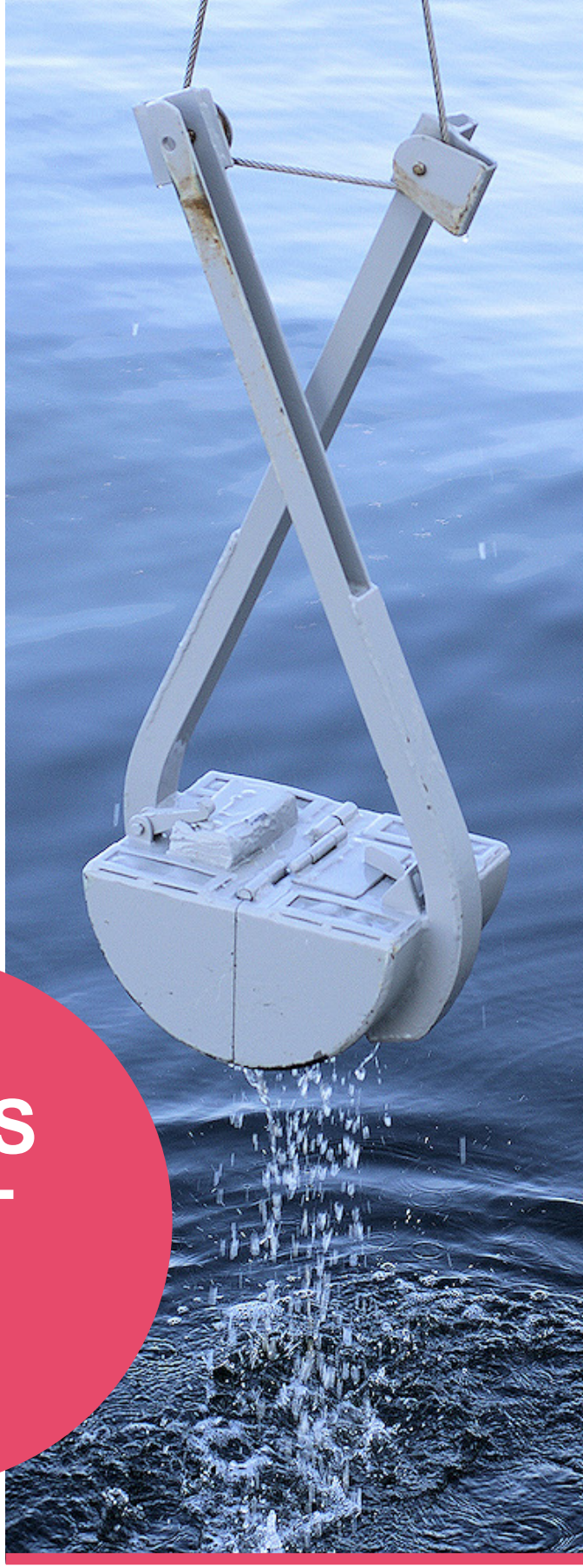
DNA-BASED MONITORING OF SOFT SEDIMENTS



GEANS PILOT 1.

Interreg North Sea Region GEANS
European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region

DNA-BASED MONITORING OF SOFT SEDIMENTS

Version: 3.0

Project:

GEANS– Genetic tools for Ecosystem health Assessment in the North Sea region

Authors:

Sofie Derycke, Kevin Beentjes, Magdalini Christodoulou, Jørgen LS Hansen, Sahar Khodami, Ingrid Kröncke, Pedro Martinez Arbizu, Rumakanta Sapkota, Peter AU Stæhr, Laure Van Den Bulcke, Anne Winding, Annelies De Backer

Date:

17 May 2023

Lead Partner:

Flanders Research Institute for Agriculture, Fisheries and Food (ILVO)

Funding:

Funding for this research was received through the North Sea Program of the European Regional Development Fund of the European Union. Cofinancing was provided by the Sand Fund of the Federal Public Service Economy (Belgium), Senckenberg Research Institute (Germany), and from the Danish EPA.

Main contact:

sofie.derycke@ilvo.vlaanderen.be

How to cite this report?

Derycke S., Beentjes K., Christodoulou M., Hansen J.L.S., Khodami S., Kröncke I., Martinez Arbizu P., Sapkota R., Stæhr P.A.U., Van Den Bulcke L., Winding A., De Backer A. (2023). DNA-based monitoring of soft sediments.

Contents

1. Summary

DNA-based monitoring has the potential to provide high resolution biodiversity data from the marine environment at a fraction of the time and costs associated with morphology based monitoring. Yet, proper ground truthing of DNA-based monitoring is needed across different member states to reach harmonization of methods and to pave the way for implementation in national and European monitoring programs. Within the GEANS soft sediment pilot, three case studies were selected in which bulk DNA-based monitoring was conducted in parallel with morphology-based monitoring: 1/ national ecological impact monitoring for aggregate extraction in the Belgian part of the North Sea, 2/ long term monitoring of a North Sea benthos observatory sampled since 1978 in Germany and 3/ impact assessment of mechanical harvesting of lugworms (*Arenicola marina*) in The Netherlands. The first two case studies showed very consistent results with identical ecological patterns in alpha diversity (number of species, Shannon-diversity) and beta diversity (community composition) compared to traditional morphology based sampling despite the very different objectives of the two studies. Differences between the two methods were situated in the number of species detected - significantly more species were detected with DNA-based monitoring - and in their identity - only 37% (case study 1) and 25% (case study 2) of the species were detected by both methods thereby emphasizing the complementarity of the two methods. The species responsible for the observed differences linked to sand extraction or to the seasonal fluctuations in the long term monitoring study were to some extent shared between the two methods, but each method also had unique indicator species. For the lugworm case study, no data could be obtained because of a failure in the PCR amplification step. This illustrates that DNA-based monitoring may also come with a risk of getting no data and highlights the importance of following accurate field and lab protocols. In a fourth case study, eDNA from the sediment instead of bulk DNA was used to determine macrobenthos as part of the national MSFD monitoring in the Danish part of the North Sea. Here, similar ecological patterns were observed between eDNA-based and morphological data with both methods showing significant differences between the locations and between the two depth zones. However, species lists were completely different between the two methods, illustrating that eDNA from the sediment is not suitable for macrobenthos monitoring related to MSFD assessments. The case studies all tracked time and costs associated with the two methods, and provide empirical evidence that bulk DNA-based monitoring is 26-27% cheaper and 46-66% faster than morphology based monitoring. A SWOT analysis for DNA-based monitoring of macrobenthos was conducted based on our experience with the case studies, and the contribution of GEANS to solve some of the weaknesses and threats are provided. Finally, we provide practical recommendations for performing DNA-based monitoring and highlight the next steps towards implementation of the method for marine soft-bottom environmental monitoring.

2. Introduction

Soft-bottom animals bigger than 1 mm, also known as macrobenthos, are good indicators to assess marine environmental quality. Consequently, soft sediment samples are taken all around the North Sea in the framework of the European Marine Strategy Framework Directive (MSFD), environmental impact assessments or for monitoring long-term environmental trends. Despite the different objectives of these monitoring programs, the overall sample processing is largely

similar: macrobenthos individuals are sorted, counted and identified under a stereo-microscope by taxonomic experts. This is a time-consuming process, and accuracy of species detection relies on the skills of the identifier. Furthermore, conventional monitoring of soft bottom fauna is based on grabs (Fig 1) with a small sampling surface (0.1 to 0.5 m^2) compared to the total area under study and thus a single sample only represents a fraction of the species pool present in the surroundings. Hence, the number of detected species depends on the number of samples collected, and only a limited part of the diversity of the sampled area will be described. Marine benthic environmental assessments thus depend on the number of collected samples and the expertise of the taxonomist.

DNA-based approaches, where species can be distinguished based on their unique DNA barcode, may speed up and complement the conventional methods for species identification. The approaches used in this project, either use so-called bulk DNA where the macrobenthos species are used as starting material or use DNA molecules released by the benthos in the water or sediment through e.g. tissue, faeces, slime, cells, gametes (so-called eDNA or environmental DNA). Common for the DNA-based approaches is that they do not rely on sorting and identifying each individual specimen, and can therefore easily process many samples without similar upscaling of the costs associated with sample processing. DNA-based methods may therefore provide more representative biodiversity data for a study area.

The GEANS soft-bottom pilot was designed to compare traditional morphology based processing of macrobenthos samples with bulk DNA or sediment eDNA processing of the same samples in four different monitoring case studies in the North Sea: national ecological impact monitoring for aggregate extraction in the Belgian part of the North Sea, national MSFD monitoring in the Danish part of the North Sea, long term monitoring of a North Sea Benthos Observatory sampled since 1978 in Germany and impact assessment of mechanical harvesting of lugworms (*Arenicola marina*) in the Wadden Sea near Texel (The Netherlands). The different case studies were specifically chosen to evaluate whether DNA-based monitoring reveals similar ecological patterns in alpha diversity (e.g. number of species, Shannon-diversity) and beta diversity (e.g. community composition) compared to traditional morphology based sampling. We also investigated whether both approaches detected the same species and whether the indicator species identified by each approach were similar. A SWOT analysis based on the results of these case studies supports our final recommendations on how DNA-based monitoring can be implemented in marine soft-bottom environmental monitoring.

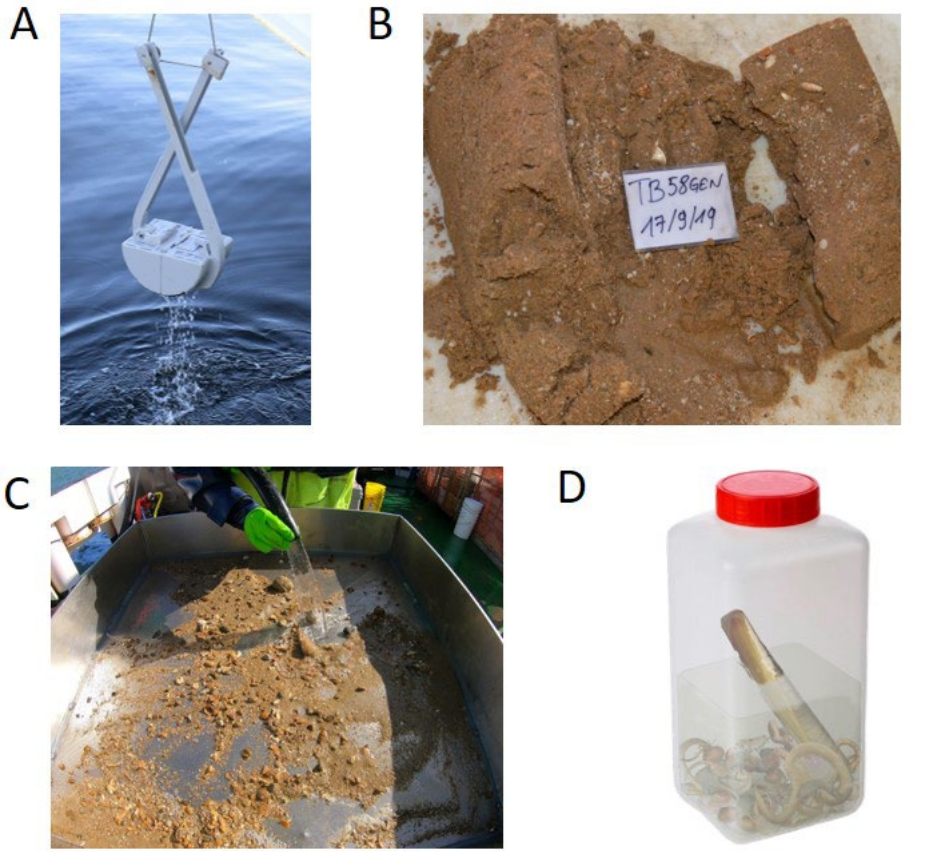


Figure 1. Overview of typical steps involved in conventional monitoring of soft sediment habitats. A/ Soft-bottom grab; B/ sediment sample; C/ sieving of sediment; D/ collection of macrobenthos in formalin.

3. Pilot design

The soft sediment pilot consists of four different monitoring case studies: a/national ecological impact monitoring for aggregate extraction in the Belgian part of the North Sea, b/ long term monitoring of a North Sea Benthos Observatory sampled since 1978 in Germany, c/ impact assessment of mechanical harvesting of lugworms (*Arenicola marina*) in the Wadden Sea near Texel (The Netherlands), and d/national MSFD monitoring in the Danish part of the North Sea.

a. Case study on environmental impact assessment of marine aggregate extraction (Belgium)

i. Design and monitoring objective

This pilot was designed to evaluate whether DNA-based monitoring can be used for environmental impact assessment (EIA) of marine aggregate extraction. Marine aggregate extraction activities are extensive throughout the North-East Atlantic and alter the seafloor through sediment removal and sediment (re)suspension which in turn affect the benthic communities (de Jong et al., 2015, Wyns et al., 2021). Monitoring changes in soft-sediment

macrobenthos communities is part of a legally obliged environmental monitoring program for aggregate extraction activities in the Belgian part of the North Sea (BPNS).

We selected three concession zones in the BPNS which are characterized by different degrees of extraction intensity: the Thortonbank, which is the epicenter of extraction since 2015 with continuous high extraction intensities of ca 150 000 m^3/month for industrial use, the Oostdyck with continuous but low extraction intensities of around 30 000 m^3/month for industrial use as well and the Hinderbanken with periodically high amounts of extraction (sometimes up to 500 000 m^3/month) for coastal protection. Depending on the amount of sand extracted during one year in a 50 m buffer area (i.e. 7800 m^2) around our sampling location, three impact groups (high: > 2000 m^3, medium: 500 – 2000 m^3, low: < 500 m^3) and a reference group (0 m^3) were delineated in each of the sandbanks. Initially five locations were foreseen in each impact zone, but as the extracted volume of sand was only determined after sampling, between one and six locations for the impact zones were obtained, together with six to twelve reference locations for each sandbank (Oostdyck: 12, Hinderbanken: 12 (2019) or 6 (2021) and Thorntonbank: 9). These locations were sampled in autumn 2019 onboard the RV Belgica and again in autumn 2021 onboard RV Simon Stevin (Oostdyck), GEO-Ocean V (Hinderbanken) and GeoSurveyor XI (Thortonbank).

ii. Collection of samples

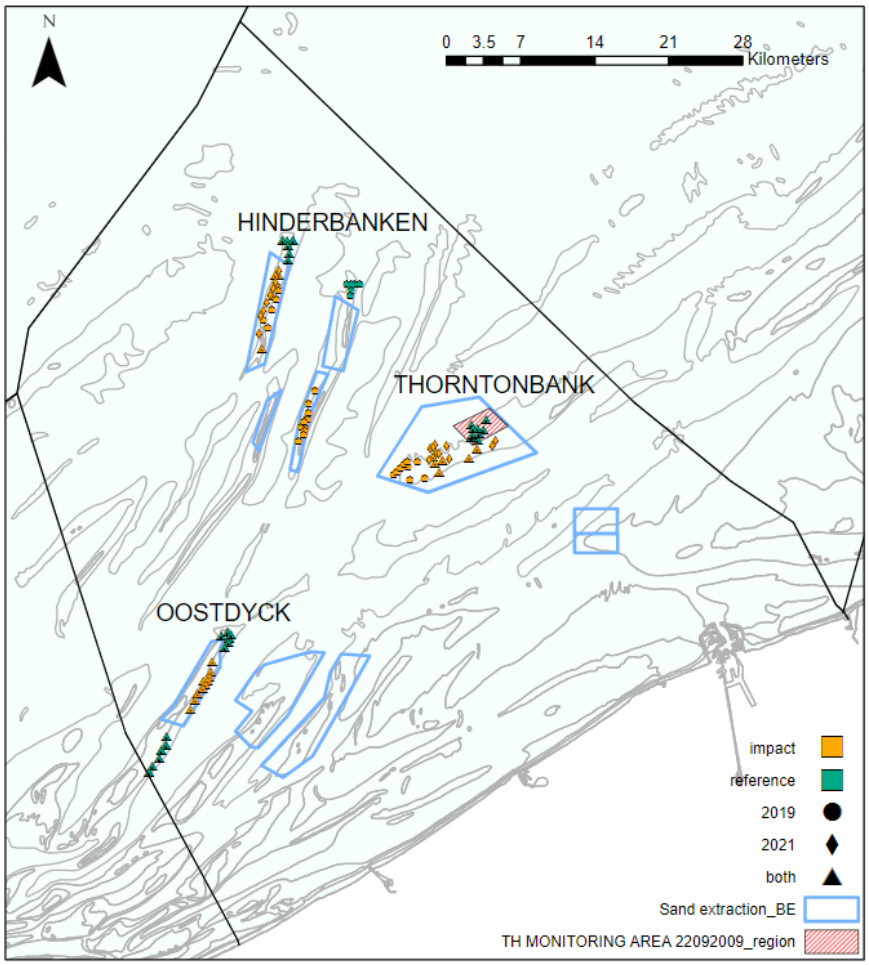


Figure 2: Map of Belgian Part of the North Sea (BPNS) with the different sampling locations in three sandbank systems (Hinderbanken, Oostdyck, Thorntonbank) in 2019 (78 samples, circles) and 2021 (65 samples, diamonds). Locations sampled in both years are indicated by triangles. Green: reference locations; orange: impact locations.

Sediment samples were collected using a Van Veen grab (0.1 m²). During the 2019 sampling campaign, two replicate Van Veen grabs were collected in 78 locations distributed in impact and reference areas in the three sandbanks (Thontonbank, Hinderbanken and Oostdyck; Fig 2). For each grab, the sediment was sieved on a 1 mm sieve and the remaining animals were fixed in absolute ethanol and stored at -20°C until further DNA-based processing for the first Van Veen, or fixed in formalin for morphological identification in the lab for the second Van Veen. The ethanol-preserved samples of the Thorntonbank (n=24) were first identified morphologically and then bulk DNA processing took place. During the 2021 campaign, one Van Veen was collected in 65 locations distributed in impact and reference areas in the three sandbanks. The sediment was sieved on a 1 mm sieve and the remaining animals were fixed in absolute ethanol and stored at -20°C. Ethanol preserved specimens were first identified morphologically and then processed in the lab for bulk DNA-based identification.

iii. Lab processing

Specimens were recovered from the formalin or ethanol samples by the decanting process using a 1 mm sieve and tap water (each sample was ten times decantated). Heavier specimens were added to the decanted material in ethanol/ formalin after screening the remaining material (e.g. shells). Specimens were identified morphologically up to species level, except for juveniles, which were identified up to genus level and specimens belonging to Nemertea, Anthozoa and Oligochaeta, which were identified up to phylum, class and order level, respectively. Molecular processing of the ethanol fixed samples was done according to the GEANS COI metabarcoding lab protocol (https://northsearegion.eu/geans/output-library/). Samples from 2019 were pooled into one library and samples from 2021 were pooled in a second library. Both libraries were sent to Admera Health (US) for paired-end 250bp Illumina Miseq sequencing. Each library was sequenced on a separate Miseq run.

iv. Bioinformatic processing

Bioinformatic processing of the sequencing data was separately conducted for each dataset (2019 or 2021) in R v4.0.2 (Core Team, 2022). The quality of demultiplexed reads was checked with MultiQC (Ewels et al. 2016), and forward and reverse primers were removed using Trimmomatic (Bolger et al. 2014). Amplicon sequence variants (ASVs) were generated using the Dada2 pipeline in the Dada2 v1.17.0 package (Callahan et al. 2016). Reads were further trimmed by removing parts with a quality score lower than 30. Forward and reverse reads were merged for each sample. Chimeras were removed with the removeBimeraDenovo function, and taxonomy was assigned with the assignTaxonomy function in the Dada2 package using the Ribosomal Database Project (RDP) Classifier (Wang et al. 2007). Standard settings were employed, except for the minimum bootstrap confidence parameter, which was set to 80. The GEANS reference database version 4 containing 1993 COI sequences from 565 species was used. Contaminants were removed from the dataset with the "prevalence" method (using the detected ASVs in the negative controls) of the decontam package v1.18.0 (Davis et al., 2017).

v. Data-analyses

Samples in the DNA-based analysis were rarified at 10 000 reads for the 2019 dataset and at 13 000 reads for the 2021 dataset. This number was a tradeoff between reaching the plateau of the rarefaction curves and removing a minimum number of samples. Diversity was assessed using the lowest taxonomic classification (i.e. species level for bulk DNA and higher taxonomic classification for difficult-to-identify taxa using morphology). A square root transformation on the reads (bulk DNA datasets) and on the counts (morphological datasets) was performed to minimize effects of highly abundant species. The number of species and the Shannon diversity index were determined using the diversity function of the vegan package (v2.6.4, Dixon 2003) and visualized using the ggplot2 package (v3.4.0, Wickham 2016). The formalin-fixed samples collected in the Thorntonbank in 2019 were excluded from subsequent analyses.

