Quantification of marine benthic communities with metabarcoding

Lise Klunder\textsuperscript{1,2} | Judith D. L. van Bleijswijk\textsuperscript{3} | Loran Kleine Schaars\textsuperscript{1} | Henk W. van der Veer\textsuperscript{1} | Pieternella C. Luttikhuizen\textsuperscript{1} | Allert I. Bijleveld\textsuperscript{1}

\textsuperscript{1}Department of Coastal Systems, NIOZ Royal Netherlands Institute for Sea Research, AB Den Burg Texel, The Netherlands
\textsuperscript{2}Marine Evolution and Conservation, Groningen Institute of Life Sciences, University of Groningen, CC Groningen, The Netherlands
\textsuperscript{3}Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, AB Den Burg Texel, The Netherlands

Correspondence
Lise Klunder, Department of Coastal Systems, NIOZ Royal Netherlands Institute for Sea Research, AB Den Burg Texel, The Netherlands.
Email: lise.klunder@nioz.nl

Abstract
DNA metabarcoding methods have been implemented in studies aimed at detecting and quantifying marine benthic biodiversity. In such surveys, universal barcodes are amplified and sequenced from environmental DNA. To quantify biodiversity with DNA metabarcoding, a relation between the number of DNA sequences of a species and its biomass and/or the abundance is required. However, this relationship is complicated by many factors, and it is often unknown. In this study, we validate estimates of biomass and abundance from molecular approaches with those from the traditional morphological approach. Abundance and biomass were quantified from 126 samples of benthic intertidal mudflat using traditional morphological approaches and compared with frequency of occurrence and relative read abundance estimates from a molecular approach. A relationship between biomass and relative read abundance was found for two widely dispersed annelid taxa (\textit{Pygospio} and \textit{Scoloplos}). None of the other taxons, however, showed such a relationship. We discuss how quantification of abundance and biomass using molecular approaches are hampered by the ecology of DNA i.e. all the processes that determine the amount of DNA in the environment, including the ecology of the benthic species as well as the compositional nature of sequencing data.

KEYWORDS
abundance, biomass, eDNA, metabarcoding, next-generation sequencing, quantification

1 | INTRODUCTION

Marine benthic ecosystems are characterized by high biodiversity and are of global importance to climate, nutrient cycling and primary and secondary productivity (Austen et al., 2002; Covich et al., 2004; Snelgrove, 1997). From the intertidal to the deep sea, benthic fauna are central to the maintenance of ecosystem services, whereby a high diversity is thought to maintain a positive interaction among species and promoting stability and resistance to ecosystem functioning (Danovaro et al., 2008; Leduc and Pilditch, 2013; Levin et al., 2001).

Effective monitoring of the benthic fauna is a first crucial step towards conservation of the marine benthic ecosystems (Patricio et al., 2009). Traditionally, benthic biodiversity assessments are based on the morphological identification of species (see for instance Compton et al., 2013; Diaz et al., 2004). These morphological...
inventories are time-consuming, require taxonomic expertise which is scarce and are often limited to macrofauna species (Beukema & Dekker, 2020; Bucklin et al., 2011; Cardoso et al., 2011; Cowart et al., 2015). Thus, there is a need for methods that can assess marine benthic biodiversity in a rapid and cost-effective yet detailed and accurate manner (Aylagas et al., 2018).

In recent years, DNA metabarcoding methods have been successfully implemented in various studies to assess marine benthic biodiversity. For instance, Chariton et al. (2015) used this approach to assess the ecological condition of estuaries, whereas Lanzén et al. (2016) was able to identify effects of offshore oil-drilling activities using eukaryotic metabarcoding. DNA metabarcoding provides the opportunity to assess the benthic community in a replicable manner that allows for the simultaneous recovery of a wide variety of taxa from all size classes without first isolating any organisms, facilitating rapid biodiversity monitoring (Taberlet, Coissac, et al., 2012). DNA metabarcoding most often relies on the extraction of DNA from a matrix of choice - sediment, water, air or a mixture of organisms - by, for example, the shedding rates of DNA by the source organism. Specifically, shedding rates depend on morphological and physiological characteristics or seasonal patterns and can increase up to 100-fold at certain times (Barnes & Turner, 2016; Harrison et al., 2019). Furthermore, the variability of DNA per gram tissue varies due to variable tissue cell density (Pompanon et al., 2012) or the transport of eDNA through the environment (Kelly et al., 2018). Lastly, there is no true independence between the quantitative measurements of species within an eDNA data set due to the compositional base of sequencing data. This compositional base is due to the maximum limit of reads which can be translated during sequencing, and leads to a negative correlation bias between species abundancies (Gloor et al., 2017). The frequency of occurrence approach is only based on detections and not on the compositional data set, and, this is hypothetically not under influence of the negative correlation bias.

