eDNA-BASED MONITORING OF THE MARINE ENVIRONMENT

GEANS PILOT 4.

Interreg North Sea Region GEANS
European Regional Development Fund

EUROPEAN UNION

Genetic tools for Ecosystem health Assessment in the North Sea region

eDNA-BASED MONITORING OF THE MARINE ENVIRONMENT

Version: 4.0

Project:

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

Authors:

Derycke S, Cornelis I, Doorenspleet K, Hablützel P, Heyndrickx H, Nijland R, Uhlir C, De Backer A

Date:

16 June 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.

Main contact:

sofie.derycke@ilvo.vlaanderen.be

How to cite this report?

Derycke S, Cornelis I, Doorenspleet K, Hablützel P, Heyndrickx H, Nijland R, Uhlir C, De Backer A (2023). eDNA-based monitoring of the marine environment.

Contents

2.

1. Summary

The use of environmental DNA for monitoring marine biodiversity has rapidly gained interest because of the ease and non-invasive nature of sample collection. Several studies have shown that eDNA adequately captures fish diversity and reflects known spatial patterns, but whether this is true for the shallow North Sea system is currently not known. Moreover, to investigate whether eDNA is a valuable monitoring technique, comparison with morphology based analyses is needed. The eDNA pilot aimed to address these gaps and consisted of three case studies, one in offshore wind farms targeting fish and invertebrate diversity, one in harbors to monitor non-indigenous species, and one near artificial reefs to monitor fish communities. The eDNA results for the offshore wind farm study detected considerably more fishes and invertebrates than morphological identification of beam trawl catches. Patterns in species numbers and community structure of fishes and invertebrates found with eDNA were similar to those observed with morphological analyses. The majority of the fishes detected in the beam trawl catches (83%) were also detected with 12S eDNA, while only a small fraction of the invertebrate species found in the catches (27%) were detected with COI eDNA metabarcoding. Detailed time and cost tracking illustrated that beam trawl analyses are 42.5% faster and 53% cheaper than eDNA metabarcoding, however, eDNA monitoring is much less invasive for the environment and for the species of interest and can be applied in areas where bottom trawl sampling is not allowed (e.g. marine protected areas, offshore wind farms). The two case studies investigating the occurrence of NIS in harbors both showed that eDNA outperformed morphology based analyses since three to four times more NIS were detected with eDNA. However, bulk DNA metabarcoding of zooplankton samples was able to detect 10 additional NIS compared to eDNA, suggesting that this method is the preferred choice for detecting NIS. The Ostend case study further illustrated that rapid, real time sequencing devices such as the Minion from Oxford Nanopore can be used for DNA-based monitoring of NIS which is expected to substantially reduce time and costs associated with NIS monitoring. Finally, the case study near artificial reef sites showed that long read sequencing of a 2kb barcode allowed the detection of fish species for which the shorter fragment (163-169 bp) lacks species level resolution. However, comparison with short read data and populating reference databases with long read barcodes of North Sea fishes are needed to fully grasp the potential of this 2kb fragment for detecting fishes with eDNA. The overall findings of the eDNA pilot underscore the significance of DNA-based methods (eDNA and bulkDNA) in advancing our knowledge, monitoring practices, and management strategies of marine ecosystems, but also highlight that standardization and harmonization of protocols across member states are needed to fully grasp the benefits of this non-invasive method.

2. Introduction

The North Sea experiences a variety of human activities such as the construction of offshore wind farms for renewable energy, exploitation of living and mineral resources, fisheries, and transportation of goods. These activities are driven from an economical perspective and their sustainable exploitation requires that the health of the North Sea ecosystem remains in a good status. The North Sea is one of the most invaded marine ecoregions in the world due to the high level of transportation in this area (Molnar et al., 2008). In 2020, Verleye et al. reported that 79 non-indigenous species (NIS) had established in the Belgian part of the North Sea. These numbers are expected to increase in the coming decades with a continued rise in population size, global trade and as a result of global climate change (Sardain et al., 2019). The construction of offshore wind farms in the North Sea also increases, resulting in more artificial hard substrates. This provides a new habitat for colonizing epifaunal species, including non-

indigenous species (De Mesel et al., 2015), but also offers increased food and shelter opportunities, which attracts various fish and epibenthic species, the so-called artificial reef effect (Degraer et al. 2020). In contrast, many naturally occurring reefs consisting of flat oysters, large rocks or sunken trees have been removed from the North Sea bottom, as such depleting the "reef" habitat, and therefore also the reef-associated marine species. Reef restoration activities are currently in place by reintroducing flat oysters and adding reef substrate in some areas of the North Sea.

Proper management measures need to be taken for safe keeping North Sea ecosystem health, and such measures benefit from fast and accurate monitoring of marine biodiversity. Depending on the taxonomic group of interest, different sampling techniques are being used, such as beam trawls for epibenthos and fish monitoring, plankton nets for zooplankton monitoring and scrape samples for invertebrates associated with hard substrates. The morphological identification of taxa requires expert knowledge, which is declining (Wheeler et al., 2004), and processing of samples is time consuming, labor intensive and expensive. Consequently, the datasets resulting from these traditional methods are often spatially and temporally sparse.

The use of environmental DNA (eDNA) may overcome the slow and expensive traditional monitoring using morphology based identification. Organisms shed DNA by releasing scales and skin tissue, mucus, blood, faeces and gametes which leave a blueprint of the species present in that region. This eDNA is collected from the marine environment by filtering seawater, and is in this way a non-invasive sampling technique. The DNA molecules are extracted from the filters and then PCR amplified using universal primers that target many species at once (metabarcoding) or using species specific primers that only target one species (dPCR, qPCR or ddPCR). The metabarcoding approach has the advantage that it provides a good overview of all species that have released eDNA, providing that the used primers match the eDNA and that there is an extensive reference database available to link the DNA sequence to species names. The primers often target short mitochondrial DNA barcoding regions (cytochrome oxidase subunit I gene (COI), the ribosomal RNA genes 12S and 16S, or the cytochrome B gene) or the nuclear encoded 18S rRNA gene. The COI gene has the most complete reference database and is often used to target invertebrate species, while the 12S rRNA and 18S rRNA genes are used to investigate fishes and plankton, respectively. Ideally, the amplified gene fragment should have low variation within species and a high number of differences between species so that the DNA sequence can be used to identify up to species level. Species resolution may increase with increasing fragment length, and long read high throughput sequencing provided by Oxford Nanopore may be a promising method to increase species resolution in metabarcoding studies. Recent studies have shown that eDNA typically detects more fish and phytoplankton species than what is reported based on traditional data (Thomsen et al., 2012; Fraija-Fernandez et al. 2020; Kim et al 2019). Smaller fish species (van Bleijswijk et al, 2019) or species that are difficult to catch (Bakker et al, 2017) can also be detected using eDNA. Studies show that different habitat types in the marine environment have different eDNA profiles, even at a local scale of 5 km (Jeunen et al. 2019, Staehr et al., 2022), which shows the potential of eDNA for biomonitoring (Sigsgaard et al., 2020, Fraija-Fernandez et al. 2020). The potential of eDNA for various monitoring questions in shallow and highly dynamic systems such as the North Sea currently remains unclear.

The eDNA pilot within the GEANS project aimed to explore whether eDNA would be applicable for NIS monitoring and for characterizing biodiversity in offshore wind farms, shipwrecks and oyster reefs in the North Sea. Three case studies were conducted: 1/ fish and invertebrate monitoring in and near offshore wind farms in the Belgian part of the North Sea, 2/ NIS detection

in harbors in Belgium and Germany and 3/ fish monitoring near artificial reefs. The first two studies explicitly compared eDNA results with morphology based analyses of the same set of samples to investigate how eDNA performs compared to traditional monitoring methods. Time and costs associated with the two methods were tracked to investigate whether eDNA is faster and cheaper than morphological identification of beam trawl samples. The last case study focused on designing new primers to target a 2000 bp long fragment and explored whether this long fragment can be used for eDNA monitoring of fishes.

3. Pilot design

The eDNA pilot consisted of two case studies that simultaneously generated eDNA metabarcoding and morphology based data for a/ environmental impact monitoring of offshore wind farms (OWF) in the Belgian part of the North Sea (BPNS) and b/ monitoring of non-indigenous species (NIS) in harbors in Belgium and Germany. A third case study involved the application of long read metabarcoding data to monitor fish diversity near artificial reefs in The Netherlands.

a. Case study on environmental impact assessment of offshore wind farms (Belgium)

i. *Design and monitoring objective*

The offshore wind farm area in the eastern part of the BPNS, with an installed capacity of 2.26 GW, covers 238 km² (Figure 1) in which no fishery activities are allowed. Effects on soft sediment epibenthos and demersal fish in the two oldest wind farms C-Power and Belwind have been investigated since 2008 using a beam trawl (De Backer *et al.* 2020). These beam trawl samples are sorted and fishes and epibenthic invertebrates are morphologically identified onboard the research vessel. This pilot was designed to evaluate whether eDNA reflects fish and epibenthic invertebrate communities detected with morphology based analyses around offshore wind farms and to determine whether eDNA-based monitoring is cheaper and faster than morphology based analyses of beam trawl samples.

ii. *Collection of samples*

In total, 12 coastal and 18 offshore sites located within and outside C-power (zone 1) and Belwind OWF (zone2) were sampled in autumn 2021 during two different North Sea field campaigns (Figure 1). All samples at the coastal sites, with the exception of site ft230, were collected with the research vessel Simon Stevin in September 2021. The samples from the offshore locations and the coastal location ft230 were collected during the field campaign with the Geo Ocean V in November 2021.

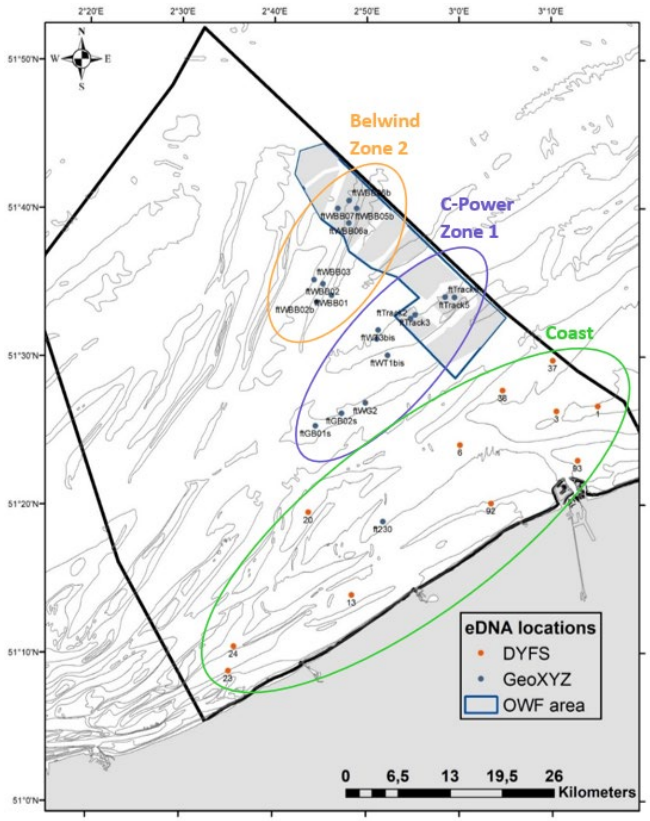


Figure 1: Map of the sample locations within the BPNS. During the DYFS campaign (orange dots) in September 2021 samples from 11 coastal locations were collected (green oval). In November 2021, 19 samples were collected from 18 offshore locations and 1 coastal location, ft230, during the GeoXYZ campaign (blue dots). The offshore zone encompasses the Belwind OWF (orange oval), the transition zone is located around the C-Power OWF (purple oval).

At each location, seawater samples were collected circa 1 m above the seafloor using a 10 L Niskin bottle deployed from the research vessel. A single Niskin bottle was used on the Geo Ocean V with five biological replicates taken one after the other, while a Niskin carousel was used during the Simon Stevin campaign where three biological replicates were collected from one carousel deployment. From each 10 L Niskin bottle, a subsample of 2 L was collected in clean commercial drinking water bottles using a sterilized 200 µm mesh nylon prefilter to remove bigger pieces of debris. Between samples, the Niskin bottles were rinsed with drinking water. Nine field negative controls were collected on different sampling days by collecting drinking water from the Niskin bottles after rinsing. During the Geo Ocean V campaign, all water samples were immediately filtered, while the bottles from the coastal campaign were stored in the dark at -20 °C until further processing in the lab. After collection of the water samples, beam trawling was done to obtain morphology-based identification of the epibenthos and demersal fish communities. Samples were collected using an 8-meter shrimp beam trawl (22 mm mesh size in the cod-end equipped with a bolder chain in front of the ground rope. The beam trawl was towed for 15 min at an average speed of 4 knots over the ground along with the current. Epibenthos and fish were counted and identified to species level. Due to technical difficulties with the beam trawl, no morphological samples were taken at the locations ftGB01, ftGB02, ftWBB01, ftWBB02b, ftTrack2, ftTrack3, ftWBB05b and ftWBB06b.

ii. *Lab processing*

Each water sample was filtered over a Sterivex 0.45 µm filter (Sterivex-HV Filter, with Luer outlet, Merck – Millipore) using a Masterflex pump with double pump head. After filtering, the Sterivex filters were sealed with a sterile Luer-lock™ cap and stored at -20 °C until further processing. Between samples, the tubes of the Masterflex pump were treated with 10 % bleach and rinsed with 125 ml bottled water. Six filter negatives were included by filtering bottled water over a Sterivex 0.45 µm filter. eDNA extraction of the filtered samples was conducted in the laminar flow cabinet in a PCR-free designated clean room. Before and after use, a 15 min UV-treatment was in place and all surfaces were successively cleaned with 10 % bleach and 70 % ethanol. Extraction was performed using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's protocol, except for the addition of 800 µl lysis buffer to the Sterivex filter. DNA negative controls existed of three blank Sterivex filters that were treated the same way.

A one-step PCR amplification targeting the COI or 12S gene fragment was conducted on all eDNA extracts. The PCR reactions were performed in triplicate in a total volume of 25 µl containing 12.5 µl KAPA HiFi Hotstart 2x ReadyMix (Roche), 0.5 µl Bovine Serum Albumin (BSA) (10 mg/µl), 1 µl of each primer (2.5 µM), 7 µl UltraPure™ water (UltraPure™ DNase/RNase-Free Distilled Water, Invitrogen™) and 3 µl extracted eDNA. PCR negative controls were included using 3 µl of UltraPure™ water instead of eDNA. The 12S sequences were amplified using the MiFish primers developed by Miya et al. (2015) which target a 163–185 bp region of the mitochondrial 12S rDNA. The universal forward and reverse primer pair (MiFish_U) was degenerated to simultaneously target Osteichthyes and Elasmobranchs (MiFish_U/E_F: 5'-GT(C/T)GGTAAA(A/T)CTCGTGCCAGC-3', MiFish_U/E_R: 5'-CATAGTGGGGTATCTAATCC(C/T)AGTTTG-3'). The COI target sequence (313 bp) was amplified using the mICOIintF (5'-GG(A/T)AC(A/T)GG(A/T)TGAA(A/T)GT(A/T)TA(C/T)CC(C/T)CC-3') and jgHCO2198 (5'-TANAC(C/T)TCNGG(A/G)TGNCC(A/G)AA(A/G)AA(C/T)CA-3') primers designed by Leray (2013). PCR cycling conditions began with 3 min of denaturation at 95 °C, 40 cycles of denaturation for 20 s at 98 °C, annealing for 15 s at 62 °C and elongation for 15 s at 72 °C, and ended with a final elongation step of 5 min at 72 °C for the 12S barcode. The COI barcode was amplified with an initial denaturation of 3 min at 95 °C, 40 cycles of denaturation for 30 s at 98 °C, annealing for 30 s at 54 °C and elongation for 30 s at 72 °C, and ended with a final elongation of 5 min at 72 °C. For a subset of each of the technical replicates, the quality was checked on the Bioanalyzer (2100 Bioanalyzer Instrument, Agilent). The three PCR replicates were pooled separately, resulting in three pools. Each pool of uniquely tagged PCR products was purified using magnetic CleanNGS beads (CleanNA), by adding 1x (12S) or 0.8x (COI) the total volume of the pool and quality checked with the Bioanalyzer. The three PCR pools of 12S and COI were ligated with different Illumina TruSeq adapters, pooled and sent for sequencing to Admera Health Biopharma Services (Miseq, 2 x 300bp). For each marker gene, one flow cell was used.

