UNIVERSITY OF BIRMINGHAM

UNDERSTANDING SPATIOTEMPORAL CHANGES IN FRESHWATER BIODIVERSITY USING ENVIRONMENTAL DNA

By

Niamh Eastwood

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

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

September 2023

Abstract

Human driven environmental change is having a widespread negative impact upon the planet's biodiversity. Biodiversity is key to delivering the ecosystem functions and services which maintain a healthy environment. However, existing methods for monitoring biodiversity and understanding the interrelations between environmental change and biodiversity change are lacking. Traditional methods for biodiversity monitoring are taxonomically limited, labour intensive and low throughput, all of which results in a lack of whole community biodiversity data.

In this thesis, I set out a novel framework which utilises big data science to combine whole community biodiversity data with multiple environmental matrices and applies machine learning approaches to uncover the relationships between environmental and biodiversity change. I then applied this framework to a pilot lake to find correlations between the historical community, as measured using sedimentary environmental DNA, biocide usage and climate change. I show that the combination of these pressures can explain a large proportion of the variation in the lake community over time, underlining the value of longitudinal data analysis.

I improve on the processing of environmental DNA samples by developing a novel multiplexed metabarcoding method which utilises an early pooling approach and validate the method across biological matrices. This method offers a large reduction in library preparation cost and labour.

I then apply the multiplex method to two different sample types (water and biofilm) from over 50 lakes across England to assess spatial variation in freshwater communities. I showed that established regulatory methods under the water framework directive to classify lakes do not explain all of the variation in community diversity.

This thesis shows that whole community approaches, which capture variation in prokaryotic and eukaryotic biodiversity, can better reflect responses to environmental change and changes in ecosystem function and service delivery. This can be utilised to identify the factors, or combination of factors, most disruptive to biodiversity and therefore potential targets for regulation and remediation.

Author Contributions

This thesis is formatted as an 'alternative format' thesis, containing published works, manuscripts under review and manuscripts in preparation. I am first author or joint first author on all publications and have led data analysis and writing for all chapters. My contribution (highlighted with **Eastwood N** in the author list and **NE** in the author contribution statements) is as follows:

Chapter 1 – Adapted from preprint on Authorea: **Eastwood N**, Stubbings WA, Abdallah MA, Durance I, Paavola J, Dallimer M, Pantel JH, Johnson S, Zhou J, Brown JB, Ullah S, Krause S, Hannah DM, Crawford SE, Widmann M, Orsini L. "The wicked problem of biodiversity and ecosystem services in a changing world" 2020. DOI: 10.22541/au.160133650.04034920

Author Contributions: **NE** drafted the initial preprint with LO. **NE** adapted and added to the preprint to write this chapter. All authors made contributions to the intellectual content and interpretation of the literature review in the initial preprint.

Chapter 2 – Published in *Trends in Ecology and Evolution* as **Eastwood N,** Stubbings WA, Abou-Elwafa Abdallah MA, Durance I, Paavola J, Dallimer M, Pantel JH, Johnson S, Zhou J, Hosking JS, Brown JB, Ullah S, Krause S, Hannah DM, Crawford SE, Widmann M, Orsini L. "The Time Machine framework: monitoring and prediction of biodiversity loss" 2022 Feb;37(2):138-146. doi: 10.1016/j.tree.2021.09.008

Author Contributions: **NE** wrote the publication with LO. All authors made contributions to the content, the framework and editing of the manuscript.

Chapter 3 – Published in *eLife* as **Eastwood N**, Zhou J, Derelle R, Abdallah MA, Stubbings WA, Jia Y, Crawford SE, Davidson TA, Colbourne JK, Creer S, Bik H, Hollert H, Orsini L. “100 years of anthropogenic impact causes changes in freshwater functional biodiversity” 2023 doi: 10.7554/eLife.86576

Author contributions: **NE** carried out all laboratory work, sequencing data processing, bioinformatic analysis and interpretation of all results. First authorship is shared between **NE** and JZ. JZ created the code and ran the machine learning analyses. RD completed preliminary bioinformatics analyses. MA-EA and WS generated the DDT data. YJ, SC and HH optimised chemical assays. TAD provided the sediment, and environmental data. HB provided the barcode design. LO conceived and coordinated the study. **NE** and LO wrote the manuscript and all authors edited and approved the final manuscript.

Chapter 4 – Adapted from initial draft of published article in Environmental DNA as **Eastwood N**, Kissane S, Campbell L, Briscoe A, Egeter B, Orsini L “Single metabarcoding multiplex captures community-level freshwater biodiversity and beyond” 2024 doi: 10.1002/edn3.515

Author contributions **NE** developed the novel multiplexing method. **NE** and SK optimised the experimental conditions for the multiplex. **NE** completed all data and statistical analysis. LC, AGB and BE provided materials and protocols. LO

conceived the study and supervised data analysis. **NE** and LO wrote the manuscript and all authors edited and approved the final manuscript.

Chapter 5 – In preparation as **Eastwood N**, Watson A, Walsh K, Pitt JA, Warren J, Zhou J, Orsini L "Spatial freshwater biodiversity dynamics across English lakes"

Author Contributions **NE** carried out sequencing data processing, data analysis and statistical analysis. KW, JAP and JW provided material and data. **NE**, LO and KW designed the study. AW and JZ performed exploratory analyses (not shown). **NE** drafted the manuscript. **NE** LO JZ and AW contributed to manuscript editing. KW JAP and JW will have the opportunity to edit the manuscript.

Chapter 6 – Discussion

NE wrote this chapter.

Acknowledgements

I am deeply grateful for the incredible support and encouragement I've received, both throughout my PhD and everything which led up to it. Thank you to MIBTP for funding my PhD and the brilliant training and support I've received throughout my studentship.

Thank you to my inspirational supervisor Prof Luisa Orsini who has made this brilliant journey possible. Luisa, you have nurtured me as a scientist and helped me to develop all round. Your incredible ability to keep track of everything, steadfast belief in doing what is right and sheer enthusiasm for science over the past 6 years has had a huge impact on who I want to be both as a scientist and as a woman. To Dr Albert Zhou, thank you for so patiently explaining the most complicated elements of your work and always being so positive. To Dr Mohamed Abdallah, thank you for your support and guidance even when things didn't go to plan.

I've also made wonderful friends who have helped me within and outside of work: thank you to Steve for all of your help, your mad chats and office diversions, to all of the friendly faces (past and present) that make coming into the office so positive: Xiaojing, Archana, Hollie, Rose, Erin it is (or was!) a delight to work alongside you. Thank you to Katie for the lifesaving coffee breaks, advice on being a good scientist and all-round friendship.

Thank you to Marianne, you have motivated me to come to the office, been a cheerleader when everything felt like it was going wrong, given the best hugs

and converted me to whatsapp voicenotes. I can’t thank you enough for just being you and I am so lucky to have made a wonderful friend for life.

To Sarah, thank you for 15 (!) years of friendship and weirdness, and your support. We match now! Thank you too to Miss Rollason for hooking me on Biology so many years ago.

Finally thank you to my family. Thank you Nanny and Grandad for always being so supportive and proud. Dad, from when I was little you would always say that so long as I tried my best you would never ask for anything more - your steadfast support been so important (and the daily packed lunches have helped too). Mom, I’m not sure that I can convey all the ways I’m grateful for your endless love, unwavering support, kindness and deep strength. I wouldn’t be here today without you, from lunchbox notes in primary school to listening to me whitter on about my work now – I can’t wait for all of the happy days that we still have yet to come. Love you to the moon and back.

Table of Contents

Chapter 1: INTRODUCTION

1.1 BIODIVERSITY, ECOSYSTEM FUNCTIONS AND ECOSYSTEM SERVICES

Ecosystems are self-sustaining and self-regulating systems that provide ecosystem services (ES), including *provisioning* (food), *regulating* (e.g. climate), *supporting* (nutrient cycling, primary production), and *cultural* (e.g. aesthetic and recreational) services (Mace, Norris, and Fitter, 2012). A sustained delivery of services while maintaining the integrity of natural ecosystems is challenging because natural capital is finite and the impact of human intervention on ecosystems is uncertain and/ or unknown across different spatial, temporal, and economic scales (DeFries and Nagendra, 2017). Ecosystem management is complicated because complex and interdependent interactions among the components of an ecosystem create non-linear feedback (Hannah *et al.,* 2007), increasing the uncertainty of risks and unpredictable consequences (Fig. 1). Additionally, the ecological and economic values of these risks can diverge; limited resources, competing objectives and the need for economic profitability are at the core of the so-called 'wicked problem' – complex, open-ended and intractable - of ecosystem services (Jax *et al.,* 2018).

In the last 50 years, 60% of ES have deteriorated or been overused (Mace, Norris, and Fitter 2012). Rapid and severe biodiversity loss has been identified as its main cause (e.g. Cardinale *et al.,* 2012). International governing bodies have set targets to preserve biodiversity and ES (e.g. the Convention of Biological Diversity 2012 and the Aichi Biodiversity Targets 2019). However,

despite these efforts, the first Global Assessment of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) in 2020 found an accelerating decline in Earth's biodiversity and associated services caused by an increasing use of natural (finite) resources to meet human needs (Ruckelshaus *et al.,* 2020).

Mitigation interventions aimed at conserving and restoring natural capital have so far been inefficient and inadequate. This is in part because research on biodiversity and ES is constrained by disciplinary boundaries, whereas truly cross-disciplinary solutions are needed (Barnosky, Ehrlich, and Hadly, 2016). Disciplinary approaches may neglect process interactions, result in research undertaken at inappropriate or disconnected scales, or use discipline-specific tools which are inadequate to address cross-disciplinary questions (e.g. Roux *et al.,* 2017). Scientific knowledge has also rarely been translated into tools for stakeholders that address ecological, economic and social issues within the same framework (Ehrlich, Kareiva, and Daily, 2012). Decision-making frameworks that enable the prioritization of intervention mechanisms based on objective evidence weighting ecological as well as socio-economic priorities simply do not exist. Relative to other forms of capital assets, natural capital is undervalued by governments, businesses and the public and recognized only upon its loss (Kinzig *et al.,* 2011).

There is wide agreement that biodiversity has some intrinsic value, however there is no consensus on the explicit value of biodiversity. Biodiversity can be valued by the services derived from the organisms in the environment (the

'ecosystem services perspective') or simply due to the existence of the diversity of species and genes within the environment (the 'conservation perspective') (Mace, Norris, and Fitter, 2012). In the former, a functional role is implicit but does not reflect the values that people hold for biodiversity beyond its functional role in ecosystem processes, therefore somewhat neglecting the cultural value of biodiversity. In the latter, the role of biodiversity in underpinning ecosystem processes is ignored, and conservation efforts often focus on a subset of charismatic and/or threatened species (Smith *et al.,* 2012). Neither of these perspectives captures the complexity and dynamics of biodiversity and its links to ES.

Linking biodiversity to ecosystem function and service is challenging because of the inherent complexity of biodiversity (e.g Durance *et al.,* 2016). Growing evidence shows that biological diversity (the variation in genes, species, functional traits) is positively correlated to overall ecosystem function (resource capture, biomass production, decomposition, nutrient cycling) (Cardinale *et al.,* 2012). Higher levels of biodiversity thus imply higher levels of ecosystem function and ecosystem services. This is self evident when considering provisioning services (e.g. provision of fish), or regulating services (decomposition), but is less obvious with services of a cultural nature. These services are defined by the beneficiary and they may thus differ across individuals, communities and generations (e.g. wars over water; Shiva, 2002). However, because biodiversity can affect multifunctionality of ecosystems in non-additive ways compared to its effects on individual ecosystem processes (Perkins *et al.,* 2015), preservation of increased volumes of biodiversity can be

effective for maintaining ecosystem services. Ecologists have progressed to better understand what aspects of biodiversity - richness or evenness, in traits, genes, or phylogenies (e.g. Cadotte, 2013); community structure and composition (Spaak *et al.,* 2017) - contribute to ecosystem functions. Studies that span environmental change from community shifts to altered ecosystem functions are needed to gain a mechanistic understanding of current and future dynamics (Stockwell *et al.,* 2020).

Establishing links between biodiversity and ES is further complicated by environmental factors affecting biodiversity dynamics, including 'traditional' (e.g., overexploitation, pollution, agriculture/land-use change) and 'horizon' (e.g., climate change) threats (Fig. 1) (Bonebrake *et al.,* 2019). The main drivers of biodiversity loss include land use change, exploitation of natural resources, chemical pollution and climate change (Jaureguiberry *et al.,* 2022). However, the drivers of biodiversity loss are not necessarily the same as the drivers of biodiversity change (e.g. community composition alteration) (Jaureguiberry *et al.,* 2022), complicating the conservation of biodiversity based simply on species presence and instead underlining the need for an EF/ES approach to preserve ecosystems as a whole.

There are increasing efforts to understand the impact of chemical pollutants on biodiversity beyond simple pollution scenarios (e.g. acid rain) (Cardinale *et al.,* 2012). Ultimately, the combined effect of multiple threats has to be considered, knowing that synergistic effects among environmental threats are responsible for at least 50 and up to 68% of species or community changes (Jackson *et al.,* 2016). However, understanding biodiversity responses to multiple threats is

challenging because they affect biotic interactions – e.g. insecticides impact on freshwater invertebrates increases nutrient pollution by (toxic) algal blooms by reducing grazing pressure (Alexander *et al.,* 2013); ecological trade-offs – e.g. habitat modification can exacerbate the impacts of invasive predators (Doherty *et al.,* 2015); and patterns of local adaptation - e.g. extinction most commonly occurs when threats are outside the evolutionary experience of species or outpace adaptation (Brook, Sodhi, and Bradshaw, 2008). Moreover, causes of biodiversity loss can operate on different spatial and temporal scales (Bonebrake *et al.,* 2019) and are all context-dependent outcomes from processes operating over many years (Nogues-Bravo *et al.,* 2018). Indeed, research on the impact of chemical pollution upon biodiversity change as a whole is still lacking in comparison to the impact of other environmental stressors upon biodiversity (Sylvester *et al.,* 2023)

1.2 CURRENT METHODS OF ASSESMENT

Existing assessments of the impact of environmental stressors, including chemical pollution and climate change, tend to consider a single stress. For example, ecotoxicological assessments of new compounds are carried out in isolation, with no regulatory requirement for mixture assessment, therefore missing any synergistic effects. Indeed, research upon mixture effects typically considers the mixture of a very small number (as few as two) of chemicals (Martin *et al.,* 2021). Furthermore, the impact of these chemicals is assessed upon species in isolation, with no direct assessment of trophic interactions and therefore impacts upon larger communities. Keystone or species are used for

such assessments, however this will still not capture any dynamics of species interactions or feedback between organisms.

Ecological processes, biodiversity dynamics, and environmental change occur over time. For example, synthetic chemical production has increased at a higher rate than most other stressors between 1970 and 2010 (Bernhardt, Rosi, and Gessner, 2017). Currently, regime shifts are measured in comparison to already shifted baselines, failing to elucidate the causes and processes causing the shifts. Long-term data also provide access to baseline observations predating major environmental impact to better understand long-term dynamics leading to current patterns (Nogues-Bravo *et al.,* 2018). However, collecting these data requires commitments beyond research projects, and often beyond human lifespans.

A valuable source of longitudinal data are historical documents and paleo and archeo data (Lotze and Worm, 2009), such as species assemblages, pollen records, nutrient loads, and historical contaminant profiles in dated sediment and ice cores. This data can be used to improve environmental reconstructions that examine the impacts of land use changes, environmental stressors and other contaminants (Korosi *et al.,* 2017). Yet, retrospective characterization of past effects of contaminants in aquatic ecosystems are poorly understood which may be explained by fossil remains analysis (morphotaxonomy) requiring specialist skills (e.g. light microscopy and taxonomy), being low throughput and limited to taxonomic groups with well-preserved remains in environmental matrices (e.g. pollen, exoskeletons, bones).

To compensate for scarce longitudinal data, researchers often use space-for-time-substitutions, studying the relationships between ecological variables at sites that are assumed to be at different stages of the same succession (e.g. clines) (Damgaard, 2019). This is useful when the drivers that control biological turnover in time are the same as drivers that operate in space (Wogan and Wang, 2018), and when the dynamics studied occur over short time-frames (Banet and Trexler, 2013). However, it is problematic because it assumes static environments, requires that the observed spatial patterns are due to different realizations of the equilibrium properties of the studied ecological process, and that the temporal dynamics at the site level are ignored (Damgaard, 2019).

Moreover, current standard toxicology tests tend to consider only short term, acute effects upon organisms, missing longer term (within organism lifetime) or intergenerational effects. Some chemical compounds can persist for decades, even after regulatory changes (Silva *et al.,* 2019) meaning long term biological phenomena, such as adaptation and evolution, can come into play in the biological response. However, the relationship between evolution and ecosystem services remains largely unexplored and omitted in model forecasts (Rudman *et al.,* 2017).

Surveillance of biodiversity itself is biased towards a few taxa, often chosen for their distinctive morphology and existing knowledge of their ecological, alongside known responses to specific environmental stressors (de Olde *et al.,* 2017; Cooper, Gilmour, and Fabricius, 2009). Monitoring of a small number of taxa is attractive due to the reduced cost, versus multiple species monitoring. Morphological identification of these taxa can also give additional information

such as body condition or behaviour. Such bioindicator species are typically habitat or pressure specific, with limited overlap or relevance in other environments and stressors (Puig-Girones and Real, 2022). Morphological identification of these species can be complicated by cryptic diversity and changes in appearance throughout life stages, and also requires specialist skills which are typically not transferable to other taxonomic groups. Morphological dentification of live organisms is low throughput and prone to operator bias, as it is labour intensive for another operator to independently identify the same specimen, if it is even still available (e.g. if the specimen is released back into the environment).

Environmental stressors are also spatially heterogenous, with variances across regions, for example due to weather patterns or agricultural practices. Geographic features, such as watershed boundaries, further cause regional differences, as pollution events from sewage treatment works or farming may be concentrated in specific watercourses. Measurements of environmental factors often have limited spatial resolution, for example of the state-of-the-art General Circulation Models (GCMs) that are used to generate future climate change projections (IPCC. 2018)) are limited to approx. 100km^2. Whilst the change in global mean temperature can be simulated with relatively small uncertainties, regional changes in temperature and precipitation can carry substantial uncertainties due to biases in the simulated atmospheric circulation (e.g. Woollings, 2010).

In this thesis I focus on freshwater ecosystems and their services because these ecosystems are ecologically and societally valuable, delivering important

ecosystem services (e.g. clean water, food provision and recreation) and under increasing threat of destruction and degradation (Ruckelshaus *et al.,* 2020). Freshwater habitats are diverse, including drylands, wetlands, streams/rivers and ponds/lakes, and geographically widely distributed, ranging from the poles to the equator. Many water-dependent ecosystems are of high conservation value. Freshwater ecosystems are also among the most impacted by biodiversity loss, with an 83% reduction in the global freshwater living planet index since 1970 (WWF, 2018). Lakes, and their communities, are particularly vulnerable to chemical pollution as the pollutants within the watershed of the lake can be concentrated into the lake (Gloege *et al.,* 2020).

1.3 LONG TERM DATA

Longitudinal data are paramount to explain present-day patterns and to model cause-effect relations between biodiversity and environmental change (Baert *et al.,* 2016). They are valuable for studying biodiversity dynamics bridging the gap between the timescale of empirical observations (at most one year) and the scales at which the underlying biological processes take place (at least multiple decades) (Balint, Pfenninger, *et al.,* 2018).

Coupling longitudinal biodiversity data with the continuous assessment of environmental change offers enormous potential to establish the cause-effect relationships between environmental variables and biodiversity change or loss. Continuous biodiversity longitudinal data spanning evolutionary timescales coupled with environmental change records are needed to explicitly test hypotheses concerning the drivers of long-term biodiversity change. These data

help understanding population continuity, migration, ecosystem structure and evolution (Nogues-Bravo *et al.,* 2018; Orsini, Schwenk, *et al.,* 2013). Moreover, they enable one to disentangle the impact of extreme events (e.g. droughts, heat waves, toxic algal blooms) from the one of trends occurring on longer time scales (e.g. average temperature) (Jiguet *et al.,* 2006). Continuous longitudinal data enable us to establish the cause-effect relationship between specific pollutants (e.g. pesticides) and biodiversity decline (Ashfaq *et al.,* 2019).

1.4 HOLISTIC COMMUNITY ASSESSMENT

DNA extracted from environmental matrices is referred to as 'Environmental DNA' or eDNA; this term emerged at the beginning of the 2000s (Cristescu and Hebert, 2018). eDNA has the potential to provide whole community data in biodiversity (Balint, Pfenninger, *et al.,* 2018), proving particularly advantageous when experimental manipulations are not feasible, early warning of 'regime-shifts' critical for ecosystem management have to be identified (Pace, Carpenter, and Cole 2015), and a broad taxonomic coverage is required to reconstruct community dynamics (Cristescu and Hebert, 2018).

DNA extracted from environmental matrices and macrofossils combined with genomics and high throughput sequencing technologies is used in conservation biology to estimate human impact on biodiversity (Hofman *et al.,* 2015), in invasion biology to identify timing and severity of alien species invasion (Ruppert, Kline, and Md Rahman, 2019) and biodiversity science by enabling the census of species/taxa on a global scale in real time (Garlapati *et al.,* 2019).

In addition, eDNA can provide continuous temporal data that are key to identifying temporal baselines and to assess changes in biodiversity over time. Quantitative identification of single species can be achieved using targeted quantitative PCR. It is a low-cost/high accuracy approach, suitable for single species identification such as for protected species surveillance programmes, e.g. Great Crested Newts (*Triturus cristatus*) (Rees *et al.,* 2014). However, quantitative PCR requires knowledge of the genetics of the species of interest for detection probe design. Semi-quantitative, whole community biodiversity data can be obtained in a high throughput fashion by using markers gene regions (metabarcoding) which can capture specific broad taxonomic groups (Deiner *et al.,* 2017) or even total eDNA sequencing (metagenomics). As eDNA-based approaches move from single-marker to multi-marker approaches, eDNA becomes an increasingly viable method for whole community surveillance (Cordier *et al.,* 2019). However, the poor representation of some ecosystems, taxa and genes in reference databases used for taxonomic assignment can lead to unassigned or incorrectly assigned sequencing reads (Kvist 2013). There are open questions surrounding the persistence of such eDNA with environmental conditions such as temperature, humidity, pH and matrix, alongside the impact of biological processes in eDNA production and environmental degradation. Technical challenges associated with incomplete reference databases also limit analysis of some data. Multiple marker metabarcoding of eDNA has the potential to explore ecosystem-level processes and community diversity, identifying tipping points of ecosystem stability and nuanced or whole community responses to environmental pressures.

Environmental DNA can be analysed in a high throughput manner by processing hundreds of samples using robotics coupled with high throughput sequencing methods. This enables the comparison of multiple sites simultaneously, and with standardised replicated methods, sites can be compared across time over multiple seasons or years. Taxonomic identification of eDNA samples does not require specialist taxonomic knowledge, but instead the comparatively widespread molecular biological expertise, which is increasingly commercially accessible.

Well preserved DNA (such as from sediment, termed sedimentary ancient DNA) can be utilised to address the challenges previously identified in lack of long term biological data and reconstruct the past occurrence of some communities comparably with standard methods (Thorpe *et al.,* 2022).

A potential benefit of eDNA based approaches for environmental monitoring is the analysis of compositional shifts and relative abundance of taxonomic units which may not have a Linnaean name. Many taxa have poorly described ecology and therefore may be neglected in typical conservation approaches due to a lack of knowledge of the particular species. Instead, by analysing operational taxonomic units responsible for ecosystem stability or delivery of ecosystem services, lack of traditional knowledge may be a lower barrier to conservation. Interventions based upon ecological function could favour the re-establishment of natural complexity and feedback, accelerating ecological restoration of natural ecosystems. This system-level approach using requires the modernization of current environmental practice and a shift from bioindicator-based (morpho taxonomic of fossil remains (Boere *et al.,* 2011) or

bioindicators, such as the Trophic Diatom Index (TDI) (Kelly and Whitton, 1995) to whole-system, holistic monitoring at community scale.

Lack of understanding of the processes that underpin ecosystem services has often led to mismanagement with clear dis-benefits for the environment, the economy and human well-being. Approaches and tools constrained by discipline boundaries have been unable to capture process interactions and different spatio-temporal scales necessary to address the so-called wicked problem of biodiversity and ES (Ottoni *et al.,* 2023; Schutter and Hicks, 2021). The divide between academia and practitioners as well as the lack of training for stakeholders and practitioners on state-of-the-art tools hinders the translation of science into practice, and results in outdated tools for environmental practice.

In this thesis chapter 2 outlines a novel framework combining long term community level biodiversity data, such as that derived from lake sediment cores and matrices of multiple environmental factors to forecast the impact of environmental change upon biodiversity. This enables the prioritisation of intervention mechanisms based on objective evidence weighting ecological as well as socio-economic priorities. Mitigating the loss of biodiversity and the targeted regulation of pollutants with severe adverse effects on biodiversity improves ecosystem integrity and resilience. Resilience enables sustained delivery of ecosystem functions with a positive cascading effect on ES e.g, provisioning (food, clean water) and regulating services (climate mitigation, nutrient cycling). Chapter 3 is a pilot study of this framework, applying multimarker metabarcoding and machine learning methods to a sediment core

from a Danish lake with a well-documented history of human impact. This chapter demonstrates the depth of knowledge gained from whole community surveillance, as opposed to traditional taxa specific methods and the ability to identify chemical targets for mitigation. Chapter 4 builds on the multimarker metabarcoding approach to develop a novel multiplex method which reduces cost and validates its use across sample matrices. Chapter 5 applies the multiplex developed in chapter 4 to investigate spatial variance in freshwater communities across England and how well this variance is captured by the standard regulatory surveillance.

1.5 REFERENCES

Alexander, A. C. , A. T. Luis, J. M. Culp, D. J. Baird, and A. J. Cessna. 2013. 'Can nutrients mask community responses to insecticide mixtures?', *Ecotoxicology* 22: 1085-100.

Ashfaq, M. Y., M. A. Al-Ghouti, H. Qiblawey, D. F. Rodrigues, Y. D Hu, and N. Zouari. 2019. ' Isolation, identification and biodiversity of antiscalant degrading seawater bacteria using MALDI-TOF-MS and multivariate analysis', *Science of the Total Environment*, 656: 910-20.

Baert, J. M., C. R. Janssen, K. Sabbe, and F. De Laender. 2016. 'Per capita interactions and stress tolerance drive stress-induced changes in biodiversity effects on ecosystem functions', *Nat Commun*, 7: 12486.

Balint, M., M. Pfenninger, H. P. Grossart, P. Taberlet, M. Vellend, M. A. Leibold, G. Englund, and D. Bowler. 2018. 'Environmental DNA Time Series in Ecology', *Trends Ecol Evol*, 33: 945-57.

Banet, A.I. , and J.C. Trexler. 2013. 'Space-for-time substitution works in everglades ecological forecasting models', *PLoS One*, 8: e81025.

Barnosky, A. D. , P. R. Ehrlich, and E. A. Hadly. 2016. 'Avoiding collapse: Grand challenges for science and society to solve by 2050', *Elementa: Science for the Anthropocene*, 4: 000094.

Bernhardt, E. S., E. J. Rosi, and M. O. Gessner. 2017. 'Synthetic chemicals as agents of global change', *Frontiers in Ecology and the Environment*, 15: 84-90.

Boere, A. C., W. I. Rijpstra, G. J. De Lange, J. S. Sinninghe Damste, and M. J. Coolen. 2011. 'Preservation potential of ancient plankton DNA in Pleistocene marine sediments', *Geobiology*, 9: 377-93.

Bonebrake, T. C., F. Guo, C. Dingle, D. M. Baker, R. L. Kitching, and L. A. Ashton. 2019. 'Integrating Proximal and Horizon Threats to Biodiversity for Conservation', *Trends Ecol Evol*, 34: 781-88.

Brook, B.W. , N.S. Sodhi, and C.J.A. Bradshaw. 2008. 'Synergies among extinction drivers under global change', *Trends Ecol Evol*, 23: 453-60.

Cadotte, M.W. . 2013. 'Experimental evidence thatevolutionarily diverse assemblages result inhigher productivity', *Proceedings of the National Academy of Science USA*, 110: 8996–9000.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, D. A. Wardle, A. P. Kinzig, G. C. Daily, M. Loreau, J. B. Grace, A. Larigauderie, D. S. Srivastava, and S. Naeem. 2012. 'Biodiversity loss and its impact on humanity', *Nature*, 486: 59-67.

Cooper, T. F., J. P. Gilmour, and K. E. Fabricius. 2009. 'Bioindicators of changes in water quality on coral reefs: review and recommendations for monitoring programmes', *Coral Reefs*, 28: 589-606.

Cordier, T., F. Frontalini, K. Cermakova, L. Apotheloz-Perret-Gentil, M. Treglia, E. Scantamburlo, V. Bonamin, and J. Pawlowski. 2019. 'Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy)', *Marine Environmental Research*, 146: 24-34.

Cristescu, M. E., and P.D.N. Hebert. 2018. 'Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation', *Annual review of ecology, evolution, and systematics*, 49: 209–30.

Damgaard, C. 2019. 'A Critique of the Space-for-Time Substitution Practice in Community Ecology', *Trends Ecol Evol*, 34: 416-21.

de Olde, E. M., H. Moller, F. Marchand, R. W. McDowell, C. J. MacLeod, M. Sautier, S. Halloy, A. Barber, J. Benge, C. Bockstaller, E. A. M. Bokkers, I. J. M. de Boer, K. A. Legun, I. Le Quellec, C. Merfield, F. W. Oudshoorn, J. Reid, C. Schader, E. Szymanski, C. A. G. Sorensen, J. Whitehead, and J. Manhire. 2017. 'When experts disagree: the need to rethink indicator selection for assessing sustainability of agriculture', *Environment Development and Sustainability*, 19: 1327-42.

DeFries, R., and H. Nagendra. 2017. 'Ecosystem management as a wicked problem', *Science* 356: 265-70

Deiner, K., H. M. Bik, E. Machler, M. Seymour, A. Lacoursiere-Roussel, F. Altermatt, S. Creer, I. Bista, D. M. Lodge, N. de Vere, M. E. Pfrender, and L. Bernatchez. 2017. 'Environmental DNA metabarcoding: Transforming how we survey animal and plant communities', *Molecular Ecology*, 26: 5872-95.

Doherty , T.S. , C. R. Dickman, D.G. Nimmo, and E. G. Ritchie. 2015. 'Multiple threats, or multiplying the threats? Interactions between invasive predators and other ecological disturbances', *Biological Conservation*, 190: 60-68.

Durance, I., M. W. Bruford, R. Chalmers, N. A. Chappell, M. Christie, B. J. Cosby, D. Noble, S. J. Ormerod, H. Prosser, A. J. Weightman, and G.

Woodward. 2016. 'The challenges of linking ecosystem services to biodiversity. ', *Advances in Ecological Research*, 54: 87-134.

Ehrlich, P. R., P. M. Kareiva, and G. C. Daily. 2012. 'Securing natural capital and expanding equity to rescale civilization', *Nature*, 486: 68-73.

Garlapati, D. , B. Charankumar, K. Ramu, P. Madeswaran, and M. V. Ramana Murthy. 2019. 'A review on the applications and recent advances in environmental DNA (eDNA) metagenomics', *Review in Environmental Science and Biotechnology*, 18: 389–411.

Gloege, L., G. A. McKinley, R. J. Mooney, J. D. Allan, M. W. Diebel, and P. B. McIntyre. 2020. 'Lake hydrodynamics intensify the potential impact of watershed pollutants on coastal ecosystem services', *Environmental Research Letters*, 15.

Hofman, C. A., T. C. Rick, R. C. Fleischer, and J. E. Maldonado. 2015. 'Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene', *Trends Ecol Evol*, 30: 540-9.

IPCC. 2018. "Global Warming of 1.5°C." In An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, edited by V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla and *et al.,*

Jackson, M. C., C. J. G. Loewen, R. D. Vinebrooke, and C. T. Chimimba. 2016. 'Net effects of multiple stressors in freshwater ecosystems: a meta-analysis', *Global change biology*, 22: 180-89.

Jaureguiberry, P., N. Titeux, M. Wiemers, D. E. Bowler, L. Coscieme, A. S. Golden, C. A. Guerra, U. Jacob, Y. Takahashi, J. Settele, S. Diaz, Z. Molnar, and A. Purvis. 2022. 'The direct drivers of recent global anthropogenic biodiversity loss', *Science Advances*, 8.

Jax, K., Eeva Furman, H. Saarikoski, D. N. Barton, B. Delbaere, J. Dick, G. Duke, C. Görg, E. Gómez-Baggethun, P.A. Harrison, J. Maes, M. Pérez-Soba, Sanna-Riikka Saarela, F. Turkelboom, J. van Dijk, and A. D. Watt. 2018. 'Handling a messy world: Lessons learned when trying to make the ecosystem services concept operational', *Ecosystem Services*, 29: 415-27.

Jiguet, F., R. Julliard, C. D. Thomas, O. Dehorter, S. E. Newson, and D. Couvet. 2006. 'Thermal range predicts bird population resilience to extreme high temperatures', *Ecol Lett*, 9: 1321-30.

Kelly, M. G. , and B. A. Whitton. 1995. 'The Trophic Diatom Index: a new index for monitoring eutrophication in rivers', *Journal ofApplied Phycology*, 7: 433-44.

Kinzig, A. P., C. Perrings, F. S. Chapin, 3rd, S. Polasky, V. K. Smith, D. Tilman, and B. L. Turner, 2nd. 2011. 'Sustainability. Paying for ecosystem services--promise and peril', *Science*, 334: 603-4.

Korosi, J. B., J. R. Thienpont, J. P. Smol, and J. M. Blais. 2017. 'Paleo-ecotoxicology: What Can Lake Sediments Tell Us about Ecosystem Responses to Environmental Pollutants?', *Environ Sci Technol*, 51: 9446-57.

Kvist, S. 2013. 'Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge', *Mol Phylogenet Evol*, 69: 39-45.

Lotze, H. K., and B. Worm. 2009. 'Historical baselines for large marine animals', *Trends Ecol Evol*, 24: 254-62.

Mace, G. M., K. Norris, and A. H. Fitter. 2012. 'Biodiversity and ecosystem services: a multilayered relationship', *Trends Ecol Evol*, 27: 19-26.

Martin, O., M. Scholze, S. Ermler, J. McPhie, S. K. Bopp, A. Kienzler, N. Parissis, and A. Kortenkamp. 2021. 'Ten years of research on synergisms and antagonisms in chemical mixtures: A systematic review and quantitative reappraisal of mixture studies', *Environment International*, 146.

Nogues-Bravo, D., F. Rodriguez-Sanchez, L. Orsini, E. de Boer, R. Jansson, H. Morlon, D. A. Fordham, and S. T. Jackson. 2018. 'Cracking the Code of Biodiversity Responses to Past Climate Change', *Trends Ecol Evol*, 33: 765-76.

Orsini, L., K. Schwenk, L. De Meester, J. K. Colbourne, M. E. Pfrender, and L. J. Weider. 2013. 'The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments', *Trends Ecol Evol*, 28: 274-82.

Ottoni, F. P., South, J., Azevedo-Santos, V. M., Henschel, E., & Bragança, P. H. N. (2023). Editorial: Freshwater biodiversity crisis: Multidisciplinary approaches as tools for conservation. *Frontiers in Environmental Science*, *11*. https://doi.org/10.3389/fenvs.2023.1155608

Pace, M.L , S. R. Carpenter, and J.J. Cole. 2015. 'With and without warning: managing ecosystems in a changing world', *The Ecological Society of America*, 13: 460-67.

Perkins, D. M., R. A. Bailey, M. Dossena, L. Gamfeldt, J. Reiss, M. Trimmer, and G. Woodward. 2015. 'Higher biodiversity is required to sustain multiple ecosystem processes across temperature regimes', *Global change biology*, 21: 396-406.

Puig-Girones, R., and J. Real. 2022. 'A comprehensive but practical methodology for selecting biological indicators for long-term monitoring', *PLoS One*, 17.

Rees, H. C., K. Bishop, D. J. Middleditch, J. R. M. Patmore, B. C. Maddison, and K. C. Gough. 2014. 'The application of eDNA for monitoring of the Great Crested Newt in the UK', *Ecology and Evolution*, 4: 4023-32.

Roux, D.J., J. L. Nel, G. Cundill, P. O'Farrell, and C. Fabricius. 2017. 'Transdisciplinary research for systemic change: who to learn with, what to learn about and how to learn', *Sustainable Science*, 12.

Ruckelshaus, M. H., S. T. Jackson, H. A. Mooney, K. L. Jacobs, K. S. Kassam, M. T. K. Arroyo, A. Baldi, A. M. Bartuska, J. Boyd, L. N. Joppa, A. Kovacs-Hostyanszki, J. P. Parsons, R. J. Scholes, J. F. Shogren, and Z. Ouyang. 2020. 'The IPBES Global Assessment: Pathways to Action', *Trends Ecol Evol*, 35: 407-14.

Rudman, S. M., M. Kreitzman, K. M. A. Chan, and D. Schluter. 2017. 'Evosystem Services: Rapid Evolution and the Provision of Ecosystem Services', *Trends Ecol Evol*, 32: 403-15.

Ruppert, K. M. , R. J. Kline, and S. Md Rahman. 2019. 'Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA', *Global Ecology and Conservation*, 17: e00547.

Schutter, M. S., & Hicks, C. C. (2021). Speaking across boundaries to explore the potential for interdisciplinarity in ecosystem services knowledge production. *Conservation Biology*, *35*(4), 1198–1209. https://doi.org/10.1111/cobi.13659

Shiva, V. . 2002. *Water wars: privatization, pollution and profit.* (South End Press: Cambridge).

Silva, V., H. G. J. Mol, P. Zomer, M. Tienstra, C. J. Ritsema, and V. Geissen. 2019. 'Pesticide residues in European agricultural soils - A hidden reality unfolded', *Science of the Total Environment*, 653: 1532-45.

Smith, R. J. , D. Veríssimo, N. J.B. Isaac, and K. E. Jones. 2012. 'Identifying Cinderella species: uncovering mammals with conservation flagship appeal', *Conservation Letters*, 5: 205–12

Spaak, J.W., J.M. Baert, D.J. Baird, N. Eisenhauer, L. Maltby, F. Pomati, V. Radchuk, J.R. Rohr, P.J. Van den Brink, and F. De Laender. 2017. 'Shifts of community composition and population density substantially affect ecosystem function despite invariant richness', *Ecology Letters*, 20: 1315-24.

Stockwell, J.D., J.P. Doubek, R. Adrian, O. Anneville, C.C. Carey, L. Carvalho, L.N. De Senerpont Domis, G. Dur, M.A. Frassl, H.P. Grossart, and B.W. Ibelings. 2020. 'Storm impacts on phytoplankton community dynamics in lakes', *Global change biology*, 26: 2756-84.

Sylvester, F., F. G. Weichert, V. L. Lozano, K. J. Groh, M. Balint, L. Baumann, C. Bassler, W. Brack, B. Brandl, J. Curtius, P. Dierkes, P. Doll, I. Ebersberger, S. Fragkostefanakis, E. J. N. Helfrich, T. Hickler, S. Johann, J. Jourdan, S. Klimpel, H. Kminek, F. Liquin, D. Mollendorf, T. Mueller, J. Oehlmann, R. Ottermanns, S. U. Pauls, M. Piepenbring, J. Pfefferle, G. J. Schenk, J. F. Scheepens, M. Scheringer, S. Schiwy, A. Schlottmann, F. Schneider, L. M. Schulte, M. Schulze-Sylvester, E. Stelzer, F. Strobl, A. Sundermann, K. Tockner, T. Troger, A. Vilcinskas, C. Volker, R. Winkelmann, and H. Hollert. 2023. 'Better integration of chemical pollution research will further our understanding of biodiversity loss', *Nature Ecology & Evolution*.

Thorpe, A., A. Anderson, T. Goodall, S. Thackeray, S. Maberly, J. Bendle, H. Gweon, and R. Read. 2022. 'Sedimentary DNA records long-term changes in a lake bacterial community in response to varying nutrient availability', *Environmental DNA*, 4: 1340-55.

Wogan, G.O.U. , and I.J. Wang. 2018. 'The value of space-for-time substitution for studying fine-scale microevolutionary processes.', *Ecography* 41: 1456–68.

Woollings, T. 2010. 'Dynamical influences on European climate: an uncertain future', *Philos Trans A Math Phys Eng Sci*, 368: 3733-56.

WWF. 2018. "Living Planet Report: aiming higher." In, edited by M. Grooten and R.E.A. Almond. Gland, Switzerland.

Chapter 2: THE TIME MACHINE FRAMEWORK: MONITORING AND PREDICTION OF BIODIVERSITY LOSS

N. Eastwood*[1], W.A. Stubbings[2], M.A. Abou-Elwafa Abdallah[2], I. Durance[3], J. Paavola[4], M. Dallimer[4], J.H. Pantel[5], S. Johnson[6], J. Zhou[1], J.S. Hosking[7, 8], J.B. Brown[9], S. Ullah[2], S. Krause[2], D. M. Hannah[2], S.E. Crawford[10], M. Widmann[2] and L. Orsini[1, 8].

* First Author

[1]Environmental Genomics Group, School of Biosciences, The University of Birmingham, Birmingham, B15 2TT, UK

[2]School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[3]School of Biosciences and Water Research Institute, Cardiff University, Cardiff, CF10 3AX, UK

[4]Sustainability Research Institute, School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK

[5]Department of Computer Science, Mathematics, and Environmental Science, The American University of Paris, 6 rue du Colonel Combes, 75007 Paris, France

[6]School of Mathematics, University of Birmingham, Birmingham, B15 2TT, UK; and Alan Turing Institute, London, NW1 2DB, UK

[7]British Antarctic Survey, Natural Environment Research Council, Cambridge, CB3 0ET, UK

[8]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK

[9]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[10]Institute of Ecology, Evolution and Diversity, Dept. of Evolutionary Ecology and Environmental Toxicology, Goethe University Frankfurt, 60438, Germany

Author Contributions: Author Contributions: **NE** wrote the publication with LO. All authors made contributions to the content, the framework and editing of the manuscript.

2.1 ABSTRACT

Trans-disciplinary solutions are needed to achieve the sustainability of ecosystem services for future generations. We propose a framework to identify the causes of ecosystem function loss and to forecast the future of ecosystem services under different climate and pollution scenarios. The framework i) applies an artificial intelligence time-series analysis to identify relationships among environmental change, biodiversity dynamics and ecosystem functions; ii) validates relationships between loss of biodiversity and environmental change in fabricated ecosystems; and iii) forecasts the likely future of ecosystem services and their socio-economic impact under different pollution and climate scenarios. We illustrate the framework by applying it to watersheds,

and provide system-level approaches that enable natural capital restoration by associating centennial biodiversity changes to chemical pollution.

2.2 THE CHALLENGE OF SUSTAINED ECOSYSTEM SERVICES

Biodiversity is directly linked to healthy ecosystems which provide provisioning (e.g. food), regulating (e.g. climate), supporting (e.g. nutrient cycling, primary production), and cultural (e.g. aesthetic and recreational) services (Cardinale *et al.,* 2012; Durance *et al.,* 2016; Mace, Norris, and Fitter 2012; Spaak *et al.,* 2017). At the global level, rapid and severe biodiversity loss has been identified as the main cause of deterioration of more than 60% of ecosystem services (Cardinale *et al.,* 2012; Mace, Norris, and Fitter 2012)), affected by various factors (Bonebrake *et al.,* 2019). Chemical pollution, habitat loss, unsustainable use of resources, invasive species and climate change are among the main acknowledged threats to biodiversity (Backhaus, Snape, and Lazorchak 2012; Jackson *et al.,* 2016).

