

100 years of anthropogenic impact causes changes in freshwater functional biodiversity

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2

August 31, 2023 (this version)

Reviewed preprint version 1

June 13, 2023

Posted to bioRxiv

February 27, 2023

Sent for peer review

February 26, 2023

Niamh Eastwood , Jiarui Zhou, Romain Derelle, Mohamed Abou-Elwafa Abdallah, William A. Stubbings, Yunlu Jia, Sarah E. Crawford, Thomas A. Davidson, John K. Colbourne, Simon Creer, Holly Bik, Henner Hollert, Luisa Orsini

Environmental Genomics Group, School of Biosciences, the University of Birmingham, Birmingham B15 2TT, UK • School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK • Department Evolutionary Ecology & Environmental Toxicology, Faculty of Biological Sciences, Goethe University Frankfurt, Max-von-Laue Straße 13, 60438 Frankfurt am Main, Germany • Lake Group, Department of Ecoscience, Aarhus University, C.F. Møllers Allé, 8000 Aarhus, Denmark • School of Natural Sciences, Environment Centre Wales, Deiniol Road, Bangor University, Gwynedd LL57 2UW, UK • Department Marine Sciences and Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA • LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany • Department Media-related Toxicology, Institute for Molecular Biology and Applied Ecology (IME), Max-von-Laue Straße 13, 60438 Frankfurt am Main, Germany • The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK

 https://en.wikipedia.org/wiki/Open_access

 <https://creativecommons.org/licenses/by/4.0/>

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.

eLife assessment

This **fundamental** study advances the analytic toolset and understanding of long-term series of biological (freshwater) communities, and the impact of humans on these. The authors highlight the value of including not only spatiotemporal scales in biodiversity assessments but also some of the possible drivers of biodiversity loss. Analyzing their joint contribution as environmental stressors, the authors provide **compelling** evidence that ecosystem assessment methods currently used by environmental regulators throughout Europe are not fit-for-purpose, and they identify several alternatives, more robust indicators of freshwater ecosystem health. The work is timely and will be of interest to ecologists, modelers and global warming scientists in general.

Introduction

Biodiversity is the foundation of provisioning, regulating, supporting, and cultural ecosystem services ¹, which underpin economic prosperity, social well-being and quality of life ². 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 ³. Biodiversity is threatened by agricultural land use, climate change, invasive species, pollution and unsustainable production and consumption ⁴. Freshwater ecosystems have suffered the greatest biodiversity loss because of these anthropogenic drivers ⁵. Experimental manipulation of biodiversity has demonstrated the causal links between biodiversity loss and loss of ecosystem functions ⁶. 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, ⁷. 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 ⁸.

Biodiversity action plans have been devised since the 1990s. However, most strategies have failed to stop or even reduce biodiversity decline ⁹. 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 ^{10, 11}. 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 ^{10, 11}. Recent initiatives like BioTIME started collating databases with species presence and abundance recorded from time series across different ecosystems ⁷. However, freshwater ecosystems are poorly represented in these studies which at most encompass the last 10-25 years ¹². 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 ^{8, 12}. Moreover, the taxonomic species assignment in these databases is oftentimes derived from traditional observational methods (e.g. microscopy), which cannot resolve cryptic diversity ¹². 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 ¹³. More recently, *sedaDNA* (environmental DNA extracted from sediment) has emerged as a promising tool to study decade-long

biological dynamics ¹⁴. However, these studies focus on specific taxonomic groups e.g. microbes ¹⁵; ciliates ¹⁶, 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 ¹⁷. Studies are emerging that investigate the impact of chemicals ¹⁸ or climate change ¹⁹ 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 ⁸, ¹⁷. 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 ²⁰. 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 ²¹ 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 ²². 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 ²¹. 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 ²³, ²⁴. 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 ²⁵ (**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.

Results

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

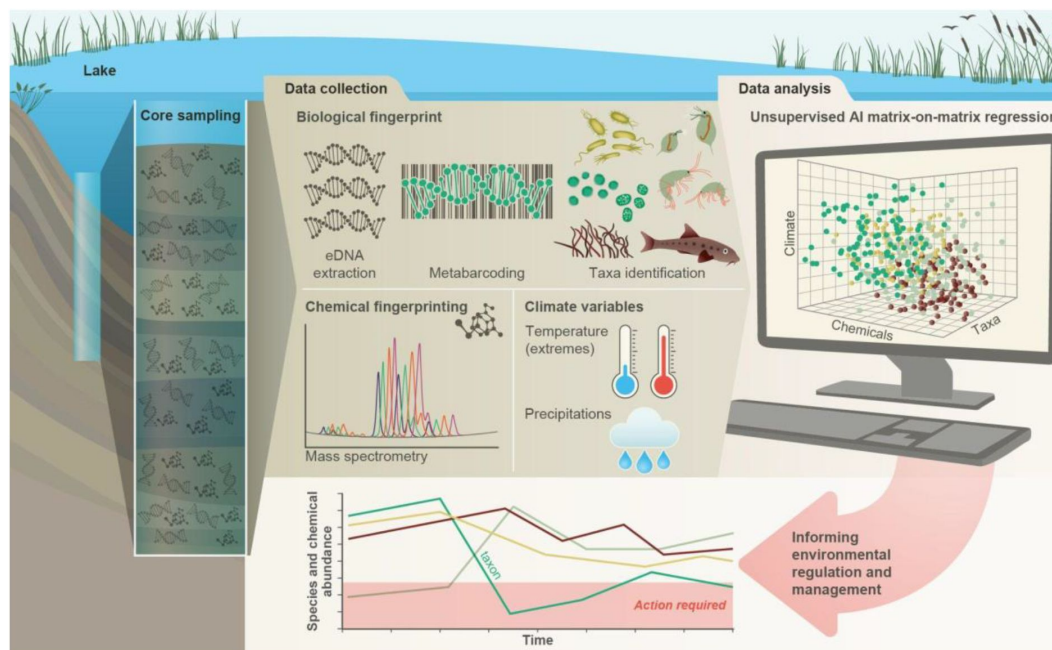


Figure 1.

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.

in the area surrounding the lake. In modern times (>1999), the lake partially recovered (see methods for more details)²¹. 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 *Brachionidae* (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 rivers²⁶. We used our rbcL data to derive a Lake Trophic Diatom Index (LTDI2) for Lake Ring following²⁷. This result confirmed our beta diversity analysis of non-significant changes over time of the diatom community (Supplementary Fig. 3).

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, $p\text{-adj} < 0.05$) 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).

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).

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

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.

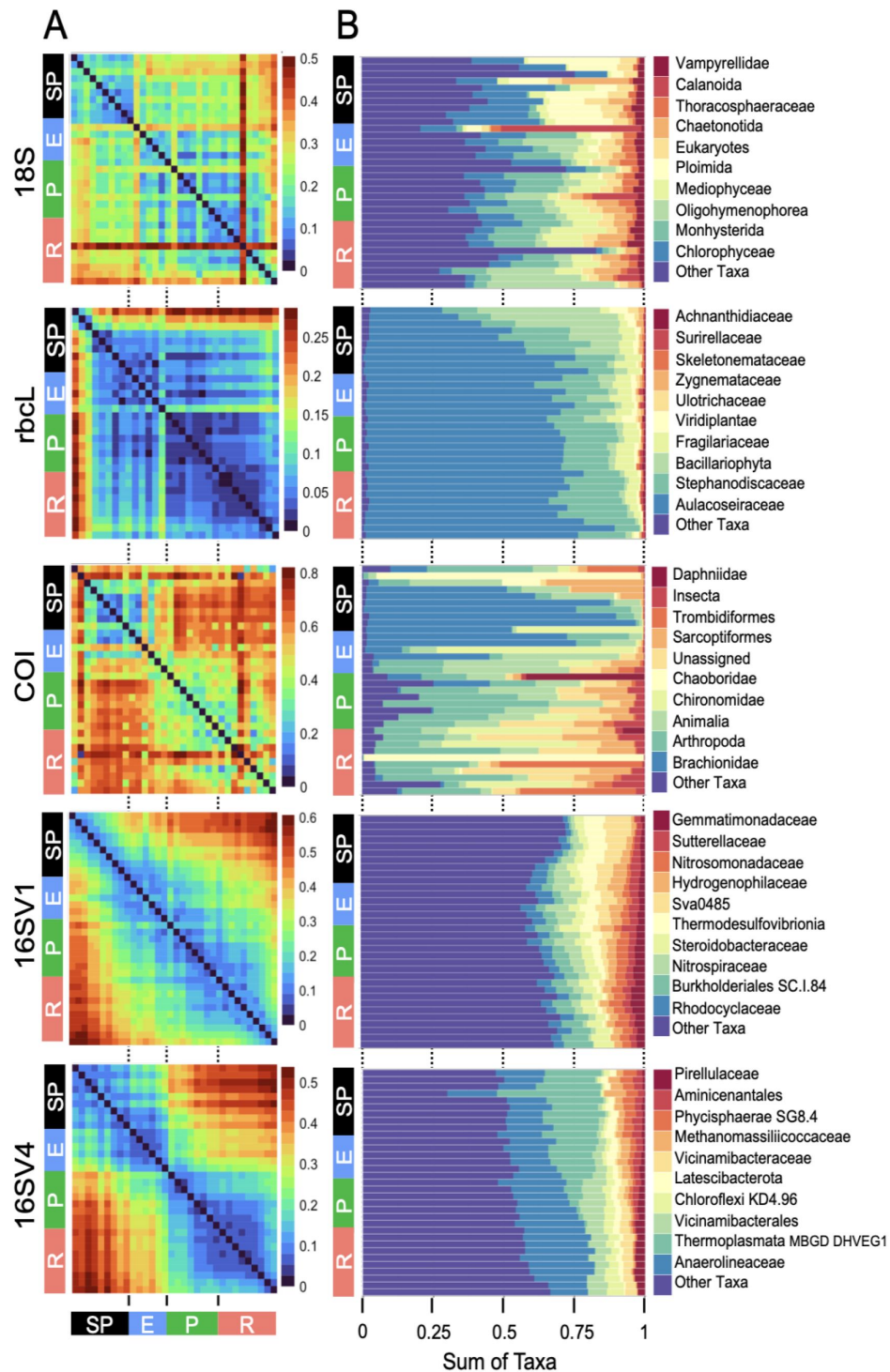


Figure 2.

