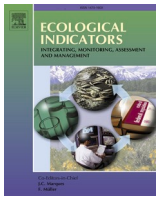
Ecological Indicators 150 (2023) 110207

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Original Articles

DNA metabarcoding on repeat: Sequencing data of marine macrobenthos are reproducible and robust across labs and protocols

Laure Van den Bulcke [a,f,*], Annelies De Backer [a], Jan Wittoeck [a], Kevin Beentjes [d], Sara Maes [a], Magdalini Christodoulou [b,h], Pedro Martinez Arbizu [b], Rumakanta Sapkota [e], Berry Van der Hoorn [d,g], Anne Winding [e], Kris Hostens [a], Sofie Derycke [a,c]

[a] Flanders Research Institute for Agriculture, Fisheries and Food – Marine Research, Jacobsenstraat 1, 8400 Oostende, Belgium
[b] Senckenberg am Meer, German Centre for Marine Biodiversity Research, Südstrand 44, 26382 Wilhelmshaven, Germany
[c] University of Ghent, Department of Biology - Marine Biology Research Group, Krijgslaan 281, 9000 Gent, Belgium
[d] Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands
[e] Aarhus University, Department of Environmental Science, Frederiksborgvej 399, 4000 Roskilde, Denmark
[f] University of Ghent, Department of Data Analysis and Mathematical Modelling - Knowledge-based Systems Research Group, Coupure Links 653, 9000 Gent, Belgium
[g] Inholland University of Applied Sciences, Landscape and Environment Management, Rotterdamseweg 141, 2628 AL Delf, The Netherlands
[h] OÖ Landes-Kultur GmbH, Biologiezentrum, Johann-Wilhelm-Klein-Straße 73, 4040 Linz, Austria

ARTICLE INFO

Keywords:
Marine environmental monitoring
North Sea
COI
Standardized operational protocol
Ring test
Standard operating procedures (SOPS)

ABSTRACT

DNA metabarcoding can be used in marine environmental monitoring if results are reproducible between labs and robust against modifications to the lab protocol. In this interlaboratory study, we conducted a ring test where subsamples of blended macrobenthos samples were distributed to four laboratories located in Belgium, the Netherlands, Germany and Denmark. Samples were processed by a standardized lab protocol and by an adapted protocol, and the resulting datasets were analyzed with the same bioinformatics pipeline. Different biodiversity indicators were calculated. Our results show that bulkDNA metabarcoding of marine macrobenthos offers a highly reproducible assessment of alpha diversity patterns when using a standardized protocol, since comparable species numbers, Shannon indices and Inverse Simpson indices were found between laboratories. Especially high abundant species and species with large body sizes were shared between the laboratories. The need for using a standardized protocol to enhance comparability in alpha diversity between different studies was shown. Beta diversity patterns are less subjected to changes in the metabarcoding protocol and were almost identical between different laboratories, as the main clustering was always based on the macrobenthic community, independent of the used protocol or the laboratory that conducted the work. We conclude that DNA metabarcoding for marine environmental monitoring is an appropriate method when the aim is to study changes in community patterns and advocate its implementation in routine monitoring programs of national and European authorities, providing that a standardized protocol is implemented and/or a detailed description of the protocol is available.

1. Introduction

Biodiversity of marine and coastal environments is under pressure due to different stressors like human activities, pollution and climate change. At the same time, these environments deliver many ecosystem services to society (Daily et al., 2009; Duncan et al., 2015). The European Marine Strategy Framework Directive (MSFD) (2008/56/EC), aiming to achieve good environmental status of marine waters, and other monitoring programs like Biodiversity strategy 2030 have been adopted to safeguard the marine environment. These monitoring programs assess the health status of marine environments using multiple indices, which translate complex ecological information into a numerical value that can be easily interpreted by governments and other stakeholders (Aubry and Elliott, 2006). Often multiple indices are used in monitoring studies, each focusing on a slightly different aspect of biodiversity (Purvis and Hector, 2000). Next to the total number of species present in a sample (= species richness), the Shannon (Shannon, 1948) and inverse Simpson (Simpson, 1949) indices also take species abundance into account (evenness). Both indices are widely used in ecological studies, where the Shannon index reflects species richness of a

* Corresponding author.

https://doi.org/10.1016/j.ecolind.2023.110207
Received 6 February 2023; Received in revised form 28 March 2023; Accepted 30 March 2023
Available online 9 April 2023

sample while the inverse Simpson index additionally places a greater weight on dominant species.

