Received: 2 September 2022 | Accepted: 8 March 2023

DOI: 10.1002/edn3.409

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

SPECIAL ISSUE ARTICLE

Maximizing sampling efficiency to detect differences in fish community composition using environmental DNA metabarcoding in subarctic fjords

Gledis Guri[1,2] | Jon-Ivar Westgaard[1] | Nigel Yoccoz[3] | Owen S. Wangensteen[2,4] | Kim Præbel[2] | Jessica Louise Ray[5] | Ryan P. Kelly[6] | Andrew Olaf Shelton[7] | Tanja Hanebrekke[1] | Torild Johansen[1]

[1]Norwegian Institute of Marine Research, Framsenteret, Tromsø, Norway

[2]Norwegian College of Fishery Science, UiT The Arctic University of Norway, Tromsø, Norway

[3]Dept. of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway

[4]Dept. of Evolutionary Biology, Ecology & Environmental Sciences, Universitat de Barcelona, Barcelona, Spain

[5]NORCE Norwegian Research Centre AS, Climate & Environment Division, NO-5838, Bergen, Norway

[6]School of Marine and Environmental Affairs, University of Washington, Seattle, Washington, USA

[7]Northwest Fisheries Science Center, NOAA Fisheries, Seattle, Washington, USA

Correspondence
Gledis Guri, Norwegian Institute of Marine Research, Framsenteret, Tromsø, Norway.
Email: gledis.guri@hi.no

Funding information
Norges Forskningsråd, Grant/Award Number: NRC301691

Abstract

Environmental DNA (eDNA) has gained popularity as a tool for ecosystem biomonitoring and biodiversity assessment. Although much progress has been made regarding laboratory and fieldwork protocols, the issue of sampling efficiency requires further investigation, particularly in three-dimensional marine systems. This study focuses on fish community composition in marine ecosystems and aims to analyze the efficiency of sampling design given the sampling effort for distinguishing between different communities. We sampled three fjords in Northern Norway, taking samples along fjord transects and at three different depths, and amplified a fragment of the mitochondrial 12S rRNA gene of bony fishes using the MiFish primers. We evaluated the effect of (i) the number of sampling stations, (ii) samples' spatial distribution, and (iii) the data treatment approach (presence/absence versus semiquantitative) for maximizing the efficiency of eDNA metabarcoding sampling when inferring differences of fish community compositions between fjords. We found that the manner of data treatment strongly affected the minimum number of sampling stations required to detect differences among communities; because the semiquantitative approach retained some information about abundance of the underlying reads, it was the most efficient. Furthermore, we found little-to-no difference of fish communities in samples from intermediate depths when comparing vertical fish communities. Lastly, we found that the differences between fish communities at the surface were the highest across the horizontal distance and overall, samples ~30 km apart showed the highest variation in the horizontal distribution. Boosting sampling efficiency (reducing sampling effort without compromising ecological inferences) can significantly contribute to enhanced biodiversity management and efficient biomonitoring plans.

KEYWORDS
eDNA, fish, heterogeneity, metabarcoding, sampling design, spatial distribution

1 | INTRODUCTION

Biological monitoring, as a key component of the assessment of ecological resources (Gold et al., 2021), allows decision makers to adjust policies and management plans to achieve environmental targets and sustainable use of marine resources. Many studies have attested to the importance of biological monitoring for a robust decision-making (Borja & Elliott, 2013; Bourlat et al., 2013; Salter et al., 2019; Stat et al., 2017). Moreover, recent changes toward ecosystem-based fisheries management rely heavily on biodiversity assessments and multispecies identification where quantification of community metrics is crucial for reliable monitoring (Gullestad et al., 2017).

Multiple studies have encouraged the use of environmental DNA (eDNA) in biomonitoring for management policies as the DNA expelled from organisms can be captured and sequenced (Taberlet et al., 2012), enabling the detection of the community composition present at the study site (Ji et al., 2013). As a noninvasive sampling technique (Cilleros et al., 2019), the reliability of eDNA metabarcoding results is an attractive alternative to traditional capture sampling for biomonitoring (Barnes & Turner, 2016; Cantera et al., 2019; Gilbey et al., 2021). However, metabarcoding approaches for biomonitoring have only recently been developed and face several obstacles (e.g., amplification bias, DNA transport) that hinder the direct use of metabarcoding data for inference of ecological communities (Banerjee et al., 2021). An essential component of any eDNA metabarcoding study is to provide an adequate sampling design to address specific ecological questions. Moreover, minimizing the DNA sampling effort can enable stakeholders and policymakers to increase sampling events and thus broaden the monitoring scale of marine biodiversity assessments.

Optimizing biological and technical replication as a trade-off between decreasing false negatives and costs has been investigated (Buxton et al., 2021; Cantera et al., 2019). Furthermore, optimization of eDNA sampling design has been laid on water volume filtered, filter types, and effort distribution among methods for maximizing univariate measures of diversity (such as α diversity) and single/rare species detection while decreasing sampling effort (Andres et al., 2022; Bessey et al., 2020; Sanches & Schreier, 2020; Wood et al., 2021). One additional study focused on the spatial distribution of eDNA samples in lentic systems to evaluate sampling design strategies for inferring patterns of fish biodiversity (α and β diversity; Zhang et al., 2020), where they suggested a systematic spatial sampling protocol and concluded that eDNA samples autocorrelated up to 2 km on the horizontal distance. However, the latter study circumvented issues related to sampling effort; thus, at present, no research has been conducted on sampling design accounting for community compositions and heterogeneity as a function of the sampling effort. Additionally, in contrast to lentic and lotic systems, marine systems are more complex in DNA transport due to multidirectional and three-dimensional water flow (Hansen et al., 2018).

Obtaining an efficient sampling design for eDNA-based studies, especially in marine systems, is a challenging task (convoluted question due to the multitude of research questions, hence every context has a specific sampling design; Yoccoz et al., 2001) that still requires attention. In this study, we focus maximizing the sampling efficiency explicitly on detecting ecosystems' fish community composition (hereafter communities); thus, circumventing univariate diversity metrics and single/rare species detection as such questions have been resolved. We study this problem in two different levels, that is, by investigating how well the sampled communities segregate given the unit of sampling effort, thus how much can we reduce the sampling effort to still obtain significant differences between these fjord communities, and second, which samples contribute the least to the dissimilarity metrics, thus can potentially be removed without affecting the variability of the ecosystem.

To achieve these objectives, we formulated the following questions: (i) What is the metabarcoding data treatment approach, presence/absence versus semiquantitative, that most efficiently represents the differences between fjords ecosystems, and what is the minimal number of sampling sites for each approach; and (ii) where to effectively collect the samples (in three-dimensional space) to maximize our efficiency to detect differences in ecological communities while minimizing sampling effort.

2 | METHODS

2.1 | Study sites

Study samples were collected on research cruises on March 2021 by R/V Kristine Bonnevie. Three Northern Norwegian fjords: Balsfjord, Frakkfjord, and Olderfjord, were selected based on their size, bathymetry, and distance from the open ocean. Balsfjord has a length of 40 km and a maximum depth of 190 m in the middle of the fjord. There is a sill at the fjord's entrance, and the archipelago limits the water exchange and categorizes it as protected from the open ocean. Frakkfjord and Olderfjord are relatively similar regarding their size and shape, with a fjord length of 8 and 4 km and a maximum depth is 86 and 74 m, respectively. These two fjords are more exposed to the open ocean compared with Balsfjord; however, Frakkfjord has the highest degree of exposure to open water. Nearly all high latitude Norwegian fjords are ice-free and depicted with an Arctic light regime (Reigstad & Wassmann, 1996). Although a subarctic water climate characterizes all fjords in the area, the degree of openness to the open sea can create differences regarding community composition. In total 17 eDNA sampling stations were deployed for the three fjords, of which 10 were located in Balsfjord, four in Frakkfjord and three in Olderfjord (Figure 1). GPS coordinates and other metadata of sampling stations are provided in Table S1.

2.2 | Water sampling, eDNA filtration, and extraction

To investigate the spatial heterogeneity of the eDNA signals for inference of fish communities, we established three sampling transects

