# High resolution species detection: accurate long read eDNA metabarcoding of North Sea fish using Oxford Nanopore sequencing

Karlijn Doorenspleet*[1], Lara Jansen*[1], Saskia Oosterbroek*[1], Oscar Bos[2], Pauline Kamermans[1,3], Max Janse[4], Erik Wurz[1], Albertinka Murk[1] & Reindert Nijland[1]†

1) Marine Animal Ecology group, Wageningen University and Research, P.O. Box 338, 6700 AH, Wageningen, The Netherlands

2) Wageningen Marine Research, PO Box 57, 1780 AB Den Helder, The Netherlands

3) Wageningen Marine Research, PO Box 77, 4400 AB Yerseke, The Netherlands

4) Royal Burgers' Zoo, Antoon van Hooffplein 1, 6816 SH Arnhem, The Netherlands

*) Contributed equally

†) Corresponding author: reindert.nijland@wur.nl

## Abstract

To monitor the effect of nature restoration projects in North Sea ecosystems, accurate and intensive biodiversity assessments are vital. DNA based techniques and especially environmental DNA (eDNA) metabarcoding from seawater is becoming a powerful monitoring tool. However, current approaches are based on genetic target regions of <500 nucleotides, which offer limited taxonomic resolution. This study aims to develop and validate a long read nanopore sequencing method for eDNA that enables improved identification of fish species.

We designed a universal primer pair targeting a 2kb region covering the 12S and 16S rRNA genes of fish mitochondria. eDNA was amplified and sequenced using the Oxford Nanopore MiniON. Sequence data was processed using the new pipeline Decona, and accurate consensus identities of above 99.9% were retrieved. The primer set efficiency was tested with eDNA from a 3.000.000 L zoo aquarium with 31 species of bony fish and elasmobranchs. Over 55% of the species present were identified on species level and over 75% on genus level. Next, our long read eDNA metabarcoding approach was applied to North Sea eDNA field samples collected at ship wreck sites, the Gemini Offshore Wind Farm, the Borkum Reef Grounds and

a bare sand bottom. Here, location specific fish and vertebrate communities were obtained. Incomplete reference databases still form a major bottleneck in further developing high resolution long read metabarcoding. Yet, the method has great potential for rapid and accurate fish species monitoring in marine field studies.

Keywords

eDNA, North Sea, metabarcoding, nanopore, long read, marine fish

# Introduction

North Sea fish populations are sensitive to disturbances such as fisheries, nutrient run offs and increasing sea water temperatures (Andersen et al., 2020; Capuzzo et al., 2018; Hofstede, Hiddink, & Rijnsdorp, 2010; O'Brien, Dafforn, Chariton, Johnston, & Mayer-Pinto, 2019). Combined management strategies such as reduced fishing (Couce, Schratzberger, & Engelhard, 2020), designation of marine protected areas (MPA), and placing artificial hard substrates such as off-shore wind parks are suggested to facilitate rehabilitation of the North Sea ecosystem (Claudet, 2018; Degraer et al., 2020; Didderen, Lengkeek, Bergsma, & Dongen, 2019; Kamermans, van Duren, & Kleissen, 2018). To understand how North Sea fish population dynamics are affected by these strategies, development and validation of methods that map fish population diversity and density is crucial. Conventional marine fish biomonitoring practices largely rely on destructive methods that involve netting and trapping (Daan, Gislason, Pope, & Rice, 2005; Reiss et al., 2010). These methods are costly, time-consuming and require expert taxonomic visual identification skills (Mateos-Rivera et al., 2020; Teletchea, 2009). In addition, conventional methods have limited sampling efficiencies and may be disruptive to the environment (Eggleton, Depestele, Kenny, Bolam, & Garcia, 2018). Thus, it is crucial to develop precise and non-invasive biomonitoring solutions that are time and cost efficient (Goodwin et al., 2017).

Environmental DNA (eDNA) based fish species identification has gained substantial attention in the last decade, as it can detect the presence of fish species based on a small amount of DNA present in e.g. 1 liter of seawater. It has been shown to be highly sensitive for non-indigenous species detection (Ficetola, Miaud, Pompanon, & Taberlet, 2008) and identification of spawning and migration patterns (Thalinger, Wolf, Traugott, & Wanzenböck, 2019). Short read eDNA metabarcoding has become an increasingly popular tool to perform fish community assessment for identification of ecological relevant fish species

from an array of ecosystems (Deiner et al., 2017; Miya et al., 2015; Ruppert, Kline, & Rahman, 2019; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012; Thomsen et al., 2012). Also in the North Sea, eDNA metabarcoding results from a sampling effort close to fykes showed to be comparable to the fyke catches themselves. (Bleijswijk et al., 2020).

The standardization of eDNA metabarcoding as monitoring strategy is still under development. Species-related differences occur in e.g. degree of skin cell shedding, ambient dependent (seasonal) degradation rates and unknown dilution factors depending on currents make quantification of the results challenging (Beng & Corlett, 2020; Lacoursière-Roussel, Côté, Leclerc, & Bernatchez, 2016; Sassoubre, Yamahara, Gardner, Block, & Boehm, 2016; Seymour et al., 2018). The sample preparation and metabarcoding technique and workflow will determine the quality of the results and thus the species detection quality and possible biases (Beng & Corlett, 2020; van der Loos & Nijland, 2020). Important steps in the protocol include decisions about methods of sampling and DNA extraction (Bessey et al., 2020; Hunter, Ferrante, Meigs-Friend, & Ulmer, 2019), primer and PCR settings (Doi et al., 2019; Sard et al., 2019; Zhang, Zhao, & Yao, 2020), sequencing technology (Egeter et al., 2020; Singer, Fahner, Barnes, McCarthy, & Hajibabaei, 2019; Truelove, Andruszkiewicz, & Block, 2019), post sequencing data handling (Santos, van Aerle, Barrientos, & Martinez-Urtaza, 2020) and reference databases used (Hestetun et al., 2020; McGee, Robinson, & Hajibabaei, 2019).

Especially choice of primer and targeted DNA region are crucial for successful fish detection with eDNA (Beng & Corlett, 2020). Several universal fish primers are available that mostly target regions of the mitochondrial genome as there is a high copy number of this genome per cell (Schon, 2000). The most used primers target different short regions from 100 to 500 nucleotides of the 12S rRNA (Miya et al., 2015; Riaz et al., 2011; Taberlet, Bonin, Zinger, & Coissac, 2018) 16S rRNA (DiBattista et al., 2017; Evans et al., 2016), cytochrome B (Thomsen et al., 2012) and COI (Balasingham, Walter, Mandrak, & Heath, 2018) gene. Although primers targeting short 12S regions are the most commonly used and considered as a standard (Shu, Ludwig, & Peng, 2020), longer target amplicon size yield higher taxonomic resolution (Zhang et al., 2020). Also the use of multiple primer sets are suggested to increase taxonomic resolution (Evans et al., 2016; Miya et al., 2015; Zhang et al., 2020) and has been shown to increase species level detectability in lakes (Sard et al., 2019). A method to relatively increase the presence of less abundant taxa is applying oligo blocking primers to reduce the amount of unwanted DNA in samples such as human DNA or DNA from very abundant taxa (Liles et al. 2003; Vestheim & Jarman, 2008 ). In addition, using longer reads and multiple marker sets can enhance the necessary species resolution and specificity

although it remains unclear how the use of a blocking primer may affect marine vertebrate species detection.

Long read sequence analysis with commonly used Illumina platforms is not possible due to its abilities to sequence with high accuracy but with a read length maximum of 500bp (Tan, Opitz, Schlapbach, & Rehrauer, 2019). Fortunately, third generation sequencing as available from Oxford Nanopore Technologies (ONT) and Pacific Biosciences enables the generation of ultra-long sequences (Bleidorn, 2016). This can be used for eDNA studies that are based on primers targeting longer regions, covering several mitochondrial marker genes. Such approach showed successful in microbial metabarcoding studies, which considerably increased taxonomic resolution up to species level (Johnson et al., 2019; Shin et al., 2016). Historically, the main limitation of nanopore sequencing was the relatively large error rate of 5 to 10% (Jain et al., 2015). This error rate can be overcome with bioinformatics tools to generate reliable consensus sequences and thus increase sequence accuracy (Baloğlu et al., 2020; Carradec et al., 2020; Egeter et al., 2020). To our knowledge, a bioinformatics pipeline is not yet available to generate such consensus sequences from raw sequence data in multiplexed metabarcoding experiments. Such a pipeline would greatly facilitate the development of long read metabarcoding in marine molecular biomonitoring.

This study aims to investigate the possibilities of long read metabarcoding of eDNA with a newly designed fish specific primer pair targeting a 2kb amplicon containing both the 12S and 16S mitochondrial rRNA genes, using Oxford nanopore MinION sequencing. To increase species identification accuracy, we developed a sequence data processing pipeline. The identification resolution of the primers were tested *in silico* on several genetically similar species from the genus *Ammodytes* (sandeels). eDNA sampled from a large marine zoo aquarium was used to analyse the accuracy of species identification together with technical evaluations such as the use of replicates to better understand the species detection possibilities of the amplicon. Further, we assessed the effectiveness of a human-specific oligo blocking primer to reduce the influence of human DNA in the sample on target species detection. Finally, we applied our newly developed long read metabarcoding approach to field samples collected at different locations and timepoints in the North Sea.

# Materials and Methods

## Sample collection

Samples were collected on three separate sampling sites with varying sampling approaches; the Ocean aquarium of Royal Burgers' Zoo, Arnhem, the Netherlands; around North Sea shipwrecks; and in the North Sea on a transect from Gemini Wind Park to Borkum Reef Grounds (figure 1).

For eDNA samples from the Ocean aquarium in Burgers' Zoo, 2L water samples were collected just under the water surface using a 1L plastic container pre-sterilized with bleach. After filtering, filters were cut in quarters to serve as filter replicates (figure 1a). Demineralized tap water from Burgers' Zoo was also filtered to serve as negative control. The Ocean aquarium has a volume of 3000 $m^3$ artificial seawater and represents the edge of a tropical coral reef with only fish, in total 31 species varying in size from ~ 5 cm to 2.5 m.–The water temperature is kept at 24.5-26.0°C, the salinity at 33.0 ‰ and the pH at 8.00-8.10. In the North Sea, samples were collected around three different ship wrecks while SCUBA diving: 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 (figure 1b, dx.doi.org/10.17504/protocols.io.6yfhftn). Tap water was used as negative control. From Borkum Reef Grounds/Gemini Wind Park, samples were collected from inside Gemini Wind Park (54.0109, 6.0781), halfway 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. Three 1L replicates were collected at each location by sampling seawater using 2.5 L Niskin bottles at 0.5-1m above the seafloor on 2, 17 and 31 July 2020 (figure 1c). Demi water was used as negative control.