The sustained delivery of ecosystem services in the face of these threats is challenging because natural capital is finite and the impact of human interventions on ecosystems is uncertain and/ or unknown across different spatial, temporal, and economic scales (DeFries and Nagendra 2017). Ecosystem management that ensures the delivery of services while preserving natural capital is a complex, open-ended problem because of limited resources, competing objectives and the need for economic profitability (Jax *et al.,* 2018). This is because:

i) biodiversity loss happens on different spatial and temporal scales, and dynamics are context-dependent outcomes stemming from processes operating over many years (Nogues-Bravo *et al.,* 2018; Orsini, Schwenk, *et al.,* 2013). The cumulative effect of processes over time can cause delayed dynamics also

known to cause so-called extinction debts (Figueiredo *et al.,* 2019). Even state-of-the-art environmental monitoring fails to capture effects that may arise from cumulative effects over time of multiple threats (Bonebrake *et al.,* 2019)). Only by quantifying trajectories of abiotic, biotic and functional systemic change before, during and after pollution events, can we identify the causes of biodiversity and ecosystem function loss;

ii) research on biodiversity and ecosystem services is often constrained by disciplinary boundaries, whereas biodiversity loss has ecological, social and economic implications (Barnosky, Ehrlich, and Hadly 2016). Discipline-constrained approaches may neglect process interactions, result in research undertaken at inappropriate or disconnected scales, or use discipline-specific tools that are inadequate to address cross-disciplinary questions (Roux *et al.,* 2017).

iii) decision-making frameworks that enable the prioritization of interventions for the sustainable use of ecosystems typically require multiple lines of evidence from different disciplines, making decisions by stakeholders challenging, especially when relationships between socio-economic and ecological priorities are not linear (e.g. (Grêt-Regamey *et al.,* 2017; Pinto *et al.,* 2014)). While such decision-making frameworks are being developed, they still often fail to cover all types of ecosystem services, particularly the cultural ones (Grêt-Regamey *et al.,* 2017).

We propose a ‘Time Machine’ framework that: 1) establishes spatiotemporal correlations among biotic, abiotic and ecosystem functional changes using multidecadal to millennial continuous data; 2) provides evidence for cause-

effect relations through experimental validation in fabricated ecosystems from correlations identified in point 1; and 3) generates likely predictions of future ecosystem services under different pollution and climate scenarios, driven by localised and regional environmental change and mediated by changes in overall biotic interactions (Fig. 1).

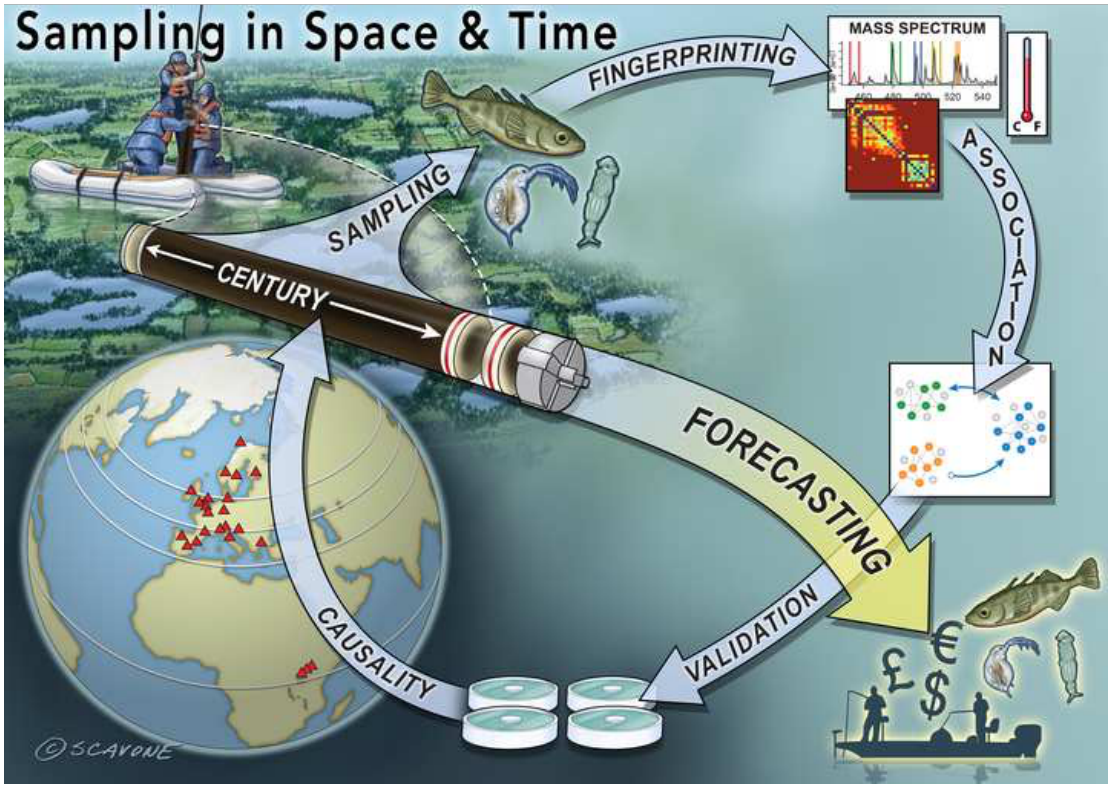


Figure 1. The Time Machine framework

The Time Machine Framework is shown to establish past correlations between biodiversity, ecosystem functions and environmental change, and generate the likely forecasts of ecosystem services under different pollution and climate scenarios. Sedimentary archives from watersheds are sampled at continental scale and across climatic regimes. Sampled and dated sedimentary archives are used to *fingerprint* biodiversity, chemical pollution and ecosystem functions. Other environmental factors (e.g. temperature) are collected from weather stations or retrieved from public databases. Associations between biodiversity, ecosystem functions and environmental factors are established with the use of AI combined with explainable network models. Associations are then validated in fabricated ecosystems that are perturbed to identify cause-effect relations between biotic and abiotic changes. Model forecasts that 'learn' from past

dynamics and feedback are tested iteratively against real data and refined to predict the future of biodiversity, ecosystems services and their economic value.

2.3 THE TIME MACHINE FRAMEWORK

We illustrate the framework in five main *Steps* for applications in freshwater ecosystems, because they are diverse, geographically distributed and of high conservation value (Fig 1); they deliver important ecosystem services such as clean water, food and recreation, and are under increasing threat of destruction and degradation (Durance *et al.,* 2016; Ruckelshaus *et al.,* 2020).

Step 1 - *Sampling* through time and space, using Lake sedimentary archives (Fig. 1; *Sampling*). Sedimentary archives preserve biological and environmental signals, providing a continuous record of changes from a reference baseline predating major human impact on climate and biodiversity (Anthropocene) to impacted environments (Baert *et al.,* 2016); these characteristics allow better understanding of temporal dynamics of biotic, abiotic and ecosystem functional data leading to current patterns (Nogues-Bravo *et al.,* 2018). To disentangle patterns driven by stochasticity from patterns driven by environmental change, semi-pristine sites (e.g. alpine lakes) can be used as reference. Whereas all natural communities experience changes over time, dynamics in semi-pristine sites will likely be driven by demographic and environmental stochasticity, which results in predictable community dynamics (Shoemaker *et al.,* 2020; Vellend *et al.,* 2014). Conversely, dynamics at impacted sites will likely be driven by exogenous environmental factors, which leave signatures that depart from stochasticity (e.g. (Wang *et al.,* 2018)).

Step 2 - Biochemical and ecosystem functions *fingerprinting*. Spatiotemporal biotic assemblages are established by applying **metabarcoding** to **environmental DNA** or **eDNA** (see Glossary) (Cristescu and Hebert 2018) extracted from dated sediment layers of the biological archives (Fig. 1; *Fingerprinting*). DNA extracted from environmental matrices provides the means to collect continuous temporal data over time and space (Balint, Pfenninger, *et al.,* 2018). These data inform conservation biology by estimating human impact on biodiversity (Tse *et al.,* 2018), invasion biology by identifying timing and severity of alien species invasion (Ruppert, Kline, and Md Rahman 2019) and biodiversity science by enabling the census of species/taxa on a global scale in real time (Garlapati *et al.,* 2019). Through DNA sequence similarity, molecular operational taxonomic units (MOTUs) are identified by matching sequence similarity to records in public databases (e.g. NCBI, SILVA), allowing the analysis of compositional shifts and relative abundance of known and unknown taxonomic units. MOTUs enable the retrospective characterization of past community-level dynamics (e.g. microbes, plants and animals) without requiring specialist skills (e.g. light microscopy and taxonomy) and without being limited to taxonomic groups with well-preserved remains in environmental matrices (e.g. pollen, exoskeletons, bones) (Gillson and Marchant 2014). On the same sediment samples, abiotic changes (e.g. presence and concentrations of chemical mixtures) are quantified using high resolution mass spectrometry (MS), combining **suspect and non-target screening** of chemical pollutants in sediment samples (Abdallah *et al.,* 2019) (Fig. 1; *Fingerprinting*). The spatiotemporal biochemical fingerprinting is complemented by the analysis of

ecosystem functions [e.g. biogeochemical functions measured as the accumulation rate of total organic carbon (C), nitrogen (N) and phosphorus (P)] via bulk stoichiometry of sediments to help elucidate long-term dynamics in productivity as influenced by nutrient availability, and the relationships of stoichiometric ratios, productivity, decomposition and biological attributes. A potential caveat of reconstructing temporal community records from sediment matrices is the preservation state of DNA that may be affected by climatic conditions (e.g. warmer/humid climates influence mineralization; Outstanding Questions). However, metabarcoding has been successfully applied to fossil or remnant DNA as far back as the Holocene (e.g. (Hirai *et al.,* 2017)). In addition, only relative abundance of MOTUs can be quantified from community assemblages. This enables the reconstruction only of relative changes in MOTU abundance between sites and time points.

Step 3 - Establishing *associations* between past biodiversity, abiotic and functional changes with the use of Artificial Intelligence (AI) (Fig. 1; A*ssociation*). Combining **explainable network models** with **multi-view learning** (Tenenhaus *et al.,* 2014), co-varying elements (nodes) within and between networks are identified, where networks can represent MOTUs, environmental variables and pollutants (Box 1). This leads to the identification of interacting environmental factors putatively driving changes in MOTUs and ecosystem functions. These correlations are then validated experimentally in fabricated ecosystems as explained in Step 4. In a pilot study conducted on a natural lake, we applied the AI approach and determined that the decline in a specific taxonomic group of primary producers (e.g. green algae) was inversely

correlated with ten herbicides among the hundreds that were quantified in the sediment (Box 1).

Step 4 - Establishing *causality*. Fabricated ecosystems are used to establish causal links between the associations identified in Step 3 (Fig. 1; *Causality*). Focusing on short-term dynamics (weeks to months), fabricated ecosystems, such as artificial ponds, are used to isolate effects, explore controlled interactions among multiple factors, and determine causality among MOTUs' relative abundance (quantified via eDNA metabarcoding), ecosystem functions (P, N, C content measured through nutrient stoichiometry), climate variables and chemical pollutants (quantified with MS). Natural communities are used to inoculate fabricated ecosystems, which are then exposed to the environmental factors identified in Step 3 to co-vary with MOTUs. For example, the effect of global warming on biodiversity and ecosystem functions can be quantified by measuring MOTUs dynamics in fabricated ecosystems exposed to current temperature and in ecosystems exposed to temperature plus 2.5°C, representing the IPCC mid-range forecasts (RCP 6.0). While experimental results investigating short-term dynamics may not be directly comparable to natural dynamics, they provide a validation of observed trends in natural ecosystems for extrapolations to long-term dynamics using e.g. **machine learning** algorithms. The fabricated ecosystems serve a dual role of providing an experimental validation of a) observed past correlative patterns, and b) predictions of biodiversity and ecosystem functional changes in different climate and pollution scenarios (see Step 5).

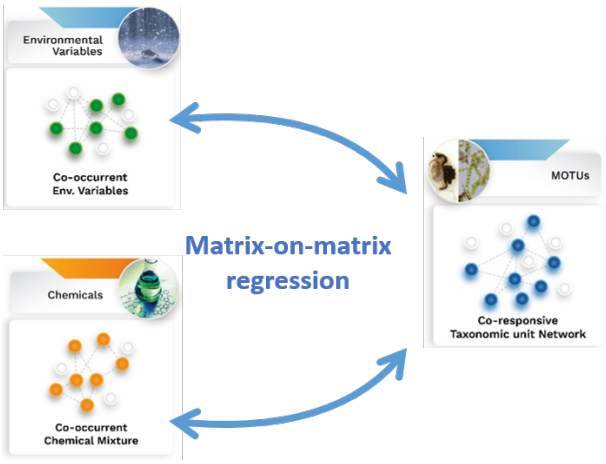
Step 5 - *Forecasting* biodiversity and ecosystem services. **Ecological process-based models**, informed by the associations identified by the AI in Step 3, are used to generate predictions about projected future states of freshwater ecosystems (Fig. 1; *Forecasting*). These predictive models are incorporated into simulations that project solutions for ecosystem services (Fig. 2). Although it is likely that a range of complex processes interact to determine how environmental drivers and associated biodiversity shifts influence ecosystem functions, we propose to start with a simple model that illustrates the impacts of community composition and structure, species interactions, and the covariation of these with environmental factors on freshwater ecosystem functions. Process-based models with these components include PCLake (Janssen *et al.,* 2019), a process-based model that links species composition, environmental drivers (e.g. nutrient loading, temperature), and ecosystem responses (e.g. water quality; (Janse *et al.,* 2010)). Alternatively, ELCOM–CAEDYM, a coupled hydrodynamic and biological model of phytoplankton dynamics and their impacts on water quality can be used (Robson and Hamilton 2004). For reviews on **process-based models** see (e.g. (Connolly *et al.,* 2017)). Inputs for these models can be provided from correlative approaches (Brophy *et al.,* 2017; Ovaskainen *et al.,* 2017), such as **correlative ecological models** (e.g. (Peterson, Papeş, and Soberón 2015)), or AI-based inferences such as described in Step 3.

2.4 BOX 1

Artificial Intelligence (AI) and eDNA metabarcoding on lake sediment: a case study

AI approaches that use explainable network models combined with multi-view learning (Li, Yang, and Zhang 2019) allow the simultaneous interrogation of different data matrices, to learn what components co-vary within a matrix (e.g. environmental variables), and among matrices (e.g. environmental variables and Molecular Operational Taxonomic Units or MOTUs) (Fig. I). These approaches are often used in a systems biology context in which e.g. protein-protein (within matrix) or protein-gene (between matrices) interaction are investigated. We use the AI approach on a pilot study to establish correlations between environmental factors and biodiversity, measured with eDNA metabarcoding. For this pilot study, eDNA data were obtained from the sedimentary archive of a watershed with a well-known history of human-driven environmental change (Lake Ring, Denmark; 55°57′51.83″N, 9°35′46.87″E) (Cambronero, Marshall, *et al.,* 2018). The history of Lake Ring can be separated into four main phases: semi-pristine (PR; <1950); eutrophication (E; 1960-1970); pesticides (P; 1980-1990); and partial recovery (R; >1999). The beta diversity (community diversity between each pair of sediment layers) from the dated sedimentary archive showed that whereas the prokaryotic community was shaped by the redox state of the sediment (Fig. IIA), the eukaryotic freshwater community composition changed with the major lake phases (Fig. IIB). Applying the explainable multi-view learning algorithm combined with matrix-on-matrix regression (Fig. I) we identified the top 10 herbicides with

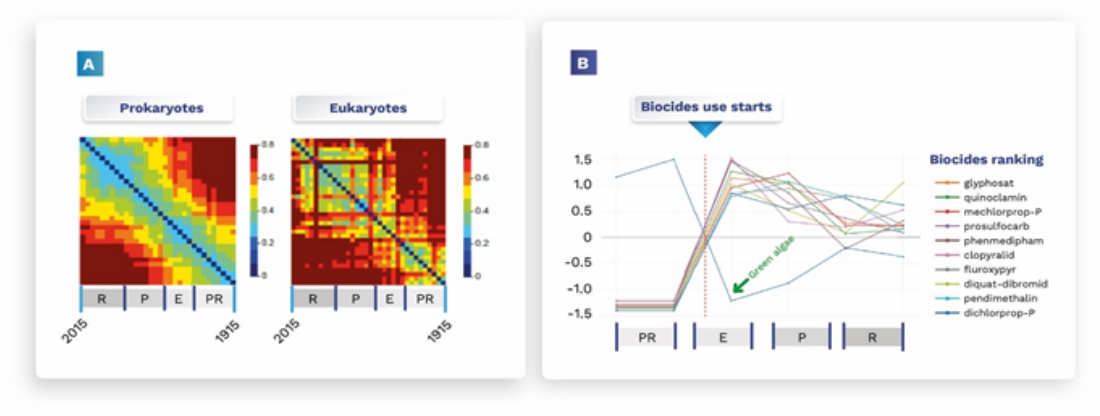
significant adverse effects on primary producers, specifically identifying green algae as target MOTUs (Fig. IIB). A top-down Pearson correlation analysis (vector-on-vector regression) validated this approach.



Box 1 Figure I. Illustration of multi-view learning combined with network analysis. Combining multi-view learning and network analysis, symmetric matrix-on-matrix regressions are obtained. The matrices represented in Figure I are Molecular Operational Taxonomic Units (MOTUs); Environmental Variables (e.g., temperature); and chemical pollutants (chemicals). Co-varying elements within a matrix as well as co-varying elements between matrices are identified. For example each node in the MOTUs network is a molecular taxonomic unit. Both co-variation in relative abundance of different MOTUs and their co-variation with environmental factors and chemical pollutants is identified using AI.

Box 1 Figure II. Biodiversity changes through time and correlations with chemical pollution. A) Eukaryote and prokaryote beta diversity through time (between each pair of layers) quantified with metabarcoding applied to eDNA extracted from lake sediment. Similarity in the composition of each sample (sediment layer) is measured through heat maps. Similarity decreases from blue to red. PR: pristine phase; E: eutrophication phase; P: pesticides phase; R: recovery phase; B) Inverse correlation between herbicides and primary

producers (green algae) established using the explainable multi-view learning algorithm combined with matrix-on-matrix regression shown in Fig. I.



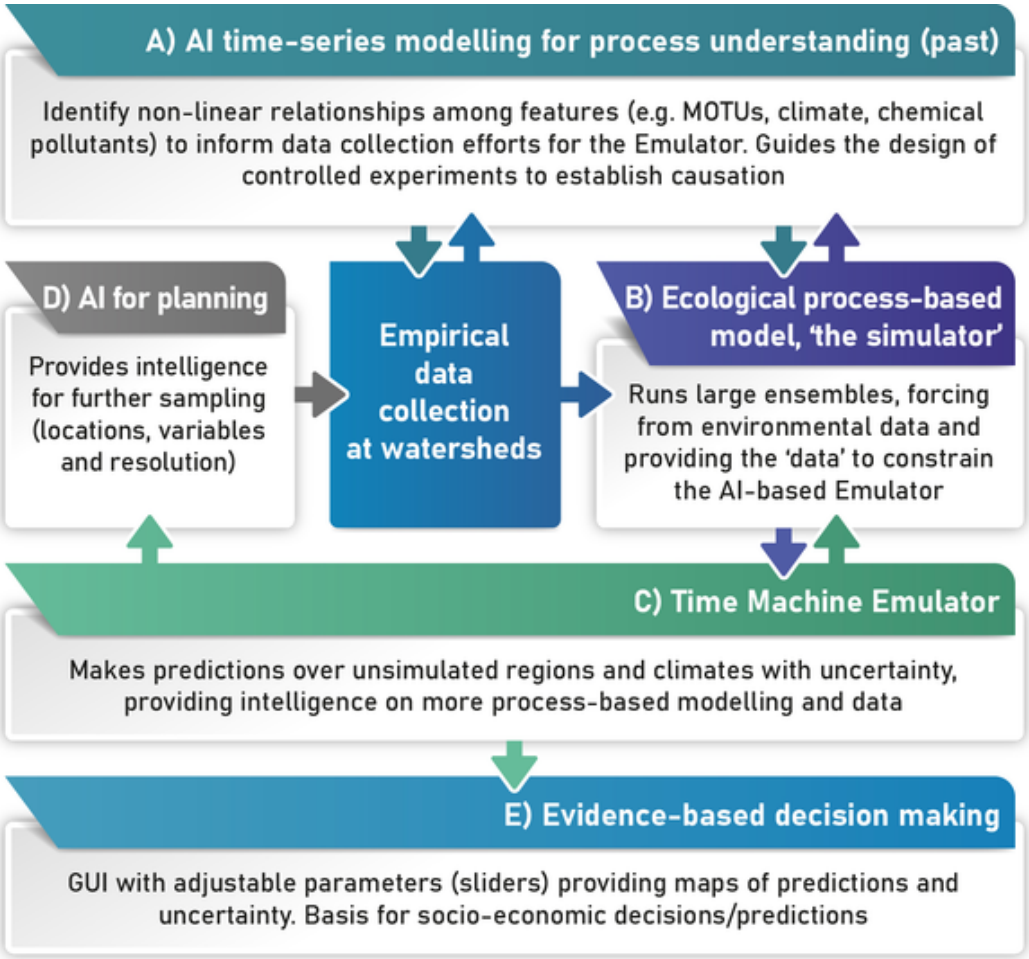


Figure 2. The Time Machine Emulator

Mapping out the various components of the data pipeline required for the Time Machine Emulator and the interoperability with data collection, process-based research, and the decision-making user interface. Arrows indicate the flow of data and information between the components. Empirical data (e.g. from watersheds) are analysed with an AI time series approach to establish non-linear relationships among multi-dimensional features (past) (A). The associations identified by the AI (A) and the process-based simulators (B) inform each other and are supported by empirical data. The Emulator (C) provides predictions and intelligence on regions that require further sampling to

reduce prediction uncertainties (D) and intelligence on more process-based modelling (B). A consensus model is obtained using a generalized additive framework that generates predictions through an AI-based emulator dashboard with a graphical user interface (GUI) (E).

The Time Machine framework is widely applicable because it finds signatures that recapitulate community dynamics (e.g. loss of MOTUs) driven by environmental change (e.g. warming, pollution) that alter ecosystem function (e.g. nutrients ratio). For example, it is applicable to physical, chemical and biological long-term observations of oceanic and coastal areas available from Continuous Plankton Recorders Survey, collecting records from the North Atlantic and the North Sea since 1931 (Survey 2020). However, the ecological models that best describe these ecosystems may differ from the ones described here.

2.5 THE TIME MACHINE EMULATOR

Continuous temporal data are paramount to explain present-day patterns and to model relationships between biodiversity and environmental change (Baert *et al.,* 2016). In particular, palaeoecological data are a valuable source of temporal data and can inform the conservation management of future biodiversity (e.g. (Fordham *et al.,* 2020)). However, past dynamics may differ from that caused by future environmental changes or threats, introducing uncertainties in model estimates. To reduce uncertainty in forecasts, realistic dynamic interactions

among several biotic and abiotic variables should be used (e.g. (Leeds, Wiklec, and Fiechter 2014)). Generating predictions that account for all these variables in different scenarios (e.g. projected IPCC climate scenarios) is computationally intensive and time consuming. **Emulators** can provide robust predictions with calculated uncertainties across multiple scenarios while reducing computational cost and time. An 'emulator' is a low-order, computationally efficient model which emulates the specified output of a more complex model in function of its inputs and parameters. Emulators work with both structure-based methods, where the mathematical structure of the original model is manipulated to a simpler, more computationally efficient form; and data-based approaches, where the emulator is identified and estimated from empirical/experimental data (Castelletti *et al.,* 2012). Emulators are widely applied in big data science, such as i) climate science to generate predictions under different socio-economic scenarios in long-range simulations (e.g. (Chantry *et al.,* 2021) and references therein); ii) ecology to predict the status of ecological processes in changing environments using e.g. long-term remote sensing data (Leeds, Wiklec, and Fiechter 2014); and iii) environmental science to predict e.g. the hydrological status of water reservoirs (Asher *et al.,* 2015).

We propose a *Time Machine Emulator* (TME) to generate forecasts of biodiversity and ecosystem functions. The TME 'learns' from past correlations, it is tested iteratively against long-term empirical data (e.g. collected in Step 3) and refined to predict the future biotic, abiotic and functional associations (Fig. 2). In our applications, it emulates ecological process-based models (e.g.

(Janse *et al.,* 2010; Mooij *et al.,* 2010)) to generate predictions across a range of historical and future climates and ecosystem states, albeit in a more computationally scalable and efficient manner. Empirical data collected from sedimentary archives, including biodiversity changes over time and pollution, as well as climate data, are used to establish past correlations that inform ecological process-based models (Fig. 2A). These models integrate historical contamination trends that disrupt ecosystem functions mediated by changes in MOTUs, to identify risk trajectories, with measured uncertainties (Fig. 2B). The TME is applied to produce the same output as the process-based models without the time consuming and computationally expensive running of the vast number of possible model setups for every possible climate and pollution scenario (Fig. 2C). One of the biggest challenges for AI on medium to long-range timescales is the lack of high-quality data for training model predictions. The TME uses empirical past biodiversity-ecosystem dynamics (collected in Step 3), as well as temperature projections from regional climate simulations, as 'training sets' to reduce uncertainty in forecasting the future of ecosystem states under different climate scenarios (Fig. 2C). Predictions are also made over non-simulated regions and climates with higher uncertainty, informing strategies for additional sampling, in this way mitigating uncertainties when new data are introduced in the model (Fig. 2D). This directs sampling efforts improve the accuracy of forecasts. The output of the TME are predictions for changes in ecosystem functions driven by localised and regional climate change and mediated by changes in overall biotic/abiotic interactions, including on partial training sets. The robustness of the TME predictions is assessed by removing

data (i.e. locations and conditions) from complete training sets and comparing predictions on whole and partial datasets. The future associations predicted by the TME are validated using fabricated ecosystems as described in Step 4. By identifying the MOTUs and ecosystem functions altered by abiotic factors and/or climate variables, the TME provides probabilistic decision-support information for ecosystem services management (Fig. 2E).

2.6 CONCLUDING REMARKS

Lack of understanding of the processes that underpin ecosystem services has often led to mismanagement with clear dis-benefits for the environment, the economy and human well-being. Systemic approaches, like the framework presented here, enable the prioritization of interventions that accelerate ecological restoration, and mitigate environmental factors that cause harm to MOTUs associated with key ecosystem functions and services. However, it requires critical changes in current environmental practice and a shift to whole-system evidence-based approaches. To overcome socio-technical barriers to adoption, stakeholder-enabled platforms are needed that include: 1) experimental protocols for the routine systemic biomonitoring of biodiversity; 2) a TME dashboard relying on Data Visualization Technology tested for ‘usability’ and ‘utility’; and 3) changes in legislation and regulatory practices to facilitate the adoption of novel tools and frameworks.

The use of the Time Machine framework comes with caveats. Process-based models cannot be easily validated in future climate regimes that have not been observed. To mitigate this, ecosystems over diverse climatic regions (e.g.,

warmer and drier environments) can be included as analogues of future climate states (see Outstanding Questions). In addition, known biotic interactions can be included in the model with information metrics, such as **Granger causality** (De Vico Fallani *et al.,* 2014). This approach infers biotic interactions using the time series empirical data collected from sedimentary archives and/or other known associations (e.g. predator/prey interactions), with estimated uncertainties. Uncertainties in the TME can be reduced by constraining the models with available past and projected climate data collected from local weather stations, globally gridded ERA5-Land datasets (Cao, Gruber, *et al.,* 2020), and Earth Observation data (https://earthdata.nasa.gov/earth-observation-data). Downscaling methods that use either physically-based regional climate models or statistical models to project future large-scale climate (Maraun *et al.,* 2017; Maraun and Widmann 2018), can be used to increase spatial granularity down to <10km and even to single point locations (e.g. (Gordon *et al.,* 2019; Maraun and Widmann 2018)). Although it is expected that uncertainty increases with increasing granularity, it may be reduced by using spatial scales for which the temporal variability is well simulated.

The TME applications can be, in principle, extended beyond predictions based on the ecological and functional status of ecosystems. By coupling ecological and economic modelling, the TME can also enable the alignment of socio-economic and ecological outcomes under different climate and pollution scenarios (see Outstanding Questions). To overcome adoption barriers by stakeholders, an AI-based Emulator dashboard can be developed, accessible to regulators and policy makers through data visualizations techniques. These

tools can be adapted for probabilistic predictions of ecosystem services to aid decision-making and socio-economic trade-offs.

2.7 GLOSSARY

eDNA or environmental DNA is DNA extracted from environmental matrices, such as soil, sediment, water, ice and aerosol without any obvious signs of biological source material.

Metabarcoding is high throughput sequencing of PCR-amplified taxonomic marker genes.

MOTU or Molecular Operational Taxonomic Units. Units of taxonomic diversity that, differently from Operational Taxonomic Units, do not necessarily correspond to species, but they are treated as such when measuring diversity. Clustering within MOTU is based on similarity of DNA sequences.

Suspect and non-target screening of chemical compounds are techniques to screen for chemical pollutants using mass spectrometry (MS). The suspect screening uses a library of catalogued data such as chromatographic retention times in lieu of reference standards. Non-target screening uses no pre-existing knowledge for comparison before analysis.

Mass Spectrometry (MS) is an analytical technique used to quantify concentrations of chemical compounds in different matrices by measuring the mass-to-charge ratio of the chemicals.

AI or Artificial Intelligence refers to the simulation of human intelligence in machines.

Machine learning or ML is a subset of artificial intelligence and focuses on the development of algorithms that can access data and use it to learn for themselves

Explainable network model (ENM) is a recent advancement in Machine Learning algorithms designed to identify which features in the data are responsible for driving a certain output, providing more trustable predictions. A typical application of ENM is for diagnosing breast cancer based on observed patterns on pathology slides.

Multi-view learning is an emerging direction in machine learning, also called deep learning, which learns from multiple networks (or views). Typical applications of multi-view learning include systems biology where functional links between e.g. gene networks and metabolite networks are established.

An **Emulator** is a hardware or software that permits programs written for one computer to be run on another computer. This enables to increase the efficiency of time-consuming simulations by parallelising resources. In climate science, emulators are used to evaluate the realism of the warming signal in different models on both global and regional scales, by comparing global trends and regional response parameters to observations.

Granger causality is a statistical test for determining whether one time series is useful in forecasting another.

Process-based models are models that characterize changes in a system's state as explicit functions of the events that drive those state changes (*sensu* (Connolly *et al.,* 2017))

Correlative ecological models are models based on environmental associations derived from analyses of geographic occurrences of species.

2.8 OUTSTANDING QUESTIONS

Is there a bias in the recovery of eDNA from fossil or remnant DNA in aged sediment and in warm climates?

While we found stable composition of primary producers across 100 years, a bias might still exist in sedimentary archives from warmer climates where mineralization of fossil remains is influenced by higher temperature and humidity. For these archives shorter time series may be used.

How to deal with high levels of uncertainty in the Emulator projections?

While the Emulator makes predictions over non-simulated regions and climates, guiding data collection efforts, it carries a level of uncertainty. This has to be accounted for in decision- making.

How does the Time Machine Emulator forecast the potential economic value of ecosystems?

Monetary and non-monetary estimates of values associated with ecosystem services can be generated with the current monetary valuation strategies and the predicted changes in ecosystem functions, uncertainty and risk provided by the Emulator. However, an outstanding challenge for decision-making is incorporating how values held by society for different services (e.g. carbon sequestration, food production, biodiversity, disease regulation) might change through time.

2.9 ACKNOWLEDGMENTS

We thank William Scavone, Kestrel Studio, USA, for the artwork in Figure 1 and Chantal Jackson for the artwork in Figure 2. This work was supported by The

Alan Turing Institute under the EPSRC grant EP/N510129/1, the Natural Environment Research Council (NE/N005716/1) and a MIBTP-BBSRC PhD fellowship to NE (BB/M01116X/1).

2.10 REFERENCES

Abdallah, M., H. Nguyen, J. Ebele, N. Atia, H. Ali, and S. Harrad. 2019. 'A single run, rapid polarity switching method for determination of 30 pharmaceuticals and personal care products in waste water using Q-Exactive Orbitrap high resolution accurate mass spectrometry', *Journal Chromatography A*, 1588: 68-76.

Asher, M. J. , B. F. W. Croke, A. J. Jakeman, and L. J. M. Peeters. 2015. 'A review of surrogate models and their application to groundwater modeling', *Water Resources Research* 51: 5957–73.

Backhaus, T., J. Snape, and J. Lazorchak. 2012. 'The impact of chemical pollution on biodiversity and ecosystem services: the need for an improved understanding', *Integr Environ Assess Manag*, 8: 575-6.

Baert, J. M., C. R. Janssen, K. Sabbe, and F. De Laender. 2016. 'Per capita interactions and stress tolerance drive stress-induced changes in biodiversity effects on ecosystem functions', *Nat Commun*, 7: 12486.

Balint, M., M. Pfenninger, H. P. Grossart, P. Taberlet, M. Vellend, M. A. Leibold, G. Englund, and D. Bowler. 2018. 'Environmental DNA Time Series in Ecology', *Trends Ecol Evol*, 33: 945-57.

Barnosky, A. D. , P. R. Ehrlich, and E. A. Hadly. 2016. 'Avoiding collapse: Grand challenges for science and society to solve by 2050', *Elementa: Science for the Anthropocene*, 4: 000094.

Bonebrake, T. C., F. Guo, C. Dingle, D. M. Baker, R. L. Kitching, and L. A. Ashton. 2019. 'Integrating Proximal and Horizon Threats to Biodiversity for Conservation', *Trends Ecol Evol*, 34: 781-88.

Brophy, C., A. Dooley, L. Kirwan, J. A. Finn, J. McDonnell, T. Bell, M. W. Cadotte, and J. Connolly. 2017. 'Biodiversity and ecosystem function: making sense of numerous species interactions in multi-species communities', *Ecology*, 98: 1771-78.

Cambronero, Cuenca M. , H. Marshall, L. De Meester, A. Davidson, A. P. Beckerman, and L. Orsini. 2018. 'Predictability of the impact of multiple stressors on the keystone species Daphnia ', *Scientific Reports*, 8: 17572.

Cao, B., S. Gruber, D. Zheng, and X. Li. 2020. 'The ERA5-Land soil temperature bias in permafrost regions', *The Cryosphere*, 14: 2581–95.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, D. A. Wardle, A. P. Kinzig, G. C. Daily, M. Loreau, J. B. Grace, A. Larigauderie, D. S. Srivastava, and S. Naeem. 2012. 'Biodiversity loss and its impact on humanity', *Nature*, 486: 59-67.

Castelletti, A. , S. Galelli, M. Ratto, R. Soncini-Sessa, and P.C. Young 2012. 'A general framework for Dynamic Emulation Modelling in environmental problems', *Environmental Modelling & Software*, 34 5e18.

Chantry, M., H. Christensen, P. Dueben, and T. Palmer. 2021. 'Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI', *Philos Trans A* 379: 20200083.

Connolly, S. R., S. A. Keith, R. K. Colwell, and C. Rahbek. 2017. 'Process, Mechanism, and Modeling in Macroecology', *Trends Ecol Evol*, 32: 835-44.

Cristescu, M. E., and P.D.N. Hebert. 2018. 'Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation', *Annual review of ecology, evolution, and systematics*, 49: 209–30.

De Vico Fallani, F., J. Richiardi, M. Chavez, and S. Achard. 2014. 'Graph analysis of functional brain networks: practical issues in translational neuroscience', *Philos Trans R Soc Lond B Biol Sci*, 369.

DeFries, R., and H. Nagendra. 2017. 'Ecosystem management as a wicked problem', *Science* 356: 265-70

Durance, I., M. W. Bruford, R. Chalmers, N. A. Chappell, M. Christie, B. J. Cosby, D. Noble, S. J. Ormerod, H. Prosser, A. J. Weightman, and G. Woodward. 2016. 'The challenges of linking ecosystem services to biodiversity. ', *Advances in Ecological Research*, 54: 87-134.

Figueiredo, L., J. Krauss, I. Steffan-Dewenter, and J. Sarmento Cabral. 2019. 'Understanding extinction debts: spatio–temporal scales, mechanisms and a roadmap for future research', *Ecography*, 42: 1973–90.

Fordham, D. A. , S.T. Jackson, S.C. Brown, B. Huntley, B. W. Brook, D. Dahl-Jensen, M. Thomas, P. Gilbert, B.L. Otto-Bliesner, A. Svensson, S. Theodoridis, J. M. Wilmshurst, J.C. Buettel, E. Canteri, M. McDowell, L. Orlando, J. Pilowsky, C. Rahbek, and D. Nogues-Bravo. 2020. 'Using paleo-archives to safeguard biodiversity under climate change', *Science*, 369: eabc5654.

Garlapati, D. , B. Charankumar, K. Ramu, P. Madeswaran, and M. V. Ramana Murthy. 2019. 'A review on the applications and recent advances in environmental DNA (eDNA) metagenomics', *Review in Environmental Science and Biotechnology*, 18: 389–411.

Gillson, L., and R. Marchant. 2014. 'From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology', *Trends Ecol Evol*, 29: 317-25.

Gordon, J., Bruinsma W. P., A. Y. K. Foong, J. Requeima, Dubois Y., and R. E. Turner. 2019. "Convolutional Conditional Neural Processes." In *International Conference on Learning Representations (ICLR)* New Orleans.

Grêt-Regamey, A., B. Weibel, S.-E. Rabe, and B. Burkhard. 2017. 'A tiered approach for ecosystem services mapping.' in B. Burkhard and J. Maes (eds.), *Mapping Ecosystem Services* (Pensoft Publishers).

Hirai, J., S. Katakura, H. Kasai, and S. Nagai. 2017. 'Cryptic Zooplankton Diversity Revealed by a Metagenetic Approach to Monitoring Metazoan Communities in the Coastal Waters of the Okhotsk Sea, Northeastern Hokkaido', *Frontiers in Marine Science*, 4.

Jackson, M. C., C. J. G. Loewen, R. D. Vinebrooke, and C. T. Chimimba. 2016. 'Net effects of multiple stressors in freshwater ecosystems: a meta-analysis', *Global change biology*, 22: 180-89.

Janse, J. H., M. Scheffer, L. Lijklema, L. Van Liere, J. S. Sloot, and W. M. Mooij. 2010. 'Estimating the critical phosphorus loading of shallow lakes with the ecosystem model PCLake: sensitivity, calibration and uncertainty', *Ecological Modelling*, 221: 654-65.

Janssen, A. B.G. , S. Teurlincx, A.H.W. Beusen, M.A.J. Huijbregts, J. Rost, A. M. Schipper, L.M.S. Seelen, W.M. Mooij, and J.H. Janse. 2019. 'PCLake+: A process-based ecological model to assess the trophic state of stratified and non-stratified freshwater lakes worldwide', *Ecological Modelling*, 396: 23-32.

Jax, K., Eeva Furman, H. Saarikoski, D. N. Barton, B. Delbaere, J. Dick, G. Duke, C. Görg, E. Gómez-Baggethun, P.A. Harrison, J. Maes, M. Pérez-Soba, Sanna-Riikka Saarela, F. Turkelboom, J. van Dijk, and A. D. Watt. 2018. 'Handling a messy world: Lessons learned when trying to make the ecosystem services concept operational', *Ecosystem Services*, 29: 415-27.

Leeds, W.B. , C.K. Wiklec, and J. Fiechter. 2014. 'Emulator-assisted reduced-rank ecological data assimilation for nonlinear multivariate dynamical spatio-temporal processes', *Statistical Methodology*, 17: 126-38.

Li, Y., M. Yang, and Z. Zhang. 2019. 'A Survey of Multi-View Representation Learning', *IEEE Transactions on Knowledge and Data Engineering*, 31: 1863–83.

Mace, G. M., K. Norris, and A. H. Fitter. 2012. 'Biodiversity and ecosystem services: a multilayered relationship', *Trends Ecol Evol*, 27: 19-26.

Maraun, D. , and M. Widmann. 2018. *Statistical downscaling and bias correction in climate research* (Cambridge University Press).

Maraun, D., T. G. Shepherd, M. Widmann, G. Zappa, D. Walton, J. M. Gutiérrez, S. Hagemann, I. Richter, P. M. M. Soares, A. Hall, and L. O. Mearns. 2017. 'Towards process-informed bias correction of climate change simulations', *Nature Climate Change*, 7: 764-73.

Mooij, W.M. , D. Trolle, E. Jeppesen, G. Arhonditsis, P. V. Belolipetsky, D.B.R. Chitamwebwa, Degermendzhy A. G., and *et al.,* 2010. 'Challenges and opportunities for integrating lake ecosystem modelling approaches', *Aquatic Ecology*, 44: 633-67.

Nogues-Bravo, D., F. Rodriguez-Sanchez, L. Orsini, E. de Boer, R. Jansson, H. Morlon, D. A. Fordham, and S. T. Jackson. 2018. 'Cracking the Code of Biodiversity Responses to Past Climate Change', *Trends Ecol Evol*, 33: 765-76.

Orsini, L., K. Schwenk, L. De Meester, J. K. Colbourne, M. E. Pfrender, and L. J. Weider. 2013. 'The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments', *Trends Ecol Evol*, 28: 274-82.

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B. E. Sæther, and N. Abrego. 2017. 'How are species interactions structured in species-rich communities? A new method for analysing time-series data', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170768.

Peterson, A. T., M. Papeş, and J. Soberón. 2015. 'Mechanistic and Correlative Models of Ecological Niches', *European Journal of Ecology*, 1: 28-38.

Pinto, R. , M. da Conceição Cunha, C. Roseta-Palma, and J. Carlos Marques. 2014. 'Mainstreaming Sustainable Decision-making for Ecosystems: Integrating Ecological and Socio-economic Targets within a Decision Support System', *Environmental Processes*, 1: 7–19.

Robson, B. J., and D. P. Hamilton. 2004. 'Three-dimensional modelling of a Microcystis bloom event in the Swan River estuary, Western Australia', *Ecological Modelling*, 174: 203-22.

Roux, D.J., J. L. Nel, G. Cundill, P. O'Farrell, and C. Fabricius. 2017. 'Transdisciplinary research for systemic change: who to learn with, what to learn about and how to learn', *Sustainable Science*, 12.

Ruckelshaus, M. H., S. T. Jackson, H. A. Mooney, K. L. Jacobs, K. S. Kassam, M. T. K. Arroyo, A. Baldi, A. M. Bartuska, J. Boyd, L. N. Joppa, A. Kovacs-Hostyanszki, J. P. Parsons, R. J. Scholes, J. F. Shogren, and Z. Ouyang. 2020.

'The IPBES Global Assessment: Pathways to Action', *Trends Ecol Evol*, 35: 407-14.

Ruppert, K. M. , R. J. Kline, and S. Md Rahman. 2019. 'Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA', *Global Ecology and Conservation*, 17: e00547.

Shoemaker, L.G. , L.L. Sullivan, I. Donohue, J.S. Cabral, R.J. Williams, M.M. Mayfield, J.M. Chase, C. Chu, W. S. Harpole, A. Huth, J. Hillerislambers, A.R. M. James, N.J. B. Kraft, F. May, R. Muthukrishnan, S. Satterlee, F. Taubert, X. Wang, T. Wiegand, W. Yang, and K.C. Abbott. 2020. 'Integrating the underlying structure of stochasticity into community ecology', *Ecology*, 101: e02922.