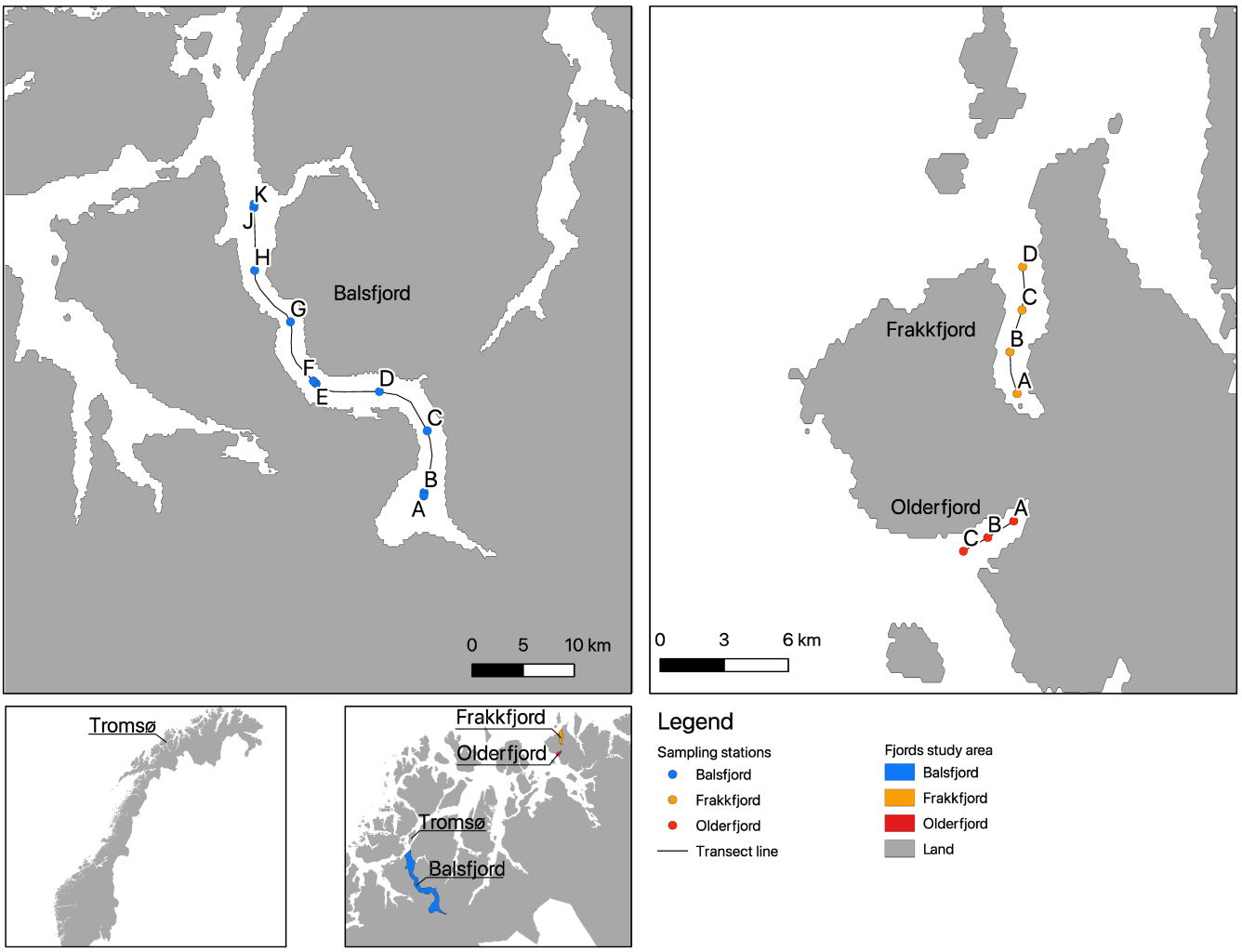

FIGURE 1 Map of the eDNA sampling station distributed among three localities (Balsfjord—blue, Frakkfjord—orange, and Olderfjord—red) and the radius of the area represented by the sampling station.

in three localities (i.e., three fjord ecosystems) where samples were assigned using fixed-distance nonoverlapping radii scaled by the size of the locality (i.e., 3000 m in Balsfjord, 1000 m in Frakkfjord, and 750 m in Olderfjord). In alphabetical order, sampling stations were designated by letters starting from the innermost station (Figure 1). To constrain the background level of spatial heterogeneity from metabarcoding, we double-sampled three locations in Balsfjord using "twin" sampling stations next to each other (i.e., A/B, E/F, and J/K). At each sampling station, 12 5-L Niskin bottles mounted on a stainless-steel frame with CTD were deployed to desired sampling depths for water collection. The three depths investigated were surface (10 m), pycnocline (depth of highest density, ~50 m), and bottom (10 m above bottom). At each sampling depth, triplicate 2-L water samples, derived from three distinct Niskin bottles, were filtered onboard the research vessel through 0.22 µm Sterivex filters (MerckMillipore) using a peristaltic pump (multichannel flow Heidolph™ Hei-Flow Advantage 01). After removing the remaining water drops by pumping air, filters were transferred to sterile 50-mL Falcon centrifuge tubes and immediately stored at −20°C until they were transported to the laboratory for −80°C storage until DNA extraction. DNA was extracted using DNeasy PowerWater Sterivex Kit (Qiagen GmbH) following the manufacturer's protocol with slight modifications (all steps involving PowerBead Tubes were omitted). DNA extraction from water samples was conducted randomly to minimize the possible biases from contamination during laboratory workflow.

2.3 | Contamination control

To control for contamination occurring at each step of the workflow, we included field-negative controls (one air and one water blank per sampling station), laboratory-negative controls (one extraction blank per block of 12 extracted samples; three PCR blanks per block of 96 extracted samples), and PCR-positive controls (one positive control PCR sample per block of 96 extracted samples) as described by Shu et al. (2020). To reduce the risk of cross-contamination during the sampling event, all sampling equipment was decontaminated with 20% (v/v) sodium hypochlorite solution (household bleach) and then rinsed with Milli-Q water onboard before and after fieldwork. Sterile nitrile gloves were used when in contact with water samples.

To minimize the risk of sample cross-contamination in the laboratory, before and after each round of DNA extraction, all the workbenches (extraction and PCR hood) were decontaminated through UV exposure for 30 min, as suggested by Goldberg et al. (2016). Additionally, the air in the laboratory was decontaminated regularly (once a week) with hydrogen peroxide. Lastly, pre- and post-PCR tasks were performed in different laboratories to reduce the risk of cross-contamination.

2.4 | Library preparation and sequencing

A total of 192 samples (divided into two libraries), including eDNA samples ($n = 150$), PCR blanks ($n = 6$), positive controls ($n = 2$), extraction blanks ($n = 6$), fieldwork water and air blank ($n = 16$ and 12 respectively), were amplified using the MiFish-U universal primer set (Forward: 5′-GTCGGTAAAACTCGTGCCAGC-3′; Reverse: 5′-CATAGTGGGGTATCTAATCCCAGTTTG-3′; Miya et al., 2015) targeting the mitochondrial 12S rRNA gene region (169–172 bp fragment). Fusion primers containing adaptor, index, and primer-specific sequences were used to allow one-step PCR amplification, where each 20 μL PCR reaction consisted of 3 μL eDNA template, 1 μL of primer mix (0.5 μL each of 5 μM stock solutions of MiFish-U-forward and MiFish-U-reverse primers), 10 μL of QIAGEN Multiplex PCR Master Mix, 0.16 μL of BSA (concentration 20 μg/mL), and lastly 5.84 μL of RNase-free water. All samples were amplified in triplicate. The thermocycler program included an initial denaturing step of 95°C for 10 min, 40 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s and a final extension step of 72°C for 5 min.

PCR amplification success and product size were assessed using capillary electrophoresis (QIAxcel; Qiagen GmbH). Products from triplicate PCR reactions were pooled into a single tube before 100 μL from each PCR product pool were combined to generate two sequencing libraries, each consisting of 96 samples/library. Sequencing libraries were electrophoresed on a 2% (w/v) agarose gel in 1× TAE buffer and stained with SYBR safe (Qiagen GmbH). Bands of the expected size (300 bp) were excised and purified using GeneJet Gel Extraction and DNA Cleanup Micro Kit (Thermo Fisher Scientific). DNA concentration of the purified sequencing libraries was measured using Qubit dsDNA HS kit (Thermo Fisher Scientific), diluted to a final concentration of 50 pM and spiked with 4 μL of Ion S5 Calibration Standard prior to loading onto the Ion Chef instrument (Thermo Fisher Scientific). Libraries were sequenced using the Ion 530-sequencing chip with the 200 bp protocol (Thermo Fisher Scientific).

2.5 | Bioinformatics

Sequences were automatically demultiplexed and quality filtered after the sequencing process using Torrent Suite™ inbuilt in the sequencer following their inbuilt standard settings. The sampled sequence dataset was thereafter filtered for chimeric sequences using a *uchime-denovo* algorithm in VSEARCH (Rognes et al., 2016). Second, we clustered the sequences into Molecular Operational Taxonomic Units (MOTU) using SWARM v2 (Mahé et al., 2014) with a distance of $d = 3$. Third, singletons were removed, and the remaining MOTUs were taxonomically annotated using the *ecotag* algorithm (Boyer et al., 2016) and a well-curated local database assembled from data available from EMBL and NCBI (access date: 14 March 2022).

To select only high-quality sequences and remove pseudogenes and artifacts, we arbitrarily customized a low-quality dataset filtering process where MOTUs with an identity match ≥97% (Li et al., 2018; Nakagawa et al., 2018; Sales et al., 2021) or higher number of reads than 1/10,000 of the total library reads were retained (Coguiec et al., 2021). MOTUs unassigned to species taxonomic rank after *ecotag* algorithm were manually checked by blasting in NCBI online database (date: March 14, 2022) where the same selection criteria were used. For downstream biodiversity analysis, we removed biological replicates with sequencing depth lower than 500 reads as an indication of low-quality samples. Amplicon sequences whose occurrence in negative controls exceeded 10% of their total read abundance in the entire dataset were removed from the dataset as an indication of contamination. Additionally, we summed (pooled) all MOTU reads across the biological replicates of the same depth taken in the same sampling station. Lastly, we removed all taxa that were not assigned to fish (class: Actinopterygii or Chondrichthyes).

2.6 | Statistical analysis

All analyses and plots were performed in R (R Core Team, 2022). To see that our sequencing depth was not a limiting factor for representing the taxonomic diversity of the study area, we examined the DNA sequencing effort through species accumulation curves (hereafter sequencing effort curves), using *rarecurve* function in the vegan package. The curves plotted the average number of species identified as a function of sequencing depth, indicating if the latter was sufficient for covering the biodiversity. Additionally, species accumulation curves (hereafter DNA collection curves), the number of species against the number of samples, were plotted to examine whether more samples would yield more species, thus determining whether our DNA collection effort was sufficient for covering the taxonomic diversity of the sampled area. The DNA collection curves were drawn using *specaccum* function in *vegan* package.

2.7 | Efficiency of data treatment approaches and minimal number of sampling stations for detecting community differences

To explore the strength of two commonly used metabarcoding data treatment approaches on the ability to detect differences between

fish communities, we selected and compared qualitative (presence/absence) and semiquantitative approaches. We outline the qualitative approach as the binary transformation of MOTU reads (after quality filtering) into the presence and absence where MOTUs with reads ≥1 were defined as a presence. For the semiquantitative approach, we selected the eDNA index proportion model developed by Kelly et al. (2019), also called inverse "Wisconsin double-standardization," a simplified way to account for species-specific differences in amplification efficiency. The following analyses were performed on both qualitative and semiquantitative treatment of the eDNA reads of pooled replicates. We used the Jaccard and Bray–Curtis distance matrices, respectively. The dissimilarity matrices between ecosystems were initially represented visually through nonmetric multidimensional scaling (nMDS) ordination using *metaMDS* in *vegan* package (Oksanen et al., 2022), with two dimensions ($k = 2$) while using 20 random starts in search of a stable solution. Differences in communities between ecosystems were computed through permutational analysis of variances (PERMANOVA) using *adonis* function in *vegan* package (Oksanen et al., 2022) with factor location (Balsfjord, Frakkfjord, Olderfjord, and positive control) and 999 permutations. For significant PERMANOVA results, we conducted permutational multivariate dispersion test (PERMDISP) to check whether the significance was due to the differences in centroids or due to the differences in dispersion of samples in principal coordinate space of dissimilarity (PCoA). Additionally, we used permutational pair-wise comparisons with the Benjamini–Yekutieli FDR correction—False Discovery Rate (Benjamini & Yekutieli, 2001) to indicate the paired location comparisons that were significantly different. We used *betadisper* and *pairwise.adonis2* functions for both tests, respectively, supplied by *vegan* package (Oksanen et al., 2022). To indicate the species driving the differences between localities, we used indicator species analysis (Dufrêne & Legendre, 1997) on both approaches using *indval* function in *labdsv* package (Roberts, 2019) with 999 permutations.