Traditionally, macrobenthic diversity indices rely on species identification based on morphological characteristics, but methodological advancements in DNA metabarcoding and the possibility of high throughput processing of samples have triggered interest in using DNA metabarcoding for marine environmental monitoring (Hering et al., 2018). In the last decade, many studies have investigated the effects of specific steps of the metabarcoding workflow on species detection, for example sampling strategy (Elbrecht and Leese, 2017), the DNA source (Derycke et al., 2021), the DNA extraction kit (Vasselon et al., 2017), preservation of DNA (Yoder et al., 2006), number of technical replicates (Feinstein et al., 2009; Lanzén et al., 2017; Van den Bulcke et al., 2021), primer choice (Braukmann et al., 2019; Derycke et al., 2021; Elbrecht and Leese, 2017; Lobo et al., 2017) and the bioinformatics pipeline (Brannock and Halanych, 2015; Pawlowski et al., 2018b). Yet, a high variation of these methodological steps exists between different studies for marine macrobenthos (van der Loos and Nijland, 2021). As each step may introduce variability and bias in the output, the need for standardization of DNA metabarcoding is high, and is currently perceived as an important drawback to implement metabarcoding for regulatory monitoring (Darling et al., 2017; Goodwin et al., 2017). Therefore, empirical studies are needed to investigate whether DNA metabarcoding data produce reproducible and robust biodiversity results (Darling et al., 2017; Goodwin et al., 2017; Hering et al., 2018; Pawlowski et al., 2018a; Zinger et al., 2019).

In this study, we designed an interlaboratory test using 12 field samples from four well-known macrobenthic communities in the Belgian Part of the North Sea (BPNS) with high (*Abra alba* community), medium (*Hesionura elongate* community) or low (*Macoma balthica* community) diversity (Breine et al., 2018). Sieved macrobenthos specimens were mixed and subsamples from the bulk soup were distributed to four different laboratories across Europe. The samples were processed by a standardized lab protocol to assess how reproducible metabarcoding results are between different laboratories. In addition, three laboratories processed the samples with their own lab protocol to assess how robust metabarcoding results are when changes in the laboratory workflow are included. The resulting seven datasets were bioinformatically processed using the same pipeline, and alpha and beta diversity patterns were compared. First, we investigated whether metabarcoding of bulk samples shows high reproducibility when processing samples in different laboratories using the same fixed lab protocol. Second, when changes were made to this lab protocol, we investigated whether comparable patterns in alpha and beta diversity between the different lab protocols were observed. This assessment of reproducibility and robustness of metabarcoding results is pivotal to evaluate whether metabarcoding of bulk samples is a reliable method for regulatory environmental monitoring.

2. Material and methods

2.1. Sample collection

Four sampling locations in the BPNS were selected, covering macrobenthic communities with low, medium (two locations) and high diversity (Breine et al., 2018) (ESM Fig. 1). These samples have been used in previous studies to optimize the DNA metabarcoding protocol (Derycke et al., 2021; Van den Bulcke et al., 2021). In short, the low diversity community in location ZVL with around six macrobenthic species per sample is dominated by *Macoma balthica* (Bivalvia, Tellinidae), occurring in fine muddy sediment. The medium diverse communities in locations 840 and 330 are defined by *Hesionura elongata* (Polychaeta, Phyllodocidae) with around 14 species per sample, typical for an offshore coarse sandy habitat. The highly diverse community in location 120 with around 26 macrobenthic species per sample is dominated by *Abra alba* (Bivalvia, Semelidae) and is characterized by coastal fine muddy sand. In each location, samples were taken in triplicate with Van Veen grabs (biological replicates A, B and C). After sieving the sediment on a 1 mm sieve, the remaining animals were fixed in 100% ethanol and stored at −20 °C until further processing.