All samples were immediately filtered using Thermo Scientific Nalgene Rapid-Flow sterile disposable Filter Units CN (Cellulose nitrate) with a pore size of 0.8µm. Filters were then individually placed in 2mL screwcap Eppendorf tubes. The tubes were prefilled with 400µL tissue lysis buffer (ATL, Qiagen, USA) when sampled in Burgers' Zoo and 400µL Zymo DNA/RNA shield (Zymo, USA) preservative when sampled at the North Sea. Samples were immediately stored at -20°C for a maximum of one month before further processing.

## Primer design

Primer design is 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. Blocking primers were designed for *Homo sapiens*. The complete 12S and 16S region was extracted from all available mitochondrial fish genomes. A consensus sequence for each species was constructed when multiple genomes were available from the same species using default settings of the MAFFT alignment tool (v7.450, Katoh & Standley, 2013) incorporated in Geneious. Consensus sequences of all species were aligned and primers were designed manually by locating regions with low genetic variation between target species. This resulted in a long read universal fish primer pair (table 1) targeting a 2kb fragment from 450bp downstream the 12S rRNA gene in forward direction and 300bp upstream the 16S rRNA gene in reverse direction (figure 2a). The 5' ends of the primers were extended with an ONT tag to allow for direct PCR based sample barcoding in downstream library preparation.

The additional *Homo sapiens* specific oligo blocking primer is designed based on the alignment with the universal forward primer and extended with a human specific sequence (table 1). The 3' end of the blocking primer was extended with a phosphoramidite C3 spacer to chemically prevent polymerase activity. Primers were synthesized by Integrated DNA Technologies and purified by standard desalting only (Integrated DNA technologies Inc., USA).

### DNA extraction

To pre-process the samples, Proteinase K was added to the samples, together with silica (Biospec products, USA) to facilitate crushing of the filters with a pestle to promote DNA extraction. Further DNA extraction was performed using a DNeasy Blood and Tissue extraction kit (Qiagen, USA) in low bind Eppendorf tubes. DNA concentrations were measured using a Qubit 2.0 Fluorometer (Invitrogen, USA). DNA from filters from the two North Sea datasets were extracted using the Quick-DNA miniprep kit (Zymo, USA) according to the manufacturer's instructions. Details of both protocols are also given at protocols.io (dx.doi.org/10.17504/protocols.io.6yfhftn).

### Amplification

For PCR amplification of the target sequences of the samples from the Ocean aquarium, 10µL 2x Phire Tissue Direct PCR Master Mix (ThermoFisher Scientific, USA) was used. To the master mix, 0.4µL of each primer (10mM), 0.5µL template and nuclease free water was added to a total volume of 20µL. To reduce the effect of stochastic heterogeneity in PCR amplification, samples were amplified using 4 separate PCR

replicates, accept for filter 4, from which only 1 replicate could be obtained after amplification with PCR. When applied, blocking primer was added in a 5:1 ratio to the universal forward primer. A thermocycler was used with the following settings: initial denaturation at 98°C for 180s, 35 cycles of 98°C for 8s, 59.6°C for 8s, 72°C for 30s followed by a final extension at 72°C for 180s. Prior to sequencing, quality of the amplicons was checked on gel. A sample was also amplified at the Ocean aquarium field site using miniPCR 8 thermal cycler (miniPCR bio, UK) and 4 replicates were pooled to be able to skip downstream barcoding PCR. Samples were purified using a pre-prepared solution of Sera-Mag™ Magnetic SpeedBeads™ (Sigma-Aldrich, USA). Beads were prepared according to the original protocol (Deangelis, Wang, & Hawkins, 1995). Amplicon replicates were not pooled and consequently sequenced separately. For the North Sea samples, three PCR replicates were used in combination with a hot start and an annealing temperature of 59°C for 8s. In addition, field samples from Borkum Reef Grounds transect were amplified with 35-45 cycles, depending on sample (S2) to increase amplification yield. PCR replicates were pooled prior to purification with beads.

### Nanopore sequencing

Amplicon sequence library preparation was performed using the SQK-LSK109 kit and PCR barcoding kit 96 (PCB-096) (Oxford Nanopore Technologies Ltd., UK), according to the manufacturer's instructions, with the following adaptations: barcoding PCR was performed in a total volume of 15µL containing 0.3µL PCR barcode primer and 10-50ng amplicon. The applied barcode PCR program was as follows: initial denaturation at 95°C for 180s, 15 cycles of 95°C for 15s, 62°C for 15s, 65°C for 90s, followed by a final extension at 65°C for 180s.

After the barcoding PCR, sample concentration was estimated using the Qubit HS kit on the non-purified barcoded PCR products, and samples were pooled in equimolar ratios. The pooled amplicon sequence library was cleaned using magnetic beads, washed once with 70% ethanol and once with a mixture of Long Fragment Buffer (LFB) and Short Fragment Buffer (SFB) (2:1) to enrich for the 2kb target size fragments. During final clean-up, the library was again washed in a mixture of LFB and SFB in a ratio of 2:1. A maximum of 100ng DNA was loaded on a primed flow cell to prevent overloading of the flow cell. If necessary, the flow cell was refuelled using a mixture of Sequencing Buffer (SQB) and nuclease free water (1:1). Sequencing was performed until an estimated average sequencing depth of 30k reads per barcode was achieved. Base-calling was performed using Guppy (Version 4.2.2, Oxford Nanopore Technologies Ltd., UK) in high accuracy (HAC) mode. For the North Sea field samples, the estimated sequencing depth

differed, as 25-50k reads were obtained per barcode for From Borkum Reef Grounds/Gemini Wind Park samples and 70-100K per barcode for the shipwreck samples.

## Bioinformatics

The bioinformatic analysis was performed with our new pipeline called 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). Decona demultiplexes different barcodes with Qcat v1.1.0 (2018 Oxford Nanopore Technologies Ltd., UK). Furthermore, Decona uses Nanofilt v2.3.0 to filter raw fast reads on quality and read length (De Coster, D'Hert, Schultz, Cruts, & Van Broeckhoven, 2018). It then uses CD-hit v4.8.1, a program clustering reads based on short words rather than sequence alignment, to cluster the reads based on a set percentage of similarity (W. Li, Jaroszewski, & Godzik, 2002). The clustered reads are subsequently aligned using Minimap2 v2.17 (H. Li, 2018). Based on these alignments, Racon v1.4.13 is used to build the initial draft consensus sequence of each cluster (Vaser, Sović, Nagarajan, & Šikić, 2017) which is then polished by Medaka v1.1.2 (2018 Oxford Nanopore Technologies Ltd., UK). For read identification Decona is integrated with NCBI's BLAST+ command line tool.

For taxonomic identification, an in-house compiled reference database was used based on sequences available in the NCBI database. Separate databases were built to analyse the datasets of the Ocean aquarium (last search April 2018) and for the North Sea (last search April 2021) datasets. 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. To validate correct species identification, closely related and tropical fish species that do not occur in the Ocean aquarium were also added to the respective in-house databases, as were frequently occurring contaminants.

Species identification of the 2kb reads were obtained with the taxonomic identifier Centrifuge v1.0.4 (Kim, Song, Breitwieser, & Salzberg, 2016) with a minimal alignment length of 200 nucleotides and using only 1 primary assignment for each consensus sequences. 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 on species level, were labelled NO_ID. In addition, consensus sequences labelled with NO_ID were later verified using BLASTn with the online database of NCBI. No hits with >99% identity were missed but if there was a hit with lower sequence identity (>94%), the sequence was assigned at genus

level. Initially, the Decona build-in function BLASTn was used for taxonomic identification and to assess the accuracy of the amplicons. However, due to the use of the longer 2kb amplicon, Centrifuge considerably outperformed BLAST in terms of overall correct species identification (S4) and was therefore used for taxonomic identification.

### Data analysis

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 one (or less) successful PCR amplification replicate, were removed from the dataset. This resulted in the removal of several samples from 31$^{st}$ of July upon which was decided that all datapoints of that day were removed. Non-parametrical Wilcoxon rank test was performed to analyse the statistical effect of blocking primer. Non-metric multidimensional scaling ('jaccard') was performed in combination with PERMANOVA to analyse the effect of location in field samples.

## Results & Discussion

### Designed 2kb primers cover sufficient length to distinguish between closely related *Ammodytes* species

Prior to utilizing, the designed primers were tested *in silico* using the available full mitogenomes of North Sea and tropical marine fish species (data not shown) which indicated that both primers could anneal to both the North Sea and tropical fish and further vertebrate mitogenomic sequences. *In silico* comparison of the region covered by the 2kb primer of this study (12S-16S_ONT) and other commonly used 12S and 16S primers (Zhang et al., 2020), shows that the regions obtained from most of the previously used primer pairs are also obtained by the primer pair from this study (figure 2a). Exceptions are the commonly used Mifish and teleo2 primer pairs that both anneal to a region upstream of the forward primer binding site of this study. Alignment of 5 *Ammodytes* (sandeel) species with low genetic variability, *A. marinus*, *A. tobianus, A. hexapterus*, *A. personatus*, *A. dubius* and *Hyperoplus lanceolatus* shows a species specific pattern of single nucleotide polymorphisms (SNPs). Alignment of the previously used primer pairs shows that a limited amount of SNPs can be detected ranging from 0 to 5 between all species (figure 2b). In

contrast, when our newly designed primer pair is used, the SNPs from all primer pairs together can be detected while also detecting SNPs upstream of the 16S rRNA gene (figure 2b). In the region of the 16S rRNA gene with a relatively high amount of variability and high primer coverage, most of the *Ammodytes* species only have one or two unique SNPs, which is not sufficient for identification on species level (figure 2c). This *in silico* comparison therefore shows that enhanced identification resolution on species level can be obtained when using long read metabarcoding.

