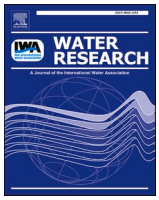
Water Research 300 (2026) 125959

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



# Discovery of nitrogen-responsive microbial indicators as metrics of freshwater ecosystem health

Jonathan Warren [a,*], Caitlin de Vries [a,b], Laura H. Hunt [a], Amy C. Thorpe [c], Susheel Bhanu Busi [c], Martyn G. Kelly [d,e], Dina-Leigh Simons [a], Joe D. Taylor [c], Daniel S. Read [c], Kerry Walsh [a]

[a] Environment Agency, Horizon House, Bristol, BS1 5AH, UK
[b] Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[c] UK Centre for Ecology & Hydrology (UKCEH), Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, UK
[d] Bowburn Consultancy, Durham, DH6 5QB, UK
[e] Nottingham University, Nottingham, NG7 2RD, UK

## HIGHLIGHTS

- Development of a framework to discover microbial indicators of ecosystem health.
- Identification of nitrogen-responsive bacterial predictive indicators.
- Determination of ecological change points for sensitive and tolerant bacteria.
- Foundation for a microbial-based metric of oxidised-nitrogen for ecosystem health.

## ARTICLE INFO

*Keywords:*
Microbiome
Machine learning
Biofilm
Ecosystem health
Nitrate
Nitrogen

## ABSTRACT

Microorganisms represent the most taxonomically and functionally diverse components of freshwater environments. Whilst distinct microbial communities exist across freshwater habitats, such as the water column and sediments, epilithic and epiphytic biofilm communities are critical in performing key roles in biogeochemical cycling and freshwater food webs. Despite their biogeochemical and ecological importance, microorganisms are underrepresented in freshwater monitoring programmes and lack metrics to interpret complex ecological communities and assess ecosystem health.

We developed a framework for identifying microbial indicators of ecosystem health by analysing 16S rRNA gene sequences from bacterial communities in 1574 freshwater biofilm samples collected from 694 sites across England's river networks. Tree-based machine learning regression was used to assess taxa importance, and threshold-indicator analysis was applied to identify pollutant concentrations that alter the composition of freshwater biofilm communities, providing a foundation for incorporating microbial data into ecosystem health metrics.

We applied this framework to measure the response of bacterial communities within English freshwater biofilms to an oxidised nitrogen gradient. Our results demonstrate that bacterial taxa can predict a large proportion of the variance in oxidised nitrogen concentrations, and we identified specific concentrations at which sensitive and tolerant taxa respond. This study represents a step toward developing a microbial metric of ecosystem health, advancing the potential use of microbial indicators in future monitoring programs. This framework could enable investigation of new and emerging pressures by examining how environmental perturbations affect functional processes, potentially across various trophic levels, providing a more comprehensive view of environmental dynamics in biomonitoring.

* Corresponding author.
*E-mail address:* jonathan.warren@environment-agency.gov.uk (J. Warren).

https://doi.org/10.1016/j.watres.2026.125959
Received 16 February 2026; Received in revised form 7 April 2026; Accepted 16 April 2026
Available online 17 April 2026

## 1. Introduction

Addressing the decline in water quality driven by human pressures such as land-use changes and global climate change is a significant societal challenge (van Vliet et al. 2023; Shi et al. 2024). Legislation such as the Water Framework Directive (WFD) (European Union 2000; HM Government 2017) attempts to address this challenge by setting a target of "good ecological status" (GES) across water bodies to ensure future sustainable water resources. However, 25 years after its adoption, 63% of water bodies in the European Union fail to meet GES (EEA 2024). Both point-source and diffuse pollution from phosphorus and nitrogen, as well as harmful chemicals including pharmaceuticals, pesticides, and personal care products (Whelan et al. 2022; Wilkinson et al. 2022), contribute to unachieved ecological quality targets. Nutrient enrichment, in particular, is recognised as a key factor contributing to the failure to achieve GES (EEA 2024). Additionally, there is growing recognition that nutrients rarely act in isolation and their effects can intensify when combined with other environmental stressors such as hypoxia, increased water temperature, or co-exposure to pesticides (Folt et al. 1999; Gomez Isaza et al. 2020; Morrissy et al. 2021; Akinnawo 2023). For example, nitrate-N has been shown to interact synergistically with pesticides such as malathion, significantly reducing survival rates in aquatic taxa compared to either stressor alone (Krishnamurthy and Smith 2010). Furthermore, the negative effects of nitrate can be compounded by natural stressors such as abrupt temperature fluctuations or acidic pH (Gomez Isaza et al. 2018).

Nutrient management goals are often not linked with ecological targets (Carvalho et al. 2019; EEA 2024). For example, there is no systematic use of nitrogen-based criteria for the ecological assessment of rivers in the UK (Poikane et al. 2019), despite nitrate-N levels in England not decreasing relative to the decline in dissolved phosphorus concentrations (Environment Agency, 2024b). Across Europe, nitrogen thresholds exist alongside phosphorus, with total nitrogen thresholds ranging from 0.25 to 4 $mgL^{-1}$ in lakes and 0.25 to 35 $mgL^{-1}$ in rivers (Poikane et al. 2019). While setting new thresholds could be useful, there could be implications for key economic sectors, such as agriculture. Therefore, a balance is required between the willingness to set ecologically relevant thresholds (Howden et al. 2013; Wiering et al. 2023) and the practical challenges of achieving them.

There is a long-held assumption that within inland waters nitrogen plays a less important role than phosphorus in limiting primary productivity. This assumption is based on work that shows phosphorus has the most significant effect on the community composition of phototrophs in freshwaters (Biggs 2000; Hilton et al. 2006; Neal et al. 2006; Stevenson et al. 2012; Yuan et al. 2023), leading to a focus on controlling phosphorus loads to address eutrophication. This dogma is now being challenged as more evidence of co-limitation by nitrogen emerges (Maberly et al. 2002, 2020; Jarvie et al. 2018). Thus, scientific opinion now recognises the need for a dual strategy that addresses both phosphorus and nitrogen (Schindler 2012; Dolman et al. 2016; Schindler et al. 2016; Paerl et al. 2020).

Nutrient thresholds are typically derived from statistical methods that examine the strength of associations between biological communities and are largely driven by phosphorus gradients (Poikane et al., 2019; Kelly et al., 2022). For example, current monitoring under the WFD uses biological quality elements (BQEs) which focus on specific biological groups (phytoplankton, macrophytes, phytobenthos, benthic macroinvertebrates, and fish), none of which show a strong or distinct response to nitrogen. To improve nitrogen management, organisms or processes that respond directly to nitrogen availability, such as microbial communities and their associated biogeochemical functions, could be monitored and better reflect true ecological change (Kelly et al. 2022).

Microbial communities in freshwater ecosystems have potential to act as additional ecological indicators. They are the most taxonomically and functionally diverse component of freshwater ecosystems (Hug et al. 2016; Thompson et al. 2017) and form epilithic biofilms that play a crucial role in the global biogeochemical cycling of essential nutrients (Falkowski et al. 2008; Battin et al. 2016). When existing tools for the assessment of GES were developed, available methods were not practical to characterise the diversity and functional potential of microbial communities. High-throughput sequencing methods, including metabarcoding and metagenomic approaches, now enable characterisation of microbial communities in situ (Gilbert et al. 2014; Shaffer et al. 2022; Thorpe et al. 2025). Recent work has shown that alkalinity and related variables (pH and conductivity) is the largest driver in shaping bacterial communities in rivers, followed by nitrate-N and orthophosphate (Thorpe et al. 2026). Furthermore, the sensitivity of microbes to nitrate-N pollution and their active involvement in nitrogen processing demonstrate their potential as candidates for nitrate-N indicators (Kuypers et al. 2018; Zhang et al. 2022).

To maximise the potential of large high-throughput sequencing datasets, machine learning techniques have been developed and adopted. These techniques have allowed complex interactions between biological communities and environmental variables to be discovered by leveraging predictive regression or classification models (Cordier et al. 2019; McElhinney et al. 2022). Two such methods are random forests and extreme gradient boosting, both of which use bootstrapping approaches to create many decision trees, resulting in highly interpretable models (Breiman 2001; Geurts et al. 2006; Chen and Guestrin 2016). Crucially, these models enable the extraction of important features contributing to model predictions, which could identify microbes most affected by environmental pressures and useful as indicators of ecological health. Such models have been widely applied to other microbiome data, particularly in healthcare settings (Pasolli et al. 2016; Ai et al. 2019; Zhou and Gallins 2019; Kosciolek et al. 2021). They have also been applied in environments such as agricultural soils (Chang et al. 2017), leaf litter (Thompson et al. 2019), bioreactors (Dutta et al. 2022; Yu et al. 2024), marine sediments (Lanzén et al. 2021), ballast waters (Gerhard and Gunsch 2019), and aquaculture (Frühe et al. 2021). However, their application in microbial ecology in natural freshwater environments remains limited, and they have not yet been used to develop microbial metrics. To our knowledge, the only work developing microbial metrics focuses on predicting the Infaunal Quality Index (IQI: Kennedy et al. 2011) using bacterial communities and random forest approaches which are already used for regulatory compliance (Wyness et al. 2026). However, the IQI is only for use in transitional and coastal waters and the microbial variant serves as a replacement for macroinvertebrate sampling, rather than a direct measure of a pressure. Collectively, molecular technologies and machine learning approaches present new opportunities to better understand microbial ecology in aquatic environments, particularly in the development of novel microbial indicators and metrics of water quality and ecosystem health (Ghannam and Techtmann 2021; Codello et al. 2022; Guseva et al. 2022; McElhinney et al. 2022; Hallin 2023; Orr et al. 2025).