Spaak, J.W., J.M. Baert, D.J. Baird, N. Eisenhauer, L. Maltby, F. Pomati, V. Radchuk, J.R. Rohr, P.J. Van den Brink, and F. De Laender. 2017. 'Shifts of community composition and population density substantially affect ecosystem function despite invariant richness', *Ecology Letters*, 20: 1315-24.

Survey, CPR. 2020. "The continuous plankton recorder survey." In, edited by Marine Biological Association. Plymouth, UK: Marine Biological Association.

Tenenhaus, A., C. Philippe, V. Guillemot, K. A. Le Cao, J. Grill, and V. Frouin. 2014. 'Variable selection for generalized canonical correlation analysis', *Biostatistics*, 15: 569-83.

Tse, T. J., L. E. Doig, S. Tang, X. Zhang, W. Sun, S. B. Wiseman, C. X. Feng, H. Liu, J. P. Giesy, M. Hecker, and P. D. Jones. 2018. 'Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial Communities', *Environ Sci Technol*, 52: 6842-53.

Vellend, M., D. S. Srivastava, K.M. Anderson, C.D. Brown, J.E. Jankowski, E.J. Kleynhans, N. J. B. Kraft, A.D. Letaw, A.A.M. Macdonald, J.E. Maclean, I.H. Myers-Smith, A.R. Norris, and X. Xue. 2014. 'Assessing the relative importance of neutral stochasticity in ecological communities', *Oikos*, 123: 1420–30.

Wang, Y., S. Wen, M. D. Farnon Ellwood, A. D. Miller, and C. Chu. 2018. 'Temporal effects of disturbance on community composition in simulated stage-structured plant communities', *Ecol Evol*, 8: 120-27.

.

Chapter 3: 100 YEARS OF ANTHROPOGENIC IMPACT CAUSES CHANGES IN FRESHWATER FUNCTIONAL BIODIVERSITY

Niamh Eastwood*[1], Jiarui Zhou*[1], Romain Derelle[1], Mohamed Abou-Elwafa Abdallah[2], William A. Stubbings[2], Yunlu Jia[3], Sarah E. Crawford[3], Thomas A. Davidson[4], John K. Colbourne[1], Simon Creer[5], Holly Bik[6], Henner Hollert[3,7,8] and Luisa Orsini[1, 9]

*Shared first authorship

[1]Environmental Genomics Group, School of Biosciences, the University of Birmingham, Birmingham B15 2TT, UK

[2]School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[3] Department Evolutionary Ecology & Environmental Toxicology, Faculty of Biological Sciences, Goethe University Frankfurt, Max-von-Laue Straße 13, 60438 Frankfurt am Main, Germany

[4]Lake Group, Department of Ecoscience, Aarhus University,C.F. Møllers Allé, 8000 Aarhus, Denmark

[5] School of Natural Sciences, Environment Centre Wales, Deiniol Road, Bangor University, Gwynedd LL57 2UW, UK

[6] Department Marine Sciences and Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA

[7]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

[8]Department Media-related Toxicology, Institute for Molecular Biology and Applied Ecology (IME), Max-von-Laue Straße 13, 60438 Frankfurt am Main, Germany

[9]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK

Author contributions

NE produced and analysed the metabarcoding data. JZ created the code and ran the machine learning analyses. RD completed preliminary bioinformatics analyses. MA-EA and WS generated the DDT data. YJ, SEC and HH optimised chemical assays. TAD provided the sedimentary archive, the climate and the biocides sales data. HB provided the 96x96 unique barcode design. LO conceived and coordinated the study and data analysis. All co-authors contributed to paper writing and approved the final manuscript.

3.1 ABSTRACT

Despite efforts from scientists and regulators, biodiversity is declining at an alarming rate. Unless we find transformative solutions to preserve biodiversity, future generations may not be able to enjoy nature's services.

We have developed a conceptual framework that establishes the links between biodiversity dynamics and abiotic change through time and space using artificial intelligence. Here, we apply this framework to a freshwater ecosystem with a known history of human impact and study 100 years of community-level biodiversity, climate change and chemical pollution trends. We apply explainable network models with multimodal learning to community-level functional biodiversity measured with multilocus metabarcoding, to establish correlations with biocides and climate change records. We observed that the freshwater community assemblage and functionality changed over time without returning to its original state, even if the lake partially recovered in recent times. Insecticides and fungicides, combined with extreme temperature events and precipitation, explained up to 90% of the functional biodiversity changes. The community-level biodiversity approach used here reliably explained freshwater ecosystem shifts. These shifts were not observed when using traditional quality indices (e.g. Trophic Diatom Index).

Our study advocates the use of high throughput systemic approaches on long-term trends over species-focused ecological surveys to identify the environmental factors that cause loss of biodiversity and disrupt ecosystem functions.

3.2 INTRODUCTION

Biodiversity is the foundation of provisioning, regulating, supporting, and cultural ecosystem services (Baert *et al.,* 2016), which underpin economic prosperity, social well-being and quality of life (Cardinale *et al.,* 2012). Global biodiversity has been lost at an alarming rate in the past century, leading to what some have called the sixth mass extinction - biodiversity loss caused by human population growth and activities (Naggs 2017). Biodiversity is threatened by agricultural land use, climate change, invasive species, pollution and unsustainable production and consumption (Bonebrake *et al.,* 2019). Freshwater ecosystems have suffered the greatest biodiversity loss because of these anthropogenic drivers (Ruckelshaus *et al.,* 2020). Experimental manipulation of biodiversity has demonstrated the causal links between biodiversity loss and loss of ecosystem functions (Eisenhauer *et al.,* 2019). However, studies on multi trophic levels are scarce and largely focus on terrestrial and marine ecosystems; freshwater ecosystems, especially lakes and ponds, are not well represented in multitrophic experimental manipulations, (Dornelas *et al.,* 2018). These holistic studies are critical to understand the context-dependency of biodiversity-ecosystem functions relationships and to implement management measures to conserve biodiversity. However, a better understanding of the environmental factors with the largest impact on biodiversity, and their cumulative effect over time is urgently needed (Halpern *et al.,* 2015).

Biodiversity action plans have been devised since the 1990s. However, most strategies have failed to stop or even reduce biodiversity decline (Rounsevell *et al.,* 2020). This is because:

i) Biodiversity loss occurs at different spatial and temporal scales, and dynamic changes in community composition are the result of long-term ecological processes (Eastwood *et al.,* 2022; Nogues-Bravo *et al.,* 2018). State-of-the-art environmental and biological monitoring typically captures single snapshots in time of long-term ecological dynamics, failing to identify biodiversity shifts that may arise from cumulative impacts over time (Eastwood *et al.,* 2022; Nogues-Bravo *et al.,* 2018). Recent initiatives like BioTIME started collating databases with species presence and abundance recorded from time series across different ecosystems (Dornelas *et al.,* 2018). However, freshwater ecosystems are poorly represented in these studies which at most encompass the last 10-25 years (Blowes *et al.,* 2019). Although the large geographic breath of these studies is good to understand overall trends of biodiversity change, they are inadequate to identify drivers of biodiversity dynamics (Blowes *et al.,* 2019; Halpern *et al.,* 2015). Moreover, the taxonomic species assignment in these databases is oftentimes derived from traditional observational methods (e.g. microscopy), which cannot resolve cryptic diversity (Blowes *et al.,* 2019). High cryptic diversity is common in freshwater invertebrates and primary producers, potentially impacting the assessment of biodiversity in these ecosystems more severely than in terrestrial or marine ecosystems (Hirai J 2017). More recently, *seda*DNA (environmental DNA extracted from sediment) has emerged as a promising tool to study decade-long biological dynamics (Domaizon *et al.,*

2017). However, these studies focus on specific taxonomic groups e.g. microbes (Capo *et al.,* 2019); ciliates (Barouillet *et al.,* 2022), failing to capture the community-level changes in any given ecosystem.

ii) Biodiversity is threatened by multiple factors. Only by quantifying trajectories of abiotic, biotic, and functional systemic change over time, can we begin to identify the causes of biodiversity and ecosystem function loss (Bonebrake *et al.,* 2010). Studies are emerging that investigate the impact of chemicals (Groh *et al.,* 2022) or climate change (Pecl *et al.,* 2017) on biodiversity. Yet, understanding the combined effect of these abiotic factors on biodiversity is still challenging.

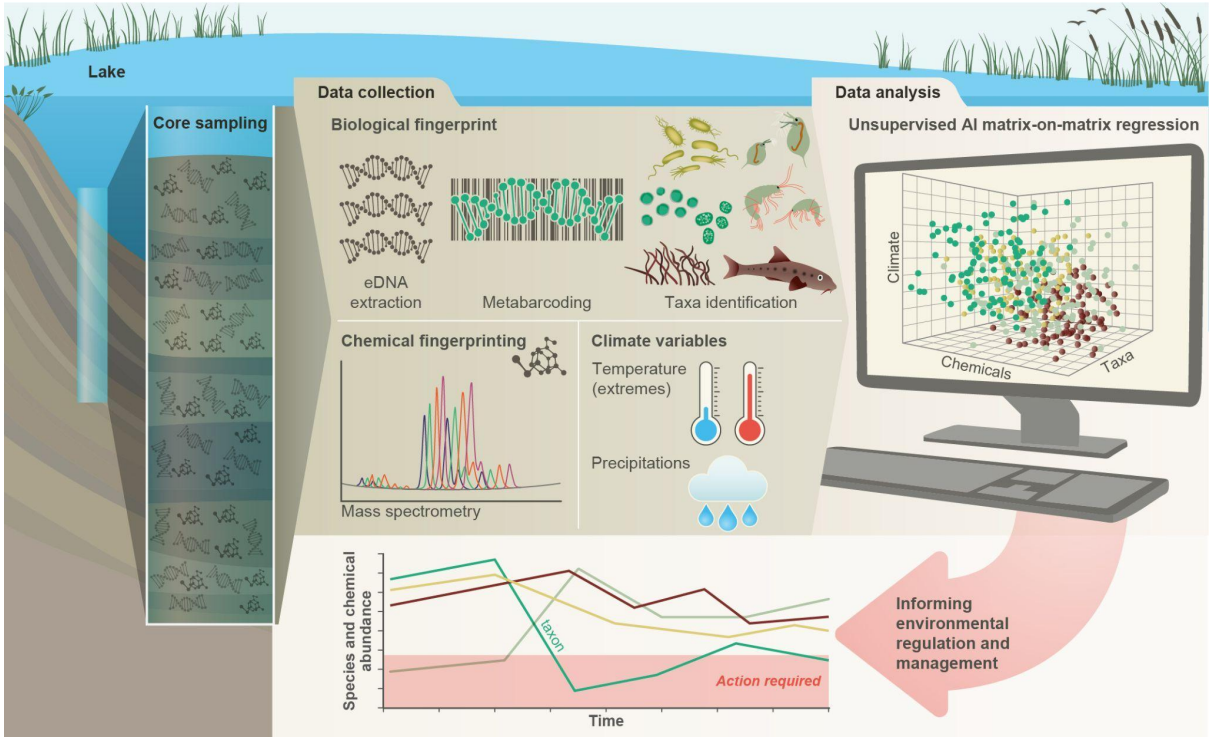
iii) The lack of paired biological and abiotic long-term monitoring data is a limiting factor in establishing meaningful and achievable conservation goals. Even well-monitored species have time series spanning a few decades at best (Bonebrake *et al.,* 2010; Halpern *et al.,* 2015). Moreover, conservation efforts have historically focused on ecological surveys of few indicator species, the identification of which require specialist skills (e.g., light microscopy and taxonomy) and are low throughput (Gillson and Marchant 2014). High throughput system-level approaches providing biological, abiotic and functional changes over multiple decades are needed to understand links between biodiversity loss, drivers of changes and potential consequences on ecosystem functionality [10].

Recently, we have developed a conceptual framework that helps establish the links between biodiversity dynamics and abiotic environmental changes using artificial intelligence, examines emergent impacts on ecosystem functions, and

forecasts the likely future of ecosystem services and their socioeconomic impact under different pollution and climate scenarios [10]. Here, we illustrate the first component of this framework in a freshwater ecosystem (Lake Ring, Denmark) with a well-documented human-impact over 100 years (Cuenca - Cambronero *et al.,* 2018) by quantifying the interrelations between community-level functional biodiversity, biocides and climate (Fig. 1). Historical records, supported by empirical evidence show that Lake Ring experienced semi-pristine conditions until the early 1940s (Cambronero, Beasley, *et al.,* 2018). In the late 1950s, sewage inflow caused severe eutrophication. When the sewage inflow was diverted at the end of the 1970s, agricultural land use intensified, leading to substantial biocides leaching (Cuenca - Cambronero *et al.,* 2018). The lake partially recovered from eutrophication and land use in modern times (>1999) but, as with every lake ecosystem in Europe, it experienced an increase in average temperature (Cuenca Cambronero *et al.,* 2018; Cuenca Cambronero, Zeis, and Orsini 2018). We apply multilocus metabarcoding and mass spectrometry analysis to a dated sedimentary archive of Lake Ring. These data, complemented by biocides sale records and climate records, were studied with explainable network models with multimodal learning to identify drivers of functional biodiversity changes across major ecosystem shifts (Baltrusaitis, Ahuja, and Morency 2019) (Fig. 1). The combination of explainable networks and multimodal learning allow the simultaneous interrogation of data matrices describing different types of data. A symmetric matrix-on-matrix regression is typically used to identify the components that covary within a matrix (e.g., environmental variables), and among matrices (e.g., environmental variables

and eDNA taxonomic units). Given the well-documented human-impact over time, Lake Ring represents an excellent natural system to demonstrate the power of systemic approaches in biological and functional monitoring.

Figure 1. Conceptual framework. A sedimentary archive spanning 100 years was sampled from Lake Ring, Denmark and dated using radioisotopes. Both biotic and abiotic changes were empirically quantified through time: 1) community-level biodiversity was reconstructed by applying multilocus metabarcoding to environmental DNA isolated from sediment layers (biological fingerprinting); 2) chemical signatures were quantified from the same sediment layers using mass spectrometry analysis (chemical fingerprinting); 3) climate data were collected from publicly available databases. Explainable network models with multimodal learning were applied to identify significant correlations between system-level biodiversity, chemical fingerprinting, and climate variables. Taxonomic units (families) impacted by environmental factors were identified and environmental factors ranked based on their effects on community biodiversity. This approach enables the prioritisation of conservation and mitigation interventions.


Lake
Core sampling
Data collection
Biological fingerprint
eDNA extraction
Metabarcoding
Taxa identification
Chemical fingerprinting
Mass spectrometry
Climate variables
Temperature (extremes)
Precipitations
Data analysis
Unsupervised AI matrix-on-matrix regression
Climate
Chemicals
Taxa
Species and chemical abundance
taxon
Action required
Time
Informing environmental regulation and management

3.3 RESULTS

3.3.1 Freshwater community dynamics across 100 years

A sedimentary archive was collected from Lake Ring in November 2016 with a gravity corer; the core was sliced in 34 layers of 0.5 cm, which corresponded to a temporal resolution of about 3 years per layer across 100 years. This estimate was based on a radiometric chronology of the core completed in 2018 (see Methods). Lake Ring has a well-known and documented history of human impact over the past century. The lake transitioned over time from a semi-pristine environment to eutrophication, and later to high pesticide pollution due to intensification of agricultural land-use in the area surrounding the lake. In modern times (>1999), the lake partially recovered (see methods for more details)(Cuenca - Cambronero *et al.,* 2018). Hereafter, we refer to the lake transitions across these statuses as lake phases.

We quantified community-level biodiversity over a century (1916 - 2016) by applying high throughput multilocus metabarcoding (18S, 16SV1, 16SV4, COI and rbcL barcodes) to bulk environmental DNA (eDNA) extracted from layers of a dated sedimentary archive from Lake Ring. After denoising, the number of unique ASVs and total number of reads across all samples (including median number of reads per sample) found per barcode were as follows: 18S - 2,023 ASVs, 569,761 total reads (median 12,893 reads); 16SV1 - 4,022 ASVs, 842,619 total reads (median 20,798 reads); 16SV4 - 5,270 ASVs, 552,064 total reads (median 13,816 reads); COI - 822 ASVs, 362,616 total reads (median 9,595 reads); rbcL - 417 ASVs, 366,489 total reads (median 9,443 reads). Alpha diversity did not significantly vary across the lake phases for both

prokaryotes and eukaryotes (Supplementary Fig. 1) and was proportionally higher in the prokaryotic (16S barcodes) than in the eukaryotic community (18S barcode). Conversely, the invertebrate community (COI barcode), and the diatom community (rbcL barcode), showed significant changes over time across the lake phases, reflecting taxon-specific patterns over time (Supplementary Fig. 1). Even though the alpha diversity varied over time, it was not consistently lower in historical than modern communities across the barcodes, allowing us to exclude bias in the preservation state of environmental DNA.

The community composition (beta diversity) changed significantly in the transition between lake phases (Table 1; Fig. 2A; Supplementary Fig. 2). The overall eukaryotic community composition changed over time across all lake phases (Table 1; Fig. 2A; 18S). However, the composition of the primary producers (e.g. rbcL) changed significantly only in the transition between the pesticide and the eutrophic phases, whereas the invertebrate's community (e.g. COI) changed significantly only between the pesticide and the recovery phases (Table 1; Fig. 2A; rbcL, COI). The significant changes in community composition identified by the PERMANOVA analysis were driven by two families of primary producers [*Chlorophyceae* (green algae), *Mediophyceae* (diatoms)] and seven families of invertebrates, [Monhysterida (nematode worms), Oligohymenophorea (ciliates), Calanoida (zooplankton), Ploimida (rotifers), Chaetonotida (gastrotrichs), *Thoracosphaeraceae* (dinoflagellates) and Calanoida (copepods)] (Fig. 2B; 18S). In the transition from the semi-pristine to the eutrophic phase, the relative abundance of rotifers and green algae declined in favour of calanoids and diatoms (Fig. 2B; 18S). The proportion of

diatoms, worms and nematodes increased in the transition from the eutrophic to the pesticide phase, while the proportion of calanoids and gastrotricha declined (Fig. 2B; 18S). The taxonomic composition of the recovery phase showed a relative increase in ciliates and gastrotricha as compared to the pesticide environment (Fig. 2B; 18S). *Vampirellidae* (Vampire amoebae feeding on algae) were relatively more abundant in the eutrophic than in the other phases, in which primary producers were also more abundant (Fig. 2B, 18S). The composition of the recovery and semi-pristine phases differed significantly, suggesting an incomplete recovery of the lake over time to this date (Table 1; Fig. 2A;18S).

The prokaryotic community significantly changed at each major transition between lake phases, consistently across the two barcodes (Table 1; 16SV1 and 16SV4). We observed two patterns in the prokaryotic community composition over time: some taxonomic groups changed with the redox status of the sediment [e.g. acidophilus archaea (Thermoplasmata) and methanogenic archaea (*Methanomassiliicoccaceae*), which declined from the semi-pristine to the recovery phase (Fig. 2B, 16SV4)]; others changed over time consistently with the nutrient levels of the ecosystem. For example *Nitrospiraceae* (nitrite oxidizers) were more abundant in high nutrient environments (eutrophic and pesticides) than in lower nutrient environments (semi-pristine and recovery) (Fig. 2B; 16SV1)].

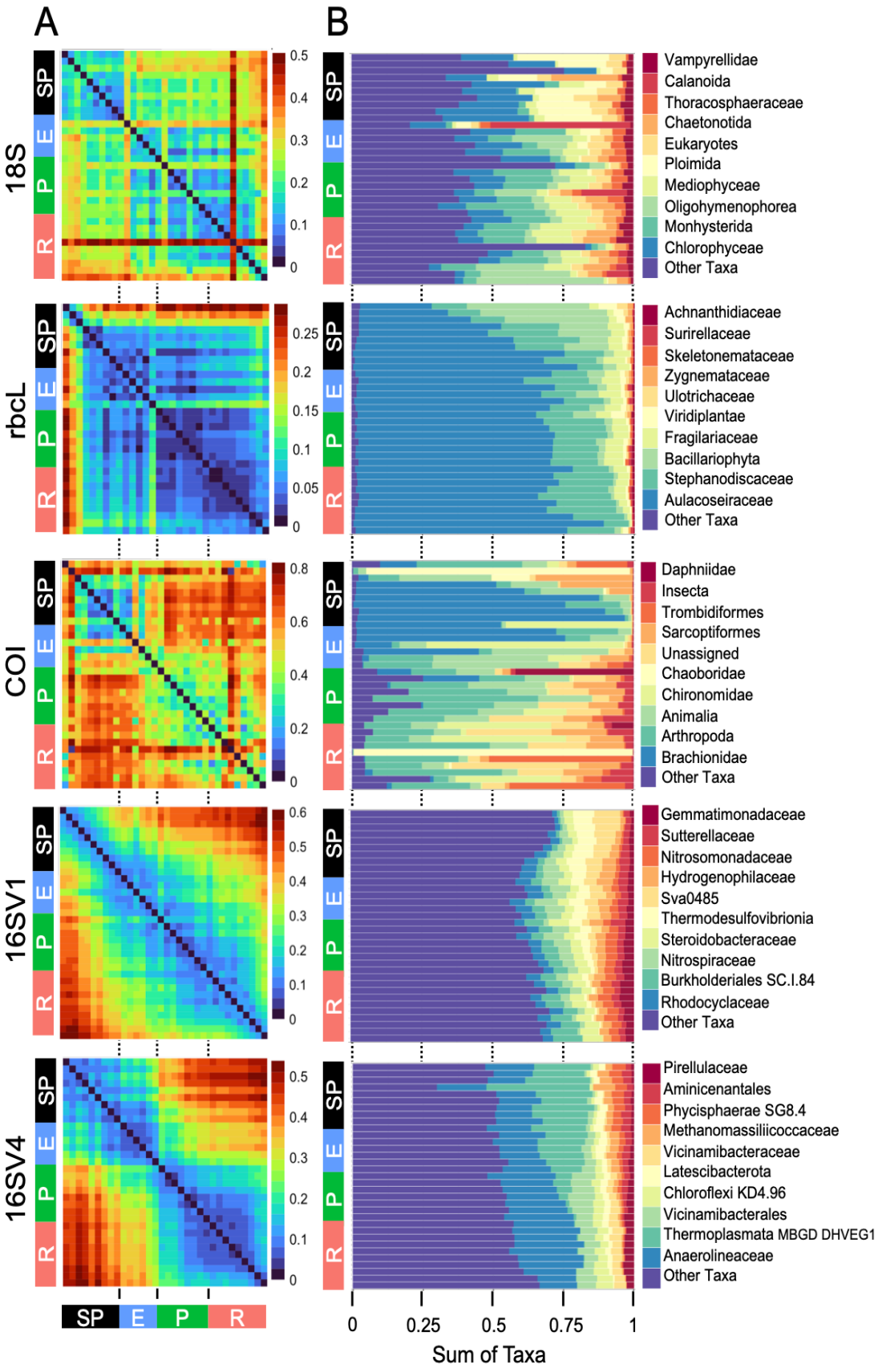
Changes in the invertebrate community were driven by *Brachionideae* (rotifers) that were most abundant in the semi-pristine phase and declined over time; *Chironomidae* (lake flies) that were proportionally more abundant in the

eutrophic and recovery phases and showed the lowest abundance in the pesticides phase; *Chaoboridae* (phantom midge larvae) that were only present in the semi-pristine and recovery phases; and *Daphniidae* (waterfleas) that were most abundant in the pesticide phase, but present throughout the 100 years of sampling (Fig. 2B; COI). The diatom composition was stable over time, with only the semi-pristine phase having a more distinctive diatom assemblage profile dominated by *Bacillariophyta* (Fig. 2B; rbcL). Diatoms are commonly used by regulators to derive the status of freshwater within the Water Framework Directive both for lakes and rivers (Agency 2020). We used our rbcL data to derive a Lake Trophic Diatom Index (LTDI2) for Lake Ring following (Bennion *et al.,* 2014). This result confirmed our beta diversity analysis of non-significant changes over time of the diatom community (Supplementary Fig. 3).

Table 1. PERMANOVA on beta diversity. Permutational Multivariate Analysis of Variance using weighted Unifrac distances ASV matrices testing for pairwise differences between lake phases across the five barcodes used in the study (16SV1, 16SV4, 18S, COI, rbcL) with 999 permutations. Significant terms (p-values <0.05 after applying Benjamini & Hochberg correction for multiple testing) are in bold. The lake phases are as follows: SP - semi-pristine; E - Eutrophic; P - pesticides; R - recovery.

Phase		16SV1		16Sv4		18S		COI		rbcL	
1	2	R2	p adj	R2	p adj	R2	p adj	R2	p adj	R2	p adj
SP	E	0.4349	**0.0067**	0.5533	**0.0017**	0.2968	**0.0033**	0.0432	0.705	0.2879	0.0914
SP	P	0.6290	**0.0025**	0.8515	**0.0017**	0.4459	**0.0033**	0.3868	**0.0033**	0.3920	**0.0125**
SP	R	0.6956	**0.0025**	0.9026	**0.0017**	0.3841	**0.0033**	0.3178	**0.0033**	0.5084	**0.0033**
E	P	0.3959	**0.006**	0.7399	**0.0017**	0.1249	0.15	0.3198	**0.005**	0.1555	0.1511
E	R	0.5656	**0.0025**	0.8520	**0.0017**	0.1816	**0.0075**	0.2806	**0.0033**	0.6019	**0.0033**
P	R	0.3026	**0.0025**	0.3724	**0.0017**	0.1029	0.15	0.1924	**0.012**	0.3605	**0.0033**

Figure 2. Biodiversity compositional changes. (A) Weighted unifrac beta diversity heatmaps between each pair of sediment layers spanning a century (1916-2016) for the five barcodes used in this study (18S, rbcL, COI, 16SV1 and 16SV4). The PERMANOVA statistics in Table 1 support these plots. The scale used may be different among the heatmaps. (B) Taxonomic bar plots including the top 10 most abundant families identified across five barcodes (18S, rbcL, COI, 16SV1 and 16SV4). shown per lake phase: SP - semi-pristine; E - eutrophic; P - pesticides; R - recovery.

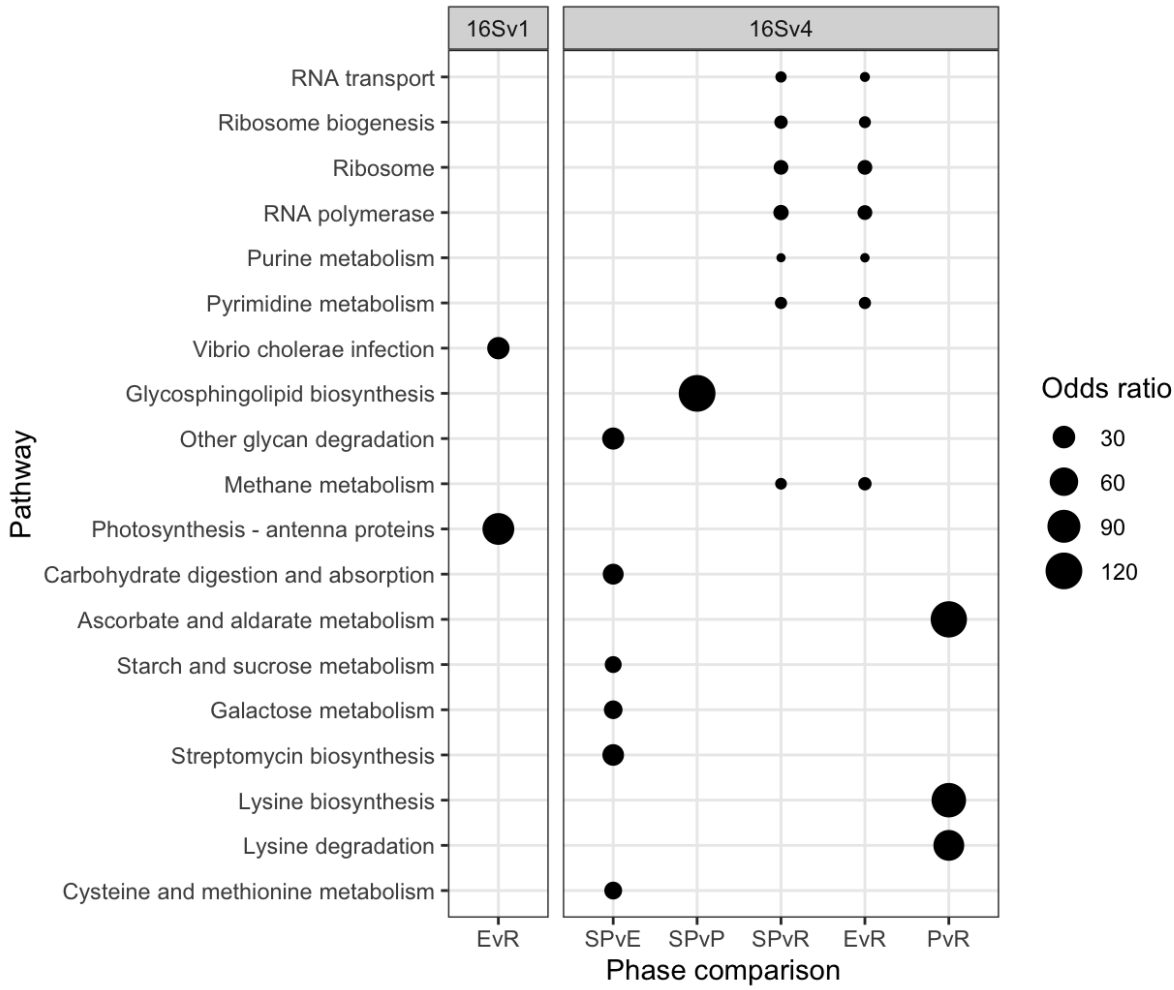

A
B
18S
rbcL
COI
16SV1
16SV4
SP
E
P
R
0.5
0.4
0.3
0.2
0.1
0
0.25
0.15
0.05
0.8
0.6
0.75
1
Sum of Taxa
Vampyrellidae
Calanoida
Thoracosphaeraceae
Chaetonotida
Eukaryotes
Ploimida
Mediophyceae
Oligohymenophorea
Monhysterida
Chlorophyceae
Other Taxa
Achnanthidiaceae
Surirellaceae
Skeletonemataceae
Zygnemataceae
Ulotrichaceae
Viridiplantae
Fragilariaceae
Bacillariophyta
Stephanodiscaceae
Aulacoseiraceae
Other Taxa
Daphniidae
Insecta
Trombidiformes
Sarcoptiformes
Unassigned
Chaoboridae
Chironomidae
Animalia
Arthropoda
Brachionidae
Other Taxa
Gemmatimonadaceae
Sutterellaceae
Nitrosomonadaceae
Hydrogenophilaceae
Sva0485
Thermodesulfovibrionia
Steroidobacteraceae
Nitrospiraceae
Burkholderiales SC.I.84
Rhodocyclaceae
Other Taxa
Pirellulaceae
Aminicenantales
Phycisphaerae SG8.4
Methanomassiliicoccaceae
Vicinamibacteraceae
Latescibacterota
Chloroflexi KD4.96
Vicinamibacterales
Thermoplasmata MBGD DHVEG1
Anaerolineaceae
Other Taxa

3.3.2 Functional changes linked to community compositional shifts

Changes in freshwater community composition corresponded to significant shifts in the predicted functioning of the prokaryotic community (Fig 3). We predicted different functions between each pair of lake phases by identifying molecular functions enriched as quantified by functional orthologs (KO terms). A functional ortholog was defined from functions experimentally assigned to the prokaryotes identified with the 16S rRNA in each lake phase. We predicted a total of 6,257 Kegg Orthologs (KO) terms from the 16SV1 and 6,828 from the 16Sv4 barcode across the lake phases. Of the total number of KO terms, 1,418 were significantly differentially abundant across the lake phases in the 16SV1 and 1,064 terms in the 16SV4 dataset, respectively. The functional KEGG pathways enriched within these KO terms and significantly differentially enriched between lake phases (Fisher's exact test, p-adj < 0.05) were 19 (17 for the 16SV4 and 2 for the 16SV1) (Fig. 3). Seven differentially enriched pathways were found between the semi-pristine and recovery phases and seven were found between the eutrophic and recovery phases (Fig. 3; 16SV4). These pathways were linked to catabolic functions (purine and pyrimidine metabolism), RNA transport and biogenesis, fundamental for gene expression and protein folding. Six functional pathways were differentially enriched between the semi-pristine and the eutrophic phases that were linked to metabolism (including methane metabolism), degradation and biosynthesis (Fig. 3; 16SV4). Three functional pathways that underpin carbohydrates metabolism, lysine biosynthesis and degradation were differentially enriched between the pesticide

and recovery phases. The latter two functions are critical for mitochondrial function. A single pathway was differentially enriched between the semi-pristine and the pesticide phases, linked to lipid metabolism (glycosphingolipid biosynthesis; Fig. 3; 16SV4). Two differentially enriched pathways were identified between the eutrophic and the recovery phases and underpin infection response and photosynthesis (Fig. 3; 16SV1).

Figure 3. Functional analysis. Functional pathways that are significantly differentially enriched between lake phases are shown for the 16SV1 and the 16SV4 barcodes. The lake phases are as in Figure 2: SP - semi-pristine; E - eutrophic; P - pesticides; R - recovery. Odds ratios indicate the representation of each pathway in the pairwise comparisons.



3.3.3 Drivers of biodiversity change

To discover drivers of biodiversity change we applied sparse canonical correlation analysis (sCCA) to community biodiversity data and other parameters measured from Lake Ring, namely climate records collected from a weather station proximal to the lake, and sales records of biocides in Denmark between 1955 and 2015 from the Danish national archives. The biocide sales records proved to be a good representation of persistent chemicals in the lake sediment, as the quantification of the persistent halogenated pesticide DDT in the sliced sedimentary archive showed by producing a very similar profile as the sales records over time (see methods section).

We discovered that insecticides and fungicides best explained changes in overall biodiversity, possessing the highest CCA loadings across the barcodes, followed by pesticides and herbicides (Supplementary Table 1A). Among the climate variables, yearly minimum temperature explained the largest biodiversity changes, whereas other climate variables had a variable impact across the barcodes and hence taxonomic groups (Supplementary Table 1B).

Having ranked biocides and climate variables that best explained changes in overall biodiversity, we identified correlations between taxonomic groups (assigned at family level where possible) and individual abiotic variables. Correlations were identified between a total of 36 eukaryotic families and abiotic variables; of these correlations, 28 were with biocides and 25 with climate variables (some correlations involved the same taxonomic group correalting with multiple environmental factors). Of the 28 families negatively correlated with biocides, the largest proportion co-varied significantly with insecticides (21

families - 75%) and fungicides (14 families - 50%), followed by herbicides (7 families - 25%) and pesticides (2 families - 7.1%) (Supplementary Table 2). Of the 25 families correlated with climate variables, the largest proportion co-varied with summer precipitation (12 families - 37%); of these, 8 families were positively correlated and 4 were negatively correlated with summer precipitation. An equal number of families (8 families - 32%) co-varied with mean minimum temperature (6 positive and 2 negative correlations), highest recorded temperature (7 positive and 1 negative correlations), and summer atmospheric pressure (6 positive and 2 negative correlations) (Supplementary Table 2).

The number of unique prokaryote families significantly negatively correlated with biocides was 99, 19 of which were identified by both 16S barcodes. Following from the sCCA analysis, significant negative correlations were observed between 60 (60.6%) families and insecticides, followed by 59 families and fungicides (59.6%), 40 families and herbicides (40.4%), and 25 families and pesticides (25.3%) (Supplementary Table 2; overall). A total of 105 non-redundant correlations were identified between prokaryotic families and climate variables, 6 of which were found in both 16S barcodes. Of the total families correlating with climate variables, 69 (65.7%) significantly correlated with mean minimum temperature. Of these, 38 were positive and 31 were negative correlations. Thirty-five families (33.3%) significantly correlated with summer precipitation; of these, 11 were positively and 23 were negatively correlated. Twenty-nine families (27.6%) significantly correlated with the lowest recorded temperature; of these 20 were positive and 9 were negative correlations.

Twenty-six families (24.8%) significantly correlated with mean summer temperature; of these 13 were positively and 13 negatively correlated. Twenty-three families (21.9%) significantly correlated with maximum daily precipitation; of these, 3 were positively and 20 were negatively correlated. Eleven families (10.4%) significantly correlated with highest recorded temperature; of these 3 were positively and 8 were negatively correlated (Supplementary Table 2).

We applied sCCA to identify families that correlated both with climate variables and biocides (Fig. 4). As biocides were introduced only in 1960, only the most recent three lake phases were included in this analysis. The eukaryotic biodiversity compositional change was predominantly explained by biocides (Fig. 4; 18S; Biocides: 44%), followed by climate variables (Fig. 4; 18S; climate variables: 22%). Up to 22% of the diatoms compositional change was explained by biocides (44%) and climate variables (36%). However, the abiotic variables only separated the recovery from the other two lake phases (Fig. 4), supporting significant biodiversity compositional shifts observed in the beta diversity analysis (Fig. 2A; Table 1). Similarly, the invertebrate community compositional changes were explained prevalently by biocides (47%), followed by climate variables (30%), which only separated the recovery phase from the other two lake phases. Climate and biocides almost equally explained up to 36% of the prokaryote biodiversity compositional change across the lake phases (16SV1 - biocides: 44%, climate variables 47%; 16SV4 - biocides 45%, climate variable 38%). Following from this analysis, joint effects of biocides and climate variables were observed for 23 prokaryote (16S) and two eukaryote (18S) families (Fig. 5A), whereas no joint effects were identified on the diatom (rbcL)

and the invertebrate (COI) communities (Fig 5A; Supplementary Table 3). The most frequent joint effects on prokaryotes involved insecticides and mean minimum temperature (Fig. 5A; Supplementary Table 3). Joint effects between herbicides and maximum daily precipitation or between herbicides and lowest recorded temperature were rare (Fig. 5A; Supplementary Table 3). The joint effects on the eukaryotic community were observed between insecticides and summer precipitation (Fig. 5A; Supplementary Table 3).

The biocide types showing joint effects with environmental variables were ranked based on their correlation coefficient over time (Supplementary Table 3). The top ranked insecticides most frequently showing these joint effects with climate variables and an adverse effect on both prokaryotes and eukaryotes were: oxydemeton-methyl (organothiophosphate insecticide, primarily used to control aphids, mites, and thrips), mevinphos (organophosphate insecticide used to control insects in a wide range of crops) and dicofol (organochlorine miticide pesticide chemically related to DDT). Additionally, parathion (organophosphate insecticide and acaricide), carbaryl (1-naphthyl methylcarbamate used chiefly as an insecticide), dieldrin (organochlorine insecticide, developed in alternative to DDT) and thiometon (organic thiophosphate insecticide) showed adverse effects with only the prokaryotic community. Examples of joint effects on specific families are shown in Figure 5B and 5C. The temporal dynamics of Isochrysidales, a coccolith-producing microalgae, was affected by the joint effect of summer precipitation and insecticides (Fig. 5B), whereas the temporal dynamics of the PeM15 group of

Actinobacteria was affected by the joint effect of insecticides and mean minimum temperature (Fig. 5C).

Figure 4. sCCA 3D plots. Sparse canonical correlation analysis 3D plots for the five barcodes used (18S, rbcL, COI, 16SV1 and 16SV4), showing the proportion of biodiversity variance explained by the biocides and climate variables. As biocides were introduced around the 1960s, this analysis spans the most recent three lake phases (Eutrophic, Pesticide and Recovery). Interactive version available: https://environmental-omics-group.github.io/Biodiversity_Monitoring/

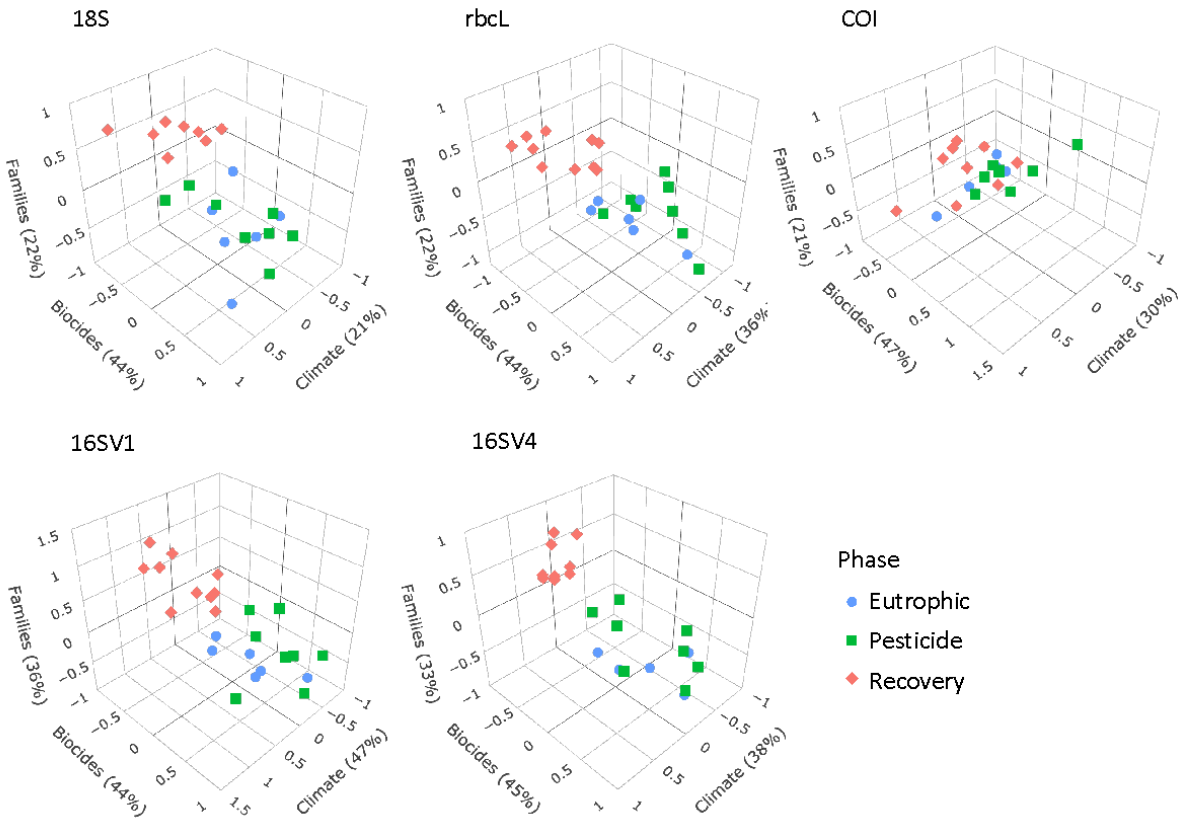
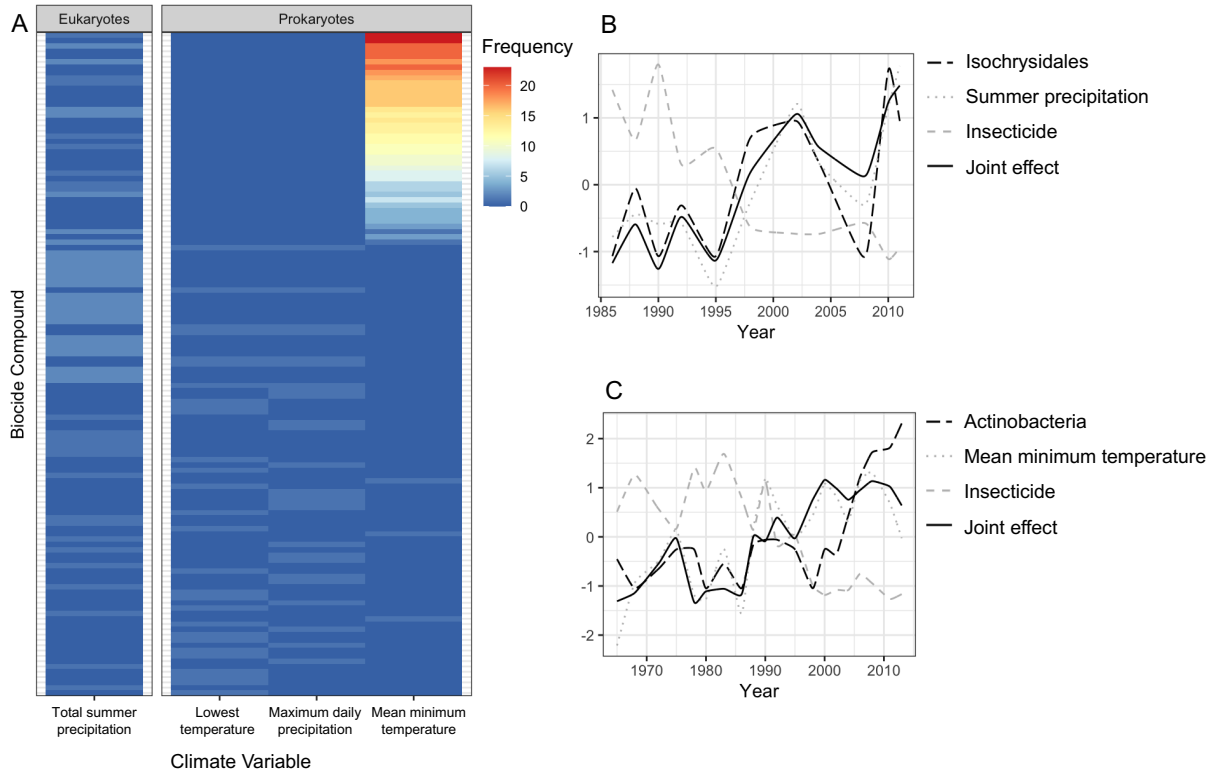


Figure 5. Joint effects of environmental variables on biodiversity. A) heatmap showing the frequency of joint effects of biocides and climate variables in eukaryotes (data from the 18S barcode) and prokaryotes (combined data from 16Sv1 and 16Sv4 barcodes). The biocides are ranked based on their correlation coefficient with taxonomic units and climate variables. Ranking of biocide types is provided in Table S3; B) temporal correlation between the family Isochrysidales, summer precipitation and insecticides. The joint effect of summer precipitation and insecticides is also shown; C) temporal correlation between Pleosporales, insecticides and mean minimum temperature. The joint effect of insecticides and mean minimum temperature is also shown. The families' relative abundance over time in plots B and C are standardized values.



