



Original Article

Metabarcoding as a quantitative tool for estimating biodiversity and relative biomass of marine zooplankton

E. A. Ershova ^{1,2,3,*}, O. S. Wangensteen ¹, R. Descoteaux¹, C. Barth-Jensen ¹, and K. Præbel¹

¹Faculty for Biosciences, Fisheries and Economics, Department for Arctic and Marine Biology, UiT The Arctic University of Norway, P.O. Box 6050, Langnes, 9037 Tromsø, Norway

²Institute of Marine Research, P.O. Box 1870, Nordnes, 5817 Bergen, Norway

³Shirshov Institute of Oceanology, Russian Academy of Sciences, 36 Nahimovskiy Prospekt, 117997 Moscow, Russia

*Corresponding author: tel: +47 458 44 964; e-mail: elizaveta.ershova@hi.no

Ershova, E. A., Wangensteen, O. S., Descoteaux, R., Barth-Jensen, C., and Præbel, K. Metabarcoding as a quantitative tool for estimating biodiversity and relative biomass of marine zooplankton. – ICES Journal of Marine Science, 0: 1–14.

Received 10 January 2021; revised 29 July 2021; accepted 16 August 2021.

Although metabarcoding is a well-established tool for describing diversity of pelagic communities, its quantitative value is still controversial, with poor correlations previously reported between organism abundance/biomass and sequence reads. In this study, we explored an enhanced quantitative approach by metabarcoding whole zooplankton communities using a highly degenerate primer set for the mitochondrial marker cytochrome oxidase I and compared the results to biomass estimates obtained using the traditional morphological approach of processing zooplankton samples. As expected, detected species richness using the metabarcoding approach was 3–4 times higher compared to morphological processing, with the highest differences found in the meroplankton fraction. About 75% of the species identified using microscopy were also recovered in the metabarcoding run. Within the taxa detected using both approaches, the relative numbers of sequence counts showed a strong quantitative relationship to their relative biomass, estimated from length-weight regressions, for a wide range of metazoan taxa. The highest correlations were found for crustaceans and the lowest for meroplanktonic larvae. Our results show that the reported approach of using a metabarcoding marker with improved taxonomic resolution, universal coverage for metazoans, reduced primer bias, and availability of a comprehensive reference database, allow for rapid and relatively inexpensive processing of hundreds of samples at a higher taxonomic resolution than traditional zooplankton sorting. The described approach can therefore be widely applied for monitoring or ecological studies.

Keywords: biodiversity, COI, ecosystem monitoring, Leray-Xt, meroplankton, mesozooplankton, metabarcoding, pelagic communities

Introduction

Mesozooplankton are critical components of the world oceans, providing an essential link from the ocean's microbial primary and secondary producers to the upper levels of the food web. Monitoring zooplankton communities provides us with critical insight on the state of the pelagic ecosystem, as well as the implications for consumers, including harvestable species. Compared to organisms in other marine and terrestrial habitats, most planktonic species are characterized by relatively short generation times and react very rapidly to shifts in the physical environment, acting as first sentinels of climate change (Richardson, 2008; Beaugrand *et al.*, 2009).

Additionally, since plankton drifts passively with currents, distinct planktonic communities and individual species can serve as useful markers of currents and water mass types (Beaugrand *et al.*, 2002). Monitoring of zooplankton communities is a common way to quantify the feeding stocks of higher trophic levels, including commercially valuable species (Orlova *et al.*, 2002). It can also be used to assess the general state of the ecosystem, such as introduction of alien species (Couton *et al.*, 2019; Sepulveda *et al.*, 2020) or changes in the community due to anthropological forcing (Andújar *et al.*, 2018). The need for such surveys globally is growing, as marine systems are placed under increasing strain through development of shipping and industry, pollution, overfishing, and

changing climate patterns (Chiba *et al.*, 2018; Klunder *et al.*, 2020).

Although zooplankton is easy to sample from both small and large vessels, and the methods of collection do not require special skills or expensive equipment, the post-processing of samples continues to be a bottleneck in most zooplankton monitoring surveys. For example, a recent study examining long-term ecosystem shifts in the Norwegian Sea included more than 1500 bulk zooplankton biomass measurements, yet merely one hundred samples were analysed taxonomically due to “limited capacity at the laboratory to analyze samples” (Toresen *et al.*, 2019). Despite significant advances in the development of optical, acoustic and molecular approaches in zooplankton ecology, these methods still often lack sufficient power to discriminate between taxa, and/or are not quantitative. Microscopic processing of preserved zooplankton samples remains therefore the main method for quantitative assessment of zooplankton communities at the species level. This time-consuming approach requires a trained taxonomist, and the processing of a single sample can take from several hours to several days. Additionally, this method is inadequate for quantifying organisms that deform or disintegrate in fixatives (i.e. gelatinous plankton) or those that lack sufficient morphological features for visual identification (many early larval stages of both planktonic and benthic animals). The data produced via manual sorting are also highly prone to human bias, with the education and background of the person performing the analysis playing a key role. As a result of the high time and expertise requirements, many months or years generally pass between sample collection and data acquisition, which may cause significant delays in management actions. This drives the need for the development of rapid, alternative methods of bulk assessment of planktonic communities.

In recent years, deoxyribonucleic acid (DNA) barcoding of zooplankton taxa has greatly reduced the time and expertise needed for zooplankton identification by taking advantage of the divergence in DNA sequences across taxa (Hebert *et al.*, 2003). Advances in DNA-based identification have challenged some established paradigms in zooplankton ecology, such as species distribution ranges (Choquet *et al.*, 2017), and highlighted the widespread existence of cryptic or pseudo-cryptic species complexes in the marine realm (Miyamoto *et al.*, 2010; BurrIDGE *et al.*, 2015; Kolbasova *et al.*, 2015; Cornils *et al.*, 2017; Ershova, 2020; Ershova *et al.*, 2021). However, most molecular studies to this day have focused on individual species or groups, and biodiversity or community analyses via metabarcoding are only now becoming widely applied in zooplankton research (Bucklin *et al.*, 2016, 2019; Zhang *et al.*, 2018; Santoferrara, 2019; Questel *et al.*, 2021). Unlike optical-based technologies, metabarcoding can potentially detect all species in a community, regardless of developmental stage or preservation of distinguishing features. Metabarcoding can therefore be used for the rapid detection of shifts in biodiversity and community composition, for monitoring of rare species that are unlikely to be captured in sufficient numbers to be identified visually, and for the detection of invasive species. One challenge that has limited the application of metabarcoding is finding a suitable DNA target region and a set of primers to amplify it. Ideally, the primers need to be truly universal, binding to all taxa equally, but the amplified DNA regions themselves need to be sufficiently genetically divergent across species while conserved within species, to allow species-level identification. Finally, to ensure the identification of all taxa present, the target DNA region needs to have a complete database of reference sequences, containing all species that may

be found in the community. To this day, the majority of metabarcoding plankton studies have used various regions of the 18S gene (Bucklin *et al.*, 2016), the primer-binding sites for which are well conserved across most taxa, but also provide relatively poor taxonomic resolution for metazoan groups (Mohrbeck *et al.*, 2015). Other commonly used markers include regions of the nuclear 28S rRNA and the mitochondrial 16S rRNA gene (Bucklin *et al.*, 2016). Although the mitochondrial cytochrome oxidase I (COI), the commonly used “barcoding” gene for metazoans, is generally too long for next-generation sequencing, good success has been demonstrated using smaller fragments of this gene, such as the one amplified by the Leray primer set (Leray and Knowlton, 2015).

Although metabarcoding is now a well-established tool for describing diversity of marine biological communities, its quantitative value is still controversial, with poor correlations previously reported between organism abundance and sequence reads (Bucklin *et al.*, 2016; Lamb *et al.*, 2019; Santoferrara, 2019; van der Loos and Nijland, 2020). Since the size of planktonic organisms spans ten orders of magnitude, the amount of DNA per organism is extremely variable, with even organisms of the same species often having vastly different numbers of DNA copies at different life stages. Better correlations have been obtained for biomass within some groups of organisms, typically, copepods (Clarke *et al.*, 2017; Yang *et al.*, 2017). In addition to an initial discrepancy in DNA copy numbers, extraction and polymerase chain reaction (PCR) methods introduce additional biases, resulting in better DNA recovery and amplification of some organisms than others. Given the exponential nature of PCR, these errors can compound to a very high degree.

In this study, we aimed to establish a relatively inexpensive and rapid protocol that would amount in a maximum recovery of diversity in environmental zooplankton samples, as well as provide quantitative information on relative biomass of organisms in the sample. To accomplish this, we used the highly degenerate modification of the Leray primers (Leray and Knowlton, 2015), Leray-XT (Wangensteen *et al.*, 2018), to amplify a 313-base pair (bp) region of the COI gene in whole zooplankton samples. Degeneration ensures high attachment across multiple groups of metazoans, and the availability of the Barcode of Life Database (BOLD) allows for a good rate of success in species-level identification, despite the existing gaps in databases for marine taxa (Wangensteen *et al.*, 2018). The results of metabarcoding using the Leray-XT primer set have been shown to reflect the quantitative composition of rocky reef benthic communities (Wangensteen *et al.*, 2018), freshwater invertebrates (Elbrecht and Leese, 2015), and trophic contents (Siegenthaler *et al.*, 2019), but have not yet been tested on marine zooplankton. Simultaneously, we processed the sample using traditional microscope sorting and estimated the biomass of all organisms present using published length-mass relationships. We hypothesized that (i) metabarcoding will recover a higher diversity of zooplankton, including most of the species identified via microscope sorting, and (ii) the relative proportion of sequence reads will be correlated with estimated relative biomass of the respective organisms. As such, this method could allow rapid and relatively inexpensive bulk processing of samples while limiting human bias, and provide a valuable tool in zooplankton monitoring studies.

