Ecological quality in freshwater streams is reflected across all three domains of life

de Jonge, Nadieh; Kuntke, Franziska; Hesselsøe, Martin; Nielsen, Jeppe Lund

Published in:
Ecological Indicators

DOI (link to publication from Publisher):
10.1016/j.ecolind.2021.108059

Publication date:
2021

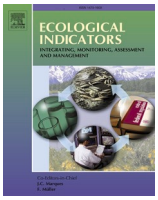
Document Version
Publisher's PDF, also known as Version of record

Ecological Indicators 130 (2021) 108059

Contents lists available at ScienceDirect

Ecological Indicators

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



Ecological quality in freshwater streams is reflected across all three domains of life

Nadieh de Jonge [a,b], Franziska Kuntke [a,c], Martin Hesselsøe [b,c], Jeppe Lund Nielsen [a,*]

[a] Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, DK-9220 Aalborg East, Denmark
[b] NIRAS A/S, Østre Havnegade 12, DK-9000 Aalborg, Denmark
[c] Amphi Consult ApS, Niels Jernes Vej 10, DK-9220 Aalborg East, Denmark

ARTICLE INFO

Keywords:
Ecological quality
Water stream quality
Metabarcoding
Bacteria
Archaea
Eukaryotes

ABSTRACT

Assessment of ecological quality in streaming surface water is often based on different Biological Quality Elements (BQE) such as plants, fish or invertebrates. Conventional stream-water quality assessment based on invertebrates as BQE relies on taxonomic expertise, which is costly and time consuming. Next-generation sequencing approaches for high-throughput analyses of diverse ecosystems are increasingly used for environmental monitoring and holds a great potential for application in stream-water quality assessments. This approach is to some extent hampered by the currently available reference databases representing freshwater invertebrates. In the present study we apply metabarcoding simultaneously targeting the 16S (prokaryotes) and 18S (eukaryotes) rRNA genes to capture a snapshot of the ecosystem composition across the three domains of life. Results based on the analysis of 50 selected Danish streams showed that the combined, as well as the domain-specific profiles can separate the samples into their respective ecological quality categories as reflected by the parallel conventional assessment based on macroinvertebrates as BQE. Furthermore, it was possible to suggest potential indicator organisms, from all three domains, which correlated specifically to the conventional data e.g. organisms with a strong correlation to ecological status across all categories. The results clearly showed that community structure in all three domains of life reflect the ecological status of the sample location. Hence, when applying a molecular approach for water-quality assessment we are not limited to the composition of visible BQE, such as macroinvertebrates. The microbial community composition in the streams may often capture an even better and more comprehensive and sensitive snapshot of the ecological quality of stream waters.

1. Introduction

The Water Framework Directive (WFD) outlines the requirements of the European Commission (EEC) for regular quality assessment of European surface waters (Bengtsson et al., 2012). The WFD target is to ensure that natural water bodies of the member states achieve at least a good ecological status. A large number of different protocols are applied for quality assessment of water bodies in the EEC member states. The implemented methods are based on an array of conventional biological quality elements such as macroinvertebrates, algae, plants, fish etc. combined with chemical/physical parameters (Birk et al., 2012). The diversity in methodology has shown the need for careful intercalibration and standardisation to obey the EEC directives. Furthermore, current assessment protocols rely heavily on invasive sampling and fauna identification by taxonomic experts, which is both time-consuming and costly (Leese et al., 2016). One example of stream-ecosystem assessment protocols is the Danish Stream Fauna Index (DSFI) (Skriver et al., 2000). This method uses macroinvertebrate diversity as Biological Quality Elements (BQE) (Agouridis et al., 2015; Elbrecht and Leese, 2017). The water quality of the stream ecosystems are categorised into one of seven groups, ranging from poor (1), to very good (7) (Skriver et al., 2000). The DSFI categories translate to the ecological status classes as described by the WFD (Baattrup-Pedersen et al., 2004).

It has been proposed that assessment methods can be improved through the use of molecular techniques such as metabarcoding (Blackman et al., 2019; Pawlowski et al., 2018). Next generation sequencing (NGS) has become an increasingly mainstream and convenient approach to perform analysis of diverse ecosystems. The use of standardised DNA barcoding based on universal genetic markers allows for the identification of species through sequence data (Leese et al.,

* Corresponding author.
E-mail address: jln@bio.aau.dk (J.L. Nielsen).

https://doi.org/10.1016/j.ecolind.2021.108059
Received 4 January 2021; Received in revised form 24 July 2021; Accepted 29 July 2021
Available online 3 August 2021

2016). Previous studies have yielded a number of protocols for sampling and DNA extraction for macroinvertebrate analyses in relation to freshwater quality assessment (Blackman et al., 2019), as well as broad range primer sets targeting invertebrate biomarkers (Elbrecht and Leese, 2017), laying the foundation for high-throughput assay development. Recently, an amplicon sequencing approach based on targeting the cytochrome *c* oxidase I (*COI*) gene was developed and successfully applied as an alternative method of performing water-quality assessments in streams (Kuntke et al., 2020).

The drawback of molecular approaches for bioassessment is that design of barcoding targets can be problematic. Often not all the desired diversity can be captured in a single barcode. Therefore, no consensus on genes and specific barcodes has been reached so far (Leese et al., 2016). Samples collected for fauna indexing according to water quality assessments contain DNA from all materials present within the ecosystem, including soil, water, plants and other organisms inhabiting the sampling site. By choosing a universal approach that captures the majority of diversity present in a sample, it could potentially be possible to report on the state and dynamics of entire ecosystem, rather than focusing on a single source of DNA in the sample. This approach has previously been used to characterise relationships between the prokaryotic and eukaryotic fractions of plankton in a temperate lake (Wurzbacher et al., 2017a).

Microbial communities of stream-water ecosystems have previously been shown to correlate with land usage, and the status of the stream environment (Lear et al., 2013). Furthermore, quality assessment of ground and stream water via detection of selected microbial indicators such as *Escherichia coli* and *Clostridium perfringens* has also shown promising results (Francy et al., 2000). Moreover, the superior size of bacterial and archaeal diversities have previously been linked to changes in ecosystem quality after an oil spill (Urakawa et al., 2012). Collectively, these findings suggest that microbial community studies may present a promising approach for fast, accurate and cost effective quality assessment of freshwater ecosystems and it has been suggested to include the microbiome in analysis to expand the scope of environmental assessment (Sagova-Mareckova et al., 2021).

In the present study we characterise the complete biome profile of 50 bulk samples of invertebrates collected from freshwater streams across Denmark. The results are aligned with the pre-determined ecological quality status based on macroinvertebrate composition as BQE (the DSFI protocol). A high-throughput metabarcoding approach using a universal primer set that targets all three domains (*Bacteria*, *Archaea* and *Eukaryota*) was chosen for this purpose. The potential of whole biome analysis for ecosystem quality determination is investigated. Furthermore, taxa from all three domains with potential as indicators of ecological status were explored.