The effects of methodology and sand extraction impact on the number of species and Shannon index were tested using a mixed ANOVA model with fixed factors methodology (two levels: bulk

DNA, morphology) and impact (four levels: Reference, Low, Medium, High) and the interaction factor impact*methodology. Sandbank and sampling stations were considered random factors, with sampling stations nested within sandbank. Variation caused by sampling year was included as a fixed factor in the model because the number of levels (two: 2019 and 2021) was too low to be defined as a random factor. When the interaction effect was not significant, it was removed to only look at the main effects 'methodology' and 'impact'. Model assumptions were checked by looking at plots of the residuals to investigate the homogeneity of the variances and the normality of the data. When model assumptions were not met, data were log transformed. If significant effects were observed, posthoc pairwise comparisons were calculated using the package emmeans v1.8.2.

The effect of aggregate extraction and sandbanks on species numbers and Shannon index were studied using a mixed ANOVA model with fixed factors impact (four levels: Reference, Low, Medium, High) and sandbank (three levels: Hinderbanken, Oostdyck, Thornonbank). Variation due to sampling stations and sampling year were included as random factor or fixed effect (see above), respectively. This ANOVA was conducted on the bulk DNA and morphology-based datasets separately to investigate whether observed patterns were similar for both methods. When the interaction effect was non-significant, it was removed from the model to investigate the main effects of impact and sandbank. ANOVA assumptions and post hoc testing were tested as described above. The species detected with bulk DNA and morphology were compared using only the samples with data from both methods (so samples removed by rarifying in the bulk DNA dataset were also excluded from the morphology based dataset) and visualized with Venn diagrams for each sandbank and sampling year separately.

For each sandbank and identification method separately, the variation in community composition between impact groups and years was investigated using non-metric multidimensional scaling (nMDS) plots based on the Bray-Curtis (Edward 1984) dissimilarity index using PRIMER 6. A two-way PERMANOVA was conducted in R with main effects 'impact' (four levels: reference, low, medium, high) and year (two levels: 2019 and 2021), performed with 9999 permutations. A distance dispersion test and permutest were used to test the homogeneity of dispersion between samples with the R package "vegan" v2.5-6. For each sandbank, a SIMPER analysis (in PRIMER 6, 90% abundance cut-off level) for both the morphological and bulk DNA identification method was performed to investigate which species contributed most to within-group similarity of the different sand extraction impact groups. This allowed us to observe whether both methods revealed similar characteristic species groups.

b. *Case study on long-term soft bottom monitoring at Norderney* island (Germany)

i. Design and monitoring objective

The study was performed north of Norderney Island in the Wadden Sea (Fig 3, South North Sea, Germany), one of the few long-term benthic study sites in the North Sea and the world. Senckenberg has a 45 year time series in this site (Zeiss & Kröncke, 1997; Kröncke et al., 1998;

2001, 2013; Dippner & Kröncke, 2015), where samples have been collected since 1978 aiming to monitor temporal variability on macrobenthic communities.

For this pilot, the same sampling plan was followed as the one performed for the long-term monitoring since 1978 in order for our results to be comparable (Kröncke et al., 2001). Sampling was carried out on board of RV *Senckenberg* three times within the years 2019-2020. Samples were taken from five stations (stn 11, stn 12, stn 13, stn 14, stn 15) north from Norderney Island in water depths ranging between 12 and 20 m (Fig 3). The stations at the study site are characterized by fine sand (grain size: 63-250 µm) and a *Fabulina fabula* community (Kröncke et al., 2001; Kröncke & Reiss, 2010) in which the bivalve *F. fabula*, species of the polychaete genera *Nephtys* and *Magelona*, amphipod species of the genus *Bathyporeia* are dominant. No fishing by big vessels is allowed, while seldom smaller vessels fish in the area.

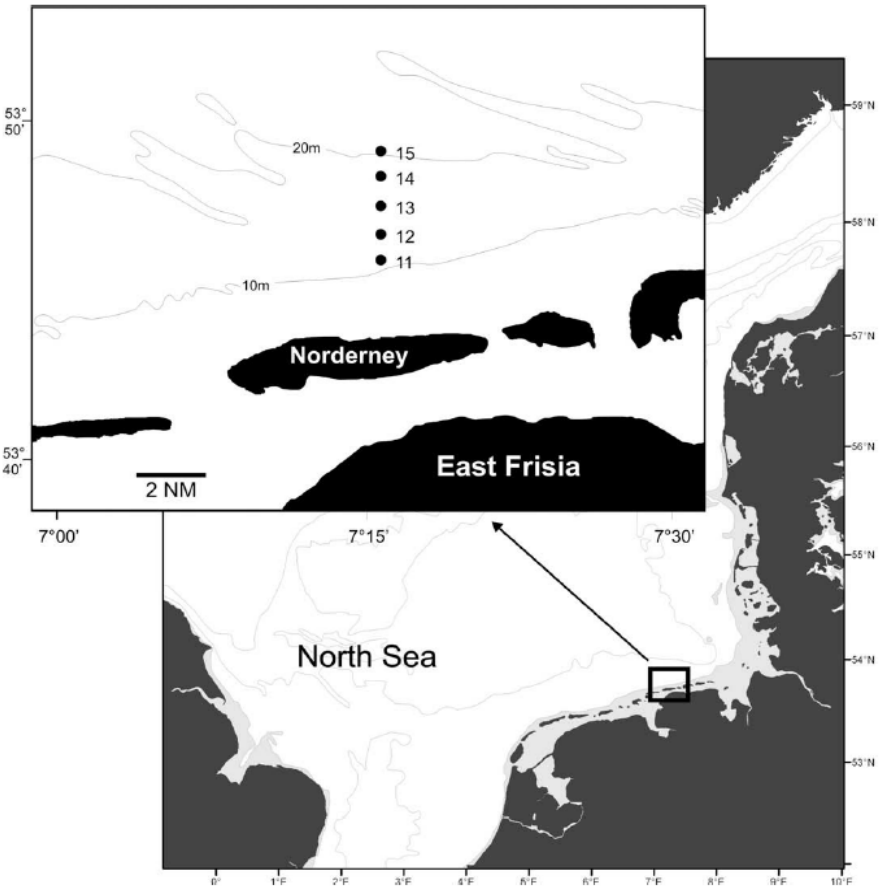


Figure 3. Map indicating the sampling locations in the long-term monitoring station of Senckenberg off Norderney island, Wadden Sea (North Sea).

ii. Collection of samples

Sampling was carried out in September 2019, March 2020 and June 2020 in parallel with the morphology-based sampling. Soft sediment samples were collected using a 0.2 m^2 van Veen grab following the sampling protocol of the Norderney long-term monitoring (Kröncke et al.,

2013) and were washed on board over a sieve of 0.63 mm mesh size (Fig 4). From each of three stations (stn 11, stn 13, stn 14), two replicate Van Veens were taken, of which one was fixed in formalin 4% and the other in 96% undenatured ethanol. From the remaining two stations (stn 12, stn 15) six replicates were taken, of which three were fixed in formalin and three in ethanol. The ethanol from the samples was exchanged the following day and always kept in -20°C pending their morphological examination and DNA-based processing. Both ethanol and formalin preserved samples were sorted and then morphologically identified, except for the samples from June where only the formalin samples have been identified morphologically. Specimens from one replicate per station of the ethanol preserved samples have been pooled together (September and March) for bulkDNA metabarcoding, while for June, this one replicate was used directly for metabarcoding. The rest of the ethanol preserved replicates have been used for building the barcode reference library. In summary, for metabarcoding one sample per station from each season (with three DNA extraction pseudoreplicates) has been sequenced and analyzed (n=45); for morphology, from spring and autumn, one replicate from stations 11, 13 and 14, and 3 replicates from stations 12 and 15 have been processed from each of the formalin and ethanol preserved samples. For summer, only one replicate from only formalin samples has been identified morphologically from all the 5 stations.

Figure 4. Soft sediment sampling on board the RV Senckenberg off Norderney Island, Wadden Sea (North Sea).

iii. Lab processing

Specimens were sorted and separated first in higher taxonomic groups from both the formalin and the ethanol preserved samples. For the samples collected in June, morphological analyses have been only carried out on the formalin preserved samples. In order to make the specimens easier visible, rose bengal stain was used for the formalin kept specimens but not for the ethanol preserved so as to not interfere with downstream genetic analyses. Specimens were identified morphologically down to species level, whereas for juveniles genus level identification was possible. Morphologically difficult groups such as Nemertea, Cnidaria and Oligochaeta, were only identified to phylum level.

For metabarcoding, one ethanol-preserved sample from each station/season (total of 15 samples) was homogenized for 60 seconds using a lab mixer. From each homogenized sample, three subsamples (pseudoreplicates) of 1.5 ml in sterile 1.5 ml eppendorf tubes were taken; the ethanol was completely evaporated using a SpeedVac centrifuge at 60°C for 1 hour. Genomic DNA were extracted using the DNeasy PowerSoil Pro Kit (Qiagen) following the manufacturer's protocol and eluted in 100 ul of elution buffer. Using a real-time PCR cycler, two qPCR were performed to amplify the COI gene fragment and attach unique dual indexes and Illumina universal adapters to each library. The first qPCR was performed using 1µl of mlCOIintF (Leray et al., 2013), 1 µl of jgHCO2198 (Geller et al., 2013), 10ul of SsoAdvanced Universal Inhibitor-Tolerant SYBR Green Supermix (BIO-RAD), 6µl of DNase-free water and 2µl of template DNA. Initial denaturation of 98°C for 3 minutes followed by 30 cycles of denaturation at 98°C, annealing at 54°C, extension of 72°C (each step for 30 seconds) and a final extension of 72°C for 3 minutes were set up for the first qPCR. Same supermix has been used for the second qPCR with 2µl of IDT® for Illumina® DNA/RNA UD Indexes, 6 µl of DNase-free water and 2ul of 1st PCR product as template. The cycling program followed: initial denaturation at 98°C for 1 minute, 15 cycles of 98°C denaturation (10 sec), annealing and extension of 72°C (15 sec) and a final extension of 72°C for 5 minutes. The Relative Fluorescence Unit, Ct values, amplification and melting curves have been checked during the qPCRs to ensure the amplification of target fragment and attachment of the indexes. The indexed products were purified and normalized using SequalPrep™ Normalization kit (Thermofisher) prior pooling. A test run has been performed on pooled libraries consisting of 47 samples (15 samples * 3 seasons plus a positive and a negative control) using MiSeq Reagent Nano Kit v2 (250 cycles, 1 million paired end reads) followed by a final run of 25 million reads using v3 reagent kit (300 cycles, paired end) on MiSeq platform.

iii. Bio-informatic processing

The de-multiplexed NGS reads have been trimmed for primer sequences using BBmap (sourceforge.net/projects/bbmap/). Further the Illumina reads have been de-noised, truncated, merged to make contigs, filtered by length and quality scores, chimera detected and de-replicated to ASVs using DADA2 pipeline (Callahan et al., 2016). A costume script (SGN Metabarcoding pipeline) has been used to blast the ASVs against the NCBI database incorporating BLASTn pipeline. The sequences of ten best blast hits were retrieved and pooled with the GEANS reference library v4 and this merged dataset was used as database (db) to assign the best and closest taxonomic assignment to each ASV including the percentage identity, query

coverage, length of the fragment, GenBank/reference library accession number and number of reads per ASV. The taxonomic assignments of ASVs (which have been retrieved from GenBank) were checked against WoRMS to ensure accurate taxonomic assignment. The target macrofauna ASVs have been further aggregated into potential species using a similarity threshold of 97% (Neighbor Joining method). Only for the comparison between morphology and DNA-based identification the unique species names have been considered (the aggregated species have been merged into unique taxonomic species names).

iv. Data-analyses

The community analyses were performed at species (aggregated ASVs at 97% similarity) levels using different R packages: vegan, Limma, dada2pp, pairwiseAdonis and DECIPHER. Compositional differences of macrobenthos communities were examined between the five stations and three seasons illustrated by non-metric MDS plots applying Hellinger dissimilarity index and logarithmic transformation. Permutational multivariate analysis of variance was performed to investigate differences between seasons using the Hellinger distance applying logarithmic transformation and euclidean method. Multilevel pattern analysis using "Indicspecies" R package was performed to assess the strength and statistical significance of the relationship between species occurrence/abundance and groups of sites (here morphology and metabarcoding results).

c. Case study on impact assessment of mechanical lugworm harvesting in the Wadden Sea (The Netherlands)

i. Design and monitoring objective

This study was set up to assess the impact of mechanical lugworm (*Arenicola marina*) harvesting in the Wadden Sea on macrobenthos communities. During high tide, a trawler removes and sieves the top layer of sediment along a transect, using their anchor as a winch. Because the trawler removes the top layer of sediment and discards it back into the sea after filtering, the seabed is turned, which may leave macrobenthos vulnerable to predation or buries them under sediment. Such disturbances may be lethal for several species, such as the common cockle (*Cerastoderma edule*), which can only survive in the upper 10 cm of the seabed. Previous studies have shown changes in macrobenthos communities in both species abundance and composition after physical disturbance. From dredging studies, where effects have been extensively studied, it is known that dredging is followed by a decline in species numbers, population density and biomass of macrobenthos (Newell et al., 1998; Queirós et al., 2006; Piló et al., 2019). Shortly after dredging, opportunistic species with a relatively fast reproduction cycle and growth are prevalent. Species with a longer lifespan and slow growth need more time for recovery, and recolonization in areas of low current velocity can take up to five to ten years (Van der Veer et al., 1985; Newell et al., 1998). The aim of this case study was therefore to analyse the impact of lugworm harvesting on benthic diversity using morphology based and bulk DNA-based analyses. For the latter, a detailed evaluation of the best primer set to use was conducted using mock communities of known macrobenthos composition.

ii. *Collection of samples*

Three pairs of transects were sampled near the island of Texel in the Dutch part of the Wadden Sea (Fig 5), where each transect pair consisted of one transect that was undisturbed and one transect that was used for mechanical lugworm harvesting. Sampling was carried out by the Royal Netherlands Institute for Sea Research (NIOZ), using core samples (177cm2, 25-30cm depth), which were sieved on a 1 mm sieve. Along each sampling transect, a total of six samples were collected. The transects were sampled a total of seventeen times, starting on 20-3-2016 (the morning before the lugworm harvesting) with intervals ranging from a few days to several months (Fig 5). Sieved macrobenthos samples were freeze dried using the Alpha 2-4 LSCbasic (Martin Christ) and stored at -30°C. Of each transect, three samples were morphologically identified by NIOZ.

Sampling location and dates

T0	20-3-2016		
T1	21-3-2016	**T9**	4-7-2016
T2	31-3-2016	**T10**	31-8-2016
T3	5-4-2016	**T11**	14-11-2016
T4	11-4-2016	**T12**	13-3-2017
T5	20-4-2016	**T13**	9-5-2017
T6	9-5-2016	**T14**	23-5-2017
T7	23-5-2016	**T15**	6-6-2017
T8	6-6-2016	**T16**	26-6-2017

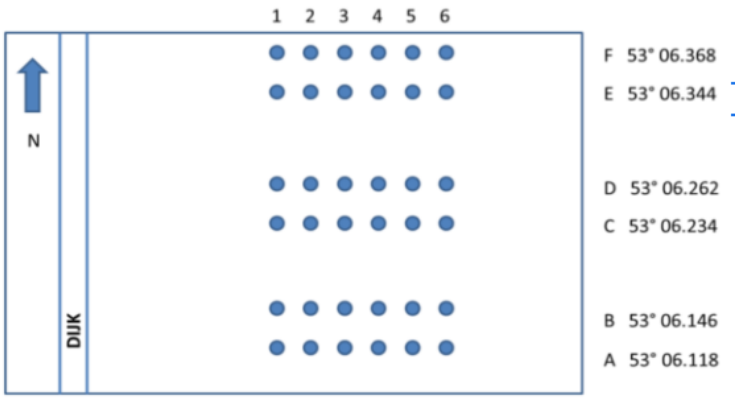


Figure 5. Sampling moments and transect locations of the coast of Texel.

ii. *Lab processing*

The freeze dried samples were homogenized using liquid nitrogen and a mortar and pestle before DNA extraction, which was performed in triplicate using the MagAttract PowerSoil DNA Kit (Qiagen), according to the manufacturer protocol. The DNA extraction triplicates were then combined and measured on the DropSense 96 (Trinean). DNA was further cleaned with the OneStep PCR Inhibitor Removal Kit (Zymo Research), again following the manufacturer protocol, and measured on the DropSense 96 again.

PCR was performed using the primer combination mlCOIintF (Leray et al., 2013) and jgHCO2198 (Geller et al., 2013). PCR amplification was performed using different polymerases; KAPA HiFi HotStart ReadyMix (Roche), Taqman Environmental Master Mix 2.0 (ThermoFisher), and Phire II Hotstart (ThermoFisher). Further processing for Next-Generation Sequencing was not performed, as no usable PCR products were obtained from these bulk DNA extracts.

Various COI primer sets were tested on mock communities prepared from DNA extracts from recent collections deposited at the Naturalis Biodiversity Center. A total of 50 commonly occurring species were collected, including 33 BISI (Bentic Indicator Species Index) species. For each species, DNA was normalized to 5ng/µl and combined, in order to create a mock sample with equal proportions of all DNA extracts. A second mock sample was created the same way, except that eight species were diluted 1000 times before adding to the sample, to simulate rare species. These mock samples were amplified using different primer combinations from the literature, which were evaluated *in silico* before being applied to the samples. All species in the mock samples had been sequenced prior to this study, so COI barcodes were available for each of them. Five different forward primers were tested *in silico*: BF1, BF2 (Elbrecht & Leese 2017), BF3 (Elbrecht et al., 2019), mlCOIintF (Leray et al., 2013) and mlCOIintF-XT (Wangensteen et al., 2018). In the amplification evaluation, the standard forward primer was compared to the best primer according to the *in silico* test (mlCOIintF and mlCOIintF-XT, respectively), which were combined with jgHCO2198 (Geller et al., 2013) and Fol-Degen-Rev (Yu et al., 2012). A dual indexed MiSeq amplicon library was prepared using two rounds of PCR amplifications, using Nextera-tailed primers for the first round. All PCRs in the first round contained 20 µl reactions with 10 µl 2x TaqMan Environmental Master Mix 2.0 (Thermo Fisher), 1 µl of template, 7 µl MiliQ and 1 µl (10 pMol) of both a forward and a reverse primer. Initial denaturation was done at 96 °C for 10 min, followed by 30 cycles of 96 °C for 30 s, 50 °C for 30 s, 72 °C for 20 s and followed by a final elongation 72 °C for 7 min. Twelve replicates were performed per mock with 1 µl template DNA, and one additional PCR was performed for each mock with 12 µl of template DNA. To ensure the PCR was successful the product was checked on an E-Gel 96 pre-cast agarose gel (Thermo Fisher), and the PCR products were cleaned with NucleoMag NGS-Beads (Machery-Nagel) using a 1:0,9 ratio. The second round of PCRs was done in 20 µl reactions using 10 µl 2x TaqMan Environmental Master Mix 2.0 (Thermo Fisher), 1 µl of MiSeq Nextera XT labels, 4 µl PCR product and 4 µl MiliQ. The initial denaturation was performed at 96 °C for 10 min, followed by 8 cycles at 96 °C for 30 s, 55 °C for 60 s, 72 °C for 30 s and followed by a final elongation at 72 °C for 7 min. PCR products were then cleaned with NucleoMag NGS-Beads using a ratio of 1:0,9 and then quantified on the QIAxcel (Qiagen) and pooled equimolarly with the QIAgility (Qiagen). The pool was quantified on the Bioanalyzer 2100 (Agilent Technologies) with the DNA High Sensitivity Kit, and run on the Illumina MiSeq at BaseClear (Leiden, the Netherlands).

iii. Bioinformatic processing

No bio-informatic processing and analyses were performed for the impact study, due to failure of the PCR amplification. Consequently, a comparison between morphology based and bulk DNA-based biodiversity analyses could not be performed.