2.2. Sample processing and morphological identification

The samples were further processed following the protocol by Aygalas (2016). In short, specimens were recovered from the samples by the decanting process using a 1 mm sieve and tap water (varying from six to 13 times) and were stored in ethanol. After screening the remaining material (e.g. shells), the heavier specimens that were not decanted properly, were added to the decanted material in ethanol. To compare with traditional morphological species identification, one replicate from each location (120-B, 840-C, 330-C, ZVL-A) was identified under a stereomicroscope up to species level, except for juveniles, which were identified up to genus level and specimens belonging to Nemertea, Anthozoa and Oligochaeta, which were identified up to phylum, class and order level, respectively. The collected specimens in ethanol were mixed with a blender or with a mortar and pestle for samples with less than 100 mL volume to obtain a homogenous bulk sample.

2.3. Experimental set-up and library preparation

These blended bulk samples were used to test the reproducibility and the robustness of the metabarcoding protocol in a ringtest involving four laboratories in Europe: Senckenberg am Meer (SGN), Naturalis, Aarhus University and Flanders research institute for agriculture, fisheries and food (ILVO).

2.3.1. Reproducibility test

For the reproducibility test, the four laboratories received three 2 mL subsamples of each biological replicate (n = 12), except the laboratories Naturalis and Aarhus University because bulk samples were limited for 840-C and ZVL-A: Aarhus University received two 2 mL and one 1.2 mL subsamples from the bulk soup of 840-C, and both institutes received two (Naturalis) or three (Aarhus University) DNA extracts taken for a previous study (Derycke et al., 2021) of the location ZVL-A. A detailed table with the sent volumes for each sample can be found in ESM Table 1. Within the framework of the Interreg North Sea Region project GEANS (https://www.geans.eu/) a laboratory protocol was developed for metabarcoding of soft sediment macrobenthos of the North Sea (GEANS, 2021). This protocol was used for processing the samples by all four institutes. In short, three DNA extractions (using 1 * 2 mL per extraction) were performed for all 12 samples (three biological replicates of four locations). The resulting DNA extracts were pooled per biological replicate and PCR amplified with primers that amplify 313 bp of the mitochondrial COI barcode region (Leray et al., 2013) in triplicate. These three PCR products were then pooled per sample (75 μL in total), cleaned and used for the index PCR using the Nextera kit set A (Illumina). The libraries were sequenced in-house by SGN and Aarhus University, while the other two laboratories sent the libraries for sequencing to different sequencing facilities (Admera Heath Biopharma Services, BaseClear BV). Three samples of SGN (330B, 840B and 120A) showed very low read numbers, so these samples were sequenced a second time in a separate run. The reproducibility test thus resulted in four datasets, one for each laboratory, and each dataset consisted of 12 samples (three biological replicates of four locations) (Fig. 1).

2.3.2. Robustness test

To assess how robust metabarcoding results are, three laboratories (Aarhus University, Naturalis and SGN) applied their own library preparation protocols using the DNA extracts of the reproducibility test (n = 12), except for SGN, as they worked with another DNA extraction kit in their own protocol. In that protocol, 1 mL (instead of 2 mL)