iii. *Bio-informatic processing*

The quality of the raw Illumina MiSeq sequencing reads was verified with FASTQC v0.11.9 (Andrews, 2010). The pair-end reads were then reorientated, demultiplexed and trimmed by using cutadapt v3.5 (Martin, 2011). After demultiplexing, DADA2 v.1.20.0 (Callahan et al., 2016) was used for denoising, merging, removing of chimeric reads, and dereplication of the demultiplexed sequences. Before taxonomic assignment, the three PCR-replicates of each sample were concatenated and the taxonomic assignment of the resulting amplicon sequence variants (ASVs) was performed against a custom made reference database using rdp in DADA2 with a minimum bootstrap of 80 %. The 12S database contained the reference sequences of 122

fishes that have been detected in beam trawls during the monitoring campaigns of ILVO in the BPNS. The GEANS COI reference database for macrobenthos was complemented with sequences from BOLD and MIDORI and contained 56 089 sequences from 53 178 unique species. Only five species found in the beam trawl surveys had no COI reference sequence available. ASVs that remained unassigned with rdp were run with BLASTn v2.12.0 (Altschul et al., 1990) against the custom made reference databases. The ASVs that remained unassigned after this were run with BLAST against the GenBank nucleotide database (from October 2022). The results from the BLASTn run were only considered if the identity matches were above 97 %. After receiving the full taxonomic assignment, the samples from both datasets were rarefied to contain 25 000 sequences (COI-metabarcoding) and 10 000 sequences (12S-metabarcoding) using the rrarefy function in vegan v2.5-7 (Ricotta et al., 2019). After rarefaction, the ASVs were cleaned by removing all ASVs with no reads left after rarefaction and contaminant ASVs. ASVs were considered as contaminants if their occurrence in any of the negative control samples exceeded 10 % of their total read abundance across all the samples. For the 12S metabarcoding only ASVs that were assigned to the Pisces order level were used for further analysis. For the COI metabarcoding only the ASVs assigned to invertebrates were used for further analysis.

iv. *Data-analyses*

All data analyses were done with the rarefied and cleaned data and samples that contained less than 10 000 (12S) or 25 000 (COI) reads were removed. To compare the species detected with eDNA and morphology based trawl surveys, a Venn diagram was constructed in R using VennDiagram v1.6.20 (Chen, 2022). The average number of species and Shannon index were calculated per location and plotted using the ggplot2 package v3.3.5 (Wickham, 2016). To investigate differences in species richness and Shannon index between zones and methods, a two-way ANOVA with factors location (5 levels: coast, C-Power OWF, C-Power control, Belwind OWF, Belwind control) and method (morphology or eDNA) was performed for 12S and COI, after testing for equality of variance with Levene's test. To analyze the differences in community composition between the locations a distance matrix based on the Bray-Curtis dissimilarity was generated using the transformed read counts (eDNA) or specimen counts (morphology). To reduce the weight of a few highly abundant species, the four datasets (12S eDNA, COI eDNA, morphology fish and morphology invertebrates) were log transformed using the decostand function in vegan v2.6-4 (Ricotta et al., 2019). For each dataset , the differences in community composition between the coastal and offshore locations inside and outside the OWF were tested using a permutational multivariate analysis of variance (PERMANOVA, adonis 2, 9999 permutations) with location as only factor (5 levels). Next, a two way PERMANOVA with main factors location (outside vs inside OWF) and the zone (zone 1 vs zone 2) and their interaction was performed to investigate whether there were differences in fish and invertebrate community compositions inside and outside the OWF area. All analyses were followed by a multivariate homogeneity of group dispersions test (BETADISPER) and pairwise multilevel comparison to detect the contribution of the levels to the statistical significance (function "pairwise.adonis2; permutations=9999, PAIRWISEADONIS package, Martinez Arbizu, 2020). The species community compositions were also visualized using a nonmetric multidimensional scaling (NMDS) with two dimensions (k = 2).

v. *Calculation of cost and time for eDNA-based and morphology based sample processing*

Time and cost tracking was done for ten locations in and around the OWFs where both sample types were collected. For beam trawl sampling, time tracking included the time of beam trawl deployment, time needed for sorting, counting and measuring fish and invertebrates onboard the research vessel and for additional processing of invertebrates under a microscope in the lab, quality control of the data and adding the data into the database. For the eDNA-based analyses, time tracking included the time needed to collect five subsequent replicate water samples in each of the ten locations using a single Niskin bottle and to filter water samples with a peristaltic pump and Sterivex filters onboard the research vessel. Two samples can be filtered simultaneously, and the filtering step was timed two times for each of the ten locations, summed and then scaled to 50 water samples. The time for the steps conducted in the lab were estimated for eDNA extraction of 24 samples, PCR amplification in triplicate, quality check of the PCR products on a Bioanalyzer, pooling of samples, library clean-up and final quality check of the libraries. Sequencing is performed by an external company, time until retrieving the sequence data was not included. Time needed for bioinformatic processing of the sequences was estimated starting from the raw reads, demultiplexing, following the DADA2 pipeline and taxonomic assignment with rdp and BLASTn until the final species list was obtained. The timings were summed for all steps as total time in hours. For eDNA analyses, a distinction was made between total time and hands-on time since overnight incubation during eDNA extraction, PCR amplification, retrieving the sequencing data, and running the DADA2 and BLASTn commands do not require personnel time. Costs include consumables, sequencing and personnel costs for each method. Shipping costs are not included since they are the same for both methods. For personnel costs, the hourly rate of the executing person was multiplied by the hands-on time spent for each method respectively.

b. Case study on NIS monitoring in harbors

i. Design and monitoring objective

This case study was designed to evaluate whether eDNA can be used to improve traditional NIS monitoring based on morphological identification in the harbor of Ostend (Belgium) and in the harbor of Rostock (Germany). Harbors provide favorable conditions for NIS to establish and thrive. These areas often have protected waters, nutrient-rich sediments, and artificial structures like docks and pilings that provide attachment surfaces. Harbors can act as "hotspots" for NIS, with high concentrations and diverse communities of invasive organisms. Ostend was chosen because of its proximity to the laboratory facilities of VLIZ and because of its long history in marine biological research. Rostock's harbor was chosen because it serves as a transshipment point for goods traveling between northern and eastern Europe. Its strategic location provides easy access to the Baltic Sea and enables efficient transportation of goods to countries in Scandinavia and the Baltic states. The port's connectivity is rendering it a potential major hub for the spread of NIS between the North Sea and the Baltic Sea.

ii. Collection of samples

In the harbor of Ostend (Belgium), water samples for eDNA analysis and scrapes from settlement plates for morphological identification were taken at three different locations during late spring and summer 2020: Vuurtorendok (51.237634°N, 2.931726°E), Ponton Overzet (51.234216°N, 2.927157°E), and Marina Mercator (51.228220°N, 2.923981°E) (Figure 2). For eDNA, five biological replicates and two field control samples were collected for each season and at each

location from 1 m depth using 2 L sterile Nalgene bottles. The water samples were stored in a cooled container in the field and immediately filtered upon return to the lab through 1.2 µm cellulose nitrate filters (47 mm diameter). After filtration, samples were stored at -80 °C until further processing. Morphological identification was done on scrapes from settlement plates that were deployed at the three locations, following the OSPAR-HELCOM protocol for NIS monitoring (OSPAR, 2019). Two sanded grey PVC plates (15 x 15 cm) were put out at 1 m and 7 m depth for each location during late spring. The limited depth range at Marine Mercator only allowed one settlement plate to be deployed at 1 m depth. Benthic communities established on the plates were sampled after two months during late summer. Samples were preserved in DESS in the field and morphological identification was done back in the lab using a stereomicroscope following the recommendations of Hayward & Ryland (2017).

Figure 2: Map of the sample locations in the Ostend harbor, Belgium. Credit: Google Maps, 2021.

The harbor of Rostock is the largest harbor area in the German Baltic Sea and is characterized by a salinity gradient influenced by the river Warnow. The sampling methodology followed the OSPAR-HELCOM recommendation whenever possible (OSPAR, 2019). Samples were taken at three different locations within the greater harbor area: Station HRO-1 (54°05'40.3"N 12°07'00.8"E; marina in city harbor, southernmost location, strongest influence of the river Warnow, lowest salinity), HRO-2 (54°08'39.3"N 12°05'49.9"E; international port), HRO-3 (54°10'50.7"N 12°05'27.7"E; marina in Warnemünde, in immediate proximity of the Baltic Sea, highest salinity; Figure 3) in 2020 and 2021. For eDNA analysis, a one liter water sample was taken in each of the three locations in each sampling year and stored at 4 °C until processed a few days after sampling. Zooplankton was sampled with nets of two different mesh sizes as follows: in 2020, two replicates were taken at each of the three stations HRO-1 to -3, resulting in a total number of 12 zooplankton samples, half of which were sampled with a 100 µm mesh (Z100) and the other half with a 335 µm mesh (Z335). Since sequencing was unsuccessful for both 335 µm samples from HRO-1, this results in a total of 10 samples examined in 2020. Again, in 2021, three replicates per station were taken with both mesh sizes for quantitative comparison between years. Successful sequencing of all samples results in a total number of 24 zooplankton samples examined in 2021. Morphological examination of the samples was performed only in 2020, and was done by the Institute for Applied Ecosystem Research (IfAÖ,

Rostock). All zooplankton samples were immediately fixed in 500 mL of 96 % Ethanol and stored at -20 °C until further processing.

Figure 3: Sampling locations of water and zooplanktonic samples for eDNA survey in the harbor of Rostock (Germany).

iii. Lab processing

For the eDNA samples in the harbor of Ostend, a quarter of the filter was cut into small pieces while the rest was kept in the -80 °C ultra freezer for long-term storage. Further sample processing included tissue ligation in proteinase K solution with CTAB buffer, followed by an extraction and clean-up of the DNA with a simple and low-cost protocol using chloroform, sodium acetate and isopropanol. Extracted DNA was stored in 10 mM Tris buffer at -20 °C until further processing. All eDNA extracted samples underwent a two step PCR approach to target the 18S rRNA V4-V5 region for amplicon sequencing. The F-566 and R-1200 primer pair proposed by Hadziavdic et al. (2014) was used for amplification of the 18S rRNA target region in the first PCR. These primers were adjusted with overhanging sequences at the 5' – end to facilitate the attachment of PCR barcodes during a second PCR (forward primer: 5'-TTTCTGTTGGTGCTGATATTG-3'; reverse primer: 5'-ACTTGCCTGTCGCTCTATCT-3'). The latter step is required to enable the pooling of individual samples into one library for amplicon sequencing. For each sample, three technical replicates were processed during the first PCR, to account for amplification biases, using the Phusion Hot Start II High-Fidelity PCR master mix (ThermoFisher Scientific). For the first PCR, each 25 µL reaction consisted of 12.5 µL Phusion master mix (2x), 5' – tailed forward and reverse primers (0.5 µM each), gDNA template (0.2 ng) and nuclease free water. The PCR cycling conditions included an initial denaturation at 98 °C for 30s, 30 cycles of denaturing (98 °C for 10 s), annealing (60 °C for 15 s), and extension (72°C for 15 s), followed by a final extension step at 72°C for 7 min. After the first PCR, the technical replicates were pooled and cleaned using a CleanPCR kit (GC Biotech, CPCR-0005). The purified amplicons were

quantified on a Qubit 3 Fluorometer using the Qubit dsDNA Broad Range Assay kit (ThermoFisher Scientific). Oxford Nanopore barcodes from the PCR Barcoding Expansion Pack (EXP-PBC096) were added to the amplicons in a second PCR step. Each PCR reaction (25 µL) contained 12 µL Phusion master mix (2x), barcode primer mix (0.5 µL), PCR1 product (0.2 ng) and nuclease free water. The PCR conditions were as followed: initial denaturation (98 °C for 30 s), 15 cycles of denaturation (98 °C for 10 s), annealing (60 °C for 15 s), and extension (72 °C for 15 s), and a final extension step at 72 °C for 7 min. After the PCR, another clean-up and quantification step was performed, and the purified amplicons were subsequently pooled in equimolar amounts. Next, a pooled barcoded library of 1 µg was prepared in 47 µL nuclease-free water followed by an end-prep step using the NEBNext Ultra II End repair reaction buffer and enzyme mix according to the manufacturer's instructions, and a third clean-up and quantification step. Finally, a Ligation Sequencing Kit (ONT, SQK-LSK109) was used to add sequence adapters to the barcoded amplicons according to the manufacturer's protocol, a final clean-up and quantification step was performed, and the samples were sequenced on the MinION Mk1B sequencer using a FLO-MIN106D flow cell (Oxford Nanopore Technologies).

For eDNA and zooplankton analyses in the harbor of Rostock, 300 mL of each water sample was filtered using a vacuum pump (filter pore size of 0.7 µm; zooplankton samples through filter pore size of 2.7 µm) and then centrifuged for 30 min to evaporate ethanol residues. DNA was subsequently extracted from the filters using the E.Z.N.A. Mollusc Kit (Omega Bio-Tek) following the manufacturers recommendations. The selected primer pair mlCOIintF (forward; Leray et al. 2013) and jgHCO2198 (reverse; Geller et al. 2013), spiked with the Illumina overhang (Illumina 2011), were used to target the COI region (~313 base pair fragment) for metabarcoding. The following settings were used for the qPCR cycle: Initial denaturation at 98 °C for 3 minutes, x25 cycles of denaturation at 98 °C for 30 seconds, primer hybridization at 50 °C for 30 seconds, elongation at 72 °C for 30 seconds, and a final elongation step at 72 °C for 5 minutes. From each successful amplification, 1 µL of product and a unique combination of dual Nextera-compatible IDT for Illumina 10 bp was used for the second PCR with identical settings, but only 13 cycles in total. Successful amplifications were purified and normalized using the SequalPrep Normalization Plate 96 kit (Invitrogen), and 2 µL of each product was added to a pooled library. The concentration of the final library was measured using the Qubit 3.0 Fluorometer (Invitrogen), and the molarity was measured using the Collibri Library Quantification Kit (Invitrogen). The pooled library was denatured and 20% PhiX genomic control DNA was added before a test sequencing run with a MiSeq Reagent Nano Kit v2 (250 cycles paired-end) and a final run with a MiSeq Reagent Kit v3 (300 cycles paired-end) on a high-throughput Illumina MiSeq sequencing platform.

iv. Bio-informatic processing

For the harbor of Ostend samples, basecalling was done in Guppy (v4.3.4, Oxford Nanopore Technologies Ltd., UK), followed by read demultiplexing with qcat (v1.0.1, , Oxford Nanopore Technologies Ltd., UK) and filtering of the reads with NanoFilt (v2.8.0; De Coster et al., 2018). Next, a density based clustering approach implemented in ASHURE (V1.0.0; Baloğlu & Chen, 2021) was used for error correction (by consensus of clusters) and potential chimeras were removed with VSEARCH (Rognes et al., 2016). Taxonomy annotation was performed using the SILVA 138.1 database with BLASTn. Only sequences with an assignment identity higher than 90

% and an alignment length of at least 500 bp were retained. Sequence variants identified as anything other than Eukaryota were not of interest and discarded. The taxonomy was checked with the WoRMS database using its REST API, and AphiaIDs were allocated to the matching scientific names. Taxonomic units that were not recognized by WoRMS and hence could not receive an AphiaID, were not considered from downstream analysis and interpretation (mostly protists). NIS species were considered if identity matches were above 98 % and all but one NIS species that we report had a similarity score of more than 99 %.

For harbor of Rostock samples, the demultiplexed NGS reads were trimmed by primer's sequences using BBmap (sourceforge.net/projects/bbmap/). Further the Illumina reads were denoised, merged to make contigs, filtered by length and quality scores, chimera detected and dereplicated to high resolution ASVs using DADA2 pipeline (Callahan et al., 2016). A custom script (SGN Metabarcoding pipeline) was used to blast the ASVs against the NCBI database incorporating BLASTn pipeline. The ten best blast hits were retrieved and pooled with GEANS reference library of macrofauna and this merged dataset was used as final custom blast 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 library. Taxonomic annotations were assumed to be correct at species level after a two-step-quality-control of the database match by filtering ASVs with ≥97 % percent identity (*pident*) and ≥90 % query coverage (*qcovs*). ASV outliers were identified according to the frequency distribution of the following parameters: sequence length (read length), *pident* and *qcovs* of the sequence length compared to the best matching GenBank sequence. Outliers are defined as values that fall outside 1.5 times the quartile distance. The lower limit for length was 226 bp, for *pident* 71.7 % and for *qcovs* 75 %. Further, all ASVs were checked against WoRMS (WoRMS Editorial Board, 2023) to ensure accurate taxonomic assignment at species level and were given an AphiaID. Taxonomic units that did have a match in the database (no AphiaID), were removed from downstream analysis and interpretation.

v. Data-analyses

The Ostend dataset was rarified to an even sequencing depth of 42.783 using the rrarefy function in vegan v2.5-7 (Ricotta et al., 2019) and a seed of 711. Statistical analysis and visualizations of the data were performed in R (v3.4.4; R Core Team, 2021), with the use of the vegan (v2.5-7; Oksanen et al., 2020), phyloseq (McMurdie & Holmes, 2013), and ggplot2 packages (Wickham, 2016). To analyze the variability in the eDNA community compositions between the two time points (late spring and summer) and across the three locations in the Ostend harbor, a two-way Permutational Multivariate Analysis of Variance was used with the adonis2 function (PERMANOVA; Anderson, 2017) with main factors time and location and using 10,000 permutations, and with a seed of 711 for reproducibility. For this, a distance matrix was calculated based on Bray-Curtis dissimilarities (reflects relative abundance). Multivariate homogeneity of the dispersions amongst the predefined groups (locations and time points) was also tested using the Betadisper function from the vegan package. The latter is required to determine if the community variability is a result of a 'location' effect (the predefined groupings) or simply a 'dispersion' effect (heterogeneity of dispersions). The community comparisons were visualized in an ordination plot using non-metric multidimensional distance scaling (nMDS),

based on the Bray-Curtis distance matrix and 2-dimensions (k = 2). Relative abundances were calculated for each sample and displayed in a barplot at phylum level.