Here, we develop an analytical framework using DNA metabarcoding data to identify features of microbial communities that could serve as bioindicators of environmental change. To achieve this, we analysed 16S rRNA gene sequences from bacterial communities in 1574 freshwater biofilm samples from 694 sites across England's river networks. We created an analysis framework utilising machine learning techniques to develop baselines for microbial indices and microbially derived ecological change points where the community responds strongest to specific water quality concentrations. Then, we demonstrate the application of the analysis framework by assessing the impact of oxidised nitrogen concentrations on the bacterial component of the river microbiome. This study also serves as a case study to illustrate how metabarcoding can support faster progress towards WFD targets, as well as contributing to the wider debate on appropriate nitrogen thresholds for inland waters.

## 2. Materials and methods

### 2.1. Site selection and biofilm sampling

To capture trends in microbial communities across England, river benthos samples collected as part of the Environment Agency's River Surveillance Network monitoring program for diatoms were utilised. This is a network of spatial sample locations that are statistically generated in an unbiased manner using a generalised random tessellation stratified probabilistic sampling design (Brown et al. 2015). Stone or macrophyte scrape samples were collected according to previously described methods (Kelly et al. 2020) from selected sites and sampled up to six times in the spring and autumn of 2021, 2022, and 2023 (Fig. 1).

Biofilm-covered stones were collected in a tray and scrubbed with deionised or tap water using a clean toothbrush. Using a pipette, 5 mL of the biofilm suspension was transferred to a 15 mL tube containing RNA*later*-based preservative (3.5 M ammonium sulphate, 17 mM sodium citrate, and 13 mM Ethylenediaminetetraacetic acid), transported to the laboratory via an overnight courier at 5 ± 3 °C, and stored frozen at −20±5 °C for up to 3 years prior to DNA extraction (Kelly et al. 2020; Warren et al. 2024).

### 2.2. Water chemistry

Water chemistry data from sampling locations collected between 2021 and 2023 were extracted from the Environment Agency Water Quality Archive (Open WIMS data). Both nitrate-N and TON were left-censored because of the abundance of values below the limit of detection (LOD). Within Python (version 3.12.3) a Kaplan-Meier Estimation (KME) was applied to impute values to the lowest available LOD (Kaplan and Meier 1958) using *lifelines* package (version 0.28) *KaplanMeierFitter* function (Davidson-Pilon 2024). To assess the overall impact of oxidised nitrogen concentrations on the microbiome, data were expressed as the mean measurement for both nitrate-N and TON within a three-month period prior to biofilm sampling. This approach accounts for the fact that microbial biofilms are integrative indicators that reflect seasonal averages and nutrient conditions leading up to sampling, rather than instantaneous concentrations which may be subject to high-frequency

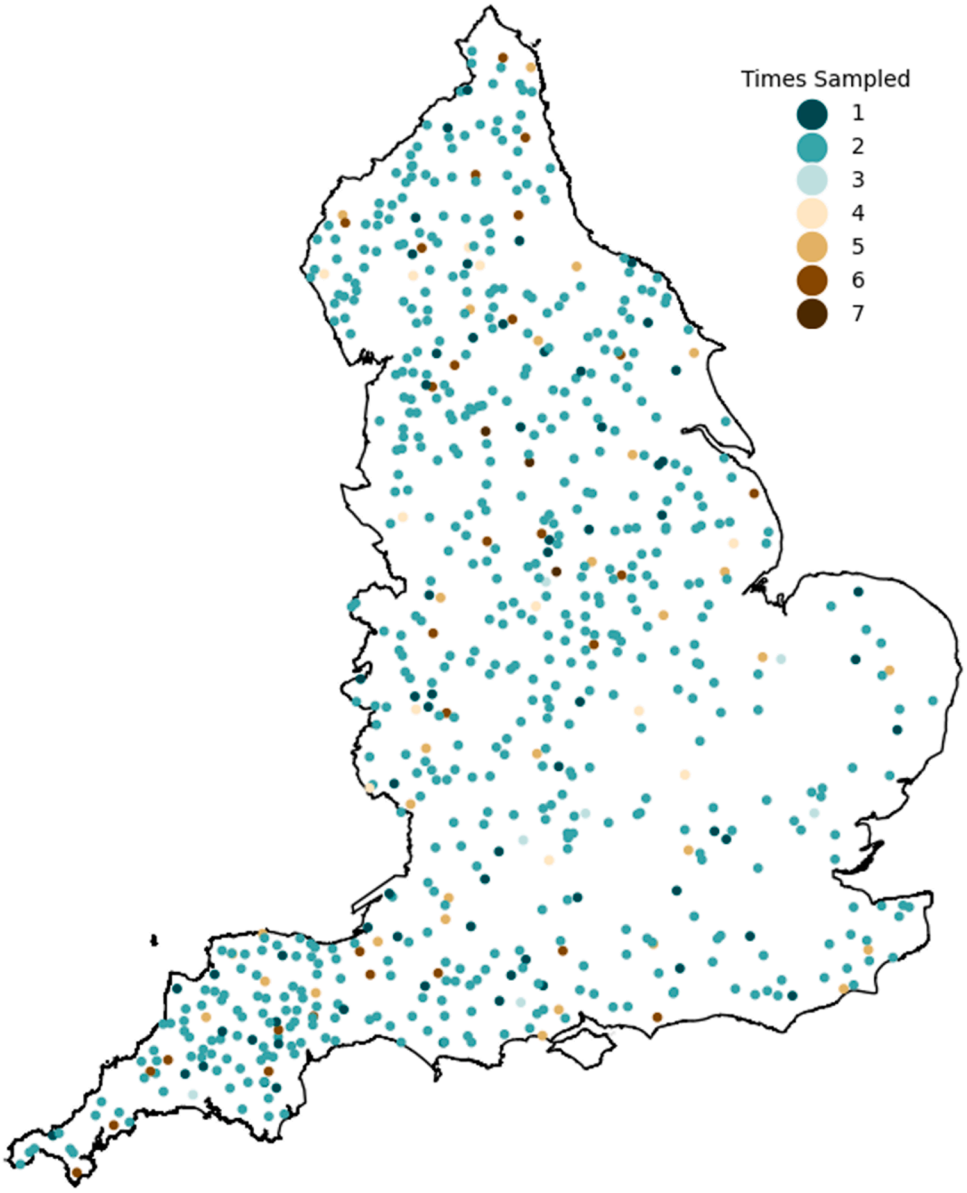



**Fig. 1.** Map of England showing the sample site locations. Colours represent the number of times each site was sampled during a three-year period from 2021 to 2023. Of the 694 sites, 66 were visited once, 547 twice, 7 three times, 12 four times, 30 five times, 32 six times, and 3 seven times.

variability (Snell et al. 2014; Kelly et al. 2020).

### 2.3. Sequence generation and processing

Bacterial communities were characterised by metabarcoding, as previously described in Environment Agency (2024a). Briefly, DNA was extracted using the DNA Fecal/Soil Microbe Kit (cat. no. D6011; Zymo Research, CA, USA) following a modified protocol to maximise DNA recovery (Newbold et al. 2023, 2025). The v4 region of the 16S rRNA gene was amplified using primers 515f: 5′-GTGYCAGCMGC CGCGGTAA-3′ and 806r-modified: 5′-GGACTACNVGGGTWTCTAAT-3′ (Walters et al. 2016) and tagged before sequencing using an Next-Seq1000 (Illumina, CA, USA) instrument. Cutadapt was used to remove primer sequences, and subsequent quality control, trimming, and denoising of the amplicon reads were performed using dada2 (version 1.35.6) (Martin 2011; Callahan et al. 2016). Taxonomy was assigned to amplicon sequence variants (ASVs) using the SILVA 138.2 database using the *AssignTaxonomy()* function, which was modified to perform 1000 bootstraps, with a minimum of 90% of bootstraps required to assign taxonomy to the genus level (Wang et al. 2007; Quast et al. 2013). Data were transformed per sample to relative abundance to normalise samples and avoid data leakage between model training and evaluation. Low-abundance taxa (<0.001% of the total aggregate relative abundance across all samples) and those with a low prevalence (found in <10 samples) were discarded ($n = 1014$ genera removed) to reduce issues of overfitting, resulting in 699 genera for further analysis. Bacterial ASVs were aggregated to the genus level at a 90% confidence threshold prior to modelling to reduce feature sparsity and improve model stability while retaining sufficient taxonomic resolution for functional ecological interpretation. This level of aggregation also ensures consistency across samples and sampling periods, which is important for identifying long-term indicators of ecosystem condition.

### 2.4. Tree-based regression analysis

The performance of two machine learning algorithms implemented within Python (version 3.12.3), *sklearn's* (version 1.4.2) *RandomForestRegressor* (Breiman 2001; Geurts et al. 2006) and *xgboost's* (version 3.0.0) *XGBRegressor* (Chen and Guestrin 2016), was compared for the ability to predict nitrate-N and TON concentrations using bacterial genera relative abundance and to extract predictive indicator genera.

Throughout models utilised a 70/30% train/test split with a random state of 42 to ensure reproducibility. Optimal hyperparameters for each model and regressor combination were determined using *RandomizedSearchCV* with 500 combinations trialled, each using a 10 cross-fold validation approach. The final model hyperparameters (Table 1) were evaluated based on the mean $R^2$ across fold validations and compared with the $R^2$ of the test data to ensure that the models were not overfitted. The final model performance was evaluated using the test set $R^2$ and the mean average percentage error.

Shapley additive explanation (SHAP) values were calculated for the test set using the Python *SHAP* package (version 0.46.0), which increased model interpretability by quantifying the contribution of each genus to the prediction. The SHAP values were derived from cooperative game theory by iteratively assessing all combinations of features and measuring how the model's prediction changed as individual features were added or removed, where a feature's final SHAP value reflected its average marginal contribution across these combinations (Lundberg et al. 2020; Li et al. 2024).