2. Material and methods

2.1. Traditional DSFI analysis and conversion to ecological status classes

Sampling of 50 Danish stream ecosystem sites was performed using the kick-sampling method (Bradey and Ormerod, 2002) and subsequent DSFI analysis was performed by the Laboratory of Fish Ecology (NIRAS A/S, Allerød, Denmark) in accordance with the Danish standardised DSFI protocol (Skriver, 1999). Briefly, the collected samples were decanted using a sieve (mesh size 0.5 mm) prior to sample sorting for DSFI analysis. All collected material was saved and recombined after DSFI index scoring. The conventional DSFI analysis was performed under sterile conditions. The samples used for the DSFI analysis were stored in 96 % ethanol at 4 °C until further analysis. After DSFI analysis, the scores were converted to ecological quality classes using a previously proposed distribution (Baattrup-Pedersen et al., 2004). An overview of all samples, their DSFI quality category and ecological quality class is shown in Table 1.

Table 1
Overview of samples, DSFI quality and ecological status.

DSFI category	Ecological status	n
1–2	Bad	2
3	Poor	3
4	Moderate	7
5–6	Good	22
7	High	16

2.2. Sample homogenisation and DNA extraction

Sample homogenisation and DNA extraction were performed in a separate laboratory dedicated to environmental sample handling, and all surfaces were decontaminated using 70 % ethanol and RNAse AWAY (Thermo Fisher Scientific, USA). Prior to DNA extraction, the bulk samples were emptied into a mesh sieve (0.5 mm) and subsequently homogenised with a blender (JB 5160 BK, Braun GmbH, Germany) using 10 s cycles at speed setting 3, to avoid warming the sample. The sieve, blender and other tools were thoroughly washed and decontaminated using 70 % ethanol and RNAse AWAY (Thermo Fisher, USA) between each sample. Homogenised samples were kept on ice prior to DNA extraction. Total genomic DNA was extracted from 0.25 g of homogenised material using the QIAamp PowerFecal DNA kit (Qiagen, USA) according to manufacturer's recommendations. Extracted DNA was eluted to 50 μL elution buffer from the kit. To account for potential contaminants stemming from the extraction kit, a DNA extraction of twice UV treated nuclease-free water was generated to function as an additional negative control. Concentration of the extracted DNA was determined using Quant-IT Broad Range DNA assay kit (Thermo Fisher Scientific, USA) and an M200 Infinite PRO plate reader (TECAN, Switzerland).

2.3. Amplicon sequencing of SSU biomarkers

Molecular ecosystem profiles of the samples were captured using an amplicon sequencing approach targeting the genes coding for the small ribosomal subunits 16S (prokaryotes) and 18S (eukaryotes) simultaneously using a previously described universal primer set: 926F (5′-AAACTYAAAKGAATTGACGG-3′) and 1392R (5′-ACGGGCGGTGTGTRC-3′) (Engelbrektson et al., 2010). PCR reactions were prepared in a sterile laminated air flow bench, and pre- and post-PCR work was performed in separate areas of the laboratory to minimise contamination. Amplicon PCR was performed on 10 ng of extracted DNA per 25 μL PCR reaction (1X Platinum High Fidelity buffer (Thermo Fisher Scientific, USA), 2 mU Platinum Taq DNA polymerase HF (Thermo Fisher Scientific, USA), 1.5 mM $MgSO_4$, 400 nM of each dNTP and 400 nM of each primer), using the following PCR programme: 2 min initial denaturation at 95 °C, 30 cycles of 20 s at 95 °C, 30 s at 52 °C, 60 s at 72 °C, final elongation for 5 min at 72 °C. A positive control of known content and the DNA extraction non-template control (twice UV treated nuclease free water), as well as a PCR non-template control were included to monitor contamination and sequencing quality. PCR reactions were performed in duplicates, pooled afterwards and subsequently purified using AMPure XP bead protocol (Beckmann-Coulter, USA) using a sample:bead ratio of 5:4. Amplicon quality was assessed using TapeStation 2200 and D1000 ScreenTapes (Agilent, USA), and quantity was determined using Quant-IT High Sensitivity DNA assay kit (Thermo Fisher Scientific, USA). The amplicon libraries were subsequently barcoded in accordance with the Nextera XT barcoding protocol (Illumina, USA). Equimolar concentrations of the libraries were sequenced on a MiSeq platform using reagent kit v3 (2x300 PE) (Illumina, USA), with a final library pool concentration of 4 pM and 20 % PhiX spike-in.

2.4. Data processing and analysis

The obtained raw sequencing reads were quality checked and processed into ZOTUs using the AmpProc pipeline (v5.1) (https://github.com/eyashiro/AmpProc), based in USEARCH11 (Edgar, 2013) and Qiime 1.9.1 (Caporaso et al., 2010). The taxonomy assignment was performed using SILVA S138 (Quast et al., 2013), and the taxonomic assignment on the eukaryotic ZOTUs was polished using BLAST (Altschul et al., 1990). Due to a significant difference in average expected amplicon length between the prokaryotic (502 ± 11 bp) and the eukaryotic amplicon length (534 ± 52 bp), it was not expected to be possible to merge the majority of the eukaryotic sequences. The data treatment was therefore performed as follows: The entire dataset was first processed in single read mode, and based on the obtained diversity of the ZOTUs, the reverse reads (13,686 ZOTUs) were chosen for analysis over the forward reads (9,416 ZOTUs). Subsequently, the prokaryotic data was also processed in paired end mode to retrieve the highest possible taxonomic resolution for *Bacteria* and *Archaea*, the eukaryotic reads were removed from this dataset. To ensure that the single read and paired end read data from the prokaryotes were comparable, their beta diversity was compared, and found to be highly similar (data not shown).

Data analysis was performed using R version 4.0.2 (R Development Core Team, 2021) wrapped by RStudio version 1.3.959 (http://www.rstudio.com/). Visualisation of the obtained data and ordination analysis were performed using the packages ampvis2 (Andersen et al., 2018) and ggplot2 (Wickham, 2016). Estimated richness per sample was calculated using the Chao1 index, as calculated by the estimateR function from the vegan package (Oksanen et al., 2016). Differences between groups were tested using the non-parametric Wilcoxon rank sum test. Beta diversity was visualised using Canonical Correspondence analysis on Hellinger transformed read counts. Heatmaps were used to visualise the community compositions, and correlation analysis was performed using Spearman's correlations.