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.

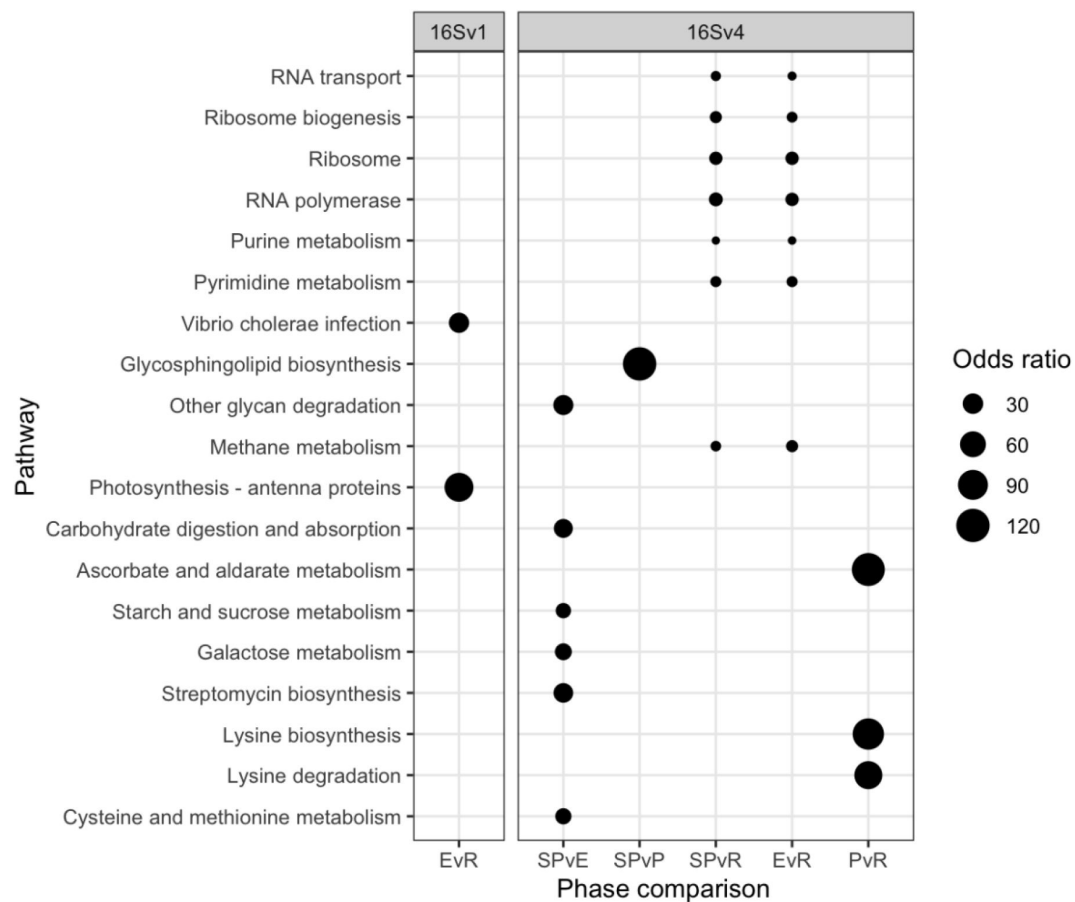


Figure 3.

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.

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 correlating 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**).

Discussion

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 skills²⁸. 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 resolved¹³. Recently, automated acquisition of microfossil data using artificial intelligence has been proposed as an alternative to human inspection for reconstructing long-term biological changes²⁹. 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 globe¹². 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 multidecadal biodiversity time series are limited to terrestrial and marine ecosystem, with freshwater ecosystems being marginally represented¹². Moreover, long-term freshwater studies tend to focus on indicator species or specific taxonomic groups (e.g. invertebrates), rather than capturing community-level patterns⁷. Developments in the field of *sedaDNA* have addressed the limitations of direct observations, utilising the properties of eDNA¹⁵. However, *sedaDNA* studies have predominantly focused on microorganisms as proxies for ecosystems' health (e.g. cyanobacteria³⁰; ciliates¹⁶; parasitic taxa³¹), 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 *sedaDNA* 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 observations¹⁶. 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

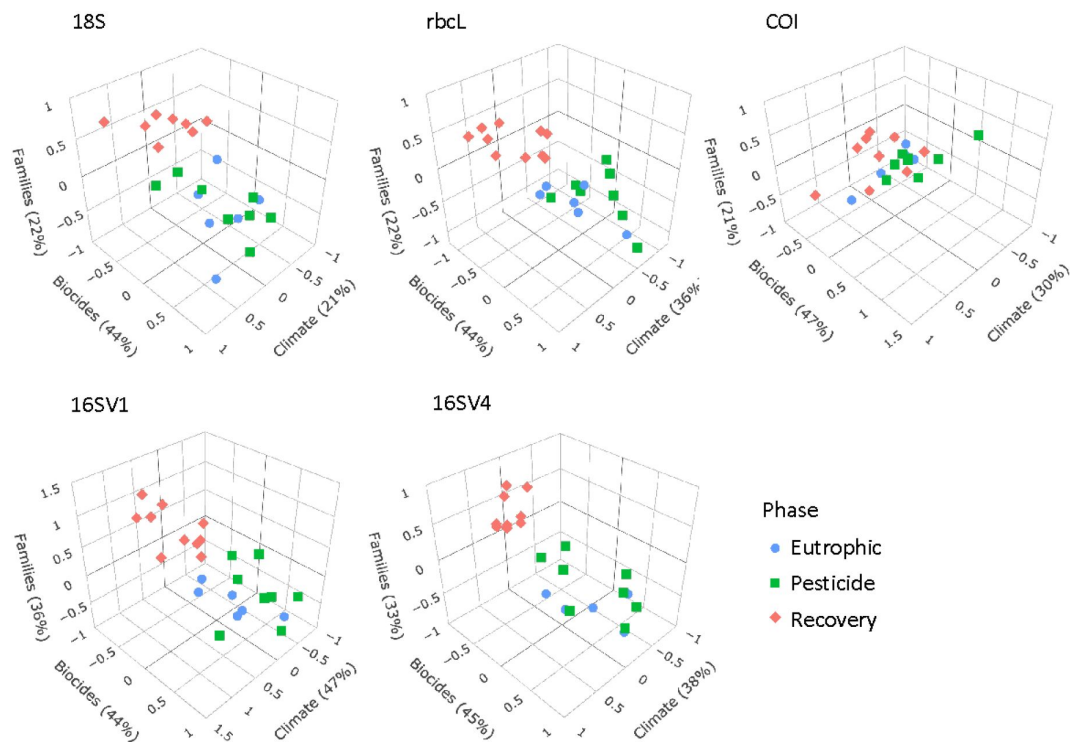


Figure 4.

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/

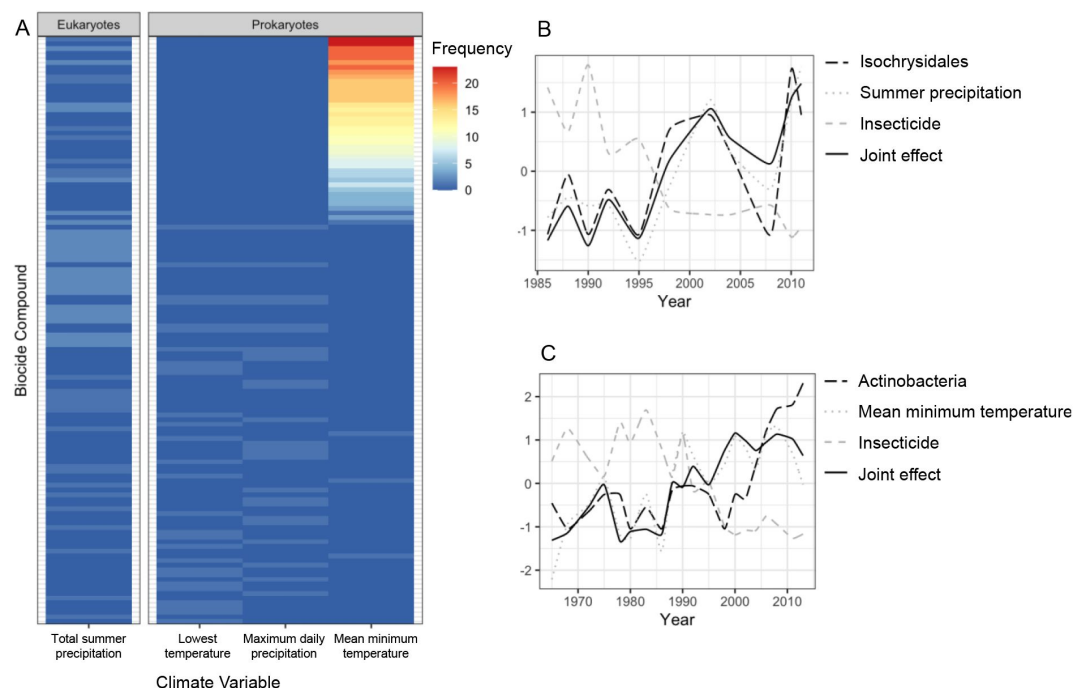


Figure 5.