## Accuracy and detection sensitivity testing using the Ocean aquarium

### Enhanced sequence accuracy up to 100% using Decona

Samples collected in the Ocean aquarium yielded 599299 reads that were assigned to 247 different consensus sequences (Doorenspleet, K., Jansen, L., Oosterbroek S., Nijland, 2021). Barcode distribution was between 5-20k reads except for barcode 35 (sample processed at the aquarium site) which contained 286108 reads (S1. When the consensus sequences were identified with BLAStn, based on complete query coverage, more than 35% of these sequences could be correctly identified based on a percentage identity of 99.9-100%. A total of 71.6% of the consensus sequences could be correctly identified based on a percentage identity of 99.5% or higher. This demonstrates the potential of the Decona pipeline to successfully increase identification accuracy of nanopore sequence data with up to 10% points compared to the raw read accuracy. This accuracy is comparable to what is currently expected from highly accurate Illumina reads (Caporaso et al., 2011). To our knowledge a combination of bioinformatics tools in one line of code have not yet been described for nanopore based sequence read processing of metabarcoding data. The script and associated tools are available as the pipeline Decona (https://github.com/saskia-oosterbroek/decona). Decona is written in such way that only one line of code suffices to correctly run the pipeline. As such, data processing also becomes possible for scientists with limited experience in the command line. The bioinformatics tools integrated in Decona are well established programs widely used in genomics and transcriptomic studies (Huang, Niu, Gao, Fu, & Li, 2010; H. Li, 2018). Currently, there is limited automated bioinformatics processing reported in nanopore based studies, especially for metabarcoding (Santos et al., 2020). For example, tools as CD-Hit (Huang et al., 2010) have previously been used in nanopore studies for clustering (Voorhuijzen-Harink et al., 2019) and consensus building of fish amplicons. Reference based polishing was successfully applied when identifying benthic organisms on autonomous reef monitoring structures (Jin et al., 2020) using *minibarcoder.py* (Srivathsan et al., 2018). The combination of both clustering and *de novo* alignment based polishing with racon (Vaser, Sović, et al.,

2017) and medaka (https://nanoporetech.github.io/medaka/benchmarks.html#evaluation-across-samples-and-depths) has been used for the correction of metagenomes in the *NANOclust* pipeline (Rodríguez-Pérez, Ciuffreda, & Flores, 2020). Decona on the other hand combines similarity based clustering based on short word tables instead of an alignment approach in combination with alignment based polishing with racon and medaka which further increase the identification accuracies.

### Blocking primer reduces *Homo sapiens* reads but also reduces species richness

The use of a blocking primer significantly (p= 0.0013) reduced the relative read abundance of human DNA (figure 3a) from the eDNA sample. Overall, 15 species could be identified both with and without the use of blocking primer (figure 3c). However, the species *Chaetodon ulititensis* (1.5%), *Chelmon rostratus* (0.1%) and *Sufflamen chryopterum* (1.2%) were only detectable without the use of blocking primer (figure 3b). Initially, oligo blocking primers have been developed to block the amplification of otherwise dominating bacterial (Liles, Manske, Bintrim, Handelsman, & Goodman, 2003) taxa or host DNA in Antarctic krill diet studies to increase the detection of rarer reads/taxa (Vestheim & Jarman, 2008). The presence of an abundant proportion of human reads as observed in this study has previously been reported (Miya et al., 2015). Using an oligo blocking primer was effective for reduction of the amount of human sequences in this study. However, a decrease in species count was found, especially in taxa with low relative read abundance. A recent eDNA study that assessed the usage of blocking primers in combination with several fish specific short read primers, have also reported decreases in species richness (Zhang et al., 2020). Hence, the findings in our study agree with the suggestion to increase sequencing depth instead of using an oligo blocking primer to obtain the highest fish and vertebrate species richness (Zhang et al., 2020). Nevertheless, sequencing depth and time can be an important consideration, especially when sequencing budget is limited. Careful usage of an oligo blocking primer can then still be relevant for limited studies where a more general species pattern is more relevant than the complete diversity.

### Replication increases species detection sensitivity

In the Ocean aquarium dataset, four pseudo replicates are used by cutting the filters in four. After metabarcoding without using a blocking primer, the species composition and richness differed (Figure 4). Noteworthy, the species uniquely found on one of the filters; *C. rostratus* (0.4%), *Epinephelus flavocaeruleus* (0.6%), *Labroides dimidiatus* (0.3%), *Odonus niger* (0.4%), *Scomber scombrus* (1.5%) *and S. chrysopterum* (1.7%) all represent species with low total biomass in the aquarium. Of these, findings

of mackerel (*S. scombrus*) likely represents the detection of the feed that is used for the sharks. Several different numbers of replicates have been used between studies ranging from no replication (Gold, Sprague, Kushner, Zerecero Marin, & Barber, 2021; Stoeckle et al., 2021), three replicates (Andruszkiewicz et al., 2017; Singer et al., 2019) to five replicates (Jeunen et al., 2019). In our study there was a considerable increase in species richness and detection of unique rare reads when using replicates. Thus, filter replication increases the detection sensitivity. This is in agreement with similar observations where the use of replication increased the detection sensitivity (Beentjes, Speksnijder, Schilthuizen, Hoogeveen, & Van Der Hoorn, 2019; Evans et al., 2017). Recently, an increase of 25% of species count was found between using one and 18 biological 1L replicates, but using only limited number of replicates already detects the optimal number of species (Macher et al., 2021). The filters extracted and amplified at the Burgers' Zoo site, showed the highest species richness. However for this sample, the sequencing depth was 10 to 40 times higher than the separate barcodes of the lab processed filter replicates. This makes direct comparison between the processing locations delicate as it cannot be determined to which extend the sequencing depth could have influenced detection of field sample unique rare species. On the other hand, sampling approach, processing and sequencing depth considerably differed from samples processed in the lab. Therefore these data also shows the robustness of the method, independent of the processing location. Nevertheless, in this study, several species remained undetected, which could be improved by using more replicates, a greater sequencing depth or a greater volume of water where possible. Accordingly, careful considerations on the sampling design both in terms of replication, sampling volume and sequencing depth are also important when conducting long read eDNA metabarcoding.

### Majority of species are detected by long reads

Of the total number of consensus sequences, 92.2% of the raw reads resembling a consensus sequence could be taxonomically identified at species or genus level. The remaining 7.8% could not be taxonomically identified with the used reference database and identification tools. However, based on individual BLASTn results of the unassigned consensus sequences 2 additional genera, *Plectorinchus* and *Glaucostegus* were identified. Of all taxonomically identified sequences, 17 genera could be identified from 23 different genera present in the Ocean aquarium (figure 5a). These 23 genera represented 31 species of which 18 species could be identified (figure 5b). Specifically, the reads obtained with this long read metabarcoding method identified more than 58% of the species and 74% of genera present in the ocean aquarium. This percentage is lower than previous single marker validation studies where 80-100% of species present in a tank could be identified (Evans et al., 2017; Kelly, Port, Yamahara, & Crowder, 2014). It must be noted however that

in these studies often a low (5-10) number of fish species per tank is used. A validation experiment of a complex aquarium with 100+ species using multiple markers, demonstrated a maximum of 50% identification on species level and 80% on genus level when using 12S, 16S and COI together (Morey, Bartley, & Hanner, 2020). Results presented here are better comparable in terms of species complexity of the aquarium but especially the technical approach of using multiple markers. This, because *in silico* validation of the primer pair of this study shows that this fragment covers both the target regions of most of the commonly used 12S and 16S primers for fish eDNA studies and large regions of the 12S and 16S rRNA gene that are usually not targeted. *In silico* validation also shows that the 2kb fragment is better in distinguishing between closely related species of the *Ammodytes* genus than using a single 12S or 16S marker. We therefore suggest that long read metabarcoding targeting this 2kb fragment can be as effective if not more effective as using a multi marker approach to detect complex fish communities. Moreover, using multiple markers is more effective than single marker approaches but considerably increases the processing costs. Thus, an additional advantage of using a long read approach with a single primer pair to cover multiple genetic regions is increased species detection sensitivity without increasing laboratory costs.

### Incomplete reference database still hampers correct identification of long reads

Of the 5 genera that were not detected in this experiment, 3 genera (*Bodianus, Chrysiptera* and *Pomacentrus*) were only represented in the reference database with the 16S reference on genus level, and from 2 genera (*Ctenochaetus* and *Siganus*) the full mitogenome was available (Table 2). Although the full mitochondrial references of these genera were present in the database, these species were not detected. The species corresponding to these genera represent very small fish species, indicating that the availability of shed DNA from these relatively small fish could have been too limited for identification. This is in line with inconsistent detection of rare taxa between filters in previous reports (Evans et al., 2017; Kelly et al., 2014; Morey et al., 2020). Species names could be assigned based on both complete reference mitogenome or partial 12S or 16S references (Table 2). Of the undetected species *Acanthurus tennentii*, *Myripristis jacobus, Plectorhinchus obscurus* and *Chrysiptera parasema* no reference was available, which makes definitive identification of these species almost impossible. Instead of the unidentified species *Abudefduf sexfasciatus* and *Myripristis murdjan*, the species *Abudefduf vaigiensis* and *Myripristis berndti* were identified. For these two identified species the full mitogenome was present whereas for the undetected species, only the 16S reference was available. For primers targeting long DNA regions, a limited reference database can be the limiting factor as several parts of the target sequence might not (yet) be stored in the reference database. This study shows that even with only a single marker available in the

reference database, identification is still possible. However, a complete reference of the full length amplified region remains preferred as this will give more certainty on correct identification, especially of closely related species as this reduces the possibility of false positives.

## Biodiversity assessment of North Sea fish from different locations

### Long read eDNA metabarcoding detects ship wreck specific spatial variation in species compositions

From 9 different filtered water samples from 3 different shipwrecks in the Dutch part of the North Sea, a total of 262732 reads were analyzed (Doorenspleet, K., Jansen, L., Oosterbroek S., Nijland, 2021), with a barcode distribution between 5-20k reads (S1). A total of 79 consensus sequences were identified, which after correction indicated 21 species. A significant difference in beta-diversity (jaccard) between the 3 locations was found (p= 0.002) (figure 6a), which resulted in clustering of water sample replicates per ship wreck (figure 6b). Despite the similarity between replicates, some species only occurred in 1 of the 3 filter replicates on all locations. For example, the species scaldfish *Arnoglossus laterna,* solenette *Buglossidium luteum,* harbour porpoise *Phocoena phocoena* and plaice *Pleuronectes platessa* from ship wreck 1 only occurred in replicate 2, while the species sprat *Sprattus sprattus* and *S. scombrus* only occurred in replicate 3. As observed with the experiment with the ocean aquarium and earlier findings (Beentjes et al., 2019; Evans et al., 2017; Macher et al., 2021) these findings confirm that using multiple sample and filter replicates to increase the detected species richness of a location. From ship wreck 1 the composition according to the 3 samples mainly exists of grey gurnard *Eutrigla gurnardus*, dab *Limanda limanda* and bull trout *Myoxocephalus scorpius*. Unique for this shipwreck is the occurrence of sardine *Sardina pilchardus.* Around wreck 2 *E. gurnardus* is commonly found, and also sand goby *Pomatoschistus minutus* and plaice *Pleuronectes platessa* are common there. Turbot *Scophthalmus maximus, P. minutus* and common dragonet *Callionymus lyra* only were found at this ship wreck. The identifications from ship wreck 3 are mainly dominated by lesser sandeel *Ammodytes marinus* and to a lower extend in terms of read count per sample, *L. limanda* and *M. scorpius.* Clearly unique species composition profiles could be indicated for the 3 different shipwrecks, and the large number of detected fish species is a good basis for defining the ecological profile. Hence, the designed primer pair is universal enough to both target tropical and North Sea fish and elasmobranch species. Interestingly, in the samples from shipwreck 1, a moderate amount of *P. phocoena*, sequences was found. Although the 2kb primer set was designed on and for bony and cartilaginous fish, the detection of harbour porpoise in our field samples suggest the applicability of the primer pair for vertebrate species. Ecologically, detection of *P. phocoena* suggest the relatively high

productivity and abundance of fish in this area, on which harbor porpoise feed. Although it is unlikely that this method has captured the entire fish and marine vertebrate biodiversity, it is still sensitive enough to find relevant spatial differences between sampling sites. Especially using location specific sampling methods (e.g. ship wreck sampling) gives a general overview of the vertebrate richness on a local scale and give insight into visiting species such as the harbor porpoise which is likely to be left undetected with other methods.