Various rarefaction analyses were conducted to assess whether sampling was thorough enough to detect all the species that were present in the study area. For this, species detections were re-sampled per sampling event (to test whether five replicates were sufficient), per sampling date (to test whether better temporal coverage would have yielded more species) and per sampling site (to test whether the selected microhabitats within the harbor were representative for the whole study area). Rarefaction was done with resampling with replacement but replacement was capped at the number of samples because a species could only be detected once per sample. Since the full analysis included a total of 30 samples, the rarefaction analysis was still expected to yield smooth saturation curves.

The identification of the presence of non-indigenous or 'alien' species in the Ostend harbor was done using a python script. Here, the AlphiaID of each taxonomic unit, resolved to species level, was used to look up the species distribution in the WoRMS register and apply that to the region of interest (Ostend harbor). Additionally, species were also compared against the list of established non-indigenous species of the Belgian part of the North Sea published by "VLIZ Alien Species Consortium" (Verleye et al., 2020).

The filtered (*pident* 71.7 %, *qcovs* 75 %) Rostock dataset was rarified to an even sequencing depth of 4.500 using the rarefy_even_depth function from the phyloseq package (v1.16.2; McMurdie & Holmes, 2013) and a seed of 711. This removed 16 samples from the dataset: two eDNA samples from 2020 and nine from 2021, four Z335 samples from 2021; and one Z100 sample from 2021. The nine eDNA samples from 2021 accounted for all three replicates taken at each location, thereby removing this whole group from the dataset for further analyses. Compositional differences of eDNA and zooplankton communities were examined for the three methods (eDNA, Z100 and Z335) with a one-way PERMANOVA using the adonis2 function on the rarified dataset that was selected for macrofaunal taxa with an AphiaID. Due to the reduced sample size after bioinformatic processing, community variation could not be investigated for time point, location, or the interaction of these groups with the three methods. Homogeneity of variances amongst the different methods was determined with the Betadisper function. Bray-Curtis dissimilarities were used to calculate the distance matrices. The community variations were also illustrated in an ordination plot using non-metric multidimensional distance scaling (nMDS) based on 2-dimensions (k = 2). Similar to the dataset from the Ostend harbor, all community diversity analyses were performed with 10,000 permutations and with a seed of 711 for reproducibility. Relative abundances were calculated for each sample and displayed in a barplot at phylum level.

c. Case study on the use of long read metabarcoding in artificial reefs in the North Sea

i. Design and monitoring objective

This case study was designed to evaluate whether eDNA based monitoring can be used to measure the biodiversity of fish near non-biogenic reefs provided by man-made structures such as shipwrecks and wind turbine monopiles. These structures provide complex shapes for a wide range of benthic and pelagic organisms to settle. The current standard approach for fish eDNA is based on the amplification of a short 150 bp fragment of the 12S rRNA gene (Mifish, Miya et. al., 2015). However, these short amplicons sometimes lack the required resolution to identify

the obtained eDNA fragments down to species level. We assessed whether Oxford Nanopore technologies (ONT) Nanopore sequencing can be used to sequence larger, 2 kb DNA fragments stretching both the 12S and 16S rRNA gene, that have the potential to allow a more accurate identification on species level. We designed new primers and tested their amplification using eDNA samples from the Burgers Zoo ocean aquarium. Next, we selected different reef systems in the North Sea to validate their applicability in North Sea settings: three different shipwrecks in the North Sea between the Netherlands and England, and three locations north of Schiermonnikoog. These were inside wind park Gemini and on the location of a flat oyster restoration project at the Borkum reef grounds. The third location was a sandy bottom non-reef site directly in between these sites. This sandy bottom halfway site acted as a non-reef control site.

ii. Collection of samples

From the Ocean aquarium in Burgers Zoo, two samples were collected with a 1L plastic sterilized container collecting a total of 2L just under the surface of the aquarium. The negative control here was 1L of demineralized tap water. In the North Sea, the samples were taken while SCUBA diving from wreck 1 (55.1821 N, 03.4446 E using the WGS84 reference system), wreck 2 (55.2609, 03.5117) and wreck 3 (55.0774, 02.5087). At each sample location, three replicates of approximately 2L of seawater were taken by pumping the seawater in a balloon using a hand pump (dx.doi.org/10.17504/protocols.io.6yfhftn). North of Schiermonnikoog, samples were collected inside Gemini Wind Park (54.0109, 6.0781), halfway between Gemini Wind Park and the Borkum Reef Grounds on sandy substrate (53.8645, 6.2145) (Sandy bottom) and at Borkum Reef Grounds (53.7016, 6.3467). All samples were taken at slack tide during neap tides. Samples were taken using a 2.5L Niskin bottle at one depth between 0.5-1 m above the seafloor on 2, 17 and 31 July 2020. Samples were taken in triplicates, resulting in a total of 27 samples. All water samples, both from the aquarium and the North Sea were immediately filtered with disposable Thermo Scientific Nalgene Rapid-Flow sterile disposable Filter Units CN (Cellulose nitrate, pore size: 0.8µm). Filters were stored in 2 ml tubes with 400 µL pre-filled DNA preservatives as ATL ((ATL, Qiagen, USA, ocean aquarium) and Zymo DNA/RNA shield (Zymo, USA, North sea samples). Samples were immediately stored at -20°C for a maximum of one month before further processing.

iii. Lab processing

The primers were designed based on the adjacent ribosomal genes 12S and 16S of the mitochondrial genome of bony and cartilaginous fish present in either the North Sea or the Ocean aquarium of Royal Burgers' Zoo, Arnhem, the Netherlands. Primers were designed in silico in Geneious prime 2019.0.4 (Kearse et al., 2012) and based on the in NCBI available mitochondrial genomes of the target species. This resulted in a long read universal fish primer pair targeting a 2 kb fragment from 450 bp downstream the 12S rRNA gene in forward direction and 300 bp upstream the 16S rRNA gene in reverse direction. The 5' ends of the primers were extended with an ONT tag (underlined) to allow for the ONT based barcode PCR:

Forward primer Fish_12S_fw1-ONT: TTTCTGTTGGTGCTGATATTGCGGATTAGATACCCYACTATGC

Reverse primer Fish_16S_rv1-ONT: ACTTGCCTGTCGCTCTATCTTCGATTGCGCTGTTATCCCTAG

Samples were extracted using the manufacturers protocol of the Quick-DNA miniprep kit (Zymo, USA). Details of the protocol are given at protocols.io (dx.doi.org/10.17504/protocols.io.6yfhftn). Three PCR replicates were processed with a master mix of 10 µL 2x Phire Tissue Direct PCR Master Mix (ThermoFisher Scientific, USA, 0.4 µL of each primer (10mM), 0.5 µL template and nuclease free water to 20 µL. The PCR program was 98 °C for 180 s, 35 cycles of 98 °C for 8 s, 59 °C for 8 s, 72 °C for 30 s followed by a final extension at

72 °C for 180 s. Amplicon sequence library preparation was performed using the SQK-LSK109 kit and PCR Barcoding Expansion 1-96 (EXP-PBC096) on a R9.4.1 flowcell (Oxford Nanopore Technologies Ltd., UK), according to the manufacturer's instructions.

iv. *Bio-informatic processing*

Basecalling was performed using Guppy (Version 4.2.2, Oxford Nanopore Technologies Ltd., UK) in high accuracy (HAC) mode. Sequences were processed using the bioinformatics pipeline Decona (v0.1.2) (https://github.com/Saskia-Oosterbroek/decona). Decona was used to demultiplex, filter read length (1800-2200 bases) and quality (q 10), cluster (at 80%) and build Medaka consensus sequences from each cluster larger than 100 sequences (Decona -d -l 1800 -m 2200 -q 10 -c 0.8 -n 100 -M). Reference databases were separately built based on sequences available in the NCBI database matching the species list for the collection held in the Ocean aquarium (last search April 2018) and for the North Sea, based on the fish species listed as present in the North Sea at https://www.nederlandsesoorten.nl/ (last search April 2021). When the whole mitochondrial genome was not available, available sequences of the 12S and/or the 16S from the species were added to the databases. Species identification was performed with the taxonomic identifier Centrifuge v1.0.4 (Kim, Song, Breitwieser, & Salzberg, 2016) with a minimal alignment length of 200 nucleotides and using only one primary assignment for each consensus sequence. If a consensus sequence was aligned with the same quality/score to two or more species, sequences were assigned to genus level. Consensus reads that could not be identified were verified using BLASTn.

v. *Data-analyses*

The Centrifuge output including the number of clusters and sequences per identification was processed in R studio v1.1.463. Sequence data were loaded as data frame in R using the packages taxize v0.9.96 (Chamberlain & Szöcs, 2013) and phyloseq v1.30.0 (McMurdie & Holmes, 2013). Prior to analysis, sequences from *Homo*, *Ovis*, *Gallus* and *Bos* genera were removed from the dataset when applicable. In the dataset of the Borkum Reef Grounds, samples that yielded no or one successful PCR amplification replicate were removed from the dataset. This resulted in the removal of several samples from 31st of July upon which was decided that all data points of that day were removed. To obtain data of overlapping species found in the Ocean aquarium of Burgers Zoo, a Venn-diagram was used. For this, the species list was obtained from the phyloseq object using the tax_glom() function. From the same species list, the species were plotted using the barplot() from phyloseq function in combination with ggplot2 (Wickham H., 2016). For the field samples, the observed number of species and Shannon index were calculated per sample and a non-metric multidimensional scaling on presence-absence data ('jaccard') was performed in combination with PERMANOVA from the pairwiseAdonis package in R (Martinez Arbizu, P. (2020) to identify significant differences between shipwrecks and between Gemini, the Borkum reef ground and the sandy bottom. For the field samples, rrarify() was used to correct for unequal sequencing depth using a rarefaction threshold of 5000 reads per sample. The latter analysis was done for each time point separately and because of the limited number of samples, samples for the two timepoints were also pooled per location. Pairwise PERMANOVA and Kruskal-Wallis posthoc tests were conducted to identify which locations were different from each other.

4. Results

a. Case study on environmental impact assessment of offshore wind farms (Belgium)

After filtering, merging and removal of the chimeric sequences and concatenation of the three PCR replicates, an average of 50.651 ± 37.400 reads per sample were obtained for the 12S dataset. DADA2 identified 3.415 ASVs, of which 1637 ASVs (48.9 %) received a taxonomic assignment. Although only 327 ASVs (9.6 %) were assigned to 64 fish species, they represented the majority (78.4 %) of the raw reads. After removing ASVs of which more than 10 % of the reads were assigned to the negative control samples and rarefaction of the data to 10.000 reads per sample, a total of 2915 unique ASVs remained of which 300 were assigned to 62 fishes at species level. An additional 9 ASVs remained unassigned at species level due to the low taxonomic resolution of the 12S barcode between closely related fish species (*Ammodytes tobianus* and *Hyperoplus lanceolatus*, and *Chelidonichthys lucerna* and *Eutrigla gurnardus*). Morphology based identification from beam trawl catches detected 42 fish species of which 35 were also detected with eDNA (Fig 4). In addition to the four species mentioned above, 12S eDNA metabarcoding failed to detect *Alosa alosa, Zeus faber* and *Buglossidium luteum* (Fig 4). The latter species was actually present in the eDNA dataset but was identified as a contaminant based on its high frequency in the negative control samples. Importantly, an additional 27 fishes were uniquely detected with eDNA metabarcoding. The detection of *Lepidorhombus whiffiagonis*, *Merluccius merluccius*, *Salmo salar*, *Salmo trutta* and *Trisopterus esmarkii* may be false positives, since they do not normally occur in BPNS and their eDNA detection was characterized by only low read numbers (<100) and detection in only 1 or 2 biological replicates.

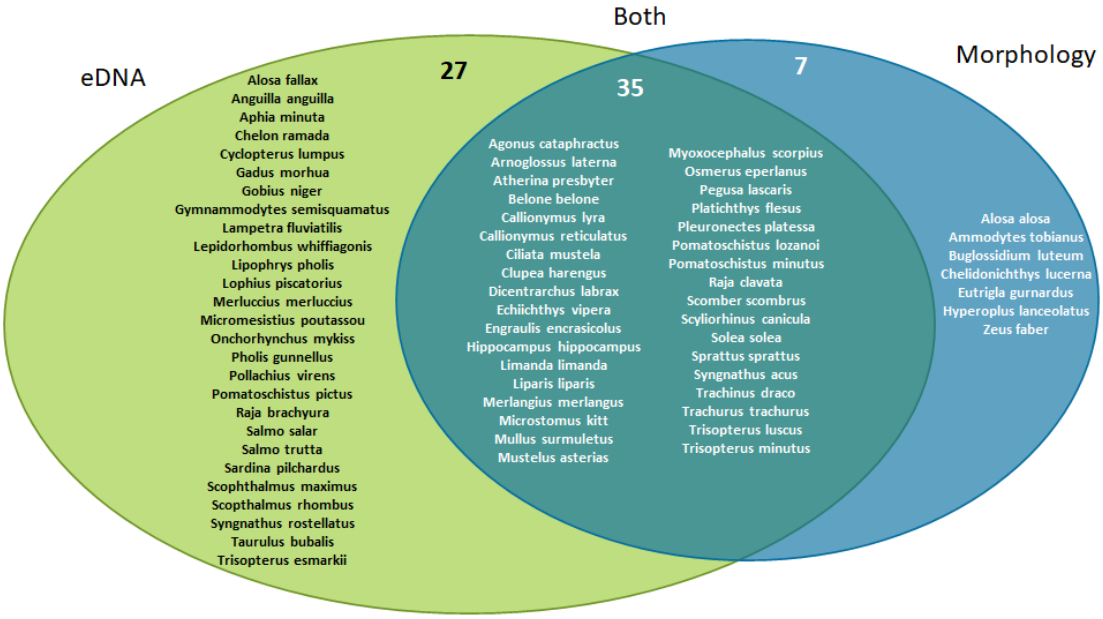


Figure 4: Overview of the fish species detected with 12S eDNA metabarcoding and bottom trawls. The Venn diagram shows the overlap between the eDNA data (green) and the trawl surveys (blue) of the coastal and offshore waters combined. Both methods were able to detect 35 species, 27 species were only detected by eDNA metabarcoding and seven species were only detected by trawling.

After filtering, merging and removal of the chimeric COI sequences and concatenation of the three PCR replicates, an average of 114.889 ± 190.739 reads per sample from November and 42.505 ± 38.792 reads per sample taken in September were available. Together these reads were allocated to 9702 ASVs, of which only 507 ASVs (5.2 %) received a taxonomic assignment. Only 311 ASVs were assigned to 148 species belonging to the kingdom Animalia, representing 17.5 % of the reads. The second most common Kingdom was Chromista (14.3 % of the reads). However, the majority (65.2 %) of the COI reads did not receive a taxonomic assignment. The cleaned and rarefied dataset contained a total of 7.918 ASVs of which 248 were assigned to 115 Animalia at

species level. Of these, 225 ASVs belonged to 94 invertebrate species. Morphology based analyses of the samples identified 52 epibenthic invertebrate species of which only 14 were also detected with eDNA metabarcoding. In contrast, eDNA metabarcoding was able to detect an additional 80 invertebrate species that were missed with morphology based beam trawl samples. Most of them were Annelida (24 species), Cnidaria (23 species) and zooplankton species (16 species) that are not targeted with the beam trawl. Moreover, COI was able to identify 19 fish species, two of which (*Ammodytes marinus* and *Zeus faber)* were missed with the 12S data. *Ammodytes marinus* was missed by the 12S barcode due to the low taxonomic resolution of the 12S barcode between closely related species of the Ammodytidae. *Zeus faber* is a rare species in the BPNS and may have been missed with the 12S barcode due to PCR-bias.

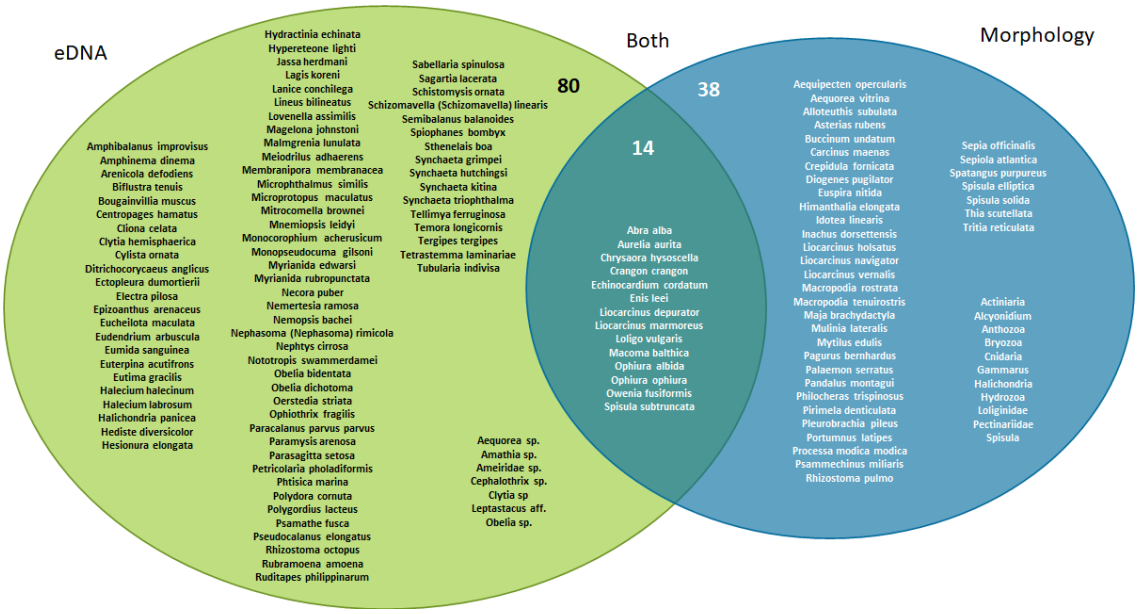


Figure 5: Overview of the invertebrate taxa detected with COI eDNA metabarcoding and bottom trawls. The Venn diagram shows the overlap between the eDNA data (green) and the trawl surveys (blue) of the coastal and offshore waters combined. Both methods were able to detect 14 species, 80 species were only detected by eDNA metabarcoding and 38 species were only detected by trawling.