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.

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 ³², biases might still exist due to the degradation of eDNA, especially over geological times ³³ and in warmer climates ³⁴. 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 ³³. 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°C. All these factors are known to reduce microbial activity, allowing a better preservation of DNA in sediment ³⁵.

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 ¹².

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 ⁴. 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 ²⁰. 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 *sedaDNA* on decennial timeframes focusing on prokaryotes (e.g. cyanobacteria ³⁶), whereas studies on eukaryotic compositional changes are just emerging to enable quantitative comparative assessments ³⁷. 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 ³⁸. However, longer-term experimental manipulations show a different perspective with irreversible changes in

biodiversity composition and function ³⁹. 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 ⁴⁰. Biomass and changes in biomass only capture productivity, while disregarding other metrics, such as decomposition or resource turnover ⁴¹. A complete assessment of biodiversity should include functionality ⁶. 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 ⁴². 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 ⁴³. 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 ¹⁹, whereas chemicals are thought to pose a negligible threat to biodiversity because living organisms can adapt and evolve ¹⁸. 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 ⁴⁴. 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 ⁴⁵. Higher temperatures increase diffusion of chemical molecules, resulting in faster uptake by living organisms and hence toxicity ⁴⁶. 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 approach⁴⁷). 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. ⁴⁸). 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 responses⁴⁸.

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 ¹⁰, providing a potentially transformative approach.

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 lacking⁴⁹, even if some countries have tried to introduce semi-quantitative indices to assess the ecological status of freshwater⁵⁰. 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 demonstrates¹⁹. 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 health⁵¹. 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.

Materials and Methods

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 impact²¹. 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 protocols⁵², 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 [52](#).

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 [21](#).

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 [53](#), [54](#). 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 [55](#).

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 - *sedaDNA* - 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), macroinvertebrates (COI), primary producers (focus on diatoms; *rbcL*), and prokaryotes (16SV1 and 16SV4) 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 96×96 dual tag barcoding to facilitate multiplexing and to reduce crosstalk between samples in downstream analyses 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 and multiqc. Sequences were then imported into QIIME2 v 2021.2, trimmed, filtered, merged and denoised using the QIIME2 DADA2 module 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; the diat.barcode v9.2 was used for the assignment of *rbcL* reads; and the Barcode of Life Database was used for the COI reads. The taxonomy was assigned using qiime feature-classifier classify-sklearn and used at family level where possible. 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 in R v4.0.2 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. 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 unifracs distances, were established with a PERMANOVA test. Alpha diversity was plotted using ggplot2 v3.3.5 with R v4.0.2. Heatmaps of weighted unifracs Beta diversity between each pair of sediment layers were plotted with the pheatmap v1.0.12 in R v4.0.2.

The function of the microbial communities across the four lake phases were predicted with PICRUST2 ⁷³ plugin in QIIME2 ⁶³, using the rarefied reads. Differentially abundant KEGG Orthology (KO) terms between pairs of lake phases were identified using the ANCOM plugin ⁷⁴ in QIIME2 ⁶³ and were mapped onto KEGG pathways with enriched pathways identified using a Fisher Exact test.

Drivers of biodiversity change

To identify correlations between biological assemblages (families identified through the *sedDNA* 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 ⁷⁵. 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 ⁷⁶.

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 ⁷⁷) 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 ⁷⁸. 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 a 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 $\text{padj} < 0.05$ (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 ; $\text{padj} < 0.05$; SWC with 10,000 permutations). Chemicals with more than 50% null values or Pearson correlation coefficients < 0.5 were discarded.

Data availability

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

Code availability

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

Acknowledgements

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, BioSequencing 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).

Author contributions

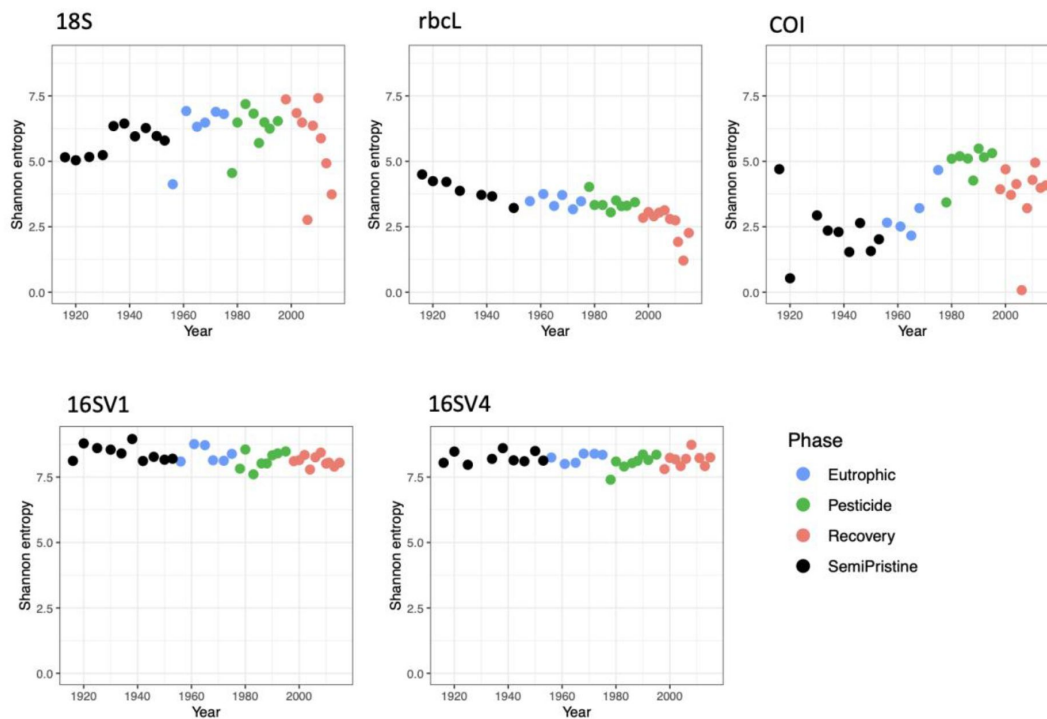
NE produced and analysed the metabarcoding data. JZ created the code and ran the machine learning analyses. RD completed preliminary bioinformatics analyses. MA-EA and WS generated the DDT data. YJ, SEC and HH optimised chemical assays. TAD provided the sedimentary archive, the climate and the biocides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

Competing interests

The authors declare no competing interests.

Supplementary information

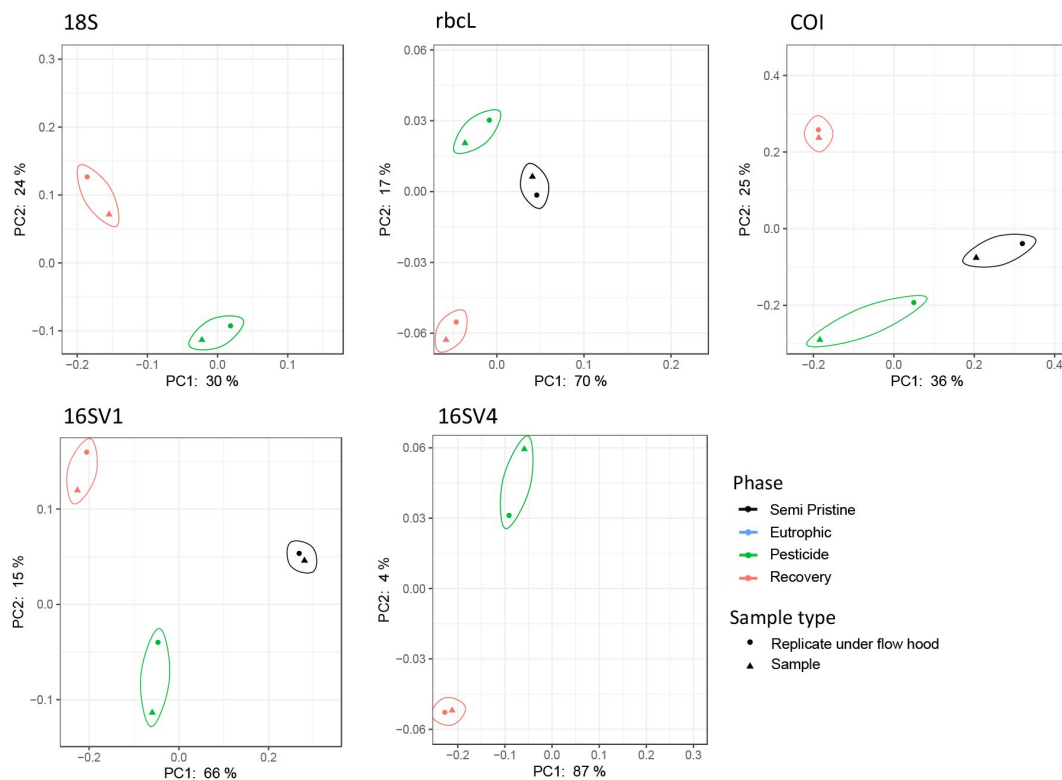
Supplementary Figures and Tables



Supplementary Figure 1.