### Long read eDNA metabarcoding detects variation in species composition in natural/artificial reef.

Overall, eDNA metabarcoding of 18 filtered water samples from 3 locations sampled at two time points at the transect from Borkum Reef Grounds to Gemini Wind Park resulted in 913119 sequences (Doorenspleet, K., Jansen, L., Oosterbroek S., Nijland, 2021), with a barcode distribution between 20-100k reads (S1). Based on 105 consensus sequences, 20 species could be identified. The species composition was significantly different between sampling locations and sampling dates ($p = 0.01$). The species composition of Gemini Wind Park differed significantly from that on the sandy bottom ($p=0.002$). The species on the Borkum Reef Grounds, on the other hand, were not significantly different from the other locations. The most abundant species at Gemini were *A. marinus, E. gurnardus* and *S. scombrus* (figure 7). For location Sandy bottom the most abundant species are *P. minutus* and sprat *Sprattus sprattus,* while for Borkum Reef Grounds this was *E. gurnardus*, *P. minutus* and *C. lyra*. In Gemini Wind Park, the common species composition hardly differed between the sampling dates 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 17$^{th}$ 2020 samples. At the Sandy bottom, more differences occurred between the 2 sampling dates. For example, the relative read abundance of *E. gurnardus* was higher on the 17$^{th}$ of July, while *S. pilchardus*, eel *Anguilla anguilla*, *B. luteum* and *L. limanda* were undetectable in the 17$^{th}$ of July samples. Also at the Borkum Reef Grounds, variation between sampling dates was found: *S. scombrus*, *S. pilchardus* and long-spined sea scorpion *Taurulus bubalis* were only detected the 2$^{nd}$ of July, while species *E. gurnardus*, *Arnoglossus laterna*, striped red mullet *Mullus surmuletus* and whiting *Merlangius merlangus only* on the 17$^{th}$ of July. Despite the location-specific species found in these samples, there is considerable difference when samples are taken at different timepoints. In this study we saw clear changes in presence of pelagic fish as *S. scombrus* and *S. pilchardus* at the Borkum Reef Grounds. On the other hand in the sandy bottom, bottom dwelling fish as *P. minutus* and *B. luteum* were found on the first sampling day but not on the second day. It is enticing to suggest that these results show short term dynamics like pelagic fish

462 movement in Borkum Reef Grounds and at the sandy bottom halfway between Gemini Wind Park and
463 Borkum Reef Grounds. However, on such short temporal scale it remains delicate to draw any meaningful
464 ecological conclusions as eDNA metabarcoding is highly sensitive (Beng & Corlett, 2020), even when
465 sampling in a relatively controlled environment. On the other end of the argument, temporal replication at
466 the shorter time scale is important to consider to obtain a more complete image of the diversity in a certain
467 area.

468

## 469 Conclusions

470 This study demonstrates and validates an eDNA metabarcoding approach using nanopore long read
471 technology, enabling increased resolution on species level. Highlighted are increased species resolution
472 due to the longer DNA fragments analyzed enhanced by our nanopore raw read processing pipeline Decona.
473 We further show that limitations as insensitivity issues from incomplete reference databases remain in long
474 read metabarcoding. Detection limitations also occur when using an oligo blocking primer. Further research
475 should focus on direct comparisons of long read nanopore approaches with other sequencing platforms.
476 Another interesting possibility is the use of long read metabarcoding to study spatial-temporal shifts based
477 on the detection of eDNA fragments of different lengths. Moreover, it is essential that additions of longer
478 reference sequences, preferably full mitogenomes, to reference databases become a high priority in marine
479 molecular ecology, as only then long read based DNA metabarcoding and metagenomics can develop to
480 its full potential to serve as monitoring tool.

481

## 482 Acknowledgements

483 We are grateful for the members of 'Duik the Noordzee Schoon' foundation for providing the trip to the
484 ship wreck and assisting during the dives and sampling. We would like to thank Linda Tonk of Wageningen
485 Marine Research, Miriam Schutter of Bureau Waardenburg and crew members of MS Vrijheid III, for
486 providing the cruise to the Borkum Reef Grounds and Gemini Wind Park and assistance with sampling. We
487 acknowledge GEANS, an Interreg project supported by the North Sea Programme of the European Regional
488 Development Fund of the European Union and JIP ECO-FRIEND (RVO reference number TEWZ118017) for
489 funding parts of this research. Especially valuable contributions are made by Aline Joustra by providing the
490 illustrations of the experimental design (http://www.alinesci.com/).

## Authors contributions

LJ and RN designed the experiment; LJ, PK, OB, MJ and EW were involved in sample collection and processing; LJ and RN did the laboratory work; SO, KD and RN designed the bioinformatics pipeline Decona; KD and RN conducted the data analysis; KD, LJ, SO, RN and TM interpreted the data; all authors wrote and revised the manuscript.

## Data availability

All raw sequence data that support these findings are available in ENA at reference XXXXX.

## References

Andersen, J. H., Al-Hamdani, Z., Harvey, E. T., Kallenbach, E., Murray, C., & Stock, A. (2020). Relative impacts of multiple human stressors in estuaries and coastal waters in the North Sea–Baltic Sea transition zone. *Science of the Total Environment*, *704*, 135316. doi: 10.1016/j.scitotenv.2019.135316

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*, *12*(4), e0176343. doi: 10.1371/journal.pone.0176343

Balasingham, K. D., Walter, R. P., Mandrak, N. E., & Heath, D. D. (2018). Environmental DNA detection of rare and invasive fish species in two Great Lakes tributaries. *Molecular Ecology*, *27*(1), 112–127. doi: 10.1111/mec.14395

Baloğlu, B., Chen, Z., Elbrecht, V., Braukmann, T., MacDonald, S., & Baloglu, B. (2020). *A workflow for accurate metabarcoding using nanopore MinION sequencing 1*. doi: 10.1101/2020.05.21.108852

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Hoogeveen, M., & Van Der Hoorn, B. B. (2019). The effects of spatial and temporal replicate sampling on eDNA metabarcoding. *PeerJ*, *2019*(7), e7335. doi: 10.7717/peerj.7335

Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodiversity and Conservation*, Vol. 29, pp. 2089–2121. doi: 10.1007/s10531-020-01980-0

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., … Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, edn3.74. doi: 10.1002/edn3.74

Bleidorn, C. (2016, January 2). Third generation sequencing: Technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, Vol. 14, pp. 1–8. Taylor and Francis Ltd. doi: 10.1080/14772000.2015.1099575

Bleijswijk, J. D. L., Engelmann, J. C., Klunder, L., Witte, H. J., Witte, J. I., & Veer, H. W. (2020). Analysis of a coastal North Sea fish community: Comparison of aquatic environmental DNA concentrations to fish catches. *Environmental DNA*, edn3.67. doi: 10.1002/edn3.67

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., … Knight, R. (2011). Moving pictures of the human microbiome. *Genome Biology*, *12*(5). doi: 10.1186/gb-2011-12-5-r50

530 Capuzzo, E., Lynam, C. P., Barry, J., Stephens, D., Forster, R. M., Greenwood, N., … Engelhard, G. H. (2018). A
531 decline in primary production in the North Sea over 25 years, associated with reductions in zooplankton
532 abundance and fish stock recruitment. *Global Change Biology*, *24*(1), e352–e364. doi:
533 10.1111/gcb.13916

534 Carradec, Q., Poulain, J., Boissin, E., Hume, B. C., Voolstra, C. R., Ziegler, M., … Wincker, P. (2020). A
535 framework for in situ molecular characterization of coral holobionts using nanopore sequencing. *BioRxiv*
536 *Genomics*, *10*(1), 1–25. doi: 10.1101/2020.05.25.071951

537 Chamberlain, S. A., & Szöcs, E. (2013). Taxize: Taxonomic search and retrieval in R. *F1000Research*, *2*. doi:
538 10.12688/f1000research.2-191.v2

539 Claudet, J. (2018). Six conditions under which MPAs might not appear effective (when they are). *ICES Journal*
540 *of Marine Science*, *75*(3), 1172–1174. doi: 10.1093/icesjms/fsx074

541 Couce, E., Schratzberger, M., & Engelhard, G. H. (2020). Reconstructing three decades of total international
542 trawling effort in the North Sea. *Earth System Science Data*, *12*(1), 373–386. doi: 10.5194/essd-12-373-
543 2020

544 Daan, N., Gislason, H., Pope, J. G., & Rice, J. C. (2005). Changes in the North Sea fish community: Evidence of
545 indirect effects of fishing? *ICES Journal of Marine Science*, *62*(2), 177–188. doi:
546 10.1016/j.icesjms.2004.08.020

547 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and
548 processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. doi:
549 10.1093/bioinformatics/bty149

550 Deangelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation
551 of PCR products. *Nucleic Acids Research*, *23*(22), 4742–4743. doi: 10.1093/nar/23.22.4742

552 Degraer, S., Carey, D. A., Coolen, J. W. P., Hutchison, Z. L., Kerckhof, F., Rumes, B., & Vanaverbeke, J.
553 (2020). OFFSHORE WIND FARM ARTIFICIAL REEFS AFFECT ECOSYSTEM STRUCTURE AND FUNCTIONING.
554 *SPECIAL ISSUE ON UNDERSTANDING THE EFFECTS OF OFFSHORE WIND ENERGY DEVELOPMENT ON*
555 *FISHERIES*, *33*(4), 48–57. doi: 10.2307/26965749

556 Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017,
557 November 1). Environmental DNA metabarcoding: Transforming how we survey animal and plant
558 communities. *Molecular Ecology*, Vol. 26, pp. 5872–5895. Blackwell Publishing Ltd. doi:
559 10.1111/mec.14350

560 DiBattista, J. D., Coker, D. J., Sinclair-Taylor, T. H., Stat, M., Berumen, M. L., & Bunce, M. (2017, December 1).
561 Assessing the utility of eDNA as a tool to survey reef-fish communities in the Red Sea. *Coral Reefs*, Vol.
562 36, pp. 1245–1252. Springer Verlag. doi: 10.1007/s00338-017-1618-1

563 Didderen, K., Lengkeek, W., Bergsma, J. H., & Dongen, U. Van. (2019). *WWF & ARK Borkum Reef Ground*
564 *oyster pilot*. (September).