### 2.5. Threshold indicator analysis

We used the threshold indicator taxa analysis within R (version 4.3.2) using the TITAN2 (version 2.4.3) package (Baker and King 2010; R Core Team 2022) to identify microbial genera that respond to changes in oxidised nitrogen concentrations and could therefore act as threshold indicators. The same data used in the machine learning models were used as input for the *titan()* function, with the default 0.95 reliability threshold, an increased number of permutations to 500, and a reduced purity cut-off of 0.8. Purity thresholds were reduced from the default to allow more community members to contribute to the filtered sensitive and tolerant taxon change points. Ecological change points were calculated for all bacterial genera and then again for all predictive indicator genera identified in the SHAP analysis of the regression models to ensure final indicators were representative of the total bacterial community.

**Table 1**
Final hyperparameter values used for the oxidised nitrogen models. Full hyperparameter definitions are available for RFR and XGB. CV-grid score is the mean $R^2$ score calculated across all cross-validation folds and the test $R^2$ score which is generated by assessing the model's performance on the test dataset using the final, optimal hyperparameters.

| Chemical | Model type | Optimal hyperparameters | CV-grid score | Test $R^2$ |
|---|---|---|---|---|
| Nitrate-N log10 | RFR | 'max_depth': None, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 300 | 0.632 | 0.596 |
| TON log10 | RFR | 'max_depth': None, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 9, 'n_estimators': 400 | 0.636 | 0.597 |
| Nitrate-N log10 | XGB | 'colsample_bytree': 0.7, 'gamma': 0.148153724, 'learning_rate': 0.1, 'max_depth': 3, 'min_child_weight': 6, 'n_estimators': 300, 'reg_alpha': 0.1, 'reg_lambda': 0.809361155, 'subsample': 0.8 | 0.668 | 0.654 |
| TON log10 | XGB | 'colsample_bytree': 0.6, 'gamma': 0.068356481, 'learning_rate': 0.1, 'max_depth': 3, 'min_child_weight': 5, 'n_estimators': 300, 'reg_alpha': 0, 'reg_lambda': 0.821926636, 'subsample': 0.6 | 0.670 | 0.650 |

### 2.6. Linking functional processes with indicator models

Reference genomes were interrogated to link the results of the indicator analysis to the nitrogen cycling functional capabilities of each genus. To do this for all genera used for modelling, complete annotated reference genomes were downloaded from NCBI (O'Leary et al. 2024). This was first performed by searching for all the genus-level reference genomes. If no genus-level reference genomes were available, family level references were used. If no reference genomes were available, the genus was omitted from the analysis. Mismatches between SILVA 138.2 and the NCBI taxonomy were resolved manually by searching for known synonyms.

Reference genome annotations were parsed within Python, and child gene ontology (GO) terms for nitrogen cycle metabolic processes (GO:0071941) and GO descriptions were matched using GOtools (Klopfenstein et al. 2018). For each reference genome, the GO terms for nitrogen cycling were marked as present or absent. A genus was determined to be capable of a child process if any of the references were capable of that process.

To link function with predictive models and genus indicator status, weighted abundance tables were created to express the proportion of the community involved in each functional process. This allowed the identification of processes that may contribute disproportionately to nitrate-N prediction. Genera were then grouped by indicator status (predictive-threshold sensitive, predictive-threshold tolerant, and non-indicator genera), and we assessed whether genera in any group were more likely to be involved in nitrogen cycling processes.

Shotgun metagenomic analysis would have been the preferred

approach for linking functional processes with changes in pressures but opted to use this approach as a more cost-effective option.

## 3. Results

### 3.1. Site chemical and microbial characterisation

A total of 1574 biofilm samples were collected from 694 sites across the river networks of England (Fig. 1). The average nitrate-N concentrations had a median value of 2.75 $mgL^{-1}$ and ranged from 0.001 to 31.85 $mgL^{-1}$. TON had a median of 2.80 $mgL^{-1}$ and ranged from 0.005 to 32.00 $mgL^{-1}$ (Fig. 2). Nitrate-N, TON, and total nitrogen were all very strongly correlated with each other (Pearson's $r > 0.99$), moderately correlated with nitrite and alkalinity ($r = 0.42$–$0.43$), and weakly correlated with phosphates and phosphorus ($r < 0.262$, Supplementary Materials 1 + 2). Bacterial communities characterised using 16S V4 region rRNA gene metabarcoding generated 51,234,432 high quality reads containing 305,821 unique ASVs, of which 30.08% of reads (11.38% of ASVs) were assigned at the genus level, representing a total of 699 genera (Supplementary Materials 3) and used to build predictive models.

### 3.2. Identification of important predictive bacterial indicators of oxidised nitrogen

Both the XGB and RFR model types performed similarly in terms of $R^2$ and mean average percentage error, with XGB demonstrating a slightly improved fit and lower error compared to the equivalent RFR models (Table 2, Supplementary Materials 4). Therefore, XGB models were selected for both nitrate-N and TON.

Genera were deemed important predictive indicators if the SHAP values exceeded the value expected under equal contribution (i.e. the total sum of SHAP values for all genera divided by the number of genera). While additional genera had non-zero SHAP values and thus contributed to predictions, they were not classified as important (above a SHAP value of 0 but below the value of 'equal importance'). The XGB models identified 193 unique genera that were important in either the nitrate-N or TON model, with 130 genera common to both models and 22 and 41 unique important predictive indicators in the nitrate-N and TON models, respectively (Supplementary Materials 5).

For each XGB model, the 20 most important genera (based on SHAP values) were ranked and plotted to demonstrate how their relative abundance influenced the model predictions (Supplementary Materials 6 and 7). The relationships between these genera and the concentrations of nitrate-N and TON were simplified to positive or negative relationships (Table 3). Of the top 20 important taxa identified for each model,

**Table 2**
RFR and XGB model comparisons in terms of model fit, error, and distribution of feature importance according to SHAP analysis. The number of features above 'equal importance' is the number of features where the contribution is greater than the total SHAP values divided by the total number of features. Peak SHAP value (%) is the peak SHAP value expressed as a percentage of the total model SHAP values.

| Oxidised Nitrogen modelled (log10) | Model type | $R^2$ | Number of features above 0 SHAP value | Number of features above equal importance | Peak SHAP value (%) |
|---|---|---|---|---|---|
| Nitrate-N | RFR | 0.596 | 695 | 109 | 16.09 |
| Nitrate-N | XGB | 0.654 | 346 | 152 | 6.83 |
| TON | RFR | 0.597 | 681 | 109 | 16.84 |
| TON | XGB | 0.650 | 416 | 171 | 5.16 |

16 genera were shared, while four genera were unique to the nitrate-N model: *Roseococcus, Paludibaculum, Acinetobacter,* and *Polaromonas,* and four were unique to the TON model: *Phormidium CYN64, Chamaesiphon PCC-7430, Phormidesmis ANT.LACV5.1,* and *Pseudanabaena PCC-7429*. Those genera important in both models displayed similar relationships with the concentration of nitrate-N and TON. Additionally, the top four most important genera were shared across all models.

### 3.3. Identification of important predictive-threshold bacterial indicators of oxidised nitrogen

Of the 699 genera used for modelling, only 193 were important predictive indicators according to XGB modelling. Of these, only 37 taxa were not threshold indicators for either nitrate-N or TON in the TITAN analysis, resulting in 156 genera which were important predictive-threshold indicators of oxidised nitrogen species (Supplementary Materials 5). For most genera (98.2%), the two models agreed on whether there was no threshold response or whether the response was increasing (tolerant) or decreasing (sensitive) (Table 4).

Using the predictive threshold indicators, TITAN identified the community change point of oxidised nitrogen: the concentration along the environmental gradient where the sensitive or tolerant taxa show the strongest change, according to the concentration which had the greatest cumulative sum of z-scores. The significant community level change points were similar for both the nitrate-N and TON models, with tolerant genera having an ecological change point of 3.78 $mgL^{-1}$ and 3.97 $mgL^{-1}$, respectively. Sensitive genera also had similar change points of 0.30 $mgL^{-1}$ for nitrate-N and 0.30 $mgL^{-1}$ for TON. In addition, similar ecological change points were identified when using all bacterial genera

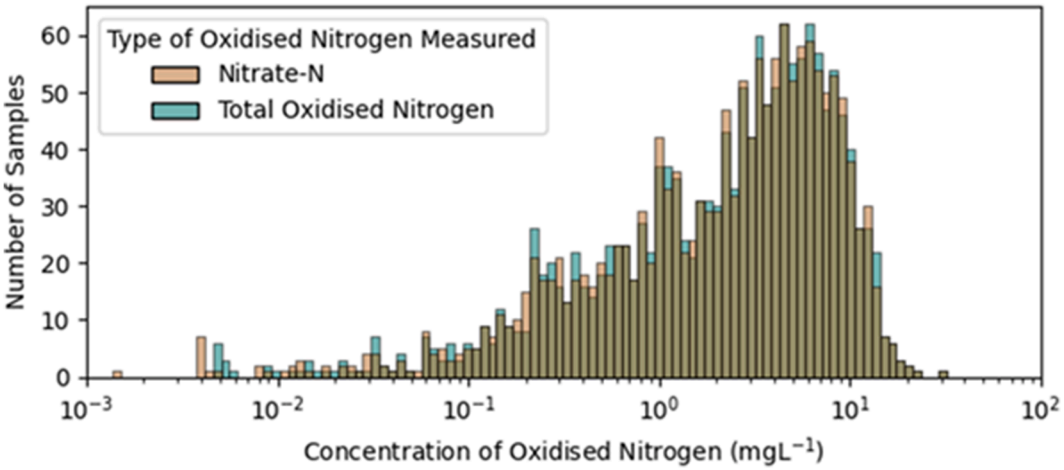



**Fig. 2.** Distribution of nitrate-N (brown) and Total Oxidised Nitrogen (green) across number of samples measured on a log10 scale.