Furthermore, to investigate the efficiency of each data treatment approach (qualitative and semiquantitative), we conducted a linear regression of dissimilarity (measured in pseudo-*F*-value) as a function of sampling effort (number of sampling stations) for each approach using *lm* function in R with the formula dissimilarity ~ sampling effort × method. We measured the dissimilarity as the pseudo-*F* and *p* values of a series of PERMANOVA tests for detecting ecosystem differences (factor locality) across a range of sampling effort. The range of sampling effort was conducted through progressively removing one sampling station from each locality (excluding positive controls) until one sampling station remained in each locality. During each removal of stations, the dataset was randomly subsampled 999 times and we estimated the mean of *p* and pseudo-*F*-values from PERMANOVA tests for each combination. Lastly, the minimal number of sampling stations required to infer differences between ecosystems was concluded as the mean plus two standard deviation of PERMANOVA *p*-values lower than 0.05. Subsequently, the data treatment approach with the highest efficiency (highest pseudo-*F* value) was selected for all analyses below.

2.8 | Sampling efficiency for maximum vertical and horizontal community difference

To investigate the spatial relationship of eDNA samples and maximum dissimilarity (sampling efficiency), we selected only samples taken in Balsfjord due to the linear shape of the fjord and the extensive number of sampling stations investigated ($n = 10$).

For exploring the sampling efficiency among depth layers, we explored whether the community composition differed significantly between each depth category using PERMANOVA analysis with Bray–Curtis dissimilarity index with factor depth category. Thereafter a post hoc test was conducted for identifying pairs of samples that differed using permutational pair-wise comparisons with the Benjamini–Yekutieli FDR correction. Subsequently, for significant PERMANOVA results, we conducted PERMDISP test. Samples of depth categories that did not differ significantly were defined uninformative, thus can be removed for reducing sampling effort.

Additionally, we explored the dissimilarity of the community among the horizontal distance of the fjord. We fitted a beta regression using *betareg* function in *betareg* package with the pairwise Bray–Curtis dissimilarity index (measured through *vegdist* function in *vegan* package) as response variable, and the log-transformed pairwise horizontal distance (measured in QGIS) and pairwise depth categories as predictor using the formula dissimilarity ~ $\log_{10}$(distance + 1) × depth category. Note that only pairwise comparisons of the similar depth categories were included (i.e., bottom–bottom, pycnocline–pycnocline, and surface–surface). We concluded that the most efficient horizontal distance for sampling was the minimum distance with the maximized community dissimilarity among samples.

3 | RESULTS

3.1 | Sequencing and eDNA sampling effort

After the standard Ion GeneStudio quality and sequence-length filtering, the run yielded 13,579,483 sequence reads from 150 eDNA samples (4,984,499 and 8,594,984, respectively, for each of the two sample pools), two positive controls, and 40 negative controls (Figure S1a). Filtering of chimeric reads and singletons resulted in the removal of 64,581 reads. After our customized low-quality removal and blank treatment subtraction, the final dataset (152 samples) used for analysis contained 13,398,370 reads. We observed very few reads in all negative controls (mean = 32.2 ± 29.3; Figure S1b), indicating a low level of contamination. As a result, no "contaminating" MOTUs were removed from the dataset. From the quality-filtered and nonchimeric reads, we detected 33 marine taxa, of which 31 were assigned to the species level, and the remaining two (*Sebastes* spp. and *Anarhichas* spp.) were assigned to the genus level. Around 99.99% of MOTU reads belonged to nine different orders of bony

fishes (Actinopterygii), while the remaining reads were assigned to *Amblyraja radiata* (Chondrichthyes).

Sequencing effort curves revealed that samples in Frakkfjord and Olderfjord had higher sequencing depth than the Balsfjord samples (Figure 2a). Overall, most curves from samples with high sequencing depth saturated, while the remaining samples achieved a slope change but did not reach a plateau. However, visual inspection indicated that over ca. 500,000 reads per sampling station (samples pooled) would be needed to approximate the full diversity of the samples. This indicated that our sequencing effort (>1,000,000 reads per sampling station) in Frakkfjord and Olderfjord successfully covered the taxonomic complexity, while in Balsfjord (~300,000 reads per sampling station), such coverage was not achieved.

DNA collection curves (Figure 2b) signaled that samples in Frakkfjord and Olderfjord detected most species present as the curve saturated around two sampling stations, stating that the taxonomic diversity of the fjord was successfully covered from the samples deployed. In contrast, the Balsfjord curve did not plateau, indicating that deploying more samples could potentially detect more species in the fjord.

3.2 | Efficiency of data treatment approaches and minimal number of sampling stations for detecting community differences

Two nonmetric multidimensional scaling ordination (nMDS) plots displayed dissimilarities of species composition among the three localities (biological replicates pooled), based on Jaccard (Figure 3a) and Bray–Curtis (Figure 3b) dissimilarity index for qualitative (presence/absence) and semiquantitative (eDNA index) representation of the metabarcoding data, respectively. The qualitative-based nMDS indicated separation of the centroids of the samples based on their locality, with a strong fish community variation and overlap among localities. Furthermore, positive control samples were embedded within 70% confidence interval ellipsoids of all locality centroids. Conversely, the semiquantitative-based nMDS revealed a strong separation of clusters by locality, stating that the semiquantitative community composition in each locality is distinct. Additionally, positive control samples are confidently separated from all localities' community composition.

Both PERMANOVAs (Table 1) indicated significant differences in the composition between localities ($p = 0.005$ and $p < 0.001$). Pseudo-F and R^2 values differed considerably between the tests (Table 1). Semiquantitative-based pair-wise tests indicated significant differences between each pair-wise comparison of locality (including the positive control samples; Table S2b). Conversely, qualitative tests failed to identify significant differences between Frakkfjord and Olderfjord (Table S2a). Additionally, all pair-wise differences between positive control and each locality were found to be nonsignificant (Table S2a,b). Semiquantitative-based PERMDISP was found significant ($p < 0.001$; Table 1), indicating a difference in dispersion of samples (average distance to its centroid). Pair-wise comparison of dispersion revealed a significant difference between Balsfjord and Frakkfjord and all pair-wise comparisons of positive control (Table S2b). Conversely, qualitative-based PERMDISP indicated a homogeneous dispersion among all localities (Table 1).

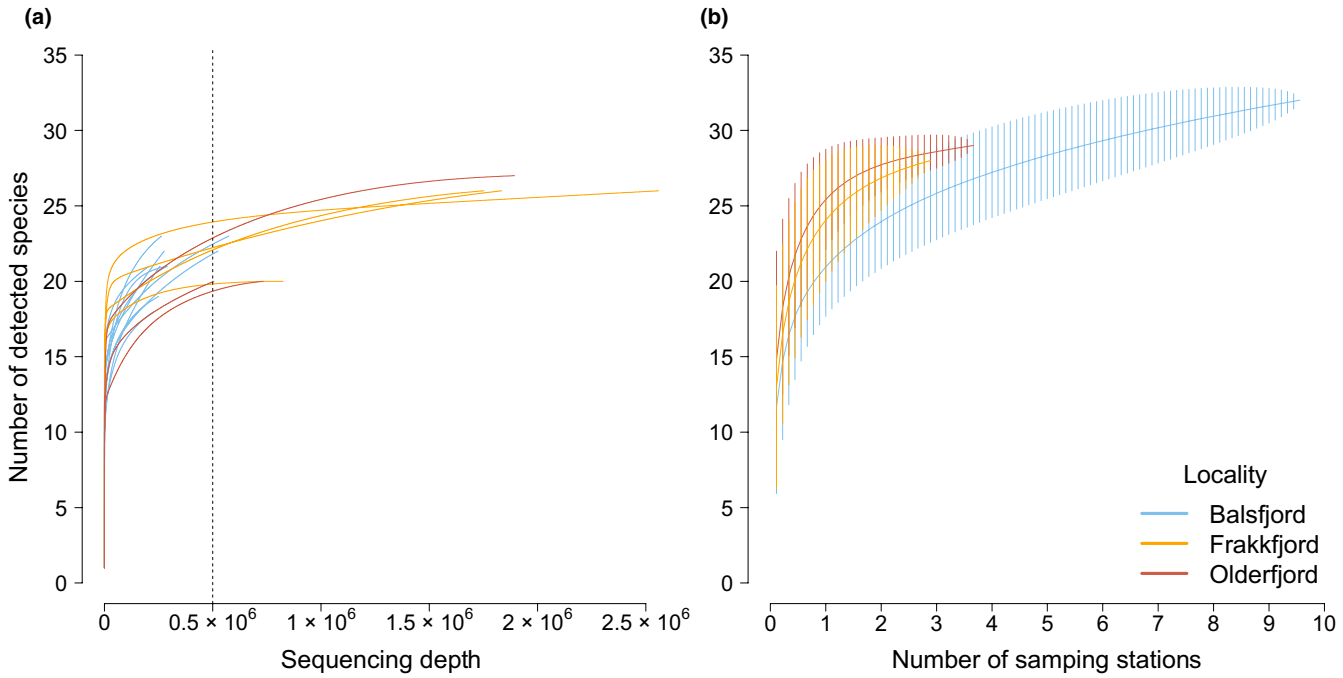

FIGURE 2 Species accumulation curves as a function of sequencing depth, thus sequencing effort curves (a; samples and their biological replicates pooled together) and as a function of number of the sampling stations, thus DNA collection effort curve (b; biological replicates not pooled) colored for each locality. The vertical dashed line indicates the sequencing depth (500,000 reads) where sample curves start to plateau.

