100 years of anthropogenic impact causes changes in freshwater functional biodiversity

Niamh Eastwood*[1], Jiarui Zhou*[2], Romain Derelle[1], Mohamed Abou-Elwafa Abdallah[3], William A. Stubbings[3], Yunlu Jia[4], Sarah E. Crawford[4], Thomas A. Davidson[5], John K. Colbourne[1], Simon Creer[6], Holly Bik[7], Henner Holler[4,8,9] and Luisa Orsini[1, 10]

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

[2] Computer Science, University of Birmingham, Birmingham, UK

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

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

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

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

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

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

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

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

*These authors contributed equally

Corresponding author:
Niamh Eastwood
Environmental Genomics Group
NXE435@student.bham.ac.uk

Keywords: sedaDNA, machine learning, freshwater, multilocus metabarcoding, functional biodiversity

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 precipitations, explained up to 90% of the functional biodiversity changes. 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.

Our study advocates the advantage 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.

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 6th 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 changes over time in natural ecosystems are only recently emerging. Moreover, they focus on terrestrial and marine ecosystems, while freshwater ecosystems are not well represented, especially lakes and ponds [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:

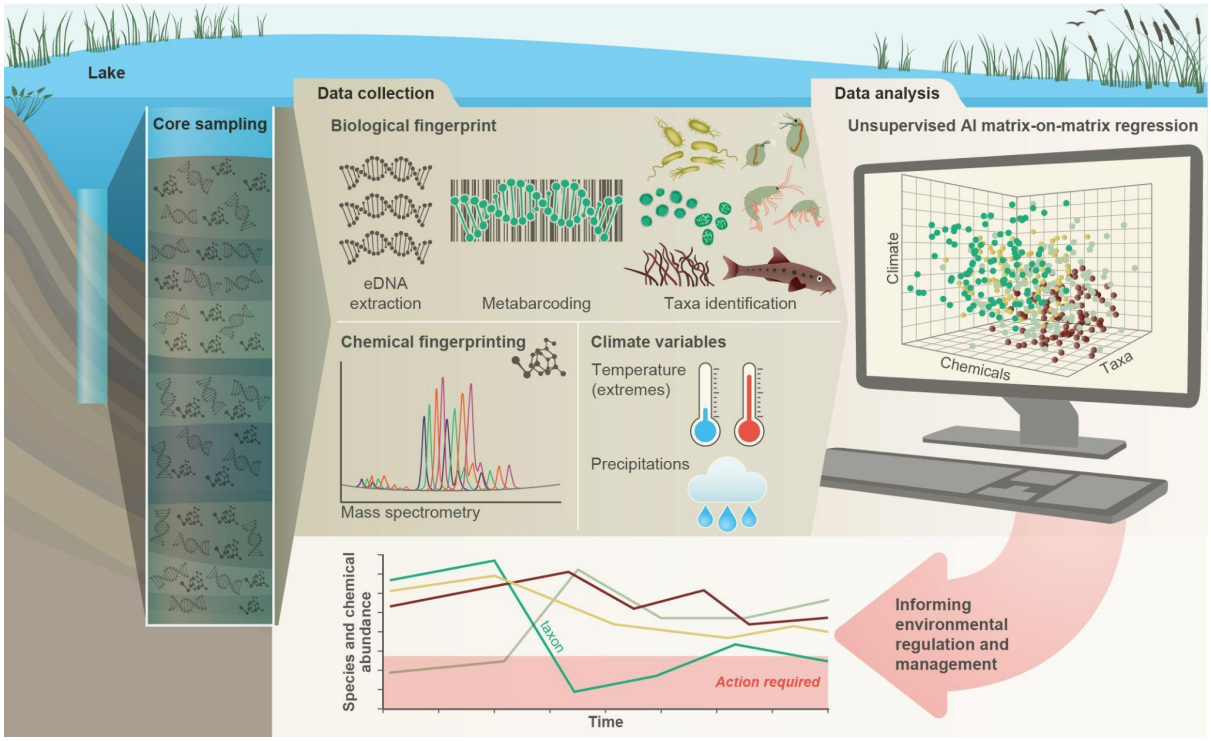
i) Biodiversity loss occurs at different spatial and temporal scales, and dynamic changes in community composition are the result of long-term ecological processes [10,11]. State-of-the-art environmental and biological monitoring typically captures single snapshots in time of long-term ecological dynamics, 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, the majority of these studies encompass the last 10-25 years and includes a minority of freshwater ecosystems [12]. Whereas 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, *seda*DNA (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 {Eastwood, 2022 #43.

Recently, we have developed a framework that helps establish the links between biodiversity dynamics and abiotic environmental changes and examines emergent impacts on ecosystem functions {Eastwood, 2022 #43}. Here, we illustrate 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 (>1999s) 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). Given the well-documented human-impact over time, Lake Ring represents an excellent natural system to demonstrate the power of systemic approaches in biological and functional monitoring.

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



Results

Freshwater community dynamics across 100 years

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. The 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 barcode) 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 primary producer's composition (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 (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 phase, 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 [e.g. *Nitrospiraceae* (nitrite oxidizers) were more abundant in high nutrient environments (eutrophic and pesticides) than in lower nutrient environments (semi-pristine and recovery) (Fig. 2B; 16SV1)].

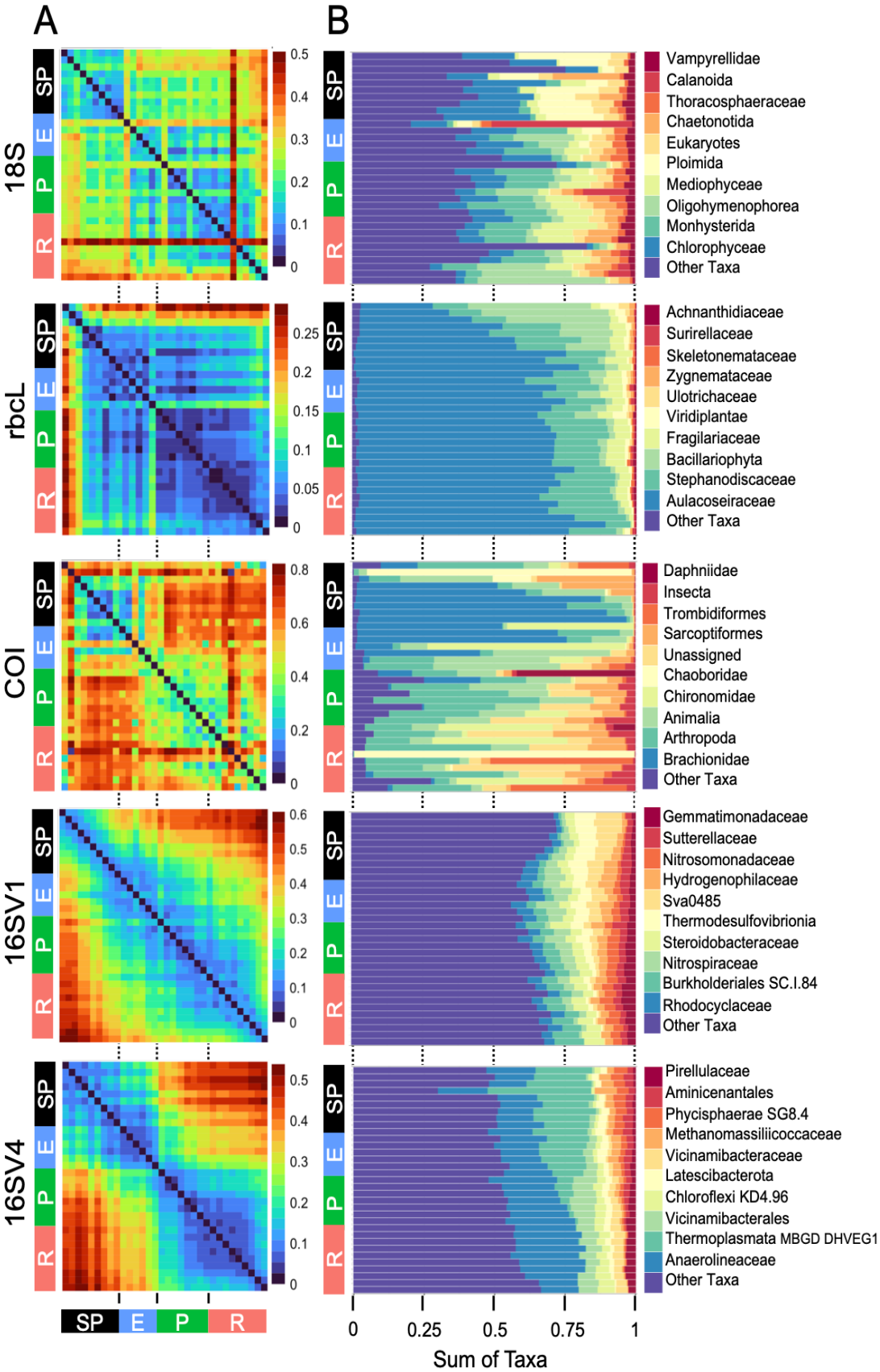
Changes in the invertebrate community were driven by *Brachionideae* (rotifers) that were most abundant in the semi-pristine phase and declined over time; *Chironomidae* (lake flies) that were proportionally more abundant in the eutrophic and recovery phases and showed the lowest abundance in the pesticides phase; *Chaoboridae* (phantom midge larvae) that were only present in the semi-pristine and recovery phases; and *Daphniidae* (waterfleas) that were most abundant in the pesticide phase, but present throughout the 100 years of sampling (Fig. 2B; COI). The diatom composition was stable over time, with only the semi-pristine phase