Alpha diversity.

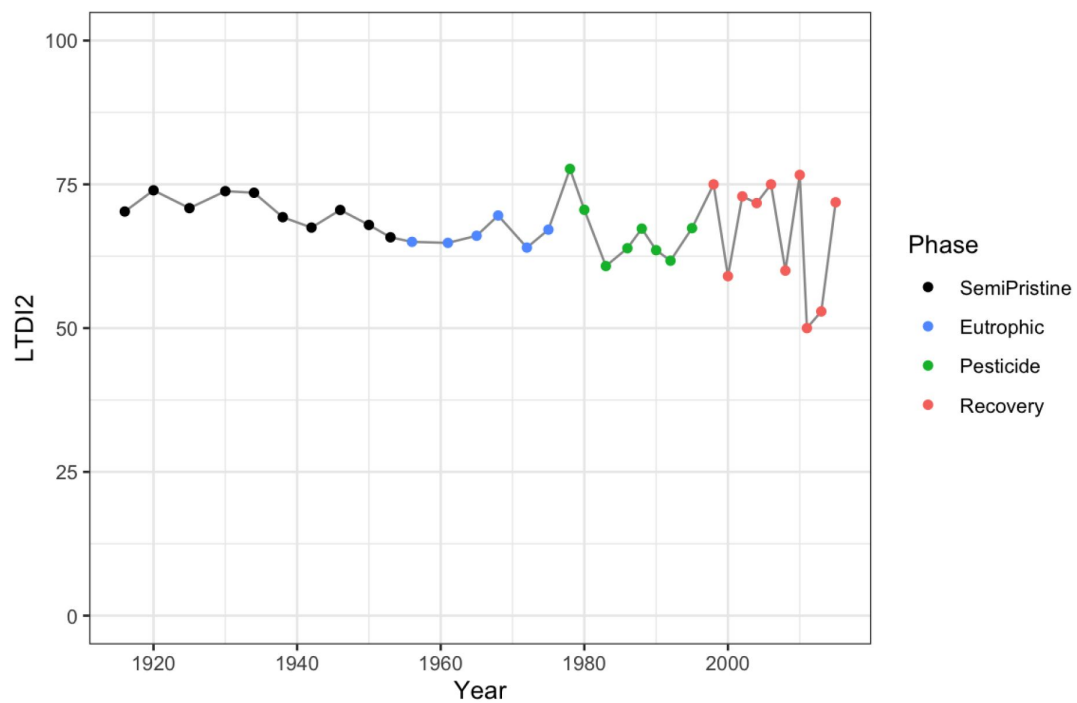
Alpha diversity, measured as Shannon entropy, is shown for the five barcodes used in this study (16SV1, 16SV4, 18S, COI and rbcL) between 1916-2016. The four lake phases are colour-coded as follows: Black - Semi-pristine; blue - Eutrophic; green - Pesticides; red - Recovery. Kruskal-Wallis test across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



Supplementary Figure 2.

Principal Coordinate Analysis.

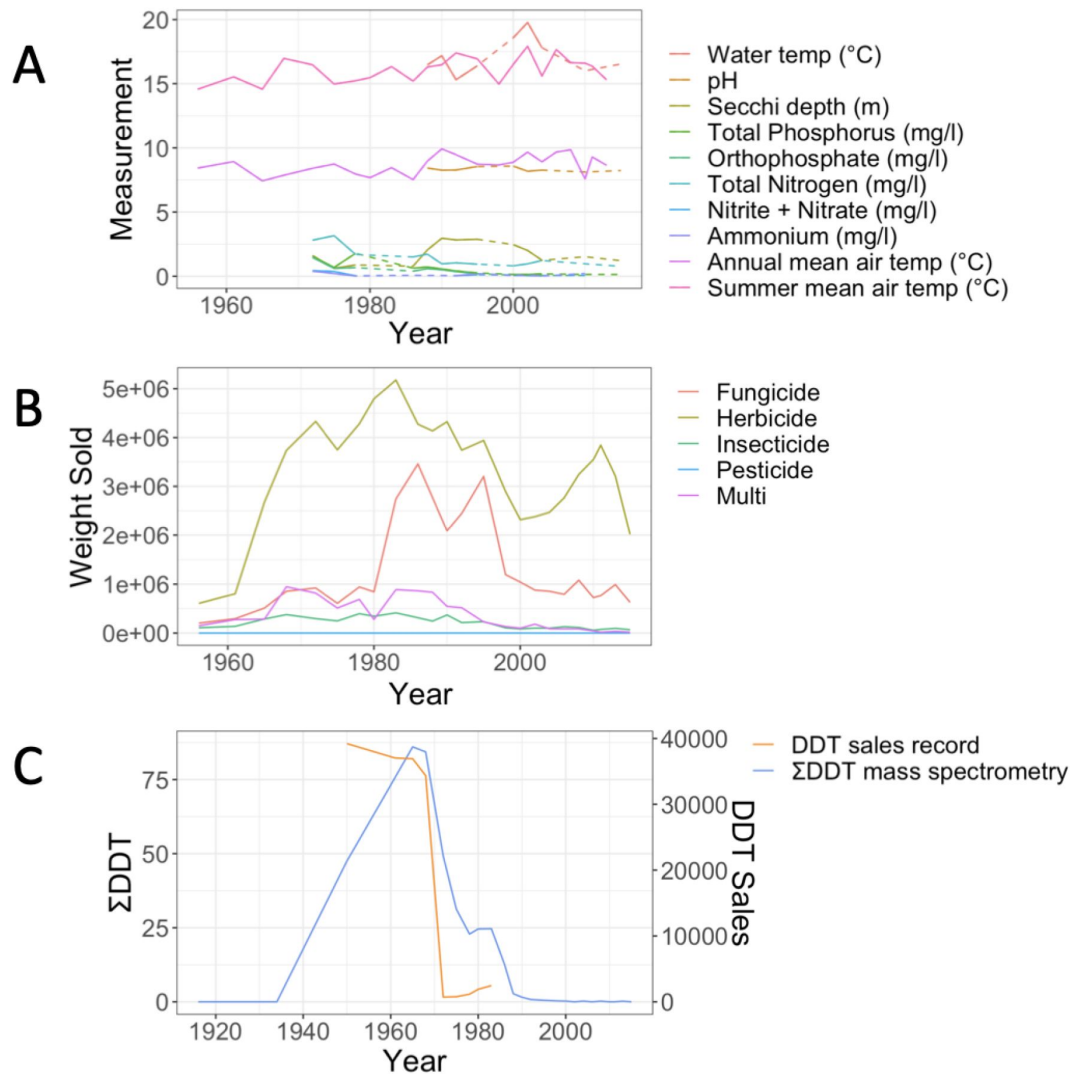
PCoA visualization of weighted unifrac distance between samples. Positive controls for PCR consist of duplicates of up to three samples from the sedimentary archive for each of the five barcodes used in the study (16SV1, 16SV4, 18S, rbcL and COI). Replicated samples are circled. The four lake phases 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.