To summarize, metabarcoding studies can aid biodiversity assessments. However, the quantitative abilities of such studies are unknown and questionable. The aim of this study is to explore to what extent traditional and metabarcoding approaches align in assessing abundances and biomass of marine macrofauna communities. Specifically, whether metabarcoding outputs from sediment samples can be used as a quantitative estimate for abundance and biomass of benthic macrofauna in the intertidal Dutch Wadden Sea. For this study, we chose to use a fast and easy sampling approach for the collection of eDNA samples in which environmental DNA is extracted from small sediment cores (Klunder et al., 2019); an approach which can readily be adopted in future monitoring programs. Simultaneously, we collected morphological-based data to estimate species occurrence and abundance. The first aim of the study was to examine whether detection rates for benthic macrofauna species were comparable between molecular and traditional analyses. Following that, we compared the morphological approach with the DNA metabarcoding approach to test the reliability of both a frequency of occurrence approach and relative read abundance approach for estimating abundances and biomass.
2 | MATERIALS AND METHODS

2.1 | Sampling

Divided over seven sampling events between June 2016 and July 2017, a total of 126 locations were sampled on tidal flats north-east of the isle of Texel in the western part of the Dutch Wadden Sea (N53°06' E4°54'). Each sampling event comprised 18 stations within a 500 m spatial range (Figure 1). At each station, two cores were taken; one larger core for the morphological identification of macrofauna (177 cm², 25–30 cm depth) and one smaller for the molecular identification (5.6 cm², 10 cm depth). The two cores were sampled directly adjacent to each other and were taken simultaneously.

2.2 | Morphological approach

The samples for morphological analyses were washed over a 1 mm mesh sieve in the field and stored in the freezer at -20°C. In the laboratory, samples were thawed and preserved in a 4% formaldehyde solution with Bengal rose (~2.5 mg/L). Species were sorted by hand and identified. Taxonomic identification was based on the NIOZ reference collection of local benthic macrofaunal species conform WoRMS Editorial Board (2019). Individuals were counted and biomass (g) was determined as ash-free dry mass of the flesh following Compton et al. (2013). To minimise weighing error for small biomasses, specimens of small taxa such as Tharyx sp., Oligochaeta and Pygospio sp. where only weighted for a minimum of four individuals and Heteromastus sp. was only weighted with a minimum of two individuals and average biomass estimates from the data set were used instead.

2.3 | Molecular approach

To remove extracellular DNA, the entire samples for molecular identification were rinsed twice with a saturated phosphate-buffer solution with Bengal rose (~2.5 mg/L). Species were sorted by hand and identified. Taxonomic identification was based on the NIOZ reference collection of local benthic macrofaunal species conform WoRMS Editorial Board (2019). Individuals were counted and biomass (g) was determined as ash-free dry mass of the flesh following Compton et al. (2013). To minimise weighing error for small biomasses, specimens of small taxa such as Tharyx sp., Oligochaeta and Pygospio sp. where only weighted for a minimum of four individuals and Heteromastus sp. was only weighted with a minimum of two individuals and average biomass estimates from the data set were used instead.

2.4 | Bioinformatics

Raw sequences with a quality score ≤30 over 75% of the nucleotide positions were discarded using the fastq_quality_filter script in the FASTX-Toolkit (https://hannonlab.cshl.edu/). Quality filtered reads were demultiplexed using the split_libraries.py script in QIIME (Caporaso et al., 2010), allowing zero mismatches in both the forward and reverse barcode label. Subsequently, reads were checked for chimeras and dereplicated in VSEARCH (Rognes et al., 2016) and unique sequences were discarded. The remaining sequences were clustered with a 98% similarity cutoff.
This cutoff was found sufficient for genus-level identification for the macrofauna species in our reference library for this area.Singletons were discarded and the remaining out clusters were taxonomically assigned using the RDP Classifier (Wang et al., 2007) with a minimum confidence of 0.8 against the SILVA 18S rRNA database (release 132, Pruesse et al., 2007) and our local reference database (Table S3 and Genbank accession numbers MZ709983–MZ710042). Our local reference database covers all macrofauna species found with the morphological approach.Only reads from taxonomic families containing macrofauna species were retained (see Table 1), reads from other families and reads assigned at a higher taxonomic level were omitted. OTUs assigned at the family level were compared against the NCBI database (https://blast.ncbi.nlm.nih.gov, accessed at 10/2019) using blastn; the OTUs which returned a match with a percent identity >99% at the genus level were assigned accordingly. Raw Illumina sequences were deposited in the European Nucleotide Archive (ENA accession number: PRJEB46793).