having a more distinctive diatom assemblage profile dominated by *Bacillariophyta* (Fig. 2B; rbcL). Diatoms are commonly used by regulators to derive the status of freshwater within the Water Framework Directive both for lakes and rivers [26]. We used our rbcL data to derive a Lake Trophic Diatom Index (LTDI2) for Lake Ring following [27]. This result confirmed our beta diversity analysis of non-significant changes over time of the diatom community (Supplementary Fig. 3).

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

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

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


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

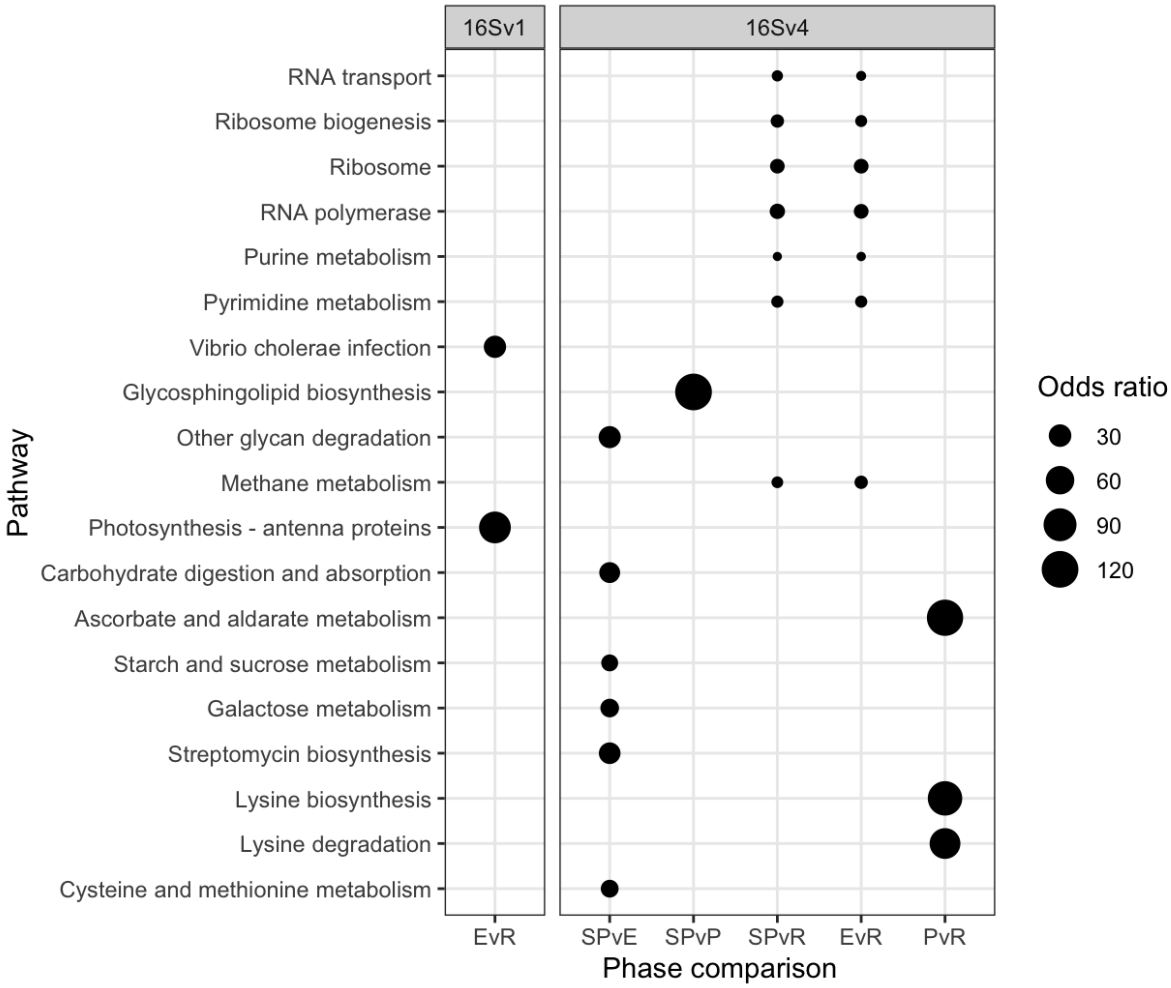
211

212

Functional changes linked to community compositional shifts

Changes in freshwater community composition corresponded to significant shifts in the predicted functional biodiversity of the prokaryotic community (Fig 3). We predicted a total of 6,257 Kegg Orthologs (KO) terms from the 16SV1 and 6,828 from the 16Sv4 barcode across the lake phases. Of the total number of KO terms, 1,418 were significantly differentially abundant across the lake phases in the 16SV1 and 1,064 terms in the 16SV4 dataset, respectively. The functional KEGG pathways enriched within these KO terms and significantly differentially enriched between lake phases (Fisher's exact test, p-adj < 0.05) were 19 (17 for the 16SV4 and 2 for the 16SV1) (Fig. 3). Seven differentially enriched pathways were found between the semi-pristine and recovery phases and seven were found between the eutrophic and recovery phases (Fig. 3; 16SV4). These pathways were linked to catabolic functions (purine and pyrimidine metabolism), RNA transport and biogenesis, fundamental for gene expression and protein folding. Six functional pathways were differentially enriched between the semi-pristine and the eutrophic phases that were linked to metabolism (including methane metabolism), degradation and biosynthesis (Fig. 3; 16SV4). Three functional pathways that underpin carbohydrates metabolism, lysine biosynthesis and degradation were differentially enriched between the pesticide and recovery phases. The latter two functions are critical for mitochondrial function. A single pathway was differentially enriched between the semi-pristine and the pesticide phases, linked to lipid metabolism (glycosphingolipid biosynthesis; Fig. 3; 16SV4). Two differentially enriched pathways were identified between the eutrophic and the recovery phases and underpin infection response and photosynthesis (Fig. 3; 16SV1).