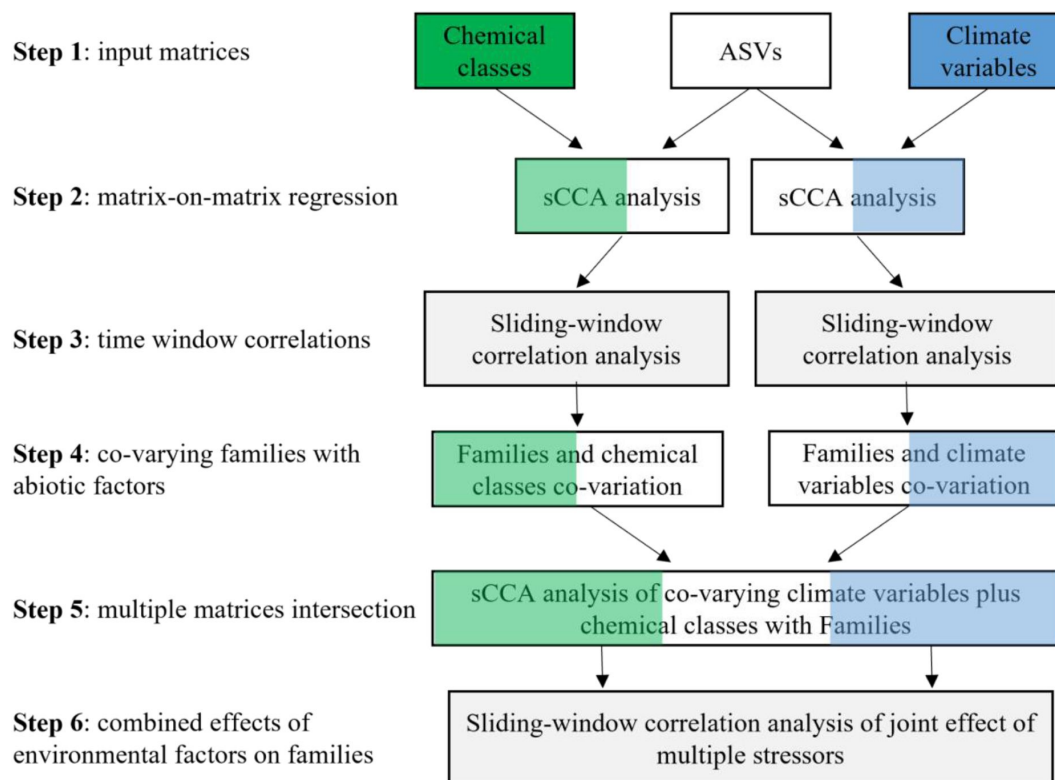


Figure 5.

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, biocides 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.

	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 temp	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	summer total precipitation	annual total precipitation	summer mean
		temperature			temperature

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.

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 Eastwood_etal_Supplementary Table 2

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 Eastwood_etal_Supplementary Table 3

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 pressure 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 Eastwood_etal_Supplementary Table 4

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 Eastwood_etal_Supplementary Table 5

References

- 1 Baert J. M., Janssen C. R., Sabbe K., De Laender F (2016) **Per capita interactions and stress tolerance drive stress-induced changes in biodiversity effects on ecosystem functions** *Nat Commun* **7** <https://doi.org/10.1038/ncomms12486>
- 2 Cardinale B. J., et al. (2012) **Biodiversity loss and its impact on humanity** *Nature* **486**:59–67 <https://doi.org/10.1038/nature11148>
- 3 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
- 4 Bonebrake T. C., et al. (2019) **Integrating Proximal and Horizon Threats to Biodiversity for Conservation** *Trends Ecol Evol* **34**:781–788 <https://doi.org/10.1016/j.tree.2019.04.001>
- 5 Ruckelshaus M. H., et al. (2020) **The IPBES Global Assessment: Pathways to Action** *Trends Ecol Evol* **35**:407–414 <https://doi.org/10.1016/j.tree.2020.01.009>
- 6 Eisenhauer N., et al. (2019) **A multitrophic perspective on biodiversity-ecosystem functioning research** *Adv Ecol Res* **61**:1–54 <https://doi.org/10.1016/bs.aecr.2019.06.001>
- 7 Dornelas M., et al. (2018) **BioTIME: A database of biodiversity time series for the Anthropocene** *Glob Ecol Biogeogr* **27**:760–786 <https://doi.org/10.1111/geb.12729>
- 8 Halpern B. S., et al. (2015) **Spatial and temporal changes in cumulative human impacts on the world's ocean** *Nat Commun* **6** <https://doi.org/10.1038/ncomms8615>
- 9 Rounsevell M. D. A., et al. (2020) **A biodiversity target based on species extinctions** *Science* **368**:1193–1195 <https://doi.org/10.1126/science.aba6592>
- 10 Eastwood N., et al. (2022) **The Time Machine framework: monitoring and prediction of biodiversity loss** *Trends Ecol Evol* **37**:138–146 <https://doi.org/10.1016/j.tree.2021.09.008>
- 11 Nogues-Bravo D., et al. (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>
- 12 Blowes S. A., et al. (2019) **The geography of biodiversity change in marine and terrestrial assemblages** *Science* **366**:339–345 <https://doi.org/10.1126/science.aaw1620>
- 13 Hirai J K. S., Kasai H, Nagai S (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**
- 14 Domaizon I., Winegardner A., Capo E., Gauthier J., Gregory-Eaves I (2017) **DNA-based methods in paleolimnology: New opportunities for investigating long-term dynamics of lacustrine biodiversity** *Journal of Paleolimnology* **58**:1–21
- 15 Capo E., et al. (2019) **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**

- 16 Barouillet C., et al. (2022) **Paleoreconstructions of ciliate communities reveal long-term ecological changes in temperate lakes** *Sci Rep* **12** <https://doi.org/10.1038/s41598-022-12041-7>
- 17 Bonebrake T. C., Christensen J., Boggs C. L., Ehrlich P. R (2010) **Population decline assessment, historical baselines, and conservation** *Conservation Letters* **3**
- 18 Groh K., Vom Berg C., Schirmer K., Tlili A (2022) **Anthropogenic Chemicals As Underestimated Drivers of Biodiversity Loss: Scientific and Societal Implications** *Environ Sci Technol* **56**:707–710 <https://doi.org/10.1021/acs.est.1c08399>
- 19 Pecl G. T., et al. (2017) **Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being** *Science* **355** <https://doi.org/10.1126/science.aai9214>
- 20 Gillson L., Marchant R (2014) **From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology** *Trends Ecol Evol* **29**:317–325 <https://doi.org/10.1016/j.tree.2014.03.010>
- 21 Cambronero Cuenca, et al. (2018) **Predictability of the impact of multiple stressors on the keystone species *Daphnia*** *Scientific Reports* **8**
- 22 Cambronero C. M., Beasley J., Kissane S., Orsini L (2018) **Evolution of thermal tolerance in multifarious environments** *Mol Ecol* **27**:4529–4541 <https://doi.org/10.1111/mec.14890>
- 23 Cuenca Cambronero M., Beasley J., Kissane S., Orsini L (2018) **Evolution of thermal tolerance in multifarious environments** *Mol Ecol* **27**:4529–4541 <https://doi.org/10.1111/mec.14890>
- 24 Cuenca Cambronero M., Zeis B., Orsini L (2018) **Haemoglobin-mediated response to hyper-thermal stress in the keystone species *Daphnia magna*** *Evol Appl* **11**:112–120 <https://doi.org/10.1111/eva.12561>
- 25 Baltrusaitis T., Ahuja C., Morency L. P (2019) **Multimodal Machine Learning: A Survey and Taxonomy** *IEEE Trans Pattern Anal Mach Intell* **41**:423–443 <https://doi.org/10.1109/TPAMI.2018.2798607>
- 26 Agency E (2020) **Phytobenthos - Diatoms for Assessing River and Lake Ecological Quality (River DARLEQ3)**
- 27 Bennion H., et al. (2014) **Assessment of ecological status in UK lakes using benthic diatoms** *Freshwater Science* **33**:639–654
- 28 Moraitis M. L., Valavanis V. D., Karakassis I (2019) **Modelling the effects of climate change on the distribution of benthic indicator species in the Eastern Mediterranean Sea** *Sci Total Environ* **667**:16–24 <https://doi.org/10.1016/j.scitotenv.2019.02.338>
- 29 Itaki T., et al. (2020) **Innovative microfossil (radiolarian) analysis using a system for automated image collection and AI-based classification of species** *Sci Rep* **10** <https://doi.org/10.1038/s41598-020-77812-6>
- 30 Picard M., et al. (2022) **Using metabarcoding and droplet digital PCR to investigate drivers of historical shifts in cyanobacteria from six contrasting lakes** *Sci Rep* **12** <https://doi.org/10.1038/s41598-022-14216-8>

- 31 Talas L., Stivirins N., Veski S., Tedersoo L., Kisand V (2021) **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** <https://doi.org/10.3390/microorganisms9040719>
- 32 Slon V., et al. (2022) **Extended longevity of DNA preservation in Levantine Paleolithic sediments, Sefunim Cave, Israel** *Sci Rep* **12** <https://doi.org/10.1038/s41598-022-17399-2>
- 33 Jia W., et al. (2022) **Preservation of sedimentary plant DNA is related to lake water chemistry** *Environmental DNA* **4**:425–439
- 34 Mauvisseau Q., et al. (2022) **The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments** *Environ Sci Technol* **56**:5322–5333 <https://doi.org/10.1021/acs.est.1c07638>
- 35 Giguët-Covex C., et al. (2014) **Long livestock farming history and human landscape shaping revealed by lake sediment DNA** *Nat Commun* **5** <https://doi.org/10.1038/ncomms4211>
- 36 Cao X., et al. (2020) **Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities** *Water Res* **183** <https://doi.org/10.1016/j.watres.2020.116077>
- 37 Zhang H., Huo S., Yeager K. M., Wu F (2021) **Sedimentary DNA record of eukaryotic algal and cyanobacterial communities in a shallow Lake driven by human activities and climate change** *Sci Total Environ* **753** <https://doi.org/10.1016/j.scitotenv.2020.141985>
- 38 Hillebrand H., Kunze C (2020) **Meta-analysis on pulse disturbances reveals differences in functional and compositional recovery across ecosystems** *Ecol Lett* **23**:575–585 <https://doi.org/10.1111/ele.13457>
- 39 Fordham D. A (2015) **Mesocosms Reveal Ecological Surprises from Climate Change** *PLoS Biol* **13** <https://doi.org/10.1371/journal.pbio.1002323>
- 40 Li F., et al. (2020) **Human activities' fingerprint on multitrophic biodiversity and ecosystem functions across a major river catchment in China** *Glob Chang Biol* **26**:6867–6879 <https://doi.org/10.1111/gcb.15357>
- 41 Gounand I., Little C. J., Harvey E., Altermatt F (2018) **Cross-ecosystem carbon flows connecting ecosystems worldwide** *Nat Commun* **9** <https://doi.org/10.1038/s41467-018-07238-2>
- 42 Choi J., et al. (2017) **Strategies to improve reference databases for soil microbiomes** *ISME J* **11**:829–834 <https://doi.org/10.1038/ismej.2016.168>
- 43 Mi H., Muruganujan A., Ebert D., Huang X., Thomas P. D (2019) **PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools** *Nucleic Acids Res* **47**:D419–D426 <https://doi.org/10.1093/nar/gky1038>
- 44 Cuenca-Cambronero M., et al. (2021) **Evolutionary mechanisms underpinning fitness response to multiple stressors in *Daphnia*** *Evol Appl* **14**:2457–2469 <https://doi.org/10.1111/eva.13258>