| Taxonomy | Family | Genus | Morpho | DNA |
|----------|--------|-------|--------|-----|
| Phylum   |        |       |        |     |
| Annelida | Nereididae | Alitta | 1      | 3   |
|          | Arenicolidae | Arenicola | 104    | 104 |
|          | Capitellidae | Capitella | 104    | 107 |
|          | Echiuridae | Echiurus | -      | 8   |
|          | Phyllophaga | Eteone | 68     | 20  |
|          | Phyllophaga | Eumida | -      | 6   |
|          | Syllidae | Exogone | -      | 6   |
|          | Glyceraidae | Glycera | -      | 10  |
|          | Nereididae | Hediste | 9      | 46  |
|          | Capitellida | Heteromastus | 42    | 98  |
|          | Terebellida | Lanice | 2      | 89  |
|          | Magelonidae | Magelona | 3     | 1   |
|          | Spionidae | Marenzelleria | 40    | 10  |
|          | Nephtyidae | Nephtys | 6      | 3   |
|          | Oligochaeta sp. | Oligochaeta | 58    | 101 |
|          | Phyllophaga | Phyllophaga | 4     | 2   |
|          | Spionidae | Polydora | -     | 3   |
|          | Spionidae | Pygospio | 76     | 94  |
|          | Orbenidae | Scoloplos | 122   | 110 |
|          | Spionidae | Spio | -      | 12  |
|          | Spionidae | Streblospio | -     | 20  |
|          | Cirratulidae | Tharyx | 46     | 123 |
| Arthropoda | Carcinidae | Carcinus | 2      | 2   |
|           | Crangonidae | Crangon | 38     | 3   |
|           | Gammaridae | Gammarus | 3      | -   |
|           | Urothoidae | Urothoe | 81     | 4   |
| Mollusca  | Tellinidae | Macomangulus | -     | 3   |
|           | Cardiidae | Cerastoderma | -     | 18  |
|           | Ostreidae | Magallana | -     | 36  |
|           | Patellidae | Ensis | 2      | 80  |
|           | Tellinidae | Limecola | 16     | 5   |
|           | Myidae | Mya | 1      | 35  |
|           | Mytilidae | Mytilus | -     | 54  |
|           | Hydrobiidae | Peringia | 2      | 29  |
|           | Veneridae | Petricolaria | -     | 13  |


2.5 | Data analysis

All statistical analyses and data visualisations were performed in R 3.5.2. For the morphological approach, count and biomass data were calculated at the genus level, resulting in a total number of individuals and a total biomass per taxonomic genus per sample. Both the abundance and biomass data showed a skewed distribution and hence were square root transformed. To prevent a bias in the molecular data set between samples due to variable sequencing depths per sample, read numbers per genus for all molecular samples were transformed into a relative abundance of reads per sample. Although read numbers per samples differed, no relation was found between sequencing depth and species richness ($t_{(124)} = -.36, p = .72$) Therefore relative abundances were chosen over rarefaction to conserve OTUs (McMurdie & Holmes, 2014) whereas relative read abundance transformations preserve all read calls (Lanzén et al., 2016). For each taxon, their presence per sample was scored and the occurrence was calculated as the sum of total detections divided by the total number of samples. Unfortunately, detection rates for all taxa within the arthropod and mollusc phyla were too low (detection rate $< 20\%$ for both approaches) and highly biased in either the morphological or the molecular methods to test these assumptions within these taxa. Therefore, the reliability of the quantitative approaches could only be tested on annelid taxa.

To test the reliability of the frequency of occurrence approach, the assumption that a higher occurrence in the morphological method corresponds to a higher detection probability in the molecular method was tested. The square root of the abundance data as derived from the morphological method was compared to the occurrence in the molecular method based on a logistic regression using the popbio package in R. The strength of the logistic regression was accentuated in Figure 3. A clear discrepancy could be seen for the mollusc and arthropod taxa. The detection rates for arthropod taxa were higher in the morphological samples. For instance, Urothoe was detected in 64% of the samples for the morphological approach and only in 3% of the samples in the molecular approach. Mollusc species were overall detected at higher rates in the molecular samples. Five mollusc taxa which were detected in the molecular samples (i.e., Macomangulus [2.4%], Cerastoderma [14.3%], Magallana [28.6%], Littorina [0.79%] and Petricolaria [10.3%]) were never detected in the morphological samples. Detection rates for the annelids were higher for both approaches compared to the detection rates of the other phyla. Except for Eteone and Marenzelleria, detection rates for the annelid taxa were higher in the molecular approach compared to the detection rates in the morphological approach.

3 | RESULTS

3.1 | Detection probability

In total, 23 and 35 macrofauna taxa (genus-level) could be identified from the morphological and molecular samples, respectively (Table 1) of which 22 taxa were identified by both approaches. The molecular approach detected seven extra taxa attributed to Annelida and five extra Mollusca taxa compared to the morphological approach, whereas the arthropod Gammarus was only identified by the morphological approach but not with the molecular approach. The annelid taxa exclusively detected by the molecular approach mostly include smaller specimens such as Echiura, Eumida, Spiol, Streblospio and Polydora. Moreover, the genus Polydora is known to include parasitic species and might therefore have been present concealed in its host species. Figure 2 shows that the average number of taxa detected per sample was significantly higher (paired t test, $t_{125} = 13.1$, $p < .001$) for the molecular approach (mean $= 10.0 \pm 2.2$) than for the morphological approach (mean $= 6.6 \pm 2.1$).