235 **Figure 3. Functional analysis.** Functional pathways that are significantly differentially
236 enriched between lake phases are shown for the 16SV1 and the 16SV4 barcodes. The lake
237 phases are as in Figure 2: SP - semi-pristine; E - eutrophic; P - pesticides; R - recovery. Odds
238 ratios indicate the representation of each pathway in the pairwise comparisons.
239



240

241

Drivers of biodiversity change

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

Having ranked biocides and climate variables that best explained changes in overall biodiversity, we identified correlations between taxonomic groups (assigned at family level where possible) and individual abiotic variables. Correlations were identified between a total of 36 eukaryotic families and abiotic variables; of these correlations, 28 were with biocides and 25 with climate variables (some correlations involved the same taxonomic group correalting with multiple environmental factors). Of the 28 families negatively correlated with biocides, the largest proportion co-varied significantly with insecticides (21 families - 75%) and fungicides (14 families - 50%), followed by herbicides (7 families - 25%) and pesticides (2 families - 7.1%) (Supplementary Table 2). Of the 25 families correlated with climate variables, the largest proportion co-varied with summer precipitations (12 families - 37%); of these, 8 families were positively correlated and 4 were negatively correlated with summer precipitations. 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 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 precipitations; 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 precipitations; 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 prominently explained by biocides (Fig. 4; 18S; Biocides: 44%), followed by climate variables (Fig. 4; 18S; climate variables: 22%). Up to 22% of the diatom's 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, additive effects of biocides and climate variables were observed for 23 prokaryote (16S) and two eukaryote (18S) families (Fig. 5A), whereas no additive effects were identified on the diatom (rbcL) and the invertebrate (COI) communities (Fig 5A; Supplementary Table 3). The most frequent additive effects on prokaryotes involved insecticides and mean minimum temperature (Fig. 5A; Supplementary Table 3). Additive effects between herbicides and precipitation/lowest recorded temperature were rare (Fig. 5A; Supplementary Table 3). The most frequent additive effects on the eukaryotic community were observed between insecticides and precipitations (Fig. 5A; Supplementary Table 3).

The biocide types showing additive effects with environmental variables were ranked based on their correlation coefficient over time (Supplementary Table 3). The top ranked insecticides most frequently showing additive 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 additive effects with only the prokaryotic community. Examples of additive effects on specific families are shown in Figure 5B and 5C. The temporal dynamics of Isochrysidales, a coccolith-producing microalgae, was affected by the additive effect of summer precipitation and insecticides (Fig. 5B), whereas the temporal dynamics of the PeM15 group of Actinobacteria was affected by the additive effect of insecticides and mean minimum temperature (Fig. 5C).

Figure 4. sCCA 3D plots. Sparse canonical correlation analysis 3D plots for the five barcodes used (18S, rbcL, COI, 16SV1 and 16SV4), showing the proportion of biodiversity variance explained by the biocides and climate variables. As biocides were introduced around the 1960s, this analysis spans the most recent three lake phases (eutrophic, pesticides and recovery).

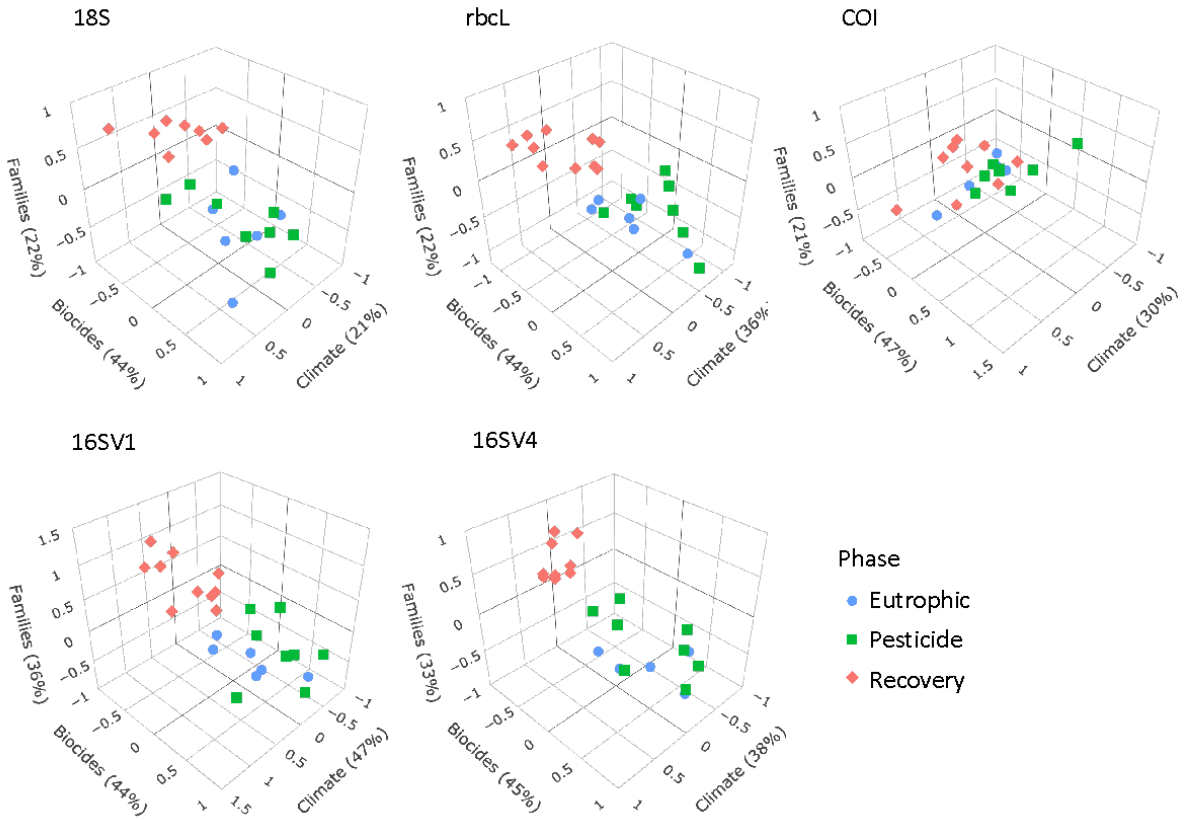
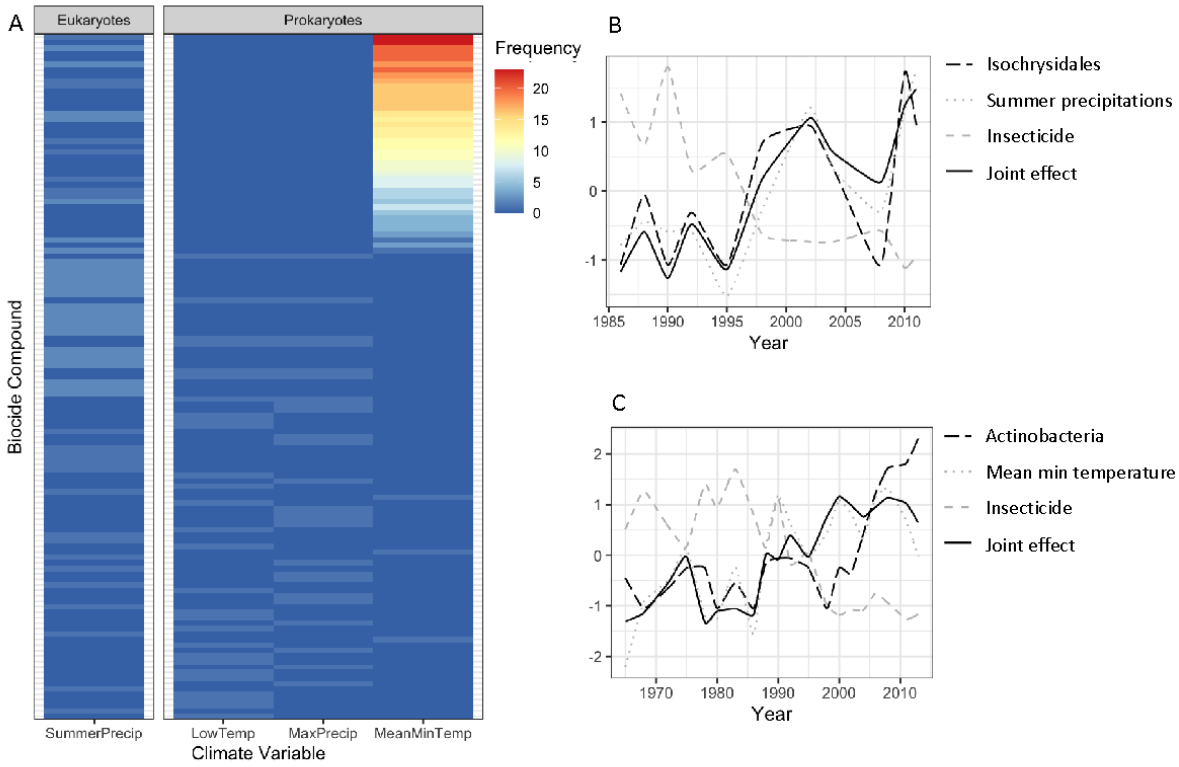