**Table 3**
Top 20 genera from both nitrate-N and TON models showing their rank of importance in model prediction according to SHAP values, as well as their relationship to the dependent variable (+ for positive, – for negative, nl for non-linear). Higher taxonomic levels are shown in brackets.

| Taxonomy (domain/phylum/class/order/family) Genus | SHAP rank (relationship with dependent variable) | |
|---|---|---|
| | Log 10 Nitrate-N | Log 10 TON |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Burkholderiales; Comamonadaceae) **Hydrogenophaga** | 1 (+) | 1 (+) |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Burkholderiales; Rhodocyclaceae) **Denitratisoma** | 2 (+) | 2 (+) |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Lysobacterales; Lysobacteraceae) **Arenimonas** | 3 (+) | 3 (+) |
| (Bacteria; Pseudomonadota; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae) **Polymorphobacter** | 4 (-) | 4 (-) |
| (Bacteria; Pseudomonadota; Alphaproteobacteria; Hyphomicrobiales; Beijerinckiaceae) **FukuN57** | 5 (-) | 10 (-) |
| (Bacteria; Bacteroidota; Bacteroidia; Cytophagales; Spirosomataceae) **Lacihabitans** | 6 (+) | 5 (+) |
| (Bacteria; Acidobacteriota; Acidobacteriae; Bryobacterales; Bryobacteraceae) **Paludibaculum** | 7 (-) | 26 |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Lysobacterales; Rhodanobacteraceae) **Pseudolysobacter** | 8 (+) | 6 (+) |
| (Bacteria; Pseudomonadota; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae) **Rhizorhapis** | 9 (-) | 11 (-) |
| (Bacteria; Acidobacteriota; Acidobacteriae; Bryobacterales; Bryobacteraceae) **Bryobacter** | 10 (-) | 15 (-) |
| (Bacteria; Myxococcota; Myxococcia; Myxococcales; Anaeromyxobacteraceae) **Anaeromyxobacter** | 11 (-) | 9 (-) |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Cyanobacteriales; Xenococcaceae) **Pleurocapsa PCC-7319** | 12 (+) | 8 (+) |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Pseudanabaenales; Pseudanabaenaceae) **Synechococcus PCC-7502** | 13 (-) | 12 (-) |
| (Bacteria; Verrucomicrobiota; Verrucomicrobiia; Verrucomicrobiales; Rubritaleaceae) **Rubritalea** | 14 (+) | 17 (+) |
| (Bacteria; Pseudomonadota; Alphaproteobacteria; Caulobacterales; Hyphomonadaceae) **UKL13–1** | 15 (-) | 18 (-) |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Rickettsiellales; Rickettsiellaceae) **Rickettsiella** | 16 (+) | 7 (+) |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Burkholderiales; Comamonadaceae) **Polaromonas** | 17 (-) | 36 |
| (Bacteria; Pseudomonadota; Alphaproteobacteria; Acetobacterales; Acetobacteraceae) **Roseococcus** | 18 (-) | 23 |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Burkholderiales; Oxalobacteraceae) **Massilia** | 19 (+) | 13 (+) |
| (Bacteria; Pseudomonadota; Gammaproteobacteria; Pseudomonadales; Moraxellaceae) **Acinetobacter** | 20 (nl) | 29 |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Leptolyngbyales; Leptolyngbyaceae) **Phormidium CYN64** | 29 | 14 (-) |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Leptolyngbyales; Leptolyngbyaceae) **Chamaesiphon PCC-7430** | 42 | 16 (-) |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Phormidesmiales; Phormidesmiaceae) **Phormidesmis ANT.LACV5.1** | 55 | 19 (-) |
| (Bacteria; Cyanobacteriota; Cyanobacteriia; Pseudanabaenales; Pseudanabaenaceae) **Pseudanabaena PCC-7429** | 100 | 20 (-) |

as input for regression and not only predictive indicator genera (Supplementary Materials 8).

### 3.4. Linking functional processes to predictive indicator models

The use of NCBI to extract functional potential was imperfect with only 340 (48.6%) of genera having a reference genome at the genus level and 271 (38.8%) at family level. However, there were significant Pearson correlations between the relative abundance of taxa performing certain nitrogen cycling metabolic processes and predicted oxidised nitrogen in the XGB models. Three nitrogen cycle processes were weakly but significantly correlated with predicted oxidised nitrogen: denitrification processes (nitrate-N: $r = 0.198$, $p < 0.0001$, TON: $r = 0.202$, $p < 0.0001$, Supplementary Materials 9), urea metabolic processes (nitrate-N: $r = -0.092$, $p = 0.046$, TON: $r = -0.100$, $p = 0.029$, Supplementary Materials 9), and other nitrogen cycling processes (nitrate-N: $r = 0.191$, $p < 0.0001$, TON: $r = 0.192$, $p < 0.0001$, Supplementary Materials 9). These suggest that taxa that perform these functional processes increase with increased levels of nitrate-N and TON. The relative abundance of genera performing nitrogen fixation, and nitrate assimilation showed no significant trends with predicted nitrate-N or TON (Supplementary Materials 9).

**Table 4**
The number of genera that had a threshold response to nitrate-N and TON which met purity (0.8) and reliability (0.95) filters. Tolerant genera were those that increased in response to a threshold change whereas sensitive genera decreased in response to a threshold change.

| | | Genera for TITAN model for nitrate | | |
|---|---|---|---|---|
| | | No threshold response | Number of sensitive | Number of tolerant |
| Genera for TITAN model for TON | No threshold response | 37 | 1 | 1 |
| | Number of sensitive | 1 | 60 | 0 |
| | Number of tolerant | 1 | 0 | 92 |

### 3.5. Linking functional processes to predictive-threshold indicators

Genera were split into three groups based on their final indicator status (Fig. 3), and the proportion of taxa involved in these pathways was compared. Some trends were observed: the predictive-sensitive indicator group had a higher proportion of genera which performed nitrogen fixation, urea metabolism, and nitrate assimilation. Whereas the predictive-tolerant indicator group had a higher proportion of genera associated with denitrification and other nitrogen cycling genes. However, none of these were significantly different.

## 4. Discussion

This study developed an analytical framework to investigate and identify potential microbial indicators of river ecosystem health. Using the framework, we identified bacterial genera that were predictive indicators of nitrate-N and TON. We subsequently identified which of these predictive indicators were responsive to specific concentrations of nitrate-N and TON and change point concentrations (i.e. where over half of the bacterial community respond) for nitrate-N and TON. We then linked these predictive-threshold indicators to metabolic pathways involved in nitrogen cycling. This integrative approach provides a more robust scientific foundation based on organisms performing important ecological functions, to inform and support environmental regulators in protecting and restoring ecosystem health.

### 4.1. Framework identifies ecological relevant microbial indicators

Our framework identified 156 bacterial genera that are potential predictive-threshold indicators of oxidised nitrogen concentration, with the models accounting for a large proportion of the observed variance in both nitrate-N ($R^2$:0.654) and TON ($R^2$:0.650), and the associated SHAP values form the basis of a bacterial index for oxidised nitrogen in rivers.

Multiple genera identified as predictive-threshold indicators play

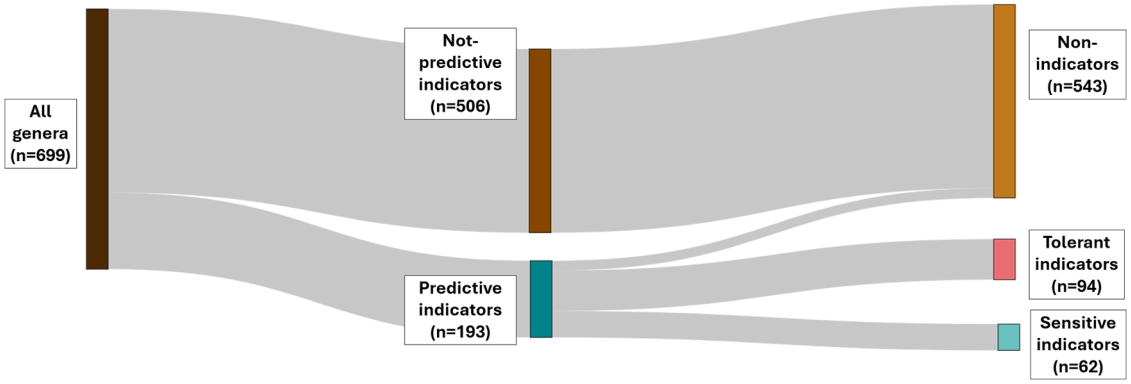



**Fig. 3.** Sankey diagram showing flow of genera and how predictive-threshold indicators were derived. Predictive indicators were derived from XGB modelling and passed to TITAN to determine indicators which were also threshold indicators.

key roles in the nitrogen cycling processes, reinforcing the ecological relevance of our framework. For example, genera involved in denitrification processes within nitrogen cycles, such as *Denitratisoma* (Fahrbach et al. 2006), were positively associated with nitrate-N and TON. Thus, we suggest that these genera are limited by oxidised nitrogen at lower concentrations and increase in abundance at higher concentrations. Furthermore, we found that the relative abundance of genera involved in denitrification pathways was weakly but positively associated with nitrate-N concentrations. Previous studies have similarly shown that denitrifying genes and denitrifying bacteria are enriched in response to nitrate-N concentrations (Zhang, S et al. 2016; Zhang, D et al. 2022). In addition, other genera involved in nitrogen cycling, such as *Anaeromyxobacter* (Masuda et al. 2020), showed negative relationships with nitrate-N and TON. Such genera are known to fix their own nitrate or utilise urea as a source of nitrogen if required, which may provide a competitive advantage at low oxidised nitrogen concentrations. This is further supported by a negative association between the relative abundance of urea metabolising genera and concentrations of nitrate-N. This competitive advantage potentially allows sensitive indicators to utilise alternate sources of nitrogen even with low externally available nitrate-N but are subsequently outcompeted by tolerant genera as nitrate-N concentrations increase. These findings suggest the framework is identifying ecologically relevant taxa as important predictive-threshold indicators of nitrate-N and TON, however studies using direct measurements of gene pathways are needed to confirm and strengthen this link.