For the mock samples in the primer and PCR evaluation, filtering and clustering of the raw data obtained from the Illumina MiSeq was performed in the bioinformatics pipeline of Naturalis Biodiversity Center through a Galaxy instance. Raw sequences were merged using FLASH and all non-merged reads were discarded. The reads were split into the four different primer combinations and the primers trimmed from the merged reads using Cutadapt. All reads with primers not present or not anchored were discarded. Sequences with a length below 310 bp and above 316 bp were discarded using PRINSEQ. Reads were clustered into OTUs with UNOISE2 with an alpha value of 0.5. A 0.02% threshold was used to omit spurious reads from the OTU tables. A custom reference was created from the individually sequenced specimens that were

used for the mock samples. Comparison between the mock samples and the custom reference was done with BLASTn (using 100% identity match).

d. Case study on MSFD monitoring in the North Sea (Denmark)

i. Design and monitoring objective

Danish monitoring conducted for the Marine Strategy Framework Directive (MSFD) in the North Sea (NOVANA) consists of 20 sets of sampling grids of 2 to 6 km², distributed evenly over the Danish part of the North Sea (Fig 6). Biannual sampling of 10 grids each year is conducted using conventional morphology based analyses. The NOVANA program has been running from 2015 to 2021. The sediment composition at the sampling sites ranges from coarse sand and gravel to fine sand and muddy sand at the deepest westernmost stations. Depth ranges from 19 to 79 m. The sample stations can be broadly divided into two categories based on depth. Sediment in coastal stations (<44 m depth) are likely more mixed (mix) than offshore stations (> 44 m depth, bottom).

The pilot study in the Danish part of the North Sea was designed to investigate if and how eDNA-based analysis of soft bottom sediments can supplement the on-going conventional Danish North Sea monitoring program (NOVANA). Specifically, the pilot investigated to what extent species identification by sediment eDNA metabarcoding represents the benthic fauna identified by conventional methods. Furthermore, our analysis was designed to investigate patterns in species composition with increasing depth in the North Sea.

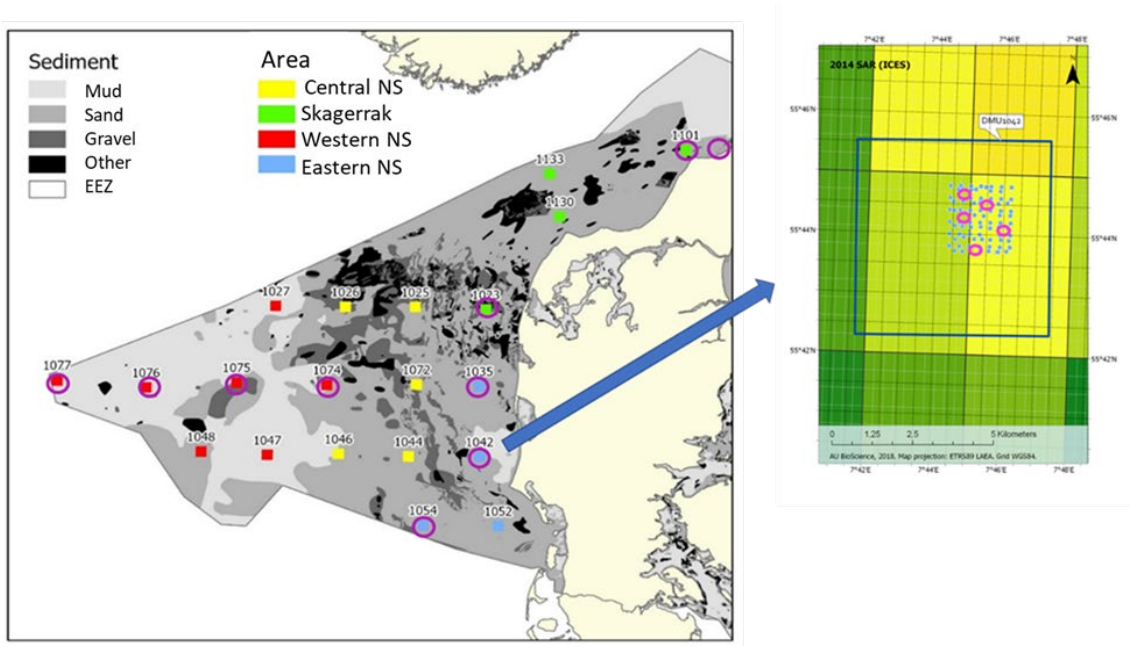


Figure 6. Left panel: map of the NOVANA sampling sites where soft bottom fauna is sampled as part of MSFD-monitoring. The stations sampled in April 2019 are encircled with pink. The gray shades indicate the substrate while the color of the sampling sites indicates how the infauna community groups together in different

areas. Right side: Example of a station grid with 42 individual sample locations (blue dots) and the subset of samples selected for eDNA analyses (purple circles).

i. Collection of samples

In April 2019, the morphology based monitoring program focused on 11 grids where sediment was collected in 42 stations within each grid (Fig 6). The field sampling uses a Haps bottom corer (Kanneworff & Nicolaisen 1973) covering a sampled area of the sea floor of 0.0143 m^2. The Haps corer is further equipped with a vibrator to ensure penetration to about 20 cm into the sediment. All sampling of the macrofauna follows a set of national guidelines (Hansen & Josefson 2014), in line with corresponding guidelines under the Sea Conventions OSPAR and HELCOM. In short, the sediment collected by the Haps corer is sieved onboard the ship using a 1-mm sieve and all retained material are preserved with 70 % ethanol (final concentration) for later analysis in the laboratory. The material is sorted under the microscope where all animals are identified to the lowest possible taxon leading to a final data format of species-specific abundance and wet weight for each individual sample and is stored in the National Marine Database.

Sediment samples for eDNA analysis were collected from the same Haps cores that were part of the morphological monitoring program and subjected for later analysis of the infauna community. Sediments were sampled from five random samples in all 11 grid areas. Upon retrieval of the Haps core, a small subsample (20 ml, sediment surface area of 2.2 cm^2) of the surface sediment was sampled with a 20-ml open-cut syringe that penetrated about 9 cm into the sediment (Fig 7). The syringe samples were immediately stored at -20°C, while the rest of Haps sample was processed for macrofaunal analysis as described above. Hence, for each Haps core one syringe sample was analyzed, resulting in five replicates of both eDNA and conventional species identification within each of the 11 grid areas. For eDNA and conventional monitoring comparison, we found syringe samples collected from haps cores for 9 grid areas out of 11, thus we only used 9 grid areas (stations) for diversity based analysis.

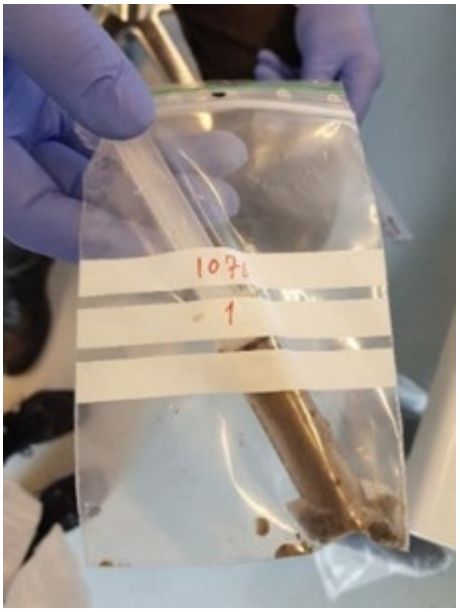


Figure 7: Example of intact sediment sample within syringe.

ii. Lab processing

Sediment in the syringes was sliced into five depth layers: 0-2 cm, 2-4 cm, 4-6 cm, 6-8 cm and > 8 cm. In order to avoid mixing of the sediment profile, slicing of the sample was completed before the sediment was fully thawed. Each slice was weighed before DNA extraction, and DNA was extracted from 0.250 g sediment using the PowerLyzer UltraClean Microbial DNA isolation kit (Qiagen) according to the manufacturer's instructions. The DNA concentration was measured with a Qubit® Fluorometer (Thermo Fisher Scientific) and stored at -20°C until the amplicon libraries were prepared. All DNA extractions and slicing of the samples were done in a clean DNA-laboratory inside a clean flow bench. Further measures to avoid contamination and mixing of DNA between samples involved alcohol and flame sterilization of scalpels used to slice the sediment and changing of gloves between each handled syringe following general recommendations.

Two sets of primers targeting the mitochondrial COI region and the ribosomal 18S rDNA were used for library preparation (Leray et al. 2013; Stat et al. 2017). The initial PCR amplicons were generated using 25 µL reaction mixture consisting of 0.25 µL PCRBIO HiFi Polymerase (2U/µL), 5 µl PCRBIO reaction buffer (PCR Biosystems), 0.5 µl BSA, 16.25 µL water, 0.5 µL of of 10µM forward and reverse primers (10µM) and 2 µL of DNA template. For COI region, the PCR thermocycler program included an initial denaturation at 95°C for 2 minutes, followed by 35 cycles of 95°C 30 sec, 48-52°C for 30 sec, 72°C for 1 min and a final elongation at 72°C for 4 min followed by 4°C for 10 minutes. Multistep annealing temperature starting from 48°C for 10 cycles, followed by 50°C for 10 cycles and by 52 °C for 15 cycles was used for the COI region. Thermal cycles were performed similarly for 18S rDNA, with the exception that the multistep annealing temperature started from 50°C for 10 cycles, followed by 52°C for 10 cycles and 54°C for 15 cycles. This was followed by a 15-cycle indexing PCR, during which unique index combinations (i7 and i5) and adaptors were added. The index PCR included 5 µl of PCR 1 product and primers with Illumina adaptors and in house dual index combinations using the 25 µl amplification mixture containing 0.25 µL PCRBIO HiFi Polymerase (2U/µL), 5 µl PCRBIO reaction buffer (PCR Biosystems), 2 µl of 10µM forward and reverse primers, 10.75 µL water and 5µl template DNA. The index PCR program included 98°C for 1 min, 13 cycles of 98°C for 10 sec, 55°C for 20 sec and 72°C for 40 sec and a final elongation of 72°C for 4 min. The PCR amplification was confirmed using 1.5% agarose gels and SYBR green staining. Subsequently, the amplicon products were cleaned using HighPrep™ magnetic beads (MagBio Genomics Inc. Gaithersburg, USA), according to the manufacturer's instructions. DNA concentrations were measured using Qubit 4.0 (Thermo-Fischer Scientific) with the High-Sensitivity DNA assay. Finally, amplicons were equimolarly pooled for equal representation in the sequencing library, and sequenced on three Illumina Miseq runs (2*250bp, Department of Environmental Science, Aarhus University, Denmark).

iii. Bioinformatic processing

The DADA2 plugin in QIIME2 was used with default parameters except reads trimmed for primer sequence, and reads truncated after 230 bp (Callahan et al., 2016; Bolyen et al., 2019). For 18S rDNA, the resulting amplicon sequence variants (ASVs) were classified using the QIIME2 naive Bayes classifier trained on 99% Operational Taxonomic Units (OTUs) from the SILVA rRNA database (v. 138) after trimming to the primer region (Quast et al., 2013). COI ASVs were assigned using the naive Bayes classifier against MIDORI2 merged with GEANS reference database (v4). Singleton ASVs and ASVs found in only one sample were filtered out before taxonomic assignment. Blast taxa with high similarity and coverage (>97%) at species were used. The ASVs for each slice of individual syringe samples were merged for further downstream analyses, giving five biological replicates for each of nine grids.

iv. Data-analyses

The ASV table and taxonomy files were imported into R and statistical analyses and data visualizations were performed in v.4.2.1 (R Core Team, 2022). Diversity based analysis (non-metric multidimensional scaling and rarefaction analyses) was carried out using the 'vegan' package (Oksanen et al., 2019) and 'phyloseq' (McMurdie et al., 2013). Patterns in community composition and alpha diversity (the number of observed species and Shannon diversity index) were investigated for the COI and 18S datasets separately. In addition we investigated similarities in species composition using presence-absence data. Here we merged the ASV table (from 18S and COI) at species level using the *tax_glom* function from the 'phyloseq' package. Only metazoan taxa were kept in the final species table. The significance of differences between alpha diversity was evaluated using a non-parametric Kruskal-Wallis test followed by Tukey HSD test for pairwise comparison. To investigate patterns in species composition, we produced ordination plots using non-metric multidimensional scaling (nMDS). To investigate differences in species composition among sample stations, we applied a non-parametric analysis of variance (ANOSIM). Community analysis was performed using Presence/Absence data, using the PRIMER software (Clarke and Gorley 2006). A pairwise PERMANOVA test was performed in order to identify significant differences between stations using the 'pairwise Adonis' package in R. Benthic species associated with depth (indicator species) were calculated using the *multipatt* function from 'indicspecies' R package. Furthermore, we used SIMPER analysis for the conventional and eDNA data using *simper* function from vegan. Presence-absence data was used to perform indicator species and SIMPER analysis.

e. *Effect of sample fixative (formalin vs ethanol) on macrobenthos diversity*

For two case studies, replicate samples have been collected from the same stations and were fixed with either formalin or ethanol. Both sample types have been identified by morphology, and the ethanol samples also by bulk DNA analysis. This allowed comparing the effect of different fixatives on the species identification of macrobenthos samples.

For case study "a" focusing on sand extraction impact, replicate Van Veen grabs from the same locations in the Thorntonbank sampled in 2019 have been fixed with ethanol or formalin. Differences in taxonomic composition between the three methods were investigated with an ANOVA with main factors methodology (three levels: bulk DNA, morphFORMOL, morphETOH) and impact (four levels: Reference, Low, Medium, High). The species detected for each method were visualized with a Venn diagram. Community composition obtained with the different methods were visualised with nMDS plots based on Bray-Curtis dissimilarity in R. A two-way PERMANOVA was performed with main effects 'identification method' (three levels: morphology fixed on ethanol, morphology fixed on formalin, bulk DNA fixed on ethanol) and impact (four levels: reference, low, medium, high), performed with 9999 permutations. Homogeneity of dispersion between samples was checked using distance dispersion test and permutest with the package "vegan" v2.5-6.

For case study "b" focusing on the long term monitoring of Norderney Island, the effect of sample fixatives (ethanol vs formalin) was investigated using a subset of stations for which species identification was available for three methods (bulk DNA from ethanol, morphology from formalin and morphology from ethanol) in autumn and spring. The number of species and

Shannon index were calculated and significant differences were assessed using a two-way ANOVA (factors season and method). The number of shared and unique species for each method/fixative was visualized with a venn diagram. Community differences were assessed using PERMANOVA, and visualized using an nMDS plot using presence/absence of species.

f. Calculation of cost and time for DNA-based and morphology based sample processing

Three case studies tracked time and costs of processing the samples by morphology or by DNA-based methodology. For case study "a" (aggregate extraction), time and cost tracking was done on one set of samples i.e. the 24 ethanol preserved samples taken on the Thorntonbank in 2019. For case study "b" (long-term monitoring), time and cost calculations were done based on two sets of five ethanol-preserved samples collected in September 2019 and October 2020 for the morphological and bulk DNA analyses. For the bulk DNA analysis an extra set of five ethanol preserved samples collected in June 2020 was also taken into account. For case study "d" (MSFD monitoring), time and cost tracking was done as averages of 40 samples taken during 2019.

Since the first two steps of sample processing (decantation and screening for remaining specimens) are similar for the bulk DNA and morphology based analyses, time-tracking started at the identification step for the morphology-based method and at the mixing step for DNA-based analyses. For case study "d", the first two steps of sample processing are different, since decantation and screening for remaining specimens are not needed for the eDNA analyses of the sediment. For this case study, time tracking of the morphology based method was done as for the two other case studies, but started from the DNA extraction step for the eDNA-based analyses. For all methods, time tracking was done up to acquiring the species list with abundance and biomass or read numbers. For the morphology-based analyses, identification time followed by weighing the species was tracked for each individual sample. Input of the data in a database was timed in batch. For the bulk DNA-based analyses, mixing time was tracked per sample while all other steps (DNA extraction, PCR amplification, library preparation and DADA2 pipeline) were each timed for the entire set of samples. The different timings were summed for all methods as total time in hours and divided by the number of samples to get time per sample. Costs were calculated starting from the lab processing. Ship and sampling costs have not been included since these are similar for both methods. The costs included both consumables/equipment and personnel costs for each method. For personnel costs, the hourly rate of the executing person was multiplied by the total time spent on each method respectively.

4. Results

a. Case study on environmental impact assessment of marine aggregate extraction (Belgium)

This pilot study investigated whether different identification methods (bulk DNA metabarcoding or morphology) yield similar results for monitoring sand extraction activity. ANOVA showed no significant interaction effect between the methodology and sand extraction impact for the number of taxa (Chisq=1.59, p-value=0.66) and for the Shannon diversity index (Chisq= 2.28, p-value=0.52) (Table 1) indicating that both methods detect similar patterns in alpha diversity related to sand extraction activity. The bulk DNA dataset detected a significantly higher number

of taxa (mean=12.9 ±4.9 taxa) than the morphology-based dataset (mean=10.9 ± 8.3 taxa). Also, a significant effect of the sand extraction impact was observed, with the medium impacted group showing higher number of taxa (mean=14.8 ± 10 taxa) compared to the reference group (mean=10.43 ± 4.5 taxa) (Table 1). Shannon indices were significantly higher using morphological identification (mean = 2.1 ± 0.6) compared to bulk DNA (mean=1.8 ± 0.5) (Table 1), and showed a significant effect of the sand extraction impact, with higher Shannon indices for medium (mean= 2.1 ± 0.7) and high (mean=2.1 ± 0.6) impacted locations compared to the the reference (mean=1.8 ± 0.5) and low (mean=1.9 ± 0.5) impacted groups when averaged over year and method.

When looking at the sand extraction impact on the different sandbanks for each method separately, alpha diversity again yielded comparable results for both methods, except for a borderline significant interaction between sandbank and impact group for number of taxa for morphology (Chisq=12.56, p-value= 0.049) that was not apparent for bulk DNA (Table 2). Nevertheless, for both methods, a significant difference in the number of taxa was observed between sandbanks (Table 2), with more species being found on the Thorntonbank compared to the two other sandbanks (Fig 8). Similarly, the Shannon index showed a significant effect of sandbank for both methods (Bulk DNA: Chisq=20.32, p-value=3.86e-05; morphology: Chisq=43.38, p-value=3.81e-10), and no effect of sand extraction impact (Bulk DNA: Chisq=4.42, p-value=0.22; morphology: Chisq=7.64, p-value=0.054) (Table 2).