Figure 5. Additive effects on biodiversity. A) heatmap showing the frequency of additivity between biocides and climate variables in eukaryotes (data from the 18S barcode) and prokaryotes (combined data from 16Sv1 and 16Sv4 barcodes). The biocides are ranked based on their correlation coefficient with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



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 skills [28]. Direct observations are inherently biased towards species that leave fossil remains; species identification is strongly reliant on well-preserved remains in environmental matrices; and cryptic species diversity cannot be resolved [13]. Recently, automated acquisition of microfossil data using artificial intelligence has been proposed as an alternative to human inspection for reconstructing long-term biological changes [29]. However, this approach relies on the completeness of reference databases and of the fossil remains, suffering from the same limitations of direct observations minus the low throughput aspects. Efforts to catalogue temporal changes in biodiversity have recently started to understand changes in species richness and assemblages in different geographic regions of the globe [12]. These efforts are important to understand the extent of overall biodiversity loss. However, there are only a handful of existing datasets that span more than 50 years and many of the multidecal biodiversity time series are limited to terrestrial and marine ecosystem, with freshwater ecosystems being marginally represented [12]. Moreover, long-term freshwater studies tend to focus on indicator species or specific taxonomic groups (e.g. invertebrates), rather than capturing community-level patterns [7]. Developments in the field of *seda*DNA have addressed the limitations of direct observations, utilising the properties of eDNA [15]. However, *seda*DNA studies have predominantly focused on microorganisms as proxies for ecosystems' health (e.g. cyanobacteria [30]; ciliates [16]; parasitic taxa [31]), 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.

Studies of temporal dynamics typically start from an already shifted baseline and rely on discrete observations [16]. Our study alleviates these limitations by providing a continuous community-level analysis of biological changes over recent evolutionary times and starting from a relatively undisturbed environment. However, eDNA-based studies suffer from limitations linked to the level of preservation of nucleotides in environmental matrices. Although it has been shown that DNA can be recovered from lacustrine and marine sediments as far back as the Holocene [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 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 did. 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 environment, 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 indicator of freshwater ecological status, was not changing 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 to identify ecosystem shifts.

Even if Lake Ring partially recovered from eutrophication and biocides pollution in modern times, both the contemporary eukaryotic and prokaryotic communities are significantly different from the semi-pristine historical community, as the PERMANOVA on beta diversity demonstrates. Our findings align with other studies using *seda*DNA on decennial timeframes focusing on prokaryotes (e.g. cyanobacteria [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.

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, 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 in correspondence of major ecosystem shifts.

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 additive effect of insecticides/fungicides and yearly extreme temperature/precipitations best explained changes in overall biodiversity. The additive effects of insecticides and extreme temperature events affected prokaryotes by altering their functionality and changing their metabolic, biosynthesis and degradation functions. The additive effect of insecticides and 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/precipitations, 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 reliability surpass state-of-the-art analysis of patterns of co-occurrence of taxa within communities [47]. A step in the right direction to capture complex interactions between biotic and abiotic variables is the network analysis of co-occurrence patterns among physicochemical and biological parameters using random forest machine learning algorithms (e.g. [48]). This approach is hypothesis-free and allows to identify synchronicity between various environmental variables and *seda*DNA sequence variation. However, even when applied to temporal trends, it does not quantify additive effects of environmental factors on biodiversity. So far, random forest machine learning algorithms have only been applied to prokaryotic communities, disregarding other taxonomic groups and providing a partial understanding of community-level patterns and responses [48].

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 lacking [49], even if some countries have tried to introduce semi-quantitative

indices to assess the ecological status of freshwater [50]. Regulators must rely on approaches ingrained into environmental law, even though they have been proven inadequate (e.g. TDI), as the continuous decline in biodiversity demonstrates [19]. Even when direct links between biological indicators and abiotic drivers can be established, these rely on indicator species (e.g. a fish, an alga and an invertebrate) used as proxies for ecosystem health [51]. 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 impact [21]. 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. To reduce potential contamination when handling older sediment layers each layer of sediment was handled in a PCR-free and DNA-free environment. A radiometric chronology of this sediment was completed in 2018 by Goldsmith Ecology Ltd following standard protocols [52], and provided an accurate dating of the sediment to the year 1916. 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 physicochemical 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 parameters 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

We applied multilocus metabarcoding or marker gene sequencing to environmental DNA (eDNA) extracted from the 34 layers of sediment from the biological archive of Lake Ring using a laminar flow hood in a PCR-free environment to obtain a fine-grained temporal quantification of taxonomic diversity and relative abundance of taxonomic groups. eDNA was extracted from the dated sediment layers - *seda*DNA - using the DNeasy PowerSoil kit (Qiagen), following the manufacturer's instructions. Negative aerial and PCR controls were used; in addition, positive controls for PCR consisting of duplicates of three random samples from the sedimentary archive, were used. The duplicated samples were very similar, providing confidence in the approach used (Supplementary Fig. 2). Triplicates of each *seda*DNA sample were amplified with a suite of five nuclear and mitochondrial PCR primers (barcodes) to capture presence and relative abundance of eukaryotes (18S) [56], macroinvertebrates (COI) [57], primary producers (focus on diatoms; rbcL) [58], and prokaryotes (16SV1 and 16SV4) [59] using Q5 HS High-Fidelity Master Mix (New England Biolabs) and following the manufacturer's instructions. A negative control in triplicate per plate was used. Paired end 250 bp amplicon libraries were obtained using a 2 step PCR protocol with 96x96 dual tag barcoding to facilitate multiplexing and to reduce crosstalk between samples in downstream analyses [60] 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 aiming for 100,000 reads per sample and amplicon. 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 [61] and multiqc [62]. Sequences were then imported into QIIME2 v 2021.2 [63], trimmed, filtered, merged and denoised using the QIIME2 DADA2 module [64] 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 at family level 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 [65]; the diat.barcode v9.2 was used for the assignment of rbcL reads [66]; and the Barcode of Life Database was used for the COI reads [67]. The taxonomy was assigned using qiime feature-classifier classify-sklearn at family level where possible [68]. 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 [69] in R v4.0.2 [70] 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 [63]. The following samples did not meet the rarefaction cutoff: 16SV1: aerial negatives A, B, C; 16SV4: aerial negatives A, B, C and sampleID 62 sample;18S: aerial negatives A,B,C, negative PCR control, sampleID 18, positive control replicate 62; rbcL: aerial negative A, B, and sampleIDs 50, 54, 60; COI sampleIDs 40, 64. Alpha diversity differences among lake phases, using shannon entropy, were tested with Kruskal-Wallis test and beta diversity differences among lake phases, calculated as weighted unifrac distances, were established with a PERMANOVA test [71]. Alpha diversity was plotted using ggplot2 v3.3.5 with R v4.0.2. Heatmaps of weighted unifrac Beta diversity between each pair of sediment layers were plotted with the pheatmap v1.0.12 in R v4.0.2 [72].