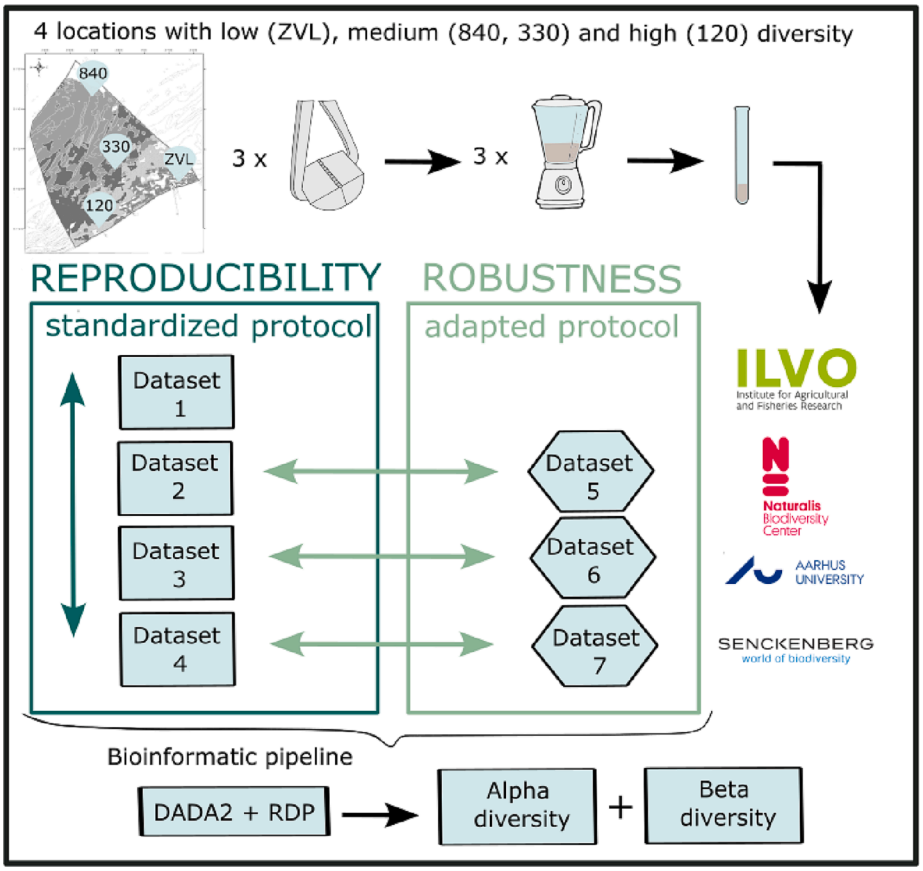


Fig. 1. Visual representation of the experimental design. Subsamples were taken from the mixed 'bulk soup' of the three biological replicates from four macrobenthic communities and distributed to different laboratories. They processed the 12 samples with a standardized protocol for the reproducibility test (left) and with their own protocol for the robustness test (right). The resulting sequencing output was processed by the same bioinformatic pipeline and used for the calculation of alpha diversity (species richness, Shannon index and Inverse Simpson index) and beta diversity.

subsamples were used, therefore, three extra subsamples of 1 mL per sample were sent to SGN, except for the samples 840-C and ZVL-A (no more blended bulk sample available). A detailed table with the sent volumes for each sample can be found in ESM Table 1. The SGN laboratory exerted the following changes: 1 mL subsample, 2 μL DNA and the EZNA Mollusc Kit (Omega Bio-tek) for DNA extraction, Phusion Green Hot Start II High-Fidelity PCR Master Mix (ThermoFisher Scientific) for the first PCR mix and the ExoSap-IT PCR product clean-up reagents (ThermoFisher)) (Table 1). The Aarhus University laboratory performed only one DNA extraction per sample instead of three and added 2 μL DNA to the first PCR mix with a PCRBIO HiFi polymerase (PCR Biosystems), while the Naturalis laboratory processed the samples with Phire Hot Start II DNA Polymerase (Thermo Fisher) with adapted PCR conditions, using 5 μL DNA template and NucleoMag NGS-Beads clean-up reagents (Macherey-Nagel) (Table 1). Except for these changes, the samples were processed as explained above, resulting in three datasets of 12 samples each (Fig. 1).

2.4. Bioinformatic processing

Bioinformatic processing was done in R (Core Team, 2020) v4.0.2. The detailed code to reconstruct these results can be found on https://gitlab.com/lvandenbulcke1/ringtest-geans-for-testing-repeatability-and-robustness-metabarcoding-data/-/tree/main.

2.4.1. Processing of raw reads

For each of the seven datasets (four for the reproducibility test and three for the robustness test, each with 12 samples), the quality of demultiplexed reads was checked with MultiQC (Ewels et al., 2016), and forward and reverse primers were removed using Trimmomatic (Bolger et al., 2014). Amplicon sequence variants (ASVs) were generated using the Dada2 pipeline in the Dada2 v1.17.0 package (Callahan et al., 2016). As the datasets were generated on different sequencing runs, the sample inference script of the Dada2 pipeline was run on separate datasets to take into account the different error profiles. Reads were further trimmed by removing parts with a quality score lower than 30. Unique reads were determined and merged for each sample. For the reproducibility test, the sequence tables of the different laboratories obtained with the fixed lab protocol were combined with the mergeSequenceTables function in the Dada2 package. For the robustness test, the resulting sequence tables for each of the lab protocols were added. After merging the datasets for the reproducibility and robustness test, chimeras were removed with the removeBimeraDenovo function. The total numbers of reads were compared between the different laboratories for the reproducibility and the robustness test and visualized by barplots in R. Taxonomy was assigned with the assignTaxonomy function in the Dada2 package, based on the Ribosomal Database Project (RDP) Classifier (Wang et al., 2007). Standard settings were employed, except for the minimum bootstrap confidence parameter, which was set to 80. Within the GEANS project, a reference database of marine invertebrates in the North Sea is being constructed, composed of in-house and public COI sequences of macrobenthos from multiple monitoring campaigns in the North Sea (GEANS, 2021). A preliminary version of this reference database was used, containing 1992 COI sequences from 565 species. The dataset of Naturalis in the reproducibility test had much higher read numbers and allowed to investigate the effect of higher sequencing depth on the detected species; an UpSet plot with the non-rarified datasets of the reproducibility test was made and the species uniquely detected in this dataset were studied in detail.