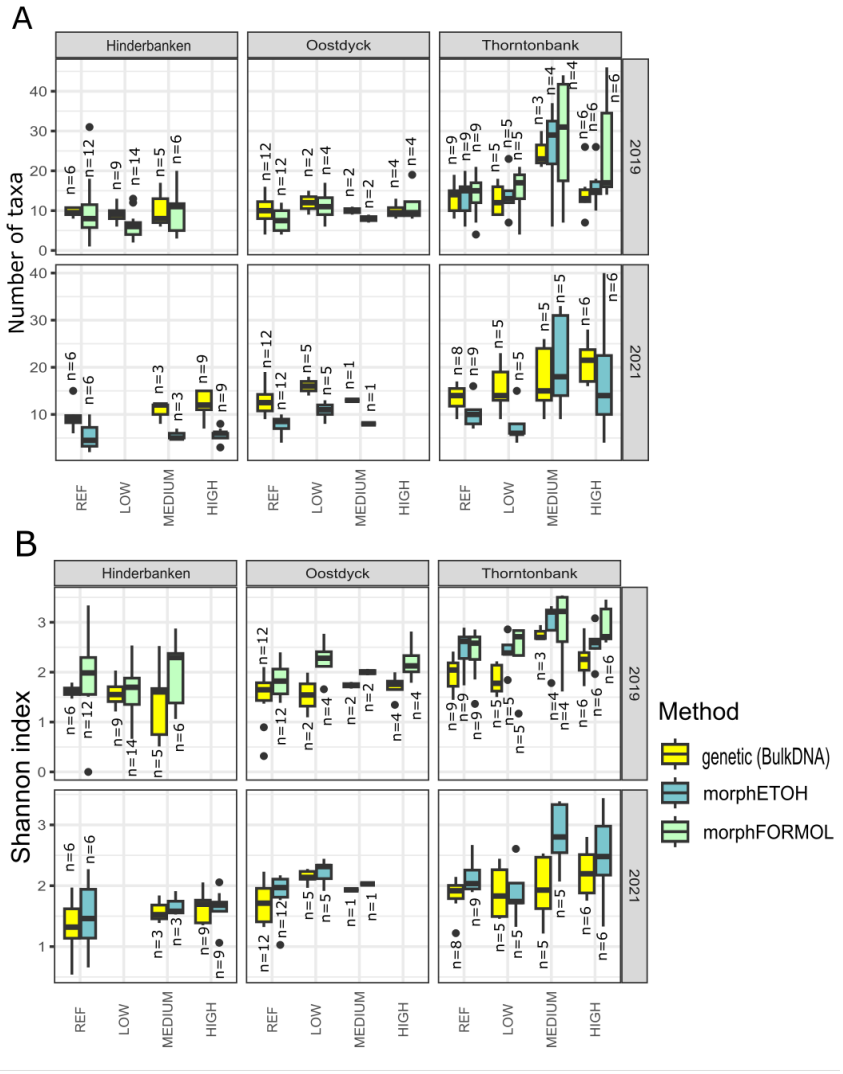


Figure 8: Alpha diversity with the median and 25 and 75 percentiles of the number of species (A) and Shannon index (B) per concession zone and per year for each impact group (reference, low, medium and high) determined using bulk DNA metabarcoding (yellow), morphological identification from ethanol (blue) and from formalin fixed samples (green). In some impact groups, less samples were used for the bulk DNA datasets due to low sequencing depth. The number of replicates is indicated above the boxplots.

Table 1: Output of the mixed ANOVA for number of species and Shannon index, with fixed factors Identification method and impact and the interaction method*impact. Significant values are indicated in bold.

	Number of species			**Shannon Index**		
	Chisq	Df	Pr(>Chisq)	Chisq	Df	Pr(>Chisq)
(Intercept)	163.1733	1	< 2.2e-16	93.0508	1	< 2.2e-16
Impact	10.9764	3	**0.0119**	9.4698	3	**0.0237**
Identification_method	28.2589	1	**1.061e-07**	31.1353	1	**2.406e-08**

Impact:Identification_method	No significant interaction, so removed from model	No significant interaction, so removed from model

Table 2: Output of the mixed ANOVAs for number of species and Shannon index in the bulk DNA and morphology-based dataset, with fixed factors impact and sandbank and the interaction impact*sandbank. Significant values are indicated in bold

Number of species						
	Bulk DNA metabarcoding			morphology		
	Chisq	Df	Pr(>Chisq)	Chisq	Df	Pr(>Chisq)
(Intercept)	844.7042	1	< 2.2e-16	201.5305	1	
Impact	4.8288	3	0.1848	2.7154	3	0.4376
Sandbank	34.4708	2	**3.272e-08**	8.2060	2	**0.0165**
Impact:Sandbank	No significant interaction, so removed from model			12.6490	6	**0.0490**
Shannon Index						
	Bulk DNA metabarcoding			morphology		
	Chisq	Df	Pr(>Chisq)	Chisq	Df	Pr(>Chisq)
(Intercept)	36.2278	1	**1.755e-09**	215.1946	1	**< 2.2e-16**
Impact	4.4230	3	0.2193	7.6404	3	0.0541
Sandbank	20.3225	2	**3.864e-05**	43.3780	2	**3.807e-10**
Impact:Sandbank	No significant interaction, so removed from model			No significant interaction, so removed from model		

.

The species identity was substantially different between bulk DNA metabarcoding and morphology since only 27.5% and 37.3% for the Hinderbanken, 26.5% and 36.4% for the Oostdyck and 28% and 37.6% for the Thorntonbank of the total detected species for 2019 and 2021 respectively, were shared between the two methods (Fig 9). Most of the uniquely found taxa with the morphological method were classifications at a higher taxonomic level (13 and 6 for the Hinderbanken, 13 and 8 for the Oostdyck, 29 and 15 for the Thorntonbank, in 2019 and 2021 respectively) or, if identified at species level, were Polychaeta (14 and 5 for the Hinderbanken, 10 and 9 for the Oostdyck, 27 and 18 for the Thorntonbank respectively). Some unique identifications for the morphology were not present in the reference database (7 and 4 for the Hinderbanken, 6 and 5 for the Oostdyck, 22 and 15 for the Thorntonbank, in 2019 and 2021 respectively) (Table 3). Most of the uniquely found taxa with bulk DNA had low read numbers (<100).

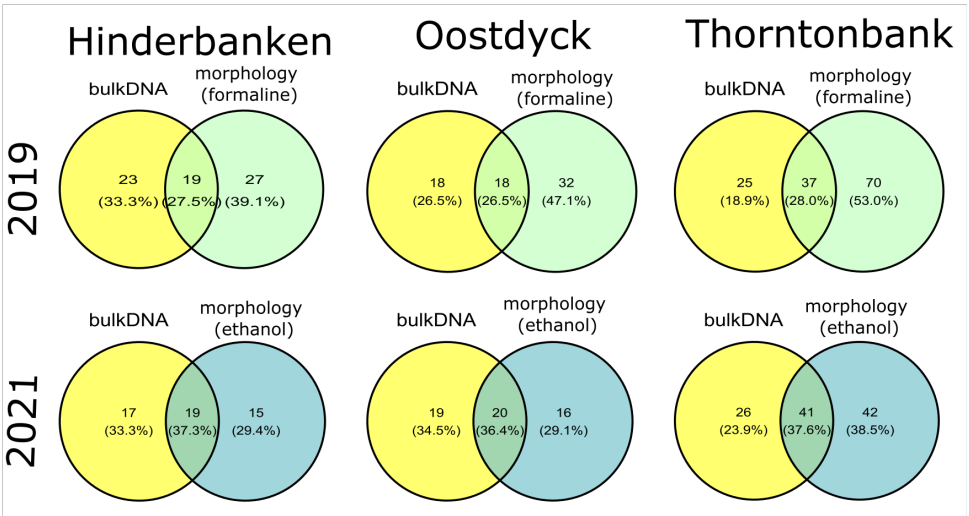


Figure 9: Venn diagrams of the taxa detected with bulk DNA metabarcoding and morphological identification, visualized per sampling year (2019 and 2021) and sandbank (Hinderbanken, Oostdyck and Thorntonbank).

Table 3: Summary of the species uniquely found by one method: identified at a higher taxonomic level in one dataset or availability of a sequence in the reference database.

	Hinderbanken		**Oostdyck**		**Thorntonbank**	
	2019	**2021**	**2019**	**2021**	**2019**	**2021**
Shared species	19	19	18	20	37	41
Unique for morphological method	27	15	32	16	70	42
➔ Higher taxonomic level	13	6	13	8	29	15
➔ No reference sequence available in the used reference database	7	4	6	5	22	15
➔ Phylum = Mollusca	5	0	7	2	13	3
➔ Class = Polychaeta	14	5	10	9	27	18
Unique for Bulk DNA metabarcoding method	23	17	18	19	25	26
➔ Low read numbers (<100)	14	13	14	14	11	20
➔ Higher taxonomic level present in morphological identified samples	2	2	4	3	11	7

Community PERMANOVA results were comparable between the bulk DNA and morphological identification method for the Oostdyck, where a significant effect of the main factors year and impact was observed (Table 4). The high impact samples cluster separately from the reference and low impact samples, but no significant post hoc tests for the bulk DNA were obtained (Fig 10). For the Hinderbanken, only a significant effect was seen for the factor year using the bulk DNA metabarcoding method, while for the morphological identification method also a significant effect of the main factor impact was observed. Posthoc tests showed that high impact samples were significantly different from the reference and low impact samples. A significant interaction effect was observed for the Thorntonbank with the bulk DNA metabarcoding method while this was (borderline) non significant for the morphology-based method. The main

factors year and impact were significantly different with both methods (Table 4). Significant PERMANOVA results of factor impact in the Oostdyck (both methods) and Thorntonbank (morphological method), the interaction impact*year in the Thorntonbank (bulk DNA) and the factor year in the Hinderbanken were caused by location as well as by dispersion effects, as significant permdisp results were seen (Table 4).

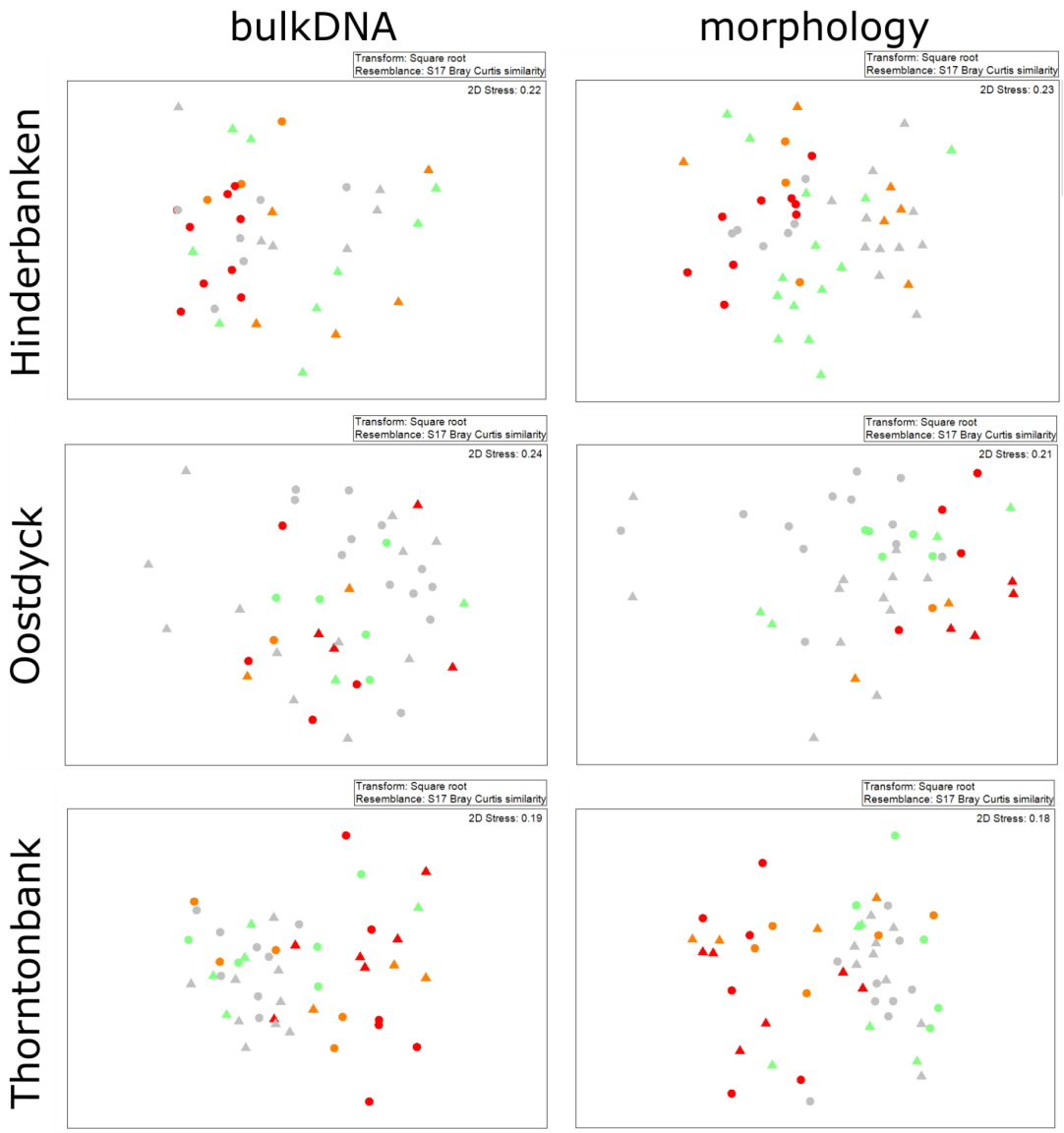


Figure 10: nMDS plots per sandbank (Hinderbanken, Oostdyck and Thorntonbank) and method (left: bulk DNA metabarcoding, right: morphology). The different colors indicate the sand extraction impact (REF: grey, LOW: green, MEDIUM: orange, HIGH: red) and the two years are visualized by a different symbol (2019: triangle, 2021: circle).

Table 4: Output Permdisp and PERMANOVA with main factors year (2019 and 2021) and impact (reference, low, medium, high) per sandbank for bulk DNA and morphology separately.

	Bulk DNA metabarcoding						Morphology					
Oostdyck	**Permdisp**	**Df**	**SumOfSqs**	**R2**	**F**	**Pr(>F)**	**Permidsp**	**Df**	**SumOfSqs**	**R2**	**F**	**Pr(>F)**
Year	0.2375	1	0.4628	0.0431	1.9059	**0.0471**	0.1247	1	0.7863	0.0801	3.9793	**0.0001**
Impact	**0.0083**	3	1.1836	0.1102	1.6249	**0.0251**	**0.0142**	3	1.4495	0.1476	2.4453	**0.0002**
Year:Impact	**2,00E-04**	3	0.8366	0.0779	1.1485	0.2714	**8,00E-04**	3	0.4719	0.0481	0.7962	0.794
Residual		34	8.2553	0.7688				36	7.1133	0.7243		
Total		41	10.7383	1.0000				43	9.8209	1		
Hinderbanken		Df	**SumOfSqs**	**R2**	**F**	**Pr(>F)**		**Df**	**SumOfSqs**	**R2**	**F**	**Pr(>F)**
Year	**1,00E-04**	1	1.1727	0.1248	4.9895	**0.0001**	**9,00E-04**	1	1.2137	0.0911	5.0392	**0.0001**
Impact	**0.0016**	3	0.5617	0.0598	0.7966	0.7419	0.1594	3	1.3862	0.1041	1.9186	**0.0016**
Year:Impact	**0.0029**	1	0.1420	0.0151	0.6042	0.7922	0.1972	1	0.1188	0.0089	0.4931	0.9312
Residual		32	7.5209	0.8003				44	10.5971	0.7958		
Total		37	9.3973	1.0000				49	13.3157	1		
Thorntonbank		Df	**SumOfSqs**	**R2**	**F**	**Pr(>F)**		**Df**	**SumOfSqs**	**R2**	**F**	**Pr(>F)**
Year	0.2636	1	0.5775	0.0482	2.6233	**0.0091**	0.5275	1	0.7005	0.0526	3.0498	**0.0014**
Impact	**0.0164**	3	1.7992	0.1500	2.7244	**0.0003**	**0.0389**	3	2.2588	0.1696	3.2778	**0.0001**
Year:Impact	**0.0293**	3	1.0302	0.0859	1.5600	**0.0385**	0.7359	3	0.9440	0.0709	1.3699	0.0728
Residual		34	8.5852	0.7159				41	9.4177	0.7070		
Total		41	11.9920	1.0000				48	13.3211	1		

SIMPER results showed that the number of species contributing most to within-group similarity of the different extraction impact groups were very similar between the two methods for the three sandbanks. The biggest discrepancy was observed for the medium and high impact zones of the Thortonbank, where more species contributed to within-group similarity for morphology (resp. 14 and 13) compared to bulkDNA (resp. 8 and 10) (Table 5). Additionally, the species contributing most to within-group similarity (i.e. most characteristic species) are similar between both methods. The species with high contributions that were not shared between the method showed rather high read numbers or density, and therefore can not be explained by a lower abundance in the sample (Table 5).

Table 5: Simper analysis showing the species contribution within-group similarity in each impact zone per sand bank (for bulk DNA metabarcoding and morphology, in %). Species were included to the list so the total contribution reached 90%. Species that were found in both datasets (bulk DNA and morphology) are marked in blue. For species uniquely found by one method, the read numbers or density is given.

	Hinderbanken				Oostdyck				Thorntonbank			
		Bulk DNA	**morphology**	**Abundance (read numbers OR density)**		**Bulk DNA**	**morphology**	**Abundance (read numbers OR density)**		**Bulk DNA**	**morphology**	**Abundance (read numbers OR density)**
rer	**Nephtys cirrosa**	55.39	57.97		**Nephtys cirrosa**	55.52	32.53		**Nephtys cirrosa**	45.13	28.11	
	Ophelia borealis	19.22	17.78		**Echinocardium cordatum**	11	4.46		**Spiophanes bombyx**	21.9	14.84	
	Scolelepis bonnieri	11.49		24988	**Gastrosaccus spinifer**	5.87	2.77		Scolelepis bonnieri	11.22		18557
	Gastrosaccus spinifer	3.92		5522	**Scolelepis bonnieri**	5.74	3.11		Echinocardium cordatum	7.54		23811
	Bathyporeia elegans		8.07	250	Thia scutellata	5.61		15081	**Urothoe brevicornis**	3.59	7.36	

	Hinderbanken				Oostdyck				Thorntonbank			
	Hesionura elongata		6.74	200	**Spiophanes bombyx**	4.52	5.06		Thia scutellata	2.53		5770
					Bathyporeia elegans	4.11	29.64		Hesionura elongata		25.85	1100
					Hesionura elongata		14.46	2650	Bathyporeia elegans		8.68	470
									Diogenes pugilator		3.53	270
									Gastrosaccus spinifer		3.1	
low	**Average similarity**	**33.22**	**24.87**		**average similarity**	**36.34**	**46.76**		**Average similarity**	**36.02**	**31.71**	
	Nephtys cirrosa	57.91	39.94		**Nephtys cirrosa**	36.99	28.33		**Nephtys cirrosa**	42.93	28.27	
	Scolelepis bonnieri	12.83	23.35		**Echinocardium cordatum**	16.7	7.62		**Spiophanes bombyx**	21.27	7.58	
	Urothoe brevicornis	5.69	18.28		**Scolelepis bonnieri**	9.8	4.1		**Urothoe brevicornis**	9.57	9.92	
	Lagis koreni	5.51			**Bathyporeia elegans**	5.36	22.69		**Thia scutellata**	7.4	4.1	
	Gastrosaccus spinifer	5.15		5765	**Diogenes pugilator**	5.27	3.85		Ophelia borealis	6.91	5.27	
	Ophelia borealis	3.34	4.69		**Ophelia borealis**	5.09	4.54		Scolelepis bonnieri	2.54		2836
	Pseudocuma simile		4.62	110	Thia scutellata	3.74		7874	Gastrosaccus spinifer		6.98	60
					Spiophanes bombyx	3.72	3.51		Bathyporeia elegans		5.18	220
					Megalurop us agilis	3.7	3.73					
medium	Average similarity	33.5	30.38		average similarity	50.19	59.49		Average similarity	37.41	35.62	
	Nephtys cirrosa	60.02	43.75		**Scolelepis bonnieri**	62.49	18.49		**Nephtys cirrosa**	46.23	11.72	
	Ophelia borealis	17.49	36.58		**Nephtys cirrosa**	10.97	18.49		**Spiophanes bombyx**	11.86	17.33	
	Scolelepis bonnieri	5.75	5.41		**Urothoe brevicornis**	6.41	11.92		**Lanice conchilega**	9.48	11.4	
	Spiophanes bombyx	5.09		1860	Megalurop us agilis	6.26		168	**Thia scutellata**	7.32	3.2	