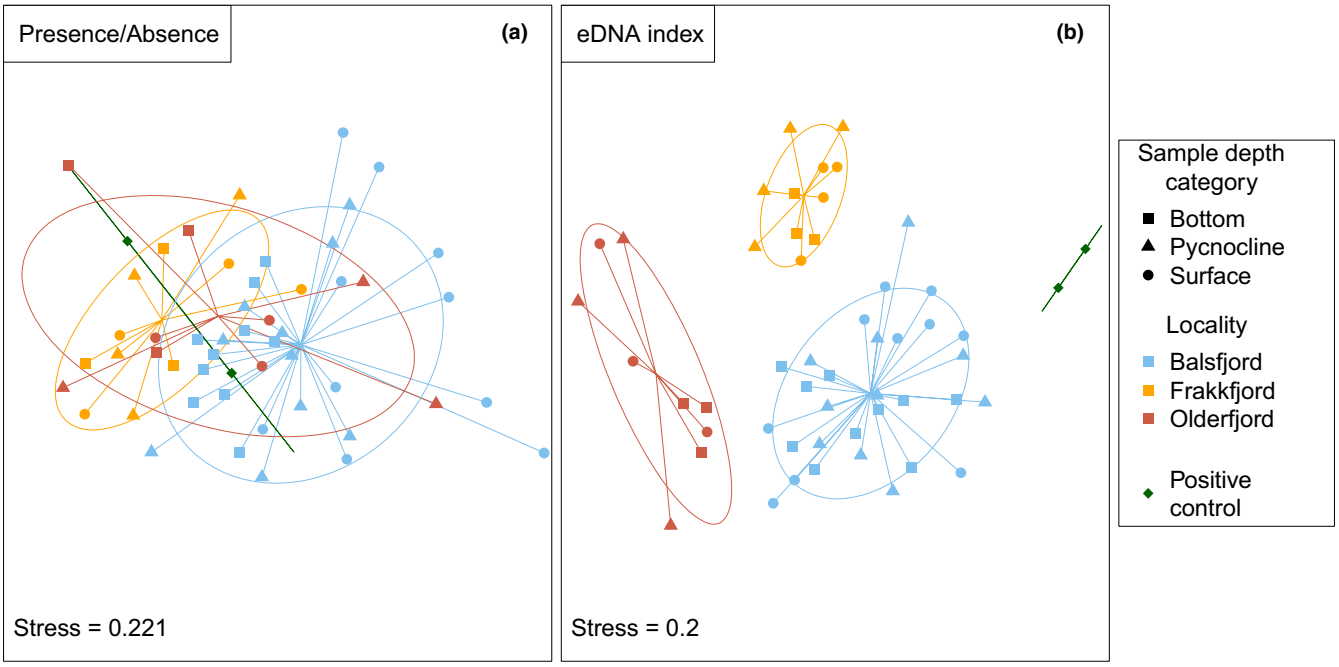

FIGURE 3 Nonmetric multidimensional scaling (nMDS) ordination of samples (biological replicates pooled) obtained using qualitative approach using Jaccard dissimilarity index based on presence/absence of species (a) and semiquantitative using Bray–Curtis dissimilarity index based on inverse Wisconsin double-standardization of amplicon reads (b). The centroids (differed by colors) for the localities (positive control included as locality) and their ellipses (95% confidence interval of the group centroid dispersion) are indicated. Higher dissimilarity of ecosystems was observed when using semiquantitative approach.

TABLE 1 PERMANOVA tests results using qualitative (Jaccard dissimilarity matrix) and semiquantitative (Bray–Curtis dissimilarity matrix) approaches comparing community composition among the factor locality (positive control included).

Approach	Factor	df	SS	MS	Pseudo-*F*	R^2	*p*-Value	PERMDISP
Qualitative	Locality (fjords)	3	0.423	0.141	2.525	0.1363	0.005	0.618
	Residuals	48	2.683	0.055		0.8637		
	Total	51	3.107			1.0000		
Semi-quantitative	Locality (fjords)	3	5.026	1.675	9.011	0.3603	0.001	0.001
	Residuals	48	8.923	0.185		0.6397		
	Total	51	13.949			1.0000		

Note: PERMDISP probabilities for homogeneity of dispersion are also shown. Significant values are indicated in bold.

Comparing the two dissimilarity matrices (qualitative and semi-quantitative based), the Mantel test revealed a significant correlation ($r=0.2$, $p<0.001$), stating that the two data treatment approaches were weakly positively associated.

The linear regression analysis (Figure 4) indicated a significant positive linear relation of ecosystems' dissimilarity with sampling effort for each data treatment approach ($p<0.001$ for both, $t=4.64$ and $t=7.51$ respectively and $R^2=0.89$; Table S4). Moreover, the analysis indicated a significant different intercept ($p<0.001$; number of stations = 3) for qualitative (pseudo-$F=1.52$) compared with semi-quantitative (pseudo-$F=4.42$; Table S4) and significant slope (0.12 and 0.39, respectively). Reflecting the greater information content in the semiquantitative data transformation, three semiquantitative samples produced a degree of resolution equivalent to 13 presence/absence samples.

Indicator species analysis differed considerably between the two approaches (Figure S2). The semiquantitative approach indicated 13 species as significant drivers of community compositional differences between the localities. Moreover, all three localities contained some species indicators (relative frequency of occurrence and relative average abundance of species was the highest within the indicative locality), whereas Balsfjord and Frakkfjord included five species each, and Olderfjord three species. Conversely, the qualitative approach only found five species as significant drivers of the community differences and no species driver defined for Balsfjord. Further investigation of the commonness of species in each community showed six species as common, three as semicommon and four rare species in the semiquantitative data (Table S3), while the qualitative data included three rare species and two semicommon species (Table S3).

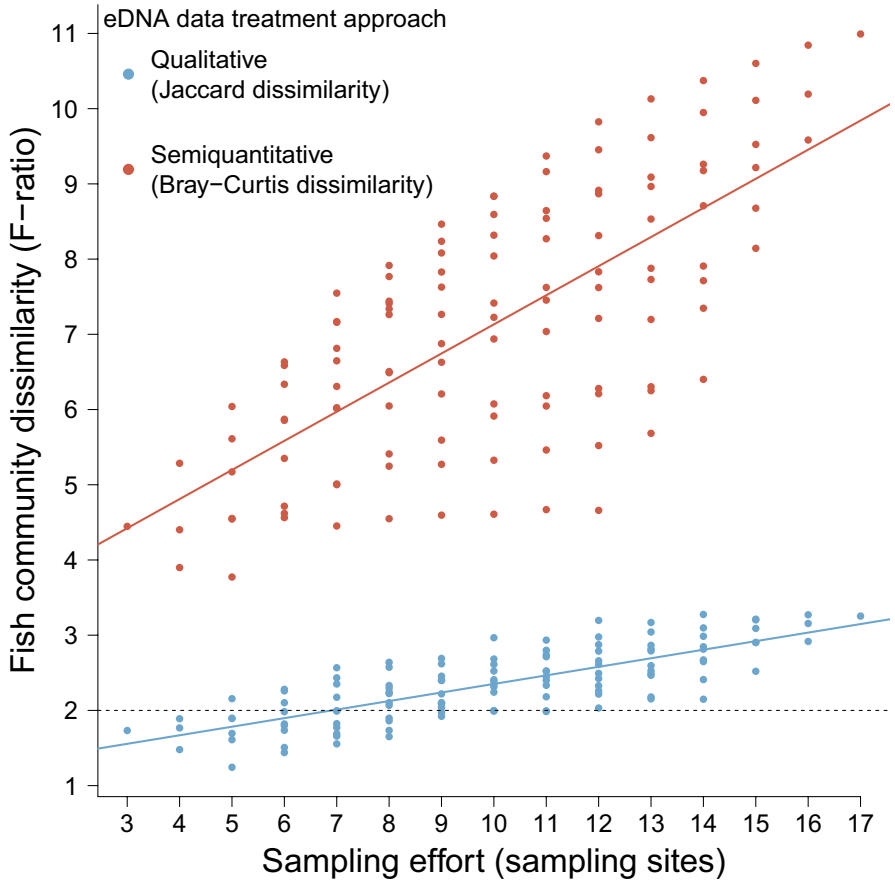

FIGURE 4 Dissimilarity of community composition between ecosystems measured as the PERMANOVA pseudo-*F* ratio using qualitative (blue; Jaccard dissimilarity matrix based on the presence/absence of species) and semiquantitative (red; Bray–Curtis dissimilarity matrix based on the inverse Wisconsin double-standardization of amplicon reads) approaches. The analyses were performed on progressively removed sampling stations, with each dot representing the mean pseudo-*F* of the sample combination (iterated 999 times) for the given number of remaining sampling stations. Regression lines for each approach are shown with their colors, respectively. The horizontal black dashed line, representing an approximation of pseudo-*F* value ($f = 2.01$) equal to $p = 0.05$, indicated that minimum three and 13 sampling stations are required for detecting significant differences between the three ecosystems respectively for semiquantitative and qualitative approach.

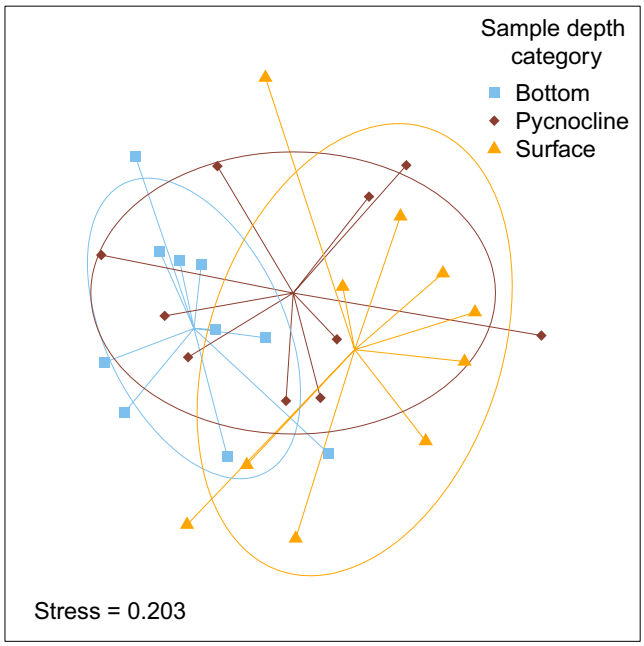

FIGURE 5 Nonmetric multidimensional scaling (nMDS) ordination of Balsfjord samples (biological replicates pooled) obtained using the Bray–Curtis dissimilarity index based on the inverse Wisconsin double-standardization of amplicon reads. The centroids for each depth category (distinguished by color) and their 95% confidence intervals (shown as ellipses) indicate a gradual shift of the vertical communities in the multivariate space.