3. Results

Amplicon sequencing of 50 bulk samples from Danish freshwater streams yielded a total of 2,416,798 high quality sequence reads. The obtained sequencing depth was assessed using a rarefaction curve (Fig. S1), and based on this, a minimum 10,000 sequences per sample was set as the criterion for inclusion in the analysis. Fifty samples across all five ecological status categories entered the analysis with an average of 41,684 ± 7,628 sequences per sample. A total of 13,446 ZOTUs were detected across all samples, with the majority representing the domain *Bacteria* (11,672 ZOTUs, 87.8 % of total reads, followed by *Eukaryota* (282 ZOTUs, 7.3 % of total reads) and *Archaea* (233 ZOTUs, 0.8 % of total reads). A total of 1,259 ZOTUs (4.1 % of total reads) could not be classified and were removed from subsequent analyses.

3.1. Diversity and composition of molecular biome profiles in Danish freshwater streams

Diversity and distribution of the obtained biome profiles was assessed using the estimated richness as calculated per domain by alpha diversity index Chao1 (Fig. 1). The lowest average number of bacterial ZOTUs was observed in the samples from good ecological status (4,243 ± 1,414 ZOTUs) (Fig. 1a), while the highest was seen in the samples of poor ecological status (5,790 ± 1,777 ZOTUs). No immediate relationship between the bacterial diversity and ecological status of the sampled locations was apparent ($p > 0.05$), and a similar pattern of diversity was observed for the *Archaea* ZOTUs. A trend towards increased diversity with better ecological status was observed for the eukaryotic ZOTUs (Fig. 1b), where an average 54 ± 3 ZOTUs were observed in the samples representing bad ecological quality, opposite 93 ± 25 ZOTUs in samples of high ecological status. However, statistical testing revealed no significant differences ($p > 0.05$) in the estimated number of ZOTUs across all 5 ecological status categories in all three domains. Bacterial and eukaryotic ZOTUs had a relatively low variation between samples of the same groups, compared to *Archaea*, where a larger variation was observed.

The composition of the molecular biome profiles across ecological status categories was further examined per domain (Figs. S2–S4). Prior to analysis, all ZOTUs associated to chloroplasts were filtered from the dataset. The majority of the abundant bacterial community members were observed across all ecological status groups (Fig. S2), including multiple representatives of *Rhodoferax, Nitrospira, Comamonadaceae* and *Anaerolineaceae*. Representatives of the genus *Bacillus*, as well as one ZOTU representing *Nitrospira* were almost exclusively detected in samples representing good and high ecological quality. Representatives of the domain *Archaea* were detected sparsely in samples across all five ecological status categories (Fig. S3). The majority of abundant archaeal ZOTUs were associated to the class *Bathyarchaeia* and the family *Nitrososphaeraceae*. For the *Eukaryota*, several abundant organisms were observed across all ecological status groups, including representatives of the amphipod family *Gammaridae* and the diatom algae *Navicula* (Fig. S4). Other eukaryotic ZOTUs were more sparsely detected in the sampled locations, and included invertebrates, fungi and annelids. Representatives of the flatworm genus *Dugesia* were primarily observed in samples of good and high ecological status.

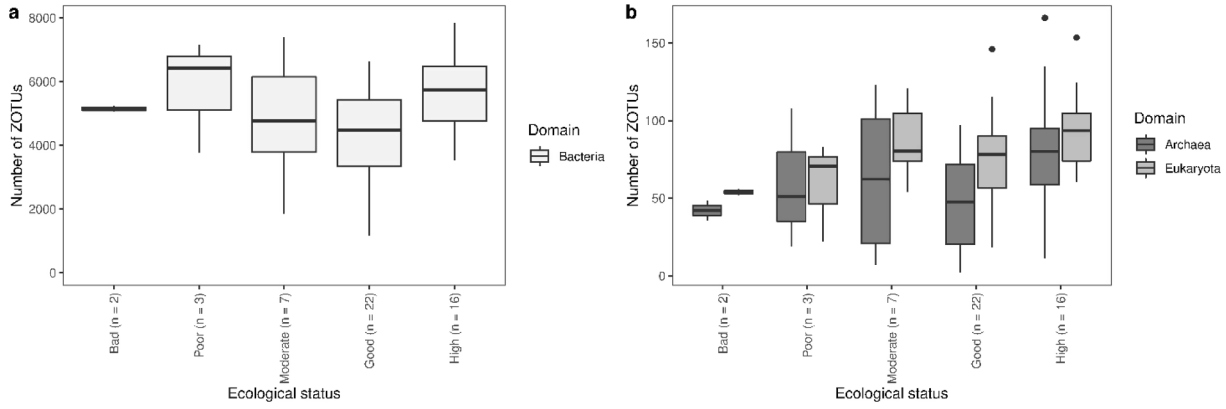


Fig. 1. Estimated richness (ChaoI) of the molecular biome profiles from 50 samples is shown as boxplots for *Bacteria* (a) and *Archaea* and *Eukaryota* (b), sorted by ecological status and divided into the three detected domains.

3.2. Ecological status is reflected in combined and domain specific molecular biome profiles

The relationship between ecological status and the composition of the molecular biome profile was investigated through beta diversity analysis with ordination. Unconstrained ordination analysis using correspondence analysis (CA) showed little to no separation between ecological status groups (data not shown). However, the use of the ecological status of the samples as a constraint for canonical correspondence analysis (CCA) revealed strong clustering and separation of the sampled locations for the whole biome data, as well as the three individual domains (Fig. 2). Near complete separation of the five ecological status groups was achieved for the whole biome data (Fig. 2a), while the *Bacteria* (Fig. 2b and *Eukaryota* (Fig. 2d) also showed clustering by ecological status but with overlap between categories. The *Archaea* ZOTUs were not able to separate the sampled locations in a meaningful way (Fig. 2c). The two lowest ecological statuses (Bad and Poor) were generally separated further away from the remaining samples, while overlap was seen primarily for locations of Good and High ecological status. Interestingly, the moderate locations showed a larger distribution for the *Archaea* along the horizontal axis as well, but it was not possible to further determine why.

3.3. Quality indicators are present in all three domains of life

Ecological indices for quality assessment rely on indicator organisms associated to the criteria described in the individual methods. To explore potential indicator organisms across the three domains analysed in the present study, three different types of relationships were investigated. The whole biome and individual domain data were both subjected to same criteria to extract ZOTUs of interest as potential indicators. The results from the whole biome dataset were comparable to the combined results of the individual domains (data not shown). The individual domains were therefore selected for detailed analysis to provide the highest taxonomic resolution possible. A selection of the best potential indicators identified across the three domains of life is shown in Fig. 3, the full result is displayed in Figs. S5–S7.