The detection rates (calculated as sum of detections divided by the number of samples) for the 21 most abundant species are visualized in Figure 3. A clear discrepancy could be seen for the mollusc and arthropod taxa. The detection rates for arthropod taxa were higher in the morphological samples. For instance, Urothoe was detected in 64% of the samples for the morphological approach and only in 3% of the samples in the molecular approach. Mollusc species were overall detected at higher rates in the molecular samples. Five mollusc taxa which were detected in the molecular samples (i.e., Macomangulus [2.4%], Cerastoderma [14.3%], Magallana [28.6%], Littorina [0.79%] and Petricolaria [10.3%]) were never detected in the morphological samples. Detection rates for the annelids were higher for both approaches compared to the detection rates of the other phyla. Except for Eteone and Marenzelleria, detection rates for the annelid taxa were higher in the molecular approach compared to the detection rates in the morphological approach.

3.2 | Frequency of occurrence and relative read abundance

To test the reliability of the frequency of occurrence approach, the abundance data as derived from the morphological method were modelled using a logistic regression against the occurrence in the molecular data set. Only the abundance data for Pygospio, as derived from the morphological approach, was a significant predictor for the detection of this taxon within the molecular method (Figure 4, Table 2). For Pygospio, per unit of increase in abundance, the odds of a detection in the molecular approach increased by a factor 1.34 ($\text{Wald}-x^2_{(124)} = 3.27, p = .001$). Most of the other annelid taxa did, however, show a positive relationship between the abundance in the morphological approach and occurrence in the molecular inventory. The only exception was Arenicola for which the odd-ratio of occurrence in the molecular approach declined with increasing abundance.

The sensitivity and specificity for each taxon as based on the logistic regression model is shown in Figure 5 as a ROC curve. The degrees of measure for the predictive ability of the model were calculated based on these plots as the area under curve percentage (AUC, Table 2). The logistic model for Pygospio was able to predict the presence of this taxon in the molecular approach at a 77% rate based on the abundance data in the morphological approach. Scoloplos had a predictability of roughly 60%, with a $x^2$ of 1.52; however, for all other species, predictability was lower. The lowest predictability was found for Heteromastus and equalled 52%, which is close to a random guess.
The relative read abundance as derived from the molecular method was modelled in a linear regression against the biomass estimates from the morphological method for six annelid taxa (Figure 6, Table 2). The steepest positive slope was found for *Pygospio*, followed by *Scoloplos* and *Tharyx*. The slope found for *Pygospio* and *Scoloplos* were significant (respectively, $F_{1,40} = 6.25, p = .017, R^2 = .14$ and $F_{1,40} = 5.69, p = .022, R^2 = .13$). *Capitella* and *Heteromastus* both showed no relationship between relative read abundance and biomass, while *Arenicola* displayed a (nonsignificant) negative slope.

The frequency of occurrence and the relative read abundance approaches both showed roughly the same results (Table 2). A strong positive relationship was obtained for *Pygospio* for both approaches. The relative read abundance approach also found a positive relationship for *Scoloplos*. Both approaches showed negative relationships for *Arenicola*.

4 | DISCUSSION

The present study offers a insight in the quantitative abilities of DNA metabarcoding methods. We compared quantitative measurements for abundance and biomass for benthic macrofauna in the intertidal Dutch Wadden Sea for DNA metabarcoding approaches with traditional morphological approaches. Although some widely dispersed annelid taxa showed positive relationships between the outcome of both methods, most taxa did not show such a relationship.

4.1 | Species detection

The combined use of a morphological as well as a molecular method to quantify the same benthic community allowed us to examine whether detection rates for benthic macrofauna taxa are comparable between the two methods. Although the overlap between species found with both methods is high, 23 and 35 taxa, respectively were found from which 22 taxa in both methods, the detection rates between methods within a taxon showed differences. The detection rates for most arthropod and mollusc taxa showed deviating results between the two methods, but the detection rates for the annelid taxa were comparable. The detection rate of arthropod taxa was higher in the morphological compared to the molecular method while the opposite was true for most molluscs. Additionally, several mollusc taxa (e.g., *Cerastoderma*, *Magallana* and *Mytilus*) detected by the molecular method were never detected by the morphological method. Explanations for the striking differences in output between the two methods for the different phyla could potentially be sought both in a methodological and an ecological context.