Of the predictive indicators, TITAN identified the sensitive 0.30 $mgL^{-1}$ (0.30 $mgL^{-1}$) and tolerant 3.78 $mgL^{-1}$ (3.97 $mgL^{-1}$) change points which represent the concentration of nitrate-N (and TON) which cause the most ecological change in bacterial communities with only small differences in change points due to differing taxonomic specificity (Supplementary materials 10). Associated nitrate-N and TON concentrations at 13.0% (12.9% of TON) of corresponding biofilm samples were below the sensitive change point, and 59.8% (61.2% of TON) were below the tolerant. In associated work, TITAN analysis using ASVs instead of genera using the same dataset found similar change points of nitrate-N reported in this study (1.60 $mgL^{-1}$ for nitrate-N sensitive ASVs and 4.07 $mgL^{-1}$ for tolerant ASVs, Thorpe et al. 2026). A similar study by Pilgrim et al. (2022), who undertook a TITAN analysis of catchment-scale riverine biofilms in the USA, also found similar change point concentrations for total nitrogen for sensitive bacterial ASVs at 0.467 $mgL^{-1}$ and 0.772 $mgL^{-1}$ for tolerant ASVs. Notably, Sundermann et al. (2015) found similar concentrations impacting freshwater macroinvertebrate communities in Germany (approximately 1 $mgL^{-1}$ and 3 $mgL^{-1}$ for sensitive and tolerant macroinvertebrates, respectively). However, it is unclear whether this impact on macroinvertebrate communities was caused by nitrate toxicity or eutrophication caused by the microbiome itself which would explain why similar change points have been observed (Harding et al. 1999; Camargo et al. 2005; Wang et al. 2021). Nevertheless, these studies find similar ecological change points for biological communities, all within the same order of magnitude, suggesting that specified change points may be widespread rather than localised, and further supports the validation of our analytic framework in a biomonitoring context.

### 4.2. *Further development of tools for managing impacts on ecosystems*

The framework developed in this study could have broad applicability for managing impacts on ecosystems, although further development is required. The framework establishes predictive indicator importance by the SHAP analysis of the XGB models, as well as using TITAN to determine initial thresholds for ecological sensitivities to oxidised nitrogen in organisms involved in nitrogen cycling. Thus, our results lay the foundations for using microbial communities as novel indicators for managing ecosystems and informing decisions, which can be built upon to develop metrics for ecosystem health.

Further work and optimisation will improve our understanding of how oxidised nitrogen impacts communities which are directly involved in nitrogen cycling in aquatic ecosystems. Understanding could then be applied to establish causal relationships with both oxidised nitrogen and other ecosystem components. For example, models may benefit from an expanded dataset using ASVs instead of genera, or exploration of other methods of aggregating ASVs such as functional trait-based classification (Krause et al. 2014; Yang 2021) or network analysis to identify groups of interacting bacteria (Ma et al. 2020; Thorpe et al. 2026). Similarly, the analysis of functional gene composition and expression using mesocosms may provide a more comprehensive understanding of ecosystem dynamics than functional inferences based on taxonomy alone (Thorpe et al. 2025). Finally, other taxonomic groups across trophic levels should be evaluated to better understand the ecological impact of pressures on the ecosystem using a more holistic approach. Based on this novel framework, future work could aim to further develop microbial biomonitoring into a tool for water quality management, that could be utilised not only in England but also worldwide.

We used TITAN analysis to detect points at which the bacterial community changed and to offer this as a possible method for deriving thresholds for oxidised nitrogen in rivers. This is an increasingly common approach for deriving or evaluating thresholds for physicochemical variables (Sundermann et al. 2015; Kelly et al. 2019; Romero et al. 2019; Pilgrim et al. 2022). However, further research is required to ensure its robustness for formal regulation. Optimisation should include trialling a typology approach to evaluate the applicability of different microbial indicators across diverse river systems and to understand the influence of interacting environmental variables. However, this study identifies approximate points along the nitrate concentration gradient at which the underlying ecology changes,

regardless of whether these are practical regulatory targets.

A further challenge is linking these ecological change points to regulatory targets, because it remains unclear how policy defines ecological ambition with respect to oxidised nitrogen in inland waters. The Nitrates Directive (European Union 1991; Nikolaidis et al. 2025) expresses this ambition as avoiding "eutrophication". The WFD is more specific, requiring the achievement of GES, formally defined by the condition of a limited number of components of the riverine ecosystem. A broad narrative framework that links nitrogen cycle processes to current regulatory ecological impact targets would support the interpretation of TITAN outputs and could stimulate alternative approaches for setting thresholds.

## 5. Conclusion

We developed and applied a novel analytical framework to initiate the development of a bacterial metric for assessing oxidised nitrogen in freshwater riverine ecosystems. The framework identified ecologically relevant bacterial indicators of oxidised nitrogen and identified concentrations of ecological change within these communities. The framework developed here could be used to investigate new and emerging pressures, assess the impact of perturbations on functional processes within the environment, and further investigate the roles that microbes play in supporting a healthy environment. This integrative approach provides a robust scientific foundation to inform and support environmental regulators in protecting and restoring ecosystem health.

## Code/Data availability

The FASTQ files are accessible at the ENA (accession number: PRJEB90117). The bioinformatics scripts are available at Defra-Data-Science-Centre-of-Excellence/EA_CSG_amplicon_seq_processing_biofilms: Processing of amplicon sequencing data for EA biofilm project and the framework with filtered genera and oxidised nitrogen concentrations available at Micro-Biology/Warren_et_al_-_microbial_indicators.

## CRediT authorship contribution statement

**Jonathan Warren:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Caitlin de Vries:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Laura H. Hunt:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Amy C. Thorpe:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Susheel Bhanu Busi:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Martyn G. Kelly:** Writing – review & editing. **Dina-Leigh Simons:** Writing – review & editing. **Joe D. Taylor:** Writing – review & editing. **Daniel S. Read:** Writing – review & editing, Funding acquisition, Conceptualization. **Kerry Walsh:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors would like to thank Dr Jonathan Porter, Sean Butler, and Alan Wan from the Environment Agency's National Monitoring Laboratories for their support collating and shipping the biofilm samples for further processing. This work was funded by the Environment Agency under research project SC220034. Jonathan Warren and Kerry Walsh were supported by Natural Environment Research Council (NERC) grant NE/X015777/1. Amy Thorpe, Susheel Bhanu Busi, and Daniel Read were supported by NERC grant NE/X015947/1.

The views expressed in this paper are the views of the authors and do not represent the position of the Environment Agency or other employer organisations.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2026.125959.

## References

Ai, D., Pan, H., Han, R., Li, X., Liu, G., Xia, L.C., 2019. Using decision tree aggregation with random forest model to identify gut microbes associated with colorectal cancer. Genes. (Basel) 10. https://doi.org/10.3390/genes10020112.

Akinnawo, S.O., 2023. Eutrophication: causes, consequences, physical, chemical and biological techniques for mitigation strategies. Environ. Chall. 12. https://doi.org/10.1016/j.envc.2023.100733.

Baker, M.E., King, R.S., 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods Ecol. Evol. 25, 25–37. https://doi.org/10.1111/J.2041-210X.2009.00007.X.

Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., Packmann, A.I., 2016. The ecology and biogeochemistry of stream biofilms. Nat. Rev. Microbiol. 14, 251–263. https://doi.org/10.1038/nrmicro.2016.15.

Biggs, B.J.F., 2000. Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. J. North Am. Benthol Soc. 19, 17–31. https://doi.org/10.2307/1468279.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brown, J.A., Robertson, B.L., McDonald, T., 2015. Spatially balanced sampling: application to environmental surveys. Procedia Environ. Sci. 27, 6–9. https://doi.org/10.1016/j.proenv.2015.07.108.

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

Camargo, J.A., Alonso, A., Salamanca, A., 2005. Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. Chemosphere 58, 1255–1267. https://doi.org/10.1016/j.chemosphere.2004.10.044.

Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk, S., Blackstock, K. L., Borics, G., Borja, A., Feld, C.K., Ferreira, M.T., Globevnik, L., Grizzetti, B., Hendry, S., Hering, D., Kelly, M., Langaas, S., Meissner, K., Panagopoulos, Y., Penning, E., Rouillard, J., Sabater, S., Schmedtje, U., Spears, B.M., Venohr, M., van de Bund, W., Solheim, A.L., 2019. Protecting and restoring Europe's waters: an analysis of the future development needs of the Water Framework Directive. Sci. Total Environ. 658, 1228–1238. https://doi.org/10.1016/j.scitotenv.2018.12.255.

Chang, H.X., Haudenshield, J.S., Bowen, C.R., Hartman, G.L., 2017. Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. Front. Microbiol. 8, 1–11. https://doi.org/10.3389/fmicb.2017.00519.

Chen T., Guestrin C. (2016) XGBoost: a scalable tree boosting system. arXiv. https://doi.org/10.48550/arXiv.1603.02754.

Codello, A., Hose, G.C., Chariton, A., 2022. Microbial co-occurrence networks as a biomonitoring tool for aquatic environments: a review. Mar. Freshw. Res. 409–422. https://doi.org/10.1071/mf22045.

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., Pawlowski, J., 2019. Embracing environmental genomics and machine learning for routine biomonitoring. Trends. Microbiol. 27, 387–397. https://doi.org/10.1016/j.tim.2018.10.012.

Davidson-Pilon C. (2024) lifelines, survival analysis in Python.

Dolman, A.M., Mischke, U., Wiedner, C., 2016. Lake-type-specific seasonal patterns of nutrient limitation in German lakes, with target nitrogen and phosphorus concentrations for good ecological status. Freshw. Biol. 61, 444–456. https://doi.org/10.1111/fwb.12718.