For the observed species richness, no significant interaction effect was observed between the two methods (morphology or eDNA metabarcoding) and the five zones (two-way ANOVA method*location: $p = 0.24$ for fish and 0.77 for invertebrates, Figure 6). However, a significantly higher number of fish species was detected with eDNA metabarcoding than with morphology (two-way ANOVA: $p < 0.001$), while no significant difference was observed for the factor location (two-way ANOVA: $p = 0.57$). Overall, eDNA metabarcoding was able to detect more fish species per sample (10 - 30 fish species) than the beam trawl survey (6 - 18 fish species). The amount of invertebrate species differed significantly between locations (two-way ANOVA: $p = 0.03$), but not between both methods used (two-way ANOVA: $p = 0.16$). COI eDNA metabarcoding detected between five to 25 invertebrate species per sample, whereas morphology detected five to 28 invertebrates.

The Shannon diversity index of the fish dataset did not show a significant interaction effect between method and location (two-way ANOVA: $p = 0.12$) suggesting that the patterns observed across locations are similar between both methods. However, there was a (borderline) significant difference between both methods (two-way ANOVA: $p = 0.03$) and between locations (two-way ANOVA: $p < 0.001$), with highest diversity observed in the coastal samples (Figure 6).

This results from a few dominant species with high relative read abundances in the offshore zones (*Merlangius merlangus*: 46.9 % and *Limanda limanda*: 17.2 %). In contrast, a significant interaction effect between method and location (two-way ANOVA: $p < 0.001$) was observed for the Shannon index of the invertebrate dataset. The COI eDNA dataset was offshore dominated by one species, *Clytia hemisphaerica* (70.2 %).

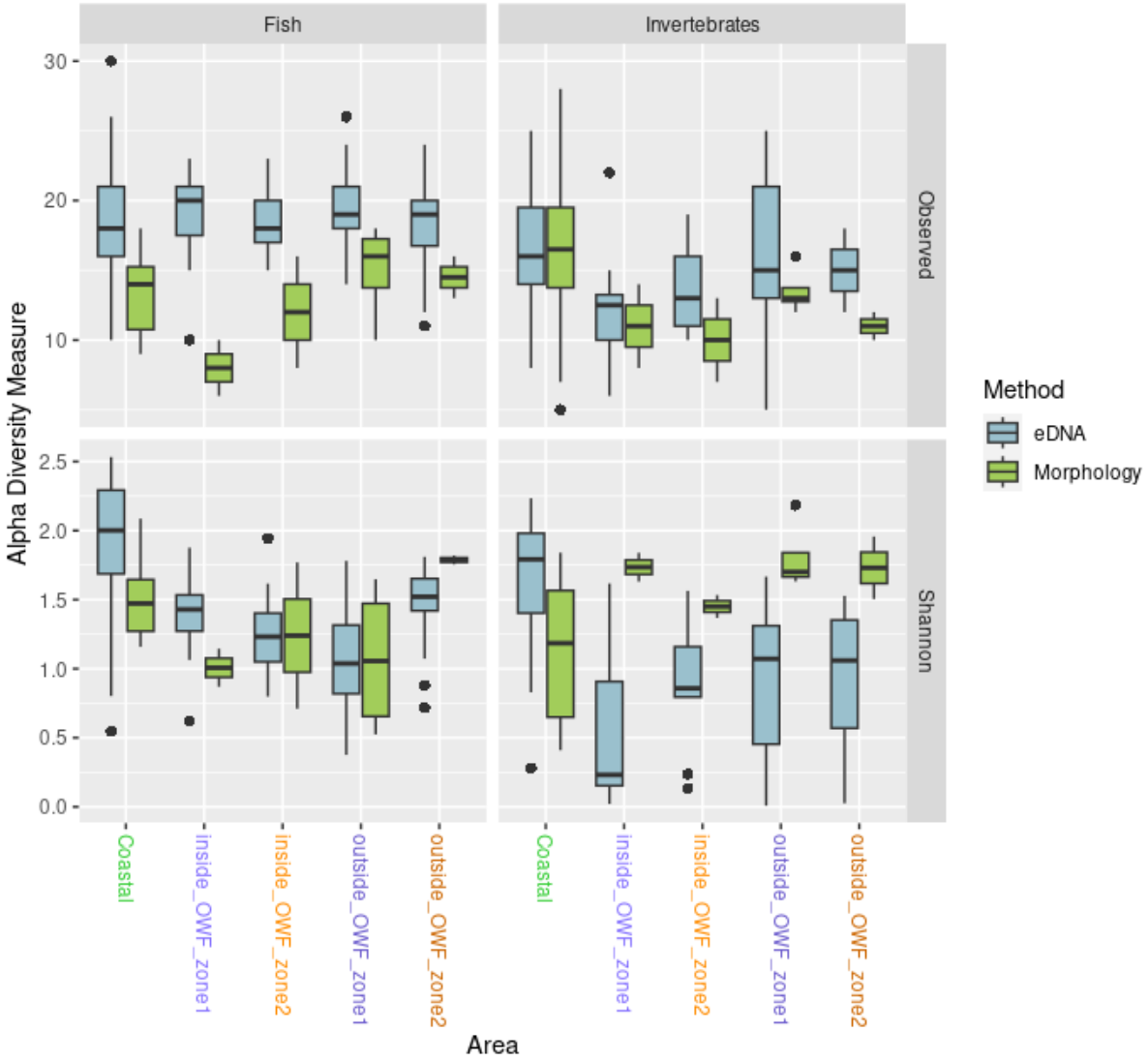


Figure 6: Observed number of species and Shannon diversity index in each of the five zones. The mean observed number of species (top) and mean Shannon diversity index (bottom) per zone was calculated based on the morphology and rarefied eDNA data for fish (right) and invertebrate (left) species. The box is drawn from the first quartile to the third quartile and the black line represents the median. The whiskers represent the values larger and smaller than 1.5 times the third and first quartile. The black dots are the outliers that lie beyond the range of the whiskers.

The fish and invertebrate eDNA community structure was significantly different (one way PERMANOVA: $p < 0.0001$, betadisper: $p = 0.01$ for 12S and $p = 0.09$ for COI) between the coast, the offshore zone around C-power (zone 1) and the offshore zone around Belwind (zone 2). The pairwise comparison revealed a significant difference between the fish and invertebrate community structures in the coastal waters and the offshore zone around C-Power (zone 1) (pairwise comparison: $p < 0.001$ for 12S and COI), and the coastal waters and offshore zone around Belwind (zone 2) (pairwise comparison: $p < 0.001$ for 12S and COI). In addition, the fish

and invertebrate communities were also significantly different between zone 1 and zone 2 (pairwise comparison: $p < 0.0001$ for 12S and $p = 0.0061$ for COI). Largely similar results were obtained when analyzing the morphology based data. The fish and invertebrate community structure was significantly different between the three zones (one way PERMANOVA: $p < 0.0001$, betadisper: $p = 0.15$ for 12S and $p = 0.02$ for COI). The fish, as well as the invertebrate community structure differed between the coast and both offshore zones (pairwise comparison: $p < 0.01$ for fish and invertebrates). In addition, the invertebrate community structure was also significantly different between zone 1 and zone 2 (pairwise comparison: $p = 0.004$ for invertebrates), while this was not the case for fish (pairwise comparison: $p = 0.579$ for fish). The nMDS-plots corroborate the PERMANOVA results in that the coastal samples are well separated from offshore samples for both metabarcoding assays and morphological datasets, while samples collected in zone 1 and zone 2 did not have a clear segregation (Figure 7).

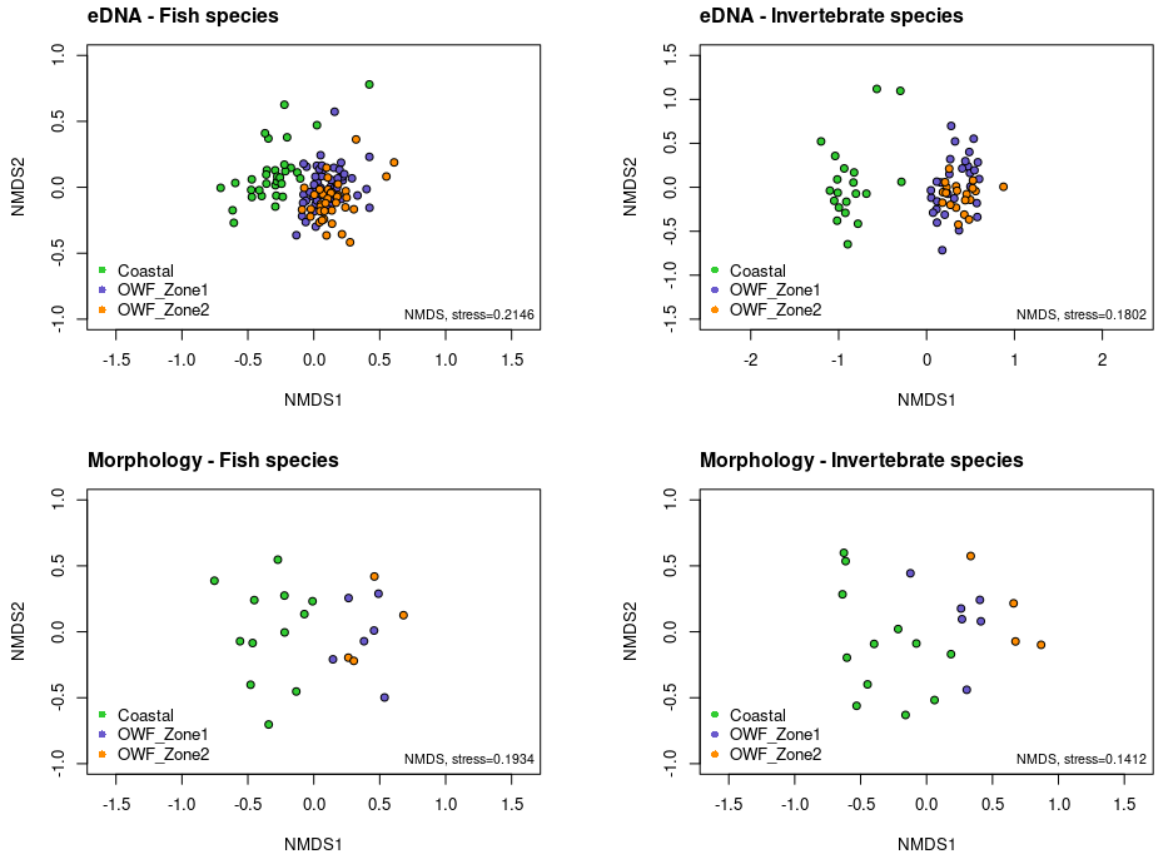


Figure 7: NMDS plot of the Bray-Curtis dissimilarities from eDNA metabarcoding and morphology based data of the three sampling zones. The Bray-Curtis dissimilarities of eDNA data were calculated based on fish (12S) and invertebrate (COI) species detected in the samples after taxonomic assignment, by DADA2 and BLAST, of the rarefied and log transformed data from samples that had 10.000 (12S) and 25.000 (COI) reads. The Bray-Curtis dissimilarities, based on the fish and invertebrate species detected in the beam trawls, were calculated with the log transformed dataset. (NMDS with k = 2)

Finally, we zoom in at the offshore windfarm scale. The fish community data showed multivariate homogeneity within the groups (betadisper: 12S $p = 0.47$). For the invertebrate community data we found different levels of dispersion by zone and impact combined (betadisper: $p = 0.01$), but not for the zone (betadisper: $p = 0.07$) and impact (betadisper: $p = 0.44$) alone. However, the ordination supports a clustering pattern based on the zone and impact suggesting that dispersion effects are not driving the patterns. For the fish and invertebrate community structures a significant interaction effect was detected between the zone (transition

versus offshore) and the impact (inside versus outside the OWF) (two-way PERMANOVA: 12S p = 0.0001, COI p = 0.0409). The nMDS ordination of eDNA data showed weak clustering of samples by impact (inside versus outside the OWF) within each zone (Figure 8).When investigating more into detail between the pairs of data, the fish and invertebrate communities found inside both OWFs were significantly different from each other and from the communities found at both reference areas outside the OWFs (pairwise comparison: $p < 0.05$ for 12S and COI). In addition, a significant difference was found between the fish communities from both reference areas outside the OWFs (pairwise comparison: p = 0.0002 for 12S) but not for the invertebrates (p = 0.4101 for COI). Hence the differences in the species community compositions are dependent on the zone in which the samples were taken and the impact. Similar to the 12S eDNA data, a significant interaction effect between the zone and impact was detected with fish data from the morphology based beam trawl survey (two-way PERMANOVA: p = 0.03, betadisper: p = 0.73) but not for the epibenthic invertebrates (two-way PERMANOVA: p = 0.34, betadisper: p = 0.89). The offshore invertebrate species communities were significantly different between both zones (p = 0.005), but not between the impacts. The nMDS ordination based on the invertebrate species caught in the beam trawl net showed a clear clustering between both zones (Figure 8).

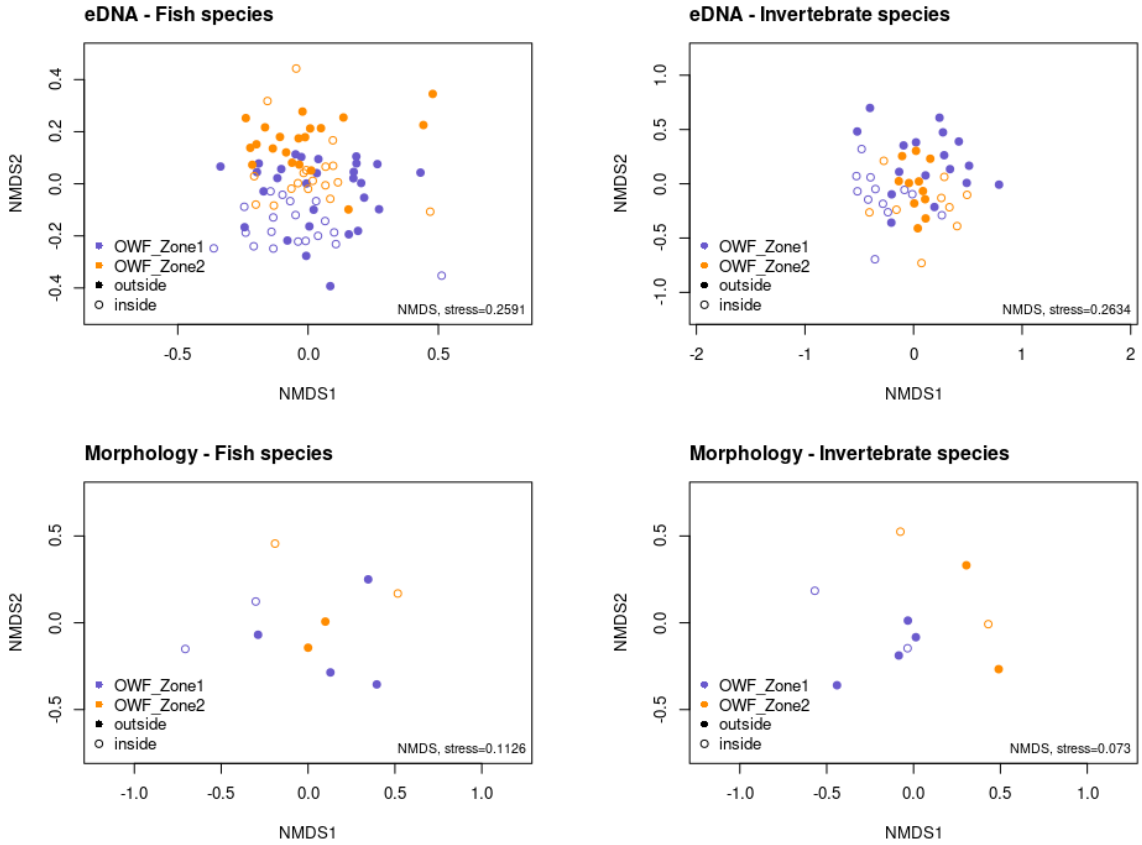


Figure 8: NMDS plot of the Bray-Curtis dissimilarities from eDNA metabarcoding and morphology based data of the offshore wind farm and its reference areas (OWF). The Bray-Curtis dissimilarities of eDNA data were calculated based on fish (12S) and invertebrate (COI) species detected in the samples after taxonomic assignment, by DADA2 and BLAST, of the rarefied and log transformed data from samples that had 10.000 (12S) and 25.000 (COI) reads. The Bray-Curtis dissimilarities, based on the fish and invertebrate species detected in the beam trawls, were calculated with the log transformed dataset. (NMDS with k = 2)

In conclusion, these results show that 12S eDNA metabarcoding is a valuable alternative to morphology based analyses of beam trawl samples for monitoring fish diversity related to

offshore wind farms. eDNA analyses detected considerably more fish species than morphological identification of beam trawl catches and only very few fishes (7/42) were missed with eDNA. Patterns in species numbers, Shannon diversity and community structure found with eDNA were not significantly different from those observed with morphological analyses. In contrast, invertebrate monitoring with COI eDNA metabarcoding detected only a small fraction of the invertebrate species found in the catches (14/52) but picked up annelids, Cnidaria and zooplankton which are not targeted with beam trawls. As such, COI eDNA metabarcoding does not accurately represent beam trawl information, but can in combination with the 12S marker increase the width of species diversity detected with eDNA. In addition, the majority of the COI reads could not be assigned to any taxon resulting in a large part of the sequencing data not being used for further analyses. Yet, both morphology and COI eDNA revealed the same pattern of the number of species across the different zones, differentiated invertebrate communities from the coastal and two offshore zones and both methods revealed no significant impact of the offshore wind farms. COI eDNA metabarcoding cannot replace the invertebrate information obtained with beam trawling, but combining eDNA and beam trawl data provides a more complete picture of the invertebrate taxa present in the different areas.