- 45 Pinheiro J. P. S., Windsor F. M., Wilson R. W., Tyler C. R (2021) **Global variation in freshwater physico-chemistry and its influence on chemical toxicity in aquatic wildlife** *Biol Rev Camb Philos Soc* **96**:1528–1546 <https://doi.org/10.1111/brv.12711>
- 46 Patra R. W., Chapman J. C., Lim R. P., Gehrke P. C., Sunderam R. M (2015) **Interactions between water temperature and contaminant toxicity to freshwater fish** *Environmental Toxicology and Chemistry* **34**:1809–1817
- 47 Yang P., et al. (2020) **Correlation-Centric Network (CCN) representation for microbial co-occurrence patterns: new insights for microbial ecology** *NAR Genom Bioinform* **2** <https://doi.org/10.1093/nargab/lqaa042>
- 48 Tse T. J., et al. (2018) **Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial Communities** *Environ Sci Technol* **52**:6842–6853 <https://doi.org/10.1021/acs.est.7b06386>
- 49 Solimini A. G., Cardoso A. C., Haiskanen A.-S (2005) **Linkages between chemical and biological quality of surface waters** :1–262
- 50 Archaimbault V., Dumont B (2010) **The normalized global biological index (IBGN): principles and evolution within the framework of the European framework directive on water** *Sciences Eaux & Territoires* <https://doi.org/10.14758/SET-REVUE.2010.1.08>
- 51 Kanno J (2016) **Introduction to the concept of signal toxicity** *J Toxicol Sci* **41**:SP105–SP109 <https://doi.org/10.2131/jts.41.SP105>
- 52 Appleby P. G (2001) **Chronostratigraphic techniques in recent sediments**
- 53 Livingstone D. M., Lotter A. F (1998) **The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palaeolimnological implications** *Journal of Paleolimnology* **19**:181–198
- 54 Preudhomme E. B., Stefan H. G (1992) **Relationship between water temperatures and air temperatures for central U.S. streams**
- 55 Wong F., Robson M., Diamond M. L., Harrad S., Truong J (2009) **Concentrations and chiral signatures of POPs in soils and sediments: a comparative urban versus rural study in Canada and UK** *Chemosphere* **74**:404–411 <https://doi.org/10.1016/j.chemosphere.2008.09.051>
- 56 Hadziavdic K., et al. (2014) **Characterization of the 18S rRNA gene for designing universal eukaryote specific primers** *PLoS One* **9** <https://doi.org/10.1371/journal.pone.0087624>
- 57 Leray M., et al. (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** <https://doi.org/10.1186/1742-9994-10-34>
- 58 Zimmermann J., et al. (2014) **Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research** *PLoS One* **9** <https://doi.org/10.1371/journal.pone.0108793>
- 59 Caporaso J. G., et al. (2011) **Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample** *P Natl Acad Sci USA* **108**:4516–4522 <https://doi.org/10.1073/pnas.1000080107>

- 60 MacConaill L. E., et al. (2018) **Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing** *BMC Genomics* **19** <https://doi.org/10.1186/s12864-017-4428-5>
- 61 Wingett S. W., Andrews S (2018) **FastQ Screen: A tool for multi-genome mapping and quality control** *F1000Res* **7**
- 62 Ewels P., Magnusson M., Lundin S., Kaller M (2016) **MultiQC: summarize analysis results for multiple tools and samples in a single report** *Bioinformatics* **32**:3047–3048 <https://doi.org/10.1093/bioinformatics/btw354>
- 63 Bolyen E., et al. (2019) **Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2** *Nat Biotechnol* **37**:852–857 <https://doi.org/10.1038/s41587-019-0209-9>
- 64 Callahan B. J., et al. (2016) **DADA2: High-resolution sample inference from Illumina amplicon data** *Nat Methods* **13**:581–583 <https://doi.org/10.1038/nmeth.3869>
- 65 Yilmaz P., et al. (2014) **The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks** *Nucleic Acids Res* **42**:D643–648 <https://doi.org/10.1093/nar/gkt1209>
- 66 Rimet F., et al. (2019) **Diat.barcode, an open-access curated barcode library for diatoms** *Sci Rep* **9** <https://doi.org/10.1038/s41598-019-51500-6>
- 67 Robeson M. S (2021) **2nd et al. RESCRIPt: Reproducible sequence taxonomy reference database management.** *PLoS Comput Biol* **17** <https://doi.org/10.1371/journal.pcbi.1009581>
- 68 Pedregosa F., et al. (2011) **Scikit-learn: Machine Learning in Python** *Journal of Machine Learning Research* **12**:2825–2830
- 69 Wickham H (2016) **ggplot2: Elegant Graphics for Data Analysis**
- 70 Team R. C (2020) **Team, R. C. (Vienna, Austria, 2020).**
- 71 Anderson M. J (2001) **A new method for non-parametric multivariate analysis of variance** *Austral Ecology* **26**:32–46
- 72 Kolde R (2019) **Kolde, R. Pretty Heatmaps, <<https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>> (2019).**
- 73 Douglas G. M., et al. (2020) **PICRUSt2 for prediction of metagenome functions** *Nat Biotechnol* **38**:685–688 <https://doi.org/10.1038/s41587-020-0548-6>
- 74 Mandal S., et al. (2015) **Analysis of composition of microbiomes: a novel method for studying microbial composition** *Microb Ecol Health Dis* **26** <https://doi.org/10.3402/mehd.v26.27663>
- 75 Lin D., et al. (2013) **Group sparse canonical correlation analysis for genomic data integration** *BMC Bioinformatics* **14** <https://doi.org/10.1186/1471-2105-14-245>
- 76 Parkhomenko E., Tritchler D., Beyene J (2009) **Sparse Canonical Correlation Analysis with Application to Genomic Data Integration** *Statistical Applications in Genetics and Molecular Biology* **8**:1–34

- 77 Nakagawa S., Cuthill I. C (2007) **Effect size, confidence interval and statistical significance: a practical guide for biologists** *Biol Rev Camb Philos Soc* **82**:591–605 <https://doi.org/10.1111/j.1469-185X.2007.00027.x>
- 78 Buckland M., Gey F (1994) **The relationship between recall and precision** *Journal of the American society for information science* **45**:12–19

Author information

Niamh Eastwood

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

For correspondence: NXE435@student.bham.ac.uk

ORCID iD: [0000-0003-2969-6091](https://orcid.org/0000-0003-2969-6091)

Jiarui Zhou

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

ORCID iD: [0000-0002-1025-718X](https://orcid.org/0000-0002-1025-718X)

Romain Derelle

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

Mohamed Abou-Elwafa Abdallah

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

ORCID iD: [0000-0002-4624-4073](https://orcid.org/0000-0002-4624-4073)

William A. Stubbings

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

ORCID iD: [0000-0002-8538-4693](https://orcid.org/0000-0002-8538-4693)

Yunlu Jia

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

Sarah E. Crawford

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

ORCID iD: [0000-0003-4341-701X](https://orcid.org/0000-0003-4341-701X)

Thomas A. Davidson

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

ORCID iD: [0000-0003-2326-1564](https://orcid.org/0000-0003-2326-1564)

John K. Colbourne

Environmental Genomics Group, School of Biosciences, the University of Birmingham,
Birmingham B15 2TT, UK
ORCID iD: [0000-0002-6966-2972](https://orcid.org/0000-0002-6966-2972)

Simon Creer

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

Holly Bik

Department Marine Sciences and Institute of Bioinformatics, University of Georgia, Athens,
Georgia, USA
ORCID iD: [0000-0002-4356-3837](https://orcid.org/0000-0002-4356-3837)

