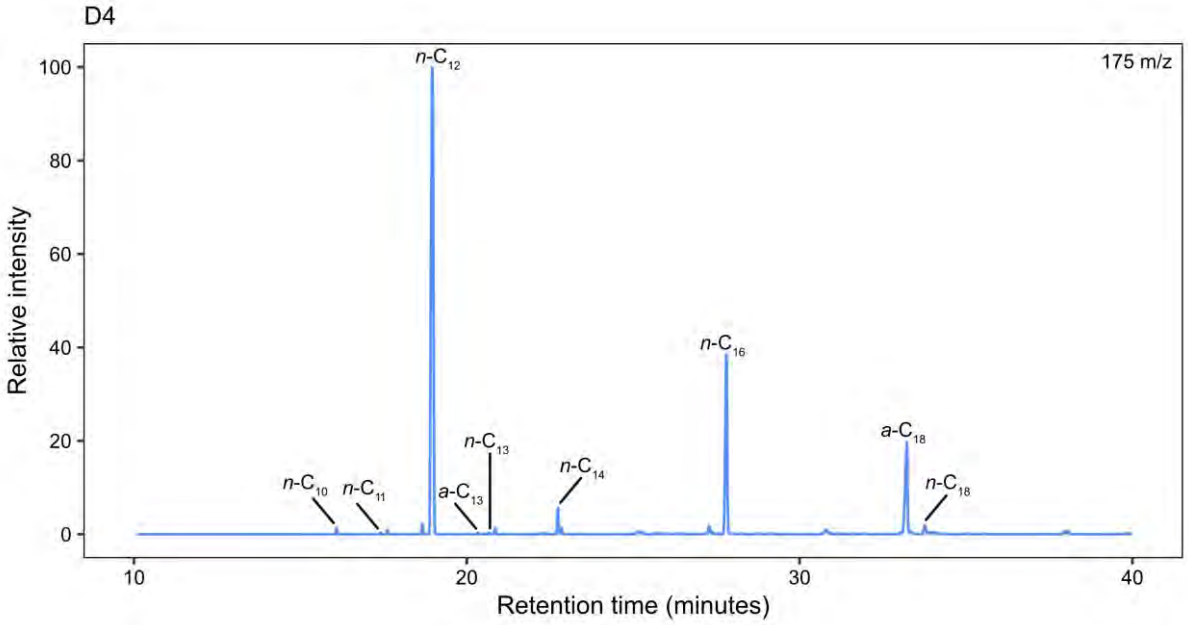

Figure S8. Chromatogram of 3-OH FA isomers extracted from bacterial isolate D4. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

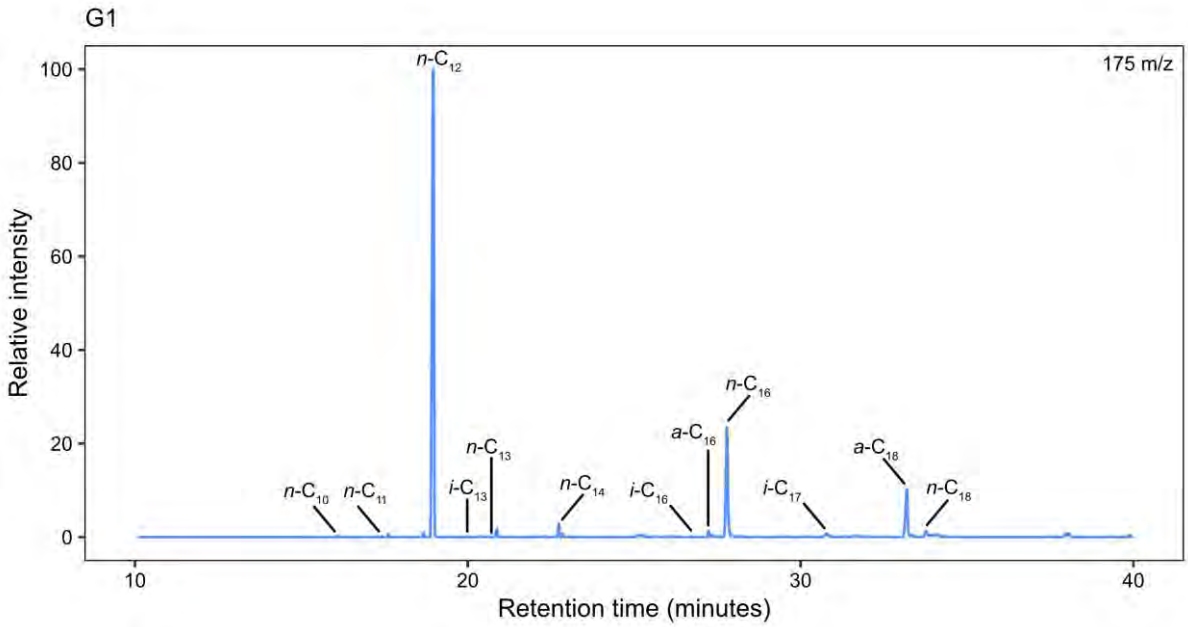

Figure S9. Chromatogram of 3-OH FA isomers extracted from bacterial isolate G1. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 175.