The time and cost calculation for the offshore wind farm case study showed that the species list was obtained much faster with the morphological analyses of beam trawl samples than with eDNA metabarcoding (38 hours vs 188 hours, Table 1). Even considering only hands-on time, beam trawling was still 42.5% faster than eDNA metabarcoding. Beam trawling was also 53% cheaper than eDNA based analyses with two marker genes.

Table 1: Summary of the time and costs associated with collecting 10 fishtracks and 50 water samples and their subsequent processing to assess fish and invertebrate diversity in 10 locations.

Processing of 10 fishtracks	Sampling	Sorting catch	counting, measuring, weighing fish	counting, measuring, weighing epibenthos	epibenthos sorting lab	QC + database input	Total	Total time (h)	Total cost (euro)
Time (h)	5	8,33	10,22	10,72	0,87	3,08	38,22	38,22	3539,50
Personel - Junior scientist	90 euro/hour						3439,5		
Consumable cost	100						100		

Processing of 50 water samples	Sampling	filtering water	eDNA extraction	12S library prep (PCR in triplicate, QC, pooling, clean-up, QC)	COI library prep (PCR in triplicate, QC, pooling, clean-up, QC)	Bioinformatic processing (12S + COI)	Total	Total time (h)	Total cost (euro)
Time (h)	2,57	3,62	42,00	14,25	14,25	17,08	93,77	187,53	
hand on time (h)	2,57	3,62	8,00	8,25	8,25	2,58	33,27	66,53	7470,00
Personel - Lab technician	75 euro/hour						2495		
Consumable cost (euro)	50	750	175	750	750		2475		
Sequencing cost (euro)	2500						2500		
Beamtrawl vs eDNA								42,56	52,62

b. Case study on NIS monitoring in harbors

Metabarcoding of the eDNA plankton communities in the harbor of Ostend generated a total of ±2.8 million reads after filtering. Sequences that were found in the negative control samples were representatives of crop plant species, human DNA, and taxa that were unique to the negative controls. Subsequently, these sequences were removed from the dataset. After selecting for taxa with an AphiaID, 243 ASVs remained across the three locations which represented 94 % of the reads (Figure 9). Of those, 142 unique ASVs (53 % of the reads) were resolved to species level but only 74 ASVs (34 % of the reads) were identified with a match greater than 98 %. Of all the reads, 1.2 % did find a match in the WoRMS database. The highest number of unique species (> 98 % match) was found at Ponton Overzet (n = 25), followed by

Marina Mercator (n = 12) and Vuurtorendok (n = 11). The remaining 26 species (> 98 % match) were found at more than one of the three locations.

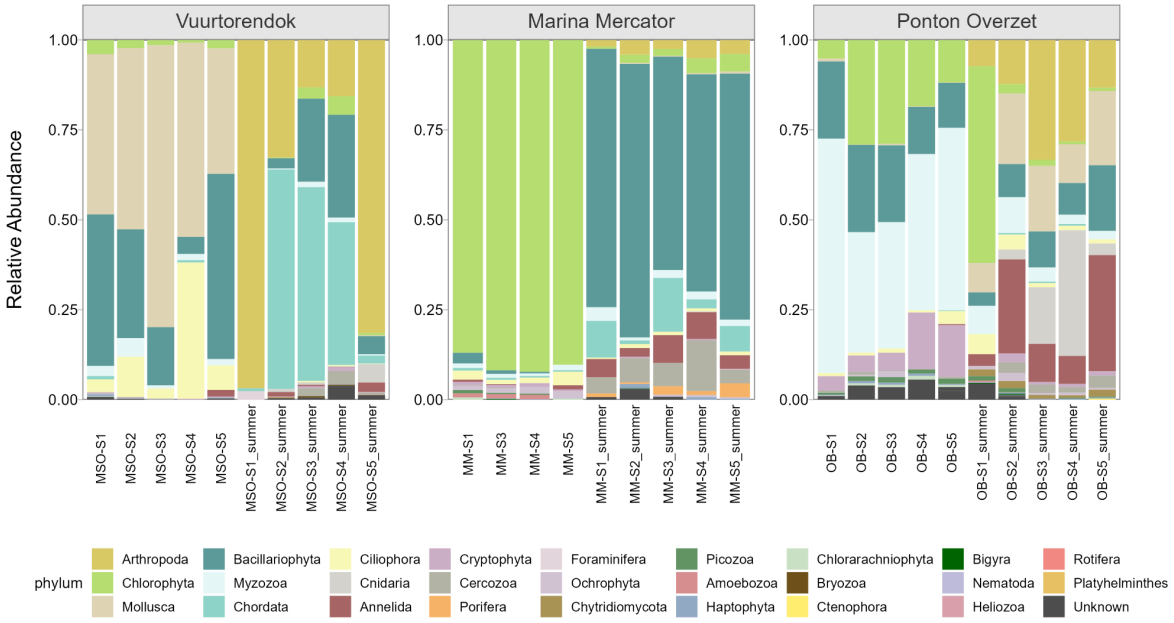


Figure 9: Relative abundance of reads assigned to phylum level for each of the three locations in the Ostend harbor. For each location, samples collected during late spring are shown first on the x-axes followed by samples from late summer.

From the rarefaction analyses, it was evident that we had enough sequencing depth to properly reflect the species diversity within each sample (Figure 10 a). However, further examination showed that the sampling design can be improved. When pooling the five replicates per sampling event, the rarefaction curves did not reach a plateau illustrating that the plankton community diversity at each sampling site for each time point, based on the 18S region (Figure 10 b), was not completely represented. Overall, the analyses indicated that additional sampling sites and sampling dates would likely have increased the observed number of species (Figure 10 c,d). The latter was also noticed in terms of NIS detection (Figure 10 e,f).

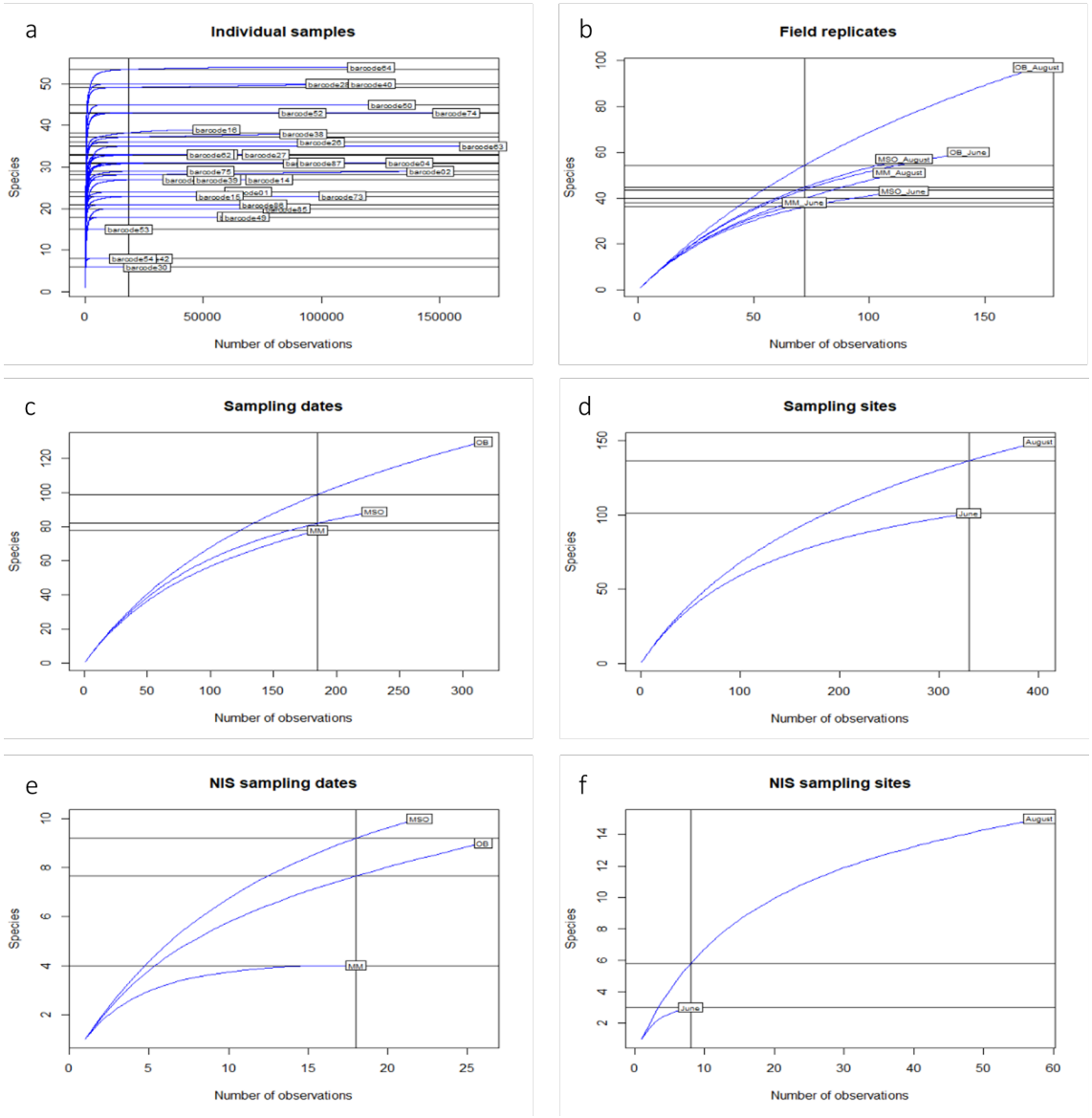
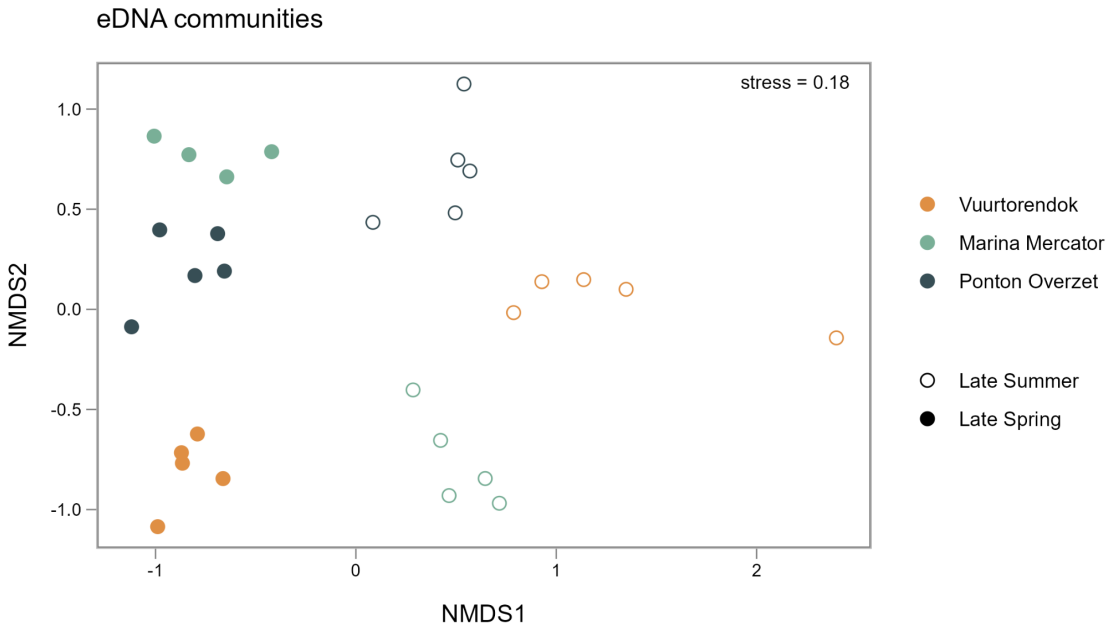


Figure 10. Rarefaction analyses. (a) Sequencing depth per sample. (b) Five field-replicates pooled per sampling date and site. (c) Two sampling dates per sampling site. (d) Three sampling sites per sampling date. (e) Sampling dates per site only including NIS. (f) Sampling sites per date only including NIS. The vertical line indicates the lowest number of reads in plot a and the lowest number of observations for plot b-f.

The species composition was significantly different between seasons, the three sampling locations, and for the interaction factor between those groups (two-way Permanova: all p-values < 0.001, betadisper: all p-values > 0.1). The clustering of these groups is also nicely illustrated in the nMDS plot, with a greater distinction between the locations during summer (Figure 11).

Figure 11: Non-metric multidimensional scaling (nMDS) of eDNA communities in the Ostend harbor. Community variability was illustrated using Bray-Curtis dissimilarities and 2-dimensions (k = 2). Colors represent the three locations (Vuurtorendok, Marina Mercator



and Ponton Overzet) that were sampled. Symbols represent the seasons when the sampling took place: summer (open circle) and spring (closed circle).

In total, eDNA analysis identified 15 taxa as non-indigenous species: eight at Ponton Overzet, four at Marina Mercator and ten at Vuurtorendok (Table 2). These species were all found during summer, only three species were also detected in late spring (Table 2). All identified NIS had a minimum identity match of 99%. Out of the 15 NIS species, two are not yet recorded in the 'non-indigenous species of the Belgian part of the North Sea' list (Verleye et al., 2020), i.e. *Oithona davisae* and *Pseudopolydora paucibranchiata*, but have already been registered from neighboring parts of the North Sea (Radashevsky et al., 2020).

Table 2: Non-indigenous species found in the Ostend harbor.

Phylum	Class	Order	Family	Scientific Name	eDNA Id.	eDNA % match	Morph Id.
Annelida	Polychaeta	Sabellida	Serpulidae	*Ficopomatus enigmaticus*	MM, MM-S, VD-S, PO, PO-S	99.7	MM-S, VD-S, PO-S
Annelida	Polychaeta	Spionida	Spionidae	*Pseudopolydora paucibranchiata*	VD-S	99.2	/
Arthropoda	Copepoda	Calanoida	Acartiidae	*Acartia tonsa*	PO-S, VD-S	99.5	/
			Pseudodiaptomidae	*Pseudodiaptomus marinus*	PO-S	98.8	/
		Cyclopoida	Oithonidae	*Oithona davisae*	MM-S, VD-S, PO-S	99.1	/
	Malacostraca	Amphipoda	Caprellidae	*Caprella mutica*	/	/	MM-S, VD-S. PO-S

	Thecostraca	Balanomorpha	Balanidae	*Amphibalanus sp.*	VD-S	99.3	/
				Austrominius modestus	PO-S	99.7	VD-S, PO-S
Chordata	Ascidiaceae	Phlebobranchia	Cionidae	*Ciona intestinalis*	MM-S, VD-S	99.2	MM-S, VD-S, PO-S
		Stolidobranchia	Molgulidae	*Molgula manhattensis/ M. retortiformis*	MM-S, VD-S	99.5	MM-S, VD-S
			Styelidae	*Botrylloides violaceus*	VD-S	99.4	VD-S, PO-S
				Styela plicata/ S. clava	VD, VD-S	99.1	/
Mollusca	Bivalvia	Myida	Dreissenidae	*Mytilopsis leucophaeata*	OB-S	99.4	/
		Ostreida	Ostreidae	*Magallana gigas*	VD-S, PO, PO-S	99.5	/
		Verenida	Mactridae	*Rangia cuneata*	PO-S	99.2	/

VD: Vuurtorendok, PO: Ponton Overzet, MM: Marina Mercator, and S: summer samples

Morphological identification of the benthic communities on the settlement plates, during summer, identified 36 taxa across the three locations of which 20 were resolved to species level (Figure 12). Six species were classified as NIS, of which five were also detected by eDNA analysis: *Austrominius modestus*, *Botrylloides violaceus, Ciona intestinalis, Ficopomatus enigmaticus*, and *Molgula manhattensis*/*M. retortiformis*. One NIS was identified through morphological examination alone, *Caprella mutica* (Table 2), a species that has been previously recorded as NIS in the Belgian part of the North Sea (Verleye et al., 2020).