The first type of indicators was assumed to be associated to a single ecological status category, and extracted by selecting all ZOTUs that were observed in one third of samples in a single category (with a minimum of n = 2). This yielded 54, 2, and 0 ZOTUs for *Bacteria*, *Archaea* and *Eukaryota*, respectively. ZOTUs representing *Rhodobacteraceae*, *Flavobacterium* and *Ferruginibacter* were associated to streams of bad and poor ecological status, while ZOTUs representing *Pirellulaceae*, *Comamonadaceae* and *Pseudonocardia* were associated to those of good and high ecological status (Fig. 3a).

The second type of indicators was assumed to be associated to either low (bad-moderate, n = 12) or high (good-high, n = 38) quality ecosystems, under the definition of the WFD which states that all ecosystems must achieve an ecological status of good or higher. ZOTUs were extracted by selecting those that occurred in at least one third of samples in either group. This yielded 60, 0 and 3 ZOTUs for *Bacteria*, *Archaea* and *Eukaryota*, respectively. Representatives of *Ferribacterium* and *Rhodobacter* were found to be associated to streams of lower ecological status, while ZOTUs representing *Solibacillus*, *Sphingomonadaceae*, *Navicula* and *Dolichopeza* were associated with high quality ecosystems (Fig. 3b).

The final type of indicator was assumed to have a strong correlation to ecological status across all categories. ZOTUs that were observed in at least half of the samples in the dataset (n = 50) were extracted and a Spearman's correlation analysis was performed. A total of 546 ZOTUs met the criteria for *Bacteria*, and additional filtering for $-0.5 < \text{rho} > 0.5$ reduced this to 34 ZOTUs with strong correlation to ecological status.

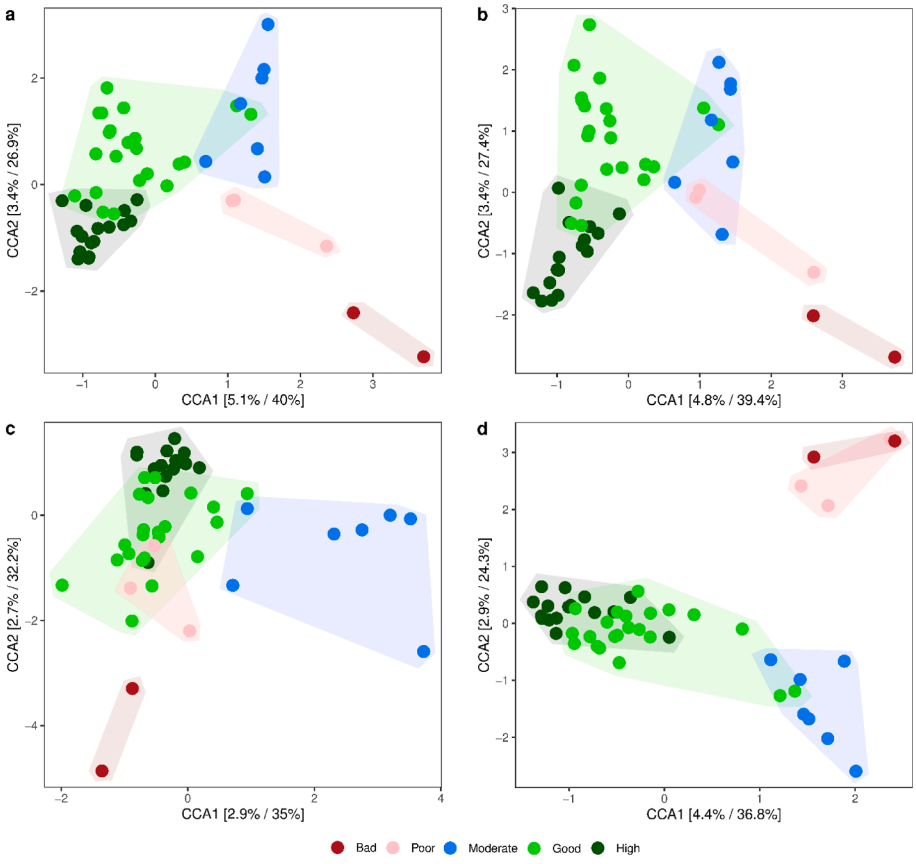


Fig. 2. Canonical correspondence analysis of freshwater stream biomes, constrained by ecological status, for the complete biome (a) *Bacteria* (b), *Archaea* (c), and *Eukaryota* (d). Samples are coloured by ecological status, and a polygon is drawn around locations with the same ecological status.