	Hinderbanken				Oostgyck				Thorntonbank			
	Gastrosaccus spinifer	4.12		1248	**Ophelia borealis**	4.64	20.51		**Ophiura albida**	5.75	3.88	
	Hesionura elongata		5.39	30	Bathyporeia elegans		23.36	100	Branchiostoma lanceolatum	3.55		505
									Urothoe brevicornis	3.31	16.62	
									Ophelia borealis	2.58		174
									Pseudocuma simile		4.81	150
									Abludomelita obtusata		4.51	320
									Aonides paucibranchiata		4	440
									Pariambus typicus		3.15	130
									Bathyporeia elegans		2.83	130
									Gastrosaccus spinifer		2.53	140
									Echinocardium cordatum		2.24	40
									Poecilochaetus serpens		2.2	240
high	**Average similarity**	**50.36**	**22.13**		**average similarity**	**40.29**	**49.39**		**Average similarity**	**33.85**	**25.2**	
	Nephtys cirrosa	66.44	45.73		**Scolelepis bonnieri**	22.44	6.68		**Lanice conchilega**	17.3	18.83	
	Ophelia borealis	14.34	22.05		**Urothoe brevicornis**	20.97	34.86		**Spiophanes bombyx**	15.29	7.26	
	Gastros accus spinifer	10.24	18.52		**Echinocardium cordatum**	18.24	5.58		**Ophiura albida**	10.62	7.47	
	Bathyporeia elegans		7.28	60	Nephtys cirrosa	11.24		14717	Nephtys cirrosa	10.13		2128
					Thia scutellata	9.98		8149	**Ophelia borealis**	8.77	9.78	
					Bathyporeia elegans	6.55		586	**Branchiostoma lanceolatum**	8.67	5.45	
					Gastrosaccus spinifer	3.83		956	Thia scutellata	7.87	3.19	
					Bathyporeia elegans		33.08	840	**Urothoe brevicornis**	5.16	7.82	

	Hinderbanken				Oostdyck				Thorntonbank			
					Spiophanes bombyx		9.26	120	Lagis koreni	3.86		627
					Diogenes pugilator		3.38	70	**Pisidia longicornis**	3.23	2.06	
									Abludomelita obtusata		10.75	1580
									Hesionura elongata		6.9	90
									Aonides paucibranchiata		5.33	270
									Owenia fusiformis		3.49	50
									Nototropis swammerdamei		2.81	320

In conclusion, bulk DNA metabarcoding and morphology-based identification revealed comparable patterns in alpha and beta diversity related to sand extraction activity. However, bulk DNA detected significantly more taxa and showed significantly lower Shannon diversity than the morphology based method. Additionally, both methods detected very different species since at most 37% of the species were shared between both methods. Despite differences in the number and identity of taxa identified by the two methods, the common characteristic species contributing to within-group similarity (reference, low, medium, high) at each sandbank were similar for the bulk DNA and morphology-based method.

b. Case study on long-term soft bottom monitoring at Norderney (Germany)

In the seasonal sampling of Norderney soft bottom macrofauna, a total of 322 clusters (after 3% cut off) were identified to 151 unique species using bulk DNA metabarcoding of the ethanol samples, while 121 species were morphologically identified from formalin samples and 111 species from ethanol preserved samples (together 140 species from morphology). Both morphology and metabarcoding indicated that collecting more samples would increase the number of captured species (Fig 11)

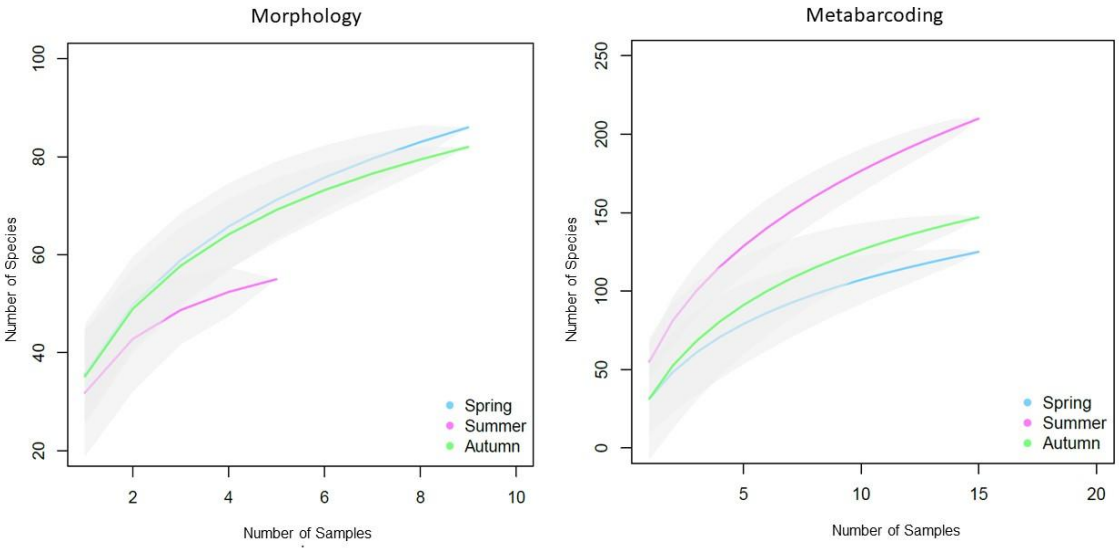


Figure 11: Rarefaction curves based on morphology (left panel) and bulk DNA (right panel) indicating that higher species richness could have been observed by increasing the number of collected samples in each season for both morphology and metabarcoding approach.

Bulk DNA metabarcoding detected a higher number of species in summer compared to the two other seasons, while morphology showed a more or less similar number of species over the seasons (Fig 12). ANOVA results supported these findings and showed that the number of species was affected by a significant interaction between method (bulk DNA/formalin morphology) and season (Table 6). For the Shannon diversity index, no significant interaction between methods and season was observed, but a significantly higher index was achieved for the morphological dataset compared to the bulk DNA dataset (Fig 12, Table 6).

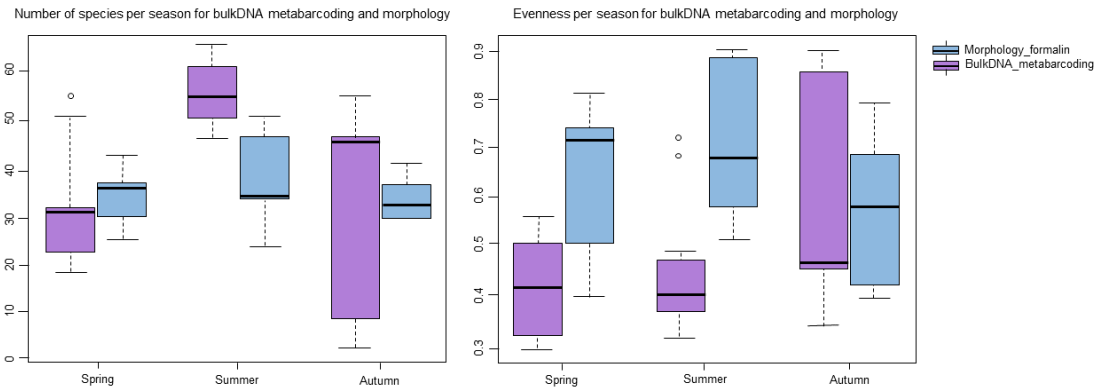


Figure 12: Number of species (left plot) and evenness (right plot) in each season for macrobenthos communities for bulk DNA metabarcoding (45 samples) and morphology (23 samples) from Norderney pilot study. Note that the formalin fixed samples are shown in this analysis, as there is no morphology data available for summer from ethanol preserved samples. The line in each box plot refers to the median.

Table 6: Analysis of variance (two way ANOVA) calculated for shannon index and the number of species (dependent variables) considering methods (bulk DNA metabarcoding and morphology formalin), the three seasons and the interaction of both factors. Only 23 samples/replicates have been analyzed here.

Two-way ANOVA for Shannon Evenness Index						
	Df	Sum	Mean	F	value	Pr(>F)
method	1	5.673	5.673	22.796	2.42E-05	***
season	2	1.263	0.632	2.538	0.0917	.
method:season	2	0.687	0.344	1.381	0.2632	
Residuals	40	9.955	0.249			
Two-way ANOVA for Number of Species						
method	1	28	28.2	0.238	0.62813	
season	2	1762	880.8	7.449	0.00178	**
method:season	2	1081	540.4	4.57	0.01631	*
Residuals	40	4730	118.3			

The number of shared species between stations was comparable between bulk DNA and morphology based analyses (31 vs 36 species, respectively), while bulk DNA metabarcoding showed a higher number of unique species in all stations and seasons compared to morphology based analyses (Fig 13). The number of shared species among the three seasons was higher for the bulk DNA analyzes than for the morphology based analyses (53 vs 46, respectively). This might be partially due to the higher number of replicates (n=45) which have been analyzed for bulkDNA compared to morphology (n=23; summer is presented with only one sample per station). Summer shows the highest number of unique macrofauna species (n=112, in bulk DNA metabarcoding), whereas in morphology spring revealed the highest diversity by means of unique species (n=25).

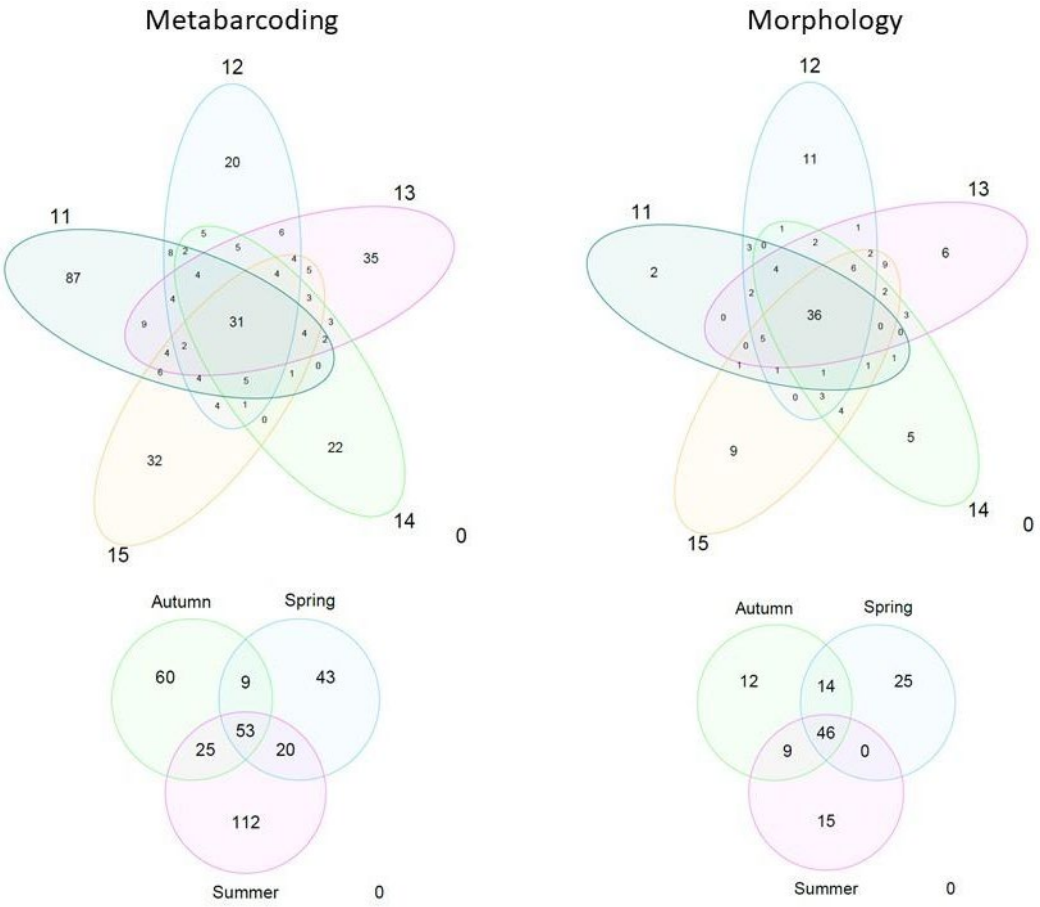


Figure 13: Venn diagrams showing the presence of unique and shared macrofauna species between the five stations (upper plots) and three seasons (lower plots) for both bulk DNA metabarcoding (45 samples) of the ethanol samples and morphology of formalin samples (23 samples) of Norderney pilot study.

When only using the same set samples for which bulk DNA and morphology (formalin) had been applied (23 samples), bulk DNA also showed a higher number of unique species (n=88) compared to formalin samples (n=66). Both methods detected very different species, since only 50 species (25%) are shared between the two datasets (Fig 14).

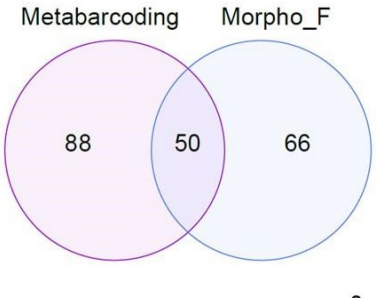


Figure 14: Venn diagram showing the number of shared and unique species between bulk DNA metabarcoding and morphology (formalin) samples from the same stations. For this analysis only 23 samples in which both methods were applied on the same samples have been considered.

The analyses of dissimilarities between the composition of macrofauna communities (Hellinger) in both morphology (formalin, 23 samples) and bulk DNA (45 samples) separated station 15, which is the deepest station, from the other stations in all three seasons (Fig 15). Both methods also revealed significant differences in community composition between the three seasons (Fig 15, Table 7).

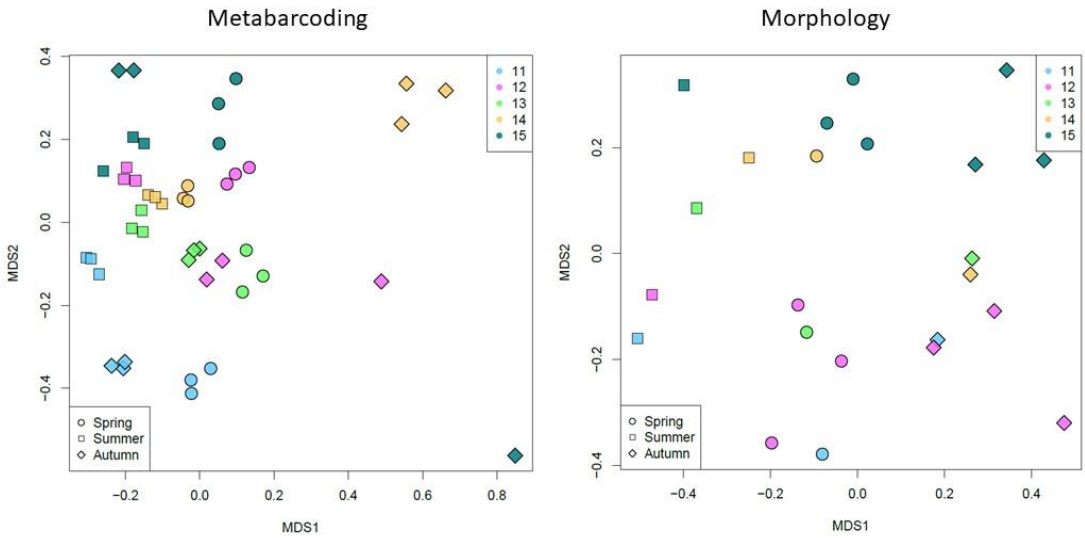


Figure 15: nMDS plot using Hellinger distance and logarithmic transformation comparing the composition of macrofauna among stations and seasons at Norderney island by bulk DNA (left plot, 15 samples with each three DNA extractions) and morphology (formalin, right plot, 23 samples). Both methods support distinct clustering between seasons, while stations showing mixed diversity with exception of station 15 (in both methods) and 11 only by metabarcoding.

Table 7: PERMANOVA comparing the community between seasons and pairwise multivariate analyses of variance using Hellinger distance ((logarithmic transformation) and Euclidean algorithm performing pairwise comparison of the three seasons of Norderney pilot study by bulk DNA metabarcoding and morphology results. Both methods support significant differences between seasons.

Permanova		Df	Sums Of Sqs	MeanSqs	F.Model	R2	Pr(>F)	sig
Morphology	season	2	0.797	0.39852	2.5078	0.2005	0.02	*
	Residuals	20	3.1783	0.15891		0.7995		
	Total	22	3.9753			1		
Metabarcoding	season	2	2.2077	1.10383	3.3749	0.13846	0.001	***
	Residuals	42	13.7371	0.32707		0.86154		
	Total	44	15.9448			1		
Pairwise Permanova		Df	Sums Of Sqs	F.Model	R2	p.value	p.adjusted	sig
Metabarcoding	Spring vs summer	1	2.010036	4.771951	0.1456108	0.001	0.003	*
	Spring vs Autumn	1	1.729555	3.154831	0.1012630	0.001	0.003	*
	Summer vs autumn	1	2.450051	4.871936	0.1482096	0.001	0.003	*
Morphology	Spring vs Summer	1	0.9286681	3.114186	0.2060439	0.003	0.009	*
	Spring vs Autumn	1	1.1121665	3.719627	0.1886256	0.001	0.003	*
	Summer vs Autumn	1	1.2865901	4.375626	0.2672036	0.002	0.006	*

Indicator species analysis showed that bulk DNA analysis had a higher number of indicator species for each season than the morphology based analysis, and only few species are indicator species for both datasets (Table 8). Lack of morphological identifications for certain groups e.g. Bryozoa, Anthozoa, Hydrozoa, Nemertea and Tanaidacea, as well as some juvenile or damaged specimens of Cumacea, Bivalvia, Ophiuroidea, Asteroidea and Polychaeta can explain these differences.

Table 8: Multilevel pattern analyzes detecting indicator species for the two approaches over the 3 seasons applied on 23 samples in which both methods were applied. Species in blue are common indicators in both approaches in each season. Names in red are only identified at high taxonomic level (morphologically) therefore, no certain statements can be provided from them in comparison to species from the same group identified to species level in bulk DNA metabarcoding.

	Metabarcoding						Morphology					
		A	B	stat	p.value			A	B	stat	p.value	
Autunm	Obelia_bidentata	0.999	0.533	0.73	0.005	**	Spisula subtruncata	1	1	1	0.005	**
	Myrianida_edwarsi	1	0.467	0.683	0.005	**	Cnidaria sp.	1	1	1	0.005	**
	Spisula_subtruncata	1	0.4	0.632	0.005	**	Golfingia sp.	0.983	1	0.991	0.005	**
	Scoloplos_cf._armiger_CB-20	1	0.333	0.577	0.005	**	Autolyte sp.	0.743	0.778	0.76	0.035	*
	Urothoe_poseidonis	1	0.267	0.516	0.035	*						
	Sigalion_mathildae	1	0.267	0.516	0.045	*						
Spring	Spio_goniocephala	1	0.4	0.632	0.005	**	Bryozoa	1	0.667	0.816	0.005	**
	Donax_vittatus	0.841	0.4	0.58	0.01	**	Pontocrates altamarinu	1	0.444	0.667	0.03	*
	Spisula_subtruncata	1	0.2	0.447	0.035	*	Hydrozoa	1	0.444	0.667	0.04	*
	Abra_alba	1	0.2	0.447	0.035	*	Nephtys cirrosa	1	0.775	0.75	0.049	*
	Amphinema_dinema	1	0.2	0.447	0.035	*						
	Lanice_conchilega	1	0.2	0.447	0.035	*						
	Nephtys_cirrosa	1	0.2	0.447	0.035	*						
	Nototropis_falcatus	0.999	0.2	0.447	0.035	*						
Summer	Liocarcinus_holsatus	1	0.8	0.894	0.005	**	Spio symphyta	0.83	0.8	0.815	0.01	**
	Diastylis_bradyi	0.746	1	0.864	0.005	**	Abra prismatica	1	0.6	0.775	0.01	**
	Microstomum_compositum	1	0.733	0.856	0.005	**	Eumida sanguinea	0.932	1	0.965	0.005	**
	Cephalothrix_sp._4_TCH-2015	1	0.733	0.856	0.005	**	Sipunculidae sp.	1	1	1	0.005	**
	Abra_prismatica	1	0.6	0.775	0.005	**	Bathyporeia sp.	1	1	1	0.005	**
	Gastrosaccus_spinifer	1	0.6	0.775	0.005	**	Corystes cassivelaunus	1	0.6	0.775	0.005	**
	Nonparahalosydna_pleiolep	0.998	0.6	0.774	0.005	**	Nereis spp.	1	0.6	0.775	0.01	**
	Ectopleura_dumortierii	0.955	0.6	0.757	0.005	**	Ophiura_albida	1	0.37	0.713	0.01	**
	Magelona_mirabilis	0.593	0.933	0.744	0.005	**	Pseudopolydora pulchr	1	0.4	0.632	0.015	*
	Leuckartiara_octona	1	0.533	0.73	0.005	**	Pseudocuma longicorne	1	0.4	0.632	0.035	*
	Pseudocuma_longicorne	1	0.467	0.683	0.005	**	Asteroidea spp.	1	0.4	0.632	0.045	*
	Ophiura_albida	1	0.467	0.683	0.005	**	Pseudocumadidae spp	0.783	0.6	0.685	0.035	*
	Electra_sp._LM-2010	1	0.467	0.683	0.005	**						
	Carcinus_maenas	1	0.467	0.683	0.005	**						
	Microplaninae_sp._MOTU23	1	0.4	0.632	0.005	**						
	Schistomysis_ornata	1	0.4	0.632	0.005	**						
	Pleopis_polyphemoides	0.854	0.467	0.631	0.005	**						
	Corystes_cassivelaunus	1	0.333	0.577	0.005	**						
	Aurelia_aurita	1	0.267	0.516	0.03	*						
		1	0.267	0.516	0.035	*						
		0.962	0.267	0.507	0.015	*						
Autumn + Spring	Donax_vittatus	0.984	0.767	0.869	0.025	*	Scoloplos armiger	0.962	1	0.981	0.005	**
	Sigalion_mathildae	0.971	0.6	0.763	0.015	*	Donax vittatus	0.9	0.889	0.894	0.03	*
Autumn + Summer	Urothoe_poseidonis	0.967	0.867	0.915	0.005	**	Scolelepis bonnieri	0.987	0.786	0.881	0.01	**
	Obelia_bidentata	0.999	0.833	0.912	0.005	**	Megaluropus agilis	0.932	0.786	0.856	0.01	**
	Obelia_sp._1_SL-2013	1	0.667	0.816	0.005	**	Spio decorata	0.964	0.5	0.694	0.035	*
	Megaluropus_agilis	1	0.433	0.658	0.005	**						
	Spio decorata	1	0.367	0.606	0.015	*						
Spring + Summer	Electra_pilosa	0.967	0.933	0.95	0.005	**	Monopseudocuma gils	0.985	1	0.992	0.005	**
	Sagartiogeton_laceratus	1	0.9	0.948	0.005	**	Pontocrates arcticus	0.867	0.929	0.897	0.025	*
	Microprotopus_maculatus	0.942	0.767	0.85	0.01	**	Diastylis bradyi	0.89	0.857	0.873	0.035	*
	Tellimya_ferruginosa	0.979	0.7	0.828	0.005	**	Owenia fusiformis	0.9	0.867	0.853	0.005	**
	Echinocardium_cordatum	0.923	0.733	0.823	0.005	**	Bathyporeia elegans	0.946	0.786	0.862	0.005	**
	Spio_decorata	0.979	0.6	0.767	0.005	**	Perioculodes longimar	0.898	0.786	0.84	0.03	*
	Owenia_fusiformis	1	0.567	0.753	0.005	**						
	Lagis_koreni	1	0.533	0.73	0.005	**						
	Eucheilota_maculata	1	0.467	0.683	0.015	*						