Methods

Material

We used nine zooplankton samples from five locations in the Arctic and north Atlantic (Figure 1, Supplementary Table 1)—in the

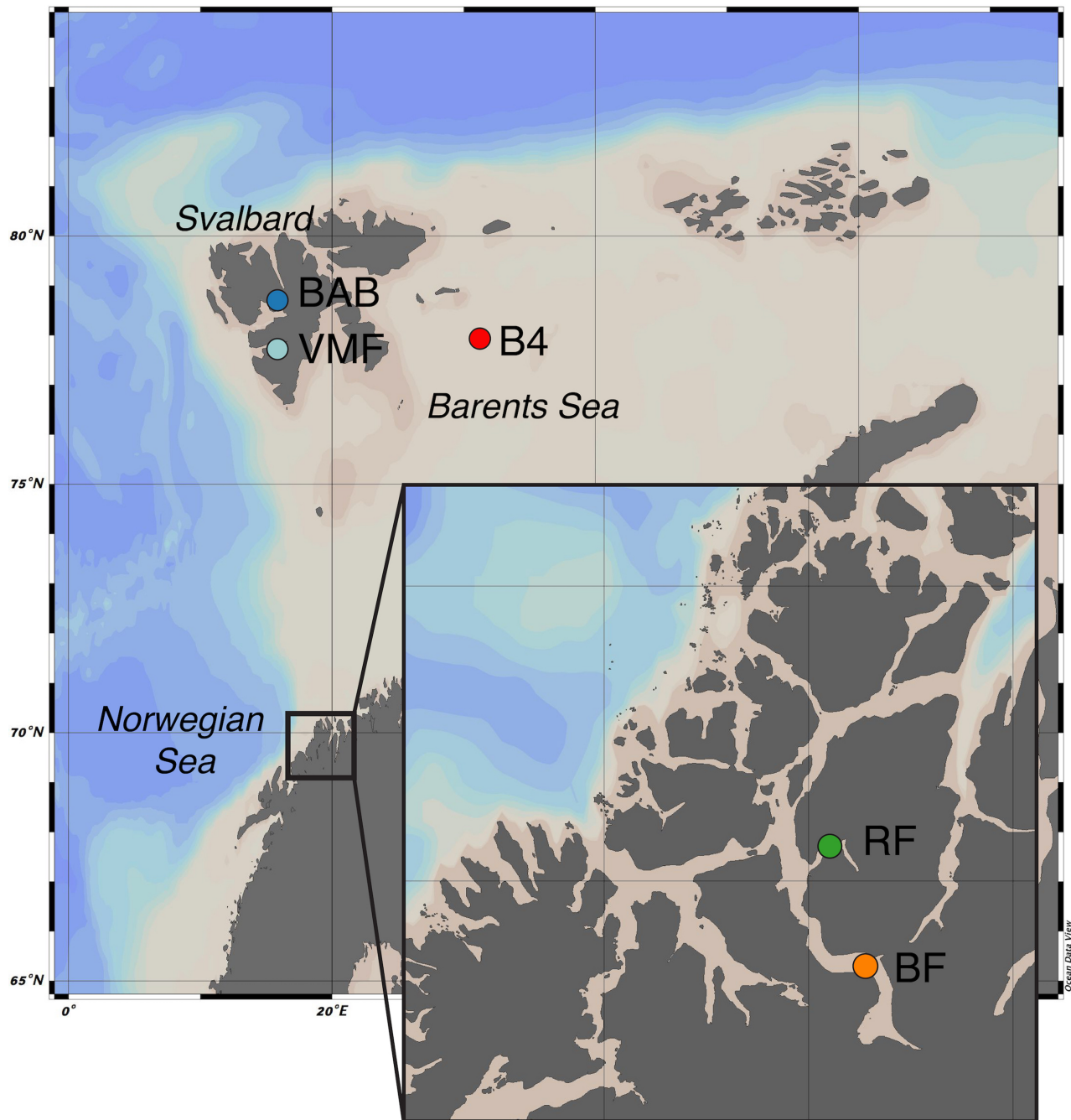


Figure 1. Locations of sampling stations used for analysis.

northern Barents Sea (one sample), Svalbard fjords (two samples), and north Norwegian fjords (six samples). The samples were collected using a WP2 net (mesh size 64 or 180 μm) and preserved with 96% ethanol. Samples were split quantitatively into two fractions. The first half was saved for traditional sorting and the other for molecular analysis. Any large (>1 cm) organisms in the sample were counted and measured prior to splitting.

Visual identification and biomass estimation

The subsample was washed from the ethanol with filtered seawater, after which it was split into consecutively smaller fractions using a

Folsom plankton splitter until approximately 100 individuals of the most common taxa remained in the final split. Increasingly larger fractions were examined for rarer groups until a total of approximately 500 individuals were counted under a Leica stereomicroscope at 25–40x magnification. The entire sample was screened for larger organisms such as chaetognaths or jellies. Each counted organism was measured using the ZoopBiom digitizing system (Roff and Hopcroft, 1986) and its biomass in milligram dry weight was estimated using a length–weight regression known from published literature for this or a similar species (Hopcroft *et al.*, 2010; Ershova *et al.*, 2015). *Pseudocalanus* spp. and *Microcalanus* spp. were iden-

tified at the genus level. *Calanus glacialis* and *Calanus finmarchicus* were identified based on prosome length and pigmentation as described in Unstad and Tande (1991). Meroplanktonic larvae were identified at the phylum or class level.

Molecular analysis

The second half of the sample was homogenized in a 2000 W blender for 3 minutes, then the excess ethanol was removed using centrifugation and three 0.3 g subsamples of the homogenate were taken for DNA extraction. Extraction was done using the Power-Soil DNA extraction kit (Qiagen) according to the manufacturer's protocol. Strict measures to avoid sample cross-contamination were observed, and all instruments were cleaned with 10% bleach and rinsed in distilled water between samples. Since DNA extracted from plankton net samples is organismal-DNA (not extraorganismal DNA), less strict measures to avoid contamination may be taken compared to eDNA samples (Rodríguez-Ezpeleta *et al.*, 2021). A ~313 bp fragment from the 5' region of the COI gene was amplified using the Leray-XT primers (Wangensteen *et al.*, 2018), which include the forward primer mCOLintF-XT 5'-GGWACWRGWTRACWITITAYCCYCC-3' and reverse primer jgHCO2198 5'-TAIACYTCIGGRTGICCRAARAAYCA-3' (Geller *et al.*, 2013). Both ends of the primers contained sample tags of 8 bp and a variable number (2–4) of leading Ns for improving sequence diversity. Amplification was conducted in a single PCR using AmpliTaq Gold DNA polymerase (Applied Biosystems), with 1 μ l of each 5 μ M forward and reverse tagged primers, 3 μ g of bovine serum albumin, and 2 μ l of extracted DNA in a total volume of 20 μ l per sample. The PCR protocol consisted of a denaturing step of 10 min at 95°C, followed by 35 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min, and a final extension of 5 min at 72°C. Two PCR blanks were added to the pool of multiplexed samples, which yielded only a few metabarcoding reads (<30 reads/blank sample). The PCR products were purified using Minelute PCR purification columns (www.qiagen.com) and pooled into a single library. The Illumina library was prepared from the DNA pool using the NextFlex PCR-free library preparation kit (Perkin-Elmer). This protocol incorporates Illumina adapters using a ligation procedure without any further PCR step, thus minimizing biases. The resulting library was sequenced in an Illumina MiSeq using $\frac{1}{2}$ of a V3 2 \times 250 bp kit (Illumina).

Bioinformatics

The first quality filtering steps of the bioinformatics pipeline were conducted using OBITools v.1.01.22 (Boyer *et al.* 2016). Paired-end reads were aligned with *illumina-pairedend* and reads with alignment score >40 were kept. Demultiplexing and removal of primer sequences were done using *ngsfilter*. Reads with length between 299 and 320 and without ambiguous nucleotides were selected using *obigrep* and dereplicated using *obiuniq*. The *uchime_denovo* algorithm (Edgar *et al.*, 2011) implemented in *vsearch* v.1.10.1 (Rognes *et al.*, 2016) was then used to remove chimeric sequences. Step-by-step clustering was performed in SWARM 2.1.13 (Mahé *et al.*, 2015) using a distance value of $d=13$ to cluster individual sequences into molecular operational taxonomic units (MOTUs). This distance value has previously been used to cluster similar datasets using the same COI fragment (e.g. Bakker *et al.*, 2019; Antich *et al.*, 2020; Atienza *et al.*, 2020). After removing singletons (MOTUs with abundance of one read), taxonomic assignment

of the representative sequence of remaining MOTUs was then performed using *ecotag* (Boyer *et al.*, 2016) against DUFA-Leray v.2020-06-10, a custom reference database (publicly available from github.com/uit-metabarcoding/DUFA), which included Leray fragment sequences extracted from BOLD and Genbank, completed with in-house generated sequences. The resulting dataset was curated for putative pseudogene sequences using the algorithm LULU (Frøslev *et al.*, 2017). The next refining steps consisted of removing MOTUs assigned to Prokaryotes and taxonomy check of selected MOTUs using BOLD (www.boldsystems.org). A species level identification was assigned with a minimum of 97% similarity, with the exception of *Parasagitta elegans*, which can have extremely high within-species divergence of the mitochondrial genome (Marlétaz *et al.*, 2017) and for which a similarity cut-off of 90% was used. A barcoding gap analysis was performed for the most common planktonic metazoan taxa with enough intra-species information in BOLD (Supplementary File F1). This analysis confirmed similar behaviours for the whole Folmer barcode fragment (~673 bp) and the Leray fragment (~313 bp). Therefore, identical similarity thresholds can be used for both markers for a species-level assignment match. All MOTUs belonging to clearly non-planktonic organisms (i.e. insects and mammals) were removed. Finally, only MOTUs observed in a minimum of two sample replicates and accounting for at least 0.01% of the total reads of any sample were kept in the final dataset.

Data analysis

All analyses were performed in R (R Development Core Team, 2011). The recovery of biodiversity via metabarcoding was estimated by rarefaction curves using the package “vegan” (Oksanen *et al.*, 2016). Relative biomass and relative sequence counts (%) were used in analyses. These values were fourth-root transformed to account for the very large spread of the biomass values and the exponential nature of PCR in the sequencing data. Overall community structure and replication was investigated using non-metric multidimensional scaling (nMDS) on Bray–Curtis dissimilarities of fourth-root transformed relative sequence counts using “vegan” (Oksanen *et al.*, 2016). Relationships between relative abundance, biomass, and sequence counts (%) of both broad taxonomic categories and individual species across samples and within each individual sample were examined using simple linear regressions, with a Bonferroni correction to the p -value when multiple comparisons were included (p_{adjust}). Significance was assumed at $p < 0.05$.