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 *seda*DNA sequencing) and drivers of change, we focused on biocides and climate variables, using sparse Canonical Correlation Analysis (sCCA; it can be thought of as consensus PCA on multiple data matrices) followed by Sliding Window (Pearson) Correlation (SWC) analysis (Supplementary Fig. 5). Physico-chemicals 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 an additive effect on a given family if the linear combination of the biocide type and the climate variable had a larger Pearson correlation coefficient than each of the correlations between the family and the biocide type and the family and the climate variable individually, in the same time interval with padj $<$ 0.05 (with 10,000 permutations in the SWC analysis).

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

17 Bonebrake, T. C., Christensen, J., Boggs, C. L. & Ehrlich, P. R. Population decline assessment, historical baselines, and conservation. *Conservation Letters* **3** (2010).
18 Groh, K., Vom Berg, C., Schirmer, K. & Tlili, A. Anthropogenic Chemicals As Underestimated Drivers of Biodiversity Loss: Scientific and Societal Implications. *Environ Sci Technol* **56**, 707-710, doi:10.1021/acs.est.1c08399 (2022).
19 Pecl, G. T. *et al.* Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, doi:10.1126/science.aai9214 (2017).
20 Gillson, L. & Marchant, R. From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology. *Trends Ecol Evol* **29**, 317-325, doi:10.1016/j.tree.2014.03.010 (2014).
21 Cuenca - Cambronero, M. *et al.* Predictability of the impact of multiple stressors on the keystone species Daphnia *Scientific Reports* **8**, 17572 (2018).
22 Cambronero, C. M., Beasley, J., Kissane, S. & Orsini, L. Evolution of thermal tolerance in multifarious environments. *Mol Ecol* **27**, 4529-4541, doi:10.1111/mec.14890 (2018).
23 Cuenca Cambronero, M., Beasley, J., Kissane, S. & Orsini, L. Evolution of thermal tolerance in multifarious environments. *Mol Ecol* **27**, 4529-4541, doi:10.1111/mec.14890 (2018).
24 Cuenca Cambronero, M., Zeis, B. & Orsini, L. Haemoglobin-mediated response to hyper-thermal stress in the keystone species Daphnia magna. *Evol Appl* **11**, 112-120, doi:10.1111/eva.12561 (2018).
25 Baltrusaitis, T., Ahuja, C. & Morency, L. P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans Pattern Anal Mach Intell* **41**, 423-443, doi:10.1109/TPAMI.2018.2798607 (2019).
26 Agency, E. Phytobenthos - Diatoms for Assessing River and Lake Ecological Quality (River DARLEQ3). (Sterling, UK, 2020).
27 Bennion, H. *et al.* Assessment of ecological status in UK lakes using benthic diatoms. *Freshwater Science* **33**, 639-654 (2014).
28 Moraitis, M. L., Valavanis, V. D. & Karakassis, I. Modelling the effects of climate change on the distribution of benthic indicator species in the Eastern Mediterranean Sea. *Sci Total Environ* **667**, 16-24, doi:10.1016/j.scitotenv.2019.02.338 (2019).
29 Itaki, T. *et al.* Innovative microfossil (radiolarian) analysis using a system for automated image collection and AI-based classification of species. *Sci Rep* **10**, 21136, doi:10.1038/s41598-020-77812-6 (2020).
30 Picard, M. *et al.* Using metabarcoding and droplet digital PCR to investigate drivers of historical shifts in cyanobacteria from six contrasting lakes. *Sci Rep* **12**, 12810, doi:10.1038/s41598-022-14216-8 (2022).
31 Talas, L., Stivrins, N., Veski, S., Tedersoo, L. & Kisand, V. Sedimentary Ancient DNA (sedaDNA) Reveals Fungal Diversity and Environmental Drivers of Community Changes throughout the Holocene in the Present Boreal Lake Lielais Svetinu (Eastern Latvia). *Microorganisms* **9**, doi:10.3390/microorganisms9040719 (2021).
32 Slon, V. *et al.* Extended longevity of DNA preservation in Levantine Paleolithic sediments, Sefunim Cave, Israel. *Sci Rep* **12**, 14528, doi:10.1038/s41598-022-17399-2 (2022).
33 Jia, W. *et al.* Preservation of sedimentary plant DNA is related to lake water chemistry. *Enviornmental DNA* **4**, 425–439 (2022).
34 Mauvisseau, Q. *et al.* The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments. *Environ Sci Technol* **56**, 5322-5333, doi:10.1021/acs.est.1c07638 (2022).

35 Giguet-Covex, C. *et al.* Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat Commun* **5**, 3211, doi:10.1038/ncomms4211 (2014).
36 Cao, X. *et al.* Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities. *Water Res* **183**, 116077, doi:10.1016/j.watres.2020.116077 (2020).
37 Zhang, H., Huo, S., Yeager, K. M. & Wu, F. Sedimentary DNA record of eukaryotic algal and cyanobacterial communities in a shallow Lake driven by human activities and climate change. *Sci Total Environ* **753**, 141985, doi:10.1016/j.scitotenv.2020.141985 (2021).
38 Hillebrand, H. & Kunze, C. Meta-analysis on pulse disturbances reveals differences in functional and compositional recovery across ecosystems. *Ecol Lett* **23**, 575-585, doi:10.1111/ele.13457 (2020).
39 Fordham, D. A. Mesocosms Reveal Ecological Surprises from Climate Change. *PLoS Biol* **13**, e1002323, doi:10.1371/journal.pbio.1002323 (2015).
40 Li, F. *et al.* Human activities' fingerprint on multitrophic biodiversity and ecosystem functions across a major river catchment in China. *Glob Chang Biol* **26**, 6867-6879, doi:10.1111/gcb.15357 (2020).
41 Gounand, I., Little, C. J., Harvey, E. & Altermatt, F. Cross-ecosystem carbon flows connecting ecosystems worldwide. *Nat Commun* **9**, 4825, doi:10.1038/s41467-018-07238-2 (2018).
42 Choi, J. *et al.* Strategies to improve reference databases for soil microbiomes. *ISME J* **11**, 829-834, doi:10.1038/ismej.2016.168 (2017).
43 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-D426, doi:10.1093/nar/gky1038 (2019).
44 Cuenca-Cambronero, M. *et al.* Evolutionary mechanisms underpinning fitness response to multiple stressors in Daphnia. *Evol Appl* **14**, 2457-2469, doi:10.1111/eva.13258 (2021).
45 Pinheiro, J. P. S., Windsor, F. M., Wilson, R. W. & Tyler, C. R. Global variation in freshwater physico-chemistry and its influence on chemical toxicity in aquatic wildlife. *Biol Rev Camb Philos Soc* **96**, 1528-1546, doi:10.1111/brv.12711 (2021).
46 Patra, R. W., Chapman, J. C., Lim, R. P., Gehrke, P. C. & Sunderam, R. M. Interactions between water temperature and contaminant toxicity to freshwater fish. *Environmental Toxicology and Chemistry* **34**, 1809–1817 (2015).
47 Yang, P. *et al.* Correlation-Centric Network (CCN) representation for microbial co-occurrence patterns: new insights for microbial ecology. *NAR Genom Bioinform* **2**, lqaa042, doi:10.1093/nargab/lqaa042 (2020).
48 Tse, T. J. *et al.* Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial Communities. *Environ Sci Technol* **52**, 6842-6853, doi:10.1021/acs.est.7b06386 (2018).
49 Solimini, A. G., Cardoso, A. C. & Haiskanen, A.-S. Linkages between chemical and biological quality of surface waters. 1-262 (Joint Research Centre, Ispra, Italy, 2005).
50 Archaimbault, V. & Dumont, B. The normalized global biological index (IBGN): principles and evolution within the framework of the European framework directive on water. *Sciences Eaux & Territoires*, doi:https://doi.org/10.14758/SET-REVUE.2010.1.08 (2010).
51 Kanno, J. Introduction to the concept of signal toxicity. *J Toxicol Sci* **41**, SP105-SP109, doi:10.2131/jts.41.SP105 (2016).