In conclusion, bulk DNA metabarcoding detected significantly more macrobenthos species and had significantly lower Shannon diversity than morphology-based analyses. Patterns of species numbers were significantly different between the two methods, while evenness patterns were identical. Both methods also detected different macrobenthos species, as only 25% of species were shared between the two methods. Despite the different species picked up by the two methods, similar patterns in beta diversity were observed between the two methods: significant differences in macrobenthos communities between all three seasons for the two methods were observed. However, the indicator species responsible for these differences were largely different between the two methods.

c. Case study on impact assessment of mechanical lugworm harvesting in the Wadden Sea (The Netherlands)

For the 90 samples processed for this case study, only five samples delivered a PCR product after the initial amplification. To improve the success, various troubleshooting methods were tried. Samples were diluted 10x and 100x to test if this would lower the effect of potential inhibitors in the DNA extracts, three different polymerase mixes were tested (KAPA HiFi HotStart ReadyMix, Taqman Environmental Master Mix 2.0 and Phire II Hotstart), and the annealing temperature was adjusted. However, all the different alterations were unsuccessful, and no more than five samples were amplified. Diversity analyses were performed on the morphological data in an internship project, but no bulk DNA-based data was obtained to compare. From the results of the PCRs, it is hard to conclude what the reason behind the failure of the samples was. The process of freeze-drying and preservation itself may have already had an impact on the integrity of the DNA of the samples. From measurements on the DNA itself, it was found that there were relatively high concentrations of inhibiting compounds present in the extracts. These were lowered after clean-up, which however also significantly lowered the concentrations of DNA measured. The reason for the failure of these samples likely comes down to a combination of low-quality, low-concentration DNA and relatively high concentrations of inhibitors that negatively affected the PCRs.

For the evaluation of the primer and PCR success, *in silico* testing was performed on five forward primers from the literature (reverse primers were not tested *in silico* as they are located outside the reference barcode region). The forward primer mlCOIintF-XT bound to the highest number of the 50 mock species (Fig 16A) and the 178 North Sea macrobenthos species (Fig 16B). The BF primers performed quite similarly to one another and all bound roughly to the same amount of mock and North Sea macrobenthos species. Primer mlCOIintF bound to noticeably less mock and North Sea macrobenthos species.

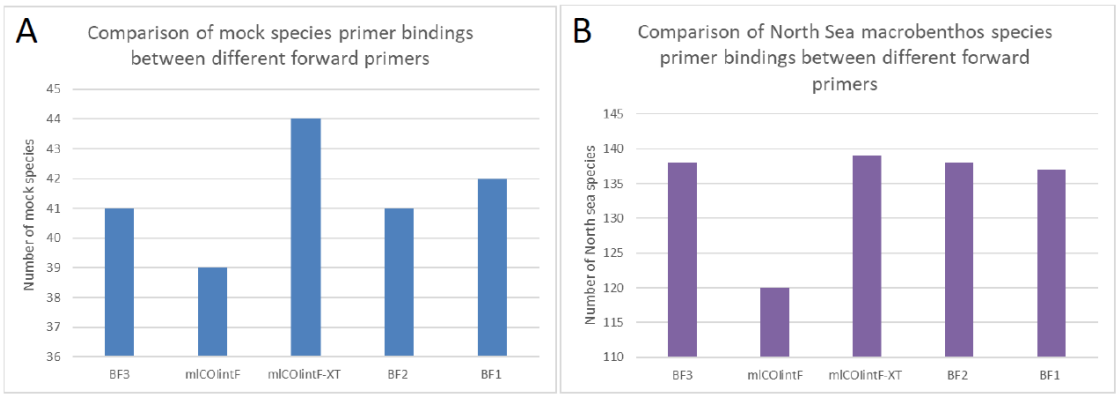


Figure 16: Different forward primers and their ability to bind in silico to the mock species sequences (A)(N = 50) and to all North Sea macrobenthos species including all mock species (B)(N = 178).

The number of species detected in the wetlab mock communities with the primer combinations where mlCOIintF was the forward primer largely exceeded the *in silico* tests in the number of

mock community species detected (Fig 17). For the mlCOIintF-XT primer combinations, *in silico* and wetlab testing were very similar. This analysis clearly illustrates that *in silico* testing of primer sets does not always accurately predict the species that will be detected in bulk DNA samples. No primer combination detected all mock community species. All combinations failed to detect the species *Arctica islandica*, *Cerianthus lloydii*, *Ensis siliqua* and *Ophiura albida* in contradiction to the *in silico* test. The combinations with Fol-Degen-Rev as reverse primer detected the species *Alcyonium digitatum* and *Aphrodite aculeata*, which were undetected by the other primer combinations. *Aequipecten opercularis* was only detected by the mlCOIintF-XT + Fol-Degen-Rev primer combination even though it was not supposed to be amplified according to the *in silico* tests.

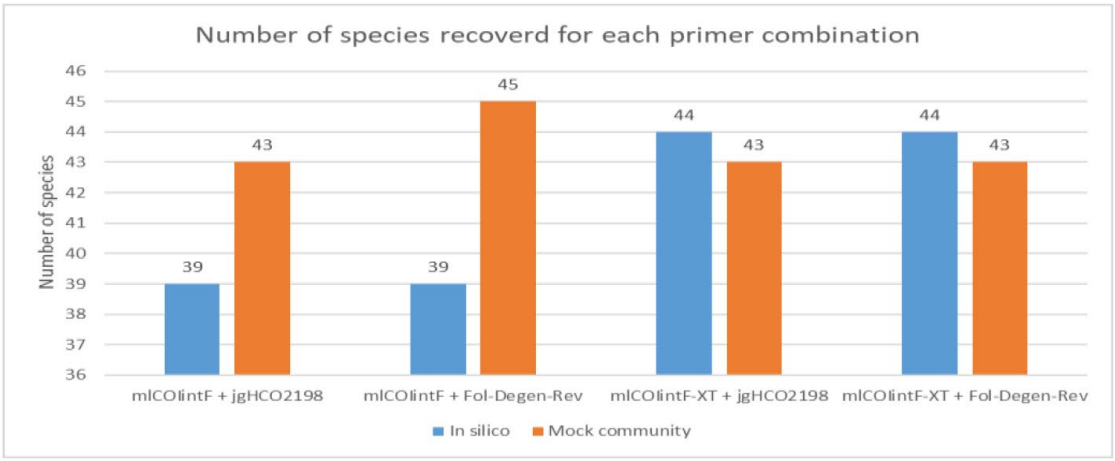


Figure 17: Different primer combinations and the number of mock community species recovered for the in silico test and the mock community samples.

The accumulation curves of the mock communities demonstrate that not all species were found in three PCR replicates (Fig 18), the number of replicates which is often used for these studies. The curves start to flatten out after three replicates, and at six replicates only two to four mock community species were still not detected. The number of PCR replicates needed to detect all the amplified species present in the mock communities was ten.

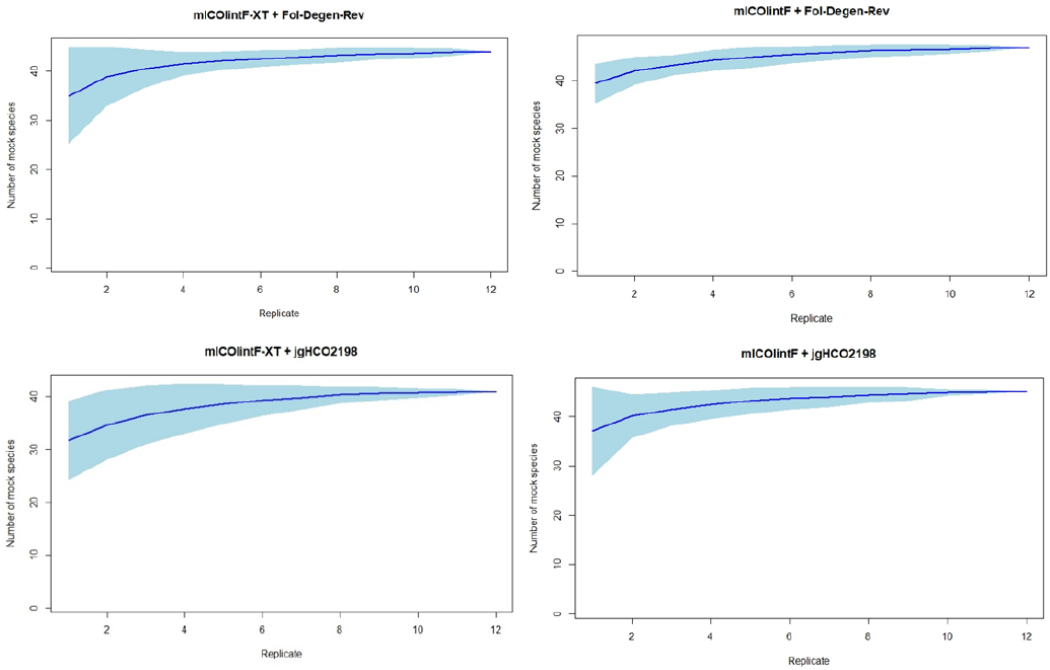


Figure 18: Species accumulation curves and their confidence intervals (p = 0,005) for all primer combinations based on the PCR replicates of the undiluted mock community.

Additionally, in the mock community with diluted DNA for several species, some of the diluted species were not detected at all in any of the 12 PCR replicates. *Lutraria lutraria*, *Pagurus bernhardus* and *Aporrhais pespecelani* have all been detected in the undiluted samples but were not detected in the diluted mock samples, whereas the number of reads for *Astropecten irregularis* and *Echinocardium flavescens* indicated that a 1000 times dilution of a species does not equal a 1000 times less reads (Fig 19).

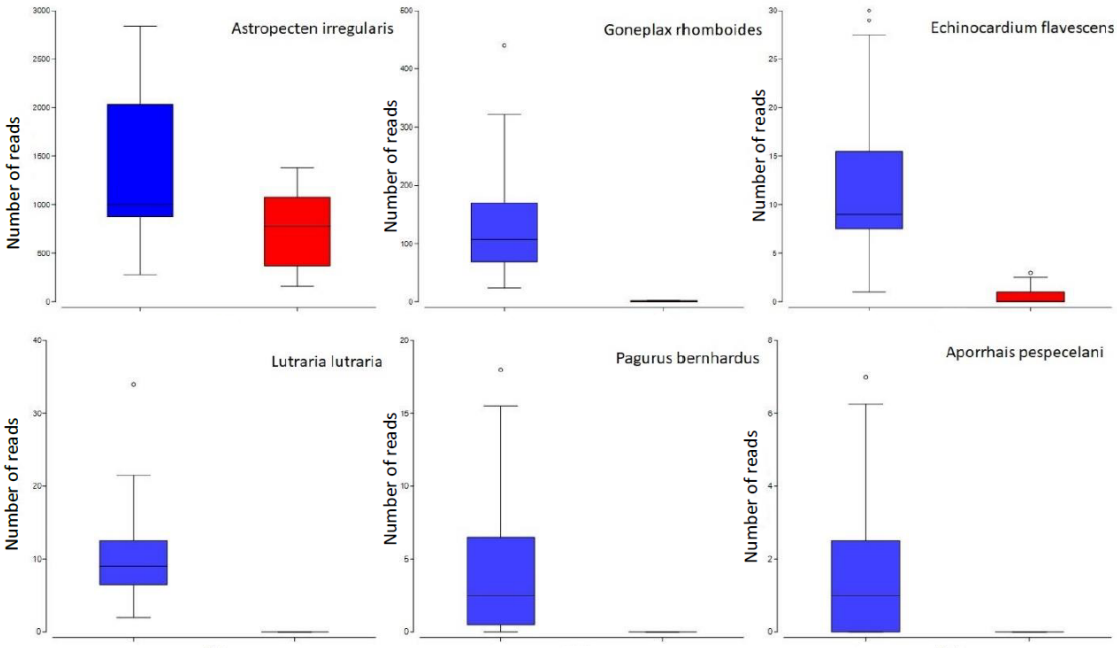


Figure 19: Boxplots displaying the number of reads found for six species in both the regular mock sample (in blue) and using a 1000-fold dilution in a second mock sample (in red).

In conclusion, this case study illustrated that DNA-based monitoring may fail and it remained unclear at which step the problem occurred (sample preservation, DNA extraction, PCR amplification). The lab tests further illustrated that the Leray primerset detected the highest number of macrobenthos species in the mock samples, and that at least three PCR replicates were needed to detect the majority of the species.

d. Case study on MSFD monitoring in the North Sea (Denmark)

For eDNA, two primer sets targeting COI and 18S rDNA were used for profiling benthic communities. The morphology based species list was dominated by Annelida, Arthropoda, Echinodermata and Mollusca and these four phyla were also abundantly detected by COI and/or 18S eDNA (Fig 20). A substantial proportion of the 18S eDNA was assigned to Platyhelminthes and a small portion of COI eDNA reads were assigned to Nematoda, two phyla that are absent from the morphology based method since they are not macrobenthos and thus pass through the sieve.

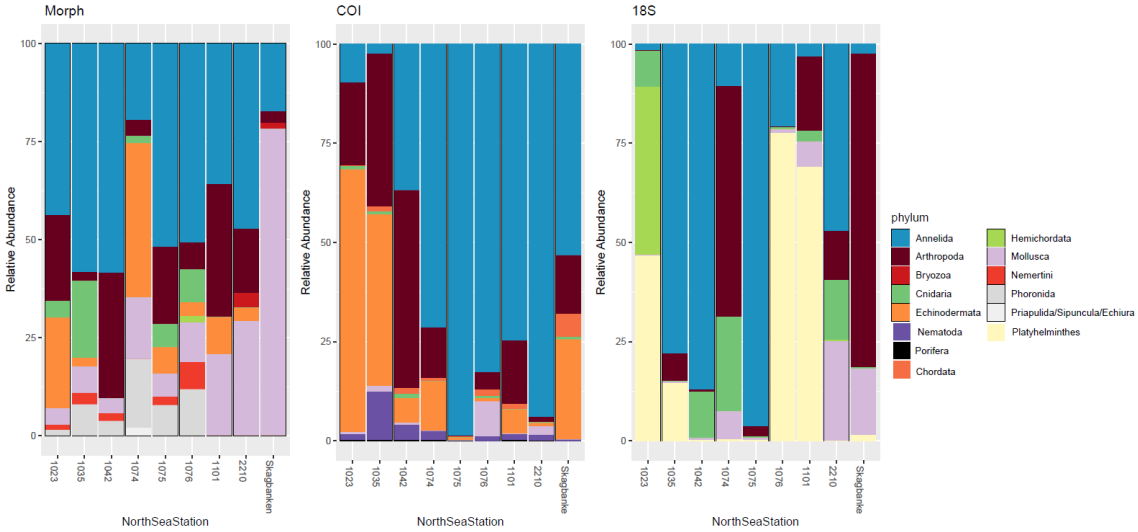


Figure 20. Barplot showing relative abundance of metazoa communities at phylum level from 2019 analyzed by conventional morphology (Morph) and eDNA (COI and 18S). The X axis shows the 9 stations, read abundance data from the 5 replicate samples per station were summed and then converted to relative abundance for each phylum.

For both eDNA and morphological methods, we determined alpha diversity as the observed species richness and the Shannon diversity index for the 9 sampling stations (Fig 21). Highest species richness was observed for the morphological method (mean 7.3 ± 2.8) followed by COI eDNA (6.1 ±2.1) and lowest for the 18S eDNA (3.1±1.8). Similarly Shannon diversity was higher for morphology based identification (mean 1.7 ± 0.5) compared to COI eDNA (0.8 ±0.5) and 18S eDNA (0.5 ±0.4). Species richness (observed) and the Shannon diversity index showed no overall significant difference between sampling stations for morphological data nor eDNA data using COI and 18S regions (Kruskal-Wallis test).

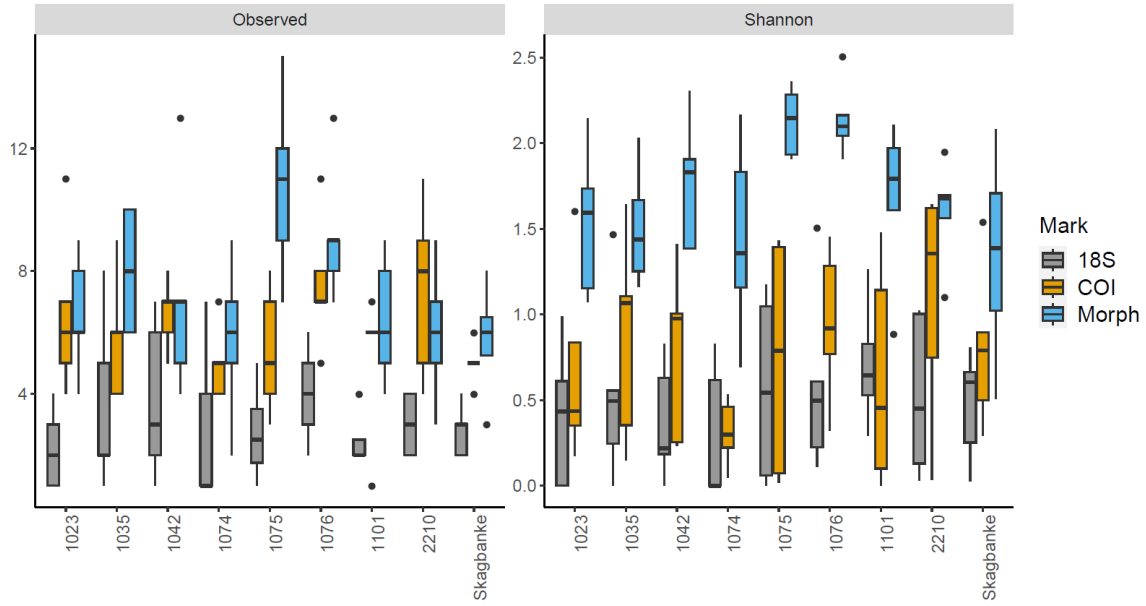


Figure 21. Benthic eukaryotic alpha diversity (observed species richness and Shannon diversity) in nine different North sea stations based on eDNA (COI and 18S) and conventional morphological (Morph) identification methods. Variability represents 5 subsamples at each station.