Results

Sequencing summary

The MiSeq run produced 4082508 paired-end raw reads for the multiplexed library. After all quality filtering steps, the final dataset consisted of 3418499 metabarcoding reads, clustered into 477 eukaryotic MOTUs. The sequencing depth per sample ranged from 45109 to 238991 reads, with an average value of 128840 reads per sample. Rarefaction curves (Supplementary Figure 1) suggest that this was enough to recover most, although not all diversity, although some individual station replicates failed to reach an asymptote.

Replication

The nMDS ordination (Figure 2) showed the three extraction replicates of each sample to be very similar in composition,

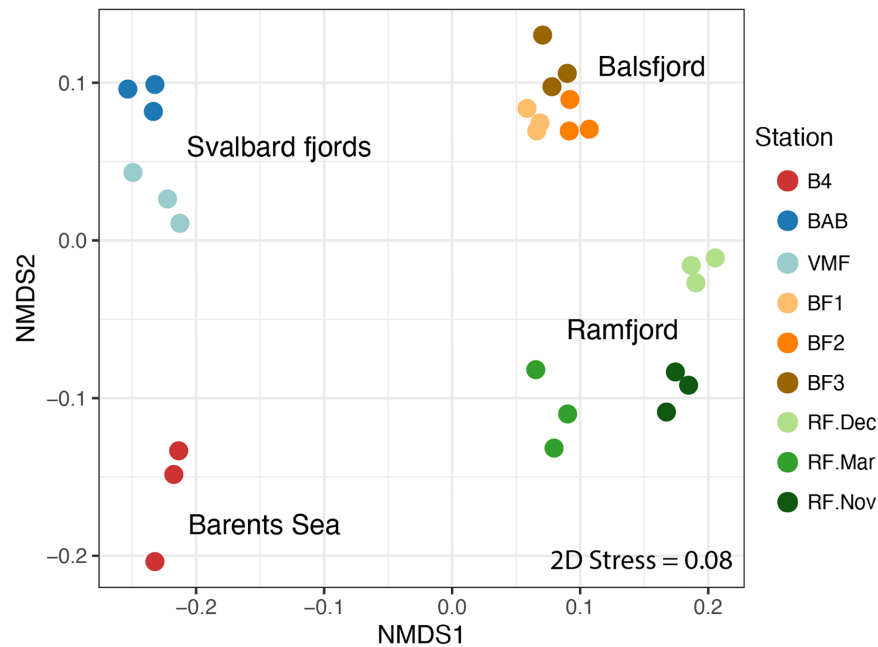


Figure 2. The nMDS ordination of fourth-root transformed sequencing data. Each symbol represents a single replicate of the sample and colours indicate stations.

with differences between stations generally greatly exceeding any within-sample variability. Despite limited sample size, the ordination also showed grouping within each location. On a broader spatial scale, the two Svalbard fjords were similar to each other while the two Norwegian Sea fjords were more divergent.

Overall biodiversity

A total of 48 unique taxa were identified visually (Supplementary Table 2). Among sequences that accounted for at least 0.01% of total reads of any sample, 477 MOTUs were identified, which corresponded to 189 unique taxonomic categories (Supplementary Table 3). In total, 243 of the MOTUs were identified to species, 22 to genus, and 27 to at least phylum level, while the remainder (188) were grouped into an unknown category. Other large groupings of MOTUs into a single taxonomic category included *Parasagitta elegans* (37 MOTUs), and unknown Calanoida (32 MOTUs). The number of unique taxa per station identified using morphological identification was 20–30; metabarcoding resulted in 80–130 unique categories per station (Figure 3a). Metabarcoding provided the largest gains in biodiversity compared to morphological identification for the meroplankton fraction, which expanded from 3–5 categories per station to 100 or more, followed by small copepods and cnidarians. Of the 48 morphologically-identified taxa, 36 were also identified via metabarcoding. The taxa that failed to appear in the metabarcoding data included the appendicularians *Oikopleura vanhoeffeni* and *Fritellaria borealis*, the pelagic polychaete *Tomopteris helgolandica*, the pteropod *Clione limacina* and several species of copepods. With the exception of *C. limacina*, all of the excluded taxa were missing from the reference databases. Among the top-ten dominant taxa by biomass and by sequence counts, five to eight overlapped between the two methods (mean 6.9 ± 1.1); of the top-10 dominant taxa by abundance four to seven overlapped (mean 5.4 ± 1.0) (Figure 3b, Supplementary Table 3). Some of the

most common species identified by morphology that consistently did not show up in the top-ten metabarcoding results include the copepods *Microsetella norvegica*, *Acartia* spp. and *Metridia* spp., while metabarcoding regularly identified among the dominant taxa Ctenophora and meroplanktonic groups, which failed to show up in the top-ten dominant lists via abundance and biomass.

Composition of biomass vs. sequence counts

The proportional contribution of each broad taxonomic group was very different among the metabarcoding, abundance, and biomass data (Figure 4). Abundance was heavily dominated by small copepods, with other groups contributing 25% or less, whereas large copepods dominated the biomass, composing 20–95% of the sample. The sequence reads were more evenly distributed between taxa, with small and large copepods, chaetognaths, and meroplankton being significant contributors. Despite the differences in absolute values, strong significant correlations ($p \leq 0.02$, Figure 5) were found between relative proportions of biomass (fourth-root transformed) and sequence reads for most taxonomic groups present in sufficient numbers (with the exception of Pteropoda). However, the ratios between sequence reads and biomass were generally not following 1:1 (Figure 5). The large copepods were consistently underrepresented in the sequencing data relative to the biomass estimates, while small copepods and meroplankton were consistently overrepresented. Abundance numbers were poorly correlated to metabarcoding data, with only euphausiids showing a significant correlation between transformed % abundance and % sequence reads ($p = 0.013$).

Species-level correlations

We observed strong correlations between estimated % biomass and % sequence reads of all taxa that were registered using both

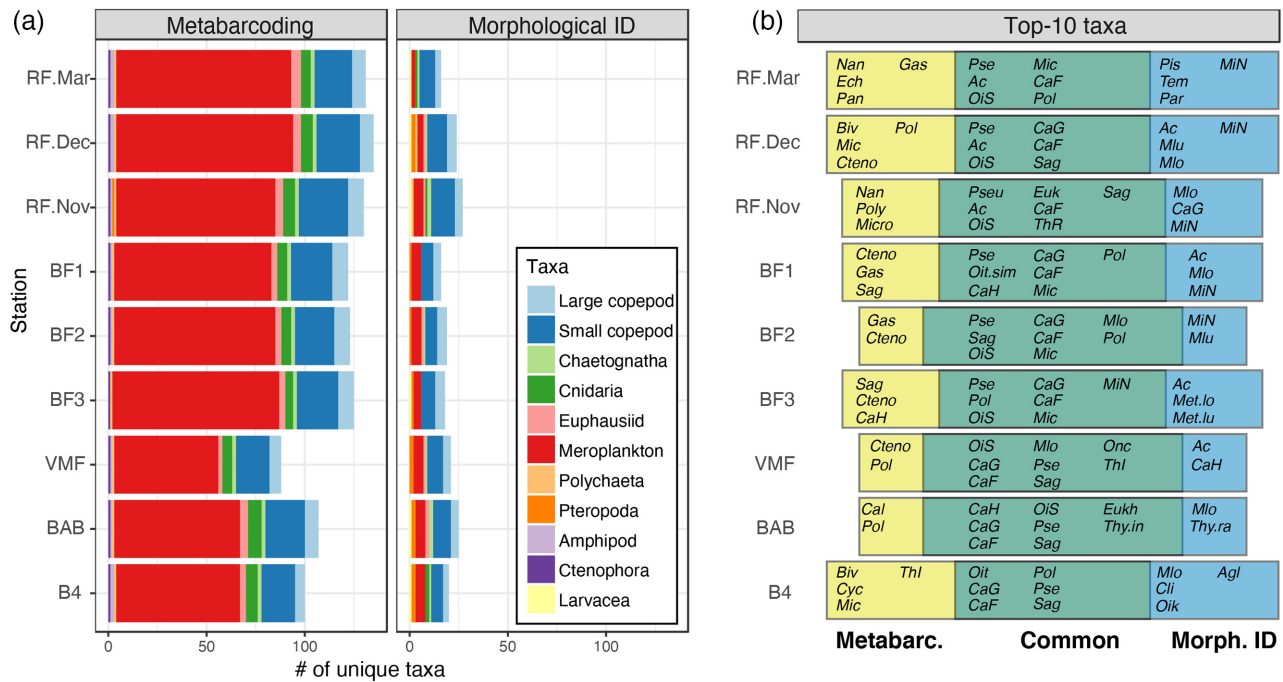


Figure 3. (a) The number of unique taxa identified using morphological identification and metabarcoding, and (b) the number of overlapping taxa among the top-ten most common taxa by metabarcoding and biomass estimation. Green colour indicates common taxa; yellow—taxa identified via metabarcoding; and blue—taxa identified using morphological ID. For species codes, see Figure 6.

methods (examined at the highest common taxonomic resolution), both within each individual sample and when all samples were pooled together ($R^2 = 0.59$, $p_{\text{adjust}} < 0.0001$, Figure 6). The correlations were strongest for copepods ($R^2 = 0.70$, $p_{\text{adjust}} < 0.0001$) and euphausiids ($R^2 = 0.78$, $p_{\text{adjust}} < 0.0001$), and weakest for meroplankton ($R^2 = 0.52$, $p_{\text{adjust}} < 0.0001$). Once again, some species/taxonomic categories were consistently overrepresented in the biomass estimate vs. the metabarcoding (i.e. *Calanus* spp.), while others were consistently underrepresented (i.e. Polychaeta larvae) in the biomass.