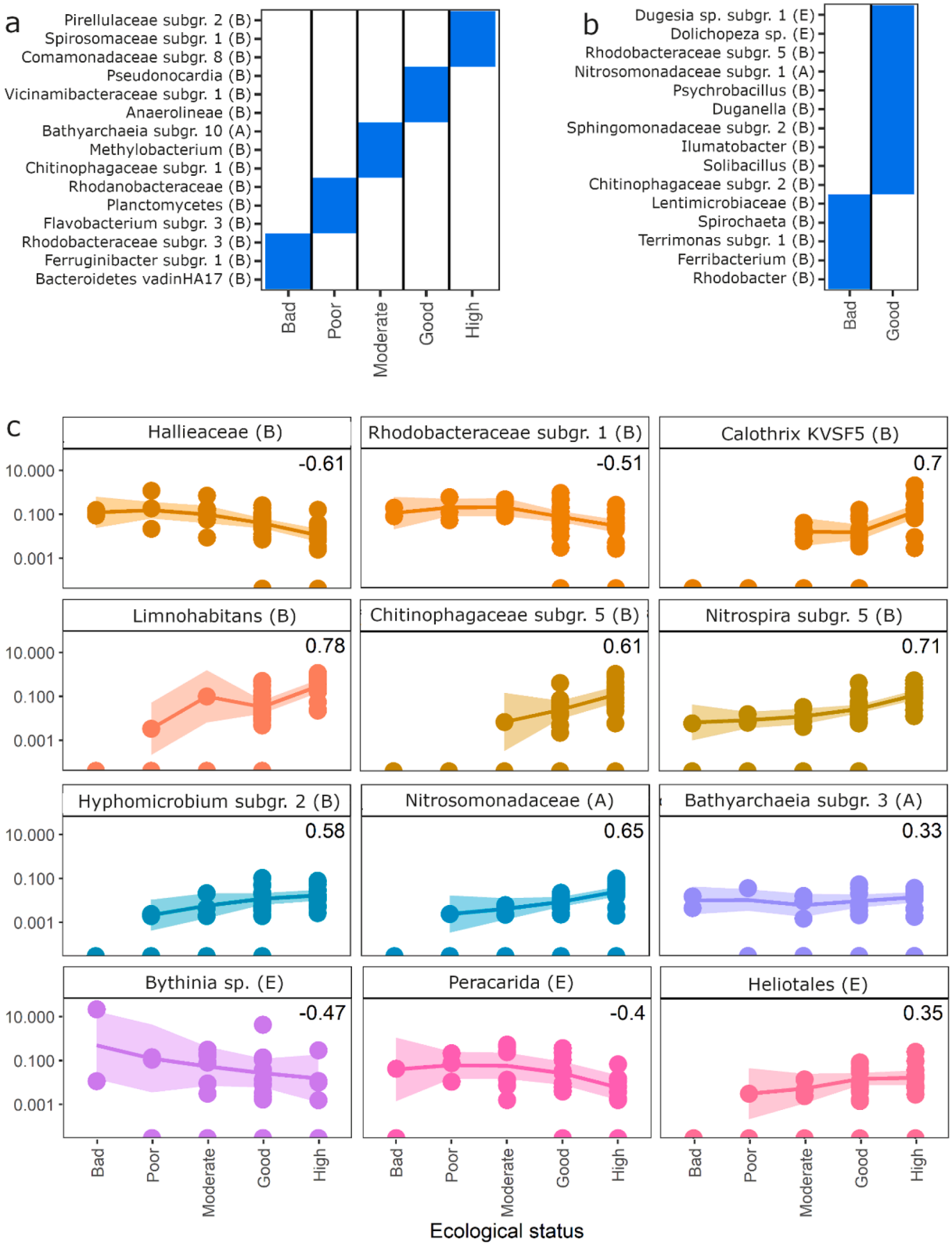


Fig. 3. Relationship between ecological status and occurrence of organisms from all three domains of life. ZOTUs from *Bacteria* (B), *Archaea* (A) and *Eukaryota* (E) with association to a single category (a), low or high quality ecosystems (b) or strong correlation to ecological status (c) were extracted from the dataset. A selection of the most influential ZOTUs for all three domains is displayed as heatmaps showing presence/absence and scatterplots with loess lines and 95 % confidence intervals with Spearman's rho displayed in each panel. The complete result of the analysis is shown in Figs. S5–S7.

The *Archaea* and *Eukaryota* did not yield ZOTUs with a strong correlation, but yielded a total of 6 and 14 ZOTUs, respectively. Strong negative correlation between occurrence and ecological status was observed for ZOTUs representing *Hallieaceae* (r = -0.61), *Rhodobacteraceae* (r = -0.51) and *Bythinia* sp. (r = -0.47), while a strong positive correlation was observed for ZOTUs representing *Calothrix* (r = 0.7), *Limnohabitans* (r = 0.78), *Nitrospira* (r = 0.71) and *Nitrosomonadaceae* (r = 0.65) (Fig. 3c).

4. Discussion

The present study aimed to characterise the biodiversity profiles of 50 Danish stream water sampling locations across three domains, using a metabarcoding approach that simultaneously targets the bacterial, archaeal and eukaryotic SSU biomarker. Furthermore, the potential of this method was assessed as a low cost and high throughput alternative to conventional stream-water quality assessments based on macroinvertebrates as BQE. The obtained combined biodiversity profiles, as well as the domain-specific community profiles were able to separate the sampled streams based on their respective ecological status and revealed potential indicator organisms for further studies from all three domains.

4.1. Biome analysis and composition of Danish freshwater streams

One aim of the present study was to assess the potential of using high-throughput technologies for stream-water quality assessment to increase sensitivity and simultaneously reduce labour intensity and cost. One of the steps in conventional bioassessment protocols is pre-treatment and sorting of the collected samples. Previous studies have developed generalised protocols for sample collection and handling for this purpose (Blackman et al., 2019). Our recently published results from a study involving metabarcoding of stream water bodies showed that direct sample homogenisation without pre-treatment was equally effective for molecular analysis of macroinvertebrate communities, compared to samples that had been washed and sorted prior to DNA extraction (Kuntke et al., 2020). The obtained sequencing results in the present study supports the validity of analysing bulk samples without extensive pre-treatment, as it was possible to obtain high quality sequences to a sufficient depth for detailed biome analysis from all three domains in 50 samples. The omission of pre-treatment steps from the bioassessment workflow has the potential to reduce the sample preparation time and the extent of the biases associated with this process. The present study applied a universal primer set in order to target all three domains, although with a reduced resolution relative to order- or family-optimized primer sets. However, the applied approach provides direct comparisons without introducing unnecessary primer biases and allows for a one-step analysis.

The obtained sequencing data was of high quality and shown to be of sufficient sequencing depth for high resolution analysis of domain-specific community profiles (Fig. S1), with the majority of the captured diversity originating from the bacterial communities present in the samples (Fig. 1a). This was as expected, as sediment material makes up a large fraction of the sample, and the sediment microbiome is one of the most complex biomes that has been studied (Battin et al., 2016). The methodology of analysing multiple domains for ecosystem quality studies has previously been attempted in a study involving the use of multiple biomarkers (*16S rRNA*, *18S rRNA* and *COI* genes) and sequencing technologies aimed to develop a framework for meta-systematic analysis of bulk samples containing arthropods (Gibson et al., 2014). Furthermore, a similar methodology to that used in the present study has previously been used for plankton characterisation in lakes, and analysis of the vertical spatial distribution of organisms present in lake sediments (Wurzbacher et al., 2017b; Wurzbacher et al., 2017a).

The sequencing data obtained in the present study reveal a measured diversity for all domains that is multiple times higher than previously reported. Furthermore, the novel approach of using a single primer set to capture a comprehensive picture of the three domains of the stream water biome simultaneously improves data handling and comparability, and highlights the quality and convenience of the chosen methodology. However, usage of universal primer sets has a drawback in terms of taxonomic coverage. So far, it has not been possible to capture all biodiversity using a single DNA barcode, and careful consideration is required when choosing the right barcode for each metabarcoding study (Creer et al., 2016). The limitation in the primers applied in the present study lead to a reduced diversity and taxonomic depth of the captured eukaryotic sequences. However, even this reduced taxonomic resolution was still sufficient to separate the sampled locations based on ecological status (Fig. 2d). Thus, the one-step analysis chosen for the present study was sufficient to address the biological question; whether or not a metabarcoding approach that simultaneously targets the SSU biomarkers in all three domains has potential for biological quality assessments in freshwater streams.