3.4 DISCUSSION

3.4.1 Continuous long-term biomonitoring from a pristine baseline

State-of-the-art paleoecological monitoring typically uses direct observations (light microscopy) of species remains to assess the ecological status of freshwater ecosystems. These approaches are low throughput and require specialist skills (Moraitis, Valavanis, and Karakassis 2019). Direct observations are inherently biased towards species that leave fossil remains; species identification is strongly reliant on well-preserved remains in environmental matrices; and cryptic species diversity cannot be resolved (Hirai J 2017). Recently, automated acquisition of microfossil data using artificial intelligence has been proposed as an alternative to human inspection for reconstructing long-term biological changes (Itaki *et al.,* 2020). However, this approach relies on the completeness of reference databases and of the fossil remains, suffering from the same limitations of direct observations minus the low throughput aspects. Efforts to catalogue temporal changes in biodiversity have recently started to understand changes in species richness and assemblages in different geographic regions of the globe (Blowes *et al.,* 2019). These efforts are important to understand the extent of overall biodiversity loss. However, there are only a handful of existing datasets that span more than 50 years and many of the multidecal biodiversity time series are limited to terrestrial and marine ecosystem, with freshwater ecosystems being marginally represented (Blowes *et al.,* 2019). Moreover, long-term freshwater studies tend to focus on indicator species or specific taxonomic groups (e.g. invertebrates), rather than capturing community-level patterns (Dornelas *et al.,* 2018). Developments in the field of

*seda*DNA have addressed the limitations of direct observations, utilising the properties of eDNA (Capo *et al.,* 2019). However, *seda*DNA studies have predominantly focused on microorganisms as proxies for ecosystems' health (e.g. cyanobacteria (Picard *et al.,* 2022); ciliates (Barouillet *et al.,* 2022); parasitic taxa (Talas *et al.,* 2021)), with other taxonomic groups less well represented. Our study addresses some of the challenges of direct observations as it is not reliant on fossil remains. However, the completeness of the community taxonomic assignment depends on the completeness of reference databases. We acknowledge that our taxonomic classification may be incomplete. Whereas the application of high throughput sequencing technologies requires training, these technologies are well established with publicly available standard operating procedures. As compared to direct observations, high throughput sequencing provides replicable results regardless of the operator. Moreover, the application of metabarcoding to *seda*DNA or more generally eDNA can be outsourced to established environmental services, removing the need for training if it is a limiting factor.

Studies of temporal dynamics typically start from an already shifted baseline and rely on discrete observations (Barouillet *et al.,* 2022). Our study alleviates these limitations by providing a continuous community-level analysis of biological changes over recent evolutionary times and starting from a relatively undisturbed environment. However, eDNA-based studies suffer from limitations linked to the level of preservation of nucleotides in environmental matrices. Although it has been shown that DNA can be recovered from lacustrine and marine sediments as far back as the Holocene (Slon *et al.,* 2022), biases might

still exist due to the degradation of eDNA, especially over geological times (Jia *et al.,* 2022) and in warmer climates (Mauvisseau *et al.,* 2022). In addition, physio-chemical changes in sediment and soil may affect the assemblage and composition of prokaryotic communities that can survive in extreme conditions, including anoxic environments. However, it has been shown that slightly alkaline water (pH 7–9) facilitates DNA preservation (Jia *et al.,* 2022). Whereas we cannot exclude that the eDNA in our study suffers from some of the mentioned biases, we expect DNA degradation not to have affected our study significantly. This is because we observed non-significant difference in species richness over time in both the prokaryotic (16S barcode) and eukaryotic (18S barcode) communities. DNA degradation would have instead resulted in lower alpha diversity with increasing age of the sediment. Preservation of DNA in our study is also favoured by the time frame studied (100 years as opposed to millennia), the stable pH since the 1960s (data prior to 1960s were not recorded), and the latitude of Lake Ring associated with average yearly temperatures below 15℃. All these factors are known to reduce microbial activity, allowing a better preservation of DNA in sediment (Giguet-Covex *et al.,* 2014).

Whereas the overall species richness did not change significantly over time, species assemblages significantly changed over time. Small changes in alpha diversity coupled with significant changes in beta diversity over time have been reported for existing time series biodiversity data in marine and terrestrial environments, even if the length of the time series rarely exceeded four decades (Blowes *et al.,* 2019).

3.4.2 Insecticides and extreme temperatures drive changes in functional biodiversity

Threats to biodiversity pose a significant challenge because they change over time and may result in additive adverse effects (Bonebrake *et al.,* 2019). Long-term continuous observations are preferable to short-term observations because they can reveal correlations and possible causation between biological changes and abiotic drivers of change (Gillson and Marchant 2014). Using eDNA-based data on multitrophic biodiversity over the past 100 years, we identified the taxonomic groups within the prokaryotic and eukaryotic communities that significantly contributed to community assemblages shifts. Whereas the prokaryotic community was overall changing at each major transition between lake phases, changes in the eukaryotic community were driven by different taxonomic groups in the transition between lake phases. The diatom community, typically used by regulators as an indicator of freshwater ecological status, did not change significantly over time, as the beta diversity and the LTDI2 index revealed. These results strongly suggest that a system-level approach, like the one proposed here, may be more appropriate than species or taxon-specific approaches. Our approach showed that diatom communities are not a reliable representation of the ecological status of freshwater ecosystems and are not good indicators of environmental change. Our approach provides a major advantage over traditional direct observations by identifying both taxonomic and functional changes of freshwater biodiversity in a high throughput fashion. The analysis of temporal trends of biodiversity from a pristine baseline through impacted environment provides a new reference point for regulators to define biodiversity in semi-pristine conditions.

Even if Lake Ring partially recovered from eutrophication and biocide pollution in modern times, both the contemporary eukaryotic and prokaryotic communities are significantly different from the semi-pristine historical community, as the PERMANOVA on beta diversity demonstrates. Our findings align with other studies using *seda*DNA on decennial timeframes focusing on prokaryotes (e.g. cyanobacteria (Cao, Xu, *et al.,* 2020)), whereas studies on eukaryotic compositional changes are just emerging to enable quantitative comparative assessments (Zhang *et al.,* 2021). Studies on prokaryotic and eukaryotic assemblages based on short experimental manipulations suggest that natural communities can return to their original state before a perturbation occurs (Hillebrand and Kunze 2020). However, longer-term experimental manipulations show a different perspective with irreversible changes in biodiversity composition and function (Fordham 2015). These long(er)-term experimental manipulations and our study suggest that empirical observation of multi trophic changes over time in natural systems are critical to understand the context-dependency of biodiversity-environmental impact relationships and assess the resilience of natural ecosystems.

Changes in community assemblages are important because they can be associated with changes in functional biodiversity. Although biodiversity variables include taxonomic, phylogenetic, and functional attributes, most studies have focused on generic taxonomic diversity measures - usually measured as species richness or abundance, ignoring functional biodiversity (Li *et al.,* 2020). Biomass and changes in biomass only capture productivity, while disregarding other metrics, such as decomposition or resource turnover

(Gounand *et al.,* 2018). A complete assessment of biodiversity should include functionality (Eisenhauer *et al.,* 2019). In particular, enzyme activities are relevant because they exhibit the functions encoded in genes and reflect the role of microbiota in the transfer of matter and energy from low to high trophic levels in ecosystems. Changes in biological assemblages over time and across lake phases in our study resulted in significant changes in functional biodiversity, observable through changes in metabolic, biosynthesis and degradation functions of the prokaryotic community demonstrated by differentially abundant KEGG pathways between lake phases. Catabolic functions, metabolism (including methane metabolism), degradation and biosynthesis were differentially enriched between the recovery and other lake phases. These are key functions for the survival of organisms. For example, change is metabolic potency and the ability to break down complex molecules into smaller ones (catabolism and degradation) may affect survival and fitness of living organisms by influencing the uptake of nutrients.

Predicting the functional profiles of prokaryotic communities based on their taxonomic composition has its limitations. Predictions of functions linked to human gut microbes tend to be more accurate than predictions on other communities because reference databases are developed on currently available genomes, which are biased towards microorganisms associated with human health and biotechnology (Choi *et al.,* 2017). Because of the bias in reference databases, functional predictions may be more accurate for basic metabolic and housekeeping functions (essential cellular functions that are evolutionary conserved), which are more commonly annotated (Mi *et al.,* 2019). Therefore, it

is possible that we underestimated the predicted changes in functional biodiversity driven by environmental change in our study. Yet, we were able to detect important functional changes (e.g., metabolism and biosynthesis essential for survival) in correspondence with major ecosystem shifts (e.g., from semi-pristine to recovery phase).

In recent years, an increasing number of studies have documented impacts on biodiversity driven by climate change (Pecl *et al.,* 2017), whereas chemicals are thought to pose a negligible threat to biodiversity because living organisms can adapt and evolve (Groh *et al.,* 2022). Adaptation to environmental change can happen, but it comes at a cost that can reduce resilience of natural populations to multiple stressors or novel stress (Cuenca-Cambronero *et al.,* 2021). Our study showed that chemicals and climate variables each explain up to 47% of biodiversity compositional changes and that the joint effect of insecticides/fungicides and yearly extreme temperature/summer precipitation best explained changes in overall biodiversity. The joint effects of insecticides and extreme temperature events affected prokaryotes by altering their functionality and changing their metabolic, biosynthesis and degradation functions. The joint effect of insecticides and summer precipitation best explained changes in primary producers and grazers. This result aligns with previous studies showing that the effect of chemicals on freshwater can be exacerbated by temperature/precipitation, because of changes in the bioavailability, adsorption, elimination and relative toxicity of chemicals by water organisms (Pinheiro *et al.,* 2021). Higher temperatures increase diffusion of chemical molecules, resulting in faster uptake by living organisms and hence

toxicity (Patra *et al.,* 2015). In some cases, higher temperatures result in effects on the organism's metabolic ability to reduce a chemical's toxicity. Our study hints at examples of both mechanisms, distinguishing between families that are negatively and positively correlated with climate variables.

The resolution and reliability of our data-driven systemic approach goes beyond current state-of-the-art, enabling us to identify the specific abiotic factors, down to the commercial name of biocides, that in isolation or combined with climate variables affected specific families of prokaryotes and eukaryotes. Our algorithm provides a high degree of confidence that surpasses state-of-the-art analysis, which predominantly identify patterns of co-occurrence of taxa within communities (e.g., Correlation-Centric Network approach(Yang *et al.,* 2020)). A step in the right direction to capture complex interactions between biotic and abiotic variables is the network analysis of co-occurrence patterns among physico-chemical and biological variables using random forest machine learning algorithms (e.g. (Tse *et al.,* 2018)). This approach is hypothesis-free and allows the identification of synchronicity between various environmental variables and *seda*DNA sequence variation. However, even when applied to temporal trends, it does not quantify joint effects of environmental factors on biodiversity. So far, random forest machine learning algorithms have only been applied to prokaryotic communities, disregarding other taxonomic groups and providing a partial understanding of community-level patterns and responses (Tse *et al.,* 2018).

A potential limitation of our approach is that correlations identified in field surveys do not demonstrate causation. However, they generate testable

hypotheses that can be proven experimentally in controlled mesocosm experiments as explained in (Eastwood *et al.,* 2022), providing a potentially transformative approach.

3.4.3 Implications for conservation and management of biodiversity

Some of the greatest challenges in biodiversity conservation faced by water resource managers is the limited information available on a time scale sufficient to assess long-term changes of aquatic ecosystems. Large scale models that link environmental drivers to biological indicators are lacking (Solimini, Cardoso, and Haiskanen 2005), even if some countries have tried to introduce semi-quantitative indices to assess the ecological status of freshwater (Archaimbault and Dumont 2010). Regulators must rely on approaches ingrained into environmental law, even though they have been proven inadequate (e.g. TDI), as the continuous decline in biodiversity demonstrates (Pecl *et al.,* 2017). Even when direct links between biological indicators and abiotic drivers can be established, these rely on indicator species (e.g. a fish, an alga and an invertebrate) used as proxies for ecosystem health (Kanno 2016). Our data-driven approach provides a novel way to address regulatory needs. However, the use of data-driven, systemic approaches requires critical changes in current environmental practice and a shift to whole-system evidence-based approaches. The transition to the novel methodologies proposed here will require changes in regulatory frameworks, following a test and acceptance phase, as well as a buy-in from regulators. Our study is a proof of concept that the drivers of biodiversity loss can be identified with higher accuracy than

currently possible, generating hypotheses that can be tested experimentally. Our data-driven approach enabled us to identify insecticides and temperature as strong drivers of biodiversity loss, both in prokaryotes and eukaryotes. The confirmation of these findings across multiple freshwater ecosystems has the potential to inform conservation and mitigation interventions, leading to an improved preservation of functional biodiversity.

3.5 MATERIALS AND METHODS

3.5.1 Environmental and paleoecological profile of Lake Ring

Lake Ring is a shallow mixed lake in Jutland, Denmark (55°57'51.83'' N, 9°35'46.87'' E) with a well-known history of human impact (Cuenca - Cambronero *et al.,* 2018). A sedimentary archive was collected from Lake Ring in November 2016 with an HTH-type gravity corer; the core was sliced in 34 layers of 0.5 cm and stored in dark and cold (-20 °C) conditions. A radiometric chronology of this sediment was completed in 2018 by Goldsmith Ecology Ltd following standard protocols (Appleby 2001), and provided an accurate dating of the sediment to the year 1916. According to this chronology the core covered 100 years at a resolution of ca. 3 years intervals. To reduce potential contamination when handling older sediment layers each layer of sediment was handled in a PCR-free and DNA-free environment. Dating of sediment was conducted by direct gamma assay, using ORTEC HPGe GWL series well-type coaxial low background intrinsic germanium detector. Sediment samples with known radionuclide profiles were used for calibration following (Appleby 2001).

We used, historical records, direct chemical analysis of sediment, and physico-chemical records to reconstruct the paleoecological environment of Lake Ring. According to historical records, the lake was semi-pristine until the 1940s. In the late 1950s, sewage inflow from a nearby town increased nutrient levels resulting in eutrophication. The sewage inflow was diverted at the end of the 1970s, but this period coincided with agricultural land-use intensification (>1980), causing biocides leaching into the lake. The lake partially recovered in modern times (>1999), experiencing a partial return to its original trophic state and reduced impact from biocides (Cuenca - Cambronero *et al.,* 2018).

Physico-chemical variables were measured in the lake between 1970 and 2016, even though data are sparse and discontinuous, limiting their use in a machine learning or statistical framework (Supplementary Fig. 4A). To complement the historical records, we obtained climate data from the Danish Meteorological Institute (Supplementary Table 4). The climate data were collected from a weather station 80 km from Lake Ring. Air and water surface temperature typically have a positive correlation for shallow streams and lakes (Livingstone and Lotter 1998; Preudhomme and Stefan 1992). Hence, we used the data from the weather station as an estimate of the lake water temperature. We also observed a tight correlation between the recorded water temperature in Lake Ring and the summer air temperature recorded by the weather station (Supplementary Fig. 4A). In addition, we procured sales records of biocides in Denmark between 1955 and 2015 from the Danish national archives (Supplementary Fig. 4B; Supplementary Table 4). To assess whether the biocide sales records were a good representation of persistent chemicals in the

lake sediment, we quantified the persistent halogenated pesticide DDT in the sliced sedimentary archive of Lake Ring, applying gas chromatography with mass spectrometry analysis (Supplementary Fig. 4C). Sediment samples were lyophilized and freeze dried in a lyophilizator using a Christ Beta 1-8 LSCplus freeze-dryer, (Martin Christ GmbH, Osterode am Harz, Germany), to avoid analyte loss during water removal. Following lyophilization, the sediment samples were sieved through 0.4 mm meshes and homogenised. Approximately 1g of dry sediment was weighed into pre-cleaned glass tubes and spiked with 100 ng of deuterated [2H8- 4,4`- DDT], used as an internal (surrogate) standard, followed by 1 g of copper powder (Merck, Dorset, UK)] for sulphur removal. The sediment samples were extracted using 5ml of hexane: acetone (3:1 v/v), vortexed for 5 min, followed by ultrasonication for 15 min and centrifugation for 3 min at 5000 rpm. The supernatant was transferred to a clean, dry tube and the process was repeated twice for each sample. The combined extract was then evaporated to dryness under a gentle stream of N2 and reconstituted in 2 mL of hexane. Sulphuric acid (3 ml) was used to wash the reconstituted crude extract. The organic phase was allowed to separate on top of the acid layer then transferred to another clean dry test tube. The remaining acid layer was washed twice, each with 2 ml of Hexane. The combined clean extract and washes was evaporated under a gentle stream of Nitrogen, reconstituted into 150 µl of iso-octane containing 100 pg/µl of PCB 131 used as syringe (recovery) standard. Quantification of target DDTs was conducted on a TRACE 1310™ GC coupled to an ISQ™ single quadrupole mass spectrometer

(Thermo Fisher Scientific, Austin, TX, USA) operated in electron ionization (EI) mode according to a previously reported method (Wong *et al.,* 2009).

3.5.2 Biodiversity fingerprinting across 100 years

eDNA extraction and metabarcoding sequencing.

We applied multilocus metabarcoding or marker gene sequencing to environmental DNA (eDNA) extracted from the 34 layers of sediment from the biological archive of Lake Ring using a laminar flow hood in a PCR-free environment to obtain a fine-grained temporal quantification of taxonomic diversity and relative abundance of taxonomic groups. eDNA was extracted from the dated sediment layers - *seda*DNA - using the DNeasy PowerSoil kit (Qiagen), following the manufacturer's instructions. Negative aerial and PCR controls were used; in addition, positive controls for PCR consisting of duplicates of three random samples from the sedimentary archive, were used. The duplicated samples were very similar, providing confidence in the approach used (Supplementary Fig. 2). Triplicates of each *seda*DNA sample were amplified with a suite of five nuclear and mitochondrial PCR primers (barcodes) to capture presence and relative abundance of eukaryotes (18S) (Hadziavdic *et al.,* 2014), macroinvertebrates (COI) (Leray *et al.,* 2013), primary producers (focus on diatoms; rbcL) (Zimmermann *et al.,* 2014), and prokaryotes (16SV1 and 16SV4) (Caporaso *et al.,* 2011a) using Q5 HS High-Fidelity Master Mix (New England Biolabs) and following the manufacturer's instructions. A negative control in triplicate per plate was used. Paired end 250 bp amplicon libraries were obtained using a 2 step PCR protocol with 96x96 dual tag

barcoding to facilitate multiplexing and to reduce crosstalk between samples in downstream analyses (MacConaill *et al.,* 2018) by EnviSion, BioSequencing and BioComputing at the University of Birmingham (https://www.envision-service.com/). PCR1 and PCR2 primers, as well as annealing temperatures per primer pair in PCR1 are in Supplementary Table 5. Excess primer dimers and dinucleotides from PCR1 were removed using Thermostable alkaline phosphatase (Promega) and Exonuclease I (New England Biolabs). PCR2 amplicons were purified using High Prep PCR magnetic beads (Auto Q Biosciences) and quantitated using a 200 pro plate reader (TECAN) using qubit dsDNA HS solution (Invitrogen). A standard curve was created by running standards of known concentration on each plate against which sample concentration was determined. PCR2 amplicons were mixed in equimolar quantities (at a final concentration of 12 pmol) using a biomek FXp liquid handling robot (Beckman Coulter). The final molarity of the pools was confirmed using a HS D1000 tapestation screentape (Agilent) prior to 250 bp paired-end sequencing on an Illumina MiSeq platform.

Bioinformatics.

The reads were demultiplexed using the forward PCR1 primer sequence using cutadapt 3.7.4 with an error rate of 0.07, equating to one allowed mismatch. The quality of sequences was assessed with FASTQC (Wingett and Andrews 2018) and multiqc (Ewels *et al.,* 2016). Sequences were then imported into QIIME2 v 2021.2 (Bolyen *et al.,* 2019), trimmed, filtered, merged and denoised using the QIIME2 DADA2 module (Callahan *et al.,* 2016) using default

parameters and trimming low quality sections and reverse primer [forward read 0-10 trimmed front, 214-225 truncation; reverse read 17-26 trimmed front, 223-247 truncation]. After denoising, the following samples had zero reads remaining: 16SV1, 16SV4, rbcL and COI negative PCR controls; COI aerial negatives A and B; 16SV1 sampleID 8. The taxonomic assignment was completed with the naive-bayes taxonomic classifiers trained using different reference databases, depending on the barcode: the SILVA v138 database was used for the assignment of the 16SV1, 16SV4 and 18S reads (Yilmaz *et al.,* 2014); the diat.barcode v9.2 was used for the assignment of rbcL reads (Rimet *et al.,* 2019); and the Barcode of Life Database was used for the COI reads (Robeson *et al.,* 2021). The taxonomy was assigned using qiime feature-classifier classify-sklearn and used at family level where possible (Pedregosa *et al.,* 2011). When classification was not possible at family level, the lowest classification possible was used. The taxonomic barplots were plotted per barcode using ggplot2 v3.3.5 (Wickham 2016) in R v4.0.2 (Team 2020) and including the top ten most abundant families. All other taxa were collapsed in the plots under 'other taxa'.

All samples were rarefied (16SV1 at 10,250 reads; 16SV4 at 10,400 reads; 18S at 9,070 reads; COI at 3,580 reads; rbcL at 4,650 reads) to achieve normalisation for calculating Alpha and Beta diversity metrics with QIIME2 (Bolyen *et al.,* 2019). The following samples did not meet the rarefaction cutoff: 16SV1: aerial negatives A, B, C; 16SV4: aerial negatives A, B, C and sampleID 62 sample;18S: aerial negatives A,B,C, negative PCR control, sampleID 18, positive control replicate 62; rbcL: aerial negative A, B, and sampleIDs 50, 54,

60; COI sampleIDs 40, 64. Alpha diversity differences among lake phases, using shannon entropy, were tested with Kruskal-Wallis test and beta diversity differences among lake phases, calculated as weighted unifrac distances, were established with a PERMANOVA test (Anderson 2001). Alpha diversity was plotted using ggplot2 v3.3.5 with R v4.0.2. Heatmaps of weighted unifrac Beta diversity between each pair of sediment layers were plotted with the pheatmap v1.0.12 in R v4.0.2 (Kolde 2019).

The function of the microbial communities across the four lake phases were predicted with PICRUST2 (Douglas *et al.,* 2020) plugin in QIIME2 (Bolyen *et al.,* 2019), using the rarefied reads. Differentially abundant KEGG Orthology (KO) terms between pairs of lake phases were identified using the ANCOM plugin (Mandal *et al.,* 2015) in QIIME2 (Bolyen *et al.,* 2019) and were mapped onto KEGG pathways with enriched pathways identified using a Fisher Exact test.

3.5.3 Drivers of biodiversity change

To identify correlations between biological assemblages (families identified through the *seda*DNA sequencing) and drivers of change, we focused on biocides and climate variables, using sparse Canonical Correlation Analysis (sCCA; it can be thought of as consensus PCA on multiple data matrices) followed by Sliding Window (Pearson) Correlation (SWC) analysis (Supplementary Fig. 5). Physico-chemical variables were not used in this analysis because of their sparsity (data rarely met the Sliding Window correlation criteria of 5 continuous values) and low variation over time (Supplementary Figure 6). sCCA is a tool for integrating and discovering

complex, group-wise patterns among high-dimensional datasets (Lin *et al.,* 2013). While most forms of machine learning require large sample sizes, sCCA uses fewer observations to identify the most correlated components among data matrices and captures the multivariate variability of the most important features (Parkhomenko, Tritchler, and Beyene 2009).

Matrices consisting of rarefied ASV reads per barcode, climate data and biocide types were used as input in the analytical pipeline summarised in Supplementary Fig 4. After the sCCA analysis the ASVs were assigned to family level where possible or at the next lowest classifier. The first step of the pipeline is preparing input matrices for ASVs, climate variables and biocides (Supplementary Fig. 5; Step 1). The following step is a matrix-on-matrix regression, applied to correlate families called from the ASVs with either biocide type or climate variables (Supplementary Fig. 5; Step 2). The top five components of the correlations, based on loading values, that explained the largest covariance between matrices were extracted from the sCCA, and the abiotic factors (climate variable and biocide type, separately) ranked according to their contribution to the overall covariance. A Sliding Window (Pearson) Correlation (SWC) analysis followed this step and was applied to each pair of vectors represented by the top ranked abiotic factor and the families. This approach was used to identify abiotic factors (either climate variables or biocide types) that significantly correlated with families over time, using the criterion that their Pearson correlation coefficient should be larger than 0.5 (i.e., large effect size (Nakagawa and Cuthill 2007)) with an FDR adjusted p-value (padj) < 0.05 following 10,000 permutations (Supplementary Fig. 5; Step 3). The minimum

sliding window size was set to 5 time points, corresponding to 15% of the total time window for which families, biocides and climate data were available (the 34 sediment layers from the sedimentary archive span 100 years). Time intervals with more than 50% zero values in either the biotic or the abiotic data were discarded from downstream analyses to reduce false positives. A recall rate was used to quantify the number of ASVs within a family that were individually significantly correlated with the abiotic variables over all ASVs in a given family (Buckland and Gey 1994). The families that co-varied with either biocide types or climate variables over time were retained if they showed a Pearson correlation coefficient > 0.5, a padj < 0.05 and a recall rate > 0.5 (90% quantile of the recall rates of all families) (Supplementary Fig. 5; Step 4). This conservative approach enabled us to reduce noise from spurious correlations and improve accuracy.

The combined effect of environmental factors may have an augmented impact on biodiversity. To identify the combined effect of climate variables and biocides on the lake community biodiversity, we applied again sCCA analysis (Supplementary Fig. 5; Step 5). For this analysis, we selected the climate variables and biocide types contributing the largest covariances in the correlation analysis in Step 4. Their combined effect on a family was considered to be significant if the biocide type and the climate variable were each significantly correlated with the family over the same time window, and their average Pearson correlation was > 0.5 with padj < 0.05 (SWC analysis with 10,000 permutations) (Supplementary Fig. 5; Step 6). The biocide type and the climate variable were interpreted to have an joint effect on a given family if the

linear combination of the biocide type and the climate variable had a larger Pearson correlation coefficient than each of the correlations between the family and the biocide type and the family and the climate variable individually, in the same time interval with padj < 0.05 (with 10,000 permutations in the SWC analysis).

Within each biocide type that significantly correlated with a family, we established their ranking based on the correlation coefficient (Supplementary Fig. 5; Step 6). Significant Pearson correlations that identified the joint effect of climate variables and individual biocides on a given family were identified with the same criteria outlined above (Pearson correlation > 0.5; padj < 0.05; SWC with 10,000 permutations). Chemicals with more than 50% null values or Pearson correlation coefficients < 0.5 were discarded.

3.6 DATA AVAILABILITY

The metabarcoding sequences generated for this project are available at Biosample ID SAMN22315717- SAMN22315798.

3.7 CODE AVAILABILITY

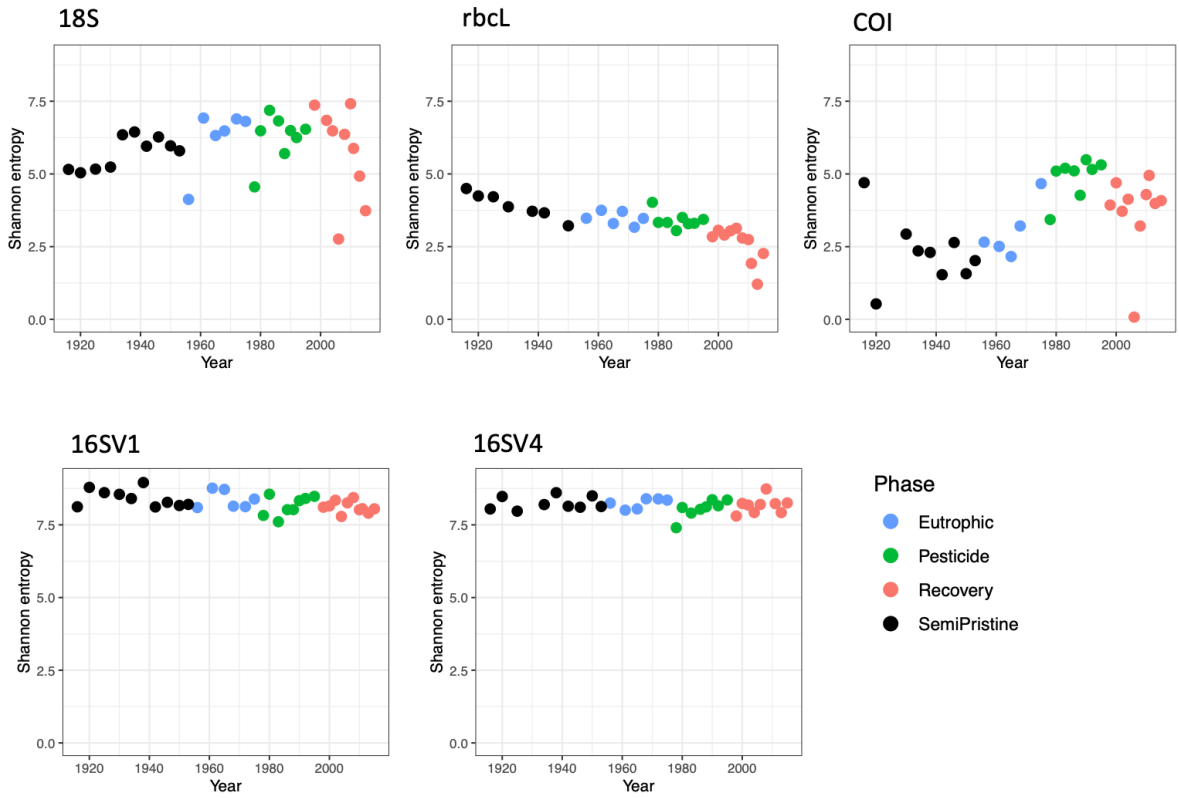
Code used to process and analyse the data in this study are available at https://github.com/Environmental-Omics-Group/Biodiversity_Monitoring

3.8 ACKNOWLEDGMENTS

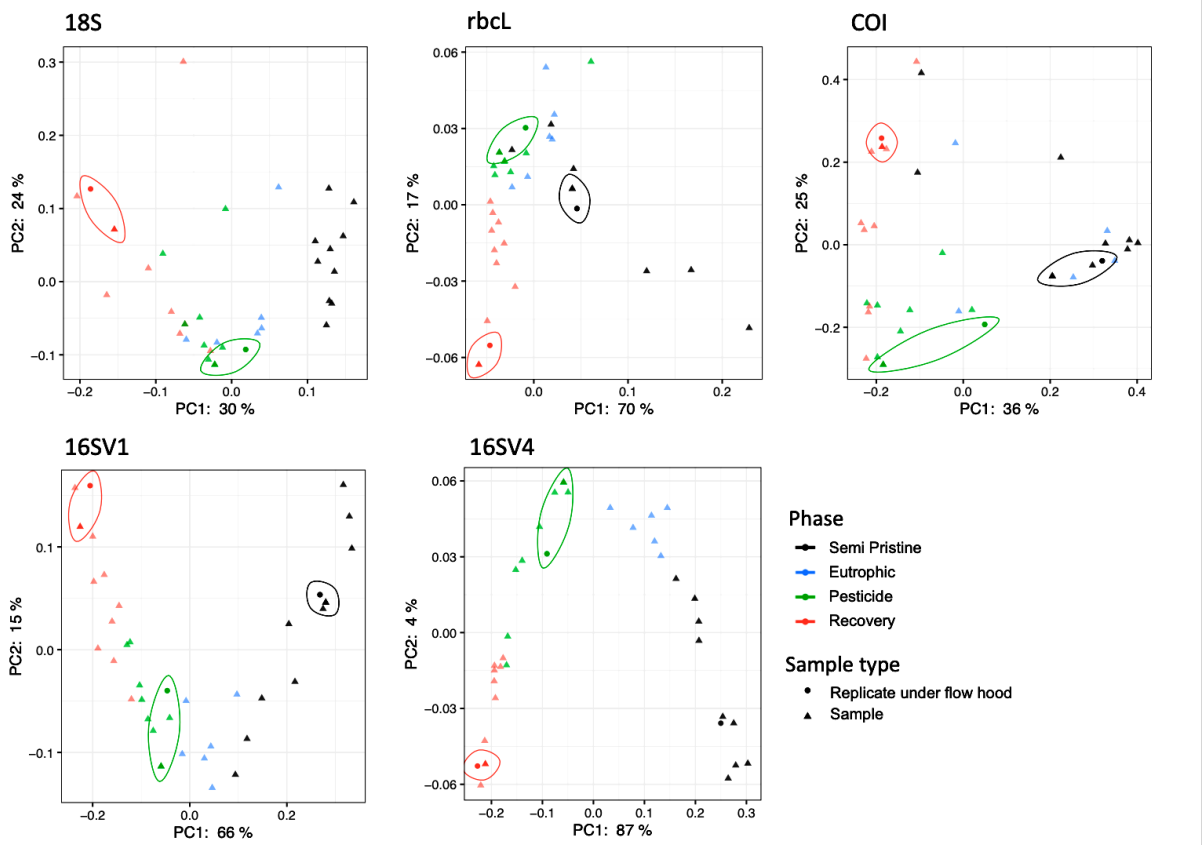
We thank Kerry Walsh and Glenn Watts, the UK Environment Agency, for helpful discussions on the application of the approach presented here within regulatory frameworks. The metabarcoding data were generated by EnviSion, BioSeqencing and BioComputing at the University of Birmingham (https://www.envision-service.com/). The DDT chemical data were generated by the GEES Mass Spectrometry Facility at University of Birmingham. Sediment sampling and dating was completed by Goldsmith Ecology, Somerset. We thank Stephen Kissane for technical assistance in generating high throughput sequencing data, Dr Xiaojing Li for helpful discussions on functional analysis and Chantal Jackson for the artwork of Figure 1. This work was funded by the Alan Turing Institute (under EPSRC grant R-BIR-001); and the NERC highlights grant LOFRESH (NE/N005716/1). Niamh Eastwood is supported by the Midlands Integrative Biosciences Training Partnership (MIBTP; BB/M01116X/1). SEC and HH have been supported by the RobustNature Cluster of Excellence Initiative (internal prefunding of Goethe University Frankfurt).

3.9 SUPPLEMENTARY INFORMATION

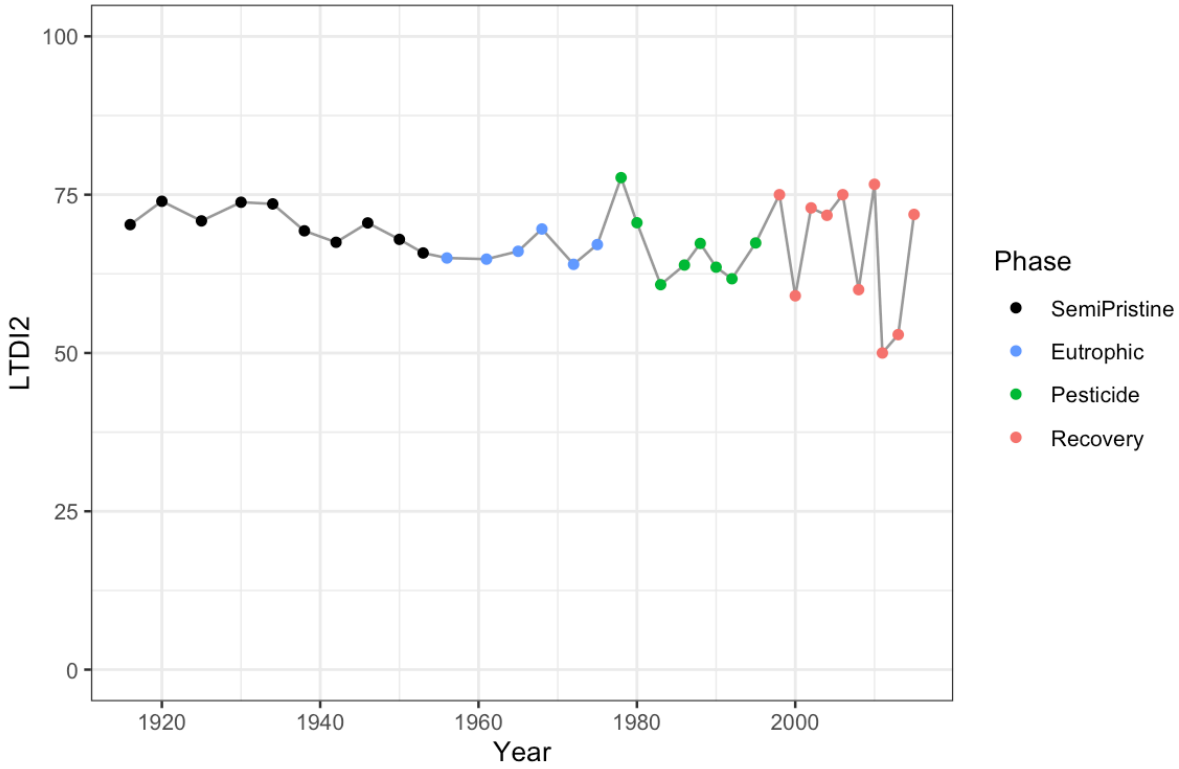
Supplementary Figure 1. Alpha diversity. Alpha diversity, measured as Shannon entropy, is shown for the five barcodes used in this study (16SV1, 16SV4, 18S, COI and rbcl) between 1916-2016. The four lake phases are colour-coded as follows: Black - Semi-pristine; blue - Eutrophic; green - Pesticides; red - Recovery. Kruskal-Wallis test across all phases: 18S: h 4.199, Pval = 0.241; rbcL: h 21.677, Pval<0.000; COI: h 16.958, Pval = 0.001; 16SV1: h 7.001, Pval = 0.072; 16SV4: h 2.220, Pval = 0.528.



Supplementary Figure 2. Principal Coordinate Analysis. PCoA visualization of weighted unifrac distance between samples. Positive controls for PCR consist of duplicates of up to three samples from the sedimentary archive for each of the five barcodes used in the study (16SV1, 16SV4, 18S, rbcL and COI). Replicated samples are circled. The four lake phases are colour-coded as follows: Black - Semi-pristine; blue - Eutrophic; green - Pesticides; red - Recovery.



Supplementary Figure 3. Trophic Diatom Index. LTDI2 calculated using the diatom species identified in our study between 1915 and 2015 with the rbcL barcode and the "DARLEQ3" (Diatoms for Assessing River and Lake Ecological Quality) tool. Mean value of 67.59, standard deviation 6.3. The four lake phases are colour-coded as follows: Black - Semi-pristine; blue - Eutrophic; green - Pesticides; red - Recovery.



Supplementary Figure 4. Biocides records. A) Records of physico-chemical parameters measured in Lake Ring. Dotted lines indicate missing data points. Summer and annual mean temperature were recorded at a weather station 80km from Lake Ring. B) Record of biocides sales in Denmark (Million Tons/Year) between 1950 and 2016, downloaded from the Danish national archives; C) empirical record of DDT measured from the sediment layers of Lake Ring using mass spectrometry analysis (ng/g; blue) and plotted against the sales record in Denmark (Million Tons/year; orange). DDT was banned in Denmark in 1986.