To ease comparison with the morphology based species detection we pooled the COI and 18S eDNA species lists. In general, there are more species captured by morphological sampling compared to eDNA-based detection. While the eDNA method detected 64 other species not reported by morphological methods, the morphological method reported 104 species not detected by eDNA. Only 10 species (5.6 % of all species found) were detected by both morphology and eDNA-based analyses (Fig 22).

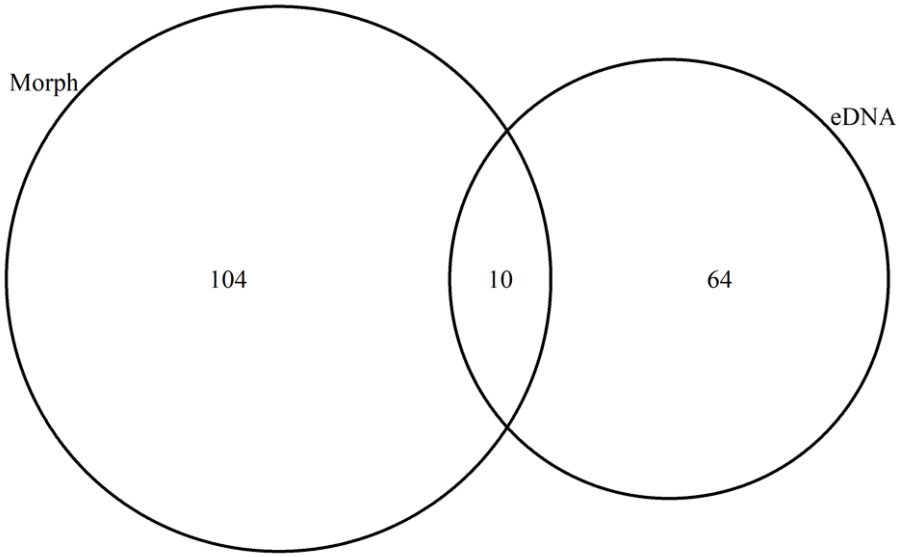


Figure 22: Venn diagram comparing the benthic invertebrate species in common between the conventional monitoring vs eDNA based metabarcoding in 2019. For eDNA, the COI and 18S species were merged.

Using a non-parametric analysis of variance (ANOSIM), we found that both the eDNA and morphological method significantly separated the nine sampling stations ($p<0.1$). Both methods furthermore indicated that samples were grouped according to sampling depth (Fig 23). The separation according to depth was also confirmed by the ANOSIM analysis ($p<0.1$).

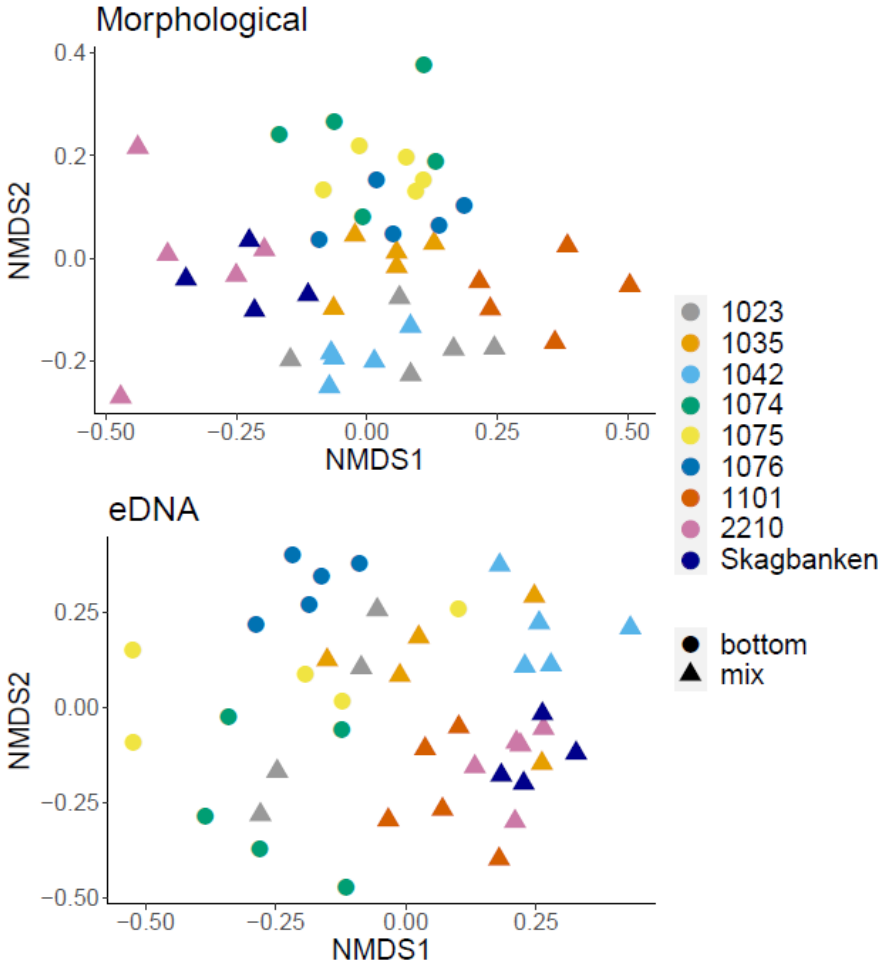


Figure 23: nMDS plots showing distribution of sediment samples from 9 stations in 2019 that were analyzed by both eDNA (COI and 18S) and morphology based methods. Analysis was based on presence/absence data and the Sørensen distance measure. Stations are indicated by different colors, depth zone is indicated by different symbols (> 44 m: bottom, < 44 m: mix).

Pairwise adonis tests on presence/absence data for morphology and eDNA (18S + COI) revealed that most stations were different between each other (Table 9). Of the 17 pairwise comparisons, only three comparisons had a different outcome: stations 1023 and 1042 were not significantly different from each other when using eDNA, and stations 1074 and 1075 and stations 2210 and 1101 were not significantly different from each other when using the morphological method.

Table 9: Pairwise permutational analysis of variance (PERMANOVA) using the 'adonis' test on Bray-Curtis distance matrices for conventional and eDNA-based eukaryotic community dissimilarity assessment using 1000 permutations using presence absence data.

Station pairs	Morphology		eDNA	
	p.value	Sig	p.value	Sig
1023 vs 1035	0.031	•	0.013	•
1023 vs 1042	0.009	••	–	–
1023 vs 1101	0.009	••	0.006	••
1023 vs Skagbanken	0.026	•	0.009	••
1035 vs 1042	0.008	••	0.007	••
1035 vs 1101	0.019	•	0.008	••
1035 vs Skagbanken	0.029	•	0.005	••
1042 vs 1101	0.008	••	0.005	••
1042 vs Skagbanken	0.007	••	0.014	•
1074 vs 1075	–	–	0.009	••
1074 vs 1076	0.007	••	0.011	•
1075 vs 1076	0.006	••	0.008	••
1101 vs Skagbanken	0.022	•	0.011	•
2210 vs 1023	0.007	••	0.014	•
2210 vs 1035	0.031	•	0.007	••
2210 vs 1042	0.005	••	0.008	••
2210 vs 1101	–	–	0.008	••

Despite the similar patterns that differentiated the stations, SIMPER analyses showed that the species contributing to the observed differences between stations were completely different for the eDNA and conventional methods. Indicator species analysis pinpointed seven species explaining the difference between the two depths for the conventional method, and seven species for the eDNA method with no overlap in the list of species (Table 10).

Table 10: Indicator species associated to two different depths for the morphological and eDNA based identification methods. Sediment samples from coastal stations (< 44 m depth) are classified as mix, while samples from offshore stations (> 44 m depth) are classified as bottom.

Group	Species	Stat	P	Sig	Method
Bottom	*Diarthrodes_sp.*	0.5	0.0005	***	eDNA
	Cyclopinodes_sp.	0.333	0.0372	*	
	Arctica_islandica	0.322	0.0424	*	
Mix	*Spisula_solidissima*	0.66	0.0001	***	
	Perophora_sagamiensis	0.605	0.0004	***	
	Acartia_hudsonica	0.522	0.0023	**	
	Paraeupolymnia_uspiana	0.408	0.0227	*	
Bottom	*Phoronis_sp.*	0.797	0.0001	***	Morphology
	Amphiura_filiformis	0.552	0.0005	***	
	Anthozoa_indet.	0.447	0.0037	**	
	Harpinia_antennaria	0.447	0.0026	**	
	Argissa_hamatipes	0.333	0.0314	*	
	Galathowenia_oculata	0.333	0.0369	*	
Mix	*Bathyporeia_pelagica*	0.37	0.0435	*	

In conclusion, the eDNA and morphology-based approaches provided similar patterns of alpha diversity (species richness and Shannon-diversity) as both methods showed no significant

differences between stations. Significantly more species were detected with morphology-based identification than with the eDNA and only 10 species were found by both methods. Despite these differences, beta diversity patterns were comparable between the two methods, as both showed that stations differed significantly in species composition, and with a clear distinction between deeper (>44m) and shallower sampling stations. However, the species explaining the differences between the two depth zones were completely different between both methods. Clearly, the eDNA extracted from the sediment obtained from a very small syringe sample does not provide a representative picture of the metazoan species detected with a larger grab sample but is able to detect ecological patterns linked to differences between stations and depth zones.

e. *Effect of sample fixative (formalin vs ethanol) on macrobenthos diversity*

Two case studies were able to compare the effect of sample fixative on alpha and beta diversity. In case study "*a*" regarding aggregate extraction, comparison between the number of detected taxa in bulk DNA samples, morphological samples fixed in ethanol or in formalin was performed for the Thorntonbank samples in 2019. Two-way ANOVA results showed that methodology did not affect the number of species detected: no significant interaction impact x method and no significant effect of the main factor method were found. Sand extraction impact did significantly affect the number of species (p-value = 0.000267), with medium impacted samples having higher numbers of taxa with all three identification methods (Fig 24 A). In contrast, the species identity differed between methods, as the three methods only shared 22.6% of the taxa. In addition to the unique taxa found by bulk DNA metabarcoding (13.5%), also 21.3% and 14.8% unique taxa were found for the morphology based method fixed in formalin or in ethanol, respectively (Fig 24 B). Both morphology based methods or fixatives shared 46.5% of the species.

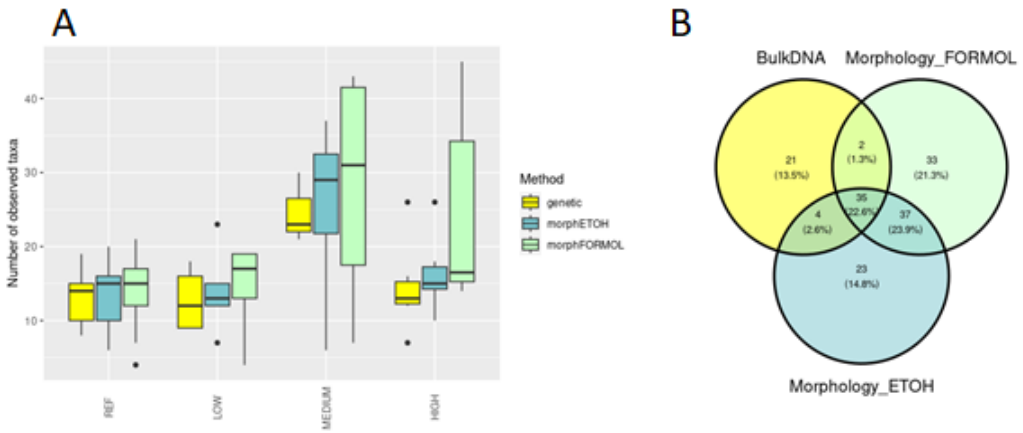


Figure 24: Boxplot (A) and Venn diagram (B) of the taxa detected in the bulk DNA dataset and in the morphological datasets fixed in ethanol or formalin in the Thorntonbank sampled in 2019.

Community patterns differed between the identification methods, with the two morphology-based methods clustering closer together (Fig 25). PERMANOVA showed no significant

interaction effect between methodologies and impact (F=1.1517, p-value=0.1229), suggesting that the three identification methods yield similar patterns of aggregate extraction impact on community composition. A significant effect of the identification method and of impact group was observed. Significant effects of the identification methods were only observed for the bulk DNA versus morphology fixed on formalin (p-value=0.006) and morphology fixed on ethanol (p-value= 0.002). As no significant effects were detected between the different fixatives for morphological identifications, only the ethanol fixative can be used.

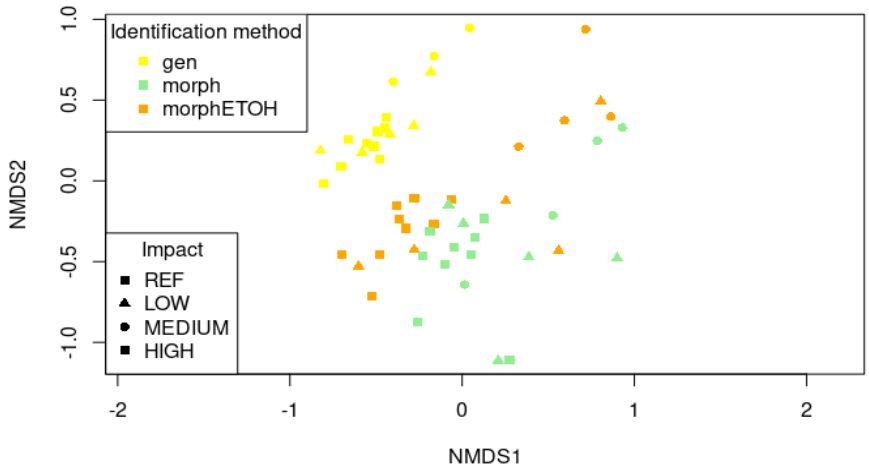


Figure 25: nMDS plot using Jaccard index (presence/absence) for samples identified with morphology from formalin (green) or ethanol (orange) fixed samples and with bulk DNA (yellow).

In case study "b" on long-term monitoring, a subset of 18 samples collected in autumn and spring was available for which formalin and ethanol fixed samples have been identified morphologically and bulk DNA data identification was obtained from the ethanol samples. Two-way ANOVA showed that the interaction between seasons and methods was not significant for evenness (p= 0.41) or for the number of species (p =0.81), indicating that the same patterns were observed irrespective of the methodology used. However, the main factor methodology was significantly different for both indices (p = 0.0006 and 0.0007 for evenness and number of species, respectively), with lower number of species and evenness for the bulk DNA dataset compared to the morphological datasets based on formalin or ethanol (Fig 26). The main factor "season" was not significantly different for the two indices.

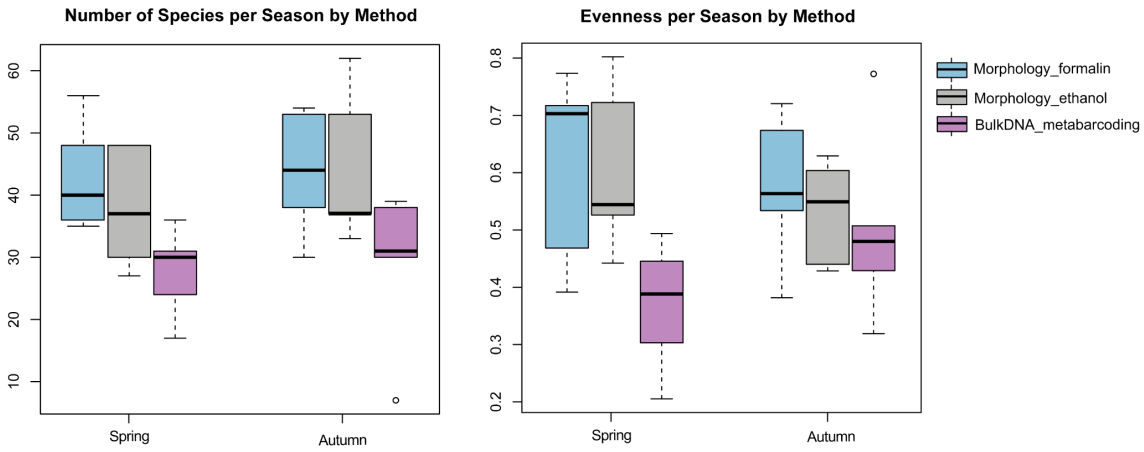


Figure 26: Number of species (left) and Shannon diversity evenness calculated from bulk DNA, morphology formalin and ethanol for spring and autumn. For each method one sample per season/station (10 in total) have been considered. For samples with 3 replicates a mean of the 3 values has been applied.

Also here, the number of shared species between the three methodologies was low (40, or 23% of all species found) (Fig 27). Bulk DNA metabarcoding shared more species (n=5) only with ethanol samples than only with formalin (n=1) and had the highest number of unique species (47). Between the two morphological approaches, 84 species were shared while formalin samples have 19 unique species and ethanol samples 18 unique species using morphology (Fig 26).

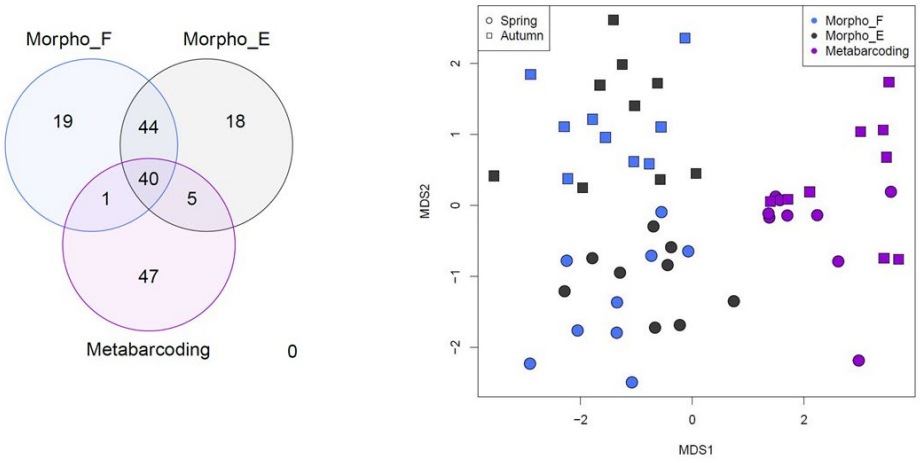


Figure 27: Venn diagram (left) showing the number of shared and unique species between bulk DNA, morphology (ethanol) and morphology (formalin) for 18 samples. Summer is excluded as ethanol morphology samples have no summer counts. nMDS plot (right) using present/absent analyses comparing the composition of Norderney soft bottom macrofauna between bulk DNA, morphology formalin and ethanol for autumn and spring (18 samples).

Two-way PERMANOVA using presence/absence (Jaccard distance) revealed significant differences in the community between methods (bulk DNA, formalin and ethanol morphology

p=0.001), seasons (autumn and spring, p=0.001) and the interaction of both factors (method * seasons, p=0.005) (Fig 27). Also, multiple pairwise PERMANOVA analyses showed no significant differences in the community between ethanol and formalin samples (morphology) for each season (autumn, p=0.18; spring p=0.39) suggesting that only ethanol-preserved samples could have been considered for morphology.

f. *Time and cost comparison between DNA and morphology based analyses of macrobenthos*

The tracked time and costs for the three case studies are summarized in Table 11. For the subpilot on sand extraction, DNA-based analysis was 46% faster and 26% cheaper compared to the morphology-based analysis. For the long term monitoring station in Norderney, DNA-based analysis was 66% faster and 27% cheaper compared to the morphology-based analysis. For the MSFD case study in Denmark using eDNA, DNA-based analysis was 43% faster and 9% cheaper than the morphology based analyses. However, it is important to note here that the first two steps of the sample processing (decantation and screening for heavier specimens) are not needed for eDNA analyses of the sediment, and therefore time and costs associated with eDNA analyses are in reality even much lower than the numbers reported here if the entire sample processing would have been taken into consideration.