565 Doi, H., Fukaya, K., Oka, S. ichiro, Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection
566 probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a
567 multispecies site occupancy model. *Scientific Reports*, *9*(1), 1–8. doi: 10.1038/s41598-019-40233-1

568 Doorenspleet, K., Jansen, L., Oosterbroek S., Nijland, R. (2021). Raw sequencing data for accurate long read
569 metabarcoding of marine eDNA. *European Nucleotides Archives*, *XXX*(XX), XXX. Retrieved from XXX

570 Egeter, B., Veríssimo, J., Lopes-Lima, M., Chaves, C., Pinto, J., Riccardi, N., … Fonseca, N. A. (2020). Speeding
571 up the detection of invasive aquatic species using environmental DNA and nanopore sequencing. *BioRxiv*,
572 *326*(1), 2020.06.09.142521. doi: 10.1101/2020.06.09.142521

573 Eggleton, J. D., Depestele, J., Kenny, A. J., Bolam, S. G., & Garcia, C. (2018). How benthic habitats and bottom
574 trawling affect trait composition in the diet of seven demersal and benthivorous fish species in the North
575 Sea. *Journal of Sea Research*, *142*, 132–146. doi: 10.1016/j.seares.2018.09.013

576 Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., … Pfrender, M. E. (2017). Fish
577 community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering.
578 *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(9), 1362–1374. doi: 10.1139/cjfas-2016-0306

579 Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., … Lodge, D. M. (2016).
580 Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding.
581 *Molecular Ecology Resources*, *16*(1), 29–41. doi: 10.1111/1755-0998.12433

582 Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA
583 from water samples. *Biology Letters*, *4*(4), 423–425. doi: 10.1098/rsbl.2008.0118

584 Gold, Z., Sprague, J., Kushner, D. J., Zerecero Marin, E., & Barber, P. H. (2021). eDNA metabarcoding as a
585 biomonitoring tool for marine protected areas. *PLOS ONE*, *16*(2), e0238557. doi:
586 10.1371/journal.pone.0238557

587 Goodwin, K. D., Thompson, L. R., Duarte, B., Kahlke, T., Thompson, A. R., Marques, J. C., & Caçador, I. (2017,
588 May 9). DNA sequencing as a tool to monitor marine ecological status. *Frontiers in Marine Science*, Vol. 4.
589 Frontiers Media S. A. doi: 10.3389/fmars.2017.00107

590 Hestetun, J. T., Bye-Ingebrigtsen, E., Nilsson, R. H., Glover, A. G., Johansen, P. O., & Dahlgren, T. G. (2020).
591 Significant taxon sampling gaps in DNA databases limit the operational use of marine macrofauna
592 metabarcoding. *Marine Biodiversity*, *50*(5), 1–9. doi: 10.1007/s12526-020-01093-5

593 Hofstede, R. Ter, Hiddink, J. G., & Rijnsdorp, A. D. (2010). Regional warming changes fish species richness in
594 the eastern North Atlantic Ocean. *Marine Ecology Progress Series*, *414*(August), 1–9. doi:
595 10.3354/meps08753

596 Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing
597 biological sequences. *Bioinformatics*, *26*(5), 680–682. doi: 10.1093/bioinformatics/btq003

598 Hunter, M. E., Ferrante, J. A., Meigs-Friend, G., & Ulmer, A. (2019). Improving eDNA yield and inhibitor
599 reduction through increased water volumes and multi-filter isolation techniques. *Scientific Reports*, *9*(1),
600 1–9. doi: 10.1038/s41598-019-40977-w

601 Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., & Akeson, M. (2015). Improved data analysis for the
602 MinION nanopore sequencer. *Nature Methods*, *12*(4), 351–356. doi: 10.1038/nmeth.3290

603 Jeunen, G., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., … Gemmell, N. J. (2019).
604 Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats
605 connected by water movement. *Molecular Ecology Resources*, *19*(2), 426–438. doi: 10.1111/1755-
606 0998.12982

607 Jin, J., Chang, M., Cheong, Y., Ip, A., Bauman, A. G., & Huang, D. (2020). MinION-in-ARMS: Nanopore
608 Sequencing To Expedite Barcoding Of Specimen-Rich Macrofaunal Samples From Autonomous Reef 2
609 Monitoring Structures 3 4. *BioRxiv*, 2020.03.30.009654. doi: 10.1101/2020.03.30.009654

610 Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., … Weinstock, G. M.
611 (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature*
612 *Communications*, *10*(1), 1–11. doi: 10.1038/s41467-019-13036-1

613 Kamermans, P., van Duren, L., & Kleissen, F. (2018). European flat oysters on offshore wind farms: additional
614 locations; Opportunities for the development of European flat oyster (Ostrea edulis) populations on
615 planned wind farms and additional locations in the Dutch section of the North Sea Wageningen, Wage.
616 *Wageningen Marine Research*, *C053/18*, 33. Retrieved from internal-pdf://0719885351/Kamermans et al,
617 2018.pdf

618 Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in
619 Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. doi:
620 10.1093/molbev/mst010

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

624 Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine
625 fishes in a large mesocosm. *PLoS ONE*, *9*(1), 86175. doi: 10.1371/journal.pone.0086175

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

628 Krehenwinkel, H., Pomerantz, A., & Prost, S. (2019). Genetic biomonitoring and biodiversity assessment using
629 portable sequencing technologies: Current uses and future directions. *Genes*, *10*(11). doi:
630 10.3390/genes10110858

631 Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with
632 eDNA: a promising tool for fisheries management. *Journal of Applied Ecology*, *53*(4), 1148–1157. doi:
633 10.1111/1365-2664.12598

634 Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. doi:
635 10.1093/bioinformatics/bty191

636 Li, W., Jaroszewski, L., & Godzik, A. (2002). Tolerating some redundancy significantly speeds up clustering of
637 large protein databases. *Bioinformatics*, *18*(1), 77–82. doi: 10.1093/bioinformatics/18.1.77

Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J., & Goodman, R. M. (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology*, *69*(5), 2684–2691. doi: 10.1128/AEM.69.5.2684-2691.2003

Macher, T.-H., Schütz, R., Arle, J., Beermann, A. J., Koschorreck, J., & Leese, F. (2021). Beyond fish eDNA metabarcoding: Field replicates disproportionately improve the detection of stream associated vertebrate species. *BioRxiv*, 2021.03.26.437227. doi: 10.1101/2021.03.26.437227

Mateos-Rivera, A., Skern-Mauritzen, R., Dahle, G., Sundby, S., Mozfar, B., Thorsen, A., … Krafft, B. A. (2020). Comparison of visual and molecular taxonomic methods to identify ichthyoplankton in the North Sea. *Limnology and Oceanography: Methods*. doi: 10.1002/lom3.10387

McGee, K. M., Robinson, C. V., & Hajibabaei, M. (2019). Gaps in DNA-Based Biomonitoring Across the Globe. *Frontiers in Ecology and Evolution*, *7*, 337. doi: 10.3389/fevo.2019.00337

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. doi: 10.1371/journal.pone.0061217

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

Morey, K. C., Bartley, T. J., & Hanner, R. H. (2020). Validating environmental DNA metabarcoding for marine fishes in diverse ecosystems using a public aquarium. *Environmental DNA*, *2*(3), 330–342. doi: 10.1002/edn3.76

O'Brien, A. L., Dafforn, K. A., Chariton, A. A., Johnston, E. L., & Mayer-Pinto, M. (2019, September 20). After decades of stressor research in urban estuarine ecosystems the focus is still on single stressors: A systematic literature review and meta-analysis. *Science of the Total Environment*, Vol. 684, pp. 753–764. Elsevier B.V. doi: 10.1016/j.scitotenv.2019.02.131

Reiss, H., Degraer, S., Duineveld, G. C. A., Kröncke, I., Aldridge, J., Craeymeersch, J. A., … Rees, H. L. (2010). Spatial patterns of infauna, epifauna, and demersal fish communities in the North Sea. *ICES Journal of Marine Science*, *67*(2), 278–293. doi: 10.1093/icesjms/fsp253

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21), e145. doi: 10.1093/nar/gkr732

Rodríguez-Pérez, H., Ciuffreda, L., & Flores, C. (2020). NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics*. doi: 10.1093/bioinformatics/btaa900

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019, January 1). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, Vol. 17, p. e00547. Elsevier B.V. doi: 10.1016/j.gecco.2019.e00547

Santos, A., van Aerle, R., Barrientos, L., & Martinez-Urtaza, J. (2020, January 1). Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Computational and Structural Biotechnology Journal*, Vol. 18, pp. 296–305. Elsevier B.V. doi: 10.1016/j.csbj.2020.01.005

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. doi: 10.1002/edn3.38

Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., & Boehm, A. B. (2016). Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science and Technology*, *50*(19), 10456–10464. doi: 10.1021/acs.est.6b03114

Schon, E. A. (2000, November 1). Mitochondrial genetics and disease. *Trends in Biochemical Sciences*, Vol. 25, pp. 555–560. Trends Biochem Sci. doi: 10.1016/S0968-0004(00)01688-1

Seymour, M., Durance, I., Cosby, B. J., Ransom-Jones, E., Deiner, K., Ormerod, S. J., … Creer, S. (2018). Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, *1*(1). doi: 10.1038/s42003-017-0005-3

Shin, J., Lee, S., Go, M. J., Lee, S. Y., Kim, S. C., Lee, C. H., & Cho, B. K. (2016). Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Scientific Reports*, *6*. doi: 10.1038/srep29681

Shu, L., Ludwig, A., & Peng, Z. (2020, March 1). Standards for methods utilizing environmental dna for detection of fish species. *Genes*, Vol. 11, p. 296. MDPI AG. doi: 10.3390/genes11030296

692 Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., & Hajibabaei, M. (2019). Comprehensive
693 biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding
694 seawater. *Scientific Reports*, *9*(1), 1–12. doi: 10.1038/s41598-019-42455-9

695 Srivathsan, A., Baloğlu, B., Wang, W., Tan, W. X., Bertrand, D., Ng, A. H. Q., … Meier, R. (2018). A MinION™-
696 based pipeline for fast and cost-effective DNA barcoding. *Molecular Ecology Resources*, *18*(5), 1035–
697 1049. doi: 10.1111/1755-0998.12890