Dutta, A., Goldman, T., Keating, J., Burke, E., Williamson, N., Dirmeier, R., Bowman, J. S., 2022. Machine learning predicts biogeochemistry from microbial community structure in a complex model system. Microbiol. Spectr. 10. https://doi.org/10.1128/spectrum.01909-21.

EEA (2024) Europe's state of water 2024 the need for improved water resilience. https://doi.org/10.2800/02236.

Environment Agency (2024a) Molecular data generation and preliminary analysis of river microbial biofilm communities.

Environment Agency, 2024b. Nitrates: challenges for the water environment - GOV.UK. Available from: https://www.gov.uk/government/publications/nitrates-challenges-for-the-water-environment.

European Union, 1991. Council Directive 91/676/EEC of 12 December 1991 Concerning the Protection of Waters Against Pollution Caused by Nitrates from Agricultural Sources (Nitrates Directive), 375. Official Journal of the European Communities L, pp. 1–8, 31.12.1991Available from. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31991L0676. December 8, 2025.

European Union, 2000. Water Framework Directive 2000/60/EC, 269. Official Journal of the European Communities L, pp. 1–15.

Fahrbach, M., Kuever, J., Meinke, R., Kämpfer, P., Hollender, J., 2006. Denitratisoma oestradiolicum gen. nov., sp. nov., a 17 β-oestradiol-degrading, denitrifying betaproteobacterium. Int. J. Syst. Evol. Microbiol. 56, 1547–1552. https://doi.org/10.1099/ijs.0.63672-0.

Falkowski, P.G., Fenchel, T., Delong, E.F., 2008. The microbial engines that drive earth's biogeochemical cycles. Science (1979) 320, 1034–1039. https://doi.org/10.1126/science.1153213.

Folt, C.L., Chen, C.Y., Moore, M.V., Burnaford, J., 1999. Synergism and antagonism among multiple stressors. Limnol. Oceanogr. 44, 864–877. https://doi.org/10.4319/lo.1999.44.3_part_2.0864.

Frühe, L., Dully, V., Forster, D., Keeley, N.B., Laroche, O., Pochon, X., Robinson, S., Wilding, T.A., Stoeck, T., 2021. Global trends of benthic bacterial diversity and community composition along organic enrichment gradients of Salmon farms. Front. Microbiol. 12, 1–17. https://doi.org/10.3389/fmicb.2021.637811.

Gerhard, W.A., Gunsch, C.K., 2019. Microbiome composition and implications for ballast water classification using machine learning. Sci. Total Environ. 691, 810–818. https://doi.org/10.1016/j.scitotenv.2019.07.053.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Ghannam, R.B., Techtmann, S.M., 2021. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Comput. Struct. Biotechnol. J. 19, 1092–1107. https://doi.org/10.1016/j.csbj.2021.01.028.

Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: successes and aspirations. BMC. Biol. 12, 1–4. https://doi.org/10.1186/s12915-014-0069-1.

Gomez Isaza, D.F., Cramp, R.L., Franklin, C.E., 2018. Negative impacts of elevated nitrate on physiological performance are not exacerbated by low pH. Aquatic Toxicology 200, 217–225. https://doi.org/10.1016/j.aquatox.2018.05.004.

Gomez Isaza, D.F., Cramp, R.L., Franklin, C.E., 2020. Living in polluted waters: a meta-analysis of the effects of nitrate and interactions with other environmental stressors on freshwater taxa. Environ. Pollut. 261, 114091. https://doi.org/10.1016/j.envpol.2020.114091.

Guseva, K., Darcy, S., Simon, E., Alteio, L.V, Montesinos-Navarro, A., Kaiser, C., 2022. From diversity to complexity: microbial networks in soils. Soil Biol Biochem. 169, 108604. https://doi.org/10.1016/j.soilbio.2022.108604.

Hallin, S., 2023. Environmental microbiology going computational—Predictive ecology and unpredicted discoveries. Environ. Microbiol. 25, 111–114. https://doi.org/10.1111/1462-2920.16232.

Harding, J.S., Young, R.G., Hayes, J.W., Shearer, K.A., Stark, J.D., 1999. Changes in agricultural intensity and river health along a river continuum. Freshw. Biol. 42, 345–357. https://doi.org/10.1046/j.1365-2427.1999.444470.x.

Hilton, J., O'Hare, M., Bowes, M.J., Jones, J.I., 2006. How green is my river? A new paradigm of eutrophication in rivers. Sci. Total Environ. 365, 66–83. https://doi.org/10.1016/j.scitotenv.2006.02.055.

HM Government, 2017. The Water Environment (Water Framework Directive) (England and Wales) Regulations 2017. Available from. https://www.legislation.gov.uk/uksi/2017/407/contents/made.

Howden, N.J.K., Burt, T.P., Worrall, F., Mathias, S.A., Whelan, M.J., 2013. Farming for water quality: balancing food security and nitrate pollution in UK river basins. Ann. Assoc. Am. Geogr. 103, 397–407. https://doi.org/10.1080/00045608.2013.754672.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree of life. Nat. Microbiol. 1, 1–6. https://doi.org/10.1038/nmicrobiol.2016.48.

Jarvie, H.P., Smith, D.R., Norton, L.R., Edwards, F.K., Bowes, M.J., King, S.M., Scarlett, P., Davies, S., Dils, R.M., Bachiller-Jareno, N., 2018. Phosphorus and nitrogen limitation and impairment of headwater streams relative to rivers in Great Britain: a national perspective on eutrophication. Sci. Total Environ. 621, 849–862. https://doi.org/10.1016/j.scitotenv.2017.11.128.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. 53, 457–481. https://doi.org/10.1080/01621459.1958.10501452.

Kelly, M., Juggins, S., Mann, D.G., Sato, S., Glover, R., Boonham, N., Sapp, M., Lewis, E., Hany, U., Kille, P., Jones, T., Walsh, K., 2020. Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. Ecol. Indic. 118, 106725. https://doi.org/10.1016/j.ecolind.2020.106725.

Kelly, M.G., Chiriac, G., Soare-Minea, A., Hamchevici, C., Juggins, S., 2019. Use of phytobenthos to evaluate ecological status in lowland Romanian lakes. Limnologica 77, 125682. https://doi.org/10.1016/j.limno.2019.125682.

Kelly, M.G., Phillips, G., Teixeira, H., Várbíró, G., Salas Herrero, F., Willby, N.J., Poikane, S., 2022. Establishing ecologically-relevant nutrient thresholds: a tool-kit with guidance on its use. Sci. Total Environ. 807. https://doi.org/10.1016/j.scitotenv.2021.150977.

Kennedy, R., Arthur, W., Keegan, B.F., 2011. Long-term trends in benthic habitat quality as determined by Multivariate AMBI and Infaunal Quality Index in relation to natural variability: a case study in Kinsale Harbour, south coast of Ireland. Mar. Pollut. Bull. 62, 1427–1436. https://doi.org/10.1016/j.marpolbul.2011.04.030.

Klopfenstein, D.V, Zhang, L., Pedersen, B.S., Ramírez, F., Vesztrocy, A.W., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., Tang, H., 2018. GOATOOLS: a Python library for gene ontology analyses. Sci. Rep. 8, 1–17. https://doi.org/10.1038/s41598-018-28948-z.

Kosciolek, T., Victor, T.A., Kuplicki, R., Rossi, M., Estaki, M., Ackermann, G., Knight, R., Paulus, M.P., 2021. Individuals with substance use disorders have a distinct oral microbiome pattern. Brain, Behav., Immun. - Health 15. https://doi.org/10.1016/j.bbih.2021.100271.

Krause, S., Le Roux, X., Niklaus, P.A., van Bodegom, P.M., Lennon, T.J.T, Bertilsson, S., Grossart, H.P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Front. Microbiol. 5, 1–10. https://doi.org/10.3389/fmicb.2014.00251.

Krishnamurthy, S.V, Smith, G.R., 2010. Growth, abnormalities, and mortality of tadpoles of American toad exposed to combinations of malathion and nitrate. Environ. Toxicol. Chem. 29, 2777–2782. https://doi.org/10.1002/etc.331.

Kuypers, M.M.M., Marchant, H.K., Kartal, B., 2018. The microbial nitrogen-cycling network. Nat. Rev. Microbiol. 16, 263–276. https://doi.org/10.1038/nrmicro.2018.9.

Lanzén, A., Mendibil, I., Borja, Á., Alonso-Sáez, L., 2021. A microbial mandala for environmental monitoring: predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay. Mol. Ecol. 30, 2969–2987. https://doi.org/10.1111/mec.15489.

Li, M., Sun, H., Huang, Y., Chen, H., 2024. Shapley value: from cooperative game to explainable artificial intelligence. Auton. Intell. Syst. 4. https://doi.org/10.1007/s43684-023-00060-8.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Ma, B., Wang, Y., Ye, S., Liu, S., Stirling, E., Gilbert, J.A., Faust, K., Knight, R., Jansson, J. K., Cardona, C., Röttjers, L., Xu, J., 2020. Earth microbial co-occurrence network reveals interconnection pattern across microbiomes. Microbiome 8, 1–12. https://doi.org/10.1186/s40168-020-00857-2.

Maberly, S.C., Pitt, J.A., Davies, P.S., Carvalho, L., 2020. Nitrogen and phosphorus limitation and the management of small productive lakes. Inland. Waters. 10, 159–172. https://doi.org/10.1080/20442041.2020.1714384.

Maberly, S.C., King, L., Dent, M.M., Jones, R.I., Gibson, C.E., 2002. Nutrient limitation of phytoplankton and periphyton growth in upland lakes. Freshw. Biol. 47, 2136–2152. https://doi.org/10.1046/j.1365-2427.2002.00962.x.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 17, 10–12. https://doi.org/10.14806/ej.17.1.200.

Masuda, Y., Yamanaka, H., Xu, Z.X., Shiratori, Y., Aono, T., Amachi, S., Senoo, K., Itoh, H., 2020. Diazotrophic anaeromyxobacter isolates from soils. Appl. Environ. Microbiol. 86, 1–12. https://doi.org/10.1128/AEM.00956-20.