A number of known sediment associated bacteria were detected abundantly across samples of all qualities (Figs. S2–S4), including the genera *Rhodoferax*, *Nitrospira* and the family *Comamonadaceae*, as well as representatives of the archaeal family *Nitrososphaeraceae* (Battin et al., 2016). The composition of plant life in forest ecosystems has previously been used to predict soil microbiome profiles in grasslands, with moderate success (Leff et al., 2018), which highlights the relationship of the soil microbes with the flora and fauna of the ecosystem they are present in. Furthermore, microbial community monitoring has previously been applied in the tracking of bioremediation after oil spills in marine sediment environments (Acosta-González and Marqués, 2016; Urakawa et al., 2012). Moreover, the quality of the soil and its microbes has previously been shown to function as a predictor for ecosystem health based on land usage (Lear et al., 2013). Thus it may be suggested that measuring the microbiome as well as the fauna and flora composition could improve assessment resolution for quality-index studies in stream waters, which is supported by a recent review which also highlighted potential benefits of including the microbiome into environmental assessments (Sagova-Mareckova et al., 2021). Not surprisingly, a number of taxonomic groups containing organisms previously used as indicator species for bioassessment were also detected (Skriver et al., 2000), including representatives of *Diptera*, *Caenogastropoda*, *Mollusca* and *Coleoptera*.

4.2. Ecological status separates whole biome and individual domain compositions

It was not possible to separate the five ecological status groups based on the composition of the analysed streams on its own (data not shown). This was expected as the highly diverse and variable composition of the biome between samples obscures the relationship between ecological status and biological diversity. This observation is supported by previous studies exploring the potential of metabarcoding as an alternative for conventional bioassessments in freshwater streams (Kuntke et al., 2020). The discrepancy might also be explained by the lack of precision in conventional assessments. Another study exploring prediction of anthropogenic activity in rivers also found that the complete observed diversity could not explain ecosystem quality, and suggested the use of indicator organisms to specialise a potential model (Li et al., 2018). One potential approach to improve the ability to model the ecological quality of freshwater stream samples based on their biological composition would be to increase the sample size, as this would increase representation of the compositional variance in these ecosystems. A consequence hereof is that the model becomes empirically better the more it is used. Broad scale environmental assessment strategies using molecular techniques have previously been identified as a major breakthrough toward implementation of DNA-based techniques as the new standard method for ecosystem evaluation (Cordier et al., 2021). In the present study, the low number of samples of a bad or poor ecological quality represent an imbalance in the total statistical strength of the dataset. However, the sampled locations in this study are representative of the overall

ecological quality status across Denmark, and this inherent imbalance is expected to remain if the sample size is increased while maintaining representativeness. The data analysis was therefore focused on the identification of potential indicator organisms representative of different ecological qualities.

To explore the relationship between ecological status and beta diversity further, a canonical correspondence model (CCA) was generated for the whole biome, as well as the individual domain data (Fig. 2). Beta diversity analysis using CCA is a well-described and widely applied method in ecological studies (ter Braak and Verdonschot, 1995), which makes the chosen approach in the present study directly compatible with existing protocols for data analysis. The CCA model, constrained by ecological status, revealed that the whole biome data (Fig. 2a) achieved the best separation, followed by the *Bacteria* (Fig. 2b), where near complete separation of all five ecological status groups was achieved. Prokaryotic communities associated with sediments and surface waters have previously been shown to be sensitive to environmental changes, and have been suggested as a tool for biomonitoring of pollution (Li et al., 2018; Mlejnková and Sovová, 2010). The bacterial communities of freshwater streams may present a relatively unexplored approach with a high potential for the discovery of new indicators for bioassessment.

A gradient like overlap between streams of bad and poor, and moderate to high ecological status was observed for the eukaryotic data. This is in line with previous studies focusing on metabarcoding of invertebrates (Elbrecht et al., 2017; Kuntke et al., 2020), as well as well-described ecological quality measurement protocols, which are based in the identification and abundance of chosen indicator species (Birk et al., 2012). The observed archaeal community was not able to separate the samples based on ecological quality in a meaningful way. However, this is likely related to the low presence and lack of differentiation to the surrounding environment and/or the coverage of the chosen primer set which might only capture a part of the archaeal taxa. It has previously been shown that archaeal communities in sediments are highly diverse, as well as sensitive to environmental change (Hoshino and Inagaki, 2019), and may be worth investigating in more details in relation to biomonitoring protocols of freshwater systems.

The domain-specific diversity analysis could potentially be extended with network analysis to reveal potential ecologically meaningful relationships within and across domains, which could strengthen the detection of indicator species and organisms associated to individual ecological status classes. A similar approach has previously been applied in paddy soils (Wang et al., 2017). Alternatively, indicator organisms could be extracted from the dataset to simplify the dimensionality of metabarcoding data and provide basis for a model describing the relationship between the biome and ecological status. This approach has previously been applied in rivers in China (Li et al., 2018). Another strategy could be to implement machine learning into the data analysis strategy to improve handling of the complex biological variation and dimensionality of metabarcoding datasets (Cordier et al., 2019). However, this last approach would require significant upscaling of the sample size.

4.3. Potential quality indicators detected in all three domains of life

The potential of microbial community members as indicators of ecosystem quality has previously been shown in contaminated sediments and marine environments (Acosta-González and Marqués, 2016; Urakawa et al., 2012), as well as rivers (Li et al., 2018). The strongest correlation to ecological status among the individual domains was observed in the ZOTUs from the bacterial community. Furthermore, correlations of interest from *Archaea* and *Eukaryotic* ZOTUs were also observed (Figs. 3, S6 and S7), but these were weaker compared to the *Bacteria*. The latter correlations may be affected by their relatively low abundance in the bulk samples, and may be more significant when implemented in a network based approach targeting cross-domain interactions as indicators of quality, due to the complex nature of stream ecosystems (Battin et al., 2016).

Bacterial communities have previously been shown to reflect environmental changes relating to land use (Lear et al., 2013), as well as (anthropogenic) pollution (Acosta-González and Marqués, 2016; Mlejnková and Sovová, 2010; Urakawa et al., 2012). Organisms previously reported in relation to ecosystem health were found among the potential indicator ZOTUs of the prokaryotes. The genus *Flavobacterium* was observed among the ZOTUs associated to lower quality streams, and has previously been linked to anthropogenic activity in surface waters (Acosta-González and Marqués, 2016; Mlejnková and Sovová, 2010). A positive correlation was observed between the ecological status and a representative of the family *Nitrosomonadaceae* and the genus *Nitrospira*. This correlation has previously been characterized in a marine ecosystem in which pollution was shown to induce significant changes in the nitrifying communities (Urakawa et al., 2012). Similar observations have described for increased abundances of the family *Rhodobacteraceae* exposed to oil pollution in sediments (Acosta-González and Marqués, 2016), however the indicator analysis in the present study yielded multiple ZOTUs representing this family, but with opposing tendencies. This highlights the importance of taxonomic resolution, as a family of bacteria can contain numerous of species with a wide range of functions and sensitivity to their environment.