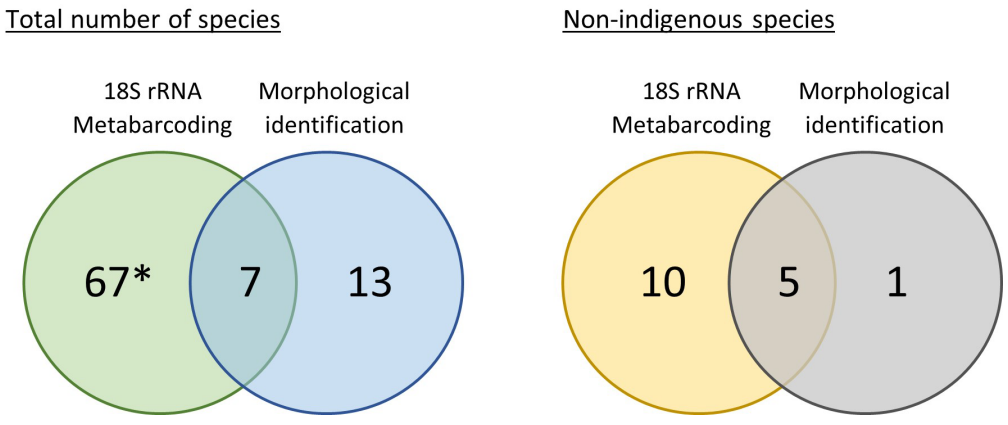


Figure 12: Overview of the observed number of species and non-indigenous species detected with eDNA (18S rRNA) and through morphological examination of the settlement plates. The Venn diagram depicts only summer samples collected from the three locations in the Ostend harbor (Ponton Overzet, Marina Mercator and Vuurtorendok). (*) at > 98 % match and only taxa with taxonomic annotation at species level were kept.

In the harbor of Ostend, more than double the amount of NIS were detected with eDNA monitoring compared to morphological examination of settlement plate samples (15 vs. 6). Three NIS detected only with eDNA are holoplankton (all three are copepods). All others are benthic species, although many of them have planktonic life stages. We did not investigate whether the eDNA came from planktonic larvae or has been shed by the benthic adults. Most NIS observed in this study reproduce during the summer months. We found five times more NIS in eDNA samples taken in late summer compared to late spring (15 vs 3), which is in line with planktonic origin of eDNA. This observation highlights that the usefulness of eDNA sampling to detect NIS may depend on the season in which the samples are taken. Rarefaction analysis found that the number of sampling replicates, sites and sampling dates was insufficient and increased sampling effort would likely have yielded additional NIS. If the goal of the monitoring is to obtain a complete list of NIS, both eDNA-based and traditional monitoring methods should be used in combination.

Sequencing of bulk metabarcoding and eDNA samples from the harbor in Rostock generated ±3.8 million reads, of which ±1.1 million reads were removed as chimeras from the dataset, resulting in roughly 2.6 million reads of interest. After bioinformatic processing, 745 ASVs were assigned to 10 phyla (Figure 13) and 182 unique macrofaunal species with an AphiaID, which represented 57 % of the reads. Only 0.24 % of the reads could not be assigned at all. The eDNA method detected many more species compared to the two size fraction processed (100 µm and 335 µm) with the bulk DNA method (126 species vs 76 or 57 species, respectively). Morphological identification conducted by the IfAÖ yielded 32 zooplanktonic taxa, of which only 13 could be identified to species level.

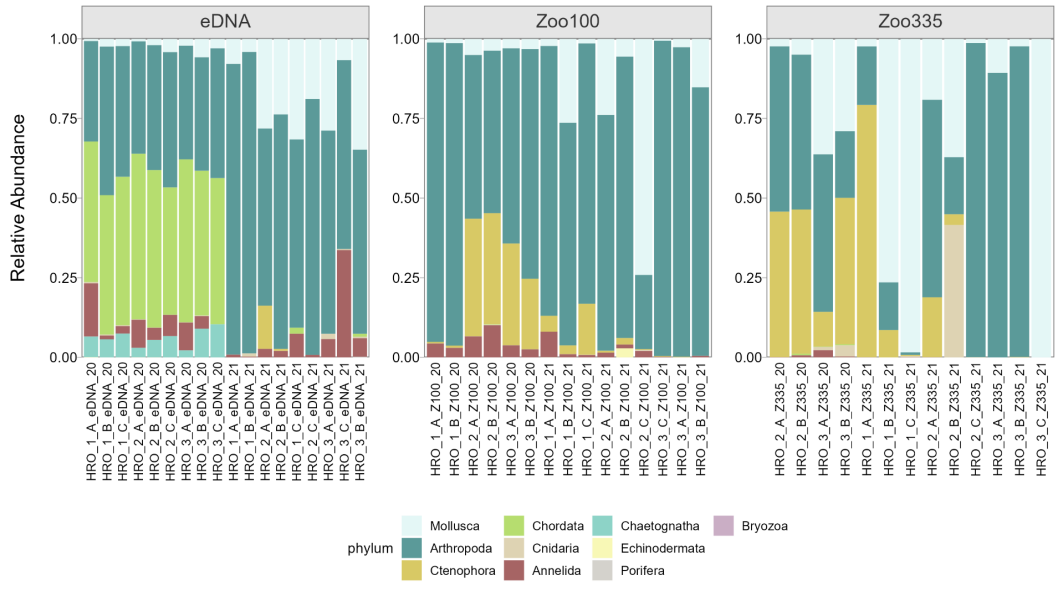


Figure 13: Relative abundance of macrofaunal reads assigned to phylum level for each of the three methods in the Rostock harbor. For each method, samples collected in 2020 are shown first on the x-axes per location (HRO-1, HRO-2 & HRO-3), followed by samples taken in 2021. The three methods are eDNA analysis, bulk DNA metabarcoding of zooplankton for the 100 µm size fraction (Z100) and the 335 µm size fraction (Z335).

Rarefaction analysis on the filtered dataset (*pident* 71.7 %, *qcovs* 75 %) showed that the sequencing depth was sufficient to cover the full species diversity for most samples except for the eDNA samples from 2021 (Figure 14).

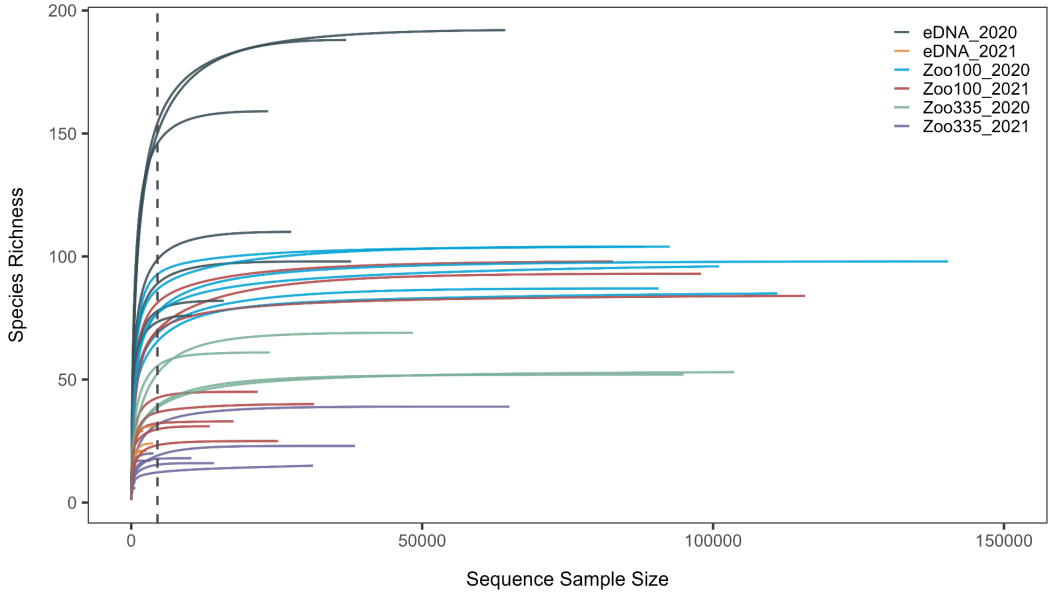


Figure 14: Rarefaction analysis showing the sequencing depth per sample. Each curve represents one sample. Colors show the different methods per year of sampling: eDNA analysis and bulk DNA metabarcoding of zooplankton for the 100 µm size fraction (Z100) and the 335 µm size fraction (Z335). The dashed line shows the sequencing depth (4.500) to which the dataset was rarified for downstream analyses.

Community composition was found to be significantly different between the three sampling methods ($p < 0.001$; Figure 15). However, the assumption of homogeneity of variances of each group (the three methods) was violated (Betadisper: $p = 0.02$). This indicates that the difference between the three methods could also be the result of a dispersion effect (variances around the mean of each group). Although, by looking at the nMDS ordination (Figure 15) there is a clear separation between the community clusters of bulk DNA metabarcoding and eDNA analysis.

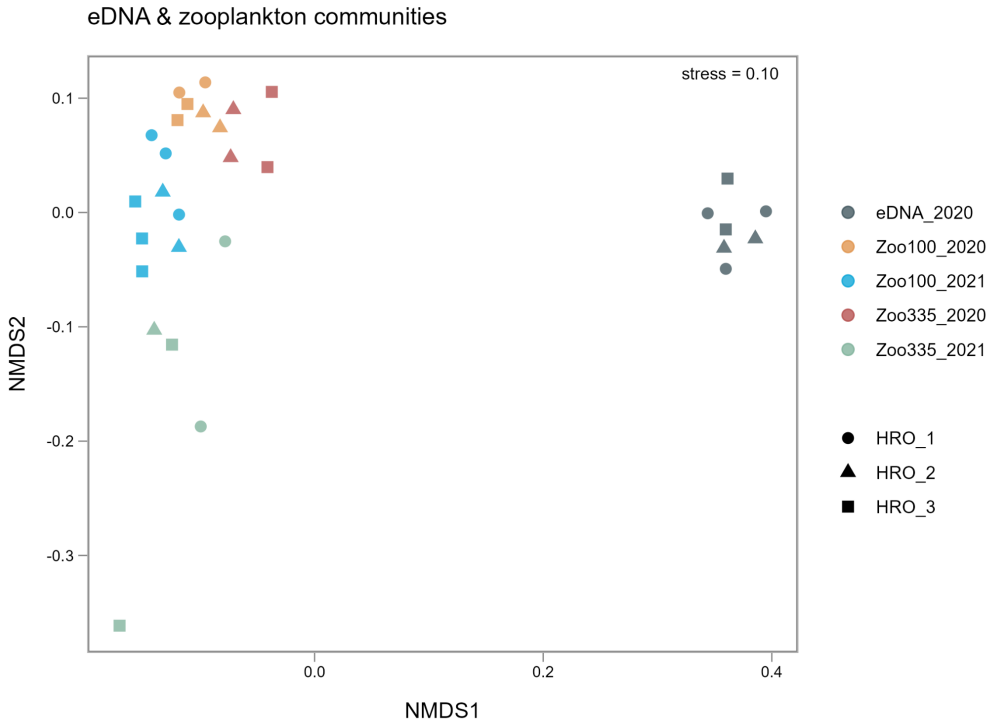


Figure 15: Non-metric multidimensional scaling (nMDS) of eDNA and zooplankton communities in the Rostock harbor. Community variability is illustrated using Bray-Curtis dissimilarities and 2-dimensions (k = 2). Colors represent the three methods per year of sampling: eDNA analysis and bulk DNA metabarcoding of zooplankton for the 100 µm size fraction (Z100) and the 335 µm size fraction (Z335). Symbols represent the three locations; circles: the marina in the city harbor (HRO-1), triangles: the international port (HRO-2), and squares: the marina in Warnemünde (HRO-3). eDNA samples of 2021 were not included because of the low sequencing depth (<4500 sequences).

In total, 20 NIS were detected in the Rostock harbor pilot (Figure 16). All of these species were found by bulk DNA metabarcoding of zooplankton samples, nine within eDNA samples and only two by morphological identification. Only taxa known as NIS that could be unambiguously identified morphologically and/or genetically to species level were counted. The exclusively genetically assigned taxa of the zooplankton samples include *Amathia gracilis, Bugulina stolonifera, Haminella solitaria, Hemigrapsus takanoi, Melita nitida, Mulinia lateralis, Mytilopsis leucophaeata, Neogobius melanostomus, Rangia cuneata* and *Sinelobus vanhaareni*. No exclusively morphologically identified zooplanktonic NIS were found, nor were any exclusively identified by eDNA analysis. Although eDNA analysis was able to identify more NIS than morphological identification, bulk DNA metabarcoding was clearly the most effective method for NIS detection at species level.

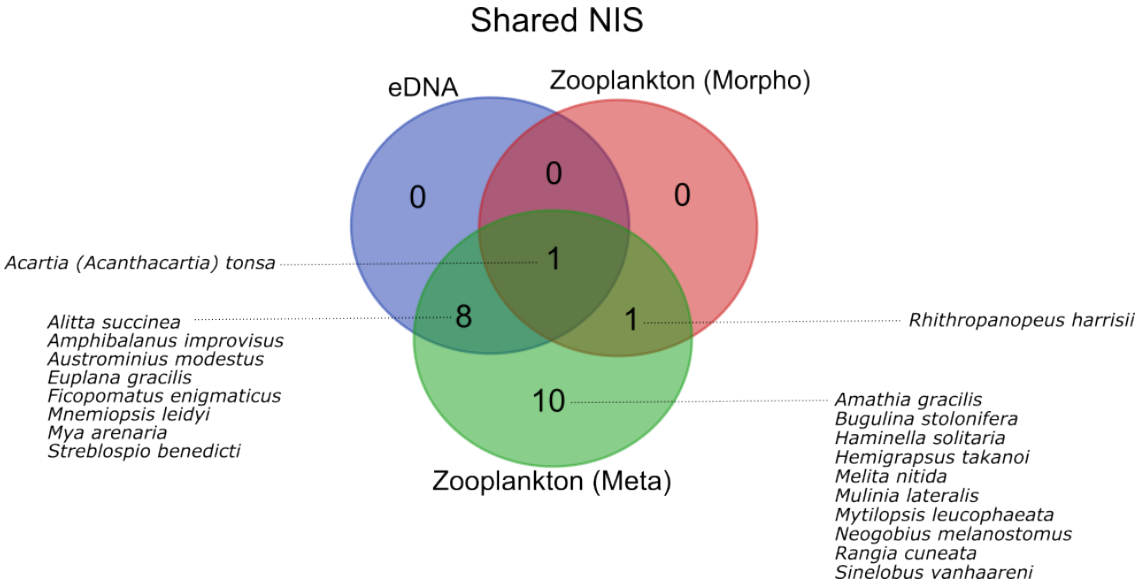


Figure 16: VENN diagram showing the total number of NIS shared between the sampling methods. Morpho: morphological identification conducted by IfAÖ; Meta: bulk DNA metabarcoding using two mesh size fractions (100 µm and 335 µm).

c. Case study on the use of long read metabarcoding in artificial reefs in the North Sea

In silico comparison of the designed primer shows that the 2 kb fragment overlaps with the DNA region obtained by most of the primer pairs that are currently used to identify fish (Zhang et al., 2020). Only Mifish and Teleo2 both anneal to a region upstream the forward primer binding site. The benefit of using a longer amplified fragment is demonstrated by the *Ammodytes* species. *Ammodytes* is known to have low genetic variability and comparison of *A. marinus*, *A. tobianus*, *A. hexapterus*, *A. personatus*, *A. dubius* and *Hyperoplus lanceolatus* shows a species specific pattern of only very few single nucleotide polymorphisms (SNPs). Only a limited amount of SNPs (0 to 5) can be found between species with commonly used primers (Fig 16c). In contrast, the designed 2kb amplicon shows that the SNPs enable unambiguous separation of *Ammodytes sp*. into different species (Figure 16b).