Methodological issues within the molecular method, which can hamper the assessment of a community, have been described extensively (e.g., Alberdi et al., 2018; Elbrecht & Leese, 2015; Kelly et al., 2019; Lanzén et al., 2017 and Piñol et al., 2019). Two key steps necessary to avoid technical biases emerge from these earlier studies: the use of PCR-replicates and the use of a mock community. Alberdi et al. (2018) showed that considerable diversity differences exist between PCR replicates for the same sample, possibly due to PCR stochasticity and/or the accumulation of PCR errors. In the present study, three PCR replicates per sample were included to minimize these biases as suggested in Grey et al. (2018). Also, a mock community, an artificial community with known species was included to be
able to identify biases induced in the metabarcoding process. This could be, for instance, biases induced during the sequencing process, during PCR due to a mismatch between a (group of) species and the universal primer, as well as biases induced in the bioinformatics pipeline (Leray & Knowlton, 2016). In our study, all species added to the mock community were recovered and the analysis of read numbers showed no indication of a bias in the detection rate of taxa from different phyla, as differences in read numbers were unrelated to phylum level. Hence, we assume that biases induced in the laboratory procedures are negligible and do not cause the discrepancy in detection rates for the molluscs and arthropods between both methods.
of detection rate of mollusc with the molecular approach compared to
originates (Harrison et al., 2019). In our data sets, we found a higher
ecology (seasonal and spatial patterns) of the taxa from which it
biology (physical appearance, ontogenetic state, physiology) and
2012). The amount of eDNA in the environment is a result of the
mucus) rather than the organism itself (Taberlet, Coissac, et al.,
molecules which can include traces of organisms (e.g., air, faeces,
ronmental sample such as marine sediments is a mixture of DNA
ical features of eDNA and the host species. The eDNA in an envi-
were not found using the morphological approach, eDNA could eas-
movements. The low detection rate of arthropods in the molecular
are known to live in high densities in intertidal beds (Folmer et al.,
S1, Philippart et al., 2014). Second, the taxa Mytilus and Magallana
between the relative read
icant positive relationship for
this study, which is a significant positive relationship for
Pygospio
quantity of one species leads to a lower abundance for the other due to lim-
any of the other species. Comparable results have been found by
Bijleveld et al., (2018) in which they showed that occurrence data
from a morphological data set on macrozoobenthic taxa in the in-
tertidal Wadden Sea did not predict abundancies for these taxa.
They discussed that the predictions in that study were highly in-
fluence by dispersal and aggregation patterns and more reliable
predictions could be made for taxa with higher dispersal rates.
The same bias might also play a role in the present study. Adult
arenicola, which has holobenthic development, has relatively low
dispersal capacity, resulting in an aggregated distribution pattern
(Günther, 1992). These Arenicola showed negative relationships in
the frequency of occurrence approach, whereas Pygospio, which
is distributed more evenly across mud flats (Gudmundsson, 1985),
showed a strong positive relationship. This may be interpreted to
indicate that species with an even spatial distribution also distrib-
ute their DNA evenly over the environment, while patchily dis-
tributed species correspondingly also have patchily distributed
DNA, despite, for example, tidal water movements. DNA-based
frequency of occurrence of the more evenly distributed species
would then typically not only be higher overall but also correlate
better with morphology-based quantifications than would be the
case for patchily distributed taxa.

Another explanation can be sought in the ecological and biological
features of eDNA and the host species. The eDNA in an envi-
ronmental sample such as marine sediments is a mixture of DNA
molecules which can include traces of organisms (e.g., air, faeces,
mucus) rather than the organism itself (Taberlet, Coissac, et al.,
2012). The amount of eDNA in the environment is a result of the
biology (physical appearance, ontogenetic state, physiology) and
ecology (seasonal and spatial patterns) of the taxa from which it
originates (Harrison et al., 2019). In our data sets, we found a higher
detection rate of mollusc with the molecular approach compared to
the morphological approach. If we break down the detection rates
of Cerastoderma, Magallana and Mya per sampling event, a poten-
tial seasonal pattern can be seen in which the peaks in detection
rate coincided with known spawning periods of these taxa (Figure
S1, Philippart et al., 2014). Second, the taxa Mytilus and Magallana
are known to live in high densities in intertidal beds (Folmer et al.,
2017) that were situated approximately 500 m from the sampled
area during the study period. Even though specimens of these taxa
were not found using the morphological approach, eDNA could eas-
ily have spread and be trapped within the sampled area due to tidal
movements. The low detection rate of arthropods in the molecular
samples compared to the morphological samples may also be ex-
plained by biological factors. Arthropods are characterized by an
exterior skeleton made of chitin. This skeleton might inhibit eDNA
exchange with the environment. Also, in contrast to mollusc taxa,
the arthropod taxa found in this study rely on internal fertilization,
which minimizes the extra production of eDNA due to spawning.
Biases induced due to ecological characteristics of eDNA and its
host species have been described scarcely but deserve a great deal
of attention in the future (Stewart, 2019).

4.2 | Quantification

The second aim of this study was to test the reliability of the fre-
quency of occurrence and the relative read abundance approaches
for quantifying abundance and biomass of marine benthic taxa
with molecular methods. For this, we tested the underlying as-
sumptions that a higher abundance or biomass in the morpho-
logical data set leads to a higher detection rate or relative read
abundance in the molecular data set for the frequency of occur-
rence and relative read abundance approach, respectively (Deagle
et al., 2019). We only found a positive relationship for Pygospio
using the frequency of occurrence approach and for Pygospio and
Scoloplos for the relative read abundance approach and not for
any of the other species. We hypothesized that the relative read
abundance approach and not for

![Figure 5](image-url)  
**FIGURE 5** Receiver operating characteristics (ROC) curve as derived from the logistic regression for six annelid taxa. The ROC curve shows the sensitivity on the y-axis, calculated as the true positive predictions by the logistic model and the specificity, calculated as the false positive predictions, on the x-axis.