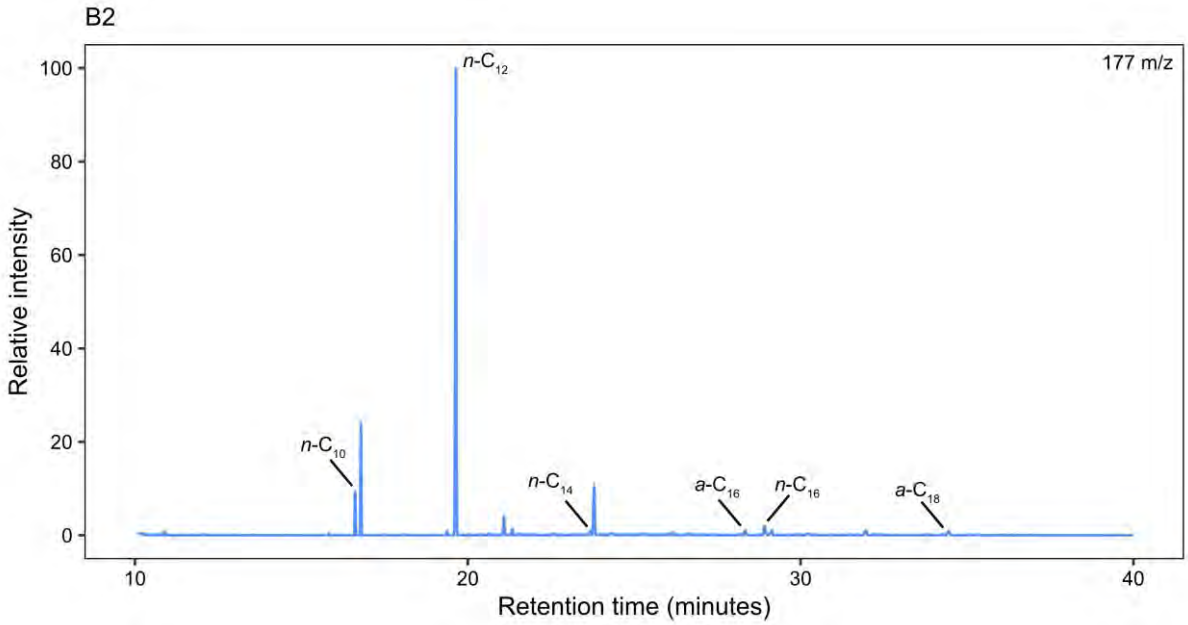

Figure S10. Chromatogram of 3-OH FA isomers extracted from bacterial isolate B2. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 177.