52 Appleby, P. G. *Chronostratigraphic techniques in recent sediments*. Vol. 1 (Kluwer Academic Publisher, 2001).
53 Livingstone, D. M. & Lotter, A. F. The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palæolimnological implications. *Journal of Paleolimnology* **19**, 181–198 (1998).
54 Preudhomme, E. B. & Stefan, H. G. Relationship between water temperatures and air temperatures for central U.S. streams. (University of Minnesota, Duluth, Minnesota, 1992).
55 Wong, F., Robson, M., Diamond, M. L., Harrad, S. & Truong, J. Concentrations and chiral signatures of POPs in soils and sediments: a comparative urban versus rural study in Canada and UK. *Chemosphere* **74**, 404-411, doi:10.1016/j.chemosphere.2008.09.051 (2009).
56 Hadziavdic, K. *et al.* Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One* **9**, e87624, doi:10.1371/journal.pone.0087624 (2014).
57 Leray, M. *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool* **10**, 34, doi:10.1186/1742-9994-10-34 (2013).
58 Zimmermann, J. *et al.* Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. *PLoS One* **9**, e108793, doi:10.1371/journal.pone.0108793 (2014).
59 Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA* **108**, 4516-4522, doi:10.1073/pnas.1000080107 (2011).
60 MacConaill, L. E. *et al.* Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* **19**, 30, doi:10.1186/s12864-017-4428-5 (2018).
61 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338 (2018).
62 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
63 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).
64 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).
65 Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**, D643-648, doi:10.1093/nar/gkt1209 (2014).
66 Rimet, F. *et al.* Diat.barcode, an open-access curated barcode library for diatoms. *Sci Rep* **9**, 15116, doi:10.1038/s41598-019-51500-6 (2019).
67 Robeson, M. S., 2nd *et al.* RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol* **17**, e1009581, doi:10.1371/journal.pcbi.1009581 (2021).
68 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
69 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
70 Team, R. C. (Vienna, Austria, 2020).

71 Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32-46 (2001).
72 Kolde, R. *Pretty Heatmaps*, <https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf> (2019).
73 Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* **38**, 685-688, doi:10.1038/s41587-020-0548-6 (2020).
74 Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **26**, 27663, doi:10.3402/mehd.v26.27663 (2015).
75 Lin, D. *et al.* Group sparse canonical correlation analysis for genomic data integration. *BMC Bioinformatics* **14**, 245, doi:10.1186/1471-2105-14-245 (2013).
76 Parkhomenko, E., Tritchler, D. & Beyene, J. Sparse Canonical Correlation Analysis with Application to Genomic Data Integration. *Statistical Applications in Genetics and Molecular Biology* **8**, 1-34 (2009).
77 Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* **82**, 591-605, doi:10.1111/j.1469-185X.2007.00027.x (2007).
78 Buckland, M. & Gey, F. The relationship between recall and precision. *Journal of the American society for information science* **45**, 12-19 (1994).

Acknowledgments

We thank Kerry Walsh and Glenn Watts, the Environment Agency of England, for helpful discussions on the application of the approach presented here within regulatory frameworks. The metabarcoding data were generated by EnviSion, BioSeqencing and BioComputing at the University of Birmingham (https://www.envision-service.com/). The DDT chemical data were generated by the GEES Mass Spectrometry Facility at University of Birmingham. Sediment sampling and dating was completed by Goldsmith Ecology, Somerset. We thank Stephen Kissane for technical assistance in generating high throughput sequencing data, Dr Xiaojing Li for helpful discussions on functional analysis and Chantal Jackson for the artwork of Figure 1. This work was funded by the Alan Turing Institute (under EPSRC grant R-BIR-001); and the NERC highlights grant LOFRESH (NE/N005716/1). Niamh Eastwood is supported by the Midlands Integrative Biosciences Training Partnership (MIBTP; BB/M01116X/1). SC 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, SC and HH optimised chemical assays. TAD provided the sedimentary archive, the climate and the biocides sales data. HB provided the 96x96 unique barcode design. LO conceived and coordinated the study and data analysis. All co-authors contributed to paper writing and approved the final manuscript.

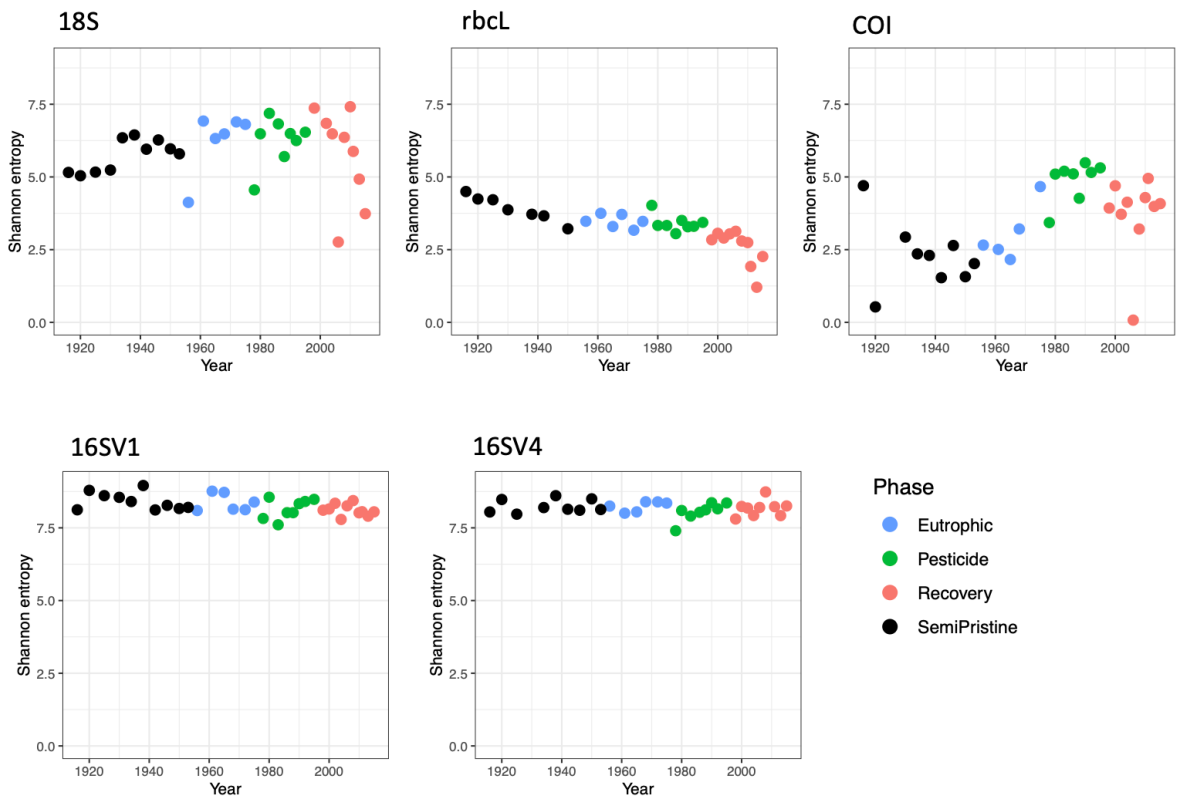
Competing interests

The authors declare no competing interests.