Henner Hollert

Department Evolutionary Ecology & Environmental Toxicology, Faculty of Biological Sciences,
Goethe University Frankfurt, Max-von-Laue Straße 13, 60438 Frankfurt am Main, Germany,
LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25,
60325 Frankfurt am Main, Germany, Department Media-related Toxicology, Institute for
Molecular Biology and Applied Ecology (IME), Max-von-Laue Straße 13, 60438 Frankfurt am
Main, Germany
ORCID iD: [0000-0001-5776-5619](https://orcid.org/0000-0001-5776-5619)

Luisa Orsini

Environmental Genomics Group, School of Biosciences, the University of Birmingham,
Birmingham B15 2TT, UK, The Alan Turing Institute, British Library, 96 Euston Road, London
NW1 2DB, UK
ORCID iD: [0000-0002-1716-5624](https://orcid.org/0000-0002-1716-5624)

Editors

Reviewing Editor

David Donoso

Escuela Politécnica Nacional, Ecuador

Senior Editor

Detlef Weigel

Max Planck Institute for Biology Tübingen, Germany

Joint Public Review

Introduction - Well written and placed within the current trends of unprecedented biodiversity loss, with emphasis in freshwater ecosystems. The authors identify three important points as to why biodiversity action plans have failed. Namely, community changes occur over large spatio-temporal scales and monitoring programs capture a fraction of these long-term dynamics (e.g. few decades) which although good at capturing trends in biodiversity change, they often fail at identifying the drivers of these changes. Additionally, most of these rely on manual sorting of samples, overlooking cryptic diversity, or state-of-the-art techniques such as sedimentary DNA (sedaDNA) which allow studying decade-long dynamics, usually focus on specific taxonomic groups unable to represent community-level changes. Secondly, the authors identify that biodiversity is threatened by multiple factors and

are rarely studied in tandem. Finally, the authors stress the need for high-throughput approaches to study biodiversity changes since historically, most conservation efforts rely on highly specialized skills for biodiversity monitoring, and even well studied species have relatively short time series data. The authors identify a model freshwater lake (Lake Ring, Denmark) - suitable due to its well documented history over the last 100 years - to present a comprehensive framework using metabarcoding, chemical analysis and climatic records for identifying past and current impacts on this ecosystem arising from multiple abiotic environmental stressors.

Results - They are brief and should expand some more. Particularly, there are no results regarding metabarcoding data (number of reads, filtering etc.). These details are important to know the quality of the data which represents the bulk of the analyses. Even the supplementary material gives little information on the metabarcoding results (e.g. number of ASVs - whether every ASV of each family were pooled etc.). The drivers of biodiversity change section could be restructured and include main text tables showing the families positively or negatively correlated with the different variables (akin to table S2 but simplified).

Discussion

The discussion is well written identifying first, some of the possible caveats of this study, particularly regarding classification of metabarcoding data, its biases and the possible DNA degradation of ancient sediment DNA. The authors discuss how their results fit to general trends showing how agricultural runoff and temperature drive changes in freshwater functional biodiversity primarily due to their synergistic effects on bioavailability, adsorption, etc. The authors highlight the advantage of using a system-level approach rather than focusing on taxa-specific studies due to their indicator status. Similarly, the authors justify the importance of studying community composition as far back as possible since it reveals unexpected patterns of ecosystem resilience. Lake Ring, despite its partially recovered status, has not returned to its semi-pristine levels of biodiversity and community assemblage. Additionally, including enzyme activity allows to assess functional diversity of the studied environment although reference databases of these pathways are still lacking. Finally, the authors discuss the implications of their findings under a conservation and land management framework suggesting that by combining these different approaches, drivers of biodiversity stressors can be derived with high accuracy allowing for better informed mitigation and conservation efforts.

Author Response

The following is the authors' response to the original reviews.

We thank the reviewers for the constrictive and detailed feedback provided. We have adopted the proposed changes to improve the manuscript clarity and accessibility. The following revisions are included in the revised manuscript:

Reviewer #1 (Public Review):

The analytical framework is not sufficiently explained in the main text.

We think the reviewer is referring to the conceptual framework mentioned in introduction. In the previously submitted manuscript, we did not provide details because the framework is published elsewhere. However, we agree with the reviewer that a short explanation may be helpful, which we have included in the resubmitted manuscript.

The significance of findings in relation to functional changes is not clear. What are the consequences of enrichment of RNA transport or ribosome biogenesis pathways between pesticides and recovery stages, for example?

We thank the reviewer for this suggestion. In the previously submitted manuscript, we included an explanation of the central functions these pathways can alter (e.g. metabolism and infection response). These functions are self-explanatory. However, we have elaborated on the consequence that the disruption of these pathways can cause in the resubmitted manuscript.

The impact of individual biocides and climate variables, and their additive effects, are assessed but there is no information offered on non-additive interactions (e.g., synergistic, antagonistic).

This was a misunderstanding based on our use of the term synergistic in this context. The approach by which we define a synergistic or joint effect of two environmental variables on a taxonomic group is explained in the methods section. This analysis is based on climate variables and biocide types contributing the largest covariances in the correlation analysis explained in Supplementary Fig. 5; Step 4. The combined effect of two environmental variables on a taxon was considered to be significant if the biocide type and the climate variable were each significantly correlated with the taxon over the same time window, and their average Pearson correlation was > 0.5 with $\text{padj} < 0.05$ (SWC analysis with 10,000 permutations). The biocide type and the climate variable were interpreted to have a joint effect on a given taxon 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 $\text{padj} < 0.05$ (with 10,000 permutations in the SWC analysis). We realise that the use of synergistic or additive was not correct in this context and have replaced the term synergistic with joint effect throughout the manuscript.

The level of confidence associated with results is not made explicit. The reader is given no information on the amount of variability involved in the observations, or the level of uncertainty associated with model estimates.

As we didn't use traditional statistical approaches, confidence level estimation in the traditional sense is not possible. Instead, we used permutation tests and adjusted P-values to identify significant correlations in the data. These approaches are more robust than traditional statistics for integrating and discovering complex, group-wise patterns among high-dimensional datasets. 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.

The major implications of the findings for regulatory ecological assessment are missed. Regulators may not be primarily interested in identifying past "ecosystem shifts". What they need are approaches which give greater confidence in monitoring outcomes by better reflecting the ecological impact of contemporary environmental change and ecosystem management. The real value of the work in this regard is that: (1) it shows that current approaches are inappropriate due to the relatively stable nature of the indicators used by regulators, despite large changes in pollutant inputs; (2) it presents some better alternatives, including both taxonomic and functional indicators; and (3) it provides a new reference (or baseline) for regulators by characterizing "semi-pristine" conditions.

We thank the reviewer for this suggestion, which we have included in the main text (L451461)

Reviewer #2 (Public Review):

Results - They are brief and should expand some more. Particularly, there are no results regarding metabarcoding data (number of reads, filtering etc.). These details are important to know the quality of the data which represents the bulk of the analyses. Even the supplementary material gives little information on the metabarcoding results (e.g. number of ASVs - whether every ASV of each family were pooled etc.).

We thank the reviewer for this suggestion. We have included a paragraph in results reporting read numbers and other statistics. The filtering criteria and handling of samples can be found in methods (L658-661; L670-675). As explained in methods the taxonomy was assigned using qiime feature-classifier classify-sklearn and used at family level where possible. When classification was not possible at family level because of incomplete/missing information in the online database or a poor match to reference database, the lowest classification possible was used.

The drivers of biodiversity change section could be restructured and include main text tables showing the families positively or negatively correlated with the different variables (akin to table S2 but simplified).

As there are over 180 unique families/taxonomic units correlated with at least one biocide or environmental variable, a simplified version of this table would be too large to include in the main text. Therefore, we prefer to keep this information in supplementary table 2 complete with correlation statistics.

We thank the reviewers for providing detailed feedback on the manuscript and respond to their suggestions as follows:

Reviewer #1 (Recommendations For The Authors):

Thank you for the opportunity to review your manuscript, which I found interesting and enjoyable to read. Here are some suggestions for improving it.

Remove spaces before citations in text.

Lines 51-53: "Community-level biodiversity reliably explained freshwater ecosystem shifts whereas traditional quality indices (e.g. Trophic Diatom Index) and physicochemical parameters proved to be poor metrics for these shifts." Seems to be the wrong way around / not clear???

Rephrased to clarify.

Line 54: Should be "...advocates the use of..." or "...demonstrates the advantages of..."

Done, thanks for the suggestion.

Line 62: Spell out numbers <10, i.e. "sixth mass extinction"

Done, thank you.

Lines 66-72: These sentences lack clarity. It's not clear that "experimental manipulation of biodiversity" hasn't involved investigation of "multi-trophic changes". By the third of these four sentences it is not clear what "they" is referring to. And in the fourth sentence, "these holistic studies" are not defined. Perhaps it would suffice to say that experiments have so far focused primarily on a single trophic level and largely neglected freshwater systems.

We have rephrased to improve clarity.

Line 81: Delete unnecessary bracket

Done, thank you.

Line 82: "a minority of freshwater ecosystems" sounds as if you're saying that few freshwater ecosystems are represented in BioTIME, which seems obvious and would also apply to terrestrial and marine systems. Do you mean that freshwater ecosystems are not well represented in the data?

We have clarified the sentence, thanks.

Line 106: Resolve issue with citation in text at the end of the sentence (repeated at line 109 and possibly other lines).