2.4.2. Alpha diversity analyses

All samples from the seven datasets were rarified at 30 000 reads to take into account the different sequencing depths. This number was a tradeoff between reaching the plateau of the rarefaction curves and

Table 1
Differences between lab protocols for the robustness test compared to the standardized GEANS protocol used in the reproducibility test.

		Standardized GEANS protocol (reproducibility test)	Adapted protocol Senckenberg am Meer (SGN)	Adapted protocol Naturalis	Adapted protocol Aarhus University
DNA extraction	**DNA extraction kit**	DNeasy Powersoil kit	EZNA Mollusc Kit	DNeasy Powersoil kit	DNeasy Powersoil kit
	Technical replicates for DNA extraction	3 × 2 mL	3 × 1 mL	3 × 2 mL	1 × 2 mL
PCR amplification	**DNA polymerase**	2× KAPA HiFi HotStart ReadyMix	Phusion Green Hot Start II High-Fidelity Master Mix	Phire Hot Start II DNA Polymerase	PCRBIO HIFI polymerase
	PCR cycling conditions:	95 °C 3 min, 35× (98 °C 30 s, 57 °C 30 s, 72 °C 30 s), 72 °C 1 min	98 °C 2 min, 30× (98 °C 15 s, 52 °C 30 s, 72 °C 30 s), 72 °C 3 min	95 °C 3 min, 35× (98 °C 30 s, 57 °C 30 s, 72 °C 30 s), 72 °C 1 min	95 °C 3 min, 35× (98 °C 30 s, 57 °C 30 s, 72 °C 30 s), 72 °C 1 min
	Volume of template DNA in 25 µL PCR reaction	2.5 µL of DNA	2.0 µL of DNA	5.0 µL of DNA	2.0 µL of DNA
	Clean-up PCR products	CleanNGS beads	ExoSAP-it PCR1 products	Macherey-Nagel NucleoMag	CleanNGS beads
Sequencing	**Sequencing method**	Illumina MiSeq (2 × 300 paired-end for ILVO, Naturalis and SGN, 2 × 250 for Aarhus University)	Illumina MiSeq (2 × 250 paired-end)	Illumina MiSeq (2 × 300 paired-end)	Illumina MiSeq (2 × 250 paired-end)
	Added percentage of PhiX	20%	20%	According to sequencing company	21%

removing a minimum number of samples. Four samples were removed in total: 840-C-Aarhus-reproducibility, 330-C-Aarhus-reproducibility, ZVL-A-Aarhus-reproducibility and ZVL-A-ILVO-reproducibility. After rarefaction, only ASVs with a taxonomic assignment were taken into account and the assigned species and read numbers were used for downstream analyses.