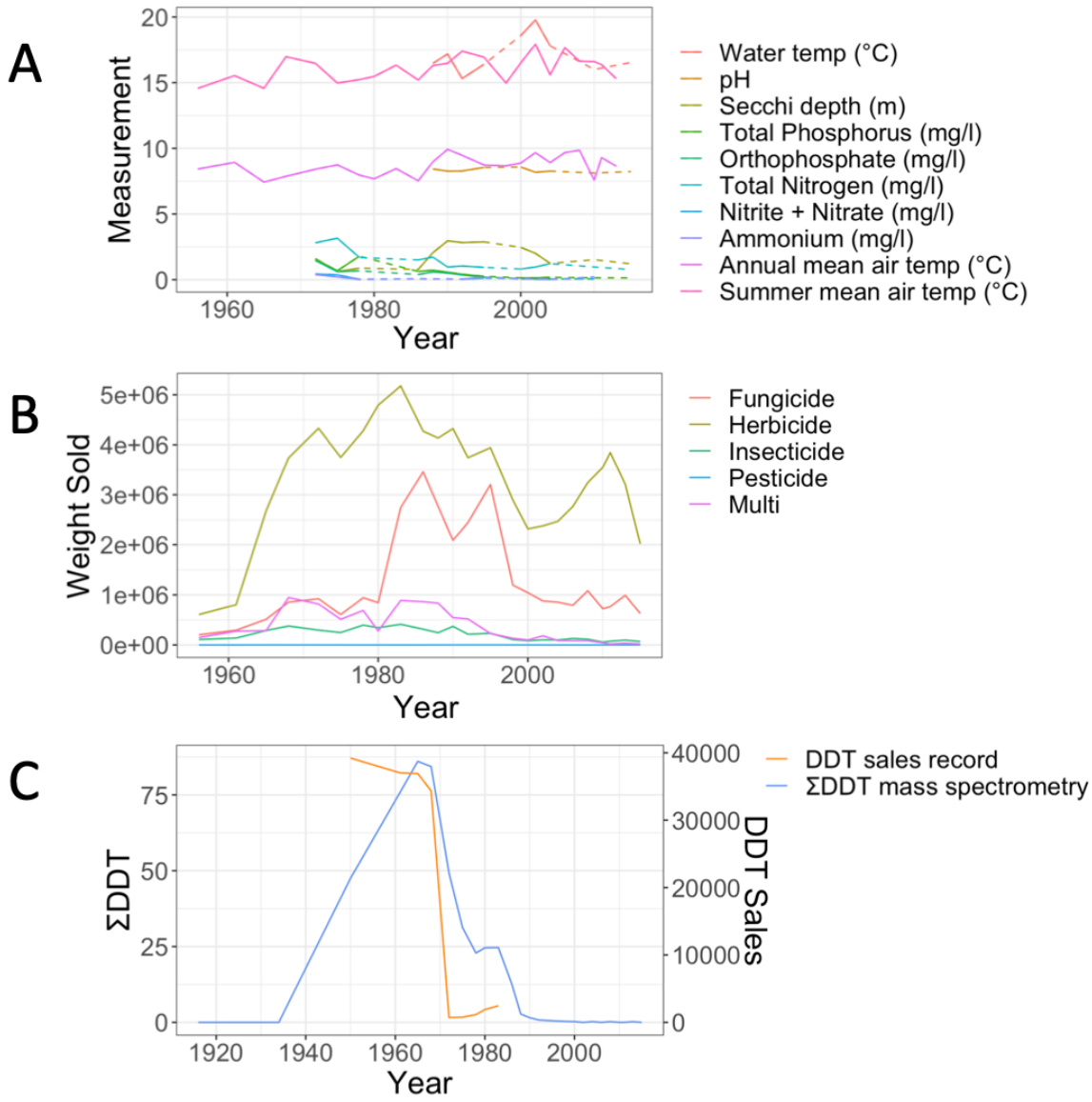
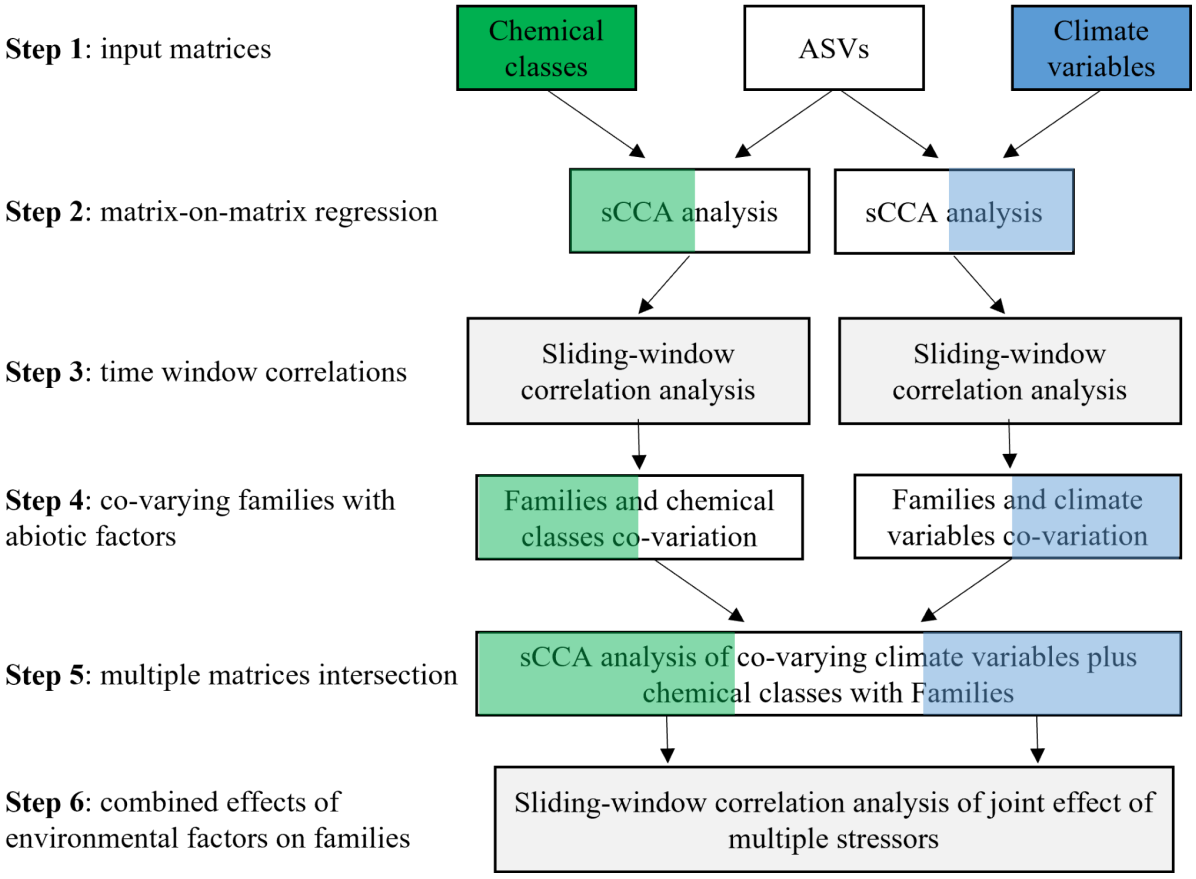

A
Measurement
20
15
10
5
0
1960
1980
2000
Year
Water temp (°C)
pH
Secchi depth (m)
Total Phosphorus (mg/l)
Orthophosphate (mg/l)
Total Nitrogen (mg/l)
Nitrite + Nitrate (mg/l)
Ammonium (mg/l)
Annual mean air temp (°C)
Summer mean air temp (°C)
B
Weight Sold
5e+06
4e+06
3e+06
2e+06
1e+06
0e+00
1960
1980
2000
Year
Fungicide
Herbicide
Insecticide
Pesticide
Multi
C
ΣDDT
75
50
25
0
40000
30000
20000
10000
0
DDT Sales
1920
1940
1960
1980
2000
Year
DDT sales record
ΣDDT mass spectrometry

Figure 5. AI pipeline. The analytical pipeline consists of six main steps: **Step 1** is the preparation of input data matrices (ASVs, biocides and climate variables) to be used in the sCCA analysis. The type of environmental data may vary with the study; **Step 2** is the matrix-on-matrix regression between the ASVs and another environmental data matrix, biocides or climate in this study. Following the sCCA analysis, the ASVs are assigned to family level (or other relevant taxonomic order); **Step 3** consists of a Sliding Window (Pearson) Correlation (SWC) analysis, used to identify significant temporal correlations between families and environmental variables from the sCCA analysis; **Step 4** identifies the families that co-vary with either biocides or climate variables independently; **Step 5** is used to perform an intersection analysis among multiple matrices (families, biodices and climate variables); **Step 6** applies a Sliding Window (Pearson) Correlation (SWC) analysis to identify families, whose relative abundance changes both with biocides and climate variables over time. The pipeline enables the ranking of environmental variables or their combination thereof that is inversely correlated to the relative abundance of families over time.


Step 1: input matrices
Chemical classes
ASVs
Climate variables
Step 2: matrix-on-matrix regression
sCCA analysis
sCCA analysis
Step 3: time window correlations
Sliding-window correlation analysis
Sliding-window correlation analysis
Step 4: co-varying families with abiotic factors
Families and chemical classes co-variation
Families and climate variables co-variation
Step 5: multiple matrices intersection
sCCA analysis of co-varying climate variables plus chemical classes with Families
Step 6: combined effects of environmental factors on families
Sliding-window correlation analysis of joint effect of multiple stressors

Supplementary Table 1 – sCCA analysis. CCA loadings calculated with sparse canonical correlation analysis for biocides (A) and climate variables (B). The categories of biocides are insecticides, fungicides, pesticides and herbicides. The environmental variables are mean minimum temperature, maximum daily precipitation, highest recorded temperature, mean summer temperature, summer precipitation, annual total precipitation, summer atmospheric pressure and lowest recorded temperature.

	18S	**16V1**	**16V4**	**rbcl**	**COI**
A) Biocides					
1	insecticide	insecticide	insecticide	insecticide	fungicide
2	fungicide	fungicide	fungicide	fungicide	insecticide
3	pesticide	pesticide	pesticide	pesticide	pesticide
4	herbicide	herbicide	herbicide	herbicide	herbicide
B) Climate variables					
1	mean minimum temperature	mean minimum temperature	mean minimum temperature	mean minimum temperature	mean minimum temperature
2	summer mean atmospheric pressure	summer total precipitation	maximum daily precipitation	summer total precipitation	annual total precipitation
3	summer mean temperature	highest recorded temperature	summer mean temperature	maximum daily precipitation	highest recorded temp
4	highest recorded temp	summer mean temperature	highest recorded temperature	summer mean temperature	summer mean atmospheric pressure
5	summer total precipitation	lowest recorded temperature	summer total precipitation	annual total precipitation	summer mean temperature

Supplementary Table 2. Correlations between biodiversity and environmental variables. Summary of correlations between taxonomic units identified through the five barcodes (18S, 16SV1, 16SV4, rbcl and COI) and environmental variables, including biocides and climate factors. The taxonomic name and the number of significant correlations between a taxonomic unit and environmental variables, is followed by a correlation value, associated p-adjusted value and recall rate for each variable. The taxonomic units are reported at the lowest taxonomic assignment possible (f – family; o – order; c-class; p – phylum; null - unassigned). Results are collated per barcode, each in a separate tab. The last tab lists only taxonomic units that significantly correlated with the environmental variables based on the combined criteria of Pearson correlation value greater than 0.5, adjusted P-value smaller than 0.05 and recall rate greater than 0.5 along with the direction of the correlation.

See Appendix A – Chapter 3, Supplementary Table 2. Also available as excel spreadsheet.

Supplementary Table 3. Joint effects between biocides and climate variable. The biocides showing significant joint effect with climate variables are ranked based on their correlation coefficient. The barcode and identified families that are affected by the joint effect of a climate variable and biocides type are shown. The order in which the biocide types are ranked is the same used to plot Figure 5.

See Appendix A – Chapter 3, Supplementary Table 3.. Also available as excel spreadsheet.

Supplementary Table 4. Lake Ring metadata. Dating record for Lake Ring, climate data collected from a weather station adjacent to the lake, and sales records for biocides are shown. The year of sampling (year), the sample ID, the depth of the sediment layer measured in centimetres (Depth), climate variables (annual mean temperature °C, summer mean temperature °C, mean minimum temperature °C, mean maximum temperature °C, highest recorded temperature °C, lowest recorded temperature °C, mean atmospheric pressure hPa, summer mean atmospheric presure hPa, annual total precipitation mm, summer precipitation mm, maximum daily precipitation mm, No. of days with snow cover, annual mean cloud cover, and summer mean cloud cover) and record of biocides sales between the 1950s and 2016 in tonnes/year and separated per class (insecticides, herbicides, fungicides and pesticides).

See Appendix A – Chapter 3, Supplementary Table 4.. Also available as excel spreadsheet.

Supplementary Table 5. PCR primers. Tab1) PCR1 primers with bibliographic references, expected fragment size (bp), annealing temperature (°C) and primer sequences (in black) with overhang to prime the sequencing flow cell; Tab2) PCR2 primers consisting of Nextera adapters, universal tail and overhang sequence.

See Appendix A – Chapter 3, Supplementary Table 5.. Also available as excel spreadsheet.

3.10 REFERENCES

1 Baert, J. M., Janssen, C. R., Sabbe, K. & De Laender, F. Per capita interactions and stress tolerance drive stress-induced changes in biodiversity effects on ecosystem functions. *Nat Commun* **7**, 12486, doi:10.1038/ncomms12486 (2016).

2 Cardinale, B. J. *et al.,* Biodiversity loss and its impact on humanity. *Nature* **486**, 59-67, doi:10.1038/nature11148 (2012).

3 Naggs, F. Saving Living Diversity in the Face of the Unstoppable 6th Mass Extinction: A Call for Urgent International Action. *Population and Sustainability* **1**, 67-81 (2017).

4 Bonebrake, T. C. *et al.,* Integrating Proximal and Horizon Threats to Biodiversity for Conservation. *Trends Ecol Evol* **34**, 781-788, doi:10.1016/j.tree.2019.04.001 (2019).

5 Ruckelshaus, M. H. *et al.,* The IPBES Global Assessment: Pathways to Action. *Trends Ecol Evol* **35**, 407-414, doi:10.1016/j.tree.2020.01.009 (2020).

6 Eisenhauer, N. *et al.,* A multitrophic perspective on biodiversity-ecosystem functioning research. *Adv Ecol Res* **61**, 1-54, doi:10.1016/bs.aecr.2019.06.001 (2019).

7 Dornelas, M. *et al.,* BioTIME: A database of biodiversity time series for the Anthropocene. *Glob Ecol Biogeogr* **27**, 760-786, doi:10.1111/geb.12729 (2018).

8 Halpern, B. S. *et al.,* Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nat Commun* **6**, 7615, doi:10.1038/ncomms8615 (2015).

9 Rounsevell, M. D. A. *et al.,* A biodiversity target based on species extinctions. *Science* **368**, 1193-1195, doi:10.1126/science.aba6592 (2020).

10 Eastwood, N. *et al.,* The Time Machine framework: monitoring and prediction of biodiversity loss. *Trends Ecol Evol* **37**, 138-146, doi:10.1016/j.tree.2021.09.008 (2022).

11 Nogues-Bravo, D. *et al.,* Cracking the Code of Biodiversity Responses to Past Climate Change. *Trends Ecol Evol* **33**, 765-776, doi:10.1016/j.tree.2018.07.005 (2018).

12 Blowes, S. A. *et al.,* The geography of biodiversity change in marine and terrestrial assemblages. *Science* **366**, 339-345, doi:10.1126/science.aaw1620 (2019).

13 Hirai J, K. S., Kasai H, Nagai S. . Cryptic zooplankton diversity revealed by a metagenetic approach to monitoring metazoan communities in the coastal waters of the okhotsk sea, northeastern Hokkaido. *Frontiers in Marine Science* **4** (2017).

14 Domaizon, I., Winegardner, A., Capo, E., Gauthier, J. & Gregory-Eaves, I. DNA-based methods in paleolimnology: New opportunities for investigating long-term dynamics of lacustrine biodiversity. *Journal of Paleolimnology* **58**, 1–21 (2017).

15 Capo, E. *et al.,* How does environmental inter-annual variability shape aquatic microbial communities? A 40-Year annual record of sedimentary DNA from a Boreal Lake (Nylandssjön, Sweden). *Frontiers in Ecology and Evolution* **7**, 245 (2019).

16 Barouillet, C. *et al.,* Paleoreconstructions of ciliate communities reveal long-term ecological changes in temperate lakes. *Sci Rep* **12**, 7899, doi:10.1038/s41598-022-12041-7 (2022).

17 Bonebrake, T. C., Christensen, J., Boggs, C. L. & Ehrlich, P. R. Population decline assessment, historical baselines, and conservation. *Conservation Letters* **3** (2010).

18 Groh, K., Vom Berg, C., Schirmer, K. & Tlili, A. Anthropogenic Chemicals As Underestimated Drivers of Biodiversity Loss: Scientific and Societal Implications. *Environ Sci Technol* **56**, 707-710, doi:10.1021/acs.est.1c08399 (2022).

19 Pecl, G. T. *et al.,* Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, doi:10.1126/science.aai9214 (2017).

20 Gillson, L. & Marchant, R. From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology. *Trends Ecol Evol* **29**, 317-325, doi:10.1016/j.tree.2014.03.010 (2014).

21 Cuenca - Cambronero, M. *et al.,* Predictability of the impact of multiple stressors on the keystone species Daphnia *Scientific Reports* **8**, 17572 (2018).

22 Cambronero, C. M., Beasley, J., Kissane, S. & Orsini, L. Evolution of thermal tolerance in multifarious environments. *Mol Ecol* **27**, 4529-4541, doi:10.1111/mec.14890 (2018).

23 Cuenca Cambronero, M., Beasley, J., Kissane, S. & Orsini, L. Evolution of thermal tolerance in multifarious environments. *Mol Ecol* **27**, 4529-4541, doi:10.1111/mec.14890 (2018).

24 Cuenca Cambronero, M., Zeis, B. & Orsini, L. Haemoglobin-mediated response to hyper-thermal stress in the keystone species Daphnia magna. *Evol Appl* **11**, 112-120, doi:10.1111/eva.12561 (2018).

25 Baltrusaitis, T., Ahuja, C. & Morency, L. P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans Pattern Anal Mach Intell* **41**, 423-443, doi:10.1109/TPAMI.2018.2798607 (2019).

26 Agency, E. Phytobenthos - Diatoms for Assessing River and Lake Ecological Quality (River DARLEQ3). (Sterling, UK, 2020).

27 Bennion, H. *et al.,* Assessment of ecological status in UK lakes using benthic diatoms. *Freshwater Science* **33**, 639-654 (2014).

28 Moraitis, M. L., Valavanis, V. D. & Karakassis, I. Modelling the effects of climate change on the distribution of benthic indicator species in the Eastern Mediterranean Sea. *Sci Total Environ* **667**, 16-24, doi:10.1016/j.scitotenv.2019.02.338 (2019).

29 Itaki, T. *et al.,* Innovative microfossil (radiolarian) analysis using a system for automated image collection and AI-based classification of species. *Sci Rep* **10**, 21136, doi:10.1038/s41598-020-77812-6 (2020).

30 Picard, M. *et al.,* Using metabarcoding and droplet digital PCR to investigate drivers of historical shifts in cyanobacteria from six contrasting lakes. *Sci Rep* **12**, 12810, doi:10.1038/s41598-022-14216-8 (2022).

31 Talas, L., Stivrins, N., Veski, S., Tedersoo, L. & Kisand, V. Sedimentary Ancient DNA (sedaDNA) Reveals Fungal Diversity and Environmental Drivers of Community Changes throughout the Holocene in the Present Boreal Lake Lielais Svetinu (Eastern Latvia). *Microorganisms* **9**, doi:10.3390/microorganisms9040719 (2021).

32 Slon, V. *et al.,* Extended longevity of DNA preservation in Levantine Paleolithic sediments, Sefunim Cave, Israel. *Sci Rep* **12**, 14528, doi:10.1038/s41598-022-17399-2 (2022).

33 Jia, W. *et al.,* Preservation of sedimentary plant DNA is related to lake water chemistry. *Enviornmental DNA* **4**, 425–439 (2022).

34 Mauvisseau, Q. *et al.,* The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments. *Environ Sci Technol* **56**, 5322-5333, doi:10.1021/acs.est.1c07638 (2022).

35 Giguet-Covex, C. *et al.,* Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat Commun* **5**, 3211, doi:10.1038/ncomms4211 (2014).

36 Cao, X. *et al.,* Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities. *Water Res* **183**, 116077, doi:10.1016/j.watres.2020.116077 (2020).

37 Zhang, H., Huo, S., Yeager, K. M. & Wu, F. Sedimentary DNA record of eukaryotic algal and cyanobacterial communities in a shallow Lake driven by human activities and climate change. *Sci Total Environ* **753**, 141985, doi:10.1016/j.scitotenv.2020.141985 (2021).

38 Hillebrand, H. & Kunze, C. Meta-analysis on pulse disturbances reveals differences in functional and compositional recovery across ecosystems. *Ecol Lett* **23**, 575-585, doi:10.1111/ele.13457 (2020).

39 Fordham, D. A. Mesocosms Reveal Ecological Surprises from Climate Change. *PLoS Biol* **13**, e1002323, doi:10.1371/journal.pbio.1002323 (2015).

40 Li, F. *et al.,* Human activities' fingerprint on multitrophic biodiversity and ecosystem functions across a major river catchment in China. *Glob Chang Biol* **26**, 6867-6879, doi:10.1111/gcb.15357 (2020).

41 Gounand, I., Little, C. J., Harvey, E. & Altermatt, F. Cross-ecosystem carbon flows connecting ecosystems worldwide. *Nat Commun* **9**, 4825, doi:10.1038/s41467-018-07238-2 (2018).

42 Choi, J. *et al.,* Strategies to improve reference databases for soil microbiomes. *ISME J* **11**, 829-834, doi:10.1038/ismej.2016.168 (2017).

43 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-D426, doi:10.1093/nar/gky1038 (2019).

44 Cuenca-Cambronero, M. *et al.,* Evolutionary mechanisms underpinning fitness response to multiple stressors in Daphnia. *Evol Appl* **14**, 2457-2469, doi:10.1111/eva.13258 (2021).

45 Pinheiro, J. P. S., Windsor, F. M., Wilson, R. W. & Tyler, C. R. Global variation in freshwater physico-chemistry and its influence on chemical toxicity in aquatic wildlife. *Biol Rev Camb Philos Soc* **96**, 1528-1546, doi:10.1111/brv.12711 (2021).

46 Patra, R. W., Chapman, J. C., Lim, R. P., Gehrke, P. C. & Sunderam, R. M. Interactions between water temperature and contaminant toxicity to freshwater fish. *Environmental Toxicology and Chemistry* **34**, 1809–1817 (2015).

47 Yang, P. *et al.,* Correlation-Centric Network (CCN) representation for microbial co-occurrence patterns: new insights for microbial ecology. *NAR Genom Bioinform* **2**, lqaa042, doi:10.1093/nargab/lqaa042 (2020).

48 Tse, T. J. *et al.,* Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial Communities. *Environ Sci Technol* **52**, 6842-6853, doi:10.1021/acs.est.7b06386 (2018).

49 Solimini, A. G., Cardoso, A. C. & Haiskanen, A.-S. Linkages between chemical and biological quality of surface waters. 1-262 (Joint Research Centre, Ispra, Italy, 2005).

50 Archaimbault, V. & Dumont, B. The normalized global biological index (IBGN): principles and evolution within the framework of the European

framework directive on water. *Sciences Eaux & Territoires*, doi:https://doi.org/10.14758/SET-REVUE.2010.1.08 (2010).

51 Kanno, J. Introduction to the concept of signal toxicity. *J Toxicol Sci* **41**, SP105-SP109, doi:10.2131/jts.41.SP105 (2016).

52 Appleby, P. G. *Chronostratigraphic techniques in recent sediments*. Vol. 1 (Kluwer Academic Publisher, 2001).

53 Livingstone, D. M. & Lotter, A. F. The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palæolimnological implications. *Journal of Paleolimnology* **19**, 181–198 (1998).

54 Preudhomme, E. B. & Stefan, H. G. Relationship between water temperatures and air temperatures for central U.S. streams. (University of Minnesota, Duluth, Minnesota, 1992).

55 Wong, F., Robson, M., Diamond, M. L., Harrad, S. & Truong, J. Concentrations and chiral signatures of POPs in soils and sediments: a comparative urban versus rural study in Canada and UK. *Chemosphere* **74**, 404-411, doi:10.1016/j.chemosphere.2008.09.051 (2009).

56 Hadziavdic, K. *et al.,* Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One* **9**, e87624, doi:10.1371/journal.pone.0087624 (2014).

57 Leray, M. *et al.,* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool* **10**, 34, doi:10.1186/1742-9994-10-34 (2013).

58 Zimmermann, J. *et al.,* Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PLoS One* **9**, e108793, doi:10.1371/journal.pone.0108793 (2014).

59 Caporaso, J. G. *et al.,* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA* **108**, 4516-4522, doi:10.1073/pnas.1000080107 (2011).

60 MacConaill, L. E. *et al.,* Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* **19**, 30, doi:10.1186/s12864-017-4428-5 (2018).

61 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338 (2018).

62 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

63 Bolyen, E. *et al.,* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).

64 Callahan, B. J. *et al.,* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).

65 Yilmaz, P. *et al.,* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**, D643-648, doi:10.1093/nar/gkt1209 (2014).

66 Rimet, F. *et al.,* Diat.barcode, an open-access curated barcode library for diatoms. *Sci Rep* **9**, 15116, doi:10.1038/s41598-019-51500-6 (2019).

67 Robeson, M. S., 2nd *et al.,* RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol* **17**, e1009581, doi:10.1371/journal.pcbi.1009581 (2021).

68 Pedregosa, F. *et al.,* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).

69 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, 2016).

70 Team, R. C. (Vienna, Austria, 2020).

71 Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32-46 (2001).

72 Kolde, R. *Pretty Heatmaps*, <https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf> (2019).

73 Douglas, G. M. *et al.,* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* **38**, 685-688, doi:10.1038/s41587-020-0548-6 (2020).

74 Mandal, S. *et al.,* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **26**, 27663, doi:10.3402/mehd.v26.27663 (2015).

75 Lin, D. *et al.,* Group sparse canonical correlation analysis for genomic data integration. *BMC Bioinformatics* **14**, 245, doi:10.1186/1471-2105-14-245 (2013).

76 Parkhomenko, E., Tritchler, D. & Beyene, J. Sparse Canonical Correlation Analysis with Application to Genomic Data Integration. *Statistical Applications in Genetics and Molecular Biology* **8**, 1-34 (2009).

77 Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* **82**, 591-605, doi:10.1111/j.1469-185X.2007.00027.x (2007).

78 Buckland, M. & Gey, F. The relationship between recall and precision. *Journal of the American society for information science* **45**, 12-19 (1994).

Chapter 4: SINGLE METABARCODING MULTIPLEX CAPTURES COMMUNITY-LEVEL FRESHWATER BIODIVERSITY AND BEYOND

Niamh Eastwood*[1], Stephen Kissane[1], Lewis Campbell[2], Andrew G. Briscoe[2], Bastian Egeter[2], Luisa Orsini[1,3,4]

[1]Environmental Genomics Group, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

[2]NatureMetrics Ltd., Surrey Research Park, Guildford, GU2 7HJ UK

[3]Institute for Interdisciplinary Data Science and AI, the University of Birmingham, Birmingham, B15 2TT, UK

[4]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB

*First author

AUTHOR CONTRIBUTIONS

NE and SK optimised the experimental conditions for the multiplex. NE completed data analysis. LC, AGB and BE provided materials and protocols. LO conceived the study and supervised data analysis. LO and NE drafted the manuscript. All authors contributed to manuscript writing.

4.1 ABSTRACT

Cost-effective and accurate quantification of biodiversity is important for biodiversity conservation, resource management, and forecasting. Traditional monitoring approaches have relied on direct observations, remote sensing, and mark-recapture techniques, providing insights into species ecology and the impact of pollution and climate change on indicator species. However, these techniques are typically expensive, and can be invasive or low throughput. In addition, they cannot detect cryptic diversity and are biased towards species that leave identifiable remains.

DNA-based methods, such as metabarcoding or single marker gene assays, have enabled high throughput screening of a wide range of taxonomic groups, including ones without well-preserved remains. As compared to traditional techniques, these approaches are high throughput, can resolve cryptic diversity, do not require taxonomic specialist skills and are non-invasive. However, although they are comparatively cheaper than traditional approaches, they are expensive when applied at community-level as single marker assays are amplified and sequenced independently. Multilocus approaches in which multiple gene markers are amplified in a single reaction are desirable to deliver community-level assessments in a cost-effective manner. Yet, they are uncommon because of technical challenges that may lead to biases in downstream analyses, such as index hopping and unbalanced representation of taxonomic groups. Here, we developed a highly multiplexed protocol that combines early pooling of marker genes that target broad taxonomic groups and taxon-specific markers in a single tube reaction. This is combined with

unique dual-indexed sequencing adapters that allow the pooling of up to 384 samples per marker gene (N=1,536 samples) in a single sequencing run. This approach dramatically reduces the costs of community-level biodiversity quantification and lowers the needs for input DNA without compromising output quality. We optimised the multiplex assay on lake freshwater sediment samples and benchmarked the assay on samples from other environmental matrices, demonstrating its direct application to river and marine communities.

Keywords: eDNA, metabarcoding, multiplex, singleplex, freshwater, marine, sediment

4.2 INTRODUCTION

Biodiversity monitoring is the foundation for effective biodiversity conservation, resource management, and forecasting. Traditional biodiversity monitoring techniques, such as direct observations, remote sensing and mark recapture techniques have provided valuable insights into species ecology and the impact of pollution and climate change on indicator species (Eveleigh *et al.,* 2007; Ropert-Coudert and Wilson, 2005; Yoccoz, 2012). However, they have significant limitations due to the difficulties associated with correct identification of (cryptic) species or life stages from the same species. These techniques require specialist taxonomic expertise, are not standardised when it comes to sampling protocols, and can be invasive. Traditional techniques are also typically low throughput and biased towards species that leave identifiable remains (Gillson and Marchant, 2014).

In the last decade, DNA-based methods (e.g. metabarcoding) have revolutionised conventional biodiversity research by enabling high throughput screening from environmental matrices without being limited to taxonomic groups with well-preserved remains (Creer *et al.,* 2016; Cristescu and Hebert, 2018). DNA-based approaches have a higher throughput than traditional approaches, do not require taxonomist specialist skills, can resolve cryptic diversity and be applied to bulk DNA extracted from environmental matrices ('Environmental DNA' or eDNA (Cristescu and Hebert, 2018)). Furthermore, large surveys can be conducted in a relatively fast and cost-effective manner, and over large geographic areas, including remote regions, non-invasively (Taberlet *et al.,* 2018). By matching sequence similarity to records in public

databases (e.g., NCBI and SILVA), molecular operational taxonomic units (MOTUs) can be identified, enabling the analysis of taxonomic compositional shifts and estimates of species richness.

To date, most metabarcoding studies have used a single marker gene approach; cytochrome c oxidase subunit I (COI) is commonly used for vertebrates and invertebrates (Krehenwinkel *et al.,* 2018; Leray *et al.,* 2013), internal transcribed spacer (ITS) is used to identify fungi (Schmidt *et al.,* 2013), plastid DNA (rbcL) is used for plants and primary producers (Chase and Fay, 2009; Tse *et al.,* 2018), and 12S for fish (Miya *et al.,* 2015). A multi-marker gene approach is highly desirable to deliver community-level assessments in a cost-effective manner (Ficetola and Taberlet, 2023). However, thus far, community-level assessments of biodiversity have only been achieved with the integration of results from individual marker genes (Eastwood, Zhou, *et al.,* 2023; Eastwood *et al.,* 2022; Li *et al.,* 2023). A multi-locus approach improves the robustness of taxonomic assignment alleviating false negatives caused by random missed amplifications of target genes caused for example by DNA degradation or mutation in primer sites (Zhan *et al.,* 2014). Species detection rate based on multiple loci can be up to 35% more accurate than when using single locus approaches (Zhang *et al.,* 2018). However, highly multiplexed approaches that enable estimates of community-level biodiversity in a cost-effective manner are largely missing because they can be challenging to optimise (Ficetola and Taberlet, 2023). A step in the right direction is 2-step PCR protocols, which include a first-round PCR (1st PCR) that amplifies a target DNA locus or marker gene region using universal primers, followed by a

second-round PCR (2nd PCR) that appends sample-specific indexes to marker gene regions (Bohmann *et al.,* 2022). After the 2nd PCR, samples are usually combined in a multiplex for high throughput sequencing. Alternatively, a sample-specific index may be added to the 1st PCR (Bohmann *et al.,* 2022). This approach improves throughput, but it requires upfront costs (Bohmann *et al.,* 2022; Caporaso *et al.,* 2011b; Ushio *et al.,* 2022).

We aimed to develop a cost-effective assay which captures community wide biodiversity by combining multiple marker gene regions whilst maintaining comparable accuracy and sensitivity to 'standard' single marker gene assays. We aimed to optimise this protocol for lake freshwater communities because these communities support humans and wildlife (Darwall *et al.,* 2018), and have high conservation value, delivering important ecosystem services (e.g. clean water, food provision and recreation) (Dudgeon *et al.,* 2006; Ruckelshaus *et al.,* 2020). In addition, we aimed to assess the suitability of the assay for applications in diverse environments, other than freshwater lakes.

In this study we developed a highly multiplexed protocol that combines early pooling at the PCR1 stage and pooling after PCR2 at the sequencing stage, significantly reducing costs of metabarcoding and lowering the amount of input DNA required to capture community-level biodiversity. We combined three marker gene primer pairs that target 18S and 16S loci, broadly capturing prokaryotes and eukaryotes, with a taxon-specific locus (rbcL) used by regulators to determine water quality in both rivers and lakes. By combining four loci, we capture community-level biodiversity in a single tube reaction.

We benchmarked the multiplex by comparing taxonomic detection rates and accuracies of the multiplex with single locus metabarcoding assays on the same samples. We validated the multiplex on independently sampled material originating from diverse environmental matrices, including river water, soil, peatland, coastal and offshore marine environments. The developed multiplex approach has the potential to significantly improve capacity for both biodiversity routine monitoring and research discoveries.

4.3 MATERIAL AND METHODS

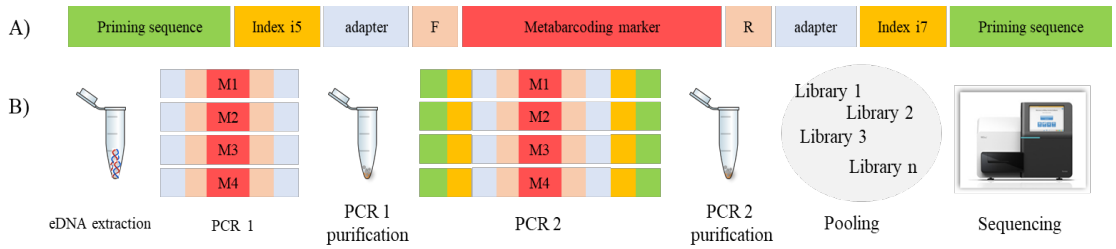
4.3.1 Multiplex optimisation on freshwater *seda*DNA samples

A standard 2-step PCR protocol includes a primary reaction for the marker gene or target locus in PCR1, in which primers with 5′ sequence overhangs are added to the marker gene of choice, and a PCR2, which carries sequencing adapters and indices to be attached to cleaned PCR1 products (Bohmann *et al.,* 2022; Ushio *et al.,* 2022). PCR2 libraries are then pooled for sequencing on an Illumina or Illumina-compatible platform, following removal of excess primers. We modified this protocol by multiplexing four loci in PCR1 (Figure 1). The multiplex protocol was optimised on bulk environmental DNA (eDNA) extracted from freshwater lake sediment from a previous study (Eastwood, Zhou, *et al.,* 2023) using DNeasy PowerSoil kit (Qiagen) (*seda*DNA), following the manufacturer instructions, in a PCR free environment. Extraction and PCR blanks were used to monitor for contamination. The metabarcoding loci used in the multiplex were: two regions targeting eukaryotes broadly [(18SV1V2) (Hadziavdic *et al.,* 2014) and (18SV8V9) (Bradley, Pinto, and Guest, 2016)],

and prokaryotes (16SV4) (Caporaso *et al.,* 2011a), plus a taxon-specific marker targeting diatoms (rbcL) (Kelly *et al.*, 2018) (Table S1). Triplicate samples were amplified in PCR1 using Q5 HS High-Fidelity Master Mix (New England Biolabs) following the manufacturer's instructions. To protect commercially sensitive information, the amplification parameters of this step will not be disclosed. After removing excess primers with High Prep PCR magnetic beads (Auto Q Biosciences), cleaned PCR1 products were pooled in a second PCR in which unique dual-indexed sequencing adapters (Table S2) allowed pooling of up to 384 samples per marker gene in a single sequencing run (N = 1,536 samples per run). Unique dual barcodes (indices) were to reduce index-misassignment and index-hopping between samples (MacConaill *et al.,* 2018). PCR2 amplicons were purified using High Prep PCR magnetic beads (Auto Q Biosciences) and quantitated using a 200 pro plate reader (TECAN) using qubit dsDNA HS solution (Invitrogen). A standard curve was created by running standards of known concentration on each plate against which sample concentration was determined. To assess whether each of the amplicons were equally represented in the PCR1 multiplex, we performed single gene marker PCRs on an aliquot of the cleaned PCR1 products (Figure S1). This approach assessed the approximate quantity of each of the four marker genes in the PCR1 multiplex product as a quality check prior to sequencing. This approach was adopted because the amplicons had overlapping lengths and could not be distinguished based on gel migration alone (Figure S1). PCR2 libraries were mixed in equimolar quantities (at a final concentration of 12 pmol) using a biomek FXp liquid handling robot (Beckman Coulter). The final molarity of the

pools was confirmed using a HS D1000 tapestation screentape (Agilent) prior to 250 bp paired-end sequencing on an Illumina MiSeq platform.

FIGURE 1 A) The composition of a dual-indexed metabarcoding Illumina library sequence, including a metabarcoding marker, forward (F) and reverse (R) primers, sequence adapters, Illumina indexes (i5 and i7) and sequences used to prime to sequencing flow cell. B) Multiplex key steps are shown, including eDNA extraction, multiplexed PCR1 with 4 metabarcoding markers plus a cleaning step to remove excess primers, PCR2 plus a second cleaning step and a final pooling step before sequencing on an Illumina or Illumina compatible platform.



4.3.2 Single and multiplex performance on freshwater sediment samples

The sequenced reads were demultiplexed per marker gene using cutadapt v4.1 (Martin, 2023), and analysed with QIIME2 v2022.8 (Bolyen *et al.,* 2019). Trimming, filtering, merging and denoising of reads was done using the QIIME2 DADA2 module (Callahan *et al.,* 2016) with default parameters. Taxonomy assignment was completed with the QIIME2 feature-classifier module with naive-bayes taxonomic classifiers trained using different reference databases,

depending on the barcode: the SILVA v138 database was used for the assignment of the 16S and 18S reads (Yilmaz *et al.,* 2014); and the diat.barcode v9.2 was used for the assignment of rbcL reads (Rimet *et al.,* 2019). The cleaned reads were rarefied and diversity indices (e.g., alpha and beta diversity) were calculated using the QIIME2 diversity module.

The performance of single and multiplex assays was assessed by comparing alpha and beta diversity, using the rarefied reads. Alpha diversity was measured as Pielou evenness and Shannon diversity, supported by Kruskal-Wallis (Kruskal and Wallis, 1952) using the function alpha-group-significance in the QIIME2 diversity module. Beta diversity was measured as Bray-Curtis distance and significant differences between single and multiplex assays assessed with a PERMANOVA test (999 permutations) using the function beta-group-significance in the QIIME2 diversity module.

4.3.3 Benchmarking the multiplex on eDNA from other environmental matrices

We benchmarked the multiplex using eDNA samples extracted from different environmental matrices and including grassland, marine coastal and marine offshore water, marine coastal and marine offshore sediment, river water, peatland and woodland (Table S3). These samples were provided by NatureMetrics (www.naturemetrics.co.uk) and are therefore anonymised to adhere to data protection requirements. eDNA from water samples (freshwater and marine) was extracted using a DNeasy Blood and Tissue Kit (Qiagen) following Spens (Spens *et al.,* 2017). The original method was modified as

described in Egeter *et al.,* (2023). Briefly, proteinase K was added directly to the disc filters on which water was filtered and stored. 1 mL of the lysate was carried forward for extraction with the DNeasy Blood and Tissue Kit (Qiagen). eDNA was extracted from soil and sediment samples using DNeasy PowerSoil Kit (Qiagen). An extraction blank was processed with each batch of extractions to assess potential contamination in the DNA extraction process. DNA was purified to remove PCR inhibitors using a DNeasy PowerClean Pro Cleanup Kit (Qiagen). Purified DNA extracts were quantified using a Qubit dsDNA HS Assay Kit on a Qubit 3.0 fluorometer (Thermo Scientific). The DNA concentration was quantified using a Qubit DNA broad range kit.

Some samples extracted from woodland, grassland and peatland did not generate a visible PCR 1 product on agarose gel. We suspected that PCR inhibitors (e.g., humic substances) were responsible for these failures. Therefore, we tested the single and multiplex assays with the addition of bovine serum albumin (BSA) (Ramalingam, Warkiani, and Gong, 2017) and compared the performance of these assays with the regular assays described above. The samples collected from different environmental matrices were amplified with the same four amplicons used in the *seda*DNA samples above (Table S1), both in single and multiplex assays. The metabarcoding libraries and sequencing strategy were the same as above. The sequence preprocessing strategy and the taxonomic assignment followed the same strategy used for the *seda*DNA samples described above.

The performance of the benchmarking samples used in single and multiplex assays was assessed by comparing overall alpha and beta diversity, as above.

Significant differences in beta diversity (Bray-Curtis distance) between single and multiplex assays was assessed with a PERMANOVA test (999 permutations) with sample type as strata using the function pairwise.adonis2 (v0.0.1)(Arbizu, 2017), wrapping the package vegan (v2.5-7) (Oksanen *et al.,* 2020) in R (v4.0.2) (R core team, 2020). In addition, we assessed alpha diversity (Pielou evenness and Shannon diversity) of sample types (e.g. river water, marine sediment) between single plex and multiplex using a Wilcoxon signed rank test with Benjamini & Hochberg correction for multiple testing in qiime2 using the pairwise-distances function in the longitudinal module in QIIME2 (Bokulich *et al.*, 2018). PCoA of Bray-Curtis distances was used to visualise the similarity between single and multiplexed samples, plotted using ggplot2 (v3.4.0) (Wickham, 2016) in R (v4.0.2) (R core team, 2020). A Venn diagram was used to visualise the overlap of species and ASVs between the single and multiplex assays for the total number of features (100%), as well as for the topmost abundant features making up 85% and 70% of the reads in the two assays. This approach was used to determine whether discrepancies, if any, between assays could be explained by the capture efficiency of rare species. To assess whether sequencing effort would lead to a convergence in the number of ASVs detected by the two assays, we performed a rarefaction analysis with resampling strategy with the function rarecurve in the package vegan (v2.6–4) (Oksanen *et al.,* 2020) using R (v4.0.2) (R Core Team, 2020), plotted with ggplot2 (v3.4.0) (Wickham, 2016).

4.4 RESULTS

4.4.1 Single and multiplex assays on sedaDNA samples

The rarefaction depth for the *seda*DNA samples was as follows: 16SV4 = 8,245; 18SV1V2 = 12,584; 18SV8V9 = 17,703; and rbcL = 6,372. The alpha diversity measured on *seda*DNA samples, both as Pielou evenness and Shannon diversity, did not significantly differ between single and multiplex assays (Table 1). The beta diversity measured as Bray-Curtis distance across the four gene markers did not differ significantly between single marker and multiplex assays (Table 2).

TABLE 1 Kruskal-Wallis test on Pielou's evenness and Shannon diversity calculated between single and multiplex assays for four marker genes (16SV4;18SV1V2; 18SV8V9 and rbcL) sequenced on the *seda*DNA samples. Significant P-values are in bold.

	Pielou Evenness (J)	Pielou Evenness P value	Shannon diversity (S-W)	Shannon diversity P value
16SV4	0.0	1.0	2.4	0.12
18SV1V2	0.0	1.0	0.33	0.56
18SV8V9	1.5	0.22	0.0	1.0
rbcL	0.6	0.44	0.6	0.44

TABLE 2 Permutational Multivariate Analysis of Variance (PERMANOVA) using Bray-Curtis distance (F) assessing differences between single plex and multiplex assays on *seda*DNA samples (999 permutations) across four gene markers (16SV4; 18SV1V2; 18SV8V9 and rbcL).

Marker gene	F	P value
16SV4	0.50	0.68
18SV1V2	0.36	0.81
18SV8V9	0.37	0.66
rbcL	0.29	0.68

4.4.2 Benchmarking the newly developed multiplex in other environmental matrices

The rarefied sequence depth for the multiplex benchmarking samples were as follows: 16SV4 =5,549; 18SV1V2= 7,734; 18SV8V9=10,900; rbcL = 1,590. The Pielou evenness index (alpha diversity) measured across all samples extracted from different environmental matrices did not differ significantly between single marker genes and multiplex assays for the 16S and both the 18S barcodes (Table 3; Figure 2). A significantly different Pielou index was observed between single and multiplex assays for the rbcL barcode, for which some samples in the single and multiplex assays showed dissimilar evenness (Table 3; Figure 2). The Shannon index showed more variability than the Pielou evenness with three out of four indices showing significant difference between single and multiplex (Table 3; Figure 2). The addition of BSA to the PCR reactions improved the amplification results, but did not significantly change the

sequencing results, as the statistical tests comparing multiplex assays with and without BSA showed (Table 3). No significant (Wilcoxon signed rank P adj. >0.05) difference in alpha diversity (both Pielou evenness and Shannon diversity) was observed between individual sample types that successfully amplified with both the single plex and multiplex (Table 4).