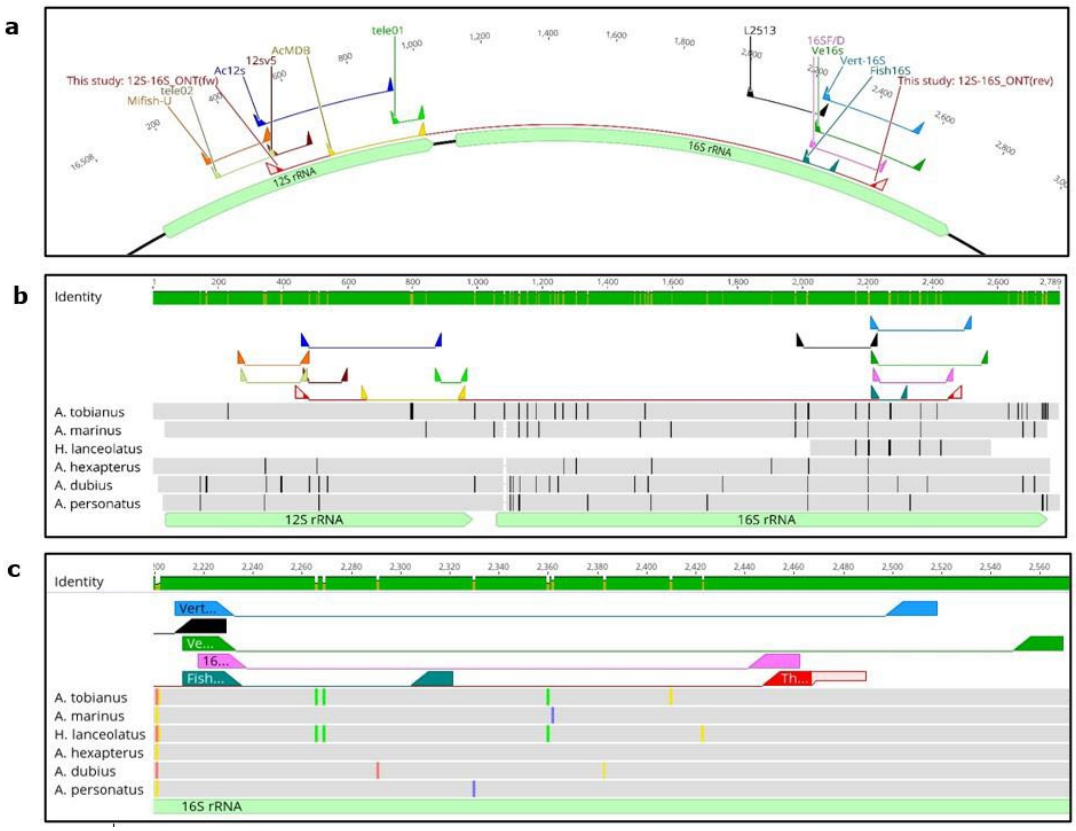


Figure 16: Comparison of primer binding sites of this study and other commonly used fish eDNA primers according to Zang et al., (2020). Where **a)** shows the specific primer binding locations (triangles) of the 2 kb primer pair used in this study (red) and other studies indicated in different colors. Red fill in the triangle indicates the binding site, empty part of the triangle indicates the ONT extension. **b)** Alignment of the 12S and 16S rRNA gene of different *Ammodytes* species that occur in both the North Sea and Atlantic Ocean. Black stripes represent SNPs. **c)** Comparison of the region with the highest variability covered by the highest amount of primers. Four colors represent different nucleotides (green = T, yellow = G, red = A and blue = C).

For the ocean aquarium we had a list of all species present in the collection, including the number of specimens, and for the larger fish, also weight and size. From 92.2 % of the raw reads used to build consensus sequences, a taxonomic identification at species or genus level was obtained using the automated Centrifuge pipeline. The remaining 7.8 % could, based on manual BLASTn results, be assigned to *Plectorinchus* and *Glaucostegus* suggesting that the automated pipeline needs further fine tuning to be able to assign all reads. Of all taxonomically identified sequences, 17 genera could be identified out of the 23 different genera present in the Ocean aquarium (Figure 17a). These 23 genera represented 31 species of which 18 species could be identified with this primer set (Figure 17b). In other words, the reads obtained with this long read metabarcoding method identified 58 % of the species and 74 % of genera present in the ocean aquarium. Of the five genera not detected with eDNA, only two genera *(Ctenochaetus* and *Siganus*) had a full mitogenome available. Additionally, of the 13 undetected species with eDNA, five species (*Acanthurus tennentii, Myripristis jacobus, Plectorhinchus obscurus* and *Chrysiptera parasema* had no reference sequence available. Instead of the unidentified species *Abudefduf sexfasciatus and Myripristis murdjan)*, the species *Abudefduf vaigiensis* and *Myripristis berndti* were identified using eDNA. For these two identified species the full mitogenome was present whereas for the two morphologically identified species, only the 16S

reference was available. This indicates that the reference database is a limiting factor in this study.

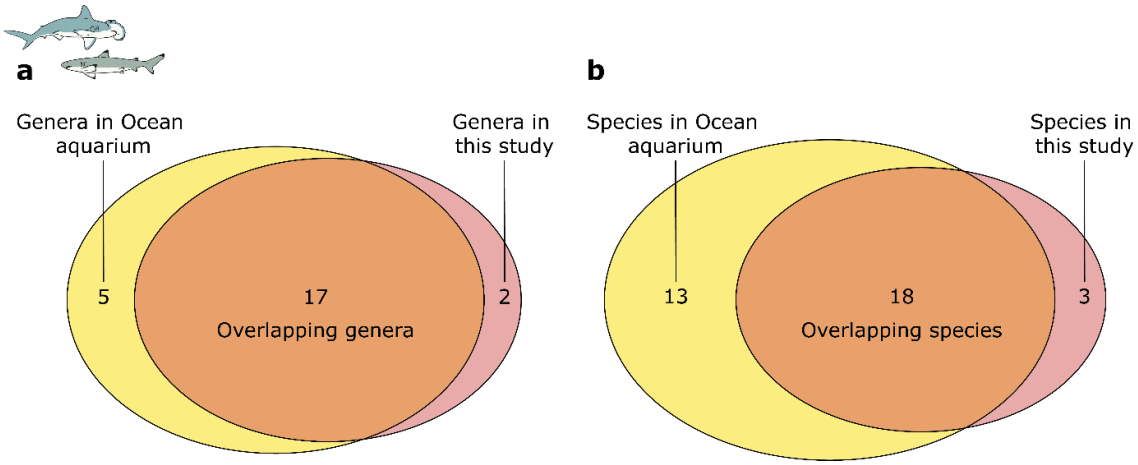


Figure 17: Euler-diagrams of the amount of overlapping and unique taxa in the samples. Yellow represents the total amount of taxa present in the Ocean aquarium but not found in this data. Orange represents the taxa found in both aquarium and eDNA method, while red represents the taxa found in the eDNA samples that are not present in the aquarium (false positives). Diagram **a)** represents the overlapping genera while **b)** represents the overlapping species.

The newly developed method was subsequently applied on field samples collected around shipwrecks in the North Sea. A total of 262.732 reads were analyzed that derived from nine samples that resembled three replicates from three different shipwrecks (average read number per sample: 11.655 ± 3.581). In total, 79 consensus sequences were identified which could be assigned to 21 different species. The nMDS ordination shows distinct clustering between the replicates of the ship wrecks that was significant (p= 0.002) (Figure 18b). The species composition of shipwreck 1 mainly consisted of gray gurnard *Eutrigla gurnardus*, dab *Limanda limanda* and bullrout *Myoxocephalus scorpius* (Fig 18a). Unique for this shipwreck is the occurrence of sardine *Sardina pilchardus*. From shipwreck 2, *E. gurnardus* is commonly found, and also sand goby *Pomatoschistus minutus* and plaice *Pleuronectes platessa* are commonly found in the sequence data. Turbot *Scophthalmus maximus, P. minutus* and common dragonet *Callionymus lyra* were unique for this ship wreck. In shipwreck 3, *A. marinus* was the most abundant species. A moderate amount of harbour porpoise *Phocoena phocoena* sequences was found in the samples from shipwreck 1 showing that this method is also able to detect mammal species.

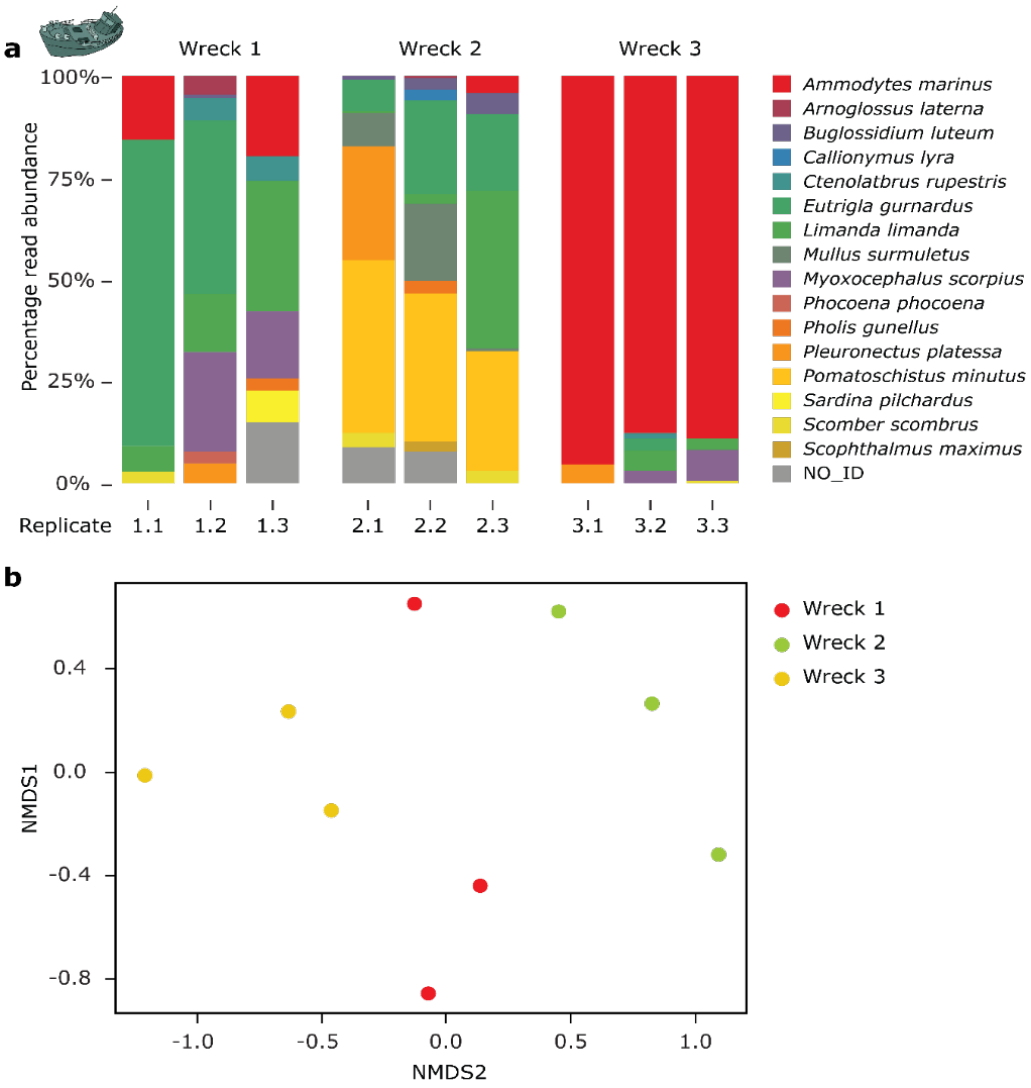


Figure 18: a) Proportional read abundance plot of marine vertebrate species associated with the different shipwrecks. Bars indicate from which ship wreck the water samples were taken, including replicate number. Colors indicate the relative read count of the identified fish species. b) NMDS plot of the beta-diversity between the different shipwrecks.

The methodology was further applied on field samples collected in a wind farm, a flat oyster restoration site and in a control sandy site in the Dutch part of the North Sea. A total of 913 119 reads could be analyzed from 18 samples from three locations. From this, 105 consensus sequences could be assigned to 20 different species. The number of species and Shannon evenness values showed similar patterns between the two time points across locations as no significant interaction was found between location and date for the two measurements (shannon: p = 0.18, Richness: p= 0.63, Fig 19). There was a significant effect of sampling date with a higher number of species detected on July 2nd compared to July 17th (shannon: p=0.016, Richness: p = 0.008). No significant effect of location was found (Shannon: p=0.54, Richness p= 0.50).

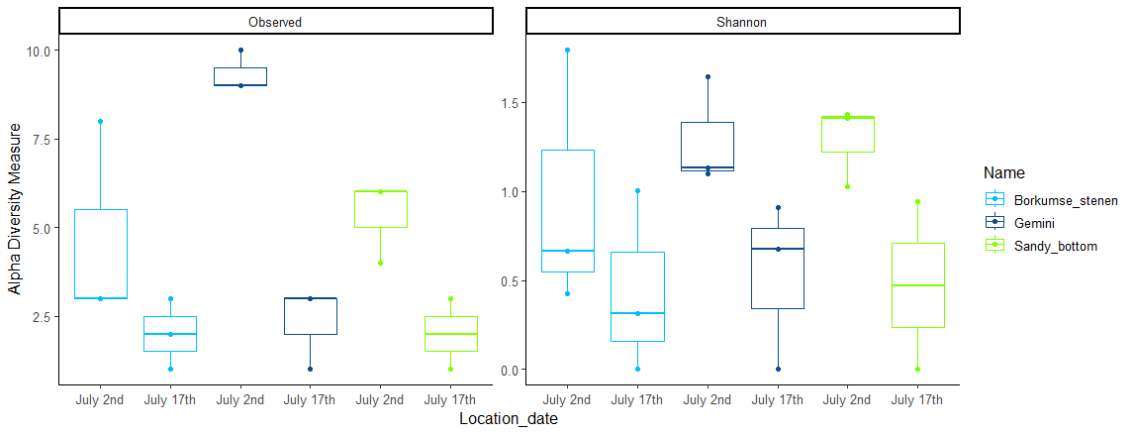


Figure 19: Boxplot of the alpha diversity using Observed and Shannon (Species Richness and Shannon diversity). Indices between locations and sampling time points. Color indicates location and X-axis label indicates the sampling timepoints.

In Gemini, sequences from *A. marinus, E. gurnardus* and *S. scombrus* were most common (Figure 20). Samples from the sandy bottom control site have *P. minutus* and *Sprattus sprattus* as the most abundant species while for Borkum Reef Grounds this was *E. gurnardus*, *P. minutus* and *C. lyra*. Hardly any difference was found between dates in gemini wind park and sequences of the thornback ray *Raja clavata* and *P. phocoena* were found. *S. pilchardus, solenette Buglossidium luteum, L. limanda, Lozano's goby Pomatoschistus lozanoi, P. phocoena* and horse mackerel *Trachurus trachurus* were not found in the July 17th samples of Gemini. At the Sandy bottom on the other hand, the relative read abundance of *E. gurnardus* was higher on the 17th of July, while *S. pilchardus*, eel *Anguilla anguilla*, *B. luteum* and *L. limanda* were undetectable in the 17th of July samples. The Borkum Reef Grounds also showed variation between dates. *S. scombrus*, *S. pilchardus* and long-spined sea scorpion *Taurulus bubalis* were only detected the 2nd of July, while species *E. gurnardus, Arnoglossus laterna*, striped red mullet *Mullus surmuletus* and whiting *Merlangius merlangus* only on the 17th of July.

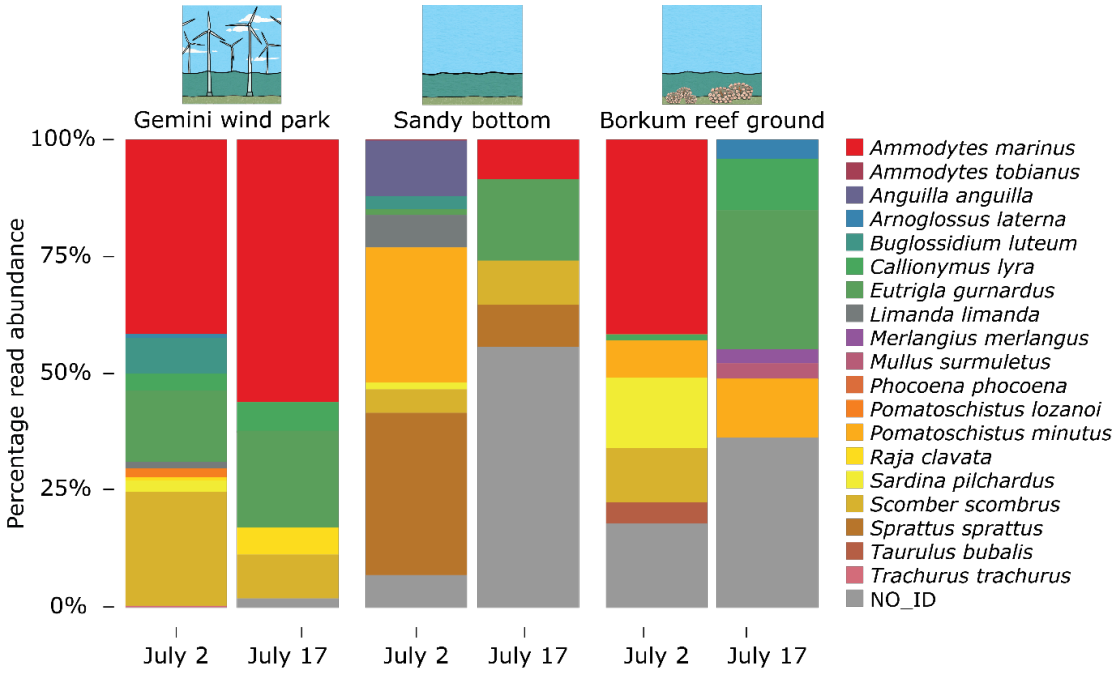


Figure 20: Percentage read abundance plot of the marine vertebrate species in the locations Gemini wind park, at sandy bottom halfway Gemini wind park and the Borkum reef grounds and at the Borkum reef grounds. Bars indicate at which date the water samples were taken. Colours indicate the proportion read abundance of the identified fish species.