Case study: pseudo-cryptic taxa

We examined the relative composition of three common pseudo-cryptic copepod genera, *Calanus* spp., *Pseudocalanus* spp., and *Microcalanus* spp. (Figures 7 and 8) in the metabarcoding data. *Calanus* was among the dominant groups both in the metabarcoding and biomass data, and was represented by the Arctic *C. glacialis* and *Calanus hyperboreus* and the boreal *C. finmarchicus* and *Calanus helgolandicus*. Overall, the results of visual and molecular identification revealed remarkably similar patterns (Figure 7). *Calanus finmarchicus* dominated at most locations both in terms of estimated biomass and sequence counts, composing 35–90% of all *Calanus* DNA, followed by *C. glacialis*. *Calanus hyperboreus* was detected visually at only the three Arctic stations and one Balsfjord station in very low numbers, and was also less abundant in the sequence counts (with the exception of B34 in the Barents Sea). Surprisingly, its highest relative contribution was not in the Arctic locations, but in Ramfjord in November, where it was not detected visually. *Calanus helgolandicus* was the rarest of all *Calanus*, and was only observed via metabarcoding in Balsfjord and Ramfjord.

Although we did not identify the species visually within *Pseudocalanus* and *Microcalanus* genera, we anticipated more than one species to be present. We found *Pseudocalanus* to consist of four species: the Arctic *Pseudocalanus acuspes* and *Pseudocalanus minutus*, and boreal *Pseudocalanus moultoni* and *Pseudocalanus elongatus* (Figure 8a). In the Norwegian fjords, all four species were present, with the relative proportions changing seasonally. *Pseudocalanus acuspes* was the dominant species in all samples except the three Balsfjord stations and Ramfjord in December, where *P. moultoni* dominated. *Pseudocalanus elongatus* was observed only in the two Norwegian fjords, where its contribution was comparable with that of the other species. *Pseudocalanus minutus* was the least common species, with its highest contribution observed in the Barents Sea. Within the *Microcalanus* genus, only *M. pusillus* was observed at all examined locations.

Case study: meroplankton

Above we already showed that estimated biomass of meroplankton using visual identification was correlated with numbers of sequence reads, but the taxonomic resolution between the two methods was vastly different (Figure 3a and b). Seven categories of meroplankton were identified using microscopic sorting (Supplementary Table 1). A total of 118 benthic species belonging to 12 taxonomic groups were observed in the DNA sequences, making meroplankton the group dominating the metabarcoding dataset in terms of biodiversity. Together, meroplankton accounted for 12% of all sequence reads and 65% of the total unique taxa observed. Benthic taxa not detected using morphological identification but present in the DNA pool included members of Ascidiacea, Porifera, Cnidaria, and Caudofoveata. The most abundant groups were polychaetes, echinoderms, and bivalves, with polychaetes being the

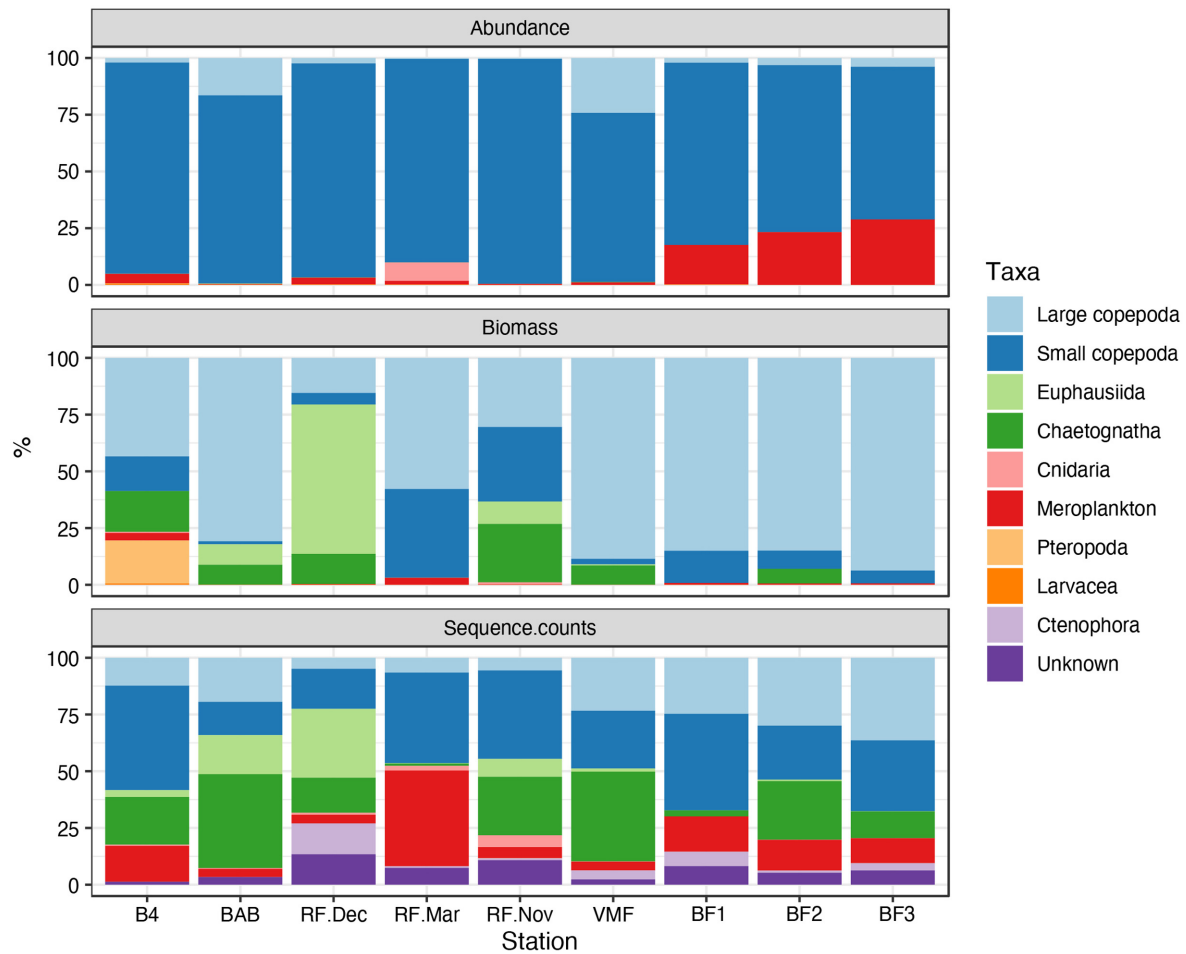


Figure 4. Relative contributions of different taxa to abundance, biomass, and sequence counts in the examined stations.

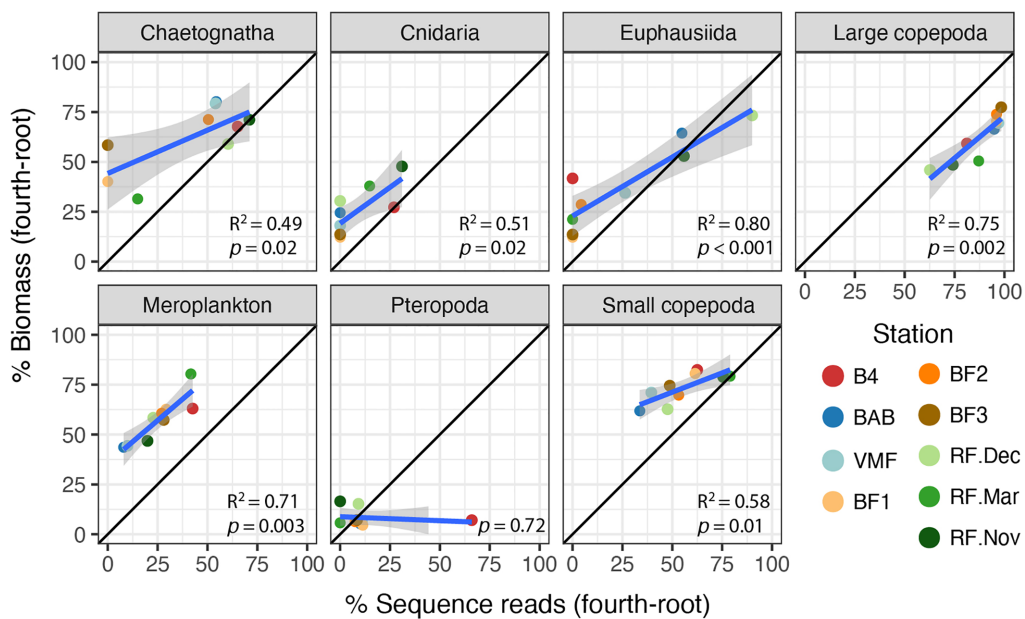


Figure 5. Linear relationships between fourth-root transformed % of sequence reads and % biomass of the different taxa in the examined stations. Black line shows a relationship of 1:1 between fourth-root transformed % of sequence reads and % biomass. Symbol colours indicate stations.