3.3 | Sampling efficiency for maximum vertical and horizontal community difference

The nMDS plot showed a gradual change of the communities along the depth as the centroids shifted aligned with the depth vector (Figure 5). PERMANOVA test indicated significant differences in communities between at least two depth categories in both approaches ($p = 0.014$; Table 2). Pair-wise comparison of communities showed significant differences between bottom and surface samples ($p < 0.001$; Table S2c). However, no significant differences were found between bottom – pycnocline and surface – pycnocline pair comparisons ($p = 0.118$ and $p = 0.396$, respectively). PERMDISP indicated homogeneous dispersion of fish communities among depth categories, which suggests that community differences are solely due to centroid differences and not due to differences in heterogeneity.

Beta regression indicated that dissimilarity of pycnocline and surface samples differed significantly across the horizontal distance ($p = 0.024$, and $p < 0.001$) but these changes were insignificant on bottom samples ($p = 0.237$; Table S5). Changes in dissimilarity across distance were significantly different for surface samples compared to pycnocline and bottom samples ($p < 0.001$ for both; Table S5). Conversely, changes in pycnocline samples dissimilarity index across the horizontal distance compared to bottom samples resulted insignificant ($p = 0.439$; Table S5). The overall maximum dissimilarity for pycnocline and bottom samples was achieved at ca. 26 km while

TABLE 2 PERMANOVA tests results using semiquantitative (Bray–Curtis dissimilarity matrix) approaches comparing community composition among the factor depth (only samples in Balsfjord selected).

Factor	df	SS	MS	Pseudo-*F*	R^2	*p*-Value	PERMDISP
Depth (categorical)	2	0.793	0.396	**2.090**	0.134	0.014	0.127
Residuals	27	5.124	0.189		0.865		
Total	29	5.917			1.000		

Note: PERMDISP probabilities for homogeneity of dispersion are also shown. Significant values are indicated in bold.

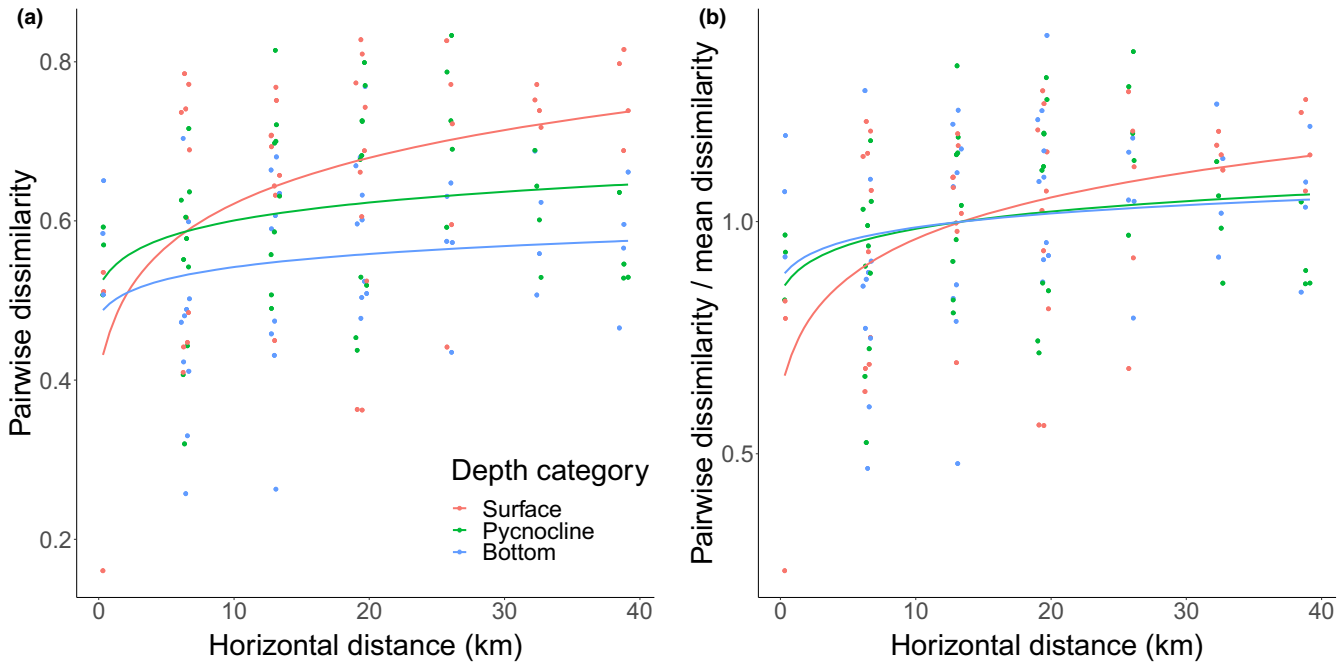

FIGURE 6 Correlation plot of samples' pairwise dissimilarity (a) and the ratio of pairwise dissimilarity over the average dissimilarity per each depth (b) along samples' pairwise horizontal distance. The dissimilarity measured using the Bary–Curtis index based on inverse Wisconsin double-standardization of amplicon reads indicated higher differences on surface communities (a) and average dissimilarity was achieved after 16 km of horizontal distance between samples (b). The standard errors are presented for both plots.

for the surface samples at 40 km (Figure 6a). Subsequently, samples' pairwise dissimilarity were divided with the average dissimilarity per each depth category to empirically measure the relative dissimilarity of samples across the horizontal distance (Figure 6b).

Additionally, the intrinsic and extrinsic variability was diagnosed via violin plots (Figure 7) on two levels, fjord ecosystem (Balsfjord) and regional ecosystem (all fjords together) using the distribution of dissimilarity between samples. The plot indicated that slightly more than half of the samples had intrinsic variability lower than the extrinsic variability within one ecosystem and the intrinsic variability of remaining samples was equal to extrinsic variability of that ecosystem (Figure 7). When looking at all ecosystems together, the intrinsic variability (within each fjord variability) was distinguishably lower than the extrinsic variability (between fjords variability; Figure 7).

4 | DISCUSSION

The sampling design distinguishing fish communities eDNA metabarcoding still poses challenges regarding the optimization of sampling depth, distance between samples and data treatment approaches. This study aimed to investigate the optimal data treatment approach, sampling effort, and spatial distribution of eDNA metabarcoding samples allowing to effectively detect differences in communities within and between fjords. Our study found that using two sampling stations (two depths, and three biological replicates), 26–40 km apart in Balsfjord, one in Frakkfjord and one in Olderfjord, when using a semiquantitative approach and omitting pycnocline depth samples, was the most efficient design to detect differences between ecosystems given the sampling effort and the fish community dissimilarity within the ecosystem.

4.1 | Sequencing and eDNA sampling effort

Exploring the unbalanced sequencing effort curves (Figure 2a) revealed that samples with higher read abundances of *Gadus morhua* and *Mallotus villosus* also had a high sequencing depth. Cumulatively, the sequence read abundance of these two species accounted for 76.6% of the total reads in all samples. Metabarcoding sequence read

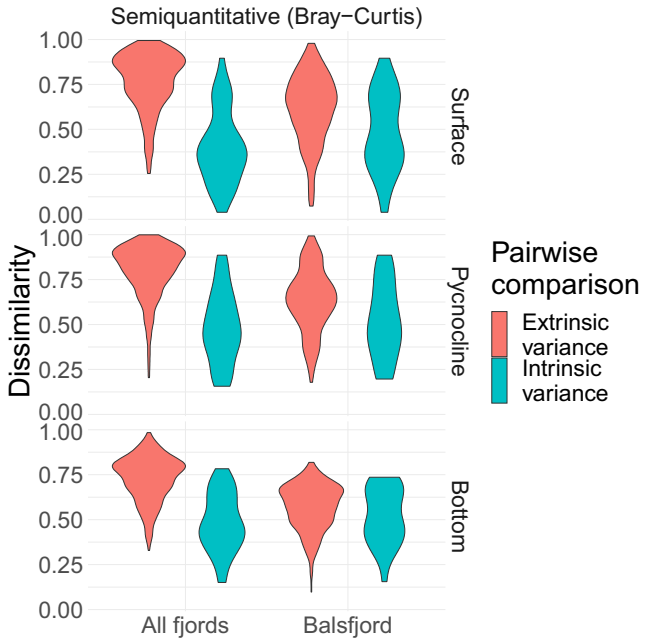

FIGURE 7 Violin plots comparing the intrinsic and extrinsic variability of dissimilarity index (using the Bray–Curtis index based on inverse Wisconsin double-standardization of amplicon reads) on two instances, fjord level (i; Balsfjord, right) and regional level (ii; all fjords, left). In fjord level (i.e., Balsfjord), variability between biological replicates was indicated as intrinsic (blue) and between all samples within the fjord as extrinsic (red). In all fjords, the variability within each fjord was indicated as intrinsic (blue) and between fjords as extrinsic (red).

abundance is a function of species DNA abundance in the samples, together with primer bias (Elbrecht & Leese, 2015). For instance, species with high abundance, but also high amplification efficiency, can result in a skewed distribution (i.e., over-represented high proportion) of sequence read abundances among samples. We consider a similar process to have caused the unbalanced sequencing depth of our localities, as *G. morhua* and *M. villosus* had high sequence read abundance predominantly in Frakkfjord and Olderfjord (the localities with higher sequencing depth; Figure S3). An additional explanation of the unbalanced sequencing effort could be the difference in the aggregation of nontarget taxa (prokaryotic and non-target eukaryotic DNAs) when competing for oligonucleotides during PCR (Díaz et al., 2020; Miya et al., 2020). Despite the nonuniform sequencing depth, the curves have similar shapes among the localities (Figure 2a), suggesting similar species detection rates among distinct fjords. Furthermore, our sequencing depth among distinct localities can be considered adequate for conducting our analysis but not adequate for covering the full diversity of the studied localities. Such limitation of sequencing depth can critically affect the quality of qualitative approach analysis as this approach strongly relies on detecting presence of non-abundant taxa.