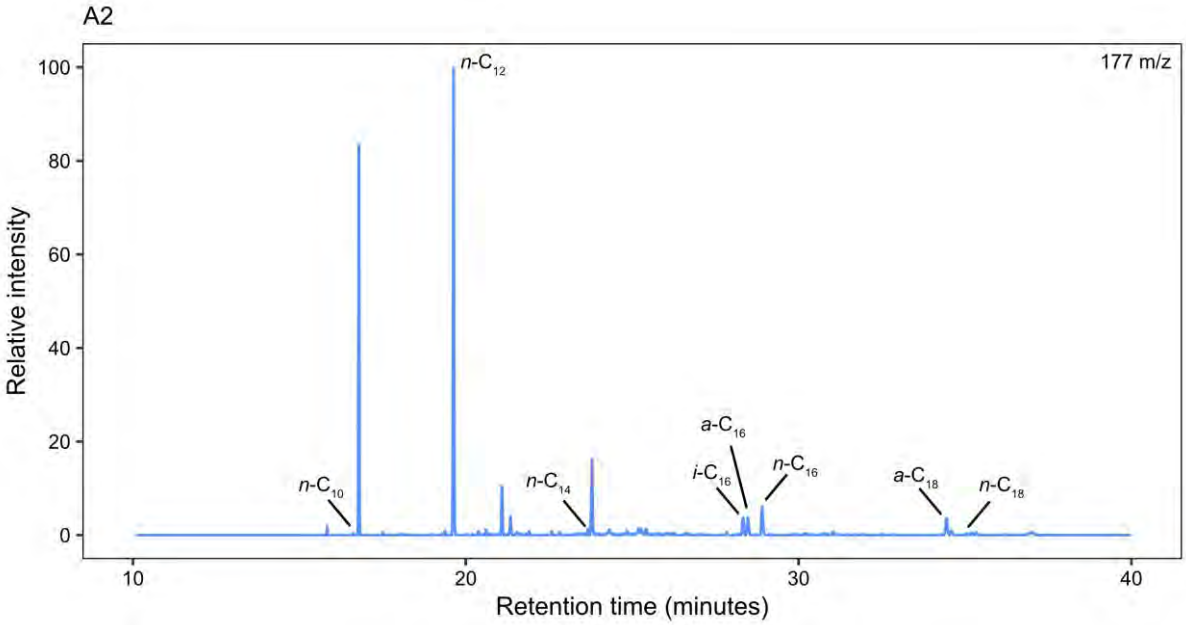

Figure S11. Chromatogram of 3-OH FA isomers extracted from bacterial isolate A2. 3-OH FA isomers are labelled, *i* = *iso*, *a* = *anteiso*, n = *normal*. m/z = 177.

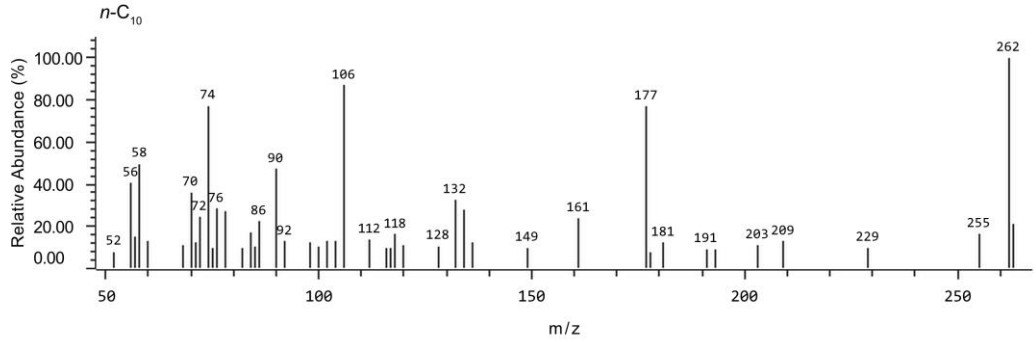

Figure S12. Mass spectrum of *n*-C_{10} 3-OH FA.

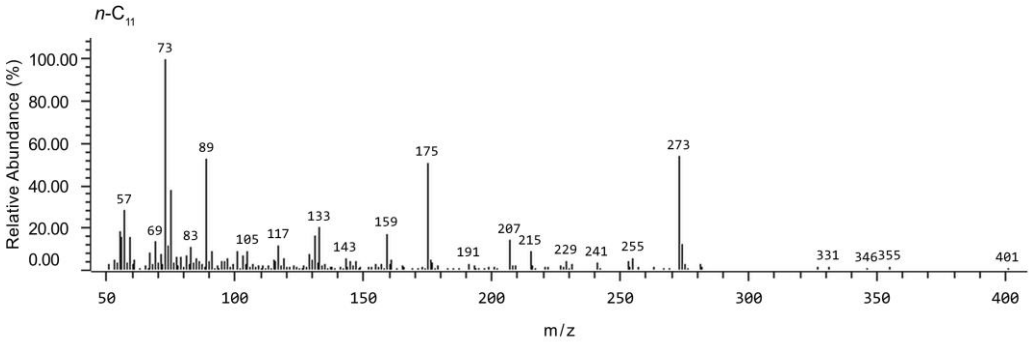

Figure S13. Mass spectrum of *n*-C_{11} 3-OH FA.