698 Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z., Dunton, K. J., Hinks, G., & VanMorter, S. M. (2021). Trawl and
699 eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey,
700 USA . *ICES Journal of Marine Science*, *78*(1), 293–304. doi: 10.1093/icesjms/fsaa225

701 Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For biodiversity research and
702 monitoring. In *Environmental DNA: For Biodiversity Research and Monitoring*. doi:
703 10.1093/oso/9780198767220.001.0001

704 Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation
705 biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. doi:
706 10.1111/j.1365-294X.2012.05470.x

707 Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in
708 Illumina paired-end sequencing. *Scientific Reports*, *9*(1), 1–7. doi: 10.1038/s41598-019-39076-7

709 Teletchea, F. (2009). Molecular identification methods of fish species: Reassessment and possible applications.
710 *Reviews in Fish Biology and Fisheries*, *19*(3), 265–293. doi: 10.1007/s11160-009-9107-4

711 Thalinger, B., Wolf, E., Traugott, M., & Wanzenböck, J. (2019). Monitoring spawning migrations of
712 potamodromous fish species via eDNA. *Scientific Reports*, *9*(1), 1–11. doi: 10.1038/s41598-019-51398-0

713 Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a
714 Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE*, *7*(8). doi:
715 10.1371/journal.pone.0041732

716 Truelove, N. K., Andruszkiewicz, E. A., & Block, B. A. (2019). A rapid environmental DNA method for detecting
717 white sharks in the open ocean. *Methods in Ecology and Evolution*, *10*(8), 1128–1135. doi:
718 10.1111/2041-210X.13201

719 van der Loos, L. M., & Nijland, R. (2020). Biases in bulk: DNA metabarcoding of marine communities and the
720 methodology involved. *Molecular Ecology*. Blackwell Publishing Ltd. doi: 10.1111/mec.15592

721 Vaser, R., Sović, I., Sović, S., Nagarajan, N., Šikić1, M., & Šikić1, Š. (2017). Fast and accurate de novo
722 genome assembly from long uncorrected reads. *Genome.Cshlp.Org*. doi: 10.1101/gr.214270.116

723 Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long
724 uncorrected reads. *Genome Research*, *27*(5), 737–746. doi: 10.1101/gr.214270.116

725 Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in
726 mixed samples - A case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, *5*(1), 1–11.
727 doi: 10.1186/1742-9994-5-12

728 Voorhuijzen-Harink, M. M., Hagelaar, R., van Dijk, J. P., Prins, T. W., Kok, E. J., & Staats, M. (2019). Toward
729 on-site food authentication using nanopore sequencing. *Food Chemistry: X*, *2*, 100035. doi:
730 10.1016/j.fochx.2019.100035

731 Zhang, S., Zhao, J., & Yao, M. (2020). A comprehensive and comparative evaluation of primers for
732 metabarcoding eDNA from fish. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210x.13485

733

734

735 **Table 1:** Primer sequence of the forward, reverse and oligo blocking primer.

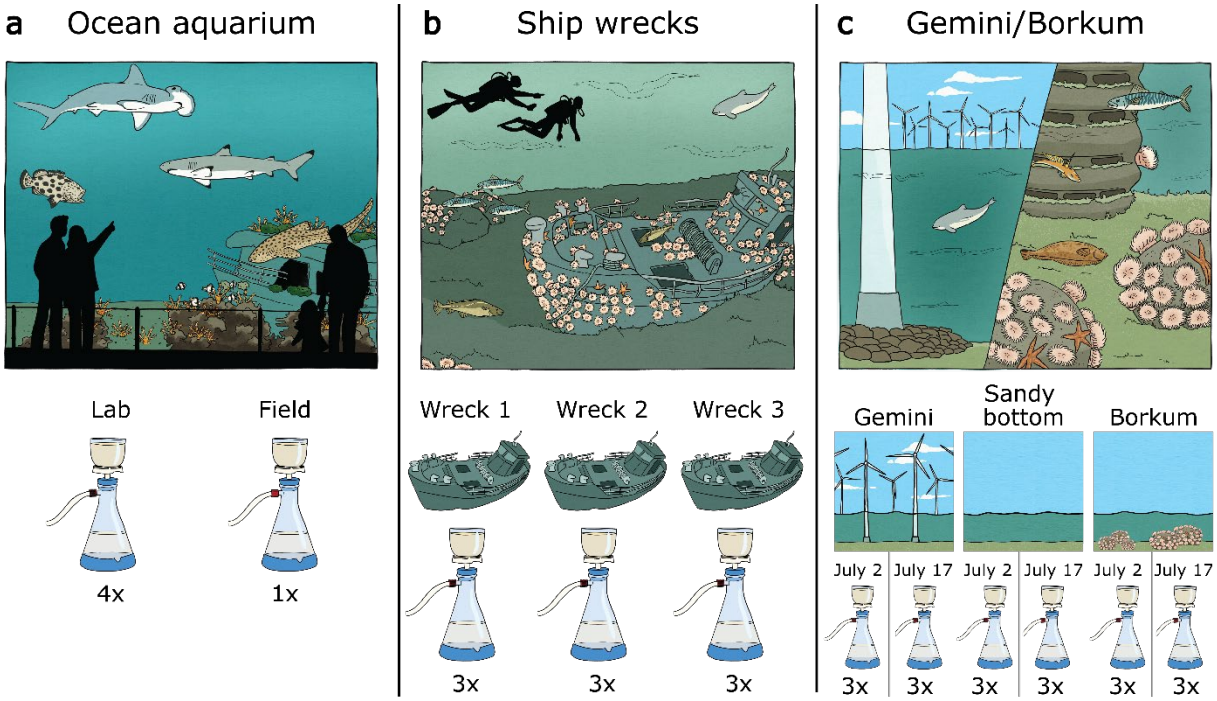
| | |
|---|---|
| Forward primer | Fish_12S_fw1-ONT:TTTCTGTTGGTGCTGATATTGCGGATTAGATACCCYACTATGC |
| Reverse primer | Fish_16S_rv1-ONT: ACTTGCCTGTCGCTCTATCTTCGATTGCGCTGTTATCCCTAG |
| Oligo blocking primer | Fish_12S_Fw-BLOCK-H.s : TATGCTTAGCCCTAAACCTCAAC/3SpC3/ |

736

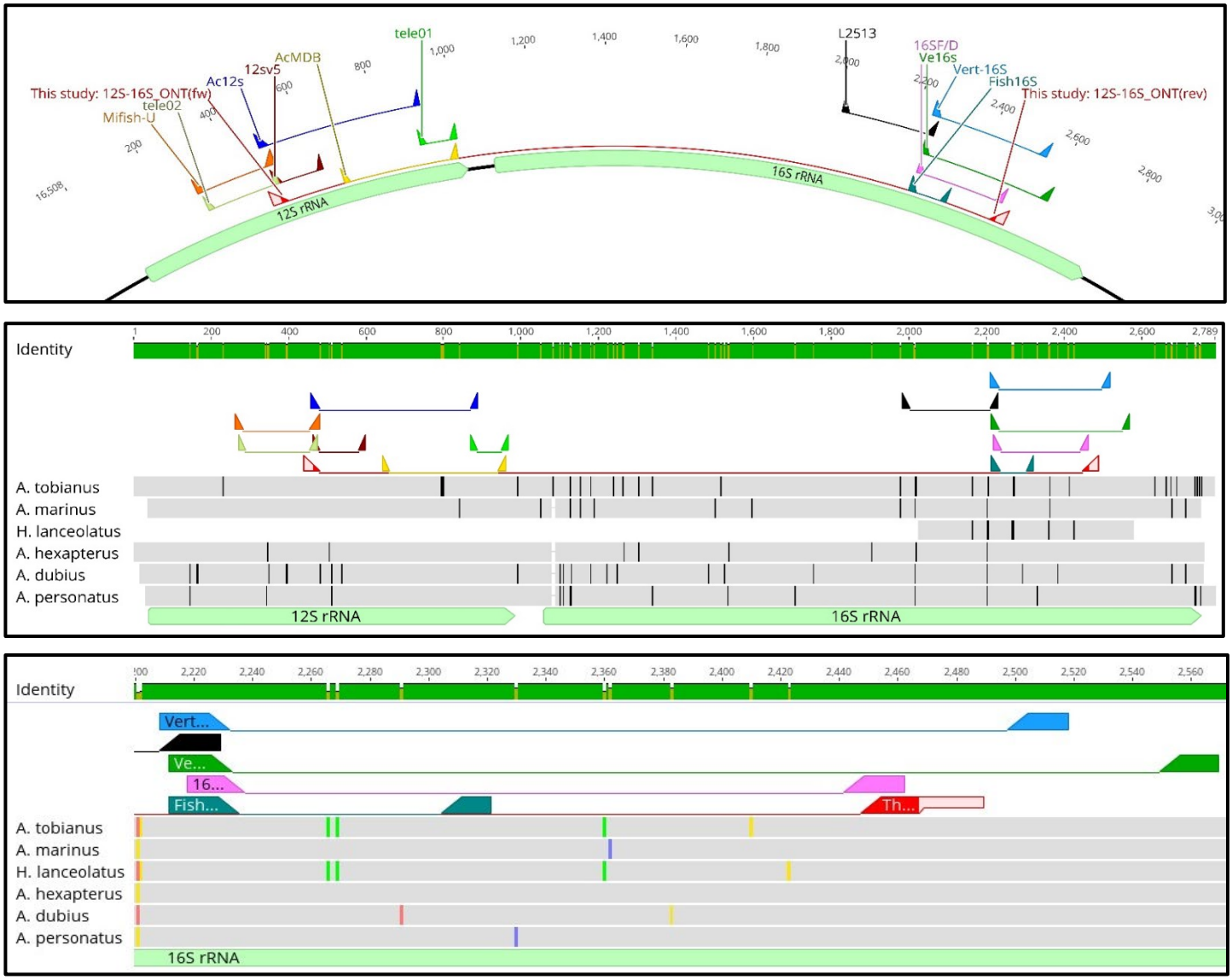
737

738

739




741 **Figure 1:** Sampling design of **a)** Ocean aquarium of Burgers' Zoo, the Netherlands **b)** North Sea ship
742 wrecks with three different ship wreck locations and **c)** Borkum/Gemini where samples were taken in
743 Gemini, halfway between Gemini and Borkum on a sandy bottom and Borkum reef grounds.