McElhinney, J.M.W.R., Catacutan, M.K., Mawart, A., Hasan, A., Dias, J., 2022. Interfacing machine learning and microbial omics: a promising means to address environmental challenges. Front. Microbiol. 13. https://doi.org/10.3389/fmicb.2022.851450.

Morrissy, J.G., Currell, M.J., Reichman, S.M., Surapaneni, A., Megharaj, M., Crosbie, N. D., Hirth, D., Aquilina, S., Rajendram, W., Ball, A.S., 2021. Nitrogen contamination and bioremediation in groundwater and the environment: a review. Earth. Sci. Rev. 222, 103816. https://doi.org/10.1016/j.earscirev.2021.103816.

Neal, C., Hilton, J., Wade, A.J., Neal, M., Wickham, H., 2006. Chlorophyll-a in the rivers of eastern England. Sci. Total Environ. 365, 84–104. https://doi.org/10.1016/j.scitotenv.2006.02.039.

Newbold, L.K., Read, D.S., Taylor, J., Warren, J., Walsh, K., 2023. The extraction of cross-kingdom biofilm DNA using Zymo Reasearch Quick-DNA fecal /soil microbe kits. Sample Preparation 1–6.

Newbold L.K., Taylor J.D., Thorpe A.C., Warren J., Walsh K., Read D.S. (2025) DNA extraction methodology has a limited impact on multitaxa riverine benthic metabarcoding community profiles. https://doi.org/10.1002/edn3.70102.

Nikolaidis, N.P., Poikane, S., Bouraoui, F., Salas-Herrero, F., Free, G., Varkitzi, I., van de Bund, W., Kelly, M.G., 2025. Comparison of eutrophication assessment for the Nitrates and water Framework Directives: impacts and opportunities for streamlined approaches. Ecol. Indic. 177, 113375. https://doi.org/10.1016/j.ecolind.2025.113375.

O'Leary, N.A., Cox, E., Holmes, J.B., Anderson, W.R., Falk, R., Hem, V., Tsuchiya, M.T. N., Schuler, G.D., Zhang, X., Torcivia, J., Ketter, A., Breen, L., Cothran, J., Bajwa, H., Tinne, J., Meric, P.A., Hlavina, W., Schneider, V.A., 2024. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI datasets. Sci. Data 11, 1–10. https://doi.org/10.1038/s41597-024-03571-y.

Orr, J.A., Armitage, D.W., Letten, A.D., 2025. Coexistence theory for microbial ecology, and vice versa. Environ. Microbiol. 27. https://doi.org/10.1111/1462-2920.70072.

Paerl, H.W., Havens, K.E., Xu, H., Zhu, G., McCarthy, M.J., Newell, S.E., Scott, J.T., Hall, N.S., Otten, T.G., Qin, B., 2020. Mitigating eutrophication and toxic cyanobacterial blooms in large lakes: the evolution of a dual nutrient (N and P) reduction paradigm. Hydrobiologia 847, 4359–4375. https://doi.org/10.1007/s10750-019-04087-y.

Pasolli, E., Truong, D.T., Malik, F., Waldron, L., Segata, N., 2016. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS. Comput. Biol. 12, 1–26. https://doi.org/10.1371/journal.pcbi.1004977.

Pilgrim, E.M., Smucker, N.J., Wu, H., Martinson, J., Nietch, C.T., Molina, M., Darling, J. A., Johnson, B.R., 2022. Developing indicators of nutrient pollution in streams using 16S rRNA. Gene Metabarcoding of Periphyton-Associated Bacteria. Water (Switzerland) 14. https://doi.org/10.3390/w14152361.

Poikane, S., Kelly, M.G., Salas Herrero, F., Pitt, J.A., Jarvie, H.P., Claussen, U., Leujak, W., Lyche Solheim, A., Teixeira, H., Phillips, G., 2019. Nutrient criteria for surface waters under the European Water Framework Directive: current state-of-the-art, challenges and future outlook. Sci. Total Environ. 695. https://doi.org/10.1016/j.scitotenv.2019.133888.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic. Acids. Res. 41, 590–596. https://doi.org/10.1093/nar/gks1219.