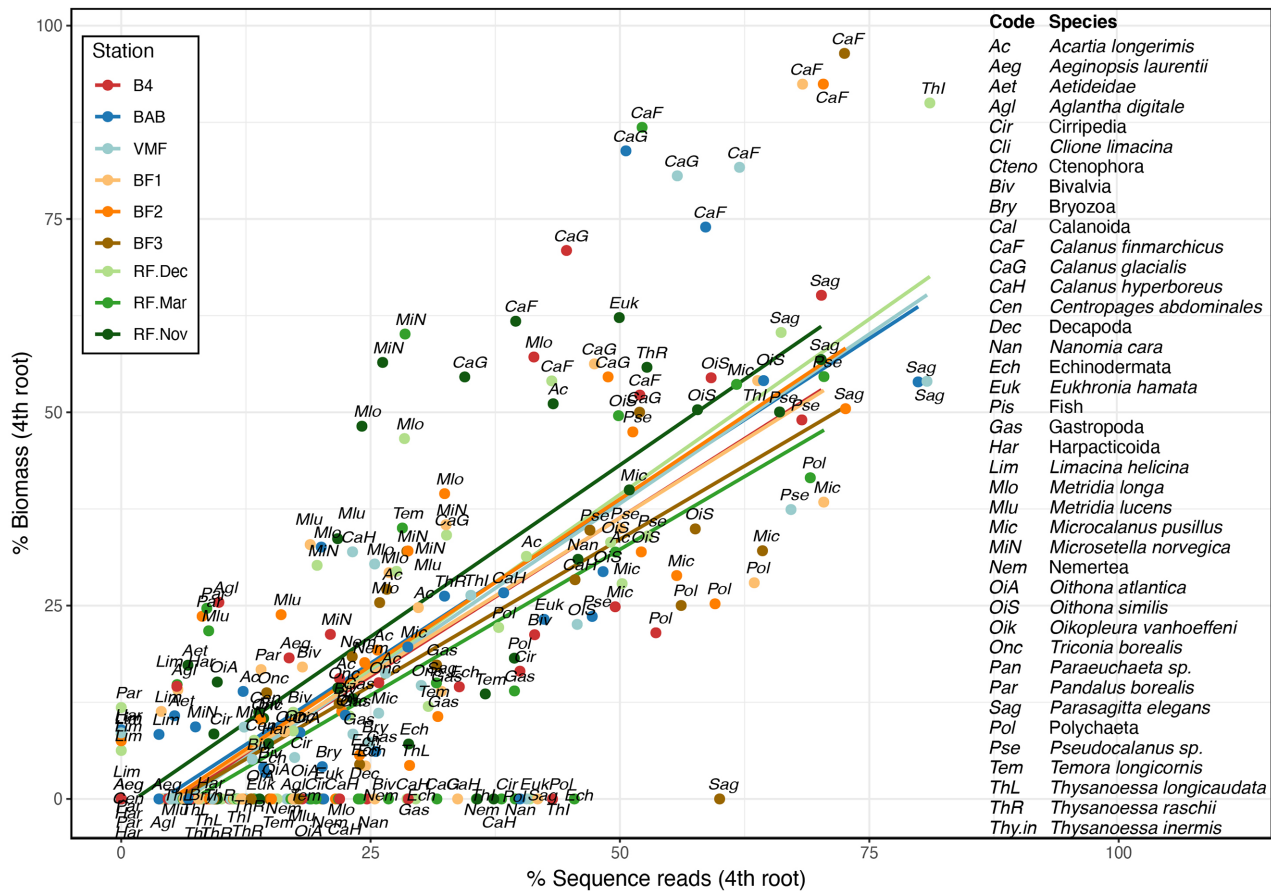


Figure 6. Relationship between the fourth-root transformed % of sequence reads and % biomass of all species/taxa that were identified using both methods of processing. Symbol colours indicate station; lines indicate best-fitting regression lines (all significant at $p < 0.01$).

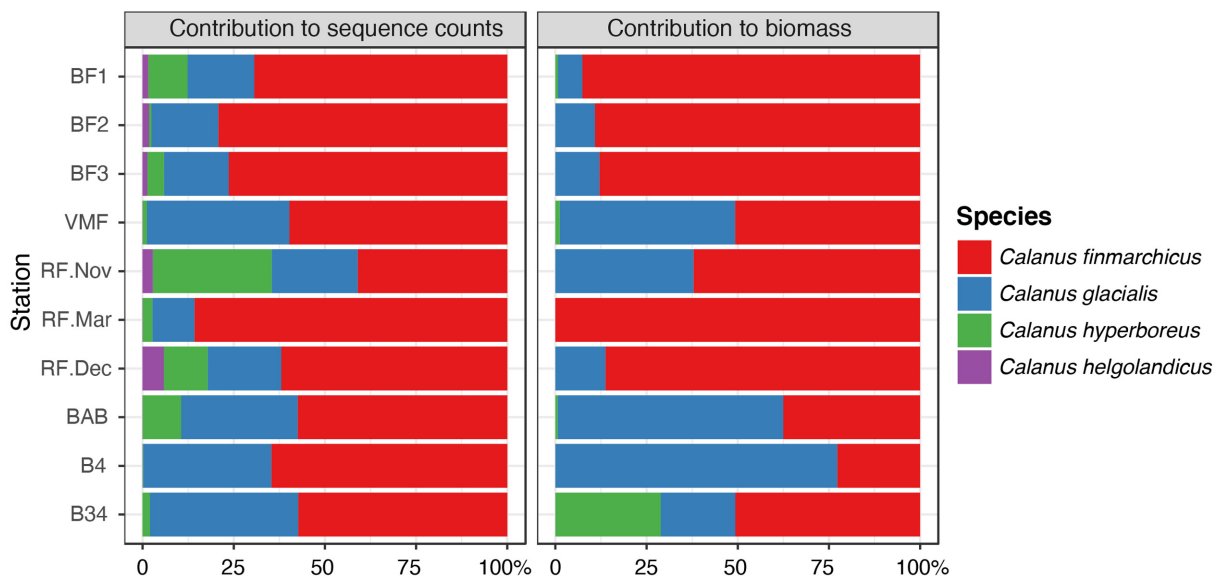


Figure 7 Relative proportions of *Calanus* species in the samples based on metabarcoding (left panel) and biomass estimates using visual identification (right panel).

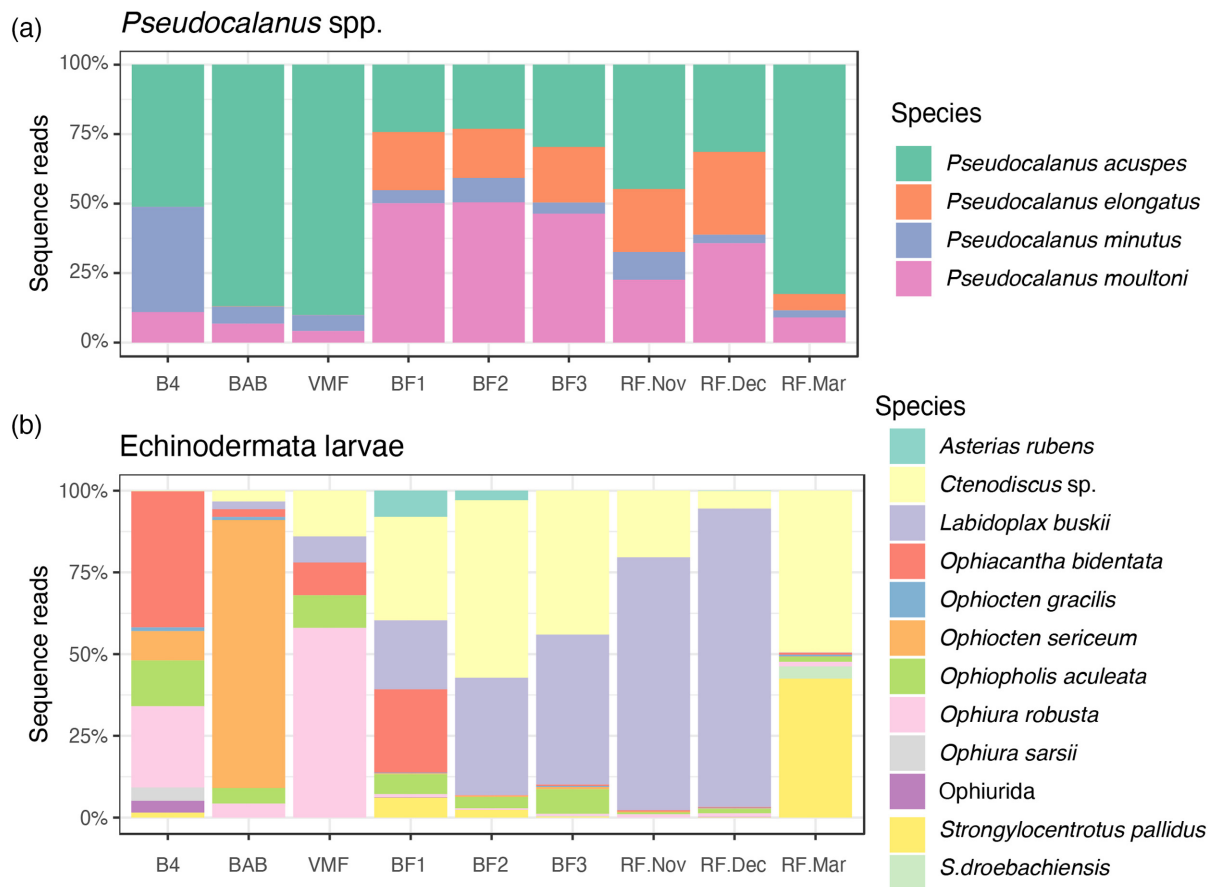


Figure 8. Proportions of DNA sequences belonging to species of (a) *Pseudocalanus* and (b) Echinodermata in the studied samples.

dominant group both in terms of diversity and sequence numbers, accounting for half of the total meroplankton species and 7.5% of total sequence reads (65% of meroplankton sequences). The highest abundance and diversity of meroplankton was observed in the Balsfjord samples. An example of the species composition of echinoderm larvae in Billefjorden were heavily dominated by *Ophiosten sericeum*, whereas this species was absent or rare in all other samples. The differences between the three Balsfjord samples (BF1–3) are notable, with *Ophiacantha bidentata* present in only one, and *Asterias rubens* in only two samples, despite the fact that these samples were collected in the same location and time. In the Ramfjorden samples (RF), the two winter samples were dominated by the holothurian *Labidoplax buskii*, but in the March sample it is replaced by *Strongylocentrotus pallidus*.

Discussion

Metabarcoding has been widely applied in ecology to investigate the biodiversity of microbial, invertebrate, and vertebrate communities, as well as food web interactions between them (Albaina *et al.*, 2016; Bucklin *et al.*, 2016; Jakubavičiute *et al.*, 2017; Zhang *et al.*, 2018; Couton *et al.*, 2019). Within the zooplankton samples examined in our study, metabarcoding yielded similar, and in many groups vastly better taxonomic resolution, than traditional microscopic sorting. Metabarcoding identified 75% of the species identified by microscopic sorting and we expect this number to continue going up as additional species are added to the sequence ref-

erence databases. Since monitoring studies often focus on the most common species, assuming them to be the most ecologically relevant, it is worth noting the 50–80% (but not 100%) overlap in the top-ten most common species in the metabarcoding and morphological approaches, both when looking at abundance and biomass (Figure 3b). It is notable that among the discrepancies between the two methods, the taxa that showed up in the top-ten of metabarcoding dataset included those that are traditionally underrepresented using morphological analysis, such as ctenophores and meroplankton.