Compared to the frequency of occurrence approach, the relative
read abundance is thought to be more influenced by methodologi-
ical issues than by ecological issues (Deagle et al., 2019; Piñol et al.,
2019). We hypothesized that the relative read abundance approach
would be more influenced by the compositional nature of the molec-
ular data set as relative abundances within an eDNA metabarcoding
data set are negatively correlated, that is, an increased abundance
of one species would then typically not only be higher overall but also correlate
better with morphology-based quantifications than would be the
case for patchily distributed taxa.

For this, we tested the reliability of the frequency of occurrence and
the relative read abundance approaches for quantifying abundance and
biomass of marine benthic taxa with molecular methods. For this, we tested the underlying as-
sumptions that a higher abundance or biomass in the morpho-
logical data set leads to a higher detection rate or relative read
abundance in the molecular data set for the frequency of occur-
rence and relative read abundance approach, respectively (Deagle
et al., 2019). We only found a positive relationship for Pygospio
using the frequency of occurrence approach and for Pygospio and
Scoloplos for the relative read abundance approach and not for
any of the other species. Comparable results have been found by
Bijleveld et al., (2018) in which they showed that occurrence data
from a morphological data set on macrozoobenthic taxa in the in-
tertidal Wadden Sea did not predict abundancies for these taxa.
They discussed that the predictions in that study were highly in-
fluence by dispersal and aggregation patterns and more reliable
predictions could be made for taxa with higher dispersal rates.
The same bias might also play a role in the present study. Adult
arenicola, which has holobenthic development, has relatively low
dispersal capacity, resulting in an aggregated distribution pattern
(Günther, 1992). These Arenicola showed negative relationships in
the frequency of occurrence approach, whereas Pygospio, which
is distributed more evenly across mud flats (Gudmundsson, 1985),
showed a strong positive relationship. This may be interpreted to
indicate that species with an even spatial distribution also distrib-
ute their DNA evenly over the environment, while patchily dis-
tributed species correspondingly also have patchily distributed
DNA, despite, for example, tidal water movements. DNA-based
frequency of occurrence of the more evenly distributed species
would then typically not only be higher overall but also correlate
better with morphology-based quantifications than would be the
case for patchily distributed taxa.

Compared to the frequency of occurrence approach, the relative
read abundance is thought to be more influenced by methodologi-
ical issues than by ecological issues (Deagle et al., 2019; Piñol et al.,
2019). We hypothesized that the relative read abundance approach
would be more influenced by the compositional nature of the molec-
ular data set as relative abundances within an eDNA metabarcoding
data set are negatively correlated, that is, an increased abundance
of one species leads to a lower abundance for the other due to lim-
ited sampling depth (Gloor et al., 2017). However, the frequency of
occurrence and relative read abundance gave comparable results in
this study, which is a significant positive relationship for Pygospio.
Moreover, relative read abundance approach also showed a sig-
nificant positive relationship for Scoloplos between the relative read
abundance and biomass.
The comparable results between the frequency of occurrence and relative read abundance methods might imply that both methods were subject to the same factors or biases in this study. We discussed that the ecological and biological features might be an important factor in this. The sampling strategy used in this study possibly induced biases in the quantitative measurements as well. Smaller species and species which are more evenly distributed such as *Pygospio* and *Scoloplos*, showed positive relationship for both approaches whereas bigger and more aggregated species such as mollusc species and *Arenicola* showed no relationship. For this study, a molecular sampling method was chosen in which only small sediment samples were collected from which DNA was extracted directly. Specimens of the smaller taxa might be physically present in the sample while larger metazoan taxa are sampled via their eDNA traces (Taberlet, Prud’homme, et al., 2012). Larger macrofauna species might have shown better quantitative relationship when bulk samples would have been used (Klunder et al., 2019). However, we chose the current method as it applies fewer treatments to the samples, which makes the method less susceptible to contamination (Aylagas et al., 2016; Elbrecht et al., 2017) as well as faster and easier to perform. In accordance with this, Elbrecht et al. (2017) found that read abundance of unsorted samples were dominated by taxa containing higher biomass within the sample and size sorting could help in preventing this bias. However, they also discuss size sorting should only be used when highly necessary to prevent cross-contamination between the samples.