FIGURE 2 Alpha diversity mean and standard error for the benchmarking samples listed in Table S1 measured in single (black) and multiplex (red) assays. For each sample type mean and standard deviation are shown. Lack of a data point indicates that the specific sample/plex failed.

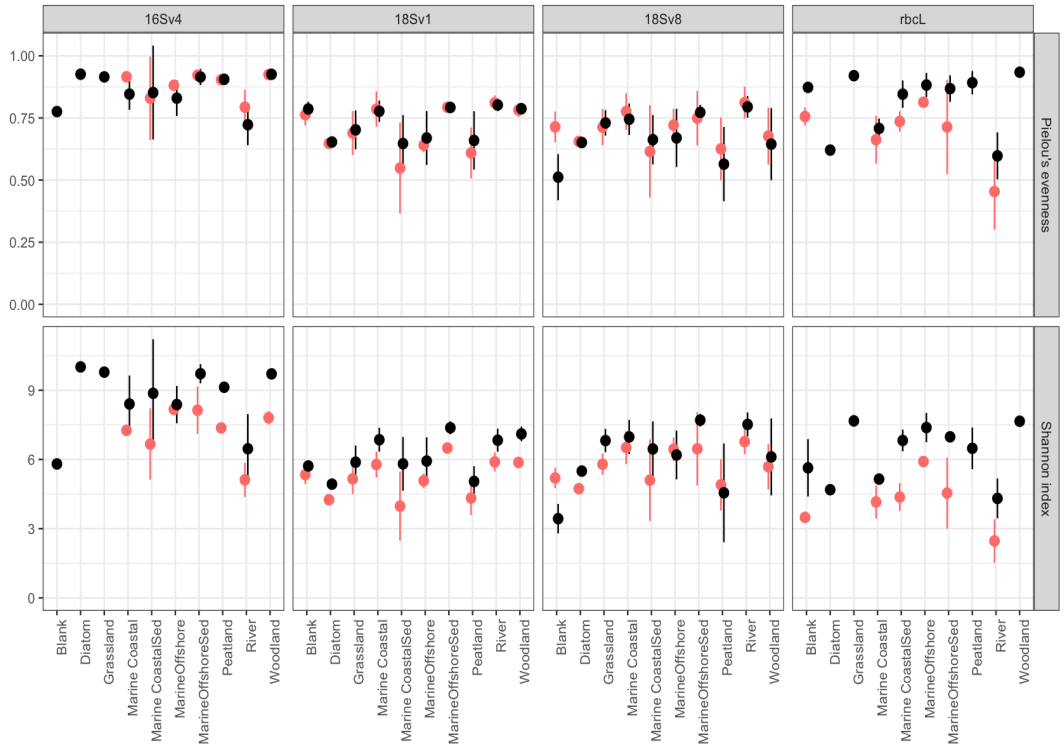


TABLE 3. Kruskal-Wallis test on Pielou's evenness (H) and Shannon diversity calculated between single and multiplexes (both regular and BSA), and between the two multiplexes assays in the benchmarking samples. P-values are Benjamini & Hochberg corrected. Significant adjusted P-values are in bold. Splex – single plex; mplex – multiplex (regular protocol); bsa – multiplex with addition of BSA.

			Pielou Evenness		Shannon Diversity	
	Group 1	Group 2	H	P adj value	H	P adj value
16Sv4	bsa (n=20)	mplex (n=19)	0.079	0.779	0.064	0.800
16Sv4	bsa (n=20)	splex (n=27)	0.778	0.766	15.504	0.000
16Sv4	mplex (n=19)	splex (n=27)	0.433	0.766	15.681	0.000
18Sv1v2	bsa (n=26)	mplex (n=27)	0.020	0.887	0.038	0.845
18Sv1v2	bsa (n=26)	splex (n=29)	0.329	0.850	11.482	0.001
18Sv1v2	mplex (n=27)	splex (n=29)	0.486	0.850	12.719	0.001
18Sv8v9	bsa (n=26)	mplex (n=27)	0.091	0.762	0.005	0.943
18Sv8v9	bsa (n=26)	splex (n=29)	0.410	0.762	3.377	0.105
18Sv8v9	mplex (n=27)	splex (n=29)	0.302	0.762	3.283	0.105
rbcL	bsa (n=15)	mplex (n=16)	0.002	0.968	0.056	0.813
rbcL	bsa (n=15)	splex (n=29)	11.422	0.001	20.194	0.000
rbcL	mplex (n=16)	splex (n=29)	12.819	0.001	21.160	0.000

TABLE 4. Wilcoxon signed rank test on Shannon Diversity and Pielou Evenness calculated between samples successfully amplified with both single plex and multiplex (regular only) assays for the benchmarking samples. P-values are Benjamini & Hochberg corrected.

		Shannon Diversity		Pielou Evenness	
Marker gene	Sample Type	W score	P adj value	W score	P adj value
16Sv4	Marine Coastal	0	1	0	1
	Marine CoastalSed	0	0.109375	4	0.765625
	MarineOffshore	0	1	0	1
	MarineOffshoreSed	0	0.875	1	1
	Peatland	0	1	0	1
	River	0	0.58333333	0	0.875
	Woodland	0	0.58333333	1	1
18Sv1v2	Diatom	0	1	0	1
	Grassland	0	0.375	0	0.5625
	Marine Coastal	0	0.375	2	1
	Marine CoastalSed	0	0.28125	0	0.28125
	MarineOffshore	0	0.64285714	1	1
	MarineOffshoreSed	0	1	0	1
	Peatland	0	0.375	0	0.5625
	River	0	0.375	0	0.5625
	Woodland	0	0.375	1	0.9
18Sv8v9	Diatom	0	1	0	1
	Grassland	0	0.75	1	1
	Marine Coastal	0	0.75	0	1
	Marine CoastalSed	0	0.28125	8	1
	MarineOffshore	1	1	0	1
	MarineOffshoreSed	0	0.9	1	1
	Peatland	3	1	3	1
	River	0	0.9	1	1
	Woodland	2	1	1	1
rbcL	Marine Coastal	0	0.41666667	1	0.625
	Marine CoastalSed	0	0.3125	0	0.3125
	MarineOffshore	0	1	0	1
	MarineOffshoreSed	0	0.625	0	0.625
	River	0	0.41666667	0	0.625

The beta diversity, measured as Bray-Curtis distance, did not significantly differ between single and multiplex across all marker genes (Table 5; Figure S2). There was no significant difference in beta diversity between multiplex assays with and without BSA (Table 5).

TABLE 5 Pairwise PERMANOVA on Bray-Curtis distance between single plex (single) and multiplex (multi, both regular and with addition of BSA) in the benchmarking samples following 999 permutations, with strata set to sample type. Beta diversity was also tested between regular multiplex and multiplex with addition of BSA.

Amplicon	Single/Multi (p val)	Single/Multiplex+BSA (p val)	Multi/Multix+BSA (p val)
16SV4	0.098	0.134	0.998
18SV1V2	0.855	0.877	1.000
18SV8V9	0.996	0.999	1.000
rbcL	0.318	0.238	0.972

The percentage of ASVs identified by both single and multiplex assays ranged between 31.3% in 18Sv1v2 and 14.0% in 16Sv4 when 100% of the features were considered (Figure 3). The performance of the two assays converged when the top 85% most abundant features were included in the analysis (Figure 3). The similarity increased more evidently for the 18S regions (96.6% in 18Sv1v2, 96.5% in 18Sv8v9), whereas similarity remained suboptimal for 16Sv4 and rbcL (55.7% and 71.1% respectively; Figure 3). When the 70% most

abundant features were considered, the ASVs identified by single and multiplex assays overlapped 98.7% in 18Sv1v2, 98.3% in 18Sv8v9, 91.2% in the 16S and 94.1% in the rbcL gene marker (Figure 3). The single and multiplex assays showed similar performance when overlap was studied at species rather than ASV level (Figure S3).

FIGURE 3 Venn diagrams showing ASVs shared between single (blue) and multiplex (red) assays, as well as unique to either assay for the total number of detected features (100%), the top 85% and 70% features.

The rarefaction analysis, aimed at understanding whether a higher sequencing effort would lead to more congruency between single and multiplex assays, showed that both assays had plateaued at the rarefied number of reads used in our analyses (Figure S4), suggesting that a higher sequencing effort would not increase the number of ASVs or species detected.

4.5 DISCUSSION

Holistic approaches that enable the quantification of community-level biodiversity are critical to research and monitoring efforts. Because environmental change affects taxonomic groups differently, ignoring the biotic interactions of a species within its food web can lead to wrong estimation of effects (Fricke *et al.,* 2022; Urban *et al.,* 2016). Only by capturing the response of entire communities to environmental change, can we begin to understand the diagnostic links between environmental drivers and loss of biodiversity (Eastwood *et al.,* 2022; Eastwood, Zhou, *et al.,* 2023; Li *et al.,* 2023; Urban *et al.,* 2016).

Highly multiplexed metabarcoding approaches have the potential to meet the challenge of capturing community-level biodiversity and help identify the causes of biodiversity loss, at comparable efforts and costs than required by single gene markers (e.g. Balint, Nowak, *et al.,* 2018). However, they have technical challenges that may lead to biases in downstream analyses. One of the most common challenges of multiplexing is finding suitable DNA regions and achieving a balanced amplification of all regions, avoiding potential competitive PCR amplifications. It is often challenging to obtain an even amplification

success rate across diverse taxonomic groups (Bohle and Gabaldon, 2012). A step in the right direction are recent efforts that successfully apply early pooling strategies after PCR1 in a 2-step traditional PCR protocol (e.g. application to fish communities (Ushio *et al.,* 2022)). Yet, these strategies are costly because each sample is tagged with a unique string of nucleotides to make the assignment of sequences to samples more robust. We overcame the limitations of combining primers with different length and amplification performance by balancing the concentration of each primer according to its amplification performance, at the same annealing temperature.

A second challenge common to multiplexing individual gene markers is the wrong assignment of reads to samples and barcodes, a phenomenon known as index hopping (MacConaill *et al.,* 2018; Taberlet, A., *et al.,* 2018). We used a paired end strategy with unique 384x384 dual tag barcoding to reduce crosstalk between samples in downstream analyses. Furthermore, we adopted downstream bioinformatics tools to reduce the number of false positives due to index-hopping and PCR and sequencing errors (Bolyen *et al.,* 2019).

A third challenge affecting multiplexing assays Is the lower accuracy in detecting taxa in each DNA sample. Single-plex metabarcoding is expected to have higher accuracy than multiplex metabarcoding, because a single target sequence is included in each reaction. Working with individual gene-markers can reduce the risk of cross-contamination between samples and the error rate introduced during amplification and sequencing of a pool of barcodes (Caroe and Bohmann, 2020). Accuracy and low cross-contamination are particularly important when working with low abundance or endangered species critical for

conservation efforts (Giebner *et al.,* 2020). We showed that alpha diversity measured with Pielou evenness and beta diversity did not significantly differ between single and multiplex assays across different sample types.

However, our results showed significant difference between single and multiplex assays when alpha diversity was measured with the Shannon index. This may be explained by the Pielou evenness index accounting for species relative abundance, as opposed to an overall assessment of richness measured by the Shannon index, and a higher sensitivity of the Shannon index to species abundance (Johnston and Roberts, 2009). This is supported by the non-significant difference between single and multiplex assay for individual sample types successfully amplified in both assays, even if this analysis could not be completed on all paired samples due to amplification failure of some samples. Furthermore, higher sequencing effort of the single plex as compared to multiplex assays would have resulted in a skewed estimate of richness, affecting the Shannon index more pronouncedly than Pielou evenness.

We also showed that the ASVs and species captured by the two assays largely overlapped. This overlap was higher for genes targeting a wide range of taxonomic groups than for taxon-specific genes. This is expected, given the variable performance of the taxon-specific gene rbcL with different sample types. For example, the rbcL performed poorly with samples originating from soil, peatland and woodland, in which freshwater diatoms are not expected. It is possible that aspecific amplification affected single and multiplex differently for this barcode. The congruence between single and multiplex assays improved for all marker genes when rare ASVs/species were excluded from the analyses,

suggesting that rare species were less efficiently captured in multiplex assays. This could be explained by the lower depth of sequencing of the multiplex as compared with single plex assays. However, the rarefaction and resampling approach we used to determine whether higher sequencing depth of the multiplex assay would capture more species, showed that both assays had already plateaued at the depth of sequencing used for the data analysis (rarefied reads). This suggests that a higher sequencing effort alone is not likely to increase the capture of rare ASVs/species in multiplex assays. The likely strategy needed to capture rare species involves a higher eDNA input and/or a higher number of biological replicates in PCR1. In particular, a higher number of replicates has been previously shown to reduce errors and biases, such as the missed amplification of rare species due to preferential amplification of abundant species (Bohmann *et al.,* 2022; Bohmann *et al.,* 2014).

In conclusion, we have shown that multiplexing gene markers in the same reaction improves throughput, reduces costs and enables the amplification of community-level biodiversity with limited input material. This is a significant advancement over previous studies using multiple primers to improve the amplification of longer marker regions (e.g., COI-5P gene (Govender *et al.,* 2022)) and the detection capacity of target taxonomic groups [e.g., zooplankton (Zhang *et al.,* 2018) and fish (Ushio *et al.,* 2018)]. Multiplexing four loci or gene markers at PCR1, combined with sample-specific dual indexes, and pooling of PCR2 libraries, provided significant savings without compromising quality and accuracy and reducing requirements on input DNA. The multiplex optimised for freshwater *seda*DNA performed comparably well on samples extracted from

rivers, marine coastal and marine offshore samples (water and sediment). A potential limitation of our assay was the lower detection of rare species. We suggest that this limitation can be overcome with a higher number of biological replicates or input eDNA. It is noteworthy that the multiplex assay presented here shows a high congruence with single marker genes, especially when targeting a wide range of taxonomic groups, which was the intended use for this tool.

4.6 DATA AVAILABILITY STATEMENT

The metabarcoding sequences generated for this project are available at Biosample ID SAMN36027245 - SAMN36027412 (benchmarking samples) and SAMN36027413 - SAMN36027427 (*seda*DNA samples).

4.7 REFERENCES

Arbizu, P. (2017). pairwiseAdonis: Pairwise multilevel comparison using Adonis.

Balint, M., Nowak, C., Marton, O., Pauls, S. U., Wittwer, C., Aramayo, J. L., Schulze, A., Chambert, T., Cocchiararo, B. Jansen, M. (2018). Accuracy, limitations and cost efficiency of eDNA-based community survey in tropical frogs. *Mol Ecol Resour, 18*(6), 1415-1426. doi:10.1111/1755-0998.12934

Bohle, H. M., & Gabaldon, T. (2012). Selection of marker genes using whole-genome DNA polymorphism analysis. *Evol Bioinform Online, 8*, 161-169. doi:10.4137/EBO.S8989

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

Bohmann, K., Evans, A., Gilbert, M. T., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol, 29*(6), 358-367. doi:10.1016/j.tree.2014.04.003

Bokulich, N. A., Dillon, M. R., Zhang, Y., Rideout, J. R., Bolyen, E., Li, H., Albert, P. S., & Caporaso, J. G. (2018). q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. *MSystems*, **3**(6). doi:10.1128/msystems.00219-18

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-

Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S. 2nd, Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J, Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R. & Caporaso, J.G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol, 37*(8), 852-857. doi:10.1038/s41587-019-0209-9

Bradley, I. M., Pinto, A. J., & Guest, J. S. (2016). Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Appl Environ Microbiol, 82*(19), 5878-5891. doi:10.1128/AEM.01630-16

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods, 13*(7), 581-583. doi:10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011a). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A, 108 Suppl 1*(Suppl 1), 4516-4522. doi:10.1073/pnas.1000080107

Caroe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *Mol Ecol Resour, 20*(6), 1620-1631. doi:10.1111/1755-0998.13227

Chase, M. W., & Fay, M. F. (2009). *Ecology*. Barcoding of plants and fungi. *Science, 325*(5941), 682-683. doi:10.1126/science.1176906

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution, 7*, 1008-1018.

Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annual review of ecology, evolution, and systematics, 49*, 209–230.

Darwall, W., Bremerich, V., De Wever, A., Dell, A. I., Freyhof, J., Gessner, M. O.,Grossart, H., Harrison, I., Irvine, K., Jähnig, S.C., Jeschke, J.M., Lee, J., Lu, C., Lewandowska, A.M., Monaghan, M.T., Nejstgaard, J.C., Patricio, H., Schmidt-Kloiber, A., Stuart, S.N., Thieme, M.L., Tockner, K., Turak, E., & Weyl, O.L. (2018). The Alliance for Freshwater Life: A global call to unite efforts for freshwater biodiversity science and conservation. *Policy, Practice and Standards, 28*, 1015–1022.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z., Knowler, D. J., Leveque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L., & Sullivan, C. A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev Camb Philos Soc, 81*(2), 163-182. doi:10.1017/S1464793105006950

Eastwood, N., Stubbings, W. A., Abou-Elwafa Abdallah, M. A., Durance, I., Paavola, J., Dallimer, M., Pantel, J. H., Johnson, S., Zhou, J., Hosking, J. S., Brown, J. B., Ullah, S., Krause, S., Hannah, D. M., Crawford, S. E., Widmann, M., & Orsini, L. (2022). The Time Machine framework: monitoring and prediction of biodiversity loss. *Trends Ecol Evol, 37*(2), 138-146. doi:10.1016/j.tree.2021.09.008

Eastwood, N., Zhou, J., Derelle, R., Abou-Elwafa Abdallah, M., Stubbings, W. A., Jia, Y., Crawford, S. E., Davidson, T. A., Colbourne, J. K., Creer, S., Bik, H., Hollert, H., & Orsini, L. (2023). 100 years of anthropogenic impact causes changes in freshwater functional biodiversity. *eLife, 12,* RP86576. doi:10.7554/eLife.86576.3

Egeter, B., Craig, H., R., H. L., E., W., & Bakker, J. (2023). Phase 1 Pilot Study Findings & Phase 2 Sampling Plan - Developing Habitat Scale DNA Monitoring in Support of Post 2020 Biodiversity Reporting Requirements. Retrieved from United Kingdom:

Eveleigh, E. S., McCann, K. S., McCarthy, P. C., Pollock, S. J., Lucarotti, C. J., Morin, B., McDougall, G. A., Strongman, D. B., Huber, J. T., Umbanhowar, J., & Faria, L. D. (2007). Fluctuations in density of an outbreak species drive diversity cascades in food webs. *Proc Natl Acad Sci U S A, 104*(43), 16976-16981. doi:10.1073/pnas.0704301104

Ficetola, G. F., & Taberlet, P. (2023). Towards exhaustive community ecology via DNA metabarcoding. *Mol Ecol*. doi:10.1111/mec.16881

Fricke, E. C., Hsieh, C., Middleton, O., Gorczynski, D., Cappello, C. D., Sanisidro, O., Rowan, J., Svenning, J. C., Beaudrot, L., & Beaudrot, L. (2022). Collapse of terrestrial mammal food webs since the Late Pleistocene. *Science, 377*(6609), 1008-1011. doi:10.1126/science.abn4012

Giebner, H., Langen, K., Bourlat, S. J., Kukowka, S., Mayer, C., Astrin, J. J., Misof, B., & Fonseca, V. G. (2020). Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. *Mol Ecol Resour, 20*(5), 1333-1345. doi:10.1111/1755-0998.13201

Gillson, L., & Marchant, R. (2014). From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology. *Trends Ecol Evol, 29*(6), 317-325. doi:10.1016/j.tree.2014.03.010

Govender, A., Singh, S., Groeneveld, J., Pillay, S., & Willows-Munro, S. (2022). Metabarcoding analysis of marine zooplankton confirms the ecological role of a sheltered bight along an exposed continental shelf. *Mol Ecol*. doi:10.1111/mec.16567

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One, 9*(2), e87624. doi:10.1371/journal.pone.0087624

Johnston, E. L., & Roberts, D. A. (2009). Contaminants reduce the richness and evenness of marine communities: a review and meta-analysis. *Environ Pollut, 157*(6), 1745-1752. doi:10.1016/j.envpol.2009.02.017

Kelly, M.G., Boonham, N., Juggins, S., Kille, P., Mann, D.G., Pass, D., Sapp, M., Sato, S., Glover, R. (2018). A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. Science Report SC140024/R, Environment Agency, Bristol. Available at: https://www.gov.uk/government/publications/a-dna-based-metabarcoding-approach-to-assess-diatom-communities-in-rivers

Krehenwinkel, H., Kennedy, S. R., Rueda, A., Lam, A., & Gillespie, R. G. (2018). Scaling up DNA barcoding – Primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods in Ecology and Evolution, 9*, 2181–2193.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks inone-criterion variance analysis. *Journal of the American Statistical Association, 47*, 583–621.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool, 10*, 34. doi:10.1186/1742-9994-10-34

Li, F., Qin, S., Wang, Z., Zhang, Y., & Yang, Z. (2023). Environmental DNA metabarcoding reveals the impact of different land use on multitrophic biodiversity in riverine systems. *Sci Total Environ, 855*, 158958. doi:10.1016/j.scitotenv.2022.158958

MacConaill, L. E., Burns, R. T., Nag, A., Coleman, H. A., Slevin, M. K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M. S., Ducar, M. D., Meyerson, M.,

& Thorner, A. R. (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics, 19*(1), 30. doi:10.1186/s12864-017-4428-5

Martin, M. (2023). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet journal, 17*, 10-12.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R Soc Open Sci, 2*(7), 150088. doi:10.1098/rsos.150088

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szöcs, E., & Wagner, H. (2020). *vegan: Community ecology package*. https://CRAN.R-project.org/package=vegan

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramalingam, N., Warkiani, M. E., & Gong, T. H. (2017). Acetylated bovine serum albumin differentially inhibits polymerase chain reaction in microdevices. *Biomicrofluidics, 11*(3), 034110. doi:10.1063/1.4983692

Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., Mann, D. G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J.,

& Bouchez, A. (2019). Diat.barcode, an open-access curated barcode library for diatoms. *Sci Rep, 9*(1), 15116. doi:10.1038/s41598-019-51500-6

Ropert-Coudert, Y., & Wilson, R. P. (2005). Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment, 3*, 437–444.

Ruckelshaus, M. H., Jackson, S. T., Mooney, H. A., Jacobs, K. L., Kassam, K. S., Arroyo, M. T. K., Báldi, A., Bartuska, A. M., Boyd, J., Joppa, L. N., Kovács-Hostyánszki, A., Parsons, J. P., Scholes, R. J., Shogren, J. F., & Ouyang, Z. (2020). The IPBES Global Assessment: Pathways to Action. *Trends Ecol Evol, 35*(5), 407-414. doi:10.1016/j.tree.2020.01.009

Schmidt, P. A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., & Schmitt, I. (2013). Illumina metabarcoding of a soil fungal community. *Soil Biology & Biochemistry, 65*, 128–132.

Spens, j., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., Sigsaard, E.E, & Hellstrom, M. (2017). Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution 8*, 635–645.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, É. (2018). *Environmental DNA: for biodiversity research and monitoring*. Oxford, UK: Oxford University Press.

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., & Jones, P. D. (2018). Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological

Techniques To Infer Historical Trends in Cyanobacterial Communities. *Environ Sci Technol, 52*(12), 6842-6853. doi:10.1021/acs.est.7b06386

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J. B., Pe'er, G., Singer, A., Bridle, J. R., Crozier, L. G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J. J., Holt, R. D., Huth, A., Johst, K., Krug, C. B., Leadley, P. W., Palmer, S. C., Pantel, J. H., Schmitz, A., Zollner, P.A., & Travis, J. M. (2016). Improving the forecast for biodiversity under climate change. *Science, 353*(6304). doi:10.1126/science.aad8466

Ushio, M., Furukawa, S., Murakami, H., Masuda, R., & Nagano, A. J. (2022). An efficient early-pooling protocol for environmental DNA metabarcoding. *Environmental DNA, 4*, 1212–1228.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T., & Kondoh, M. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding and Metagenomics, 2*, e23297.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., chweer, T., Peplies, J., Ludwig, W., & Glockner, F. O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res, 42*(Database issue), D643-648. doi:10.1093/nar/gkt1209

Yoccoz, N. G. (2012). The future of environmental DNA in ecology. *Mol Ecol, 21*(8), 2031-2038. doi:10.1111/j.1365-294X.2012.05505.x

Zhan, A., Bailey, S. A., Heath, D. D., & Macisaac, H. J. (2014). Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Mol Ecol Resour, 14*(5), 1049-1059. doi:10.1111/1755-0998.12254

Zhang, G. K., Chain, F. J. J., Abbott, C. L., & Cristescu, M. E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evol Appl, 11*(10), 1901-1914. doi:10.1111/eva.12694

4.8 SUPPORTING INFORMATION

TABLE S1. PCR1 primers, bibliographic references and sequences. Marker gene sequence in black, universal overhang in blue and adapter tag for PCR2 primer in green. Note: "|" differentiates parts of the sequence for clarity but does not indicate a separation in the nucleotide molecule.

Marker gene	Reference	Primer direction	Sequence (5'-3'): adapter tag|universal overhang|primer
18SV1 V2	Hadziavdic et al., 2014	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|GCTTGTCTCAAAGATTAAGCC
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GCCTGCTGCCTTCCTTGGA
18SV8 V9	Bradley, Pinto, and Guest, 2016	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|ATAACAGGTCTGTGATGCCCT
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|CCTTCYGCAGGTTCACCTAC
16SV4	Caporaso et al., 2011a	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|GTGCCAGCMGCCGCGGTAA
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GGACTACHVGGGTWCTAAT
rbcL	Kelly et al., 2018	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|ATGCGTTGGAGAGARCGTTTC
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GATCACCTTCTAATTTACCWACAACTG

TABLE S2. PCR2 primers consisting of nextera adapter tag, sample index and priming sequence. *Note:* From Eastwood, Zhou *et al.,* 2023

Name	i5 nextera adapter tag, sample index and priming sequence	I7 nextera adapter tag, sample index and priming sequence
NEXT_TSHT-1	AATGATACGGCGACCACCGAGATCTACACGACACAGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGACACAGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-2	AATGATACGGCGACCACCGAGATCTACACGCATAACGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCATAACGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-3	AATGATACGGCGACCACCGAGATCTACACACAGAGGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACAGAGGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-4	AATGATACGGCGACCACCGAGATCTACACCCACTAAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCACTAAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-5	AATGATACGGCGACCACCGAGATCTACACTGTTCCGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGTTCCGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-6	AATGATACGGCGACCACCGAGATCTACACGATACCTGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGATACCTGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-7	AATGATACGGCGACCACCGAGATCTACACAGCCGTAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCCGTAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-8	AATGATACGGCGACCACCGAGATCTACACCTCCTGAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTCCTGAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-9	AATGATACGGCGACCACCGAGATCTACACACGAATCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACGAATCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-10	AATGATACGGCGACCACCGAGATCTACACAATGGTCGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAATGGTCGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-11	AATGATACGGCGACCACCGAGATCTACACCGCTACATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCGCTACATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-12	AATGATACGGCGACCACCGAGATCTACACCCTAAGTCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTAAGTCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-13	AATGATACGGCGACCACCGAGATCTACACTTGCTTGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTGCTTGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-14	AATGATACGGCGACCACCGAGATCTACACCCTGTCAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTGTCAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-15	AATGATACGGCGACCACCGAGATCTACACAGCCTATCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCCTATCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-16	AATGATACGGCGACCACCGAGATCTACACTGATCACGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGATCACGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-17	AATGATACGGCGACCACCGAGATCTACACCCACATTGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCACATTGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-18	AATGATACGGCGACCACCGAGATCTACACTCGAGAGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCGAGAGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-19	AATGATACGGCGACCACCGAGATCTACACGGTCGTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGGTCGTATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-20	AATGATACGGCGACCACCGAGATCTACACACAGGCATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACAGGCATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-21	AATGATACGGCGACCACCGAGATCTACACGTGATCCAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTGATCCAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-22	AATGATACGGCGACCACCGAGATCTACACTTCGTACGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCGTACGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-23	AATGATACGGCGACCACCGAGATCTACACATGACAGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATGACAGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-24	AATGATACGGCGACCACCGAGATCTACACCGACCTAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCGACCTAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-25	AATGATACGGCGACCACCGAGATCTACACTATGGCACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTATGGCACGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-26	AATGATACGGCGACCACCGAGATCTACACATAACGCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATAACGCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-27	AATGATACGGCGACCACCGAGATCTACACGTAGTACCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGTACCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-28	AATGATACGGCGACCACCGAGATCTACACCGCGTATTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCGCGTATTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-29	AATGATACGGCGACCACCGAGATCTACACATCCACGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATCCACGAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-30	AATGATACGGCGACCACCGAGATCTACACTAACGTCGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTAACGTCGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-31	AATGATACGGCGACCACCGAGATCTACACCCTTCCATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTTCCATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-32	AATGATACGGCGACCACCGAGATCTACACGATCAAGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGATCAAGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-33	AATGATACGGCGACCACCGAGATCTACACAAGCATCGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAAGCATCGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-34	AATGATACGGCGACCACCGAGATCTACACAGGATAGCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGATAGCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-35	AATGATACGGCGACCACCGAGATCTACACGGCTCAATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGGCTCAATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-36	AATGATACGGCGACCACCGAGATCTACACTTCACGGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCACGGAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-37	AATGATACGGCGACCACCGAGATCTACACGGCGAATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGGCGAATAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-38	AATGATACGGCGACCACCGAGATCTACACAAGTGCAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAAGTGCAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-39	AATGATACGGCGACCACCGAGATCTACACGCAATTCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCAATTCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-40	AATGATACGGCGACCACCGAGATCTACACCTTCGCAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTTCGCAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-41	AATGATACGGCGACCACCGAGATCTACACCATTGACGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATTGACGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-42	AATGATACGGCGACCACCGAGATCTACACTCTGGACAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCTGGACAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-43	AATGATACGGCGACCACCGAGATCTACACGCTACAACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTACAACGTGACTGGAGTTCAGACGTGTGCTC

NEXT_TSHT-44	AATGATACGGCGACCACCGAGATCTACACCCGATGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCGATGTAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-45	AATGATACGGCGACCACCGAGATCTACACTAGGAGCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTAGGAGCTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-46	AATGATACGGCGACCACCGAGATCTACACAACAAGGCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAACAAGGCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-47	AATGATACGGCGACCACCGAGATCTACACCTCGGTAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTCGGTAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-48	AATGATACGGCGACCACCGAGATCTACACAGCTTCAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-49	AATGATACGGCGACCACCGAGATCTACACTCACCTAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCACCTAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-50	AATGATACGGCGACCACCGAGATCTACACCAAGTCGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAAGTCGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-51	AATGATACGGCGACCACCGAGATCTACACCTGTATGCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTGTATGCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-52	AATGATACGGCGACCACCGAGATCTACACAGTTCGCAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGTTCGCAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-53	AATGATACGGCGACCACCGAGATCTACACATCGGAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATCGGAGAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-54	AATGATACGGCGACCACCGAGATCTACACAAGTCCTCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAAGTCCTCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-55	AATGATACGGCGACCACCGAGATCTACACTGGATGGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGGATGGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-56	AATGATACGGCGACCACCGAGATCTACACAGGTGTTGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGTGTTGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-57	AATGATACGGCGACCACCGAGATCTACACGACGAACTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGACGAACTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-58	AATGATACGGCGACCACCGAGATCTACACGTTCTTCGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTTCTTCGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-59	AATGATACGGCGACCACCGAGATCTACACTTCGCCATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCGCCATGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-60	AATGATACGGCGACCACCGAGATCTACACCAACTCCAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAACTCCAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-61	AATGATACGGCGACCACCGAGATCTACACAACCGTGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAACCGTGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_Tag261	AATGATACGGCGACCACCGAGATCTACACCGGTTGTTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCGGTTGTTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-63	AATGATACGGCGACCACCGAGATCTACACCTAGCAGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGCAGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-64	AATGATACGGCGACCACCGAGATCTACACACCTCTTCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACCTCTTCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-65	AATGATACGGCGACCACCGAGATCTACACTACTAGCGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTACTAGCGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-66	AATGATACGGCGACCACCGAGATCTACACACAACAGCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACAACAGCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-67	AATGATACGGCGACCACCGAGATCTACACCGCAATGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCGCAATGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-68	AATGATACGGCGACCACCGAGATCTACACCAGTGCTTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGTGCTTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-69	AATGATACGGCGACCACCGAGATCTACACTCTAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCTAGGAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-70	AATGATACGGCGACCACCGAGATCTACACGATTGTCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGATTGTCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-71	AATGATACGGCGACCACCGAGATCTACACGGTACGAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGGTACGAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-72	AATGATACGGCGACCACCGAGATCTACACCTTCACTGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTTCACTGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-73	AATGATACGGCGACCACCGAGATCTACACATAGGTCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATAGGTCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-74	AATGATACGGCGACCACCGAGATCTACACACCGACAAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACCGACAAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-75	AATGATACGGCGACCACCGAGATCTACACAACACTGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAACACTGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-76	AATGATACGGCGACCACCGAGATCTACACACCATAGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACCATAGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-77	AATGATACGGCGACCACCGAGATCTACACTCGATGACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCGATGACGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-78	AATGATACGGCGACCACCGAGATCTACACGACTTGTGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGACTTGTGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-79	AATGATACGGCGACCACCGAGATCTACACCCGTTATGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCGTTATGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-80	AATGATACGGCGACCACCGAGATCTACACCAACGAGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAACGAGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-81	AATGATACGGCGACCACCGAGATCTACACTTACCGACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTACCGACGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-82	AATGATACGGCGACCACCGAGATCTACACGAGAGTACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGAGAGTACGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-83	AATGATACGGCGACCACCGAGATCTACACCTGTACCAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTGTACCAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-84	AATGATACGGCGACCACCGAGATCTACACTGAGCTGTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGAGCTGTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-85	AATGATACGGCGACCACCGAGATCTACACAGTATGCCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGTATGCCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-86	AATGATACGGCGACCACCGAGATCTACACTACTGCTCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTACTGCTCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-87	AATGATACGGCGACCACCGAGATCTACACTGCTTGCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCTTGCTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-88	AATGATACGGCGACCACCGAGATCTACACTCCACGTTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCACGTTGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-89	AATGATACGGCGACCACCGAGATCTACACAACCAGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAACCAGAGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-90	AATGATACGGCGACCACCGAGATCTACACACGAACGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATACGAACGAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-91	AATGATACGGCGACCACCGAGATCTACACATAGTCGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATATAGTCGGGTGACTGGAGTTCAGACGTGTGCTC

NEXT_TSHT-92	AATGATACGGCGACCACCGAGATCTACACCCATGAACACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCATGAACGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-93	AATGATACGGCGACCACCGAGATCTACACGAGCAATCACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGAGCAATCGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-94	AATGATACGGCGACCACCGAGATCTACACCAACTTGGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAACTTGGGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-95	AATGATACGGCGACCACCGAGATCTACACCCACAACAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCACAACAGTGACTGGAGTTCAGACGTGTGCTC
NEXT_TSHT-96	AATGATACGGCGACCACCGAGATCTACACTGGTGAAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGGTGAAGGTGACTGGAGTTCAGACGTGTGCTC

TABLE S3. Samples used for the benchmarking of the multiplex assay provided by NatureMetrics. Sample ID, sample name, and concentration (ng/μL) are provided.

Sample ID	Sample Name	Conc ng/ μL
18645	Diatom1	9.16
12607	Grassland1	116
12608	Grassland2	83.6
12609	Grassland3	54.8
18319	Marine Coastal1	1.3
18320	Marine Coastal2	1.21
14308	Marine Coastal3	21.4
11608	Marine CoastalSed1	11.1
11069	Marine CoastalSed2	44.8
11070	Marine CoastalSed3	18.1
11071	Marine CoastalSed4	15.7
17129	Marine CoastalSed5	6.28
17130	Marine CoastalSed6	5.66
17131	MarineCoastalSed7	4.12
6256	MarineOffshore1	47.8
6260	MarineOffshore2	58
15720	MarineOffshoreSed1	15.1
15723	MarineOffshoreSed2	17.6
8678	Peatland1	39.4
8681	Peatland2	16.5
8686	Peatland3	4.84
7784	River1	40.2
7791	River2	26.2
14084	River3	20.2
12606	Woodland	83
12604	Woodland1	49.8
12605	Woodland2	135

FIGURE S1. Tapestation traces. PCR1 products from multiplex and single gene markers are visualised on tapestation traces. A PCR2 trace is also shown. In the multiplex trace multiple gene markers of similar length overlap and are indistinguishable. Single marker gene amplifications on the PCR1 multiplex show that all four marker genes are present in the reaction. Note different scales on the y-axis. Multi - multiplex. The individual gene markers are: 16SV4, 18SV1V2, 18SV8V9, and rbcL.

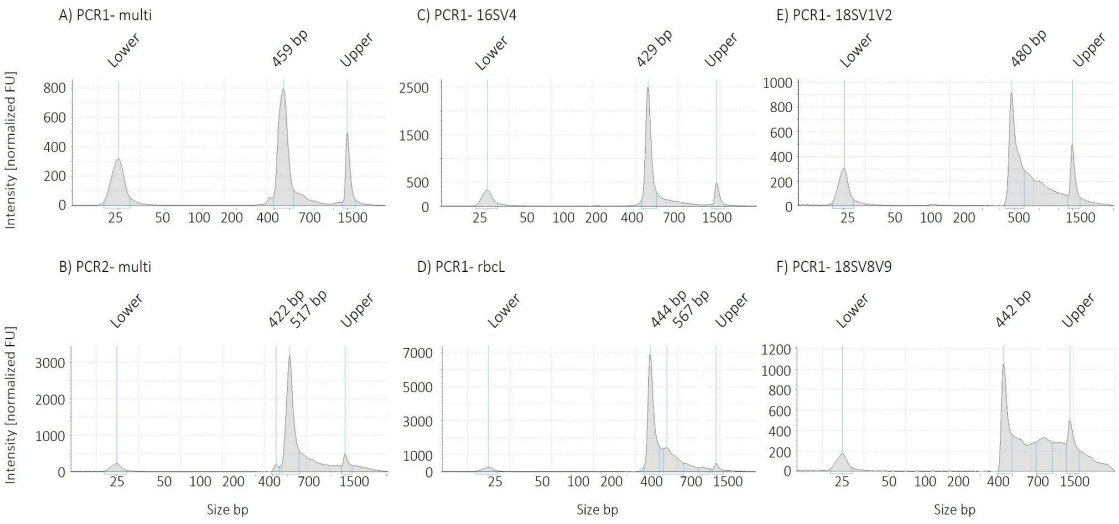


FIGURE S2 Principal Coordinate Analysis (PCoA) visualisation of Bray Curtis distance among benchmarking samples obtained with single and multiplex assays. Axes are labelled with percentage of variance explained. Single plexes are shown with triangles, multiplexes shown with circles; colour denotes sample type.

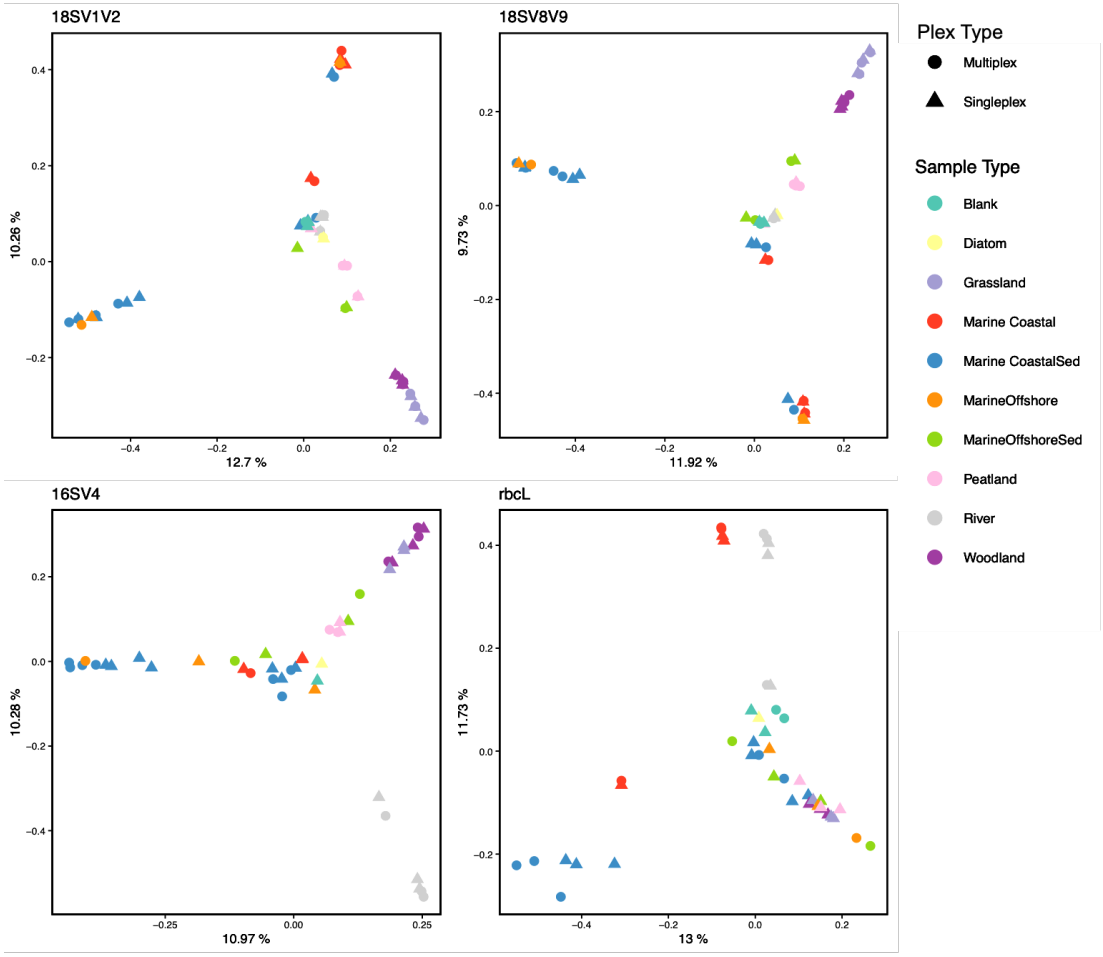


FIGURE S3 Venn diagrams showing overlap of species detected with single (blue) and multiplex (red) assays for the total number of detected features assigned to species level (100%), the top 85% and 70% features in the dataset.