Although metabarcoding is increasingly used as a tool in zooplankton ecology to investigate biodiversity, so far few studies have applied it as a quantitative tool in real environmental zooplankton samples. Nonetheless, such relationships have been explored previously, with most studies reporting poor to moderate correlations to zooplankton organism abundance (Elbrecht and Leese, 2015; Sun *et al.*, 2015; Bucklin *et al.*, 2019), which is not surprising as planktonic organisms span ten orders of magnitude in body size. Better correlations have been observed between sequence counts and biomass (Hirai *et al.*, 2015; Thomas *et al.*, 2016; Yang *et al.*, 2017). A meta-analysis of 22 metabarcoding studies that targeted a wide variety of organisms and genetic markers, found a significant, albeit weak, quantitative relationship between organism biomass and number of produced sequences (Lamb *et al.*, 2019). Similarly, our results show a poor relationship of DNA relative abundance to individual abundance, but strong, robust correlations between relative biomass and relative DNA read counts across a wide range of taxonomic groups. The correlations we observed were stronger,

and covered more taxa, than have been demonstrated in previous works (Lamb *et al.*, 2019; Hirai *et al.*, 2015; Thomas *et al.*, 2016; Yang *et al.*, 2017), likely due to our choice of a highly degenerate primer set. Since we used two separate subsamples for metabarcoding and microscopic sorting, and obtained biomass estimates indirectly (via length-weight regressions), it is important to note that these values also inevitably have a high error rate associated with them. This is especially true for groups where such length-weight relationships are poorly described, and a perfect relationship was not expected. Unsurprisingly, the best quantitative results were obtained for the well-studied crustaceans—copepods and euphausiids. Additionally, since our study included only nine samples, it is important to acknowledge the error rates that can be associated with such a small sample size, with a strong potential of outliers affecting the patterns in both directions. However, despite the absence of accurate regression formulas for many groups, such as benthic invertebrate larvae, and the small sample size within our study, the correlations were remarkably strong and statistically significant, suggesting that true patterns may be even stronger.

The potential quantitative value of metabarcoding has frequently been questioned (Bucklin *et al.*, 2016) due to the difficulty of finding truly universal primers, which amplify all groups equally well, as well as the interspecies variability in the initial amount of DNA. This is also true for our data, as the relationships that we observed between relative DNA counts and relative biomass for the different taxonomic groups rarely followed a 1:1 ratio, with certain groups being consistently overestimated either in the biomass or the sequence reads numbers. This was particularly visible in large copepods, which were the dominant group within the biomass, but had a ca. ten times lower contribution to the total DNA reads. This is not entirely unexpected, as large copepods from cold regions contain large lipid stores, which can make up more than 70% of the body mass (Falk-Petersen *et al.*, 2009), and are thus relatively poor in DNA. Other groups, such as polychaetes, had a consistently higher contribution in the sequencing data, most likely due to inaccurate estimation of their biomass or inadequate identification. For example, “unknown larvae” and eggs from the morphological counts may have belonged to polychaetes. Knowing the individual traits of the different species, conversion factors can be developed for different taxonomic groups, or even individual species/genera. In our samples, for example, the application of a conversion factor of 10x to the number of sequence reads of large copepods brought the ratio for this group closer to 1:1 and improved correlations for the other groups, as well as the overall contributions of the individual taxa.

Many metrics are used when performing ecosystem assessments, including overall abundance and biomass, presence/absence of certain indicator species, diversity estimates and quantitative community composition, and various wellness indices that take into account age, size, body composition, and population structure of key species (Miloslavich *et al.*, 2018; Canonico *et al.*, 2019). Although metabarcoding cannot readily provide information on abundance or size/population structure of organisms, it can be used to estimate the other metrics at a comparable, or better level of accuracy than traditional zooplankton sorting. We would like to emphasize, however, that the work of skilled taxonomists is still very much needed to supply reliable identification of specimens used in barcode reference databases. With a single bulk biomass measurement of the sample, and the application of taxa-specific conversion factors, metabarcoding data can be converted from relative to absolute biomass and serve as a proxy of organism biomass in the samples (Coguic *et al.*, 2021). For species with a relatively constant body

size, i.e. copepods, these values can be then transformed to estimate relative abundance, as is frequently done vice versa [e.g. Gluchowska *et al.* (2017) and references therein].

Although metabarcoding methods are increasingly coming into maturity, a series of challenges still remains to be solved before we can obtain their full potential. The most important present handicap for marine invertebrate communities is the presence of significant gaps in reference sequence databases, although significant efforts are being made to remedy this (Bucklin *et al.*, 2021). Our samples were collected in a relatively homogenous, low diversity environment (Atlantic Arctic and sub-Arctic), with relatively good reference databases (Bluhm *et al.*, 2011). Therefore, these results should be tested with communities from other parts of the world with higher species diversity and less complete databases. We chose a clustering-based method of taxonomic assignment as opposed to a pipeline based on sequence variants, since the latter are not recommended for COI (or any other similarly hypervariable marker), due to widespread intra-species variability leading to millions of sequence variants. However, the choice of the clustering algorithm to delineate MOTUs presents an additional challenge, as it can vastly impact the resulting diversity. In our pipelines, we have used Swarm v.2.1.13 (Mahé *et al.*, 2015), a step-by-step clustering algorithm, which avoids the use of arbitrary similarity thresholds. Moreover, we used taxonomic clustering when measuring diversity, which only counted uniquely identified taxonomic categories (e.g. all unknown Calanoid MOTUs were collapsed into a single group). Taxonomic clusters thus obtained are highly dependent on the completeness of the reference database, so the richness values could increase in the future when existing reference gaps are progressively filled. The 97% similarity threshold we applied for species assignments, while the accepted standard used by BOLD and applied in most ecological studies, can be significantly higher or lower in different groups of metazoans, and could also result in an under- or an over-estimation of diversity. Richness values obtained from COI metabarcoding can also be significantly overestimated by the presence of nuclear pseudogenes (nuclear mitochondrial DNA, or NUMTs) (Song *et al.*, 2008). Our pipeline included a curation step using the LULU algorithm, which allowed us to minimize the impact of these pseudogenes (Frøslev *et al.*, 2017). However, the presence of these sequences in marine invertebrate genomes are poorly understood. Future studies involving analyses of whole genomes of marine metazoans would help to understand the impact of this issue in molecular biodiversity metrics. Finally, although the Leray-XT primers are nearly universal for metazoans, and the metabarcoding datasets resulting from their use undoubtedly show enhanced quantitative value compared to others obtained using less-degenerate primer sets, it still fails to amplify some planktonic organisms, which needs to be taken into account. For example, within our study pteropods and mollusc larvae were very poorly represented in the sequencing data, which could be a fault of either the DNA extraction step, or the PCR. Molluscs in general are notoriously difficult to extract and amplify DNA from (Pereira *et al.*, 2011). Alternatively, the Leray-XT primer set may not provide optimal binding conditions for this group. Further studies including mock samples should test the extent of non-detectable planktonic taxa by this primer set. If these organisms are of particular interest, then other more specific markers can be used in conjunction with the Leray-XT set.

Despite the aforementioned limitations, our metabarcoding approach can prove invaluable to monitor cryptic/pseudocryptic species or larval communities. This is a common issue within zoo-

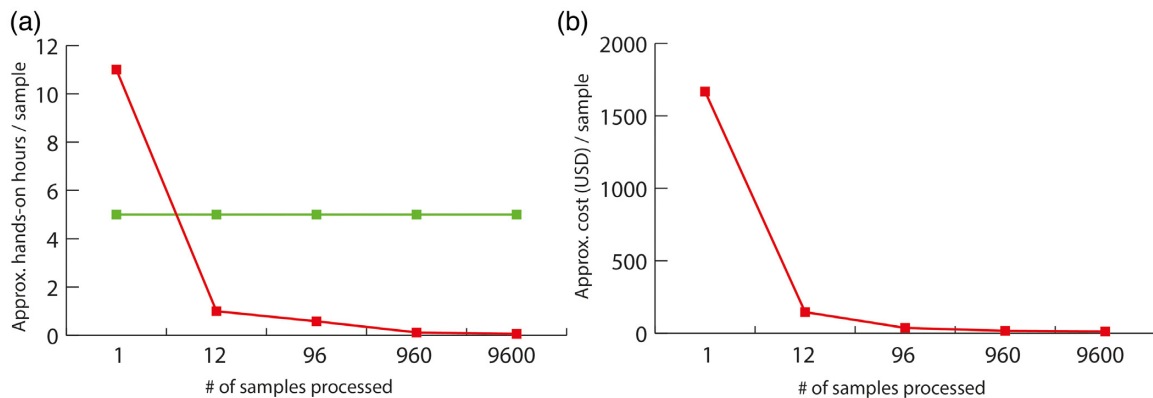


Figure 9. (a) Approximate hand-on time (hours) per sample with increasing number of samples processed via metabarcoding (red) and microscope sorting (green); (b) approximate cost per sample (USD) at time of publishing with increasing number of samples (assuming outsourced library preparation and sequencing on an Illumina MiSeq).

plankton studies, since several closely related species, sharing a very similar morphology, may be present in the water column simultaneously, and very few studies attempt to distinguish between them. Examples include complexes of the copepod “species” *Oithona similis* (Cornils *et al.*, 2017), *Acartia tonsa* (Drillet *et al.*, 2008), *Eurytemora affinis* (Devreker *et al.*, 2012) or the jellyfish *Cyanea capillata* (Kolbasova *et al.*, 2015). Despite their morphological similarity, different members of such species complexes may play different roles in the ecosystem, and often have different, slightly overlapping geographical ranges, the exact extents of which were until now difficult to parse out. Within our studied zooplankton communities, examples of such pseudo-cryptic genera included *Calanus* (Choquet *et al.*, 2017) and *Pseudocalanus* spp. (Frost, 1989; Ershova *et al.*, 2017), which were each represented by four species in our area of study. Detection of these species in our study regions is in agreement with other studies that have examined the species composition of these groups using different molecular markers (Choquet *et al.*, 2018; Coguiec *et al.*, 2021). Surprisingly, only one species of *Microcalanus*, *M. pusillus*, was observed at all examined locations, although *M. pygmaeus* has also been previously reported in these regions (Klekowski and Węśławski, 1990; Walkusz *et al.*, 2009). Similarly, meroplanktonic larvae are most frequently ignored in plankton studies, or are grouped into broad taxonomic categories. Within our study, the metabarcoding protocol identified DNA of over 100 benthic taxa, most of which presumably belong to planktonic larvae. Monitoring the timing when certain species are present in the water column can give important insight on the life cycles of benthic animals, as well as track the potential expansion of introduced bottom-dwelling species into the region (Couton *et al.*, 2019; Ershova *et al.*, 2019). Another clear advantage of metabarcoding is that the use of hypervariable markers, such as COI, has the potential to provide intra-species diversity information (population genetics) for multiple species at the same time (Turon *et al.*, 2020), which can be applied to detect subtle changes in the composition of planktonic communities over both time and space.