Analysis of the microbial communities, and its vast diversity, present in bulk samples of invertebrates from streams could provide an interesting supplement to existing metabarcoding approaches targeting biomarkers for eukaryotes such as the *COI* gene (Kuntke et al., 2020). Now, when applying a molecular approach for water-quality assessment we are no longer limited to the composition of visible BQE, such as macroinvertebrates. The microbial community composition in the streams may often capture a better and more comprehensive and sensitive snapshot of the diversity in stream waters. The value of the microbial community in freshwater streams has also previously been recognised as a potential major improvement to environmental assessment protocols (Sagova-Mareckova et al., 2021).

4.4. Complete biome capture as an alternative method for stream-water quality studies

The metabarcoding approach applied in the present study is a potential cost effective and fast alternative to existing metagenomics-based analyses for biomonitoring as it targets specific genes selected to represent a given taxonomic affiliation, thereby capturing all three domains of life. The use of domain specific indicator genes facilitates detailed analysis covering the majority of ecosystem diversity, thereby exploiting the opportunity to capture previously undescribed taxonomic diversity. While the databases for ribosomal target gene sequences from *Bacteria* and *Archaea* are good, they remain relatively poor for eukaryotic organisms, especially within the invertebrates. Attempts to fill the gaps in reference databases are ongoing as part of national and international Barcode of Life projects (e.g. NorBOL, GBOL, and SwissBOL) (Jinbo et al., 2011). However, biome profiling for fauna or flora indexing does not necessarily have to provide full taxonomic descriptions and affiliations. The results obtained in this study clearly show that members of all three domains reflect the ecological status of the sampled location. Furthermore, the results provide evidence for the presence of potential indicator organisms within all three domains of life, which may provide a potential for alternative water quality assessment.

In conclusion, simultaneous analysis of *Bacteria*, *Archaea* and *Eukaryota* using high-throughput amplicon sequencing was shown to be a promising method for convenient and fast assessment of stream-water quality. Domain-specific biome profiles separated the samples from 50 water bodies with different geological and physiochemical environments into their respective pre-determined ecological status classes. The obtained data showed that it was possible to extract potential new indicator organisms from all three domains. The strongest correlation

between ecological status and community composition was observed for the *Bacteria*. Application of metabarcoding approaches targeting diverse different taxonomic groups therefore has the potential to assist in the development of new and improved ecological quality models with reduced biases, increased sensitivity, and improved accessibility in terms of cost and expertise required.

CRediT authorship contribution statement

Nadieh de Jonge: Formal analysis, Validation, Data curation, Visualization, Writing - original draft. **Franziska Kuntke:** Investigation, Methodology. **Martin Hesselsøe:** Writing - review & editing, Funding acquisition. **Jeppe Lund Nielsen:** Conceptualization, Methodology, Funding acquisition, Supervision, Validation, Writing - review & editing.

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

This study was financed by NIRAS A/S and Amphi Consult ApS with support from Innovation Fund Denmark (grant number 1355-00144B and 9066-00046B). The authors wish to thank biologist Per Gørtz (Laboratory of Fish Ecology; NIRAS A/S, Allerød, Denmark) for his help with sampling and analysing the water stream samples with the DSFI protocol.

Data accessibility

All raw sequencing data has been made available at the European Nucleotide Archive (ENA) under the project accession number PRJEB37542.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.108059.

References

Acosta-González, A., Marqués, S., 2016. Bacterial diversity in oil-polluted marine coastal sediments. Curr. Opin. Biotechnol. 38, 24–32. https://doi.org/10.1016/j.copbio.2015.12.010.

Agouridis, C.T., Wesley, E.T., Sanderson, T.M., Newton, B.L., 2015. Aquatic macroinvertebrates: biological indicators of stream. Health 5, 1–5.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Andersen, K.S., Kirkegaard, R.H., Karst, S.M., Albertsen, M., 2018. Ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. bioRxiv 299537. https://doi.org/10.1101/299537.

Baattrup-Pedersen, A., Friberg, N., Pedersen, M.L., Skriver, J., Kronvang, B., Larsen, S.E., 2004. Anvendelse af Vandrammedirektivet i danske vandløb. Danmarks Miljøundersøgelser, Faglig rapport fra DMU.

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 (4), 251–263. https://doi.org/10.1038/nrmicro.2016.15.

Bengtsson, L., Fortin, V., Gronewold, A.D., Tanino, Y., Surridge, B., Watson, N., Herschy, R.W., Herschy, R.W., Herschy, R.W., Gergov, G., Herschy, R.W., Woo, M., Young, K., Lundin, L., Cavaleri, L., Herschy, R.W., Herschy, R.W., 2012. Water Framework Directive. Encycl. Lakes Reserv. 872–876 https://doi.org/10.1007/978-1-4020-4410-6_217.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009.

Blackman, R., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A.F., Jones, J., Macher, J., Majaneva, M., Martins, F., Múrria, C., Meissner, K., Pawlowski, J., Schmidt Yáñez, P., Zizka, V., Leese, F., Price, B., Deiner, K., 2019. Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. Metabarcoding Metagenomics 3, 49–57. https://doi.org/10.3897/mbmg.3.34735.

ter Braak, C.J.F., Verdonschot, P.F.M., 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquat. Sci. 57 (3), 255–289. https://doi.org/10.1007/BF00877430.

Bradey, D.C., Ormerod, S.J., 2002. Evaluating the precision of kick-sampling in upland streams for assessments of long-term change: the effects of sampling effort, habitat and rarity. Arch. für Hydrobiol. 155, 199–221. https://doi.org/10.1127/archiv-hydrobiol/155/2002/199.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7 (5), 335–336. https://doi.org/10.1038/nmeth.f.303.

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., Lanzén, A., 2021. Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Mol. Ecol. 30 (13), 2937–2958. https://doi.org/10.1111/mec.v30.1310.1111/mec.15472.

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 (5), 387–397. https://doi.org/10.1016/j.tim.2018.10.012.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C., Bik, H.M., Freckleton, R., 2016. The ecologist's field guide to sequence-based identification of biodiversity. Methods Ecol. Evol. 7 (9), 1008–1018.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10 (10), 996–998. https://doi.org/10.1038/nmeth.2604.

Elbrecht, V., Leese, F., 2017. Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Front. Environ. Sci. 5, 1–17. https://doi.org/10.3389/fenvs.2017.00011.

Elbrecht, V., Vamos, E.E.E., Meissner, K., Aroviita, J., Leese, F., 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods Ecol. Evol. 8, 1265–1275. https://doi.org/10.7287/PEERJ.PREPRINTS.2759V2.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. 4 (5), 642–647. https://doi.org/10.1038/ismej.2009.153.