Done, thank you.

Line 116: By ">1999s" do you mean 1990s?

This was a typo. it was supposed to be >1999

Line 120: The reader would benefit greatly from a brief explanation of explainable network models and multimodal learning in the introduction. Why are these the right tools to use? How do they work in this context? Figure 1 helps to some extent but needs more commentary in the text.

We have included an explanation of the explainable network models and multimodal learning and how their use can be beneficial to the study of diverse data types.

Line 144: Here and throughout the text the language could be much more efficient and readable. "Alpha diversity" does not require a definite article. Furthermore, when referring to significance it is convention to state the p-value, test statistic and test used.

As there are different p-values for each barcode, we have included them in legend to Supplementary Fig. 1 to avoid crowding the main text. We prefer to leave the text unchanged for this reason.

Line 155: "The primary producer's composition" is grammatically awkward and less suitable than "the composition of primary producers". This kind of awkwardness occurs again at line 285 ("diatom's") and possibly in other parts of the manuscript.

Thanks, corrected.

Line 169: The statement that this family was "relatively more abundant" needs a little more explanation. What is it relative to - other groups or to previous stages?

More abundant than in the other phases – the sentence has been modified.

Line 179: Nested brackets are unnecessary and affect readability. This could simply be a new sentence, i.e. "For example, Nitrospiraceae (nitrite oxidizers)..."

Done, thanks.

Line 215: "Functional biodiversity", which implies that some biodiversity is functional and some not, does not seem an appropriate term to describe the results you present in this section. Simply "functioning of the prokaryotic community" would suffice.

Thanks, done.

Line 214-233: This section may be inaccessible for many readers. For example, what are Kegg Orthologs and what role do they play in the functioning of a lake ecosystem? The explanation comes later in the paragraph but there needs to be a gentler introduction before diving into specific technical concepts.

We appreciate this comment and have included a short explanation of what KEGG and KO terms mean.

Supplementary Figure 3: It would be helpful to superimpose the lake stages here, as done in Figure 2.

The figure has been updated with coloured data points corresponding to each phase, as in supplementary figure 1.

Line 265: Should be "19 of which were identified..."

Done, thanks.

Line 284: "Predominantly" rather than "prominently"?

Done

Line 242-316: This section is good in that it identifies and ranks individual biocides and climate variables but there is no information on non-additive interactions (e.g., synergistic, antagonistic). Could the authors at least comment on why this was not done or not necessary, and what uncertainties this omission could introduce into the results?

This was a misunderstanding based on our use of the term synergistic in this context. the approach by which we define a synergistic or joint effect of two environmental variables on a taxonomic group is explained in the methods section. This analysis is based on climate variables and biocide types contributing the largest covariances in the correlation analysis explained in Supplementary Fig. 5; Step 4. The combined effect of two environmental variables on a taxon was considered to be significant if the biocide type and the climate variable were each significantly correlated with the taxon over the same time window, and their average Pearson correlation was > 0.5 with $\text{padj} < 0.05$ (SWC analysis with 10,000 permutations) – this is shown in Supplementary Fig. 5; Step 6. The biocide type and the climate variable were interpreted to have an additive effect on a given taxon 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 $\text{padj} < 0.05$ (with

10,000 permutations in the SWC analysis). we have replace synergistic with joint effect to avoid confusion.

Figure 4: These 3-D plots are very hard to read. Without additional features (e.g. shadows on each plane, or lines connecting points to planes) it is impossible for the viewer to tell where the points are located on each axis.

We have created interactive 3D plots here: https://environmental-omicsgroup.github.io/Biodiversity_Monitoring/.

Figure 5: Legend entry should be "summer precipitation" not "precipitations". "Additive effect" rather than "joint effect" would be more consistent with the main text.

“Precipitations” has been updated to “precipitation” where relevant throughout. We left ‘joint effect’ and unified the main text, responding to a previous comment of this reviewer on the meaning of synergistic effects in our study.

Line 348: Doesn't your approach also require specialist skills? I often feel that the "traditional" versus "molecular" monitoring debate misses this point. Some comment on the training and development needs for those interested in applying the sedaDNA approach would be welcome. Otherwise it is an unfair comparison.

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 sedaDNA or more generally eDNA can be outsourced to established environmental services, removing the need for training if it is a limiting factor. The above has been included in discussion.

Line 391: "Significantly did" what? "Did significantly change over time" would be better.

Done, thanks.

Line 407: Should be "an indicator of..." and "did not significantly change over time..."

Done, thanks.

Line 408-410: Regulators are not necessarily interested in identifying past "ecosystem shifts", so this does not seem to be the best way to contrast the capabilities of the sedaDNA approach with those of LTDI2. The real value of this work, in my opinion, is threefold. First, it shows that the reliance on diatoms as indicators of ecological status is inappropriate due to the relatively stable nature of diatom communities in the face of large environmental changes. Second, it presents some better alternatives, including both taxonomic and functional indicators. And third, it provides a new reference point for regulators by characterising "semi-pristine" conditions.

Thanks for the insightful suggestion. We agree with the reviewer on the advantages and have spelled them out in the resubmitted manuscript.

Line 445: What are "housekeeping functions"? I checked the Cuenca-Cambronero paper cited but did not find the term there.

Housekeeping functions are essential basic cellular functions that are evolutionary conserved. They are more commonly present in public databases because they have been

characterised in a number of model species (e.g. *Drosophila*, *C. elegans* and *Mus musculus*). Our reference is not to the Cuenca-Cambronero paper, but to Mi et al, describing the reference database PANTHER. We included the definition of housekeeping functions in the main text.

| *Line 449: Briefly state the main functional changes found here.*

Examples have been included.

| *Lines 451-452: Whilst this statement may be found in the cited source, most readers I suspect would not identify with it. Indeed, one could argue that most of freshwater ecology has been dedicated to this very task (documenting chemical impacts on biodiversity)! A more balanced view is needed here.*

The sentence the reviewer refers to includes also reference to climate change. Climate change and chemical pollution are the two most common causes of biodiversity loss, and not only in freshwater ecosystems.

| *Lines 463-466: These examples both point to non-additive (synergistic) effects, which were not assessed in the current study.*

Please refer to our explanation above about the inappropriate use of synergistic and, here, additive. We have altered the text throughout to use joint effects as we do not investigate synergistic, antagonistic and additive effects as traditionally described in ecology.

| *Lines 472-474: This sentence is unclear. Do you mean that this approach surpasses others in terms of reliability? If so, I don't believe this has been demonstrated in the paper.*

We apologise. The word 'reliability' should have not been in the text. We have improved the clarity of this sentence.

| *Lines 474-482: In these sentences it is unclear whether or not you are talking about your method or contrasting it with another method(s). If the latter, which method or methods are you referring to?*

We have fixed this sentence to better reflect that 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).

| *Line 631: Should be "Physico-chemical variables". I have not extensively checked the rest of the methods for such errors.*

Thank you, the text has been changed where present.

| **Reviewer #2 (Recommendations For The Authors):**

| *Introduction Line 80 remove extra ')*

Done, thank you.

| *Line 81 rephrase e.g includes few freshwater ecosystems*

We modified this sentence also following Reviewer #1

| *Line 83 although, instead of whereas?*

Done, thanks.

| *Line 106 formatting reference issue*

| *Line 109 same as above*

Thank you, noted.

| *Results*

| *Line 141 - 144 how was the sampling of the sediment performed over the 100 year core? Every year? Every 5 years? Or were they pooled to represent the (as of yet unlisted) phases?*

The reviewer is correct that details are not provided here. They are in methods. We have added some text to explain the basic concepts of how the core was obtained and sliced and refer the reader to the method section for more details.

| *Line 154 the authors have not yet explicitly listed the lake phases, so it is difficult to refer to them now.*

Noted, the addition of a short explanation at the beginning of the results section should take care of this issue.

| *Line 216 - may be worth briefly explaining KEGG orthologs and how these relate to functional biodiversity.*

We thank the reviewer. Also responding to a similar comment from Reviewer #1, we included a description of KO terms and their links to functional biodiversity.

| *Lines 249 - 260 instead of a supplementary table, it could remain in the main text*

Supplementary table 2 is a multi-tab table including information for each region amplified here. It is not possible to include this table in the main text.

| *Materials and Methods Due to the formatting of the manuscript (results & discussion before materials and methods), many of the results are not clearly understood without having to visit the M&M section. Particularly, how the biocide types were obtained (Historic records plus persistence of DDT in sediments). This could be resolved by including a few sentences on how the data was gathered in the results section. Overall, materials and methods are sufficient, however, it is not clear how many of the 37 metabarcoding samples correspond to which of the lake phases. Finally, I suggest a better organization of M&Ms by having subheadings for each section. For example, under Biodiversity fingerprinting across 100 years, one subheading could be DNA extraction and sequencing, another subheading could be bioinformatics.*

We thank the reviewer for the suggestion. To alleviate the issues linked to the methods section coming after the results section, we have introduced a short explanation of the sediments core and the lake phases at the beginning of the results section. A description of the climate and chemical data has been included at the beginning of the section 'Drivers of biodiversity change' in results. Subheadings were introduced in methods as suggested.