The differences in species composition in each of the three locations is further illustrated in the nMDS plot (Figure 21). Samples collected at the first sampling point (2nd July) clearly showed a clustering according to the three locations, while samples collected at the second sampling point (17th July) showed much higher variability but still separated the Borkum and Gemini locations from each other. One-way PERMANOVA showed a significant effect of location for 2nd July (p= 0.024) but not for the 17th of July (p=0.101). Pairwise posthoc tests between the three locations for the first time point were all non-significant. When pooling the samples from the two timepoints for each location, a significant difference between Gemini and the Sandy bottom (p= 0.03) and Gemini and the Borkum reef ground (p= 0.045) but not between the sandy bottom and and the Borkum reef grounds (p=0.537) was found.

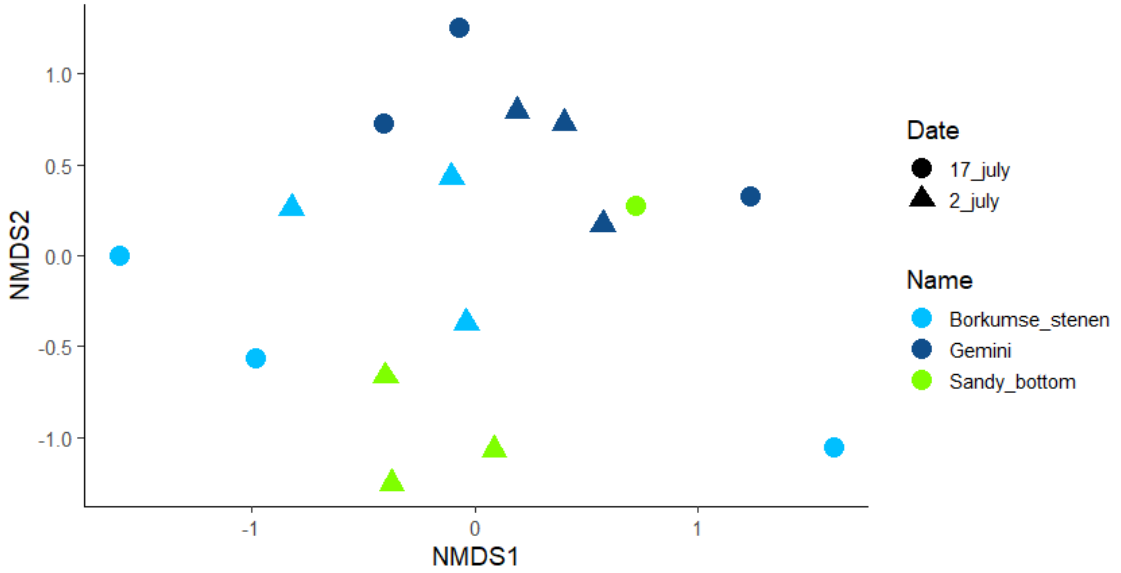


Figure 21. NMDS plot showing the three eDNA sampling locations north of Schiermonnikoog and sampled at two dates.

In conclusion, we found that the designed primer pair for fish (e)DNA using ONT nanopore sequencing can increase the resolution at species level. It is also able to detect a broad range of fish species known to be present in that area. However, there are limitations as the database is not complete for both 12S and 16S rRNA sequences for all species (about 30% missing) and not all species are detected, despite its availability in databases. Nevertheless, the field test near the ship wrecks and the reefs illustrate that distinct species compositions can be detected in different habitats. In addition, although not significant, bottom dwellers such as dab, sand goby, plaice and turbot seem to be more often picked up in samples from sandy bottoms in comparison to more pelagic fish in reef structures. To fully understand the potential of long read metabarcoding, it should be compared to a baseline metabarcoding approach (i.e. MiFish). As eDNA quickly degrades in marine systems, the collected eDNA quickly becomes too short to be picked up by these primers, which possibly accounts for not picking up all species that are present in the aquarium. Therefore, another interesting possibility is to use long read metabarcoding to study short term spatial temporal shift based on the detection of DNA fragments of different lengths.

5. General conclusions and recommendations

The case studies that used both eDNA- and morphology based analyses (the offshore wind farm study and the NIS pilots) showed that eDNA detected many more species than morphology and only a small number of fishes (7/69) or NIS (1/15) were detected only with morphology. In other words, the morphology based analyses had little added value in terms of fish species detected for impact assessments or for NIS detection. On the other hand, some of the additional species detected with eDNA may be 'false positives'. Therefore, species lists generated by eDNA should be treated with care and with local ecological knowledge in mind, as matching them against known species occurrence maps could be a way to validate the species list. Nevertheless, with respect to monitoring of fish, the eDNA datasets were able to recover similar differences in fish communities between areas in the offshore wind farm case study compared to the morphological analyses. Although no morphological data was available for the reef case study also here eDNA was able to detect spatial differences in fish communities in the North Sea. However, we showed that eDNA metabarcoding is not faster or cheaper than beam trawling. Still, the GEANS consortium considers 12S eDNA metabarcoding a valuable alternative to traditional beam trawl analyses for monitoring fish diversity because it captures more species, detects known spatial patterns in the North Sea and because of its non-destructive nature. In view of the potential automatisation of eDNA sample collection, **we foresee that 12S eDNA metabarcoding will soon gain advantage over traditional beam trawl fish monitoring**. The next step now is to create standardized protocols across member states to initiate implementation in (legislative) monitoring programmes.

With respect to invertebrates, COI eDNA metabarcoding detected only a small fraction of the invertebrate species found in the beam trawls and mostly picked up zooplankton species or planktonic stages of macrobenthos. This is in agreement with the results of the NIS study in the harbor of Rostock, which used the same COI primer set to detect zooplankton. **Based on our pilot, COI eDNA metabarcoding is not recommended for epibenthos or macrobenthos monitoring**. Twice as many NIS were detected when using bulkDNA samples of zooplankton instead of eDNA from the water column. Therefore, **we recommend bulk DNA metabarcoding of plankton net samples as the best methodology for monitoring NIS in harbors if resources are limited**. Bulk DNA metabarcoding was also cheaper and much faster than morphology based identification of NIS (Articles, Reports & Publications | GEANS).

Our case studies further demonstrated that **small and easy to handle in house sequencing devices such as the Minion sequencer from Oxford Nanopore offer an alternative to high throughput sequencing platforms** (such as Illumina) that are operated by big sequencing facilities **to detect species with eDNA**. The newly designed primer pair for fish (e)DNA using ONT sequencing can increase the resolution at species level and detected a broad range of fish species known to be present in the area. However, the method is currently limited as the reference database for this longer fragment is far from complete (about 30% of species missing) and not all species are detected despite their availability in the database. Proper ground truthing of the method with morphology based analyses and with established sequencing methods is needed to confirm its full potential for fish monitoring. We have also shown that ONT sequencing can be used to detect NIS.

Our case studies clearly showcase the potential of eDNA for various monitoring questions, but also illustrate that the field of eDNA is still in its infancy compared for example to bulk DNA

metabarcoding. Further research and international collaboration between different institutes is needed to understand systematic biases in species detections in the North Sea, to develop standardized protocols and guidelines for field and laboratory steps and to develop optimal monitoring schemes that take advantage of this non-destructive and high throughput methodology.

6. SWOT analysis

The eDNA case studies within the GEANS project were specifically chosen to explore the potential of eDNA for various monitoring questions in the marine environment. Based on our experience during the execution of the three case studies, the GEANS consortium has defined strengths and weaknesses related to eDNA-based identification of fish, invertebrates and NIS, and identified external opportunities and threats when using eDNA-based methods for monitoring in the North Sea (Table 3).

With respect to the strengths of eDNA-based monitoring, the case studies on monitoring offshore wind farms and NIS showed that eDNA-based analyses pick up more species than traditional morphology based approaches, because taxonomic groups or life stages (e.g. larvae) with few diagnostic morphological features have unique DNA barcodes allowing their identification with eDNA. Moreover a wider spectrum of species is detected by eDNA from the water column e.g. both pelagic and benthic species are picked up with eDNA, while beam trawls are mainly targeting benthic fish species (in the case of fish monitoring) and settlement plates are targeting colonising fauna (in the case of NIS monitoring). An example is the large amount of reads assigned to *Sardinus pilchardus* in the Belgian OWF case. This species had not been detected with beam trawl data over the last 20 years of monitoring in the area. The eDNA results are in agreement with survey data on fish and eggs in the wider North Sea region, and thus illustrate that eDNA is able to detect species that are either missed with beam trawling or are present in a different life stage.

In terms of weaknesses, we have shown that time and costs associated with eDNA analyses are not faster or cheaper than beam trawl data. In addition, false negative results due to identical barcode sequences can be resolved using the longer 2 kb DNA fragment as shown by the *Ammodytes* and *Hyperoplus* detection in the case study on reef monitoring. However, switching to this larger fragment did not detect all species, because of the incompleteness of the reference databases for these longer sequences. For this reason, the GEANS project has provided a protocol to mitogenomic sequences of key North Sea species (Protocols | GEANS).

In terms of opportunities, eDNA detects species that experience recent range shifts, which has been shown for the NIS species *Oithona davisae* in the Ostend harbour pilot. Furthermore, eDNA offers great opportunities for automated sampling in the near future and could allow for more standardized procedures across countries when proper data management is in place. With respect to threats, the GEANS project has taught us that local ecological knowledge is detrimental in the correct interpretation of eDNA-based results.

Table 3. List of internal strengths and weaknesses and external opportunities and threats of eDNA-based monitoring

Strengths	Weaknesses
• High throughput of samples is possible • Non-destructive method that only requires water sampling • Species from hard to identify taxonomic groups can be identified to species-level with eDNA. • Wider spectrum of taxonomic groups are detected with water column eDNA compared to traditional monitoring which targets specific groups depending on the gear used • Planktonic larval stages can be identified through sequencing eDNA, thereby making early detection of NIS possible • Typically more species detected compared to traditional NIS or bottom trawl monitoring methods • No specific taxonomic expertise is needed for eDNA analyses	• eDNA processing is not necessarily faster e.g. in the case of beam trawl fish monitoring • Shedding of DNA can differ substantially over time and between species, and species with hard exoskeleton are difficult to detect with eDNA • Quantitative information from eDNA is unreliable at this moment • No information on life stage and sex when using eDNA • Reliable, complete, curated reference libraries are needed • False negatives: not all taxa are amplified or some have identical barcode sequences at species level, so that they cannot be differentiated from each other • Non-target groups such as plankton or bacteria may take up a considerable part of the sequencing data, resulting in too few data available for the target species • Sometimes amplification of eDNA samples fails, and there is no material to fall back to. The reason for failure is not always obvious. • Expert consultation needed to exclude false positives • Lack of negative control samples may lead to false positive records of species
Opportunities	**Threats**
• Many projects in parallel possible • eDNA sampling can be automated in space and/or timethrough fixed automated samplers without the need for additional shipping time • Sample collection, lab processing and data analyses can be easily standardised across institutes and thus allows for standardisation across countries • Early detection of NIS or other species that are shifting their distribution ranges • More rapid assessments • Workflows can be generalized across different type of projects or samples	• Expectation management: important to be honest in what eDNA-based monitoring can and cannot do • • Lack of proper archiving of metadata hampers comparability across datasets from different institutes • New monitoring method for which systemic biases still need to be investigated better (i.e. systematically misses specific species) • Misinterpretation of eDNA data if insufficient local ecological knowledge is available

7. References

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

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data [WWW Document]. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 5.23.22).

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online*, 1–15. https://doi.org/10.1002/9781118445112.STAT07841

Bakker, J., Wangensteen, O. S., Chapman, D. D., Boussarie, G., Buddo, D., Guttridge, T. L., . . . Mariani, S. (2017). Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. Sci Rep, 7. doi:10.1038/s41598-017-17150-2

Baloğlu, B., & Chen, Z. (2021). BBaloglu/ASHURE: First release of ASHURE. In Methods in Ecology and Evolution (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4450611

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. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Chen, H., 2022. VennDiagram: Generate High-Resolution Venn and Euler Plots.
Chamberlain, S. A., & Szöcs, E. (2013). Taxize: Taxonomic search and retrieval in R. F1000Research, 2. doi: 10.12688/f1000research.2-191.v2

De Backer, A., Buyse, J., Hostens, K., 2020. A decade of soft sediment epibenthos and fish monitoring at the Belgian offshore wind farm area. Mem. Mar. Environ.

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., Van Broeckhoven, C., NanoPack: visualizing and processing long-read sequencing data, *Bioinformatics*, Volume 34, Issue 15, August 2018, Pages 2666–2669, https://doi.org/10.1093/bioinformatics/bty149

Degraer, S., Carey, D. A., Coolen, J. W. P., Hutchison, Z. L., Kerckhof, F., Rumes, B. & Vanaverbeke, J. (2020). Offshore wind farm artificial reefs affect ecosystem structure and functioning: a synthesis. Oceanography, 33: 48–57., https://doi.org/10.5670/oceanog.2020.405

De Mesel, I., Kerckhof, F., Norro, A., Rumes, B. & Degraer, S. (2015). Succession and seasonal dynamics of the epifauna community on offshore wind farm foundations and their role as stepping stones for non-indigenous species. *Hydrobiologia*, 756: 37–50.

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

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. doi:10.1111/1755-0998.12138

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PloS One, 9*(2), e87624. https://doi.org/10.1371/journal.pone.0087624

Hayward, P. J., & Ryland, J. S. (Eds.). (2017). *Handbook of the Marine Fauna of North-West Europe* (2nd ed.). Oxford University Press.

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647–1649. doi: 10.1093/bioinformatics/bts199

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Research, 26(12), 1721–1729. doi: 10.1101/gr.210641.116

Kim, D. K., Park, K., Jo, H., & Kwak, I. S. (2019). Comparison of Water Sampling between Environmental DNA Metabarcoding and Conventional Microscopic Identification: A Case Study in Gwangyang Bay, South Korea. Applied Sciences-Basel, 9(16). doi:10.3390/app9163272

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., 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. Front. Zool. 10, 34. https://doi.org/10.1186/1742-9994-10-34

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Martinez Arbizu, P., 2020. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4.

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. https://doi.org/10.1371/JOURNAL.PONE.0061217

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., Iwasaki, W., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. R. Soc. Open Sci. 2, 150088. https://doi.org/10.1098/rsos.150088

Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment, 6(9), 485-492. https://doi.org/10.1890/070064

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Maintainer, H. W. (2020). *Package "vegan" Title Community Ecology Package Version 2.5-7*.

OSPAR. (2019). Joint Harmonised Procedure for the Contracting Parties of HELCOM and OSPAR on the granting of exemptions under International Convention for the Control and Management of Ships' Ballast Water and Sediments, Regulation A-4. OSPAR Commission.

Radashevsky, V.I., Malyar, V.V., Pankova, V.V. *et al.* Disentangling invasions in the sea: molecular analysis of a global polychaete species complex (Annelida: Spionidae: *Pseudopolydora paucibranchiata*). *Biol Invasions* **22**, 3621–3644 (2020). https://doi.org/10.1007/s10530-020-02346-x

Ricotta, C., Acosta, A.T.R., Bacaro, G., Carboni, M., Chiarucci, A., Rocchini, D., Pavoine, S., 2019. Rarefaction of beta diversity. Ecol. Indic. 107, 105606. https://doi.org/10.1016/j.ecolind.2019.105606

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

Sardain, A., Sardain, E. & Leung, B. Global forecasts of shipping traffic and biological invasions to 2050. *Nat Sustain* **2**, 274–282 (2019). https://doi.org/10.1038/s41893-019-0245-y

Sigsgaard, E. E., Torquato, F., Froslev, T. G., Moore, A. B. M., Sorensen, J. M., Range, P., Thomsen, P. F. (2020). Using vertebrate environmental DNA from seawater in biomonitoring of marine habitats. Conservation Biology, 34(3), 697-710. doi:10.1111/cobi.13437

Staerh, P.A.U., Dahl, K., Buur, H., Göke, C., Sapkota, R., Winding, A., Panova, M., Obst, M. & Sundberg, P. (2022). Environmental DNA Monitoring of Biodiversity Hotspots in Danish Marine Waters. Frontiers in Marine Science 8, 1–15, https://doi.org/10.3389/fmars.2021.800474

Thomsen, P. F., Kielgast, J., Iversen, L. L., Moller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. PLoS ONE, 7(8). doi:10.1371/journal.pone.0041732

van Bleijswijk, J. D. L., Engelmann, J. C., Klunder, L., Witte, H. J., Witte, J. I., & van der Veer, H. W. (2019). Analysis of a coastal North Sea fish community: Comparison of aquatic environmental DNA concentrations to fish catches. Environmental DNA. doi:https://doi.org/10.1002/edn3.67

Verleye, T. J., De Raedemaecker, F., Vandepitte, L., Fockedey, N., Lescrauwaet, A.-K., De Pooter, D., & Mees, J. (Eds.). (2020). *Niet-inheemse soorten in het Belgisch deel van de Noordzee en aanpalende estuaria anno 2020*. Flanders Marine Institute. ISBN 9789464206005

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wheeler, Q. D., Raven, P. H., & Wilson, E. O. (2004). Taxonomy: Impediment or expedient? Science, 303(5656), 285-285. doi:10.1126/science.303.5656.285

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