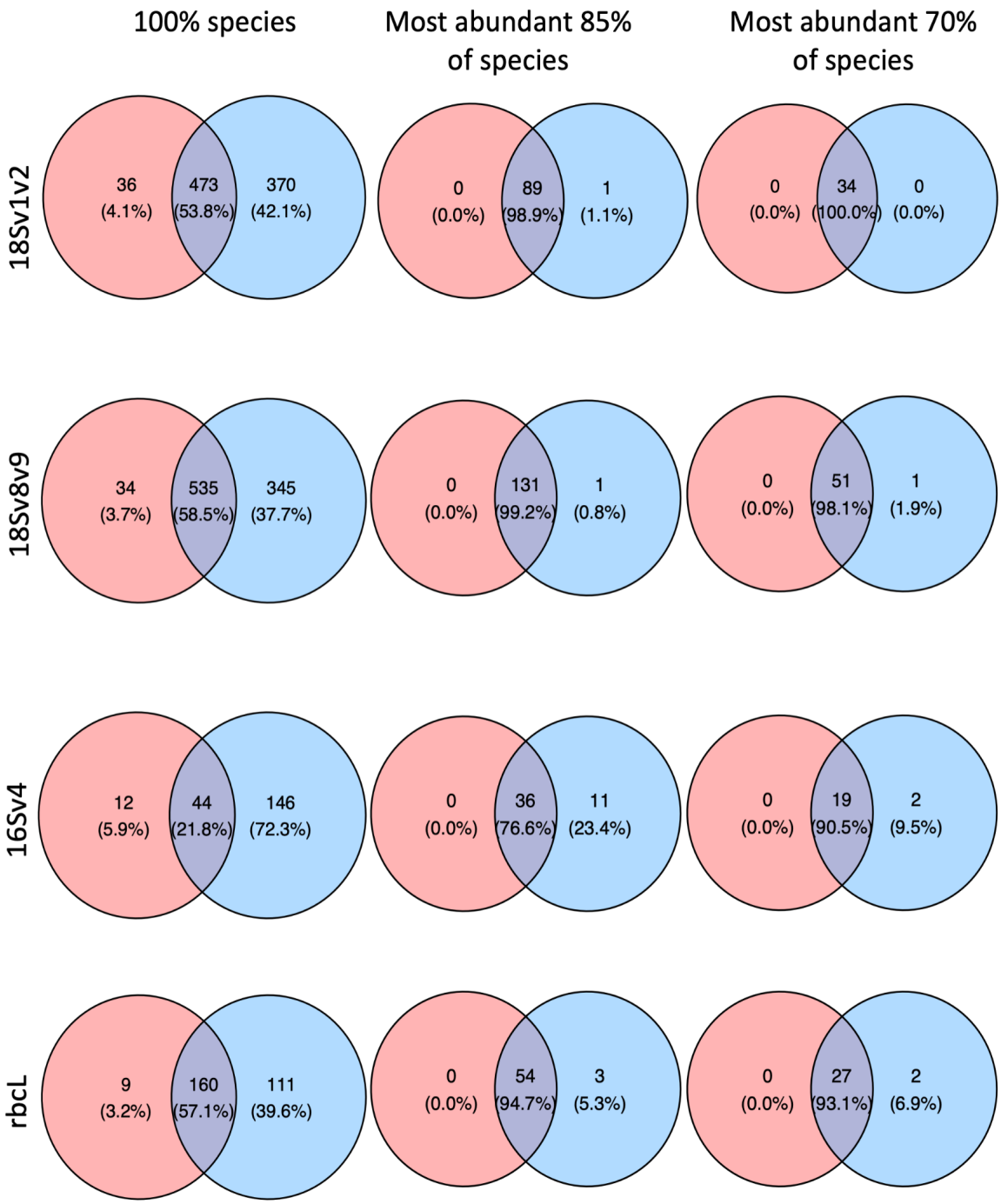
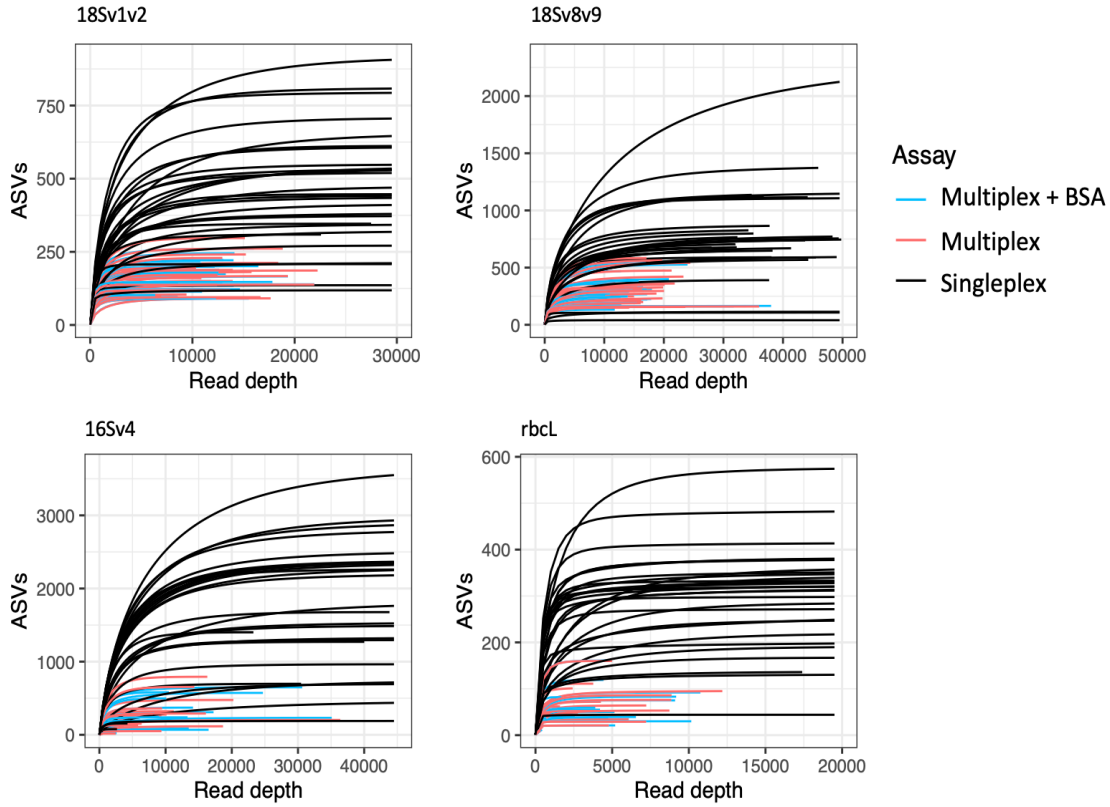


FIGURE S4 Rarefaction curves with resampling strategy on the benchmarking samples provided by NatureMetrics. Single plexes are in black, multiplex data are displayed in red and multiplex plus BSA are displayed in blue. The read depths displayed on the x-axis are limited to a maximum value (18Sv1v2: 30000 reads; 18Sv8v9: 50000 reads; 16Sv4: 45000 reads; rbcL: 20000 reads) to enhance the visualisation of multiplex assays which were run at a lower depth of sequencing.



Chapter 5: SPATIAL FRESHWATER BIODIVERSITY DYNAMICS ACROSS ENGLISH LAKES

Niamh Eastwood[1*], Arron Watson[1,2], Kerry Walsh[2], Jo-Anne Pitt[2], Jonathan Warren[2], Jiarui Zhou[1#], Luisa Orsini[1, 3,4#]

[1]Environmental Genomics Group, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

[2]Environment Agency, PO Box 544, Rotherham, S60 1BY

[3]Institute for Interdisciplinary Data Science and AI, the University of Birmingham, Birmingham, B15 2TT, UK

[4]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB

*First Author

#shared senior authorship

Author Contributions NE carried out data analysis. KW, JAP and JW provided material and data. NE, LO and KW designed the study. AW and JZ performed exploratory analyses (not shown). NE drafted the manuscript. NE LO JZ and AW contributed to manuscript writing.

Disclaimer: the opinions, findings, and conclusion or recommendations expressed in the paper are those of the authors and do not necessarily reflect the view of the Environment Agency

5.1 INTRODUCTION

In the last three decades, biodiversity has declined at a pace never observed before, driven by climate change, habitat fragmentation and chemical pollution (Dirzo *et al.,* 2014, Naggs, 2017 #367) The rapid and critical loss of biodiversity, especially among invertebrates, has been called the 'sixth mass extinction' (Naggs 2017). Loss of biodiversity has caused an unprecedented loss of ecosystem services, including clean water, food provision and climate regulation (Bonebrake *et al.,* 2019; Ruckelshaus *et al.,* 2020).

Freshwater lakes cover less than 1% of Earth surface; yet they are hotspots of biodiversity, supporting up to 10% of species on Earth and providing key ecosystem services (Durance *et al.,* 2016; Mace, Norris, and Fitter, 2012). These ecosystems are of high conservation value, but also among the most threatened by environmental change globally, having experienced 83% of biodiversity loss since 1970 (Acreman *et al.,* 2020; WWF, 2018). Decline in freshwater biodiversity has been growing despite actions to protect these ecosystems e.g. with the establishment of the Aichi Targets (https://www.cbd.int/sp/targets/). For example, climate warming combined with anthropogenic nutrient inputs in shallow lakes has been shown to promote more frequent algal blooms and oxygen depletion in lakes with direct impact on the freshwater trophic chain (Foley *et al.,* 2005). Higher vulnerability of freshwater lakes is partly explained by their landscape position as 'receivers' of land-use effluents, and their role in supporting a high number of endemic species, leading to non-substitutability. Moreover, freshwater is a commodity for many stakeholders, increasing the challenges to balance ecological and economic

priorities (Dudgeon *et al.,* 2006; Eastwood *et al.,* 2022). Despite their socio-economic value, freshwater lakes are also the least featured in conservation actions (Ahmed *et al.,* 2022). Moreover, methods used to assess impact on these ecosystems are outdated and inadequate such as use of single species for water quality assessment (Bardey, 2020).

Traditionally, biodiversity monitoring has been based on direct observations, remote sensing and mark recapture techniques (e.g. Eveleigh *et al.,* 2007; Ropert-Coudert and Wilson, 2005). Whereas these approaches have provided important insights into species ecology and persistence in the face of environmental change, they have critical limitations. Cryptic diversity cannot easily be resolved through morphological observations; life stages of the same species with different morphology can be wrongly classified; species assignment relies on visible remains in environmental matrices; species assignment requires specialist taxonomic skills; and the monitoring of biodiversity at scale is low throughput (Gillson and Marchant, 2014). Indicator species are widely used in Europe to establish the ecological status of freshwater lakes under the Water Framework Directive (WFD) (EU, 2000). However, the poor status of the large majority of European lakes may suggest that the use of indicator species to determine water quality is inadequate. The WFD methods are too variable across ecosystems, requires taxonomic specialist skills and fail to link species dynamics to drivers of biodiversity loss (Birk *et al.,* 2012).

In the last decade, DNA-based methods have become more prominent for large-scale biodiversity monitoring (Balint, Nowak, *et al.,* 2018), moving from

morphological to metabarcoding methods for water quality assessments (Zimmermann *et al.,* 2014) adopted by regulatory agencies. Their main advantage of these methods is the high throughput screening from any environmental matrix without being limited to taxonomic groups with well-preserved remains (Creer *et al.,* 2016; Cristescu and Hebert, 2018). Moreover, large-scale monitoring can be achieved in a fast, cost-effective and non-invasive manner (Taberlet, Bonin, *et al.,* 2018a). DNA-based methods do not require specialist skills and can resolve cryptic genetic diversity by matching sequence similarity to records in public databases (e.g., NCBI and SILVA). These approaches have been used to estimate human impact on biodiversity (Hofman *et al.,* 2015), the severity of alien species invasion (Ruppert, Kline, and Md Rahman, 2019) and more generally species richness and distribution at landscape level (Garlapati *et al.,* 2019). More recently, they have been applied to discover long-term patterns of biodiversity change (Eastwood, Zhou *et al.,* 2023).

As eDNA-based approaches move from single-marker to multi-marker approaches (Bohmann *et al.,* 2022; Eastwood *et al.,* 2024), eDNA can become the method of choice to study biodiversity dynamics across space, time and environmental gradients without the need of specialist taxonomic skills and evident organism's remains (Eastwood *et al.,* 2022; Eastwood, Zhou *et al.,* 2023; Fordham *et al.,* 2020). Although technical challenges associated with eDNA degradation and analytical limitations associated with incomplete reference databases exist, multiple markers eDNA metabarcoding can significantly improve our ability to explore ecosystem-level processes and

community response to environmental change (Balint, Pfenninger, *et al.,* 2018; Eastwood *et al.,* 2024; Blackman *et al.,* 2022).

Here, we apply a recently developed metabarcoding multiplex approach (Eastwood *et al.,* 2024) to quantify community-level biodiversity in 52 freshwater lakes included in the Water Framework Directive (WFD) monitoring program in England. These lakes range from high to bad quality according to the WFD ecological status classification (EU 2000). In this study, we aim to assess freshwater lake community biodiversity at national scale (England) using environmental DNA and to investigate the comparability of the community detected using different sample types (water or biofilm). We aim to understand whether differences in community composition across lakes can be explained by the geographic region of the lake. Finally, we aim to identify whether regulatory lake quality classifications explain differences in whole-community composition across lakes as captured with eDNA.

Our work reveals current limitations in monitoring frameworks and that holistic data-driven approaches may be better suited to explain dynamics of lake freshwater biodiversity across impacted landscapes.

5.2 MATERIALS AND METHODS

5.2.1 Study system

A total of 52 lake sites across the UK were included in this study (Fig. 1; Table S1) Water samples were collected from 40 of these sites and biofilm samples were collected from 42 sites. Hence, both water and biofilm samples were collected from 30 lakes, allowing a comparative analysis of benthic and pelagic communities in these lakes. These lakes range from lowland lakes to higher

altitude lakes. By sampling both benthic and pelagic communities, we were able to make a comprehensive assessment of biodiversity in freshwater lakes.

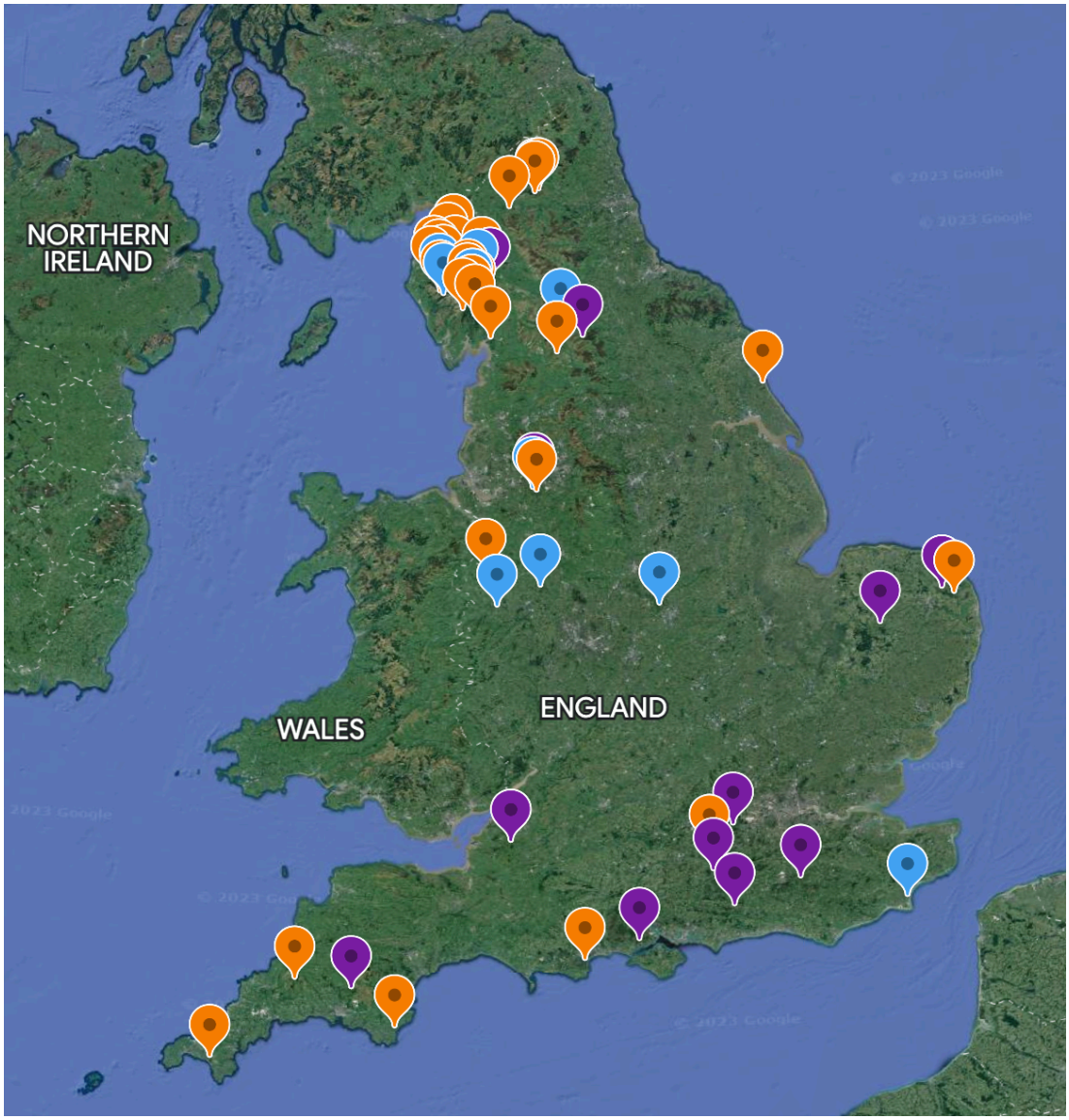


Figure 1 Map of lake sites across England in the study. Marker colour indicates sample type(s) at the site: blue – water samples, purple – biofilm samples, orange – both water and biofilm samples.

The water samples were collected following the UKTAG Environmental Standards for environmental samples (Wilby *et al.,* 2019) between March 2017-November 2019 (supplementary table 1). In brief, samples were collected from

20 shallow locations (>50 cm) around each lake perimeter at 5m intervals away from inflow hotspots. Typically, two litres of water were collected from each site and filtered on a 0.8µm filter within 24 h from collection. A field blank sample of sterile water was collected at each site. Environmental DNA (eDNA) was extracted from the filters using the Power Water DNA isolation kit (MoBio) by the Environment Agency (England) as part of their routine monitoring for lakes in the Water Framework Directive monitoring programme.

Biofilm samples were collected following standard Environment Agency sampling techniques for benthic organisms in September- November 2014 and September- November 2016 (supplementary table 1). This involves placing 5 cobbles in a tray with about 50 ml of nearshore water and then brushing the upper surface of each cobble with a toothbrush to remove the biofilm following Kelly *et al.,* (2018). After collection, the samples were transferred to the laboratory where a suspension of biofilm and water was transferred to a sterile 15ml centrifuge tube containing 5ml nucleic acid preservative (3.5 M ammonium sulfate, 17 mM sodium citrate and 13 mM EDTA) and stored frozen (-20°C). DNA was extracted at the University of Birmingham using DNeasy power biofilm kit (Qiagen), following the manufacturer's instructions.

5.2.2 Biodiversity fingerprinting

We recently developed a 2-PCR protocol that multiplexes four loci in PCR 1 and up to 384 samples per barcode (N=1,536 samples) in a single sequencing run (described in chapter 3 of this thesis; Eastwood *et al.,* 2024). The metabarcoding loci used in the multiplex were: two regions targeting eukaryotes

broadly [(18SV1V2) (Hadziavdic *et al.,* 2014) and (18SV8V9)(Bradley, Pinto, and Guest 2016)], and prokaryotes (16SV4) (Caporaso *et al.,* 2011b), plus a taxon-specific marker targeting diatoms (rbcL)(Kelly *et al.,* 2018) (supplementary table 2). Each biological sample (water or biofilm sample) was run in duplicate and handled as an independent sample. A selection of field blank samples were amplified and showed amplification when tested on agarose gels (not shown) so all field water blank samples were pooled to create a single 'field blank pool' as running each field blank independently would exceed budget. There was no equivalent field blank samples collected for the biofilm samples. A library construction blank (consisting of ultrapure water instead of template DNA) was created for the biofilm library construction and the water library construction. Each independent sample was amplified in triplicate in PCR1 using Multiplex PCR Mastermix (Qiagen). After removing excess primers with High Prep PCR magnetic beads (Auto Q Biosciences), cleaned PCR1 products were pooled in a second PCR in using unique dual-indexed sequencing adapters. Unique dual barcodes (indices) were used to reduce index-misassignment and index-hopping between samples (MacConaill *et al.,* 2018). PCR2 amplicons were purified using High Prep PCR magnetic beads (Auto Q Biosciences) and quantitated using a 200 pro plate reader (TECAN) using qubit dsDNA HS solution (Invitrogen). PCR conditions and quantities are not disclosed here to protect comerically sensitive information. A standard curve was created by running standards of known concentration on each plate against which sample concentration was determined. PCR2 libraries were mixed in equimolar quantities (at a final concentration of 12 pmol) using a biomek FXp

liquid handling robot (Beckman Coulter). The final molarity of the pools was confirmed using a HS D1000 tapestation screentape (Agilent) prior to 250 bp paired-end sequencing on an DNBSeq G400 following the conversion of linear dsDNAinto circularized libraries compatible for the DNBSeq platform using the MGIEasy Universal Library Conversion Kit.

5.2.3 Lake psycho-chemical descriptors

The Environment Agency classifies lakes under the Water Framework Directive from high to bad under a hierarchical scheme – overall status consists of ecological and chemical status, which are in turn comprised of several elements (ecological: biological quality elements, hydromorphic elements, physicochemical quality elements and specific pollutants; chemical s priority substances, priority hazardous substances and other pollutants) (UK Technical Advisory Group on the Water Framework Directive, 2007). Overall ecological scores were taken from WFD cycle 2 (years 2015-2019).

5.2.4 Bioinformatics

The sequenced reads were demultiplexed using cutadapt v4.1 (Martin, 2023), and analyzed with QIIME2 v2022.8 (Bolyen *et al.,* 2019). Trimming, filtering, merging and denoising of reads was done using the QIIME2 DADA2 module (Callahan *et al.,* 2016) with default parameters and trimming and truncation lengths as in Table 1. Taxonomy assignment was done with the QIIME2 feature-classifier module with naive-bayes taxonomic classifiers trained on the SILVA v138 database for the assignment of the 16S and 18S reads (Yilmaz *et al.,* 2014); and the diat.barcode v9.2 for the rbcL reads (Rimet *et al.,* 2019). The

taxonomic barplots were plotted per barcode per sample type using ggplot2 v3.3.5 (Caporaso *et al.,* 2011b) in R v4.0.2 (R Core Team, 2020) and including the top ten most abundant families (or lowest taxonomic level if not assigned to family). All other taxa were collapsed in the plots under 'other taxa'.

Barcode	Front trimmed forward	Front trimmed reverse	Truncated length forward	Truncated length reverse
16SV4	0	21	219	231
18SV1	0	20	224	238
18SV8	11	22	210	228
rbcL	2	28	223	230

Table 1 Trimming and truncation length parameters used in DADA2 analysis.

Initial comparisons of sample type was carried out on the cleaned reads in a combined dataset of water and biofilm samples, rarefied at an the same depth in both samples types per barcode, to the sample with the lowest read depth or 1000 reads, or whichever was higher: 16SV1 – 1000; 18SV1 – 14029, 18SV8 – 17201, rbcL – 1000. Visual inspection of comparability was carried out using a PCoA plot of beta diversity, as measured with weighted unifrac distance, calculated with qiime and plotted in ggplot v3.4.0 (Wickham, 2016) in R v4.0.2 (R Core Team, 2020).

Further analysis was carried out on the cleaned reads separated into water or biofilm sample types and rarefied to the sample with the lowest read depth or 1000 reads, whichever was higher, before downstream analyses. Rarefaction depths were as follows: Water samples: 16SV1 – 1000, 18SV1 - 18131, 18SV8 – 17201, rbcL – 1000; Biofilm samples: 16SV4 – 1000, 18SV1 – 14029, 18SV8 – 21734, rbcL – 1526. Diversity indices were measured separately on the water and biofilm samples. Alpha diversity was measured as Shannon diversity,

supported by a Kruskal-Wallis statics test (Kruskal and Wallis 1952). Beta diversity was measured as weighted unifrac distance and significant differences between lakes and categories were assessed with a PERMANOVA test (999 permutations) using qiime diversity adonis and the formula Ecological quality * Region.

5.3 RESULTS

5.3.1 Biodiversity fingerprinting

Six biofilm samples failed denoising in every barcode (Site 2, sample B; Site 11, sample B; Site 29, sample B; Site 40, sample B; Site 49, sample B; Site 50, sample B), no water samples failed denoising. Results of denoising are shown in table 2.

	Biofilm samples			Water Samples		
Barcode	Total number of reads	Median number of reads per sample	Number of ASVs	Total number of reads	Median number of reads per sample	Number of ASVs
16SV4	455,691	2,450	8,683	2,775,364	17,736	7,014
18SV1	4,670,782	57,614	7,138	4,132,951	46,891	5,458
18SV8	5,780,412	70,847	15,206	3,699,070	38,498	11,232
rbcL	3,543,100	33,186	2,127	2,153,295	20,429	1,447

Table 2. Denoising output statistics from qiime DADA2 (total number of reads across all samples, median number of reads per sample and number of unique ASVs across all samples) per barcode for biofilm samples (n = 85) and water samples (n = 83).

Samples were separated strongly by sample type (biofilm or water; figure 2) in the 18SV1 and 18Sv8 barcodes and to a lesser extent in the 16SV4, with no clear grouping effect of ecological quality score (supplementary figure 1), as assessed with weighted unifrac distance. Therefore, samples were separated by sample type for downstream analysis.

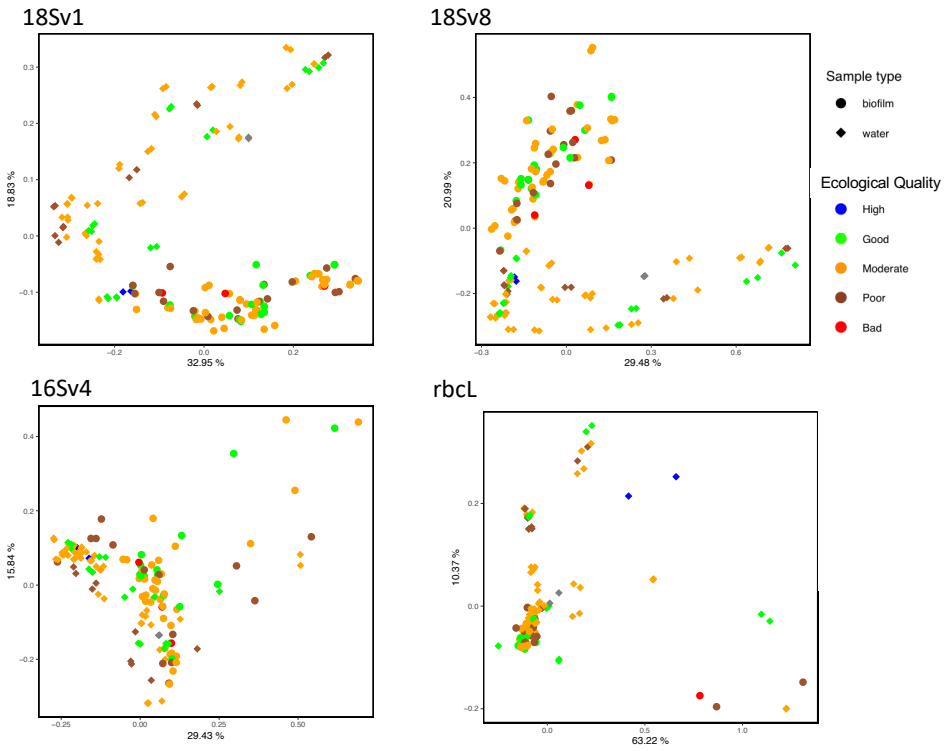


Figure 2 PCoA plots of weighted unifrac distance of all samples for each barcode. Axes are labelled with percentage of variance explained. Using equal rarefaction depth for both biofilm and water samples, per barcode: 16SV1 – 1000; 18SV1 – 14029, 18SV8 – 17201, rbcL – 1000. Biofilm sample types are circles, water samples are diamonds, coloured by WFD ecological quality score (high – blue, good - green, moderate – orange, poor - brown, bad – red, water field blank sample pool – grey diamond).

Alpha diversity as Shannon diversity and measured with pairwise kruskal wallis did not significantly differ across region in either water or biofilm samples and did not significantly differ by ecological quality in biofilm samples. There was a

significant difference (adjusted p value <0.05) in water samples for amplicon 18SV1 between Good and Poor ecological quality lakes (adjusted p value = 0.017) and between Moderate and Poor ecological quality lakes (adjusted p value = 0.004). There was also a significant difference (adjusted p value <0.05) in water samples for amplicon 18SV8 between Good and Poor ecological quality lakes (adjusted p value = 0.044) and between Moderate and Poor ecological quality lakes (adjusted p value = 0.017).

Beta diversity as weighted unifrac distance did not significantly differ ($p<0.05$) across regions or ecological quality categories, nor was there a significant interaction of region and ecological quality in biofilm samples for barcodes 16SV4, 18SV1 or 18SV8 (table 3), whereas these factors were all significant for the equivalent tests in the water samples. Ecological quality was significantly different across ecological quality categories in rbcL biofilm samples. Water samples for the barcode rbcL were significantly different across region and ecological quality categories, but the interaction of these two factors was not significant (table 3).

Barcode	Factor	Biofilm samples		Water samples	
		R^2	P value	R^2	P value
16SV4	Ecological quality	0.0718	0.233	0.1809	**0.001**
	Region	0.0705	0.786	0.2429	**0.001**
	Ecological quality:Region	0.0772	0.621	0.1121	**0.001**
18SV1	Ecological quality	0.0671	0.096	0.1869	**0.001**
	Region	0.0850	0.233	0.1574	**0.001**
	Ecological quality:Region	0.0903	0.158	0.1012	**0.002**
18SV8	Ecological quality	0.0648	0.126	0.1488	**0.002**
	Region	0.0802	0.37	0.1476	**0.001**

	Ecological quality:Region	0.0744	0.5	0.0831	**0.011**
rbcL	Ecological quality	0.1324	**0.044**	0.1396	**0.031**
	Region	0.1100	0.131	0.1945	**0.01**
	Ecological quality:Region	0.0639	0.363	0.0634	0.165

Table 3: Permutational analysis of variance (PERMANOVA) on weighted unifrac distances between biofilm samples per barcode and between water samples per barcode, 999 permutations. Test carried out using qiime diversity adonis and formula Ecological quality*Region. Significant (<0.05) p values indicated in bold.

There was limited overlap in the most abundant taxonomic groupings between the biofilm and water sample types (figure 3) with up to four groups of the ten most abundant shared between sample types: 18SV1 - Diplostroca and Chlorophyacae; 18SV8 – Oligohymenophorea; 16SV4 - Sporichthyaceae, Comamonadaceae; rbcL - Aulacoseiraceae, Bacillariaceae, Tabellariaceae, Fragilariaceae.

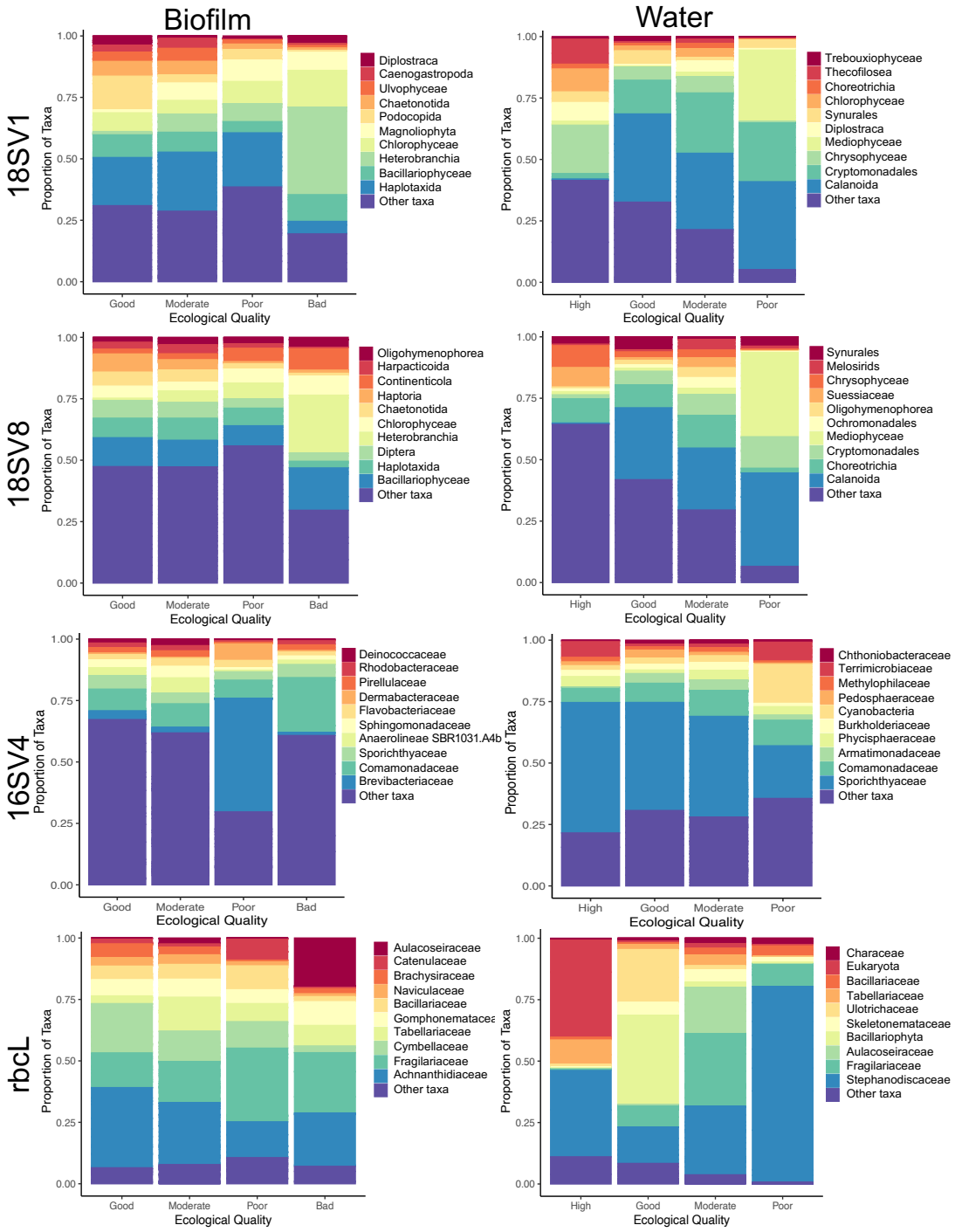

Biofilm
Water
18SV1
18SV8
16SV4
rbcL
Proportion of Taxa
Ecological Quality
Good
Moderate
Poor
Bad
High
Diplostraca
Caenogastropoda
Ulvophyceae
Chaetonotida
Podocopida
Magnoliophyta
Chlorophyceae
Heterobranchia
Bacillariophyceae
Haplotaxida
Other taxa
Trebouxiophyceae
Thecofilosea
Choreotrichia
Chlorophyceae
Synurales
Diplostraca
Mediophyceae
Chrysophyceae
Cryptomonadales
Calanoida
Other taxa
Oligohymenophorea
Harpacticoida
Continenticola
Haptoria
Chaetonotida
Chlorophyceae
Heterobranchia
Diptera
Haplotaxida
Bacillariophyceae
Other taxa
Synurales
Melosirids
Chrysophyceae
Suessiaceae
Oligohymenophorea
Ochromonadales
Mediophyceae
Cryptomonadales
Choreotrichia
Calanoida
Other taxa
Deinococcaceae
Rhodobacteraceae
Pirellulaceae
Dermabacteraceae
Flavobacteriaceae
Sphingomonadaceae
Anaerolineae SBR1031.A4b
Sporichthyaceae
Comamonadaceae
Brevibacteriaceae
Other taxa
Chthoniobacteraceae
Terrimicrobiaceae
Methylophilaceae
Pedosphaeraceae
Cyanobacteria
Burkholderiaceae
Phycisphaeraceae
Armatimonadaceae
Comamonadaceae
Sporichthyaceae
Other taxa
Aulacoseiraceae
Catenulaceae
Brachysiraceae
Naviculaceae
Bacillariaceae
Gomphonemataceae
Tabellariaceae
Cymbellaceae
Fragilariaceae
Achnanthidiaceae
Other taxa
Characeae
Eukaryota
Bacillariaceae
Tabellariaceae
Ulotrichaceae
Skeletonemataceae
Bacillariophyta
Aulacoseiraceae
Fragilariaceae
Stephanodiscaceae
Other taxa

Figure 3 Taxonomic bar plots of the top ten most abundant families (or lowest assigned taxonomic group) across the four barcodes (18SV1, 18SV8, 16SV4 and rbcL) in each sample type (biofilm or water), for each ecological quality score. Note that there are no biofilm samples from 'High' quality sites and there are no water samples from 'Bad' quality sites.

5.4 DISCUSSION

Despite large scale, high effort and high cost monitoring and remediation programmes worldwide, freshwater lakes globally are of poor quality. In the USA, 55% of lakes are too polluted for swimming, fishing or drinking (Kelderman *et al.*, 2022) and about 60% of freshwater ecosystems in Europe do not meet a 'good ecological status' (Kristiansen *et al.,* 2018)

Water framework directive ecological scores use the 'one out, all out' principle, which means that if a water body fails on any one metric, then it will fail for the whole category (UK Technical Advisory Group on the Water Framework Directive, 2007). This ensures that serious lapses in contaminants are highlighted, however it removes the room for nuance in the individual water body and may mask community level biological differences. In this study, the ecological overall quality score was used to group lakes and assess whether community composition was reflected in the quality score. However, the ecological score does not account for components within the overall chemical classification (UK Technical Advisory Group on the Water Framework Directive, 2007). Due to the 'one out, all out' principle, all water bodies in England have failed to achieve 'good' status in the chemical score (Environment Agency and

Natural England, 2023), therefore the chemical score did not vary across lakes and would not have given any explanatory power in this analysis. Furthermore, none of the lakes in this study achieved a ‘good’ or ‘high’ overall score, which reduced the explanatory power of the overall score so it was not utilised in this study.

The ecological quality of the water sampled lakes explained a significant (p value > 0.05), but small, proportion of beta diversity community variance across quality score types (R^2 less than 0.2, table 3), indicating that other factors are explaining a large proportion of the variance observed. This suggests that the WFD ecological quality score is not capturing variation in the overall community composition. There is evidence that the assessment methods used in the calculation of the ecological quality score are more sensitive to some pressures (e.g. eutrophication, organic pollution) over others pressures (eg hydromorphological degradation, priority substance pollution) or the combination of pressures (Arle, Mohaupt, Kirst, 2016; Poik āne, *et al.,* 2020). This may be because the biological quality elements of the WFD ecological quality is calculated using bioindicator species or groups of species, rather than a whole-community approach. Individual species may be a poor representative of ecosystem-level changes and the link between species dynamics and drivers of biodiversity loss are therefore not being fully captured in monitoring efforts (Birk, 2012).

Geographic region of the lake also explains a significant (p value < 0.05) proportion of beta diversity community variance in water samples. This can be

due to variation in lake geology across regions and also explained by founder effects and colonisation history (Orsini, Vanoverbeke, *et al.,* 2013).

The remaining unexplained variance may be captured by elements measured in the chemical score of the WFD framework, however it is unlikely that the binary pass/fail approach of chemical classification would further separate lakes. More granular measures, including data not measured in the WFD, may explain the variation seen in community composition. For example, it has been shown in chapter 2 of this thesis that biocide usage can explain variation in lake community composition. Ideally, future work would utilise direct measurements of biocides in the waterbody. However, this is typically conducted using mass spectrometry which is expensive at large scales. Furthermore, non-targeted screening is challenging to analyse, whereas targeted screening limits the range of compounds which can be detected and may not identify biocide transformation products (Hollender *et al.,* 2023). Therefore, direct measurements can be supplemented with biocide usage data, such as that collected by FERA in the pesticide usage survey or qualitative data of land use variation across England.

There are large changes in community composition across regions in water samples (table 3; figure S1) which preclude the differential abundance testing with ANCOM (suggested maximum 25% composition change between groups (Mandal *et al.,* 2015)). However, it is likely that the change in overall composition of the community, rather than a focus on a few individual taxa, is better able to capture the biological response to changing environmental conditions (Eastwood *et al.,* 2023).

5.4.1 Study Limitations

There was a clear difference between biofilm and water sample types, with differing taxonomic groupings recovered in each sample type. This is likely due to differences in benthic and pelagic communities, for example the most common taxonomic groups in the eukaryotic biofilm community (as measured in the 18SV1 and 18SV8) are worms and diatoms, whereas in water, zooplankton groups such as copopods and aquatic ciliates dominate (figure 3). The biofilm collection method is optimised to sample diatom communities (Bennion *et al.,* 2012). The difference in community composition captured by biofilm and water sample types may also be influenced by the difference in both sampling season and sampling year. The differing sample types were collected in separate sampling campaigns, so seasonal fluctuations in the natural community composition could alter the observed community, however this cannot be conclusively shown in this study due to lack of contemporaneous samples.

A further limitation of this study is the background levels of amplification in the water field blank samples. The individual field blank samples were shown to contain template DNA, and the pooled field blank sample (as described in the methods section) was comparable in read depth and composition to biological samples, as shown by lack of separation in the PCOA (figure 2; grey diamond). Standard methods for contamination control, such as removal of ASVs found in blanks from all biological samples, were not suitable in this study due to the high degree of overlap with biological samples which would've resulted in a strong bias of remaining ASVs.