For both tests (reproducibility and robustness), the total numbers of macrobenthic species (species richness), the Shannon index and the inverse Simpson index were determined for each sample of the different laboratories and/or protocols. Before calculating the Shannon index and the inverse Simpson index, a square root transformation was performed to account for the effect of high read numbers for some species. The diversity indices were calculated with the function diversity from the vegan package v2.5.7 (Dixon, 2003) and visualized in a barplot. To detect whether significant differences existed between the laboratories (the reproducibility test) or between the used protocol (the robustness test), two-way ANOVA tests were performed. In the reproducibility test, the laboratory effect was investigated, so main factors location (levels: 120, 840, 330 and ZVL) and laboratory (levels: ILVO, Naturalis, SGN and Aarhus University) and their interaction were tested. For the robustness test, the datasets from the robustness test (own protocol) were compared with those of the reproducibility test (fixed protocol) for each laboratory separately. Therefore, two-way ANOVA tests per laboratory with main factors location (levels: 120, 840, 330 and ZVL) and protocol (levels: fixed and own) and their interaction were performed. ANOVA assumptions were checked by plotting the residuals to investigate the homogeneity of variances and the normality of the data and, if significant effects were observed in the ANOVA, pairwise comparisons were performed using the package lsmeans v2.30-0 and displayed with the function compact letter display (cld). For both tests (reproducibility and robustness), the number of detected species was studied in more detail: the shared and unique species between laboratories (reproducibility) or between both protocols (robustness) were listed and visualized in barplots (reproducibility) with the ggplot package v3.4.0 or in VennDiagrams with the VennDiagram package v1.7.3 (robustness). Possible explanations for the observed patterns were investigated by dividing the species in different classes of body size (<10 mm, 11–20 mm, 21–100 mm, 101–200 mm, 201–500 mm, >500 mm) and listing the number of reads, phylum and, if present in the morphological identified sample, also the abundance (counted number of individuals/species/sample).

Next, for the reproducibility test where a fixed protocol was used, also intraclass correlation coefficients (ICC) were calculated to test if the results analyzed by the different laboratories were correlated. This test was performed because a non-significant ANOVA result does not say anything on how similar the results between different laboratories are. The ICC estimated the reliability between measurements of different raters, here laboratories, for the average calculated diversity index values of the biological replicates in each location. The icc function of the irr package was used, with the following three parameters: 1) “two-way model”, as the same set of samples was identified by all laboratories, 2) “single”, because we would like to use the measurements from a single rater in the future and 3) “absolute agreement”, because the absolute numbers between raters are compared (instead of the relative ratio). Based on the calculated value and the 95% confident interval (CI), the agreement between laboratories can be poor (<0.50), moderate (0.50–0.75), good (0.75–0.90) or excellent (>0.90) (Koo and Li, 2016).

Finally, the morphological and the seven metabarcoding datasets were compared using only the four morphological identified samples (120-B, 840-C, ZVL-A, 330-C) to avoid variation in detected species due to biological replicates. Also for these datasets, the different indices (species richness, Shannon index and Inverse Simpson index) were calculated and visualized in barplots. The shared and unique species between the morphology, the reproducibility and the robustness datasets were visualized in an UpSet plot with the UpSetR package v1.4.0. For the shared or unique species the taxonomic classification, the body size class (<10 mm, 11–20 mm, 21–100 mm, 101–200 mm, 201–500

mm, >500 mm), the abundance in the morphological and genetic (both reproducibility and robustness test) datasets and the availability of a reference sequence in the used reference database was listed.

2.4.3. Beta diversity analyses

To investigate variability in community composition between the different laboratories (reproducibility test) and between the fixed or own protocol (robustness test), non-metric multidimensional scaling (NMDS) plots based on the Bray-Curtis (Edward, 1984) dissimilarity index were constructed, using the R package vegan v2.5.7. A square root transformation was performed on the community data matrix prior to calculate the Bray-Curtis dissimilarity index. To compare the species communities between the laboratories (reproducibility test), a two-way PERMANOVA was conducted, consisting of two main effects location (levels: 120, 840, 330 and ZVL) and laboratory (levels: ILVO, Naturalis, Aarhus University and SGN) and their interaction, performed with 9999 permutations. To compare the species communities between the two protocols (fixed versus own, robustness test), two-way PERMANOVAs (one for each laboratory) were conducted, with location (levels: 120, 840, 330 and ZVL) and protocol (levels: fixed and own) as main effects and the interaction term location*protocol. A distance dispersion test and permutation test were used to test the homogeneity of dispersion in the samples with the R package vegan v2.5.7.