Although molecular approaches have always been associated with a high cost, they are rapidly becoming more affordable. With metabarcoding in particular, an increasing number of simultaneously processed samples will reduce the cost dramatically (Figure 9a; Supplementary Table 4). Similarly, the time required for process-

ing per sample will be reduced substantially when using a metabarcoding approach, in contrast with traditional sorting, which follows a linear progression of effort per sample (Figure 9b; Supplementary Table 4). For this reason, we suggest that this protocol can serve as a useful alternative to traditional sorting for monitoring studies, which generally involve hundreds of samples and require thousands of technician hours for processing. As zooplankton is inherently patchy, this would also allow for increased replication (i.e. processing more than one sample from the same location), the lack of which is a serious, but generally overlooked issue in zooplankton ecology (Skjoldal *et al.*, 2013). The variable nature of zooplankton was apparent already in our results, where three net tows taken during the same day in the same location produced similar, but far from identical results, with several species showing up in one or two, but not all three samples. The development of sequencing technology and advances in metabarcoding approaches makes the protocol developed herein scalable and affordable. The protocol uses metabarcoding primers with unique eight-base sample-tags attached to each primer pair (Wangenstein *et al.*, 2018), and 96 such primer pairs allow multiplexing 96 samples in a single library. A library-tag can subsequently be ligated to each library, allowing for up-scaling for simultaneous sequencing of 100–1000’s of samples in a single sequencing run. For example, a project that wants to analyse 1440 samples using our protocol, will have to extract the samples in 15 96-well plates, run PCRs in 15 plates, pool the samples in each plate, purify each pool, and prepare 15 libraries by ligating Illumina sequencing adaptors and library-tags on each pool. These 15 libraries can then be pooled into a single tube and sequenced on an Illumina Novaseq using 250 bp paired-end chemistry. With an output of 400 million paired-end reads, the Novaseq is cheaper, and provides more sequencing depth per sample compared an Illumina MiSeq. These reductions in costs and time consumption will only increase within the next few years for the metabarcoding approach, as sequencing run time keeps reducing and semi-automated and fully automated pipelines are starting to emerge for markers like COI. Similarly, we expect reference databases to vastly improve in their coverage in the coming years to allow for maximum taxonomic accuracy in the sequencing data. Therefore, by reducing processing time, metabarcoding will allow analysis of more samples than was possible before with prospects for improved replication and/or spatiotemporal resolution.

Conclusions

Our results highlight the quantitative value of a metabarcoding approach using the highly degenerate Leray-XT primer set, which provides improved taxonomic resolution, universal coverage for metazoans, reduced primer bias, and availability of a comprehensive reference database, to investigate zooplankton biodiversity, community composition, relative species biomass, and the presence of rare, poorly preserved, or cryptic species. We suggest that relative sequence counts can be used to estimate individual organism biomass via a single bulk biomass measurement, particularly within a single study, which will facilitate monitoring of spatial, seasonal, and inter-annual dynamics of zooplankton communities. Our protocol can allow rapid and relatively inexpensive processing of hundreds to thousands of samples at a higher taxonomic resolution than traditional zooplankton sorting, and can be widely applied for monitoring or ecological studies.

Data availability statement

The data underlying this article are available in the online supplementary material. The raw sequencing data and the MOTU tables are publicly available at Mendeley Data (Dataset reference 5p37bgckjf. DOI: 10.17632/5p37bgckjf.1).

Funding

This research has been funded by UiT The Arctic University of Norway and the Tromsø Research Foundation under the project “Arctic Seasonal Ice Zone Ecology,” project number 01vm/h1, and partially supported by RFBR grant 19–04–00955. This study was done within the framework of state assignment of IO RAS (theme no. 0128–2021–0007)

Supplementary Data

Supplementary material is available at the ICES/JMS online version of the manuscript.

Acknowledgements

We thank the two anonymous reviewers whose valuable feedback helped improve the manuscript.

References

- Albaina, A., Aguirre, M., Abad, D., Santos, M., and Estonba, A. 2016. 18S rRNA V9 metabarcoding for diet characterization: a critical evaluation with two sympatric zooplanktivorous fish species. *Ecology and Evolution*, 6: 1809–1824.
- Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D. W., and Vogler, A. P. 2018. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, 27: 146–166.
- Antich, A., Palacín, C., Cebrian, E., Golo, R., Wangensteen, O. S., and Turon, X. 2020. Marine biomonitoring with eDNA: can metabarcoding of water samples cut it as a tool for surveying benthic communities? *Molecular Ecology*, 30: 3175–3188.
- Atienza, S., Guardiola, M., Præbel, K., Antich, A., Turon, X., and Wangensteen, O. S. 2020. DNA metabarcoding of deep-sea sediment communities using COI: community assessment, spatio-temporal patterns and comparison with 18S rDNA. *Diversity*, 12: 123.
- Bakker, J., Wangensteen, O. S., Baillie, C., Buddo, D., Chapman, D. D., Gallagher, A. J., Guttridge, T. L. et al. 2019. Biodiversity assessment

- of tropical shelf eukaryotic communities via pelagic eDNA metabarcoding. *Ecology and Evolution*, 9: 14341–14355.
- Beaugrand, G., Ibañez, F., Lindley, J. A., and Reid, P. C. 2002. Diversity of calanoid copepods in the North Atlantic and adjacent seas: species associations and biogeography. *Marine Ecology Progress Series*, 232: 179–195.
- Beaugrand, G., Luczak, C., and Edwards, M. 2009. Rapid biogeographical plankton shifts in the North Atlantic Ocean. *Global Change Biology*, 15: 1790–1803.
- Bluhm, B. A., Gebruk, A. V., Gradinger, R., Hopcroft, R. R., Huettmann, F., Kosobokova, K. N., Sirenko, B. I. et al. 2011. Arctic marine biodiversity. *Oceanography*, 24: 232–248.
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., and Coissac, E. 2016. Obitools: a Unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16: 176–182.
- Bucklin, A., Lindeque, P. K., Rodriguez-Ezpeleta, N., Albaina, A., and Lehtiniemi, M. 2016. Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research*, 38: 393–400.
- Bucklin, A., Peijnenburg, K. T. C. A., Kosobokova, K. N., O'Brien, T. D., Blanco-Bercial, L., Cornils, A., Falkenhaus, T. et al. 2021. Toward a global reference database of COI barcodes for marine zooplankton. *Marine Biology*, 168: 1–26.
- Bucklin, A., Yeh, H. D., Questel, J. M., Richardson, D. E., Reese, B., Copley, N. J., and Wiebe, P. H. 2019. Time-series metabarcoding analysis of zooplankton diversity of the NW Atlantic continental shelf. *ICES Journal of Marine Science*, 76: 1162–1176.
- Burridge, A. K., Goetze, E., Raes, N., Huisman, J., and Peijnenburg, K. T. C. A. 2015. Global biogeography and evolution of *Cuvierina* pteropods: phylogenetics and phylogeography. *BMC Evolutionary Biology*, 15: 1–16.
- Canonic, G., Buttigieg, P. L., Montes, E., Muller-Karger, F. E., Stepien, C., Wright, D., Benson, A. et al. 2019. Global observational needs and resources for marine biodiversity. *Frontiers in Marine Science*, 6: 367.
- Chiba, S., Batten, S., Martin, C. S., Ivory, S., Miloslavich, P., and Weatherdon, L. V. 2018. Zooplankton monitoring to contribute towards addressing global biodiversity conservation challenges. *Journal of Plankton Research*, 40: 509–518.
- Choquet, M., Hatlebakk, M., Dhanasiri, A. K. S., Kosobokova, K. N., Smolina, I., Søreide, J. E., Svensen, C. et al. 2017. Genetics redraws pelagic biogeography of *Calanus*. *Biology Letters*, 13: 20170588.
- Choquet, M., Kosobokova, K. N., Kwaśniewski, S., Hatlebakk, M., Dhanasiri, A. K. S., Melle, W., Daase, M. et al. 2018. Can morphology reliably distinguish between the copepods *Calanus finmarchicus* and *C. glacialis*, or is DNA the only way? *Limnology and Oceanography: Methods*, 16: 237–252.
- Clarke, L. J., Beard, J. M., Swadling, K. M., and Deagle, B. E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, 7: 873–883.
- Coguiac, E., Ershova, E., Daase, M., Vonnahme, T., Wangensteen, O. S., Gradinger, R., Præbel, K. et al. 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: 1049.
- Cornils, A., Wend-Heckmann, B., and Held, C. 2017. Global phylogeography of *Oithona similis* s.l. (Crustacea, Copepoda, Oithonidae) – a cosmopolitan plankton species or a complex of cryptic lineages? *Molecular Phylogenetics and Evolution*, 107: 473–485.
- Couton, M., Comtet, T., Le Cam, S., Corre, E., and Viard, F. 2019. Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions*, 10: 657–689.
- Devreker, D., Pierson, J. J., Souissi, S., Kimmel, D. G., and Roman, M. R. 2012. An experimental approach to estimate egg production and development rate of the calanoid copepod *Eurytemora affinis* in Chesapeake Bay. *Journal of Experimental Marine Biology and Ecology*, 416–417: 72–83.
- Drillet, G., Jepsen, P. M., Højgaard, J. K., Jørgensen, N. O. G., and Hansen, B. W. 2008. Strain-specific vital rates in four *Acartia tonsa*