Francy, D.S., Helsel, D.R., Nally, R.A., 2000. Occurrence and distribution of microbiological indicators in groundwater and stream. Water Environ. Res. 72, 152–161. https://doi.org/10.2307/25045352.

Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W., Hajibabaei, M., 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proc. Natl. Acad. Sci. U. S. A. 111 (22), 8007–8012. https://doi.org/10.1073/pnas.1406468111.

Hoshino, T., Inagaki, F., 2019. Abundance and distribution of Archaea in the subseafloor sedimentary biosphere. ISME J. 13 (1), 227–231. https://doi.org/10.1038/s41396-018-0253-3.

Jinbo, U., Kato, T., Ito, M., 2011. Current progress in DNA barcoding and future implications for entomology. Entomol. Sci. 14, 107–124. https://doi.org/10.1111/j.1479-8298.2011.00449.x.

Kuntke, F., de Jonge, N., Hesselsøe, M., Nielsen, J.L., 2020. Stream water quality assessment by metabarcoding of invertebrates. Ecol. Indic. 111, 105982 https://doi.org/10.1016/j.ecolind.2019.105982.

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., Lewis, G., 2013. The biogeography of stream bacteria. Glob. Ecol. Biogeogr. 22 (5), 544–554. https://doi.org/10.1111/geb.12046.

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abarenkov, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., Bones, A., Borja, Á., Bruce, K., Bursić, V., Carlsson, J., Čiampor, F., Čiamporová-Zatovičová, Z., Coissac, E., Costa, F., Costache, M., Creer, S., Csabai, Z., Deiner, K., DelValls, Á., Drakare, S., Duarte, S., Eleršek, T., Fazi, S., Fišer, C., Flot, J.-F., Fonseca, V., Fontaneto, D., Grabowski, M., Graf, W., Guðbrandsson, J., Hellström, M., Hershkovitz, Y., Hollingsworth, P., Japoshvili, B., Jones, J., Kahlert, M., Kalamujic Stroil, B., Kasapidis, P., Kelly, M., Kelly-Quinn, M., Keskin, E., Kõljalg, U., Ljubešić, Z., Maček, I., Mächler, E., Mahon, A., Marečková, M., Mejdandzic, M., Mircheva, G., Montagna, M., Moritz, C., Mulk, V., Naumoski, A., Navodaru, I., Padisák, J., Pálsson, S., Panksep, K., Penev, L., Petrusek, A., Pfannkuchen, M., Primmer, C., Rinkevich, B., Rotter, A., Schmidt-Kloiber, A., Segurado, P., Speksnijder, A., Stoev, P., Strand, M., Šulčius, S., Sundberg, P., Traugott, M., Tsigenopoulos, C., Turon, X., Valentini, A., van der Hoorn, B., Várbíró, G., Vasquez Hadjilyra, M., Viguri, J., Vitonytė, I., Vogler, A., Vrålstad, T., Wägele, W., Wenne, R., Winding, A., Woodward, G., Zegura, B., Zimmermann, J., 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Res. Ideas Outcomes 2, e11321. https://doi.org/10.3897/rio.2.e11321.

Leff, J.W., Bardgett, R.D., Wilkinson, A., Jackson, B.G., Pritchard, W.J., De Long, J.R., Oakley, S., Mason, K.E., Ostle, N.J., Johnson, D., Baggs, E.M., Fierer, N., 2018. Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits. ISME J. 12 (7), 1794–1805. https://doi.org/10.1038/s41396-018-0089-x.

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., Zhang, X., 2018. Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers. Environ. Sci. Technol. 52, 11708–11719. https://doi.org/10.1021/acs.est.8b03869.

Mlejnková, H., Sovová, K., 2010. Impact of pollution and seasonal changes on microbial community structure in surface water. Water Sci. Technol. 61, 2787–2795. https://doi.org/10.2166/wst.2010.080.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Wagner, H., 2016. vegan: Community Ecology Package.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., Barquín, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637–638, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

R Development Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sagova-Mareckova, M., Boenigk, J., Bouchez, A., Cermakova, K., Chonova, T., Cordier, T., Eisendle, U., Elersek, T., Fazi, S., Fleituch, T., Frühe, L., Gajdosova, M., Graupner, N., Haegerbaeumer, A., Kelly, A.-M., Kopecky, J., Leese, F., Nõges, P., Orlic, S., Panksep, K., Pawlowski, J., Petrusek, A., Piggott, J.J., Rusch, J.C., Salis, R., Schenk, J., Simek, K., Stovicek, A., Strand, D.A., Vasquez, M.I., Vrålstad, T., Zlatkovic, S., Zupancic, M., Stoeck, T., 2021. Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. Water Res. 191, 116767. https://doi.org/10.1016/j.watres.2020.116767.

Skriver, J., 1999. Danish Stream Fauna Index (DSFI) as an indicator of rare and threatened benthic macroinvertebrates. In: Friberg, N., Carl, J.D. (Eds.), Biodiversity in Benthic Ecology. Proceedings from Nordic Benthological Meeting in Silkeborg. National Environmental Research Institute, Denmark.

Skriver, J., Friberg, N., Kirkegaard, J., 2000. Biological assessment of running waters in Denmark: introduction of the Danish Stream Fauna Index (DSFI). Verh. Internat. Verein. Limnol. 27 (4), 1822–1830.

Urakawa, H., Garcia, J.C., Barreto, P.D., Molina, G.A., Barreto, J.C., 2012. A sensitive crude oil bioassay indicates that oil spills potentially induce a change of major nitrifying prokaryotes from the Archaea to the Bacteria. Environ. Pollut. 164, 42–45. https://doi.org/10.1016/j.envpol.2012.01.009.

Wang, H., Wei, Z., Mei, L., Gu, J., Yin, S., Faust, K., Raes, J., Deng, Y., Wang, Y., Shen, Q., Yin, S., 2017. Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. Soil Biol. Biochem. 105, 227–235. https://doi.org/10.1016/j.soilbio.2016.11.029.

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

Wurzbacher, C., Attermeyer, K., Kettner, M.T., Flintrop, C., Warthmann, N., Hilt, S., Grossart, H.-P., Monaghan, M.T., 2017a. DNA metabarcoding of unfractionated water samples relates phyto-, zoo- and bacterioplankton dynamics and reveals a single-taxon bacterial bloom. Environ. Microbiol. Rep. 9, 383–388. https://doi.org/10.1111/1758-2229.12540.

Wurzbacher, C., Fuchs, A., Attermeyer, K., Frindte, K., Grossart, H.-P., Hupfer, M., Casper, P., Monaghan, M.T., 2017b. Shifts among Eukaryota, Bacteria, and Archaea define the vertical organization of a lake sediment. Microbiome 5, 41. https://doi.org/10.1186/s40168-017-0255-9.