The DNA collection curves (Figure S2b) suggested that increasing the number of sampling stations in Balsfjord might uncover more species, indicating that the number of sampling stations deployed in Balsfjord was insufficient for representing the total biodiversity of this ecosystem. Given the excessive area engaged in Balsfjord, the spatial heterogeneity can be greater than in the other two localities; thus, increasing the sample density can potentially uncover this ecosystem's additional taxonomic complexity.

4.2 | Differences between qualitative and semiquantitative approaches

Although we expected the localities to differ in fish community complexity (with respect to species abundances), evidence for differences among locations differed drastically between the qualitative and semiquantitative approaches. Overall, both methods indicated differences in community composition between at least two ecosystems (Table 1). But the semiquantitative approach discriminated samples between all localities (including the positive control; Table S2a). In contrast, the qualitative approach failed to distinguish the samples between Frakkfjord and Olderfjord and to distinguish the positive control samples from the remaining samples in all localities (Table S2b). Focusing on only the presence/absence of species when using eDNA metabarcoding can lead to inconclusive results due to the limitations of this approach (Wang et al., 2021). Below, we discuss how multiple reasons can lead to such limitations and make true biodiversity differences indistinguishable when using a qualitative approach to eDNA metabarcoding.

First, communities with ubiquitous species make it impossible for a qualitative approach (without any quantitative information) to distinguish spatial differences in communities. In our survey, we observed that almost one-third of species were ubiquitous (Table S3). Being present in all the sampling stations, such species cannot contribute to the dissimilarity between localities measured by presence/absence metrics. Hence, less frequent species have higher contributions to the differences between localities. Moreover, rare taxa are more prone to PCR biases than common taxa (Shirazi et al., 2021). Due to the stochasticity of the PCR amplification, rare taxa are detected sporadically within samples, making their reliability low. Such an issue was observed in our positive control samples, where seven taxa were found present only in one sample despite the source of the samples being the same (Figure S4), explaining the high dissimilarity between positive control samples. Larson et al. (2022) found similar results regarding rare taxa and stated that they occurred sporadically among PCR replicates. An additional cause to such patterns could also be explained due to amplification efficiency. Sequence abundance output is subjected to amplification biases due to primer efficiency (Kelly et al., 2019), especially when using universal primers, where selected taxa have a higher match with the primer than others (Banerjee et al., 2021). Such biases influence the commonness of the species, which can lead to divergence from the true species as drivers of differences between ecosystems (since low amplification species will be encouraged to be the species indicators).

Analogously, Cilleros et al. (2019) found metabarcoding results to be less discriminant when using the qualitative approach compared to traditional methods in an attempt to segregate the

faunal composition between large rivers and nearby streams. This is explained due to DNA transport as eDNA metabarcoding samples detected species found distant from the sampling source, thus increasing the homogeneity of the ecosystems (Cilleros et al., 2019). Using a qualitative approach can uprise biases from DNA transport, especially when the survey area is small compared with the DNA transport area (i.e., sampling distance is smaller than eDNA transport) as the DNA of the same origin can be captured in multiple samples simultaneously, thus less discrimination between localities. Lastly, errors arising from tag jumping (index hopping) and potential contamination are nontrivial these issues increase the occurrence of false-positive taxa, contributing to the aforementioned biases when using a qualitative approach. Although we included negative controls during all steps and stated a relatively clean workflows, the average sequence abundance in blanks was 32.2 reads, indicating that false-positive taxa can potentially be present in samples. In an additional exploratory analysis, we accounted for such false-positive by increasing the threshold of occurrence to the highest sequence abundance in negative control samples (minimum number of reads for defining amplicon presence = 81 reads) and found a substantial increase in disparity between ecosystems (Figure S5).

Notwithstanding, the semiquantitative approach does not entirely eliminate all the aforementioned obstacles. Issues such as DNA transport are prevalent regardless of the approach used for data inference. Nevertheless, using a semiquantitative approach can minimize the biases deriving from such matters. The concentration of transported eDNA decreases with increasing distance from the source due to dilution and degradation (Goldberg et al., 2018). Having a high density right at the source and exponentially lower density at neighboring sampling stations can result in relatively lower sequence abundance in the latter samples (Zou et al., 2020). Although many biotic and abiotic factors can potentially affect the displacement and the decay rate of DNA in the water (Harrison et al., 2019; Holman et al., 2021). Andruszkiewicz et al. (2019) predicted an average DNA transport of 5 km with 30% decay rate. Moreover, semiquantitative approaches lower the importance of rare taxa in detecting differences among communities compared with a qualitative approach, as their relative abundances are generally low. This simply alleviates the contribution of stochastic processes (such as PCR amplification of rare taxa) to determine the drivers of community differences. Semiquantitative approaches also minimize the contribution of false positives from tag jumping (typically in low abundance levels) by simply devaluing their importance due to low sequence abundance. Additionally, using the eDNA index as a semiquantitative approach offers an archaic way to take amplification efficiency into consideration (Kelly et al., 2019). Issues regarding ubiquitous taxa can be easily accounted for by using a semiquantitative approach, as differences in abundance will be reflected as contributions to quantitative dissimilarity indices. Although many studies have confirmed, to some extent, the correlation of eDNA metabarcoding sequence abundance with species biomass, more studies are required to come to sound conclusions. Nonetheless, our positive control samples revealed similar sequence abundance for common species, indicating a fairly robust PCR process with relatively comparable outcomes.

The difference in the two approaches is also reflected during species indicator analysis (Figure S2) as using the semiquantitative approach recognized considerably more species as indicators. Such analysis relies on relative abundances and frequencies of occurrences; hence, by removing the former layer of information (i.e., qualitative approach), it can result in reluctant conclusions. Therefore, having relative abundances as an additional layer of information contributes to finding more species as significant drivers and offers a more realistic and accurate detection of diversity patterns between localities.

Additionally, PERMANOVA on sample removal analysis indicated that the signals of dissimilarity are maintained using the semiquantitative approach (Figure 4), even when the number of sampling stations is reduced. Although reducing the number of sampling stations can have other indirect impacts on additional information, such as the heterogeneity of the ecosystem, we conclude that strong community dissimilarity can be efficiently recognized by implementing a semiquantitative approach. In contrast, it might be masked when a presence/absence approach is used. These findings could result from the small number of fish taxa ($n = 33$) present in the relatively low-diversity subarctic fjords and applying similar analyses in more diverse ecosystems might lead to different results. Moreover, having observations from only one season could impair our conclusions, as the seasonal variability might alter the differences between the ecosystems, especially in subarctic ecosystems where the seasonal variability is high mainly due to light conditions. These factors, together with the different lifecycles of fish species (i.e., spawning or feeding cycles), could reshape our conclusions. Moreover, clustering the MOTUs into species (as we did in this study) can diminish the heterogeneity of the study, therefore, enhancing the similarities between ecosystems. Jeunen, Knapp, et al. (2019) found higher dissimilarity when using MOTUs compared with species. They indicated that higher proportion of rare MOTUs can contribute to discrepancies between ecosystems when using a qualitative approach. Although the latter study used a COI marker, such properties can also be inherited by our marker choice.

However, we cross-validated our results with previous studies and management reports conducted on these localities. We found that the fish communities detected in our study matched the species composition recorded in Artsdatabanken (access date: March 2022; https://artsdatabanken.no/). Despite that the most abundant species were ubiquitous, similarly to eDNA data, records of *Ciliata mustela* and *Scophthalmus maximus* were only found in Frakkfjord. Moreover, our eDNA index results found similar outcomes, most abundant species as trawl catches in Balsfjord (as part of annual coastal surveys by IMR, unpublished data). Similar matches between eDNA and trawling have been previously documented (Fraija-Fernández et al., 2020). Kiærbech (2017) mentioned a predominant aggregation of *Melanogrammus aeglefinus* and *Pleuronectes platessa* in Balsfjord. Meanwhile *G. morhua* and *Hippoglossoides platessoides* are more abundant in the Olderfjord area. Such differences in

assemblages could be explained by the physical and chemical characteristics of the three fjords (Jo et al., 2019), as Balsfjord has limited water exchange due to closure by a large sill at the entrance of the fjord, whereas Olderfjord and Frakkfjord are more open and have a higher water exchange. Our eDNA results matched those traditional surveys as the eDNA index for *M. aeglefinnus* and *P. platessa* was predominantly aggregated in Balsfjord (Figure S6). Additionally, exceptionally high signals of *M. villosus* were found in Frakkfjord which is highly likely due to the Barents sea population of this species having one of their major spawning grounds on the Fugløy bank just outside the fjord (Alrabeei et al., 2021). Lastly, the innermost part of Balsfjord is a known and regionally important spawning area for Atlantic cod (Aglen et al., 2020), corresponding to high eDNA index in this part of the fjord (Figure S7).

Such consistency indicates that the eDNA index can robustly be used as a semiquantitative assessment of fish communities. Many studies have now confirmed the strengths of eDNA metabarcoding for local detection of species and, thus community composition inference (Hansen et al., 2018). Similarly to analogous studies (Cilleros et al., 2019; Fraija-Fernández et al., 2020; Jeunen, Knapp, et al., 2019; Jo et al., 2019; Larson et al., 2022; Li et al., 2021; Turon et al., 2022), our eDNA metabarcoding study confirm that fish community composition and biodiversity patterns can be reliably estimated using this approach.