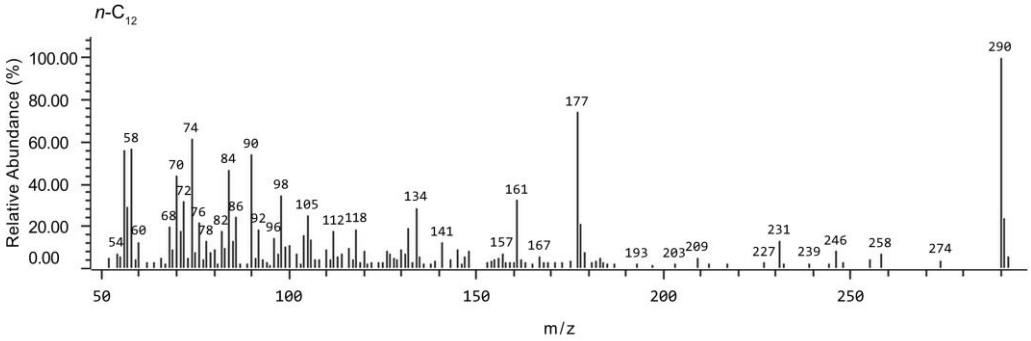

Figure S14. Mass spectrum of *n*-C_{12} 3-OH FA.

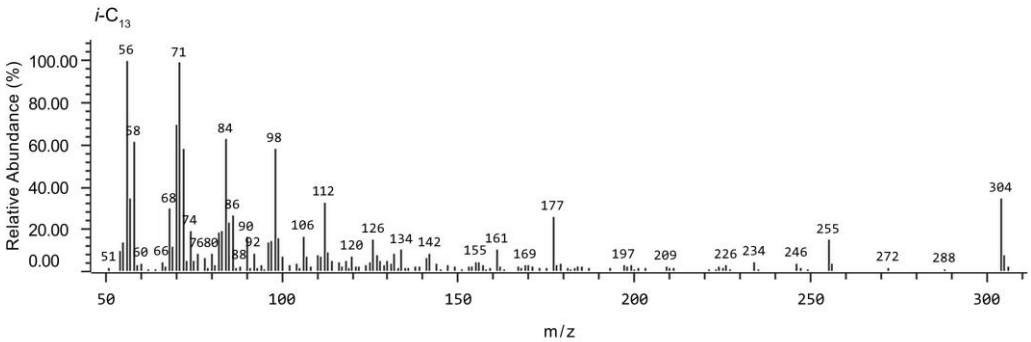

Figure S15. Mass spectrum of *i*-C_{13} 3-OH FA.

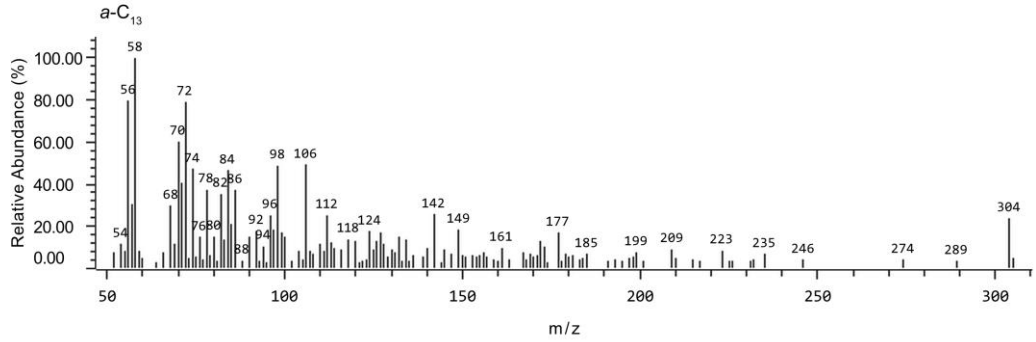

Figure S16. Mass spectrum of *a*-C_{13} 3-OH FA.

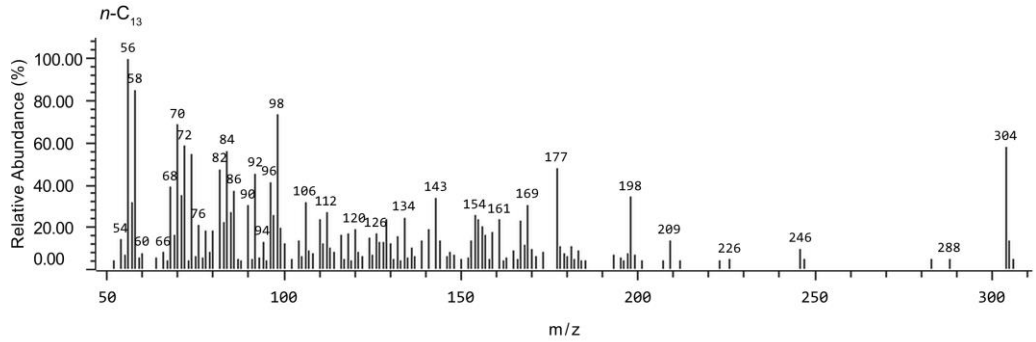

Figure S17. Mass spectrum of *n*-C_{13} 3-OH FA.