R Core Team, 2022. R: a language and environment for statistical computing. Available from. https://www.r-project.org/.
Romero, F., Sabater, S., Font, C., Balcázar, J.L., Acuña, V., 2019. Desiccation events change the microbial response to gradients of wastewater effluent pollution. Water. Res. 151, 371–380. https://doi.org/10.1016/j.watres.2018.12.028.
Schindler, D.W., 2012. The dilemma of controlling cultural eutrophication of lakes. Proc. R. Soc. B: Biol. Sci. 279, 4322–4333. https://doi.org/10.1098/rspb.2012.1032.
Schindler, D.W., Carpenter, S.R., Chapra, S.C., Hecky, R.E., Orihel, D.M., 2016. Reducing phosphorus to curb lake eutrophication is a success. Environ. Sci. Technol. 50, 8923–8929. https://doi.org/10.1021/acs.est.6b02204.
Shaffer, J.P., Nothias, L.F., Thompson, L.R., Sanders, J.G., Salido, R.A., Couvillion, S.P., Brejnrod, A.D., Lejzerowicz, F., Haiminen, N., Huang, S., Lutz, H.L., Zhu, Q., Martino, C., Morton, J.T., Karthikeyan, S., Nothias-Esposito, M., Dührkop, K., Böcker, S., Kim, H.W., Aksenov, A.A., Bittremieux, W., Minich, J.J., Marotz, C., Bryant, M.K.M., Sanders, K., Schwartz, T., Humphrey, G., Vásquez-Baeza, Y., Tripathi, A., Parida, L., Carrieri, A.P., Beck, K.L., Das, P., González, A., McDonald, D., Ladau, J., Karst, S.M., Albertsen, M., Ackermann, G., DeReus, J., Thomas, T., Petras, D., Shade, A., Stegen, J., Song, S.J., Metz, T.O., Swafford, A.D., Dorrestein, P. C., Jansson, J.K., Gilbert, J.A., Knight, R., Angenant, L.T., Berry, A.M., Bittleston, L. S., Bowen, J.L., Chavarría, M., Cowan, D.A., Distel, D., Girguis, P.R., Huerta-Cepas, J., Jensen, P.R., Jiang, L., King, G.M., Lavrinienko, A., MacRae-Crerar, A., Makhalanyane, T.P., Mappes, T., Marzinelli, E.M., Mayer, G., McMahon, K.D., Metcalf, J.L., Miyake, S., Mousseau, T.A., Murillo-Cruz, C., Myrold, D., Palenik, B., Pinto-Tomás, A.A., Porazinska, D.L., Ramond, J.B., Rowher, F., RoyChowdhury, T., Sandin, S.A., Schmidt, S.K., Seedorf, H., Shade, A., Shipway, J.R., Smith, J.E., Stegen, J., Stewart, F.J., Tait, K., Thomas, T., Tucker, Y., U'Ren, J.M., Watts, P.C., Webster, N.S., Zaneveld, J.R., Zhang, S., 2022. Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. Nat. Microbiol. 7, 2128–2150. https://doi.org/10.1038/s41564-022-01266-x.
Shi, X., Mao, D., Song, K., Xiang, H., Li, S., Wang, Z., 2024. Effects of landscape changes on water quality: a global meta-analysis. Water. Res. 260, 121946. https://doi.org/10.1016/j.watres.2024.121946.
Snell, M.A., Barker, P.A., Surridge, B.W.J., Large, A.R.G., Jonczyk, J., Benskin, C.M.H., Reaney, S., Perks, M.T., Owen, G.J., Cleasby, W., Deasy, C., Burke, S., Haygarth, P. M., 2014. High frequency variability of environmental drivers determining benthic community dynamics in headwater streams. Environ. Sci. Process. Impacts. 16, 1629–1636. https://doi.org/10.1039/c3em00680h.
Stevenson, J.J., Bennett, B.J., Jordan, D.N., French, R.D., 2012. Phosphorus regulates stream injury by filamentous green algae, DO, and pH with thresholds in responses. Hydrobiologia 695, 25–42. https://doi.org/10.1007/s10750-012-1118-9.
Sundermann, A., Leps, M., Leisner, S., Haase, P., 2015. Taxon-specific physico-chemical change points for stream benthic invertebrates. Ecol. Indic. 57, 314–323. https://doi.org/10.1016/j.ecolind.2015.04.043.
Thompson, J., Johansen, R., Dunbar, J., Munsky, B., 2019. Machine learning to predict microbial community functions: an analysis of dissolved organic carbon from litter decomposition. PLoS. One 14, 1–16. https://doi.org/10.1371/journal.pone.0215502.
Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Xu, Z.Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Song, S.J., Kosciolek, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., Agosto Rivera, J.L., Al-Moosawi, L., Alverdy, J., Amato, K.R., Andras, J., Angenent, L.T., Antonopoulos, D.A., Apprill, A., Armitage, D., Ballantine, K., Bárta, J., Baum, J.K., Berry, A., Bhatnagar, A., Bhatnagar, M., Biddle, J.F., Bittner, L., Boldgiv, B., Bottos, E., Boyer, D.M., Braun, J., Brazelton, W., Brearley, F.Q., Campbell, A.H., Caporaso, J.G., Cardona, C., Carroll, J.L., Cary, S.C., Casper, B.B., Charles, T.C., Chu, H., Claar, D.C., Clark, R.G., Clayton, J.B., Clemente, J.C., Cochran, A., Coleman, M.L., Collins, G., Colwell, R.R., Contreras, M., Crary, B.B., Creer, S., Cristol, D.A., Crump, B.C., Cui, D., Daly, S.E., Davalos, L., Dawson, R.D., Defazio, J., Delsuc, F., Dionisi, H.M., Dominguez-Bello, M.G., Dowell, R., Dubinsky, E.A., Dunn, P.O., Ercolini, D., Espinoza, R.E., Ezenwa, V., Fenner, N., Findlay, H.S., Fleming, I.D., Fogliano, V., Forsman, A., Freeman, C., Friedman, E.S., Galindo, G., Garcia, L., Garcia-Amado, M.A., Garshelis, D., Gasser, R.B., Gerdts, G., Gibson, M.K., Gifford, I., Gill, R.T., Giray, T., Gittel, A., Golyshin, P., Gong, D., Grossart, H.P., Guyton, K., Haig, S.J., Hale, V., Hall, R.S., Hallam, S.J., Handley, K. M., Hasan, N.A., Haydon, S.R., Hickman, J.E., Hidalgo, G., Hofmockel, K.S., Hooker, J., Hulth, S., Hultman, J., Hyde, E., Ibáñez-Álamo, J.D., Jastrow, J.D., Jex, A.R., Johnson, L.S., Johnston, E.R., Joseph, S., Jurburg, S.D., Jurelevicius, D., Karlsson, A., Karlsson, R., Kauppinen, S., Kellogg, C.T.E., Kennedy, S.J., Kerkhof, L. J., King, G.M., Kling, G.W., Koehler, A.V, Krezalek, M., Kueneman, J., Lamendella, R., Landon, E.M., Lanede Graaf, K., LaRoche, J., Larsen, P., Laverock, B., Lax, S., Lentino, M., Levin, I.I., Liancourt, P., Liang, W., Linz, A.M., Lipson, D.A., Liu, Y., Lladser, M.E., Lozada, M., Spirito, C.M., MacCormack, W.P., MacRae-Crerar, A., Magris, M., Martín-Platero, A.M., Martín-Vivaldi, M., Martínez, L.M., Martínez-Bueno, M., Marzinelli, E.M., Mason, O.U., Mayer, G.D., McDevitt-Irwin, J. M., McDonald, J.E., McGuire, K.L., McMahon, K.D., McMinds, R., Medina, M., Mendelson, J.R., Metcalf, J.L., Meyer, F., Michelangeli, F., Miller, K., Mills, D.A., Minich, J., Mocali, S., Moitinho-Silva, L., Moore, A., Morgan-Kiss, R.M., Munroe, P., Myrold, D., Neufeld, J.D., Ni, Y., Nicol, G.W., Nielsen, S., Nissimov, J.I., Niu, K., Nolan, M.J., Noyce, K., O'Brien, S.L., Okamoto, N., Orlando, L., Castellano, Y.O., Osuolale, O., Oswald, W., Parnell, J., Peralta-Sánchez, J.M., Petraitis, P., Pfister, C., Pilon-Smits, E., Piombino, P., Pointing, S.B., Pollock, F.J., Potter, C., Prithiviraj, B., Quince, C., Rani, A., Ranjan, R., Rao, S., Rees, A.P., Richardson, M., Riebesell, U., Robinson, C., Rockne, K.J., Rodriguezl, S.M., Rohwer, F., Roundstone, W., Safran, R. J., Sangwan, N., Sanz, V., Schrenk, M., Schrenzel, M.D., Scott, N.M., Seger, R.L., Seguinorlando, A., Seldin, L., Seyler, L.M., Shakhsheer, B., Sheets, G.M., Shen, C., Shi, Y., Shin, H., Shogan, B.D., Shutler, D., Siegel, J., Simmons, S., Sjöling, S., Smith, D.P., Soler, J.J., Sperling, M., Steinberg, P.D., Stephens, B., Stevens, M.A., Taghavi, S., Tai, V., Tait, K., Tan, C.L., Taş, N., Taylor, D.L., Thomas, T., Timling, I., Turner, B.L., Urich, T., Ursell, L.K., Van Der Lelie, D., Van Treuren, W., Van Zwieten, L., Vargas-Robles, D., Thurber, R.V., Vitaglione, P., Walker, D.A., Walters, W.A., Wang, S., Wang, T., Weaver, T., Webster, N.S., Wehrle, B., Weisenhorn, P., Weiss, S., Werner, J.J., West, K., Whitehead, A., Whitehead, S.R., Whittingham, L.A., Willerslev, E., Williams, A.E., Wood, S.A., Woodhams, D.C., Yang, Y., Zaneveld, J., Zarraonaindia, I., Zhang, Q., Zhao, H., 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457–463. https://doi.org/10.1038/nature24621.
Thorpe, A.C., Busi, S.B., Warren, J., Hunt, L.H., Walsh, K., Read, D.S., 2025. National-scale biogeography and function of river and stream bacterial biofilm communities. Nat. Commun. https://doi.org/10.1101/2025.03.05.641783, 2025.03.05.641783.
Thorpe, A.C., Busi, S.B., Warren, J., Newbold, L.K., Taylor, J.D., Walsh, K., Read, D.S., 2026. River microbiomes as sentinels of national-scale freshwater ecosystems. Glob. Chang. Biol. 32. https://doi.org/10.1111/gcb.70809.
van Vliet, M.T.H., Thorslund, J., Strokal, M., Hofstra, N., Flörke, M., Ehalt Macedo, H., Nkwasa, A., Tang, T., Kaushal, S.S., Kumar, R., van Griensven, A., Bouwman, L., Mosley, L.M., 2023. Global river water quality under climate change and hydroclimatic extremes. Nat. Rev. Earth. Environ. 4, 687–702.
Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J.A., Jansson, J.K., Caporaso, J.G., Fuhrman, J.A., Apprill, A., Knight, R., 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for Microbial Community Surveys. mSyst. 1, 1–10. https://doi.org/10.1128/msystems.00009-15.
Wang, H., García Molinos, J., Heino, J., Zhang, H., Zhang, P., Xu, J., 2021. Eutrophication causes invertebrate biodiversity loss and decreases cross-taxon congruence across anthropogenically-disturbed lakes. Environ. Int. 153, 106494. https://doi.org/10.1016/j.envint.2021.106494.
Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07.
Warren, J., Butler, S., Evens, N., Hunt, L., Kelly, M., Newbold, L., Read, D.S., Taylor, J.D., Walsh, K., 2024. Influence of storage time on the stability of diatom assemblages using DNA from riverine biofilm samples. Metabarcoding. Metagenom. 8, 169–186. https://doi.org/10.3897/mbmg.8.129227.
Whelan, M.J., Linstead, C., Worrall, F., Ormerod, S.J., Durance, I., Johnson, A.C., Johnson, D., Owen, M., Wiik, E., Howden, N.J.K., Burt, T.P., Boxall, A., Brown, C.D., Oliver, D.M., Tickner, D., 2022. Is water quality in British rivers "better than at any time since the end of the Industrial Revolution"? Sci. Total Environ. 843, 157014. https://doi.org/10.1016/j.scitotenv.2022.157014.
Wiering, M., Kirschke, S., Akif, N.U., 2023. Addressing diffuse water pollution from agriculture: do governance structures matter for the nature of measures taken? J. Environ. Manage 332, 117329. https://doi.org/10.1016/j.jenvman.2023.117329.
Wilkinson J.L., Boxall A.B.A., Kolpin D.W., Leung K.M.Y., Lai R.W.S., Wong D., Ntchantcho R., Pizarro J., Mart J., Echeverr S., Garric J., Chaumot A., Gibba P., Kunchulia I., Seidensticker S., Lyberatos G., Morales-salda J.M., Kang H. (2022) Pharmaceutical pollution of the world ' s rivers. 119: 1–10. https://doi.org/10.1073/pnas.2113947119/-/DCSupplemental.Published.
Wyness, A.J., Stoeck, T., Morrissey, B.J., Berrill, I., Davidson, K., FairBairn, C., MacIntyre, S., Webb, C., Foster, T., Brain, S., Svobodova, D., Clark, C., Ashley-Wheeler, V., Roosenmaallen, E., Pollard, P., Wilding, T.A., 2026. eDNA2IQI: an eDNA/random forest-based prediction tool that accurately predicts ecological status for regulatory-compliance assessment. Aquaculture 612, 743126. https://doi.org/10.1016/j.aquaculture.2025.743126.
Yang, Y., 2021. Emerging patterns of microbial functional traits. Trends. Microbiol. 29, 874–882. https://doi.org/10.1016/j.tim.2021.04.004.
Yu Il, S., H, Jeong, Shin, J., Shin, S.G., Abbas, A., Yun, D., Bae, H., Cho, K.H., 2024. Simulation models of microbial community, pH, and volatile fatty acids of anaerobic digestion developed by machine learning. J. Water. Process. Eng. 60, 105225. https://doi.org/10.1016/j.jwpe.2024.105225.
Yuan, L.L., Mitchell, R.M., Pollard, A.I., Nietch, C.T., Pilgrim, E.M., Smucker, N.J., 2023. Understanding the effects of phosphorus on diatom richness in rivers and streams using taxon–environment relationships. Freshw. Biol. 68, 473–486. https://doi.org/10.1111/fwb.14040.
Zhang, D., Li, M., Yang, Y., Yu, H., Xiao, F., Mao, C., Huang, J., Yu, Y., Wang, Y., Wu, B., Wang, C., Shu, L., He, Z., Yan, Q., 2022. Nitrite and nitrate reduction drive sediment microbial nitrogen cycling in a eutrophic lake. Water. Res. 220. https://doi.org/10.1016/j.watres.2022.118637.
Zhang, S., Pang, S., Wang, P., Wang, C., Guo, C., Addo, F.G., Li, Y., 2016. Responses of bacterial community structure and denitrifying bacteria in biofilm to submerged macrophytes and nitrate. Sci. Rep. 6, 1–10. https://doi.org/10.1038/srep36178.
Zhou, Y.H., Gallins, P., 2019. A review and tutorial of machine learning methods for microbiome host trait prediction. Front. Genet. 10, 1–14. https://doi.org/10.3389/fgene.2019.00579.