### 4.3 | Future outlook

Metabarcoding methods have undergone considerable developments in the past decade and the methodology is still advancing rapidly. These advancements and new insights possibly harbour solutions to the current challenges and limitations in quantifying species abundancies from eDNA studies. It is outside the scope of this study to discuss them all, but we would like to highlight some. First, in this study we tested the conventional frequency of occurrence and relative read abundance approaches to quantify macrobenthic abundancies from eDNA studies. We showed that both approaches are limited in their quantitative abilities towards this community. In recent years, approaches have been developed that better incorporate the compositional conditions of eDNA data for microbial communities (Gloor et al., 2017) which are potential useful for all eDNA data (Quinn et al., 2019). Second, in this study it became clear that ecological and biological factors are important factors which can bias species detection and quantification. Ecological factors driving the distribution of eDNA have been described sparsely in the past and only gained scientific

![Figure 6](image-url): Relative read abundance approach relationships for six annelid taxa. The square-root of biomass estimates as derived from the morphological method are shown on the x-axis and the log$_{10}$ of the relative read abundance of the same sample in the molecular data set is shown on the y-axis.
interested recently (Stewart, 2019). Moving forward, we suggest further knowledge about the factors driving eDNA distribution in the environment for different taxa and how this coincides with the sampling strategy of choice should be gained.

ACKNOWLEDGMENTS

We are grateful to all the staff and students who helped in collecting field samples, sorting them in the benthic laboratory and processing them in the molecular laboratory. We would like to thank Giovanni Heijmans, Sophie Watertor and Rowan Stavast in particular. We would also like to thank Hans Malschaert for maintaining the bioinformatic servers. Valuable comments by multiple anonymous reviewers improved the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Lise Klunder, Allert I. Bijleveld, Henk W. van derVeer and Pieterenella C. Luttikhuizen conceived the study; Lise Klunder, Loran Kleine Schaars and Pieterenella C. Luttikhuizen performed the laboratory and fieldwork; Lise Klunder analysed the data and wrote the first draft of the manuscript; Lise Klunder, Allert I. Bijleveld, Pieterenella C. Luttikhuizen, Henk W. van der Veer, Loran Kleine Schaars and Judith D.L. van Bleijswijk edited the manuscript.

DATA AVAILABILITY STATEMENT

Raw Illumina sequences and corresponding metadata have been deposited in the European Nucleotide Archive; accession number PREJB46793. Sanger sequences from our internal library can be accessed through Genbank, accession nos. MZ709983–MZ710042.

ORCID

Lise Klunder https://orcid.org/0000-0002-3072-6158
Pieterenella C. Luttikhuizen https://orcid.org/0000-0002-5735-4921

REFERENCES

Alberdi, A., Alizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9, 134–147. https://doi.org/10.1111/2041-210X.12849
Austen, M. C., Lambshde, P. J. D., Huchings, P. A., Boucher, G., Heip, C., King, G., Koike, I., Smith, C., & Snelgrove, P. V. R. (2002). Biodiversity links above and below the marine sediment-water interface that may influence community stability. Biodiversity and Conservation, 11, 113–136. https://doi.org/10.1023/A:1014098175353
Aylagas, E., Borja, A., Muxika, I., & Rodriguez-Ezepeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. Ecological Indicators, 95, 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044
Aylagas, E., Mendibil, I., Borja, A., & Rodriguez-Ezepeleta, N. (2016). Marine sediment sample pre-processing for macroinvertebrates metabarcoding: Mechanical enrichment and homogenization. Frontiers in Marine Science, 3, 1–8. https://doi.org/10.3389/fmars.2016.00203
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/s10092-015-0775-4
Berry, T. E., Osterrieder, S. K., Murray, D. C., Coghlan, M. L., Richardson, A. J., Grealy, A. K., Stat, M., Beijer, L., & Bunce, M. (2017). DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (Neophoca cinerea). Ecology and Evolution, 7, 5435–5453. https://doi.org/10.1002/ece3.3123
Beukema, J. J., & Dekker, R. (2020). Half a century of monitoring macrobenthic animals on tidal flats in the Dutch Wadden Sea. Marine Ecology Progress Series, 656, 1–18. https://doi.org/10.3354/meps13555
Bijleveld, A. I., Compton, T. J., Klunder, L., Holthuijsen, S., ten Horn, J., Koolhaas, A., Dekinga, A., van der Meer, J., & van der Veer, H. W. (2018). Presence-absence of marine macrozoobenthos does not generally predict abundance and biomass. Scientific Reports, 8, 1–12. https://doi.org/10.1038/s41598-018-21285-1
Brannock, P. M., & Halanych, K. M. (2015). Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. Marine Genomics, 23, 67–75. https://doi.org/10.1016/j.margen.2015.05.007
Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoo. Annual Review of Marine Science, 3, 471–508. https://doi.org/10.1146/annurev-marine-120308-080950
Cahill, A. E., Pearman, J. K., Borja, A., Caragati, L., Carvalho, S., David, R., Feral, J.-P., Olenin, S., Siáulis, A., Yorks, J. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. Ecology and Evolution, 8, 8908–8920. https://doi.org/10.1002/eco.43283
Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Gonzalez Pena, A., Goodrich, J. K., Gordon, J. I., Huttenhower, C. A., Kelly, S. T., Knights, D., Koenig, E. J., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., ... Knight, R. (2010). QIIME allows high throughput community sequencing data. Nature Methods, 7, 335–336. https://doi.org/10.1038/nmeth0510-335
Cardoso, P., Erwin, T. L., Borges, P. A. V., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation, 144, 2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024
Chariton, A. A., Stephenson, S., Morgan, M. J., Steven, A. D. L., Colloff, M. J., Court, L. N., & Hardy, C. M. (2015). Metabarcoding of benthic eu-karyote communities predicts the ecological condition of estuaries. Environmental Pollution, 203, 165–174. https://doi.org/10.1016/j.envpol.2015.03.047
Compton, T. J., Holthuijsen, S., Koolhaas, A., Dekinga, A., ten Horn, J., Smith, J., Galama, Y., Brugge, M., van der Wal, D., van der Meer, J., van der Veer, H. W., & Piersma, T. (2013). Distinctly variable mudscapes: Distribution gradients of intertidal macrofauna across the Dutch Wadden Sea. Journal of Sea Research, 82, 103–116. https://doi.org/10.1016/j.jmares.2013.02.002
Corinaldesi, C., Beolchini, F., & Dell’Anno, A. (2008). Damage and degradation rates of extracellular DNA in marine sediments: Implications for the preservation of gene sequences. Molecular Ecology, 17, 3939–3951. https://doi.org/10.1111/j.1365-294X.2008.03880.x
Covich, A. P., Austen, M. C., Barlocher, F., Chauvet, E., Cardinale, B. J., Biles, C. L., Inchausti, P., Dangles, O., Solan, M., Gessner, M. O., Statzner, B., & Moss, B. (2004). The role of biodiversity in the functioning of freshwater and marine benthic ecosystems. BioScience, 54, 767–775. https://doi.org/10.1641/0006-3568(2004)054[0767:TBRITE]2.0.CO;2
Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Arnaud-Haond, S. (2015). Metabarcoding is powerful yet still blind: A comparative analysis of morphological and molecular
macrobenthos of estuarine systems. *Marine Environment Research, 68*, 25–36. https://doi.org/10.1016/j.marenvres.2009.04.001