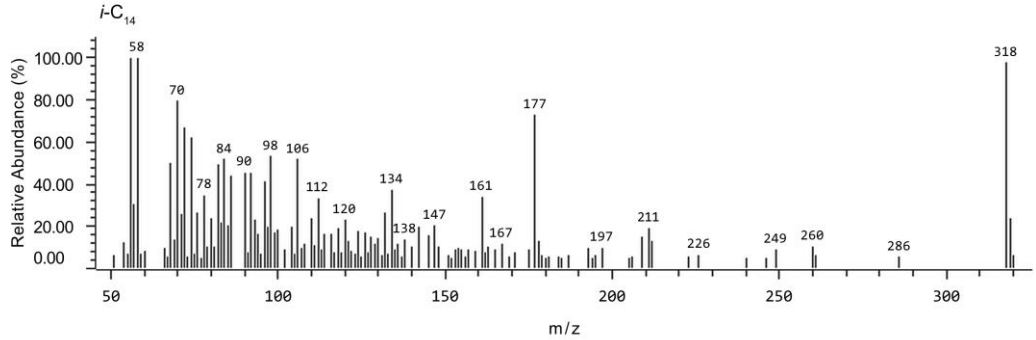

Figure S18. Mass spectrum of *i*-C_{14} 3-OH FA.

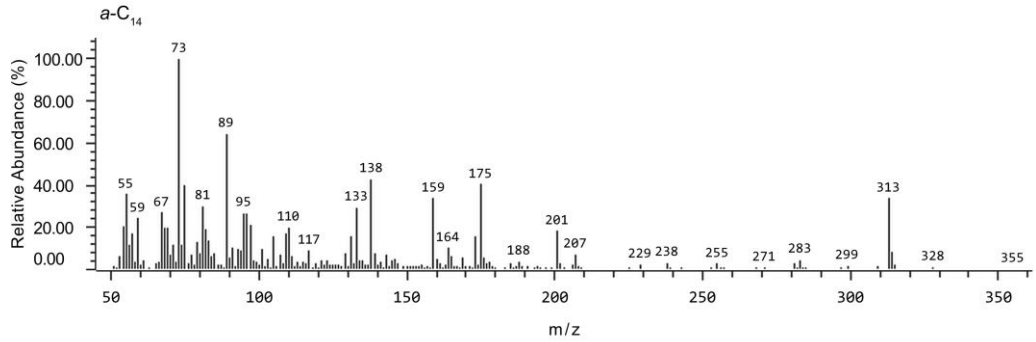

Figure S19. Mass spectrum of *a*-C_{14} 3-OH FA.

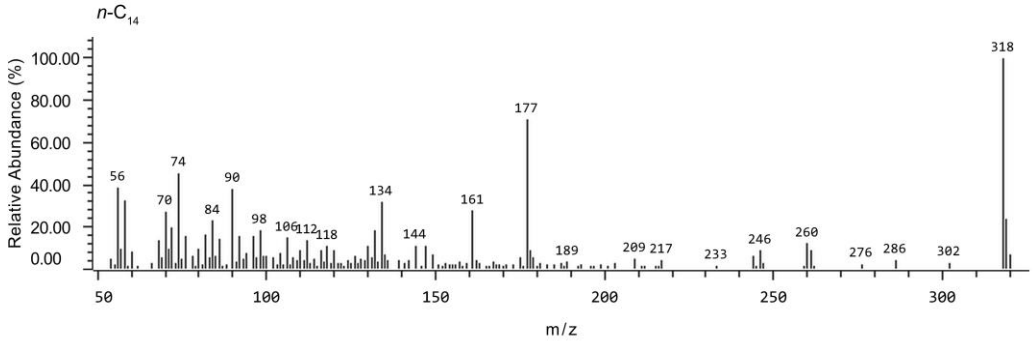

Figure S20. Mass spectrum of *n*-C_{14} 3-OH FA.