5.5 SUPPLEMENTARY INFORMATION

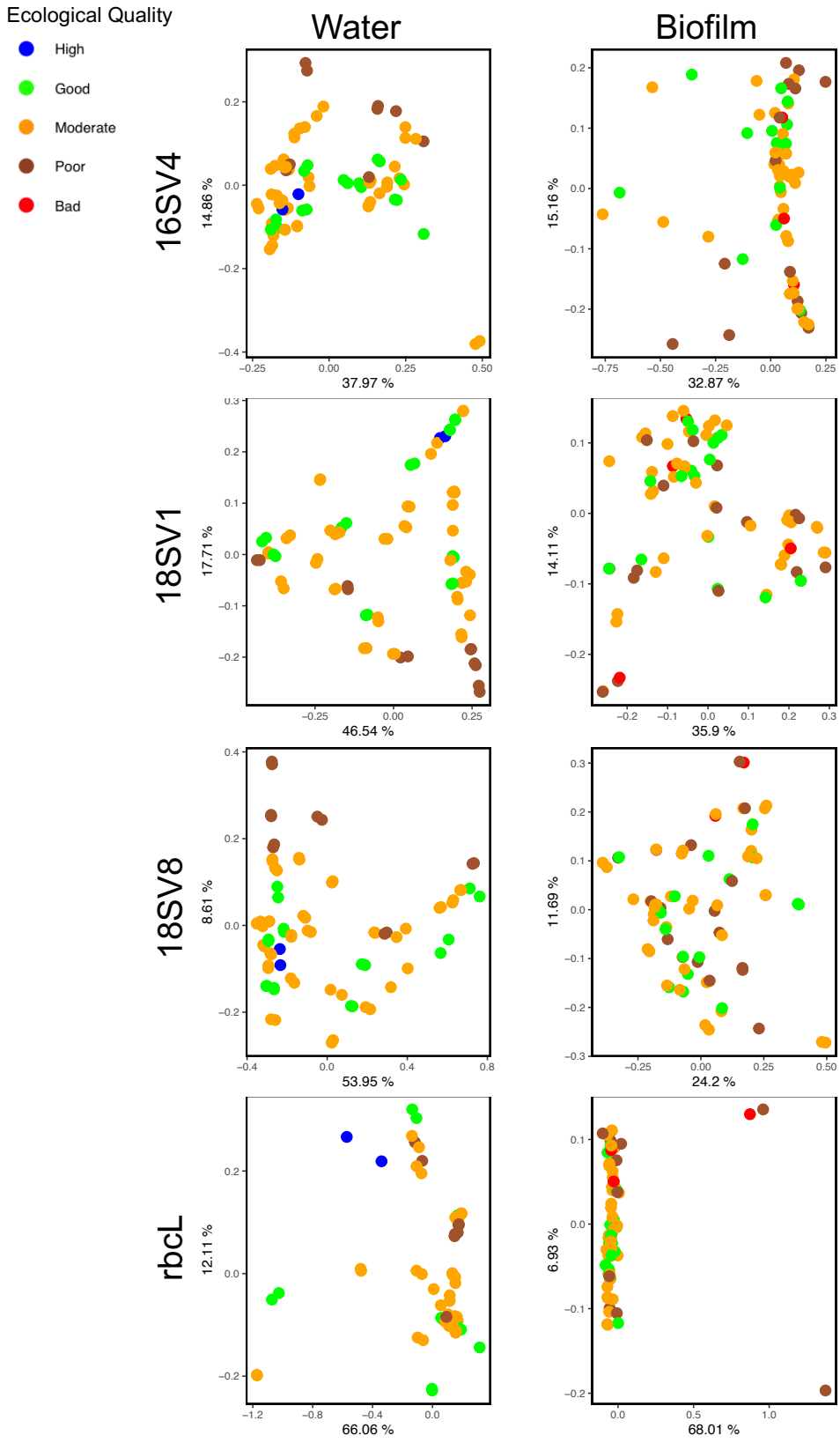
Lake name	biofilm date	water date	Region	Overall score	Ecological Quality Score
01	30/11/2016	not sampled	Anglian	Moderate	Moderate
02	20/11/2014	05/12/2018	North West	Moderate	Moderate
03	not sampled	10/01/2019	North West	Moderate	Good
04	09/09/2016	16/01/2019	North West	Moderate	Good
05	05/09/2016	23/01/2019	North West	Moderate	Moderate
06	09/09/2016	09/01/2019	North West	Moderate	Moderate
07	24/10/2016	06/12/2018	North West	Moderate	Moderate
08	06/09/2016	19/12/2018	North West	Moderate	Good
09	29/09/2016	13/12/2018	North West	Moderate	Good
10	08/09/2016	05/12/2018	North West	Moderate	Moderate
11	07/09/2016	06/12/2018	North West	Moderate	Moderate
12	02/09/2016	12/12/2018	North West	Moderate	Moderate
13	23/09/2016	not sampled	North West	Moderate	Good
14	26/10/2016	09/01/2019	North West	Moderate	Moderate
15	30/09/2016	19/12/2018	North West	Moderate	Moderate
16	01/09/2016	10/01/2019	North West	Moderate	Moderate
17	06/10/2016	29/11/2018	North West	Moderate	Good
18	08/09/2016	not sampled	North West	Moderate	Moderate
19	08/09/2016	not sampled	North West	Moderate	Moderate
20	not sampled	12/12/2018	North West	moderate	moderate
21	not sampled	29/11/2019	North West	Moderate	High
22	not sampled	13/12/2018	North West	Moderate	Moderate
23	not sampled	20/03/2019	Midlands	Poor	Poor
24	19/10/2016	not sampled	South West	Moderate	Moderate
25	08/09/2016	29/03/2019	South West	Moderate	Moderate
26	23/09/2016	03/04/2019	South West	Moderate	Moderate
27	07/09/2016	28/03/2019	South West	Poor	Poor
28	28/10/2014	07/03/2019	Anglian	Moderate	Good
29	30/10/2014	not sampled	Anglian	Moderate	Moderate
30	06/09/2016	not sampled	North West	Bad	Bad
31	20/09/2016	22/03/2019	North West	Poor	Poor
32	not sampled	21/03/2019	North West	Moderate	Moderate
33	14/11/2014	not sampled	Southern	Poor	Poor
34	not sampled	27/03/2019	Southern	Moderate	Moderate
35	06/10/2016	12/03/2019	North East	Moderate	Good
36	06/10/2016	15/06/2019	North East	Moderate	Moderate
37	06/10/2016	14/03/2019	North East	Moderate	Moderate
38	26/09/2016	13/03/2019	North East	Moderate	Good
39	not sampled	21/02/2019	Midlands	Moderate	Moderate
40	07/09/2016	22/02/2019	Midlands	Poor	Poor
41	not sampled	18/02/2019	Midlands	Moderate	Moderate
42	21/09/2016	not sampled	Southern	Moderate	Moderate
43	09/09/2016	not sampled	Southern	Poor	Poor
44	07/09/2016	not sampled	South West	Poor	Poor
45	18/10/2016	02/04/2019	South West	Moderate	Moderate
46	07/10/2016	25/03/2019	Thames	Poor	Poor
47	27/09/2016	not sampled	Thames	Bad	Bad
48	29/09/2016	not sampled	Thames	Poor	Poor
49	30/09/2014	26/02/2019	North East	Poor	Poor

50	15/09/2014	06/03/2019	North East	Moderate	Good
51	09/09/2016	not sampled	North East	Moderate	Moderate
52	not sampled	28/02/2019	North East	Moderate	Moderate

Supplementary table 1 – Lakes sampled in the study, the date sample was collected for either water and/or biofilm sample, the geographical region of the lake, the 2019 water framework directive overall quality score and 2019 water framework directive ecological quality score. Lakes are anonymised to preserve confidentiality and protect sensitive sites.

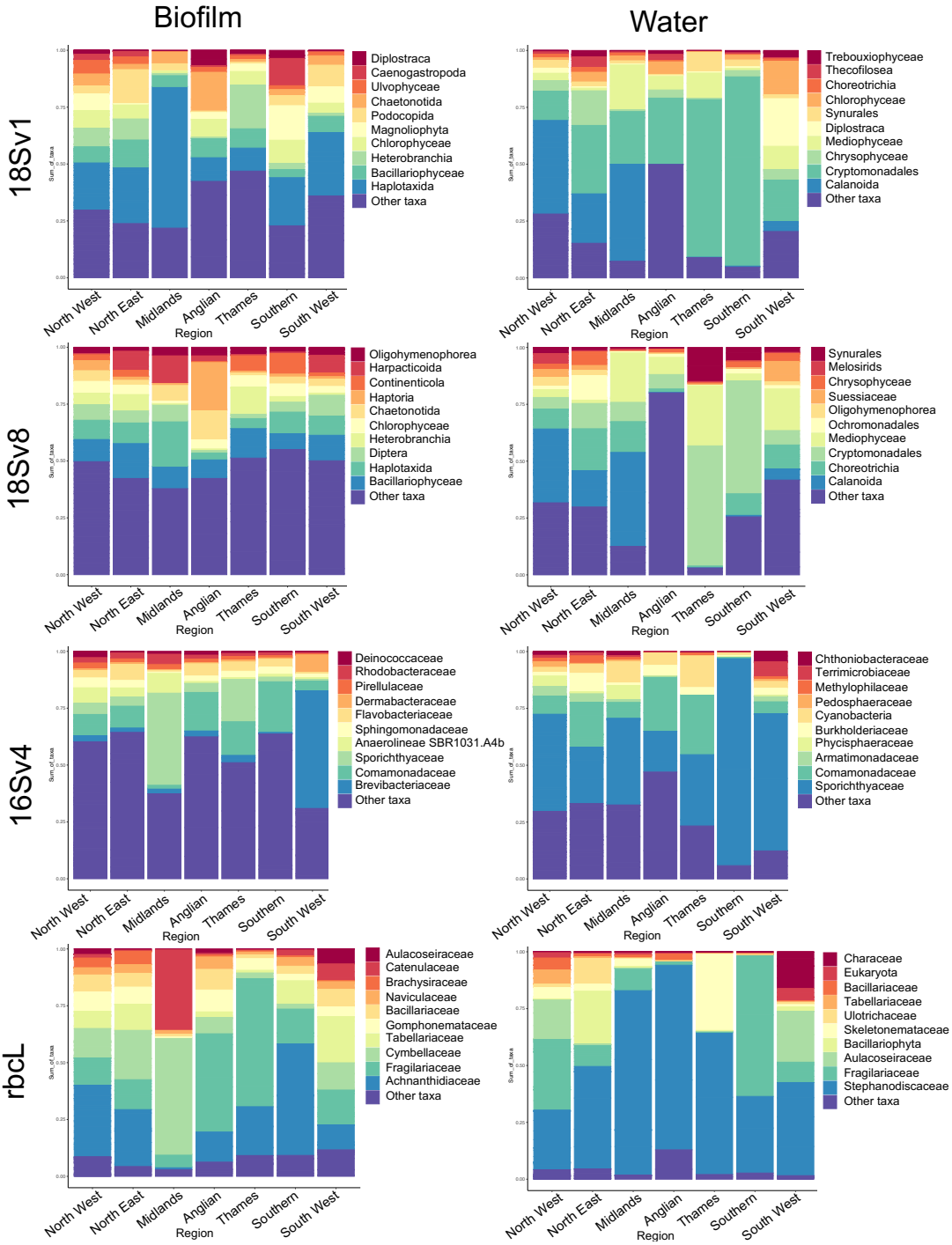
Marker gene	Reference	Primer direction	Sequence (5'-3'): adapter tag|universal overhang|primer
18SV1 V2	Hadziavdic et al., 2014	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|GCTTGTCTCAAAGATTAAGCC
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GCCTGCTGCCTTCCTTGGA
18SV8 V9	Bradley, Pinto, and Guest, 2016	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|ATAACAGGTCTGTGATGCCCT
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|CCTTCYGCAGGTTCACCTAC
16SV4	Caporaso et al., 2011a	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|GTGCCAGCMGCCGCGGTAA
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GGACTACHVGGGTWCTAAT
rbcL	Kelly et al., 2018	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT|NNNNN|ATGCGTTGGAGAGARCGTTTC
		Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT|GATCACCTTCTAATTTACCWACAACTG

Supplementary table 2 - PCR1 primers, bibliographic references and sequences. Marker gene sequence in black, universal overhang in blue and adapter tag for PCR2 primer in green. Note: “|” differentiates parts of the sequence for clarity but does not indicate a separation in the nucleotide molecule.



Supplementary Figure 1 PCoA plots of weighted unifrac distance per sample

type for each barcode. Axes are labelled with percentage of variance explained. Using equal rarefaction depth for both biofilm and water samples, per barcode: 16SV1 – 1000; 18SV1 – 14029, 18SV8 – 17201, rbcL – 1000. Points are coloured by WFD ecological quality score (high – blue, good - green, moderate – orange, poor - brown, bad – red).



Supplementary Figure 2 - Taxonomic bar plots of the top ten most abundant families (or lowest assigned taxonomic group) across the four barcodes (18SV1, 18SV8, 16SV4 and rbcL) in each sample type (biofilm or water), per region.

5.6 REFERENCES

Acreman, M., K. A. Hughes, A. H. Arthington, D. Tickner, and M-A. Dueñas. 2020. 'Protected areas and freshwater biodiversity: a novel systematic review distils eight lessons for effective conservation', *Conservation Letters*, 3: e12684.

Ahmed, S. F., P. S. Kumar, M. Kabir, F. T. Zuhara, A. Mehjabin, N. Tasannum, A. T. Hoang, Z. Kabir, and M. Mofijur. 2022. 'Threats, challenges and sustainable conservation strategies for freshwater biodiversity', *Environ Res*, 214: 113808.

Arle, J., Mohaupt, V. and Kirst, I. Monitoring of Surface Waters in Germany under the Water Framework Directive—A Review of Approaches, Methods and Results. *Water* **2016**, *8*, 217. https://doi.org/10.3390/w8060217

Balint, M., C. Nowak, O. Marton, S. U. Pauls, C. Wittwer, J. L. Aramayo, A. Schulze, T. Chambert, B. Cocchiararo, and M. Jansen. 2018. 'Accuracy, limitations and cost efficiency of eDNA-based community survey in tropical frogs', *Mol Ecol Resour*, 18: 1415-26.

Balint, M., M. Pfenninger, H. P. Grossart, P. Taberlet, M. Vellend, M. A. Leibold, G. Englund, and D. Bowler. 2018. 'Environmental DNA Time Series in Ecology', *Trends Ecol Evol*, 33: 945-57.

Bardey, D.J. 2020. 'Critically Evaluating the Consequences of a Single Species Conservation Approach', *Wildlife and Biodiversity*, 2: 555579.

Bennion, H., A. Burgess, S. Juggins, M. Kelly, G Reddihough, and M Yallop. 2012. "Assessment of ecological status in UK lakes

using diatoms." In, edited by Environment Agency. Bristol, UK.

Birk, S., W Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, S. Solimini, W. van de Bund, N. Zampoukas, and D. Hering. 2012. 'Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive', *Ecological Indicators*, 18: 31-41.

Blackman, R. C., H. C. Ho, J. C. Walser, and F. Altermatt. 2022. 'Spatio-temporal patterns of multi-trophic biodiversity and food-web characteristics uncovered across a river catchment using environmental DNA', *Commun Biol*, 5: 259.

Bohmann, K., V. Elbrecht, C. Caroe, I. Bista, F. Leese, M. Bunce, D. W. Yu, M. Seymour, A. J. Dumbrell, and S. Creer. 2022. 'Strategies for sample labelling and library preparation in DNA metabarcoding studies', *Mol Ecol Resour*, 22: 1231-46.

Bolyen, E., J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodriguez, J. Chase, E. K. Cope, R. Da Silva, C. Diener, P. C. Dorrestein, G. M. Douglas, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, S. M. Gibbons, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. A. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. D. Kaehler, K. B. Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y. X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. T. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, M. S. Robeson, 2nd, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vazquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K. C. Weber, C. H. D. Williamson, A. D. Willis, Z. Z. Xu, J. R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, and J. G. Caporaso. 2019. 'Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nat Biotechnol*, 37: 852-57.

Bonebrake, T. C., F. Guo, C. Dingle, D. M. Baker, R. L. Kitching, and L. A. Ashton. 2019. 'Integrating Proximal and Horizon Threats to Biodiversity for Conservation', *Trends Ecol Evol*, 34: 781-88.

Bradley, I. M., A. J. Pinto, and J. S. Guest. 2016. 'Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities', *Appl Environ Microbiol*, 82: 5878-91.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, and S. P. Holmes. 2016. 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nat Methods*, 13: 581-3.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', *Proc Natl Acad Sci U S A*, 108 Suppl 1: 4516-22.

Creer, S., K. Deiner, S. Frey, D. Porazinska, P. Taberlet, W. K. Thomas, C. Potter, and H. M. Bik. 2016. 'The ecologist's field guide to sequence-based identification of biodiversity', *Methods in Ecology and Evolution*, 7: 1008-18.

Cristescu, M. E., and P.D.N. Hebert. 2018. 'Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation', *Annual review of ecology, evolution, and systematics*, 49: 209–30.

Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. Isaac, and B. Collen. 2014. 'Defaunation in the Anthropocene', *Science*, 345: 401-6.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z. Kawabata, D. J. Knowler, C. Leveque, R. J. Naiman, A. H. Prieur-Richard, D. Soto, M. L. Stiassny, and C. A. Sullivan. 2006. 'Freshwater biodiversity: importance, threats, status and conservation challenges', *Biol Rev Camb Philos Soc*, 81: 163-82.

Durance, I., M. W. Bruford, R. Chalmers, N. A. Chappell, M. Christie, B. J. Cosby, D. Noble, S. J. Ormerod, H. Prosser, A. J. Weightman, and G. Woodward. 2016. 'The challenges of linking ecosystem services to biodiversity. ', *Advances in Ecological Research*, 54: 87-134.

Eastwood, E, J. Zhou, R. Derelle, M. Abou-Elwafa Abdallah, W. A. Stubbing, Y. Jia, S.E. Crawford, T.A. Davidson, J.K. Colbourne, S. Creer, H. Bik, H. Holler, and L. Orsini. in review. '100 years of anthropogenic impact causes changes in freshwater functional biodiversity', *eLife*, 6: e515

Eastwood, N., S. Kissane, L. Campbell, A. G. Briscoe, B. Egeter, and L. Orsini. 2024. 'Single metabarcoding multiplex captures community-level freshwater biodiversity and beyond ', *Environmental DNA*, 6: e515.

Eastwood, N., W. A. Stubbings, M. A. Abou-Elwafa Abdallah, I. Durance, J. Paavola, M. Dallimer, J. H. Pantel, S. Johnson, J. Zhou, J. S. Hosking, J. B. Brown, S. Ullah, S. Krause, D. M. Hannah, S. E. Crawford, M. Widmann, and L. Orsini. 2022. 'The Time Machine framework: monitoring and prediction of biodiversity loss', *Trends Ecol Evol*, 37: 138-46.

Environment Agency, and Natural England. 2023. "State of the water environment indicator B3: supporting evidence." In.

EU. 2000. "Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy." In, 1-73.

Eveleigh, E. S., K. S. McCann, P. C. McCarthy, S. J. Pollock, C. J. Lucarotti, B. Morin, G. A. McDougall, D. B. Strongman, J. T. Huber, J. Umbanhowar, and L. D. Faria. 2007. 'Fluctuations in density of an outbreak species drive diversity cascades in food webs', *Proc Natl Acad Sci U S A*, 104: 16976-81.

Foley, J. A., R. Defries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. A. Howard, C. J. Kucharik, C. Monfreda, J. A. Patz, I. C. Prentice, N. Ramankutty, and P. K. Snyder. 2005. 'Global consequences of land use', *Science*, 309: 570-4.

Fordham, D. A. , S.T. Jackson, S.C. Brown, B. Huntley, B. W. Brook, D. Dahl-Jensen, M. Thomas, P. Gilbert, B.L. Otto-Bliesner, A. Svensson, S. Theodoridis, J. M. Wilmshurst, J.C. Buettel, E. Canteri, M. McDowell, L. Orlando, J. Pilowsky, C. Rahbek, and D. Nogues-Bravo. 2020. 'Using paleo-

archives to safeguard biodiversity under climate change', *Science*, 369: eabc5654.

Garlapati, D. , B. Charankumar, K. Ramu, P. Madeswaran, and M. V. Ramana Murthy. 2019. 'A review on the applications and recent advances in environmental DNA (eDNA) metagenomics', *Review in Environmental Science and Biotechnology*, 18: 389–411.

Gillson, L., and R. Marchant. 2014. 'From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology', *Trends Ecol Evol*, 29: 317-25.

Hadziavdic, K., K. Lekang, A. Lanzen, I. Jonassen, E. M. Thompson, and C. Troedsson. 2014. 'Characterization of the 18S rRNA gene for designing universal eukaryote specific primers', *PLoS One*, 9: e87624.

Hofman, C. A., T. C. Rick, R. C. Fleischer, and J. E. Maldonado. 2015. 'Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene', *Trends Ecol Evol*, 30: 540-9.

Hollender, J., Schymanski, E.L., Ahrens, L., Alygizakis, N., Béen, F., Bijlsma, L., Brunner, A.M., Celma, A., Fildier, A., Fu, Q., Gago-Ferrero, P., Gil-Solsona, R., Haglund, P., Hansen, M., Kaserzon, S., Kruve, A., Lamoree, M., Margoum, C., Meijer, J., Merel, S., Rauert, C., Rostkowski, P., Samanipour, S., Schulze, B., Schulze, T., Singh, R.R., Slobodnik, J., Steininger-Mairinger, T., Thomaidis, N.S., Togola, A., Vorkamp, K., Vulliet, E., Zhu, L. and Krauss, M. 2023. 'NORMAN guidance on suspect and non-target screening in environmental monitoring', *Environ Sci Eur* 35:75

Kelderman, K., Phillips, A., Pelton, T., Schaeffer, E., MacGillis-Falcon, P., & Bernhardt, C. (2022). *The Clean Water Act at 50: Promises Half Kept at the Half-Century Mark*. Environmental Integrity Project. https://environmentalintegrity.org/wp-content/uploads/2022/03/CWA-report-UPDATED-8.9.23.pdf

Kelly, M., N. Boonham, S. Juggins, P. Kille, D. Mann, D. Pass, M. Sapp, S. Sato, and R. Glover. 2018. "A DNA based diatom metabarcoding approach for

Water Framework Directive classification of rivers." In. Bristol: Environment Agency of England.

Kristensen, P., Whalley, C., Zal, F.N.N., Christiansen, T. (2018) *European waters Assessment of status and pressures 2018*. rep. Luxembourg: Publications Office of the European Union.

Kruskal, W, and W Wallis. 1952. 'Use of Ranks in One-Criterion Variance Analysis', *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION*, 47.

MacConaill, L. E., R. T. Burns, A. Nag, H. A. Coleman, M. K. Slevin, K. Giorda, M. Light, K. Lai, M. Jarosz, M. S. McNeill, M. D. Ducar, M. Meyerson, and A. R. Thorner. 2018. 'Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing', *BMC Genomics*, 19: 30.

Mace, G. M., K. Norris, and A. H. Fitter. 2012. 'Biodiversity and ecosystem services: a multilayered relationship', *Trends Ecol Evol*, 27: 19-26.

Mandal, S., W. Van Treuren, R. A. White, M. Eggesbo, R. Knight, and S. D. Peddada. 2015. 'Analysis of composition of microbiomes: a novel method for studying microbial composition', *Microb Ecol Health Dis*, 26: 27663.

Martin, M. 2023. 'Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads', *EMBnet journal*, 17: 10-12.

Naggs, F. 2017. 'Saving Living Diversity in the Face of the Unstoppable 6th Mass Extinction: A Call for Urgent International Action', *Population and Sustainability* 1: 67-81.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. De Meester. 2013. 'Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization', *Mol Ecol*, 22: 5983-99.

Poikāne, S., Herrero, F. S., Kelly, M., Borja, Á., Birk, S., & Van De Bund, W. (2020). European aquatic ecological assessment methods: A critical review of their sensitivity to key pressures. *Science of the Total Environment*, *740*, 140075. https://doi.org/10.1016/j.scitotenv.2020.140075

R Core Team. 2020. "R: A language and environment for statistical computing." In. Vienna, Austria: R Foundation for Statistical Computing, .

Rimet, F., E. Gusev, M. Kahlert, M. G. Kelly, M. Kulikovskiy, Y. Maltsev, D. G. Mann, M. Pfannkuchen, R. Trobajo, V. Vasselon, J. Zimmermann, and A. Bouchez. 2019. 'Diat.barcode, an open-access curated barcode library for diatoms', *Sci Rep*, 9: 15116.

Ropert-Coudert, Y., and R.P. Wilson. 2005. 'Trends and perspectives in animal-attached remote sensing', *Frontiers in Ecology and the Environment*, 3: 437–44.

Ruckelshaus, M. H., S. T. Jackson, H. A. Mooney, K. L. Jacobs, K. S. Kassam, M. T. K. Arroyo, A. Baldi, A. M. Bartuska, J. Boyd, L. N. Joppa, A. Kovacs-Hostyanszki, J. P. Parsons, R. J. Scholes, J. F. Shogren, and Z. Ouyang. 2020. 'The IPBES Global Assessment: Pathways to Action', *Trends Ecol Evol*, 35: 407-14.

Ruppert, K. M. , R. J. Kline, and S. Md Rahman. 2019. 'Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA', *Global Ecology and Conservation*, 17: e00547.

Taberlet, P., A. Bonin, L. Zinger, and E. Coissac. 2018. *Environmental DNA for biodiversity research and monitoring* (Oxford University Press: Oxford, UK).

UK Technical Advisory Group on the Water Framework Directive. 2007. "Recommendations on Surface Water Classification Schemes for the purposes of the Water Framework Directive." In.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York}).

Wilby, N, A Law, C Bull, B Hanfling, LL Handley, and I Winfield. 2019. "A tool for classifying the ecological status of lake fish in

Britain based on eDNA metabarcoding." In.: Report to the Scottish Environment Protection Agency, UK.

WWF. 2018. "Living Planet Report: aiming higher." In, edited by M. Grooten and R.E.A. Almond. Gland, Switzerland.

Yilmaz, P., L. W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, and F. O. Glockner. 2014. 'The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks', *Nucleic Acids Res*, 42: D643-8.

Zimmermann, J., N. Abarca, N. Enke, O. Skibbe, W. H. Kusber, and R. Jahn. 2014. 'Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research', *PLoS One*, 9: e108793.

Chapter 6: DISCUSSION

Environmental change, due to human-driven climate change, chemical pollution and land use, is one of the largest challenges facing humanity and our planet (Sylvester *et al.,* 2023). Loss of ecosystem functions and services would have far reaching consequences for society and it is therefore vital that we develop a complete picture of the state of the communities which deliver these functions and services, and the priority impacts upon these communities.

However, existing methods do not encompass community level response, are low throughput and biased towards a select group of taxa, all of which can result in incomplete understanding of the impacts of environmental change (de Olde *et al.,* 2017). Large proportions of the community (especially prokaryotes) have historically been excluded from regulatory surveillance due to difficulties in identifying microscopic organisms. However, microbes provide a wide range of key ecosystem functions and due to advances in sequencing techniques and eDNA approaches, can be better integrated into environmental assessment in a high throughput and non-invasive manner (Robinson *et al.,* 2023).

Whole community assessment is challenging, with environmental DNA methods still facing questions over biased species recovery, false positive/negatives, the accuracy of quantitative analysis and universal primer selection, although there are steps forwards in technical considerations such as the fate and state of eDNA in the environment (Bista *et al.,* 2017)

For implementation in many regulatory and conservation settings, accurate taxonomic identification to species level is essential to adhere to existing policy and regulation. The monitoring of invasive and non-native species (INNS) and protected species relies upon specie levels, or sub-species level, identification which may not be possible with non-taxa specific (or 'universal') primers, such as those used in whole community metabarcoding. Furthermore, many species are absent from taxonomy assignment databases, leading to high level (e.g. order or family) being the most specific possible taxonomy assigned.

However, new initiatives, such as the Darwin Tree of Life project and UK barcode of life (UKBOL) should improve coverage of many species in such databases (The Darwin Tree of Life Project Consortium, 2022; Price et al., 2022). Although these regional initiatives are part of global projects, there is still a disparity in species coverage across different countries due to resource availability, leading to a bias towards the coverage of biodiversity which inhabits wealthier nations (Gaytán et al., 2020). Global environmental change disproportionately affects less wealthy nations (Levy and Patz, 2015), therefore the countries which would most benefit from a holistic and low cost survey of their biodiversity are the ones also least likely to be able to implement eDNA methods without further cost reductions and logistical improvements.

These cataloguing and sequencing efforts can also contribute to improving the link between specific taxa and their ecosystem function. Knowledge of the ecosystem functions delivered by certain taxa is still mostly restricted to limited taxonomic groups or narrow functions (e.g. macroinvertebrates (Cao *et al.,* 2018); feeding guilds (Thompson *et al.,* 2020)), with high throughput methods

mostly only possible for well characterised microbes (Banerjee *et al.,* 2018). Therefore, there are still open questions around the direct links between community composition changes and widespread ecosystem function delivery changes, often due to limited data availability. A potential source for such data is bulk stoichiometry of sediments cores, which can be complementary to community analysis using sedimentary eDNA. Biogeochemical functions measured as the accumulation rate of total organic, nitrogen and phosphorus can elucidate long-term dynamics in productivity as influenced by nutrient availability and the relationships of stoichiometric ratios, productivity, and biological attributes.

Environmental DNA methods for biodiversity monitoring can provide information of presence absence and increasingly on abundance data of organisms (Skelton, Cauvin and Hunter, 2022). However, other key measures for conservation or ecological understanding such as body quality or species interactions are not measurable with eDNA (Ruppert, Kiline and Rahman, 2019). Other information, whilst not directly measured, can be inferred such as behaviour (e.g. migration can be tracked through eDNA deposition (Maruyama *et al.,* 2018)) or sex ratios (theoretically possible using eDNA qPCR), although these are not typical or routine uses of eDNA due to difficulties in data interpretation and their population-specific nature.

Multi-marker metabarcoding can go some way to improve whole community coverage by including some taxa specific markers (e.g. rbcL for diatoms, ITS for fungi) to provide more detailed information on taxa of particular interest. However, as shown in chapter 4 of this thesis, there is still a disparity in

detection of rare species (chapter 4 Figure 3; Supplementary figure 3). The relevance of these species is highly context dependent, for example rare species may have functionally distinct traits and disproportionately contribute to some, but not all, ecosystem services (Dee *et al.,* 2019). The reduced cost of multimarker metabarcoding can balance this out by enabling a greater sequencing depth at an equivalent cost and detect these rarer species using metabarcoding approaches. Alternatively, qPCR methods which are more targeted than metabarcoding can be used for focussed rare species detection (Deiner, *et al.,* 2017).

Longitudinal analysis of the community level response using eDNA can capture whole community responses to environmental change (chapters 2,3 of this thesis). However, both technological and conceptual limitations remain in the use of historical eDNA. Primarily, questions over the provenance and taphonomy of eDNA cause uncertainty in the interpretation of historical community analysis (Giguet-Covex *et al.,* 2019). Ancient sedimentary DNA has been successfully recovered from sediments up to five thousand years old (Giguet-Covex *et al.,* 2014) and long term degradation leaves measurable changes in the DNA molecule, such as cytosine deamination, which can be measured to assure the ancient provenance of the molecule (Krause *et al.*, 2010), excluding modern-day contamination. However, the time scales covered in this thesis (up to 100 years) are too short for these molecular changes to sufficiently accumulate for analysis. Instead, a combination of robust extraction protocols (e.g. use of clean rooms, single use tools or nucleotide degrading cleaning chemicals) and multiple control or blank samples can be used to

minimise contamination (Bohmann *et al.,* 2021; Epp *et al.,* 2019). Data analysis steps can additionally be used to identify and remove contaminants such as known frequent contamination sequences, or sequences found in negative controls (Deiner *et al.,* 2017). These steps can minimise and identify contaminating DNA sequences, but do not in themselves assure the historical source of the DNA. Sediment characteristics such as mineral composition, surface charge and pH can impact the persistence and mobility of DNA (Sand, Jelavic and Prohaska, 2024). No sediment characterisation was performed on the samples used in this thesis, so the effect of these biogeochemical processes is unknown. However, ecological metrics were instead used to understand the changes in community characteristics which could indicate either a bias in DNA recovery due to degradation or 'smearing' of the DNA within the core. This did not appear to be the case in this study (Chapter 3; Supplementary figure 1).

Understanding biodiversity change in the context of wider environmental changes is often limited by the availability of environmental records. Some parameters (such as bulk stoichiometry, above; chemical residues, as in chapter 3; seeds or fossilised remains (Nogués-Bravo et al., 2018)) can be directly measured from sediment cores alongside the biological fingerprint of eDNA. However, this can lead to bias due to differences in degradation rates across different types of chemical, or the persistence of relatively few species remains. Furthermore, some types of data are difficult or impossible to recover from sediment cores, especially at fine temporal scales, such as yearly climatic changes or land use variation. Instead, data from weather monitoring stations,

chemical sales records or historical maps can be used, which may not reflect site specific events, may be sparse in space or time, biased (e.g. only record subgroups of chemicals or specific uses) or only cover short time periods.

To complement longitudinal studies of community response, spatial studies (such as in chapter 5) can examine gradients of pressure and combinations of multiple stressors in a way which may not be feasible with historical samples due to lack of supporting data. Contemporary surveys also may provide better coverage of changes in community response to novel pressures which are spatially heterogenous, enabling better decision making from environmental practitioners.

With current standard methods it is labour-intensive, costly and analytically challenging to survey biological responses to spatial and temporal variation in pressure and the resultant variation ecosystem function and ecosystem service delivery. However, surveys across pressure gradients (whether throughout time or space) and realistic combinations of pressures are key for understanding ecosystem resilience. The advances in eDNA approaches presented in this thesis go some way to addressing the challenges associated with this large scale data collection and interpretation.

Data driven approaches have the potential to make use of large datasets, such as those generated in whole community eDNA surveys, and detect correlations in previously under-studied taxa or environmental stressors. However, these correlations do not prove causation. Instead, they can be used to generate hypotheses which can be tested, for example in mesocosm studies, to establish causation. This level of proof is typically required for stringent regulation (e.g.

chemical product bans) however they are costly and time consuming. By establishing repeated correlations across sites or time, regulators can have improved confidence that such stressors should be targets of further study or restrictions and therefore better protect our environment and the biodiversity within it.

This thesis establishes the framework for long term biodiversity monitoring using eDNA (chapter 2) and demonstrates its use at a single site (chapter 3), building upon previous work which focussed on a more limited taxonomic scope (e.g. Capo *et al.,* 2021). A cheaper, whole community assay is developed (chapter 4) which enables the high-throughput of biodiversity across space (chapter 5).

The future direction of this work is to combine the temporal understanding gained through the use of sediment cores with the landscape-level processes investigated at a national scale to gain an understanding of the spatiotemporal variation in freshwater communities to environmental change. By using lake sites across a wide geographic area (e.g. national scale), subject to differing combinations and intensities of historical pressures such as land use, agricultural chemical pollution, industrial air pollution and climate change, patterns in response to environmental pressures can be thoroughly investigated. This wealth of data can be used as the basis for an emulator to predict responses of biodiversity to future environmental changes (chapter 2).

A first step is to identify suitable sites which have been subject to a range in both type of environmental pressures and intensity of these pressures. As environmental stress can have a non-linear impact on biodiversity (Birk *et al.,*

2020), sampling lakes which have experienced combinations of pressures is key to disentangling the effect of multiple stressors. Reference sites are also ideal to better compare the effect of environmental change upon community dynamics versus natural stochastic changes. However, there are few high-quality sites in Europe which could be used as references (Kristensen, Whalley, and Zal, 2018). Instead, lowland sites in conservation areas can be used which are likely to have lower than typical pollution.

In addition to biodiversity assessment, samples from such lakes can be used for historical environmental assessment. Through the combination of targeted mass-spectrometry (as used for DDT analysis in chapter 3 of this thesis) and non-targeted mass-spectrometry, a detailed understanding of site-specific pollution can be gained. Bulk sediment stoichiometry can complement this analysis to provide information on ecosystem functioning (such as lake productivity; Tipping *et al.,* 2016).

Whilst a nuanced understanding of long-term changes in biodiversity associated with environmental changes is a valuable source of data, its can best be used to improve the future state of the environment by using such data to train machine learning models (chapter 2). These models can then predict biodiversity changes under future pollution and climate scenarios, informing stakeholders on the potential outcomes of management or regulation decisions.

6.1 REFERENCES

Banerjee, S., Schlaeppi, K., van der Heijden, M.G.A., 2018. Keystone taxa as drivers of microbiome structure and functioning. Nat. Rev. Microbiol. 16, 567–576. https://doi.org/10.1038/s41579-018-0024-1

Birk, S., Chapman, D., Carvalho, L., Spears, BM., Andersen, HE., Argillier, C., Auer, S., Baattrup-Pedersen, A., Banin, L., Beklioğlu, M., Bondar-Kunze, E., Borja, A., Branco, P., Bucak, T., Buijse, AD., Cardoso, AC., Couture, RM., Cremona, F., de Zwart, D., Feld, CK., Ferreira, MT., Feuchtmayr, H., Gessner, MO., Gieswein, A., Globevnik, L., Graeber, D., Graf, W., Gutiérrez-Cánovas, C., Hanganu, J., Işkın, U., Järvinen, M., Jeppesen, E., Kotamäki, N., Kuijper, M., Lemm, JU., Lu, S., Solheim, AL., Mischke, U., Moe, SJ., Nõges, P., Nõges, T., Ormerod, SJ., Panagopoulos, Y., Phillips, G., Posthuma, L., Pouso, S., Prudhomme, C., Rankinen, K., Rasmussen, JJ., Richardson, J., Sagouis, A., Santos, JM., Schäfer, RB., Schinegger, R., Schmutz, S., Schneider, SC., Schülting, L., Segurado, P., Stefanidis, K., Sures, B., Thackeray, SJ., Turunen, J., Uyarra, MC., Venohr, M., von, der, Ohe, PC., Willby, N., Hering, D. 'Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems', *Nat Ecol Evol.* 2020 Aug;4(8):1060-1068. doi: 10.1038/s41559-020-1216-4.

Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., Creer, S., 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nat. Commun. 8, 14087. https://doi.org/10.1038/ncomms14087

Cao, X., Chai, L., Jiang, D., Wang, J., Liu, Y., Huang, Y., 2018. Loss of biodiversity alters ecosystem function in freshwater streams: potential evidence from benthic macroinvertebrates. Ecosphere 9, e02445. https://doi.org/10.1002/ecs2.2445

Capo, É., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., 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. E., . . . Parducci, L. (2021). Lake Sedimentary DNA Research on

Past Terrestrial and Aquatic Biodiversity: Overview and recommendations. *Quaternary*, *4*(1), 6. https://doi.org/10.3390/quat4010006

de Olde, E.M., Moller, H., Marchand, F., McDowell, R.W., MacLeod, C.J., Sautier, M., Halloy, S., Barber, A., Benge, J., Bockstaller, C., Bokkers, E.A.M., de Boer, I.J.M., Legun, K.A., Le Quellec, I., Merfield, C., Oudshoorn, F.W., Reid, J., Schader, C., Szymanski, E., Sørensen, C.A.G., Whitehead, J., Manhire, J., 2017. When experts disagree: the need to rethink indicator selection for assessing sustainability of agriculture. Environ. Dev. Sustain. 19, 1327–1342. https://doi.org/10.1007/s10668-016-9803-x

Dee, L.E., Cowles, J., Isbell, F., Pau, S., Gaines, S.D., Reich, P.B., 2019. When Do Ecosystem Services Depend on Rare Species? Trends Ecol. Evol. 34, 746–758. https://doi.org/10.1016/j.tree.2019.03.010

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., De Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Epp, L. S., Zimmermann, H., & Stoof-Leichsenring, K. R. (2019). Sampling and Extraction of Ancient DNA from Sediments. In *Methods in molecular biology* (pp. 31–44). https://doi.org/10.1007/978-1-4939-9176-1_5

Gaytán, Á., Bergsten, J., Canelo, T., Pérez-Izquierdo, C., Santoro, M., Bonal, R., 2020. DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots. Ecol. Evol. 10, 10754–10772. https://doi.org/10.1002/ece3.6733

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P., Griggo, C., Gielly, L., Domaizon, I., Coissac, É., David, F., Choler, P., Poulenard, J., & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms4211

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A., David, F., Taberlet, P., Brisset, É., Guiter, F., Sinet, R., & Arnaud, 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). https://doi.org/10.1038/s41598-019-50339-1

Jia, W., Anslan, S., Chen, F., Cao, X., Dong, H., Dulias, K., Gu, Z., Heinecke, L., Jiang, H., Kruse, S., Kang, W., Li, K., Liu, S., Liu, X., Jie, D., Ni, J., Schwalb, A., Stoof-Leichsenring, K. R., Shen, W., Tian, F., Wang, J., Wang, Y., Wang, Y.,, Xu, H., Zhang, D., Herzschuh, U. (2022). Sedimentary ancient DNA reveals past ecosystem and biodiversity changes on the Tibetan Plateau: Overview and prospects. *Quaternary Science Reviews*, *293*, 107703. https://doi.org/10.1016/j.quascirev.2022.107703

Krause, J., Briggs, A. W., Kircher, M., Maričić, T., Zwyns, N., Деревянко, А. П., & Pääbo, S. (2010). A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Current Biology*, *20*(3), 231–236. https://doi.org/10.1016/j.cub.2009.11.068

Kristensen, P., Whalley, C., Zal, F.N.N., Christiansen, T. (2018) *European waters Assessment of status and pressures 2018*. rep. Luxembourg: Publications Office of the European Union.

Levy, B.S., Patz, J.A., 2015. Climate Change, Human Rights, and Social Justice. Ann. Glob. Health, Climate Change, Global Health and Human Rights 81, 310–322. https://doi.org/10.1016/j.aogh.2015.08.008

Maruyama, A., Sugatani, K., Watanabe, K., Yamanaka, H., & Imamura, A. (2018). Environmental DNA analysis as a non-invasive quantitative tool for reproductive migration of a threatened endemic fish in rivers. *Ecology and Evolution*, *8*(23), 11964–11974. https://doi.org/10.1002/ece3.4653

Nogués-Bravo, D., Rodríguez-Sánchez, F., Orsini, L., Boer, E. de, Jansson, R., Morlon, H., Fordham, D.A., Jackson, S.T., 2018. Cracking the Code of

Biodiversity Responses to Past Climate Change. Trends Ecol. Evol. 33, 765–776. https://doi.org/10.1016/j.tree.2018.07.005

Price, B., Misra, R., Broad, G., and Clark, K., 2022 Initiation of UK Barcode of Life and filling priority gaps DEFRA Centre of Excellence for DNA Methods. Natural England Joint Publication JP041

Robinson, J.M., Hodgson, R., Krauss, S.L., Liddicoat, C., Malik, A.A., Martin, B.C., Mohr, J.J., Moreno-Mateos, D., Muñoz-Rojas, M., Peddle, S.D., Breed, M.F., 2023. Opportunities and challenges for microbiomics in ecosystem restoration. Trends Ecol. Evol. S0169-5347(23)00211–2. https://doi.org/10.1016/j.tree.2023.07.009

Ruppert, K. M., Kline, R. J., & Rahman, S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Sand, K. K., Jelavić, S., Kjær, K. H., & Prohaska, A. (2024). Importance of eDNA taphonomy and sediment provenance for robust ecological inference: Insights from interfacial geochemistry. *Environmental DNA*, *6*(2). https://doi.org/10.1002/edn3.519

Skelton, J., Cauvin, A., & Hunter, M. E. (2022). Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. *Environmental DNA*, *5*(5), 1092–1104. https://doi.org/10.1002/edn3.355

Sylvester, F., Weichert, F.G., Lozano, V.L., Groh, K.J., Bálint, M., Baumann, L., Bässler, C., Brack, W., Brandl, B., Curtius, J., Dierkes, P., Döll, P., Ebersberger, I., Fragkostefanakis, S., Helfrich, E.J.N., Hickler, T., Johann, S., Jourdan, J., Klimpel, S., Kminek, H., Liquin, F., Möllendorf, D., Mueller, T., Oehlmann, J., Ottermanns, R., Pauls, S.U., Piepenbring, M., Pfefferle, J., Schenk, G.J., Scheepens, J.F., Scheringer, M., Schiwy, S., Schlottmann, A., Schneider, F., Schulte, L.M., Schulze-Sylvester, M., Stelzer, E., Strobl, F., Sundermann, A., Tockner, K., Tröger, T., Vilcinskas, A., Völker, C., Winkelmann, R., Hollert, H.,

2023. Better integration of chemical pollution research will further our understanding of biodiversity loss. Nat. Ecol. Evol. https://doi.org/10.1038/s41559-023-02117-6

The Darwin Tree of Life Project Consortium, 2022. Sequence locally, think globally: The Darwin Tree of Life Project. PNAS 119. https://doi.org/10.1073/pnas.2115642118

Thompson, M.S.A., Pontalier, H., Spence, M.A., Pinnegar, J.K., Greenstreet, S.P.R., Moriarty, M., Hélaouët, P., Lynam, C.P., 2020. A feeding guild indicator to assess environmental change impacts on marine ecosystem structure and functioning. J. Appl. Ecol. 57, 1769–1781. https://doi.org/10.1111/1365-2664.13662

Tipping, E., Boyle, J., Schillereff, D., Spears, B. M., & Phillips, G. (2016). Macronutrient processing by temperate lakes: A dynamic model for long-term, large-scale application. *Science of the Total Environment*, *572*, 1573–1585. https://doi.org/10.1016/j.scitotenv.2015.09.129

Appendices

Appendix A

Supplementary tables for chapter 3

See supplementary files 1-5 at doi:10.7554/eLife.86576.3