Philippart, C. J. M., van Bleijswijk, J. D. L., Kromkamp, J. C., Zuur, A. F., & Herman, P. M. J. (2014). Reproductive phenology of coastal marine bivalves in a seasonal environment. *Molecular Ecology Resources, 15*(4), 819–830. https://doi.org/10.1111/plankt.12073

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources, 15*, 819–830. https://doi.org/10.1111/1755-0998.12355

Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology, 28*, 407–419. https://doi.org/10.1111/mec.14776

Plummer, E., & Twin, J. (2015). A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *Journal of Proteomics & Bioinformatics, 8*, 283–291. https://doi.org/10.4172/jpb.1000381

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology, 21*, 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Porazinska, D. L., Sung, W., Gibling-Davis, R. M., & Thomas, W. K. (2010). Reproducibility of read numbers in high-throughput sequencing analysis of nematode community composition and structure. *Molecular Ecology Resources, 10*, 666–676. https://doi.org/10.1111/j.1755-0998.2009.02819.x

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research, 35*, 7188–7196. https://doi.org/10.1093/nar/gkm864

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience, 8*, 1–14. https://doi.org/10.1093/gigascience/giz107

Richardson, R. T., Bengtsson-Palme, J., & Johnson, R. M. (2017). Evaluating and optimizing the performance of software commonly used for the taxonomic classification of DNA metabarcoding sequence data. *Molecular Ecology Resources, 17*, 760–769. https://doi.org/10.1111/1755-0998.12628

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

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A. J., Yamamoto, H., Chevaldonné, P., Cedhagen, T., Carvalho, G., & Creer, S. (2016). Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea Benthos. *Frontiers in Marine Science, 3*, 1–14. https://doi.org/10.3389/fmars.2016.00092

Snelgrove, P. (1997). The importance of marine sediment biodiversity in ecosystem processes. *Ambio, 26*, 578–583. JSTOR, www.jstor.org/stable/4314672

Stewart, K. A. (2019). Understanding the effects of biotic and abiotic factors on sources of aquatic environmental DNA. *Biodiversity and Conservation, 28*, 983–1001. https://doi.org/10.1007/s10531-019-01709-8

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology, 21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Prud’homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology, 21*, 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of RNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology, 73*, 5261–5267. https://doi.org/10.1128/AEM.00062-07

WoRMS Editorial Board. (2019). World register of marine species. Retrieved from http://www.marinespecies.org at VLIZ. Accessed 2019-12-10. https://doi.org/10.14284/170

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Klunder, L., van Bleijswijk, J. D. L., Kleine Schaars, L., van der Veer, H. W., Luttikhuizen, P. C., & Bijleveld, A. I. (2022). Quantification of marine benthic communities with metabarcoding. *Molecular Ecology Resources, 22*, 1043–1054. https://doi.org/10.1111/1755-0998.13536