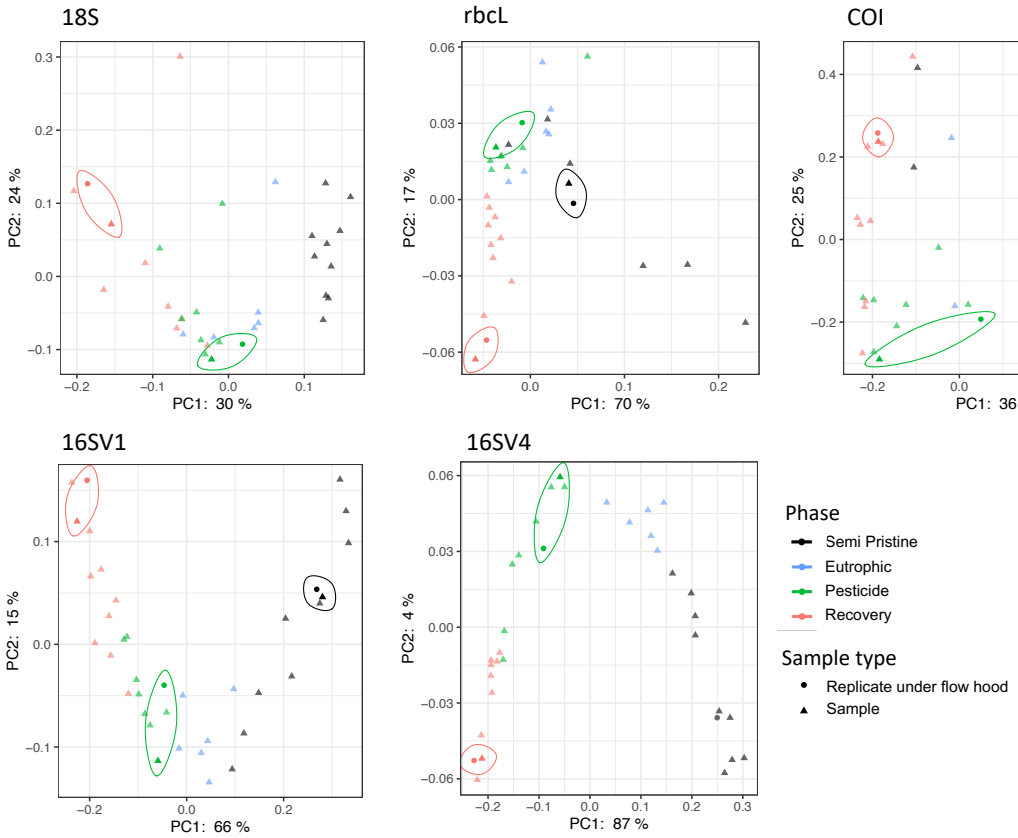
Supplementary information

Supplementary Figures and Tables

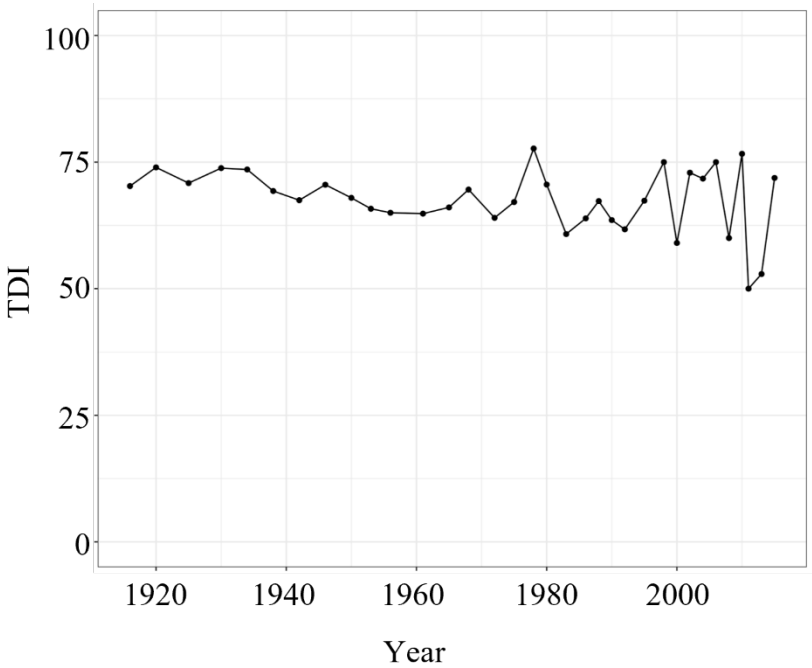
Supplementary Figure 1. Alpha diversity. Alpha diversity, measured as Shannon entropy, is shown for the five barcodes used in this study (16SV1, 16SV4, 18S, COI and rbcl) between 1916-2016. The four lake phases are colour-coded as follows: Black - Semi-pristine; green - Pesticides; blue - Eutrophic;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



Supplementary Figure 2. Principal Coordinate Analysis. PCoA visualization of weighted unifrac distance between samples. Positive controls for PCR consist of duplicates of up to three samples from the sedimentary archive for each of the five barcodes used in the study (16SV1, 16SV4, 18S, rbcL and COI). Replicated samples are circled.



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 6.31



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.