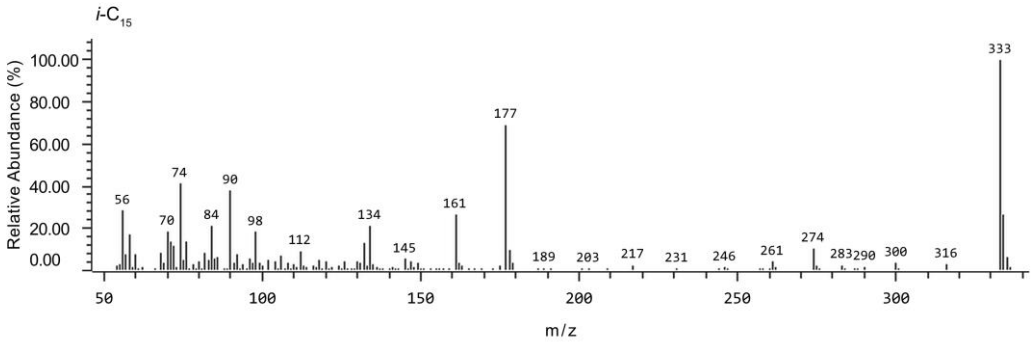

Figure S21. Mass spectrum of *i*-C_{15} 3-OH FA.

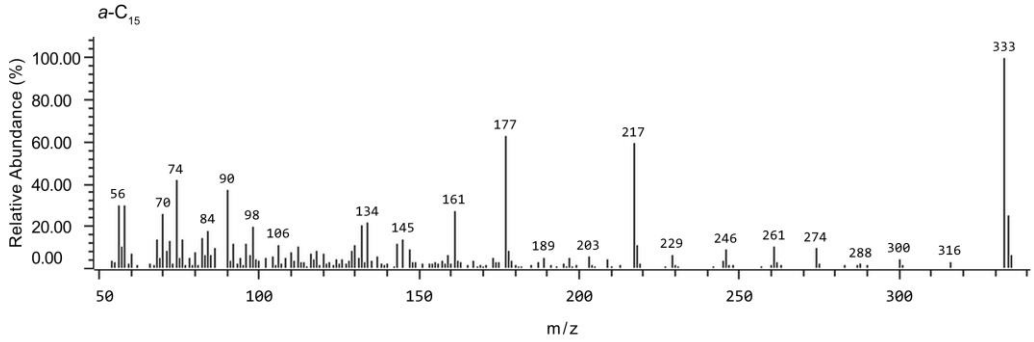

Figure S22. Mass spectrum of *a*-C_{15} 3-OH FA.

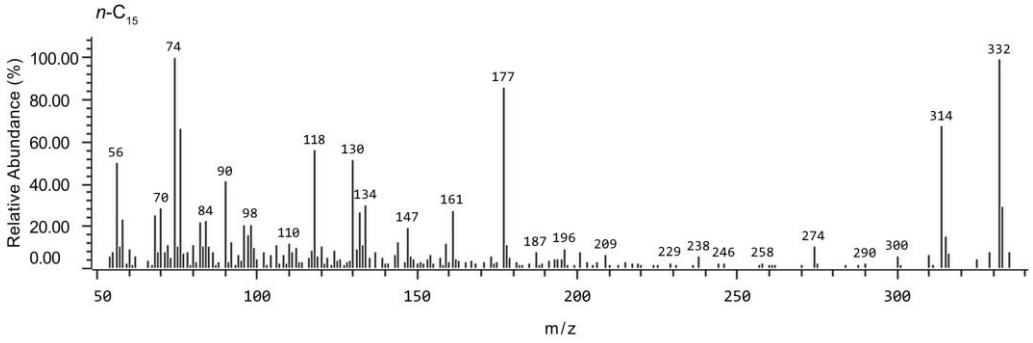

Figure S23. Mass spectrum of *n*-C_{15} 3-OH FA.

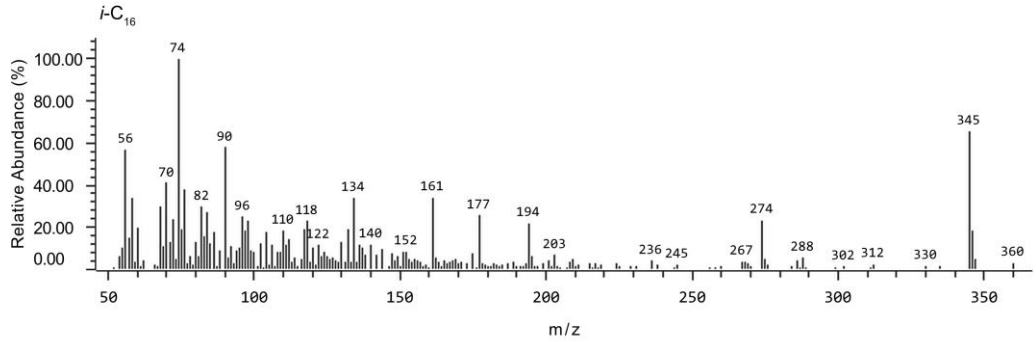

Figure S24. Mass spectrum of *i*-C_{16} 3-OH FA.