744

745 **Figure 2:** Comparison of primer binding sites of this study and other commonly used fish eDNA primers
746 according to Zang et al (2020). Where **a)** shows the specific primer binding locations (triangles) of the 2kb
747 primer pair used in this study (red) and other studies indicated in different colours. Red fill in the triangle
748 indicates the binding site, empty part of the triangle indicates the ONT extension. Alignment of **b)** of the
749 12S and 16S rRNA gene of different *Ammodytes* species that occur in both the North Sea and Atlantic
750 Ocean. Black stripes represent SNPs. **c)** Comparison of the region with the highest variability covered by
751 the highest amount of primers. Four colours represent different nucleotides (green = T, yellow = G, red =
752 A and blue = C).

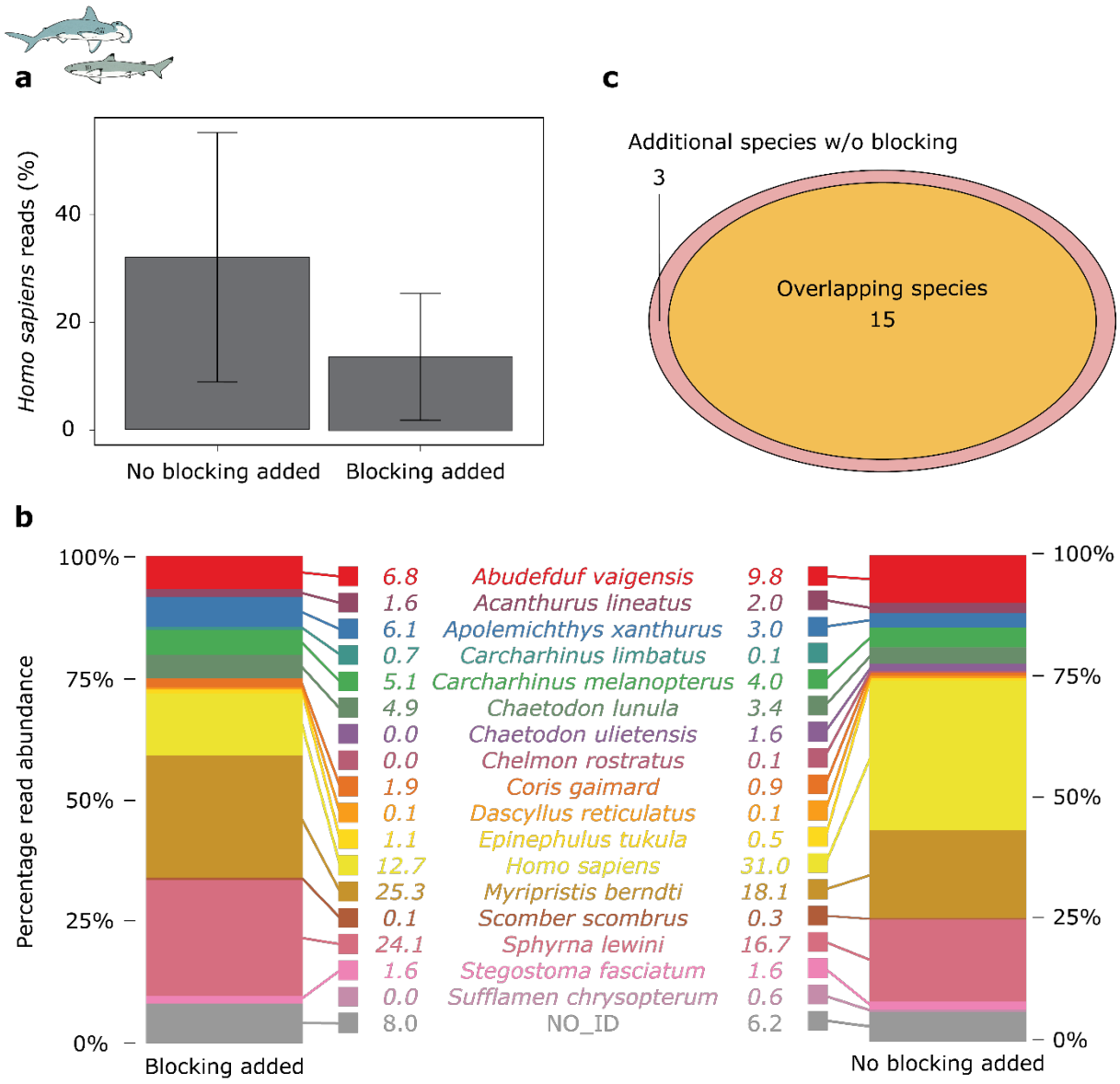



**Figure 1: a**) Represents the relative proportion of H. sapiens reads. **b)** Percentage read abundance plot of the species found when adding an oligo blocking primer (first) or without adding an oligo blocking primer. The different colors represent the species found. In the **c),** Euler-diagram the amount of overlapping and unique species are visualized. Red represents the species found when no blocking primer was added while yellow represent the found species when the blocking primer was added.

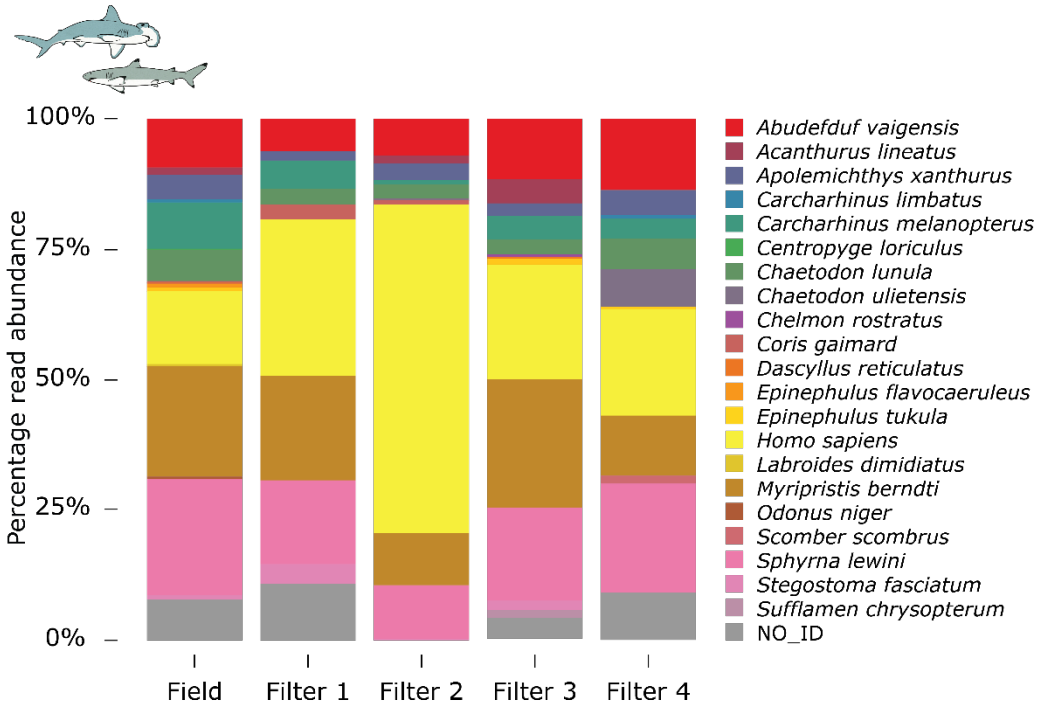
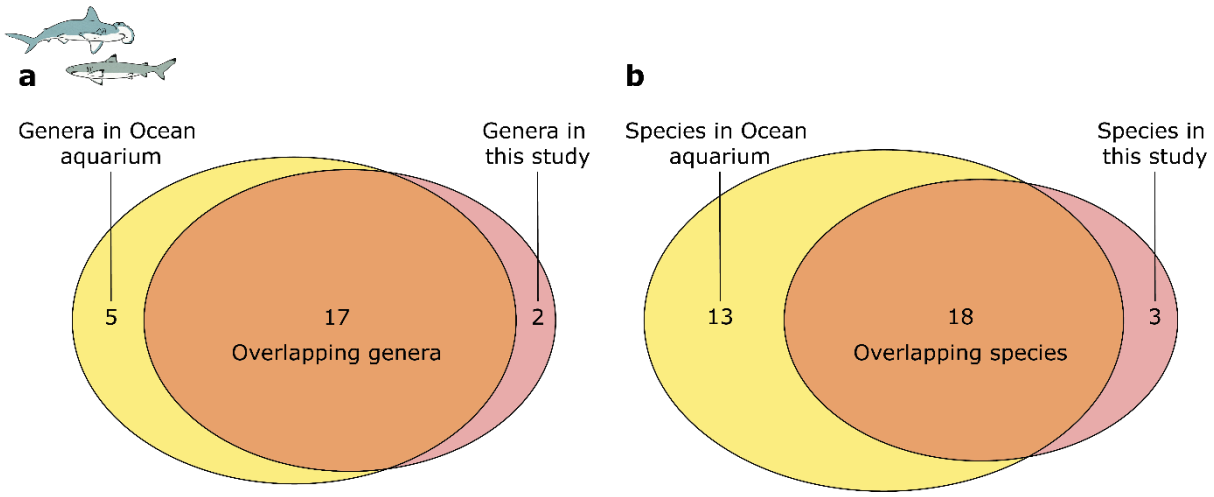



**Figure 2:** Percentage read abundance plot of the sample processed at the field site and different filter replicates processed in the lab, all collected from the curated Ocean aquarium at Burgers' Zoo. Filters represent samples where no blocking primer was added. Colours indicate the proportion of reads found for each species.




765 **Figure 3:** Euler-diagrams of the amount of overlapping and unique taxa in the samples. Yellow represents
766 the total amount of taxa present in the Ocean aquarium but not found in this data. Orange represents the
767 taxa found in both aquarium and eDNA method, while red represents the taxa found in the eDNA samples
768 that are not present in the aquarium (false positives). Diagram **a)** represents the overlapping genera while
769 **b)** represents the overlapping species.