Table 11. Summary of the time and costs associated with the processing of 24 samples from Thortonbank (case study a), 15 samples for Norderney Isle (case study b) and 40 samples for MSFD monitoring (case study d). Note that for the latter case study, time and costs for the first two processing steps differ substantially, but these have not been taken into consideration to be consistent with the two other case studies.

	# samples	Level of experience	Total time (h)	Time/sample (h)	Total cost (€)	Cost/sample (€)
Subpilot a/ impact of aggregate extraction						
Morphology	24	Senior expert	67	2.8	6514	271
Bulk DNA	24	Junior expert	37	1.5	4823	201
Bulk DNA versus morphology			**-44%**	**-46%**	**-26%**	**-26%**
Subpilot b/ Long-term soft-bottom monitoring station						
Morphology	10	Senior /Junior expert	191	19.1	5133	513,3
Bulk DNA	15	Junior expert	96	6.4	4828	322
Bulk DNA versus morphology			**-50%**	**-66%**	**-6%**	**-27%**
Subpilot d/ MSFD monitoring in Danish waters						
Morphology	40	Senior expert	140	3.5	8353	209
eDNA	40	Junior expert	80	2	7560	189
eDNA versus morphology			**-43%**	**-43%**	**-9%**	**-9%**

5. General conclusions

The two case studies (a and b) that successfully applied bulk DNA metabarcoding and morphology based analyses showed very similar conclusions despite the different objectives of both studies: 1/ bulk DNA detected significantly more taxa than morphology based identification; 2/ bulk DNA had significantly lower Shannon index than the morphology based method; 3/ the pattern of Shannon diversity was highly similar between both methods; 4/ the number of species shared by both methods was low (37 % for the sand extraction study, 25 % for the long-term monitoring study); 5/ despite differences in species identity and number of

species detected, beta diversity patterns were comparable between methods; 6/ bulk DNA identification is 46-66% faster and 26-27% cheaper than morphology based identification. For the sand extraction study, additional congruence between both methods was observed for patterns in species richness and importantly, the common species characteristic for similarity within impact groups were similar as well. Consequently, bulk DNA is a valid alternative to morphology-based analyses for monitoring sand extraction activity in dynamic coarse sandy areas like the Belgian part of the North Sea. We did, however, not look into density and biomass so this statement is valid if no density and biomass information is needed and the interest purely lies in general alpha and beta diversity patterns. For the long-term monitoring case, seasonal differences in species richness patterns were detected and the indicator species differed as well when using bulk DNA or morphology for identification of the samples. For this case study, a combination of both methods would bring the best monitoring results.

Macrobenthos samples for morphological identification are in most countries fixed with formalin, preventing subsequent DNA-based analyses. Ethanol fixed samples allow DNA-based analyses but hamper morphological identification since crustacean specimens become more fragile, and small, interstitial species seem to dissolve in ethanol over time. For this reason, a subset of samples in the sand extraction case study and in the Norderney case study were fixed with either formalin or ethanol before morphological identification. The two case studies showed consistent results when comparing formalin or ethanol fixed samples: more than half of the species were picked up by both fixatives (53% and 64% of the morphological species in case study a and b, respectively) but also a large number of unique species were identified in the ethanol and formalin samples, since these represented different replicates. However, alpha and beta diversity patterns were highly similar between the two fixatives. Bulk DNA samples shared most species with the ethanol fixed samples, which is linked to the fact that both methods can be done on the same set of animals when using ethanol.

The lugworm case study was not able to generate bulk DNA data because of PCR failure. This was the only dataset for which samples had been freeze dried instead of ethanol fixed, but it should be explicitly tested whether freeze drying was at the base of this failure. The lab testing further confirmed earlier results (Derycke et al 2021, Van den Bulcke et al. 2022) in that the used lab protocol was the optimal choice for macrobenthos: most species are detected when using three PCR replicates, the number that has been used throughout the other case studies reported here and the Leray primers (mIntCOIF and HCO2198 degenerated) detect the highest number of species of the mock community. This test also demonstrated that read numbers do not always accurately reflect DNA template amount, which hampers inferences on the abundance of species in a sample.

Finally, the eDNA approach in the MSFD monitoring case study offers a big advantage compared to bulk DNA analyses during the processing step because macrobenthos does not need to be sorted from the sediment, instead eDNA is extracted by simply taking a small amount of sediment. Results of this case study showed that significantly more species were detected with morphology based identification than with eDNA and only 10 species (5.6%) were found by both methods. In contrast to the bulk DNA case studies, the eDNA extracted from the sediment does

not provide a representative picture of the macrobenthos species in that location. Despite these differences, both eDNA and morphology methods showed that stations significantly differed in species composition, and clearly separated stations based on depth. However, the species explaining the differences between the two depth zones were completely different between both methods. Although ecological patterns linked to differences between stations and depth zones were detected with eDNA, there is absolutely no link with morphological monitoring data and therefore this kind of monitoring would completely break with long term monitoring data.

6. SWOT analysis

Based on our experience during the execution of the four case studies, the GEANS consortium has defined strengths and weaknesses related to DNA-based soft sediment monitoring, and identified external opportunities and threats when using DNA-based methods for macrobenthos monitoring in the North Sea (Table 12). The GEANS project contributed to several of the points raised in the SWOT analysis.

With respect to the strengths of DNA-based monitoring

- The different case studies have kept track of time and costs involved in the analysis of macrobenthos, and provide independent empirical evidence that DNA-based methods are cheaper and faster than traditional morphology based analyses.
- The case studies also revealed that DNA-based analyses are able to identify difficult taxonomic groups to species level which generally yielded a higher number of species in the DNA-based identification method.

In terms of weaknesses:

- The lugworm case study demonstrated that DNA-based methods can fail to yield any result and that primer choice plays a role in whether macrobenthic species will be detected or not.
- We find that sample preservation is critical to prevent failure of samples and therefore have provided guidelines for collecting samples to be used in DNA-based analyses Articles, Reports & Publications | GEANS).
- Incomplete reference databases are a drawback for DNA-based monitoring studies, since only those species that have a reference sequence in the database can be detected with DNA metabarcoding. The GEANS project generated a North Sea specific reference database with COI sequences linked to a completely updated taxonomy for macrobenthos and added 86 species that had not been barcoded before. This substantially improves the power of DNA-based metabarcoding studies in the North Sea.
- The main weakness of DNA-based methods remains in the fact that no life history information can be obtained (no information on females, males, juveniles) and that quantitative estimates are not always reliable.
- eDNA extracted from sediments provides limited information on infaunal species composition and can not replace morphological species identification

In terms of opportunities:

- The case studies show that standardisation of the bulk DNA protocol can be easily done across institutes. GEANS has produced a soft sediment protocol (https://geans.eu/protocols/sbs) that was tested in a transnational ring test and proved to be robust and reproducible (Van den Bulcke *et al.* 2023)
- The GEANS project also illustrated that taxonomic errors in traditional morphological studies can be detected when creating reference sequences. This is exemplified in the sequencing of a *Loimia ramzega* specimen that had been morphologically identified as *Lanice conchilega*. Both species are resembling and *Loimia* was previously not known from the area. Careful re-examination of the voucher specimen matched indeed the taxonomic description of *Loimia*, and allowed to trace back the taxonomic error and to update the monitoring data in retrospect.

With respect to threats:

- By implementing these different case studies alongside existing monitoring programs/studies, the GEANS project contributed to expectation management by showing how the DNA-based methods perform in comparison to traditional morphological methods. We provide evidence based results on what can and cannot be done when applying DNA-based monitoring. Additionally, the variety of these cases demonstrates the different types of monitoring questions that can be answered by DNA-based monitoring.
- Numerous discussions have taught us that local ecological knowledge is still key in the interpretation of DNA-based results. In order to acknowledge the presence of false positives or negatives, thorough knowledge on the local ecosystem is important to critically assess the obtained species list.

Table 12. List of internal strengths and weaknesses and external opportunities and threats of DNA-based monitoring of macrobenthos

Strengths	Weaknesses
• High throughput of samples is possible • High potential of detecting taxa that are taxonomically difficult to identify • In case of eDNA: same biological sample can be used without extra effort of sampling • Biodiversity estimation can be done from complex bulk samples • No taxonomic expertise needed for analysing samples thereby circumventing the loss of morphological benthic expertise • Cheaper • Faster • Area-specific reference databases such as the GEANS reference database provide ground truthing	• Sample preservation is more critical. • Quantitative information is unreliable at this moment, also no info on life stage and sex • Reliable, complete, curated reference libraries are a bottleneck. • eDNA: time of occurrence can be confounded with historical eDNA • False negatives: not all taxa are amplified • Sometimes samples fail, and there is no material to fall back to. The reason for failure is not always obvious.

Opportunities	Threats
• Automation of lab protocol • Many projects in parallel possible • More samples with same resources which can lead to higher power and/or an increased resolution • Standardisation of method • Uncover taxonomic errors • Highlight locations of specific taxonomic relevance to direct taxonomic effort to hotspots of biodiversity	• Poor pilot results may lead to negative experiences with end users • Expectation management: important to be honest in what DNA-based monitoring can and cannot do • (Legal) Framework missing (no money) • Misinterpretation of data if insufficient knowledge of taxonomic/genetic variation • Compatibility of current monitoring programs & DNA-based sampling • Long-term storage of samples in ethanol requires substantial freezer capacity

.

7. Practical recommendations for DNA-based soft sediment monitoring

The case studies reported here demonstrated that DNA-based monitoring of macrobenthos is a valuable method for ecological monitoring in geographic regions where extensive ecological knowledge is present. Such knowledge is critical to correctly interpret species information from DNA-based data, since it provides baseline information to create reliable reference databases (https://geans.eu/refrence-library) with sequence data from vouchered macrobenthos species ensuring a correct taxonomic identification of DNA sequences. When assigning taxonomy to COI sequences from southern North Sea samples, we recommend using the GEANS reference database to obtain the highest taxonomic resolution possible. Next to a good reference database, standardized field and lab protocols are key to obtaining reliable and robust results using bulk DNA metabarcoding. First of all, sample preservation is critical to obtain good quality DNA data. For DNA-based analyses, formalin preservation of samples is not an option. We have shown that species detections differ slightly between ethanol and formalin samples, but this did not affect alpha and beta diversity patterns. We therefore recommend storing macrobenthos samples in pure, undenatured ethanol and ensure that the final concentration of the macrobenthos sample is above 70%. Ideally, the concentration should be checked after 24h and if too much diluted, the ethanol should be replaced with fresh ethanol. We have provided a detailed field sampling protocol to ensure good quality samples for DNA-based analyses (Articles, Reports & Publications | GEANS). Since ethanol fixed samples can be used for both morphological and DNA-based analyses and overall diversity patterns are the same between the ethanol and formalin fixatives, we support a shift towards only collecting ethanol fixed macrobenthos samples for monitoring studies. This not only allows for genetic analyses, but also decreases exposure to carcinogenic formalin. Once samples have been adequately preserved, a standardized lab protocol offers the best strategy to obtain reliable and consistent DNA-based results. The GEANS lab protocol offers a detailed step by step procedure on how to process ethanol fixed macrobenthos samples (https://geans.eu/protocols/sbs). This protocol provides

very consistent DNA-based results across different institutes even when slight modifications are done to the library preparation step in the lab (Van den Bulcke et al., 2023), and is therefore our recommended protocol for bulk DNA metabarcoding. When the aim is to identify responses of macrobenthos communities to human impacts or to assess good environmental status in the framework of MSFD monitoring or to detect shifts over time, we recommend to use bulk DNA instead of eDNA from the sediment. Our case study in the Danish part of the North Sea illustrated that eDNA is able to detect shifts in metazoan diversity, but there was very poor alignment with species information from the morphological identification and therefore we do not recommend switching to eDNA monitoring. Finally, the storage of DNA-based monitoring data in open access databases is crucial to be able to integrate DNA data across countries and to re-use sequencing data when new bioinformatic processing pipelines become available. The GEANS datasets from the case studies have been added to the Marine Data Archive (https://mda.vliz.be/archive.php?folder=8705) and we recommend foreseeing funding in future projects to ensure proper data storage according to the FAIR principle.

8. Towards implementation of DNA-based monitoring

With the above recommendations in place we are able to get the most out of bulk DNA metabarcoding of macrobenthos. DNA-based methods definitely have proven mature enough to be considered as an equally appropriate method as the morphology based approach. The critical question at this point is whether we are ready to abandon a long tradition of morphological identification of samples and switch to DNA-based monitoring. Our results have shown that bulk DNA-based methods provide similar diversity patterns as morphology based identifications and provide better species resolution. Moreover, DNA-based methods allow a much higher throughput of samples, and therefore allow upscaling of the number of samples that can be processed. However, some species are only found with morphology, no life history information can be retrieved from DNA sequence data and abundances estimated from DNA sequence data are unreliable for many species. Although biodiversity indices may be calculated from biomass data, abundance data are the most widely used data format that goes into the calculation of most of the environmental quality indices. The intercalibration of biomass and abundances derived diversity estimates as indicators of environmental/ecological quality seems difficult as the biological traits of many species change during their lifespan (i.e. many young individuals versus few large and old individuals does not give the same information of the environmental quality of the community) and these biological traits can - for now- only be derived from morphology based analyses. As such, both methods provide complementary information and combining both yield the highest possible resolution of species information. Such high resolution biodiversity data forms the ultimate level of knowledge to adequately assess the good environmental status of the North Sea. However, combining the two methods has to be done in such a way that time and costs are not simply doubled and also in a way that we get the best out of both methods to further increase our current knowledge on ecosystem structure and function. DNA-based methods could be seen as a ‘quick’ way to keep the finger on the pulse related to benthic environmental health. In case changes are depicted using DNA, morphological or other methods can then be deployed to dive deeper in the area where change was detected to try and unravel the processes behind this change. To develop an optimal sampling strategy where morphology and DNA-based methods can be combined in an efficient way, we recommend a transition period of 3-5 years where both methods are applied

simultaneously by the different countries. This period will identify on the one hand how well DNA-based methods perform in a specific area and on the other hand, where and when it is crucial to have life history information or abundances of certain key species which can then be used to direct the efforts of morphology based monitoring. Such a transition period would allow to choose the best methodological approach for the monitoring question at hand, simultaneously contributes to further refining the reference databases and is needed to set-up a clear data management plan ensuring morphological data and DNA-based data are stored according to the FAIR principle. Minimum specifications of data collection, lab processing and bioinformatic analyses need to be provided so that information on how the species list was created can be retrieved at any point by any member state. Implementation of DNA-based monitoring should from the beginning carefully consider good data management practices.

9. References

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120. doi:https://doi.org/10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., & Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology, 37(8), 852-857.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581-+. doi:https://doi.org/10.1038/nmeth.3869

Clarke, K. R., & Gorley, R. N. (2006). Primer. PRIMER-e, Plymouth, 866.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. bioRxiv, 221499. doi:10.1101/221499

De Jong, M. F., Baptist, M. J., Lindeboom, H. J., & Hoekstra, P. (2015). Short-term impact of deep sand extraction and ecosystem-based landscaping on macrozoobenthos and sediment characteristics. Marine pollution bulletin, 97(1-2), 294–308. https://doi.org/10.1016/j.marpolbul.2015.06.002

Derycke, S., Maes, S., Van den Bulcke, L., Vanhollebeke, J., Wittoeck, J., Hillewaert, H., . . . De Backer, A. (2021). Detection of macrobenthos species with metabarcoding is consistent in bulkDNA but dependent on body size and sclerotization in eDNA from the ethanol preservative. Frontiers in Marine Science. doi:https://doi.org/10.3389/fmars.2021.637858

Dippner, J.W. & Kröncke, (2015). Ecological forecasting in the presence of abrupt regime shifts. Journal of Marine Systems, 150, 34-40.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14(6), 927-930. doi:https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Edward, W. B. (1984). Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. Advances in Ecological Research, 14, 1-55. doi:https://doi.org/10.1016/S0065-2504(08)60168-3

Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, 11.

Elbrecht, V., Braukmann, T. W., Ivanova, N. V., Prosser, S. W., Hajibabaei, M., Wright, M., ... & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ, 7, e7745.

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048. doi:https://doi.org/10.1093/bioinformatics/btw354

Geller, J., Meyer, C., Parker, M., & 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(5), 851-861.

Hansen, J.L.S. & A. Josefson (2014): Blødbundsfauna. Teknisk anvisning M19, 13 pp.

Kanneworff E, Nikolaisen W (1973) The Haps, a frame supported bottom corer. Ophelia 10: 119-129

Kröncke, I., Dippner, J.P., Heyen, H. & Zeiss, B. (1998). Long-term changes in macrofauna communities off Norderney (East Frisia, Germany) in relation to climate variability. Marine Ecology Progress Series, 167, 25-36.

Kröncke, I., Zeiss, B. & Rensing, C. (2001). Long-term variability in macrofauna species composition off the island of Norderney (East Frisia, Germany), in relation to changes in climatic and environmental conditions. Senckenbergiana marit. 31: 65-82.

Kröncke, I. & Reiss, H. (2010). Long-term variability of benthic indices off the island of Norderney in the southern North Sea. Marine Pollution Bulletin, 60, 58-68.

Kröncke, I., Reiss, H. & Dippner, J.W. (2013). Effects of cold winters and regime shifts on macrofauna communities in the southern North Sea. Estuar. Coast. Shelf Sci. 119: 79-90.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in zoology, 10(1), 1-14.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. Plos one, 8(4), e61217.

Newell, R. C., Seiderer, L. J., & Hitchcock, D. R. (1998). The impact of dredging works in coastal waters: a review of the sensitivity to disturbance and subsequent recovery of biological resources on the sea bed. Oceanography and Marine Biology: an annual review, 36(1), 127-178.

Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2020). vegan: Community Ecology Package. R package version 2.5-7.

Piló, D., Carvalho, A. N., Pereira, F., Coelho, H. E., & Gaspar, M. B. (2019). Evaluation of macrobenthic community responses to dredging through a multimetric approach: effective or apparent recovery?. Ecological Indicators, 96, 656-668.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research, 41(D1), D590-D596.

Queirós, A. M., Hiddink, J. G., Kaiser, M. J., & Hinz, H. (2006). Effects of chronic bottom trawling disturbance on benthic biomass, production and size spectra in different habitats. Journal of Experimental Marine Biology and Ecology, 335(1), 91-103.

R-Core-Team. (2022). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., ... & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Scientific reports, 7(1), 12240.

Van den Bulcke, L., De Backer, A., Ampe, B., Maes, S., Wittoeck, J., Waegeman, W., . . . Derycke, S. (2021). Towards harmonization of DNA metabarcoding for monitoring marine macrobenthos: the effect of technical replicates and pooled DNA extractions on species detection. Metabarcoding and Metagenomics, 5. doi:10.3897/mbmg.5.71107

Van den Bulcke, L., De Backer, A., Wittoeck, J., Beentjes, K., Maes, S., Christodoulou, M., . . . Derycke, S. (2023). DNA metabarcoding on repeat: Sequencing data of marine macrobenthos are reproducible and robust across labs and protocols. Ecological Indicators, 150, 110207. doi:https://doi.org/10.1016/j.ecolind.2023.110207

Van der Veer, H. W., Bergman, M. J. N., & Beukema, J. J. (1985). Dredging activities in the Dutch Wadden Sea: effects on macrobenthic infauna. Netherlands Journal of Sea Research, 19(2), 183-190.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology, 73(16), 5261-5267. doi:https://doi.org/10.1128/aem.00062-07

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. PeerJ, 6, e4705.

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

Wyns, L., Roche, M., Barette, F., Van Lancker, V., Degrendele, K., Hostens, K., De Backer, A. (2021). Near-field changes in the seabed and associated macrobenthic communities due to marine aggregate extraction on tidal sandbanks: a spatially explicit bio-physical approach considering geological context and extraction regimes. Continental Shelf Research 229: 104546. https://dx.doi.org/10.1016/j.csr.2021.104546

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution, 3(4), 613-623.

Zeiss, B. & Kröncke, I. (1997). Macrofauna long-term studies in a subtidal habitat off Norderney (East Frisia, Germany) from 1978 to 1994. I. The late winter samples. Oceanologica Acta, 20(1), 311-318.

Interreg
North Sea Region
GEANS
European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region

Interreg North Sea Region GEANS

European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region