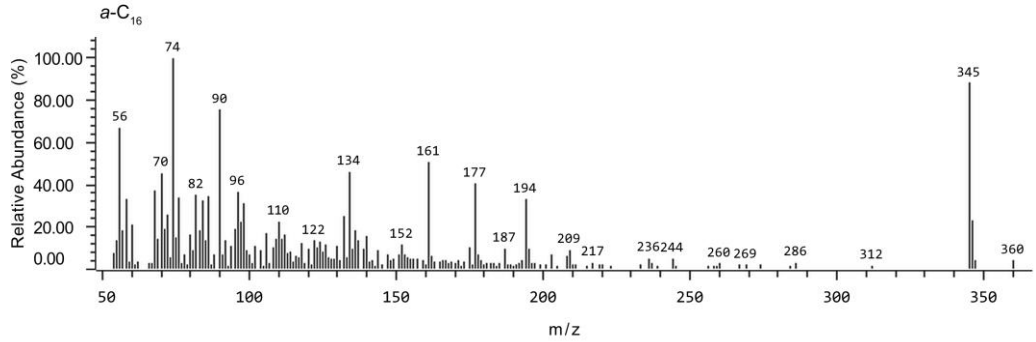

Figure S25. Mass spectrum of *a*-C_{16} 3-OH FA.

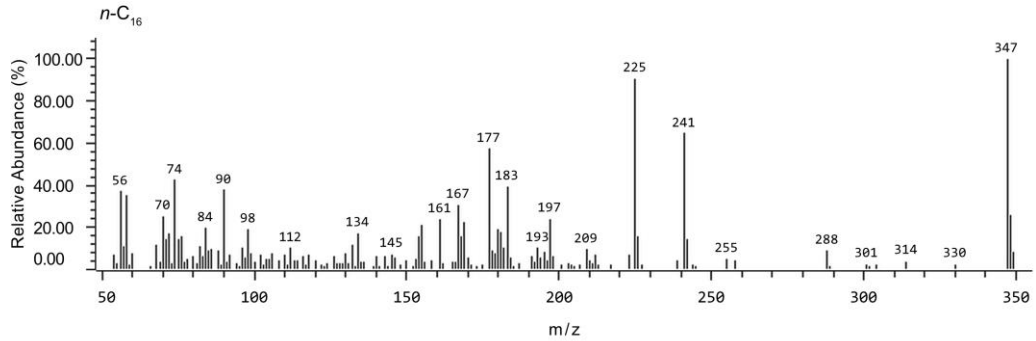

Figure S26. Mass spectrum of *n*-C_{16} 3-OH FA.

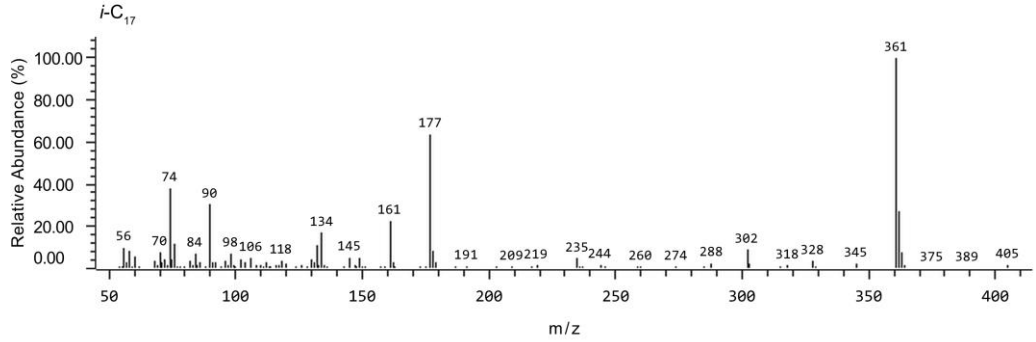

Figure S27. Mass spectrum of *i*-C_{17} 3-OH FA.