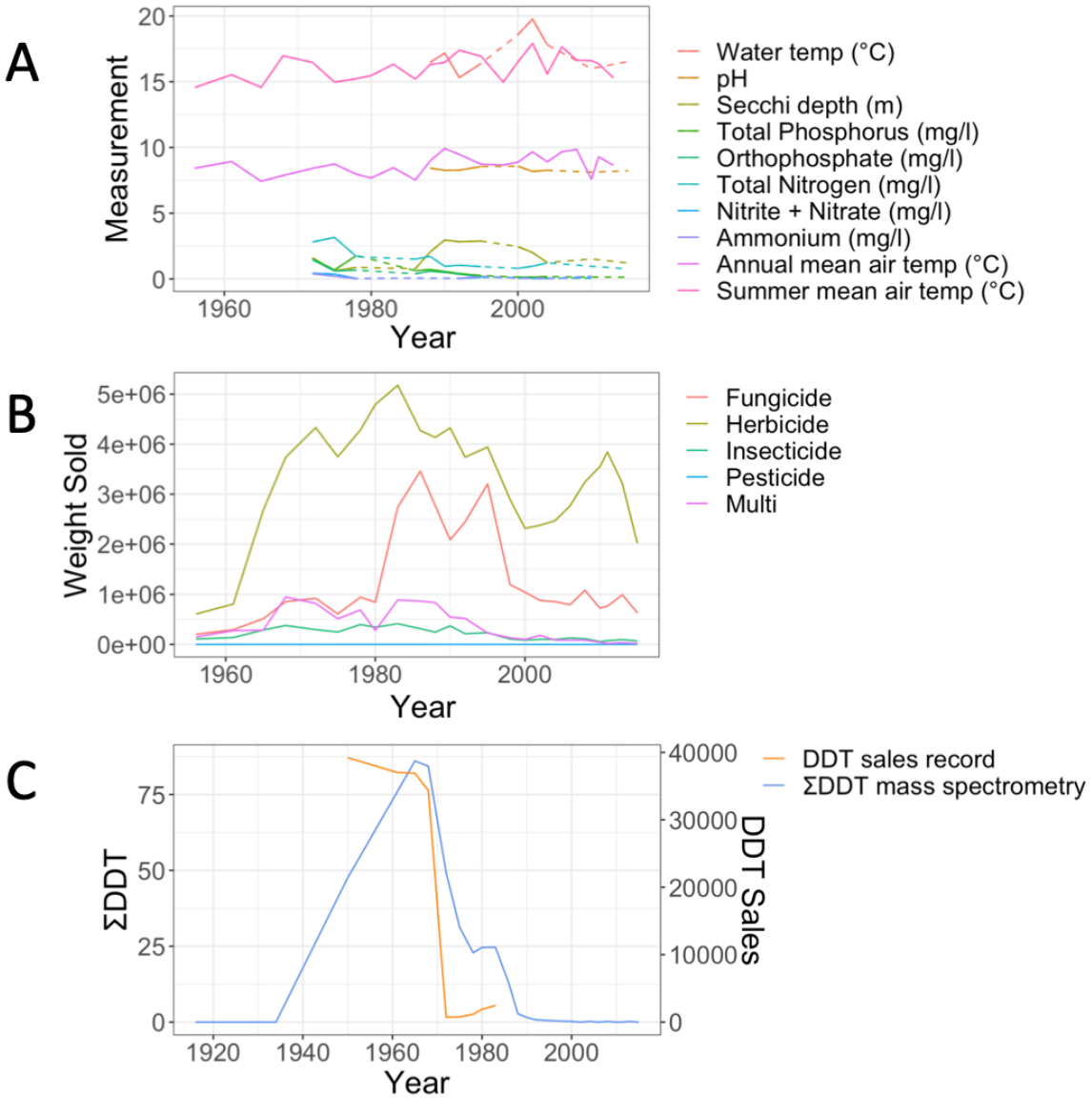
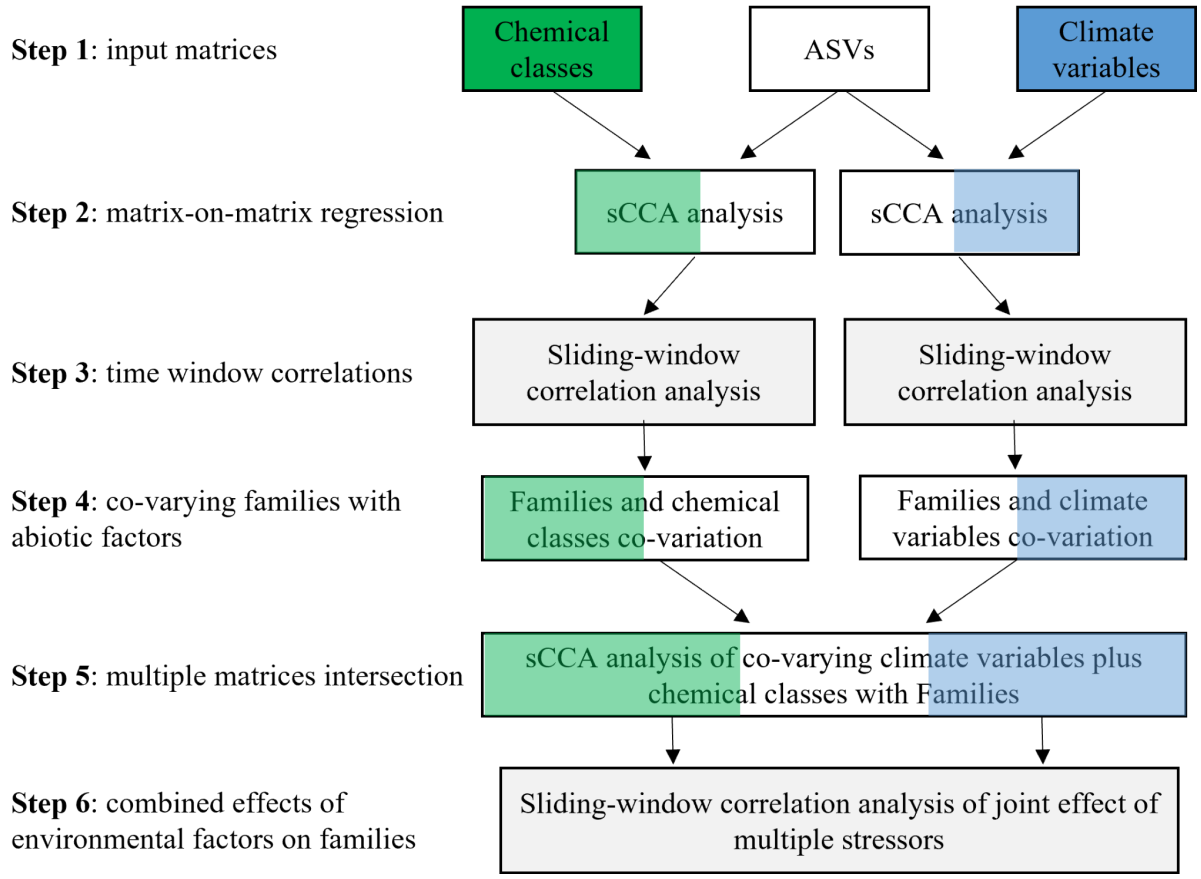


Figure 5. AI pipeline. The analytical pipeline consists of six main steps: **Step 1** is the preparation of input data matrices (ASVs, biocides and climate variables) to be used in the sCCA analysis. The type of environmental data may vary with the study; **Step 2** is the matrix-on-matrix regression 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, biodices and climate variables); **Step 6** applies a Sliding Window (Pearson) Correlation (SWC) analysis to identify families, whose relative abundance changes both with biocides and climate variables over time. The pipeline enables the ranking of environmental variables or their combination thereof that is inversely correlated to the relative abundance of families over time.



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 precipitations, highest recorded temperature, mean summer temperature, summer precipitations, annual total precipitations, summer atmospheric pressure and lowest recorded temperature.

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

		temperature			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

998 **Supplementary Table 3. Additive effects between biocides and climate variable.** The
999 biocides showing significant additive effect with climate variables are ranked based on their
1000 correlation coefficient. The barcode and identified families that are affected by the joint effect
1001 of a climate variable and biocides type are shown. The order in which the biocide types are
1002 ranked is the same used to plot Figure 5.

1003

1004 *See Eastwood_etal_Supplementary Table 3*

1005

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

See Eastwood_etal_Supplementary Table 4

1019 **Supplementary Table 5. PCR primers.** Tab1) PCR1 primers with bibliographic references,
1020 expected fragment size (bp), annealing temperature (°C) and primer sequences (in black) with
1021 overhang to prime the sequencing flow cell; Tab2) PCR2 primers consisting of Nextera
1022 adapters, universal tail and overhang sequence.

1023 *See Eastwood_etal_Supplementary Table 5*

1024

1025