4.3 | Sampling efficiency for vertical discrepancy and horizontal community dissimilarity

We showed that communities were segregated according to depth (Table 3), and this result confirms previous studies showing that eDNA metabarcoding can resolve vertical assemblages (Closek et al., 2019; Jeunen, Lamare, et al., 2019). However, in our results, the pycnocline community changed insignificantly when compared with bottom and surface samples (Table S2c), indicating a gradual transition of the communities along the depth where the significant difference could be detected only when sampling on the extremes of water column (surface and bottom samples) for our study area. Our findings are utterly aligned with those found in Closek et al. (2019), as they stated no differences in communities above and below the pycnocline samples, and these differences could only start being observed at a minimum depth distance of 80 m. This result suggests that removing pycnocline samples would not affect the robustness of eDNA metabarcoding studies when used for monitoring communities in different ecosystems. Although removing pycnocline samples might relieve the sampling effort, we note that such signal sensitivity might differ among various ecosystems and different seasons.

Research on DNA transport has repeatedly found that eDNA is deposited towards the bottom as part of the downwards transport (i.e., vertical settling; Andruszkiewicz et al., 2019; Canals et al., 2021; Turner et al., 2015). Additionally to regular settling, DNA particles can sink trough lateral advection dynamics in ice-free fjords (Canals et al., 2021; Wiedmann et al., 2016). Although sedimentation rate depends on the DNA state (Mauvisseau et al., 2021), studies have found that eDNA is 8–1800 times more concentrated in the bottom sediments than in surface samples (Turner et al., 2015). Canals et al. (2021) observed epipelagic fish eDNA in deep water samples, but not vice versa, concluding that eDNA had been transported downwards. Although it can be argued how environmental factors might affect this phenomenon (Turner et al., 2015), in our study, we observed higher species detectability in samples from the bottom waters, potentially indicating that bottom samples have a wider time span to record species occurrence at the site compared with surface samples which are mainly described as a snapshot of the present community (Díaz et al., 2020).

Among the three surveyed localities, Balsfjord had the highest heterogeneity (Figure 3). We noted that, in Balsfjord, the dissimilarity of communities increased with the distance between samples, indicating a gradual increase in fish species segregation. Coherently with other studies, our findings support the concept that eDNA can be used to measure heterogeneity of ecosystems (Wood et al., 2021). For instance, Fraija-Fernández et al. (2020) found that both methods (i.e., eDNA and trawling) tended to be more different when sampling stations were further apart.

We found significant difference of community dissimilarity in horizontal distance (Figure 6a) for surface and pycnocline communities. This indicated that with increase in spatial spread of sampling stations the dissimilarity increased for the latter communities. However, for bottom communities, these differences were not found to be significant, thus indicating that such communities remained spatially homogeneous although the spatial spreading of sampling stations increased. Such results demonstrate that the bottom layers can deposit DNA and thereby inferring a wider timespan of biological occurrence as mentioned above (Mestre et al., 2018). Subsequently, we noted that the most efficient sampling design—the highest dissimilarity within the ecosystem given the total dissimilarity of the ecosystem—was observed for samples obtained 26 km apart from each other when deploying bottom and pycnocline samples, and 40 km when deploying surface samples.

As no previous studies have been conducted in marine environments regarding spatial changes of dissimilarity, it is challenging to cross-validate our conclusions. However, Zhang et al. (2020) surveyed eDNA sampling design in a lentic system and found a spatial autocorrelation on samples up to 2 km apart. Although marine environments are more complex compared with lentic ones, their finding aligns with the concept of increased heterogeneity with increasing distances between samples. Moreover, Andruszkiewicz et al. (2019) reported that eDNA, on average, spreads until about 10 km after 7 days (when including the settling process) and furthermore estimated that, when in a moderate decay scenario, its source of origin can be <20 km. Incorporating those findings into our research, we have reasons to think that eDNA transport can influence the detection of heterogeneity in metabarcoding studies in ways that we have aforementioned. Nevertheless, we encourage follow-up studies on this topic as it can advance state of the art for eDNA metabarcoding applications in biomonitoring.

In short, we indicated that deploying sampling stations at such distance range from each other would yield the highest community dissimilarity within an ecosystem while avoiding unnecessary sampling. Note that we did not have high statistical power to precisely conclude the homogeneity of samples very close to each other ($n=3$) or very far apart from each other ($n=4$); therefore, we express the need for similar studies with an increased number of sampling stations. Moreover, our study design is limited to only three subarctic fjords with no repetition over the years and no changes in seasonality. Considering our findings, future studies should account for different ecosystems and include different seasons.

With the current rate of biodiversity loss in a climate crisis scenario, the need for more frequent sampling is increasing, therefore maximizing the efficiency of eDNA metabarcoding (i.e., reducing the unnecessary sampling while retaining the ability to detect differences in fish compositions) can enable monitoring agencies to increase their biomonitoring frequency. Reducing the amount of unnecessary sampling and the costs associated while still maintaining robust fish ecological inferences can foster the adaptation of eDNA metabarcoding into biomonitoring of coastal fish and ecosystem-based management.

5 | CONCLUSIONS

Qualitative and semiquantitative approaches to analyze eDNA metabarcoding data drastically differ in their efficiency to detect differences in communities between fjords. Thus, recommendations on a suitable number of samples and their spatial distribution to efficiently detect such differences is heavily dependent on the data treatment approach. For this reason, we stress the differences between the two approaches. Our results suggest that a semiquantitative approach has significantly higher efficiency to detect community differences from eDNA metabarcoding compared with a qualitative approach. Finally, our results showed that spreading the sampling stations ~30 km apart and avoiding sampling of the pycnocline can be sufficient for capturing fish community differences or changes in subarctic fjords when using a semiquantitative approach.

AUTHOR CONTRIBUTIONS

TJ, GG, NY, JIW, OSW, RK, JLR, AOS, TH, and KP designed the study. GG, TH, and JIW conducted the fieldwork. GG and TH did the laboratory work. GG, TH, and JIW did the bioinformatic analyses. GG did the statistical analyses. GG with support from TJ, RK, JLR, OSW, AOS, JIW, NY, and KP interpreted the statistical and ecological results. GG wrote the manuscript draft with contributions from all co-authors and all authors approved the submitted version.

ACKNOWLEDGMENTS

We thank Agneta Hansen for the significant contribution during the field and laboratory work and the crew onboard "F/F Kristine Bonnevie" for technical help with sampling instruments. This research was funded by the project FISHDIV (NRC301691) funded by the Norwegian Research Council (NRC) and the program "Marine Ecosystem processes" at Institute of Marine Research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The original raw sequences have been deposited in the NCBI SRA archive with accession number PRJNA878662.

ORCID

Gledis Guri https://orcid.org/0000-0002-3711-6758
Jon-Ivar Westgaard https://orcid.org/0000-0002-9627-8124
Nigel Yoccoz https://orcid.org/0000-0003-2192-1039
Owen S. Wangensteen https://orcid.org/0000-0001-5593-348X
Kim Præbel https://orcid.org/0000-0002-0681-1854
Jessica Louise Ray https://orcid.org/0000-0002-7305-5737
Ryan P. Kelly https://orcid.org/0000-0001-5037-2441
Andrew Olaf Shelton https://orcid.org/0000-0002-8045-6141
Tanja Hanebrekke https://orcid.org/0000-0003-3248-1681
Torild Johansen https://orcid.org/0000-0002-9858-052X

REFERENCES

Aglen, A., Nedreaas, K., Knutsen, J. A., & Huse, G. (2020). *Kysttorsk nord for 62-grader nord (Fisken og havet 2020-2)*.

Alrabeei, S., Subbey, S., Gundersen, S., & Gjøsæter, H. (2021). Spatial and temporal patterns of capelin (*Mallotus villosus*) spawning sites in the Barents Sea. *Fisheries Research*, *244*, 106117. https://doi.org/10.1016/j.fishres.2021.106117

Andres, K. J., Lambert, T. D., Lodge, D. M., Andrés, J., & Jackson, J. R. (2022). Combining sampling gear to optimally inventory species highlights the efficiency of eDNA metabarcoding. *Environmental DNA*, *5*(1) 146–157. https://doi.org/10.1002/edn3.366

Andruszkiewicz, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., & Boehm, A. B. (2019). Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Frontiers in Marine Science*, *6*, 1–14. https://doi.org/10.3389/fmars.2019.00477

Banerjee, P., Dey, G., Antognazza, C. M., Sharma, R. K., Maity, J. P., Chan, M. W., Huang, Y. H., Lin, P. Y., Chao, H. C., Lu, C. M., & Chen, C. Y. (2021). Reinforcement of environmental dna based methods (*Sensu stricto*) in biodiversity monitoring and conservation: A review. *Biology*, *10*(12), 1223. https://doi.org/10.3390/biology10121223

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*, 1–17. https://doi.org/10.1007/s10592-015-0775-4

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., Power, M., McLaughlin, J., Edgar, G. J., & Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, *2*(4), 493–504. https://doi.org/10.1002/edn3.74

Borja, Á., & Elliott, M. (2013). Marine monitoring during an economic crisis: The cure is worse than the disease. *Marine Pollution Bulletin*, *68*(1–2), 1–3. https://doi.org/10.1016/j.marpolbul.2013.01.041

Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., Griffith, J. F., Lettieri, T., Field, D., Benzie, J., Glöckner, F. O.,

Rodríguez-Ezpeleta, N., Faith, D. P., Bean, T. P., & Obst, M. (2013). Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin*, *74*(1), 19–31. https://doi.org/10.1016/j.marpolbul.2013.05.042

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182. https://doi.org/10.1111/1755-0998.12428

Buxton, A., Matechou, E., Griffin, J., Diana, A., & Griffiths, R. A. (2021). Optimising sampling and analysis protocols in environmental DNA studies. *Scientific Reports*, *11*, 11637. https://doi.org/10.1038/s41598-021-91166-7

Canals, O., Mendibil, I., Santos, M., Irigoien, X., & Rodríguez-Ezpeleta, N. (2021). Vertical stratification of environmental DNA in the open ocean captures ecological patterns and behavior of deep-sea fishes. *Limnology and Oceanography Letters*, *6*(6), 339–347. https://doi.org/10.1002/lol2.10213

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., Taberlet, P., Vigouroux, R., & Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. *Scientific Reports*, *9*, 3085. https://doi.org/10.1038/s41598-019-39399-5

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., Iribar, A., Taberlet, P., Vigouroux, R., & Brosse, S. (2019). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*, *19*, 27–46. https://doi.org/10.1111/1755-0998.12900