- cultures II: life history traits and biochemical contents of eggs and adults. *Aquaculture*, 279: 47–54.
- Edgar, R., Haas, B. J., Clemente, J. C., and Quince, C. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27: 2194–2200.
- Elbrecht, V., and 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: e0130324.
- Ershova, E. A. 2020. Routine identification of four sympatric species of Calanoid copepods *Pseudocalanus* spp. in the Atlantic Arctic using a species-specific polymerase chain reaction. *Journal of Oceanological Research*, 48: 62–72.
- Ershova, E. A., Descoteaux, R., Wangenstein, O. S., Iken, K., Hopcroft, R. R., Smoot, C., Grebmeier, J. M. *et al.* 2019. Diversity and distribution of meroplanktonic larvae in the Pacific Arctic and connectivity with adult benthic invertebrate communities. *Frontiers in Marine Science*, 6: 490.
- Ershova, E. A., Nyeggen, M. U., Yurikova, D. A., and Søreide, J. E. 2021. Seasonal dynamics and life histories of three sympatric species of *Pseudocalanus* in two Svalbard fjords. *Journal of Plankton Research*, 43: 209–223.
- Ershova, E. A., Questel, J. M., Kosobokova, K. N., and Hopcroft, R. R. 2017. Population structure and production of four sibling species of *Pseudocalanus* spp. in the Chukchi Sea. *Journal of Plankton Research*, 39: 48–64.
- Ershova, R. A., Hopcroft, R. R., and Kosobokova, K. N. 2015. Inter-annual variability of summer mesozooplankton communities of the western Chukchi Sea: 2004–2012. *Polar Biology*, 38: 1461–1481.
- Falk-Petersen, S., Mayzaud, P., Kattner, G., and Sargent, J. R. 2009. Lipids and life strategy of Arctic *Calanus*. *Marine Biology Research*, 5: 18–39.
- Frøsvlev, T. G., Kjølner, R., Bruun, H. H., Ejrnes, R., Brunbjerg, A. K., Pietroni, C., and Hansen, A. J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, 8:1–11.
- Frost, B. W. 1989. A taxonomy of the marine calanoid copepod genus *Pseudocalanus*. *Canadian Journal of Zoology*, 67: 525–551.
- Geller, J., Meyer, C., Parker, M., and Hawk, H. 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, 13: 851–861.
- Gluchowska, M., Dalpadado, P., Beszczynska-Möller, A., Olszewska, A., Ingvaldsen, R. B., and Kwasniewski, S. 2017. Interannual zooplankton variability in the main pathways of the Atlantic water flow into the Arctic Ocean (Fram Strait and Barents Sea branches). *ICES Journal of Marine Science*, 74: 1921–1936.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270: 313–321.
- Hirai, J., Kuriyama, M., Ichikawa, T., Hidaka, K., and Tsuda, A. 2015. A metagenetic approach for revealing community structure of marine planktonic copepods. *Molecular Ecology Resources*, 15: 68–80.
- Hopcroft, R. R., Kosobokova, K. N., and Pinchuk, A. I. 2010. Zooplankton community patterns in the Chukchi Sea during summer 2004. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57: 27–39.
- Jakubavičiute, E., Bergström, U., Eklöf, J. S., Haenel, Q., and Bourlat, S. J. 2017. DNA metabarcoding reveals diverse diet of the three-spined stickleback in a coastal ecosystem. *Plos One*, 12: e0186929.
- Klekowski, R. Z., and Węślawski, J. M., 1990. Atlas of the Marine Fauna of Southern Spitsbergen: Part 1 & 2. Invertebrates. Ossolineum, Wrocław.
- Klunder, L., de Stigter, H., Lavaleye, M. S. S., van Bleijswijk, J. D. L., van der Veer, H. W., Reichart, G. J., and Duineveld, G. C. A. 2020. A molecular approach to explore the background benthic fauna around a hydrothermal vent and their larvae: implications for future mining of deep-sea SMS deposits. *Frontiers in Marine Science*, 7: 134.
- Kolbasova, G. D., Zalevsky, A. O., Gafurov, A. R., Gusev, P. O., Ezhova, M. A., Zheludkevich, A. A., Konovalova, O. P. *et al.* 2015. A new species of *Cyanea* jellyfish sympatric to *C. capillata* in the White Sea. *Polar Biology*, 38: 1439–1451.
- Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., and Taylor, M. I. 2019. How quantitative is metabarcoding: a meta-analytical approach. *Molecular Ecology*, 28: 420–430.
- Leray, M., and Knowlton, N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 112: 2076–2081.
- Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3: e1420.
- Marlétaz, F., Le Parco, Y., Liu, S., and Peijnenburg, K. T. C. A. 2017. Extreme mitogenomic variation in natural populations of chaetognaths. *Genome Biology and Evolution*, 9: 1374–1384.
- Miloslavich, P., Bax, N. J., Simmons, S. E., Klein, E., Appeltans, W., Aburto-Oropeza, O., Andersen Garcia, M. *et al.* 2018. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology*, 24: 2416–2433.
- Miyamoto, H., Machida, R. J., and Nishida, S. 2010. Genetic diversity and cryptic speciation of the deep sea chaetognath *Caecosagitta macrocephala* (Fowler, 1904). *Deep Sea Research Part II: Topical Studies in Oceanography*, 57: 2211–2219.
- Mohrbeck, I., Raupach, M. J., Arbizu, P. M., Knebelberger, T., and Laakmann, S. 2015. High-throughput sequencing—the key to rapid biodiversity assessment of marine metazoa? *Plos One*, 10: e0140342.
- Oksanen, J., Kindt, R., Pierre, L., O’Hara, B., Simpson, G. L., Solymos, P., Stevens, M. H. H. *et al.* 2016. vegan: community ecology package, R package version 2.4-0. R package version 2.2-1. <http://vegan.r-forge.r-project.org>. Last accessed July 30 2021.
- Orlova, E. L., Ushakov, N. G., Nesterova, V. N., and Boitsov, V. D. 2002. Food supply and feeding of capelin (*Mallotus villosus*) of different size in the central latitudinal zone of the Barents Sea during intermediate and warm years. *ICES Journal of Marine Science*, 59: 968–975.
- Pereira, J., Chaves, R., Bastos, E., Leitão, A., and Guedes-Pinto, H. 2011. *International Journal of Molecular Sciences*, 12: 8086–8095.
- Questel, J. M., Hopcroft, R. R., DeHart, H. M., Smoot, C. A., Kosobokova, K. N., and Bucklin, A. 2021. Metabarcoding of zooplankton diversity within the Chukchi Borderland, Arctic Ocean: improved resolution from multi-gene markers and region-specific DNA databases. *Marine Biodiversity*, 51: 1–19.
- R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. <http://www.r-project.org> Last accessed July 30 2021.
- Richardson, A. J. 2008. In hot water: zooplankton and climate change. *ICES Journal of Marine Science*, 65: 279–295.
- Rodríguez-Ezpeleta, N., Zinger, L., Kinziger, A., Bik, H. M., Bonin, A., Coissac, E., Emerson, B. C. *et al.* 2021. Biodiversity monitoring using environmental DNA. *Molecular Ecology*, 21: 1405–1409.
- Roff, J. C., and Hopcroft, R. R. 1986. High precision microcomputer based measuring system for ecological research. *Canadian Journal of Fisheries and Aquatic Sciences*, 43: 2044–2048.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4: e2584.
- Santoferrara, L. F. 2019. Current practice in plankton metabarcoding: optimization and error management. *Journal of Plankton Research*, 41: 571–582.
- Sepulveda, A. J., Nelson, N. M., Jerde, C. L., and Luikart, G. 2020. Are environmental DNA methods ready for aquatic invasive species management? *Trends in Ecology & Evolution*, 35: 668–678.
- Siegenthaler, A., Wangenstein, O. S., Soto, A. Z., Benvenuto, C., Corrigan, L., and Mariani, S. 2019. Metabarcoding of shrimp stomach

- content: harnessing a natural sampler for fish biodiversity monitoring. *Molecular Ecology Resources*, 19: 206–220.
- Skjoldal, H. R., Wiebe, P. H., Postel, L., Knutsen, T., Kaartvedt, S., and Sameoto, D. D. 2013. Intercomparison of zooplankton (net) sampling systems: results from the ICES/GLOBEC sea-going workshop. *Progress in Oceanography*, 108: 1–42.
- Song, H., Buhay, J. E., Whiting, M. F., and Crandall, K. A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, 105: 13486–13491.
- Sun, C., Zhao, Y., Li, H., Dong, Y., MacIsaac, H. J., and Zhan, A. 2015. Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. *Aquatic Biology*, 24: 9–15.
- Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., and Trites, A. W. 2016. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, 16: 714–726.
- Toresen, R., Skjoldal, H. R., Vikebø, F., and Martinussen, M. B. 2019. Sudden change in long-term ocean climate fluctuations corresponds with ecosystem alterations and reduced recruitment in Norwegian spring-spawning herring (*Clupea harengus*, Clupeidae). *Fish and Fisheries*, 20: 686–696.
- Turon, X., Antich, A., Palacín, C., Præbel, K., and Wangensteen, O. S. 2020. From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, 30: e02036.
- Unstad, K. H., and Tande, K. S. 1991. Depth distribution of *Calanus finmarchicus* and *C. glacialis* in relation to environmental conditions in the Barents Sea. *Polar Research*, 10: 409–420.
- van der Loos, L. M., and Nijland, R. 2020. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*, 0: 1–19.
- Walkusz, W., Kwasniewski, S., Falk-Petersen, S., Hop, H., Tverberg, V., Wiczorek, P., and Weslawski, J. M. 2009. Seasonal and spatial changes in the zooplankton community of Kongsfjorden, Svalbard. *Polar Research*, 28: 254–281.
- Wangensteen, O. S., Palacín, C., Guardiola, M., and Turon, X. 2018. DNA metabarcoding of littoral hardbottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*, 6: e4705.
- Yang, J., Zhang, X., Xie, Y., Song, C., Zhang, Y., Yu, H., and Burton, G. A. 2017. Zooplankton community profiling in a eutrophic freshwater ecosystem-lake tai basin by DNA metabarcoding. *Scientific Reports*, 7: 1773.
- Zhang, G. K., Chain, F. J. J., Abbott, C. L., and Cristescu, M. E. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary Applications*, 11: 1901–1914.

Handling Editor: Ian Bradbury