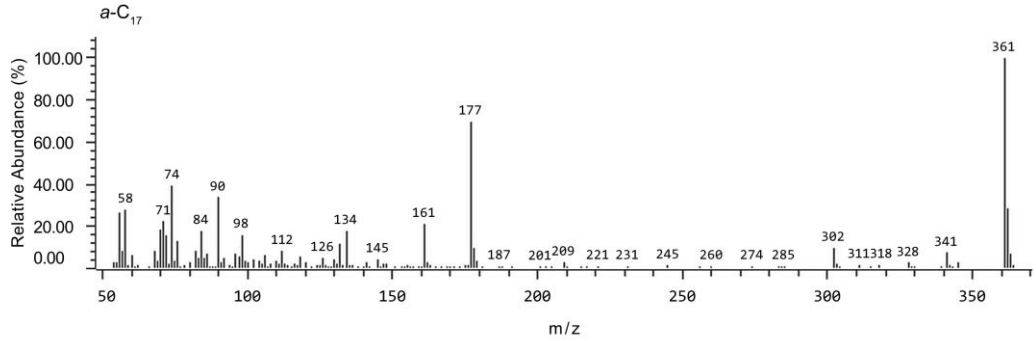

Figure S28. Mass spectrum of *a*-C_{17} 3-OH FA.

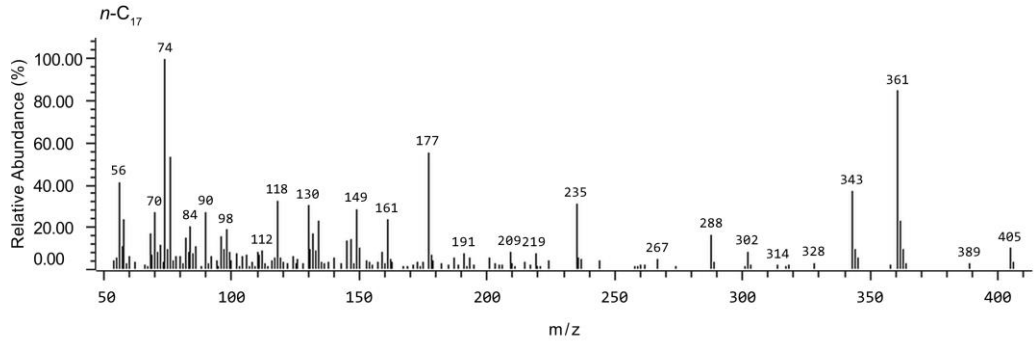

Figure S29. Mass spectrum of *n*-C_{17} 3-OH FA.

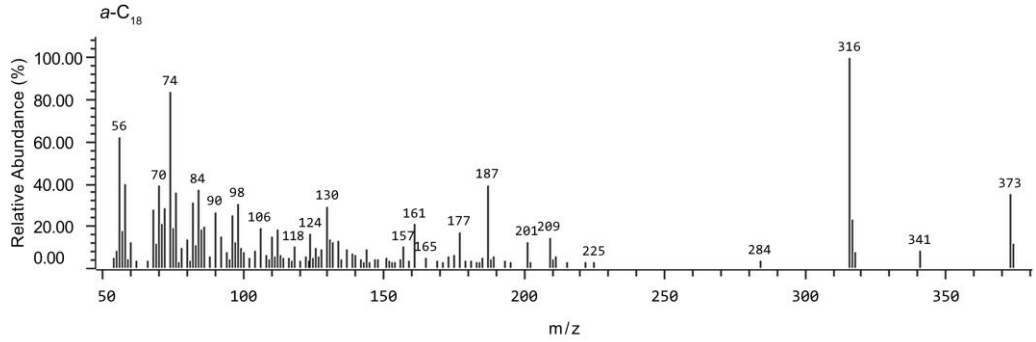

Figure S30. Mass spectrum of *a*-C_{18} 3-OH FA.

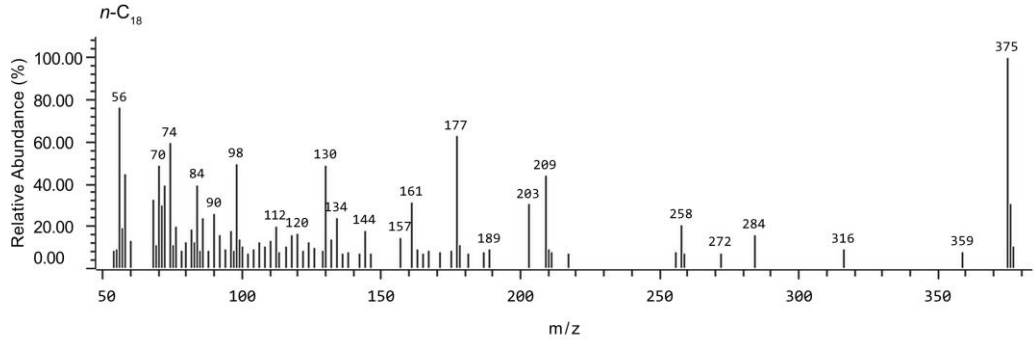

Figure S31. Mass spectrum of *n*-C_{18} 3-OH FA.