Closek, C. J., Santora, J. A., Starks, H. A., Schroeder, I. D., Andruszkiewicz, E. A., Sakuma, K. M., Bograd, S. J., Hazen, E. L., Field, J. C., & Boehm, A. B. (2019). Marine vertebrate biodiversity and distribution within the central California current using environmental DNA (eDNA) Metabarcoding and ecosystem surveys. *Frontiers in Marine Science*, *6*, 1–17. https://doi.org/10.3389/fmars.2019.00732

Coguiec, E., Ershova, E. A., Daase, M., Vonnahme, T. R., Wangensteen, O. S., Gradinger, R., Præbel, K., & Berge, J. (2021). Seasonal variability in the zooplankton community structure in a sub-Arctic Fjord as revealed by morphological and molecular approaches. *Frontiers in Marine Science*, *8*, 705042. https://doi.org/10.3389/fmars.2021.705042

Díaz, C., Wege, F. F., Tang, C. Q., Crampton-Platt, A., Rüdel, H., Eilebrecht, E., & Koschorreck, J. (2020). Aquatic suspended particulate matter as source of eDNA for fish metabarcoding. *Scientific Reports*, *10*, 14352. https://doi.org/10.1038/s41598-020-71238-w

Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, *67*(3), 345–366. https://doi.org/10.2307/2963459

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS One*, *10*(7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Fraija-Fernández, N., Bouquieaux, M. C., Rey, A., Mendibil, I., Cotano, U., Irigoien, X., Santos, M., & Rodríguez-Ezpeleta, N. (2020). Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. *Ecology and Evolution*, *10*(14), 7560–7584. https://doi.org/10.1002/ece3.6482

Gilbey, J., Carvalho, G., Castilho, R., Coscia, I., Coulson, M. W., Dahle, G., Derycke, S., Francisco, S. M., & Volckaert, F. A. M. (2021). Life in a drop: Sampling environmental DNA for marine fishery management and ecosystem monitoring. *Marine Policy*, *124*, 104331. https://doi.org/10.1016/j.marpol.2020.104331

Gold, Z., Sprague, J., Kushner, D. J., Marin, E. Z., & Barber, P. H. (2021). eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PLoS One*, *16*(2), e0238557. https://doi.org/10.1371/journal.pone.0238557

Goldberg, C. S., Strickler, K. M., & Fremier, A. K. (2018). Degradation and dispersion limit environmental DNA detection of rare amphibians in wetlands: Increasing efficacy of sampling designs. *Science of the Total Environment*, *633*, 695–703. https://doi.org/10.1016/j.scitotenv.2018.02.295

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307. https://doi.org/10.1111/2041-210X.12595

Gullestad, P., Abotnes, A. M., Bakke, G., Skern-Mauritzen, M., Nedreaas, K., & Søvik, G. (2017). Towards ecosystem-based fisheries management in Norway – Practical tools for keeping track of relevant issues and prioritising management efforts. *Marine Policy*, *77*, 104–110. https://doi.org/10.1016/j.marpol.2016.11.032

Hansen, B. K., Bekkevold, D., Clausen, L. W., & Nielsen, E. E. (2018). The sceptical optimist: Challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries*, *19*(5), 751–768. https://doi.org/10.1111/faf.12286

Harrison, J. B., Sunday, J. M., & Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1915), 20191409. https://doi.org/10.1098/rspb.2019.1409

Holman, L. E., Chng, Y., & Rius, M. (2021). How does eDNA decay affect metabarcoding experiments? *Environmental DNA*, *4*(1), 108–116. https://doi.org/10.1002/edn3.201

Jeunen, G. J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., Bunce, M., & Gemmell, N. J. (2019). Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, *19*(2), 426–438. https://doi.org/10.1111/1755-0998.12982

Jeunen, G. J., Lamare, M. D., Knapp, M., Spencer, H. G., Taylor, H. R., Stat, M., Bunce, M., & Gemmell, N. J. (2019). Water stratification in the marine biome restricts vertical environmental DNA (eDNA) signal dispersal. *Environmental DNA*, *2*(1), 99–111. https://doi.org/10.1002/edn3.49

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

Jo, H., Kim, D. K., Park, K., & Kwak, I. S. (2019). Discrimination of spatial distribution of aquatic organisms in a coastal ecosystem using eDNA. *Applied Sciences*, *9*(17), 3450. https://doi.org/10.3390/app9173450

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-48546-x

Kiærbech, K. (2017). *Do diet of long rough dab (Hippoglossoides platessoides) differ between fjords with and without red king crab (Paralithodes camtschaticus)? MSc Thesis, UiT The Arctic University of Norway, https://hdl.handle.net/1003711340*

Larson, W. A., Barry, P., Dokai, W., Maselko, J., Olson, J., & Baetscher, D. (2022). Leveraging eDNA metabarcoding to characterize nearshore fish communities in Southeast Alaska: Do habitat and tide matter? *Environmental DNA*, *4*(4), 868–880. https://doi.org/10.1002/edn3.297

Li, W. P., Liu, Z. F., Guo, T., Chen, H., & Xie, X. (2021). Using optimal environmental dna method to improve the fish diversity survey—From laboratory to aquatic life reserve. *Water*, *13*(11), 1468. https://doi.org/10.3390/w13111468

Li, Y., Evans, N. T., Renshaw, M. A., Jerde, C. L., Olds, B. P., Shogren, A. J., Deiner, K., Lodge, D. M., Lamberti, G. A., & Pfrender, M. E. (2018).

Estimating fish alpha- and beta-diversity along a small stream with environmental DNA metabarcoding. *Metabarcoding and Metagenomics*, *2*, e24262. https://doi.org/10.3897/mbmg.2.24262

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593. https://doi.org/10.7717/peerj.593

Mauvisseau, Q., Harper, L., Sander, M., Hanner, R. H., Kleyer, H., & Deiner, K. (2021). The multiple states of environmental DNA and what is known about their persistence in aquatic environments. *Environmental Science & Technology*, *56*(9), 5322–5333. https://doi.org/10.1021/acs.est.1c07638

Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., & Sala, M. M. (2018). Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(29), E6799–E6807. https://doi.org/10.1073/pnas.1802470115

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: A high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fisheries Science*, *86*(6), 939–970. https://doi.org/10.1007/s12562-020-01461-x

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

Nakagawa, H., Yamamoto, S., Sato, Y., Sado, T., Minamoto, T., & Miya, M. (2018). Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. *Freshwater Biology*, *63*(6), 569–580. https://doi.org/10.1111/fwb.13094

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2022). *Vegan: Community ecology package*. https://cran.r-project.org/package=vegan

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reigstad, M., & Wassmann, P. (1996). Importance of advection for pelagic-benthic coupling in north Norwegian fjords. *Sarsia*, *80*(4), 245–257. https://doi.org/10.1080/00364827.1996.10413599

Roberts, D. W. (2019). *labdsv: Ordination and multivariate analysis for ecology*. https://cran.r-project.org/package=labdsv

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Sales, N. G., Wangensteen, O. S., Carvalho, D. C., Deiner, K., Præbel, K., Coscia, I., McDevitt, A. D., & Mariani, S. (2021). Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. *Science of the Total Environment*, *754*, 142096. https://doi.org/10.1016/j.scitotenv.2020.142096

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., & Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Communications Biology*, *2*(1), 1–9. https://doi.org/10.1038/s42003-019-0696-8

Sanches, T. M., & Schreier, A. D. (2020). Optimizing an eDNA protocol for estuarine environments: Balancing sensitivity, cost and time. *PLoS One*, *15*(5), 1–18. https://doi.org/10.1371/journal.pone.0233522

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*(22), 15766–15779. https://doi.org/10.1002/ece3.8239

Shu, L., Ludwig, A., & Peng, Z. (2020). Standards for methods utilizing environmental dna for detection of fish species. *Genes*, *11*(3), 296. https://doi.org/10.3390/genes11030296

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*, 12240. https://doi.org/10.1038/s41598-017-12501-5

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

Turon, M., Nygaard, M., Guri, G., Wangensteen, O. S., & Præbel, K. (2022). Fine-scale differences in eukaryotic communities inside and outside salmon aquaculture cages revealed by eDNA metabarcoding. *Frontiers in Genetics*, *13*, 957251. https://doi.org/10.3389/fgene.2022.957251

Wang, S., Yan, Z., Hänfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology of fish eDNA and its applications in ecology and environment. *Science of the Total Environment*, *755*, 142622. https://doi.org/10.1016/j.scitotenv.2020.142622

Wiedmann, I., Reigstad, M., Marquardt, M., Vader, A., & Gabrielsen, T. M. (2016). Seasonality of vertical flux and sinking particle characteristics in an ice-free high arctic fjord-different from subarctic fjords? *Journal of Marine Systems*, *154*, 192–205. https://doi.org/10.1016/j.jmarsys.2015.10.003

Wood, Z. T., Lacoursière-Roussel, A., LeBlanc, F., Trudel, M., Kinnison, M. T., Garry McBrine, C., Pavey, S. A., & Gagné, N. (2021). Spatial heterogeneity of eDNA transport improves stream assessment of threatened Salmon presence, abundance, and location. *Frontiers in Ecology and Evolution*, *9*, 650717. https://doi.org/10.3389/fevo.2021.650717

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, *16*(8), 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4

Zhang, S., Lu, Q., Wang, Y., Wang, X., Zhao, J., & Yao, M. (2020). Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. *Molecular Ecology Resources*, *20*(1), 242–255. https://doi.org/10.1111/1755-0998.13105

Zou, K., Chen, J., Ruan, H., Li, Z., Guo, W., Li, M., & Liu, L. (2020). eDNA metabarcoding as a promising conservation tool for monitoring fish diversity in a coastal wetland of the Pearl River estuary compared to bottom trawling. *Science of the Total Environment*, *702*, 134704. https://doi.org/10.1016/j.scitotenv.2019.134704

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Guri, G., Westgaard, J.-I., Yoccoz, N., Wangensteen, O. S., Præbel, K., Ray, J. L., Kelly, R. P., Shelton, A. O., Hanebrekke, T., & Johansen, T. (2023). Maximizing sampling efficiency to detect differences in fish community composition using environmental DNA metabarcoding in subarctic fjords. *Environmental DNA*, *00*, 1–15. https://doi.org/10.1002/edn3.409