770

771

772

773

774

775

776

777

778

779

780

781

782

783

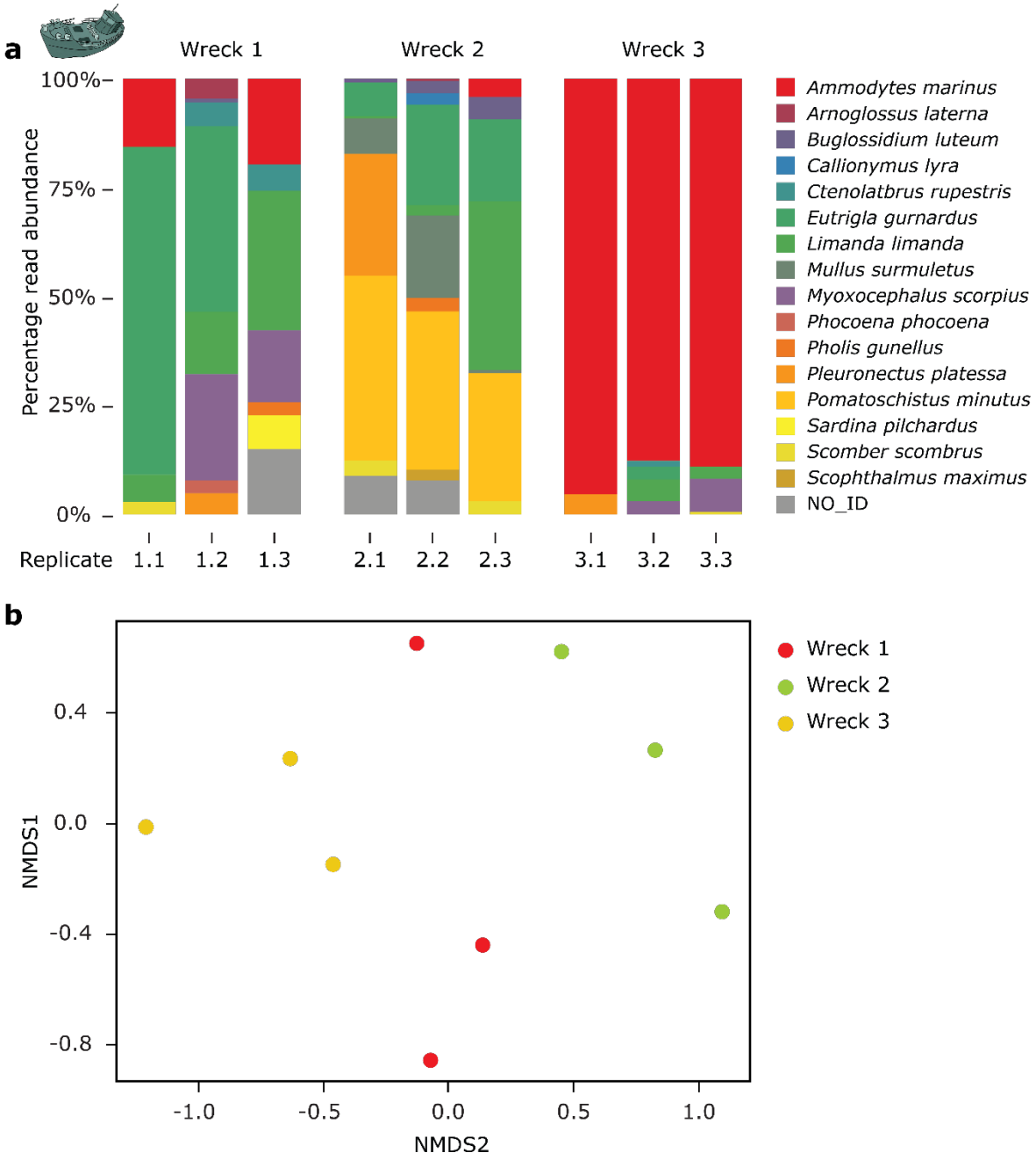
784

785

786

787

788




789

790 **Figure 6: a)** Proportional read abundance plot of marine vertebrate species associated with the different
791 ship wrecks. Bars indicate from which ship wreck the water samples were taken, including replicate
792 number. Colours indicate the relative read count of the identified fish species. **b)** NMDS plot of the beta-
793 diversity between the different ship wrecks.

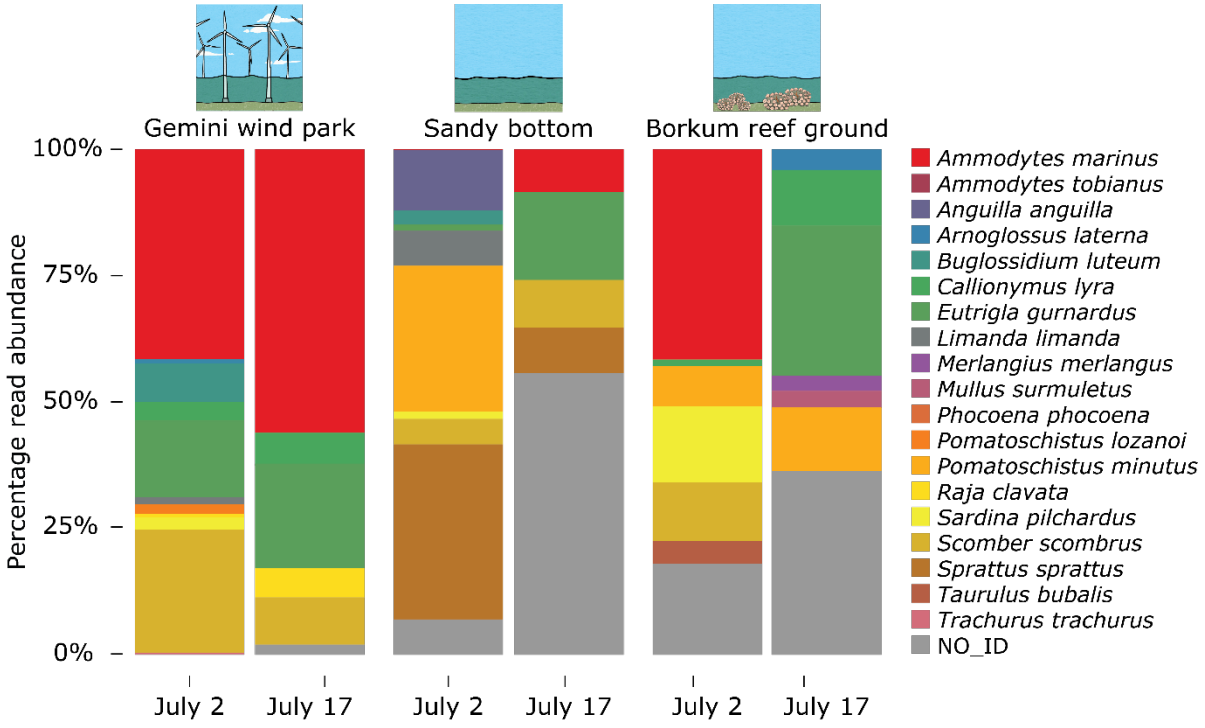



**Figure 4:** Percentage read abundance plot of the marine vertebrate species in the locations Gemini wind park, at sandy bottom halfway Gemini wind park and the Borkum reef grounds and at the Borkum reef grounds. Bars indicate at which date the water samples were taken. Colours indicate the proportion read abundance of the identified fish species.

**Table 2:** Comparison of identification sensitivity to the used reference database. The presence of the species in this study is also compared with the completeness of the database to determine the expected Identification status (expected ID status). Genus and species show genera and species present in the Ocean aquarium, ID represents whether the method detected the species or not. Database represent the availability of that taxon in the database. Expected ID status represents source of insensitivity. Misidentification represents the false p

| Genus | ID? | Reference available | Species | ID? | Reference available | expected identification status* | misidentification | reference misidentification |
|---|---|---|---|---|---|---|---|---|
| *Abudefduf* | YES | Complete mitogenome | *Abudefduf sexfasciatus* | NO | partial 16s | other species identified | *Abudefduf vaigiensis* | Complete mitogenome |
| *Acanthochromis* | NO | partial 16s | *Acanthochromis polyacanthus* | YES | partial 16s | ID based on incomplete reference | - | |
| *Acanthurus* | YES | Complete mitogenome | *Acanthurus lineatus* | YES | Complete mitogenome | ID complete | - | |
| | | | *Acanthurus tennentii* | NO | No reference | unable to identify | - | |
| *Apolemichthys* | YES | partial 16s | *Apolemichthys xanthurus* | YES | partial 16s | ID based on incomplete reference | - | |
| *Bodianus* | NO | partial 16s | *Bodianus anthioides* | NO | partial 16s | likely undetected | - | |
| *Carcharhinus* | YES | Complete mitogenome | *Carcharhinus limbatus* | YES | complete 12s t-val 16s | ID complete | - | |
| | | | *Carcharhinus melanopterus* | YES | Complete mitogenome | ID complete | - | |
| *Centropyge* | YES | Complete mitogenome | *Centropyge bicolor* | NO | partial 16s | likely undetected | - | |
| | | | *Centropyge loriculus* | YES | Complete mitogenome | ID complete | - | |
| *Chaetodon* | YES | Complete mitogenome | *Chaetodon falcula* | NO | partial 16s | ? | - | |
| | | | *Chaetodon lunula* | YES | partial 12s | ID based on incomplete reference | - | |
| | | | *Chaetodon ulietensis* | YES | partial 16s | ID based on incomplete reference | - | |
| *Chelmon* | YES | Complete mitogenome | *Chelmon rostratus* | YES | Complete mitogenome | ID complete | - | |
| *Chrysiptera* | NO | partial 16s | *Chrysiptera parasema* | NO | No reference | unable to identify | - | |
| *Coris* | YES | partial 16s | *Coris gaimard* | YES | partial 16s | ID based on incomplete reference | - | |
| *Ctenochaetus* | NO | Complete mitogenome | *Ctenochaetus strigosus* | NO | No reference | unable to identify/likely unditected | - | |
| *Dascyllus* | YES | partial 16s | *Dascyllus aruanus* | NO | partial 16s | ? | - | |
| | | | *Dascyllus reticulatus* | YES | partial 16s | ID based on incomplete reference | - | |
| | | | *Dascyllus trimaculatus* | NO | partial 16s | ? | - | |
| *Epinephelus* | YES | Complete mitogenome | *Epinephelus tukula* | YES | Complete mitogenome | ID complete | - | |
| | | | *Epinephelus flavocearuleus* | YES | Complete mitogenome | ID complete | - | |
| *Labroides* | YES | complete 12s t-val 16s | *Labroides dimidiatus* | YES | complete 12s t-val 16s | ID complete | - | |
| *Myripristis* | YES | Complete mitogenome | *Myripristis jacobus* | NO | No reference | unable to identify | - | |
| | | | *Myripristis murdjan* | NO | partial 16s | other species identified | *Myripristis berndti* | Complete mitogenome |
| *Odonus* | YES | Complete mitogenome | *Odonus niger* | YES | Complete mitogenome | ID complete | - | |
| *Plectorhinchus* | YES | Complete mitogenome | *Plectorhinchus obscurus* | NO | No reference | unable to identify | - | |
| *Siganus* | NO | Complete mitogenome | *Siganus vulpinus* | NO | Complete mitogenome | Not detected | - | |
| *Sphyrna* | YES | Complete mitogenome | *Sphyrna lewini* | YES | Complete mitogenome | ID complete | - | |
| *Stegostoma* | YES | Complete mitogenome | *Stegostoma fasciatum* | YES | Complete mitogenome | ID complete | - | |
| *Sufflamen* | YES | partial 16s | *Sufflamen chrysopterum* | YES | partial 16s | ID based on incomplete reference | - | |