3. Results

3.1. Processing of raw reads

The number of reads after each filtering step for the different samples can be found in ESM Table 2. After processing and filtering the datasets for the reproducibility and robustness tests, mean read numbers differed per laboratory and/or test: 190 745 and 119 823 for the SGN, 59 178 and 326 783 for Aarhus University, 694 250 and 143 737 for Naturalis for the reproducibility and robustness test respectively and 162 049 for ILVO (only reproducibility test) (ESM Fig. 2; ESM Table 2). Only 25%, 24%, 23%, 23%, 22% and 22% of the ASVs were assigned to phylum, class, order, family, genus and species level, respectively. However, the assigned ASVs at species level were represented by 87% of the total number of reads. The reproducibility dataset of Naturalis had a much higher sequencing depth than the other datasets, yet, the detected number of species was comparable (86 for Naturalis versus 77, 87, 70 for ILVO, SGN, and Aarhus University, respectively) and only five extra species were uniquely found (ESM Fig. 3A). Together with the rarefaction curves (ESM Fig. 3B), this illustrates that the sequencing depth per sample was sufficient to capture macrobenthos diversity in all datasets.

3.2. Alpha diversity

3.2.1. Comparison between metabarcoding and morphological identification

In each location, one biological replicate (120-B, 330-C, 840-C, ZVL-A) was identified morphologically up to species level, resulting in 57 species in total identified with the traditional method (ESM Table 3). More in detail, 39, 13, 10 and 3 species were identified in 120-B, 330-C, 840-C, ZVL-A, respectively. In these morphological identified samples, 63 species were identified with metabarcoding, of which 57 species were found with the fixed protocol of the reproducibility test (ranging from 29 to 51 for the separate laboratories), and 54 species with the adapted protocol of the robustness test (ranging from 38 to 49 for the separate laboratories). Of the 63 species identified by metabarcoding, only 33 were also found by morphological identification (ESM Fig. 4A). Even after processing the samples with different protocols by multiple laboratories, 24 morphological identified species were never picked up by metabarcoding. Despite great effort in constructing a complete reference database, 17 of these 24 species were not present in the used reference database, but COI sequences were found for 11/17 species in GenBank. All other undetected species (7/24) had low abundances (≤5 individuals), of which four species were only represented by one individual. Not only small species were missed (size classes varied from sr < 10 to sr101-200), but most missed species (6/7) belonged to the phylum "Polychaeta", known for low primer efficiency (Carr et al., 2011). Detailed information can be found in ESM Table 4. The metabarcoding method identified 30 extra species, with 18 of them having low read numbers (≤100 reads). Despite the differences in detected species, all datasets showed decreasing index values from the location with high diversity to the location with low diversity for the three indices (species richness and Shannon and Inverse Simpson after square root transformation of the read numbers), although lower Inverse Simpson values were obtained for the genetic datasets (ESM Fig. 4B).

3.2.2. Reproducibility test

For species richness, the Shannon index and the Inverse Simpson index, the two-way ANOVAs with main factors location and laboratory showed no significant effects of the interaction term location*laboratory, and no significant differences between laboratories, while a significant effect of the main factor location was observed (Table 2; Fig. 2). For all three diversity indices, the pairwise posthoc tests showed significant differences between the different locations, except between location 330 and 840, both described as locations with medium diversity (ESM Table 5, Fig. 2). In line with the non-significant effect of the main factor laboratory, the ICCs (absolute agreement, two-way random effect model and single rater) showed excellent (ICC = 0.956, CI [0.820, 0.997]), good (ICC = 0.889, CI [0.613, 0.992]) and excellent (ICC = 0.98, CI [0.911, 0.999]) agreement for the species richness and the Inverse Simpson and the Shannon index after a square root transformation on the read numbers, respectively, illustrating the high reproducibility of the indices between different laboratories.

In total, the four metabarcoding datasets generated for the reproducibility test detected 96 species. Of these 96 species, 51 were consistently found by all four laboratories. These species have high read abundance (>100 reads, 46/51 species) and/or a large body size (>20 mm, 36/51 species), expect *for Abra alba* and *Spio decorata* which were small (size class: 11–20 mm) and had very low read numbers (93 and 30 reads summed over all samples) (ESM Table 5). Only 22 species of the 96 species were found by one laboratory exclusively (ESM Fig. 5), which were typically species with low read abundance (<100 reads, 20/22 species) (Fig. 3) and none of these species were detected by the morphological analyses (ESM Table 5).

3.2.3. Robustness test

The robustness test compared the datasets obtained by using the fixed protocol and the own protocol for each laboratory. The ANOVA results per laboratory for the three diversity indices (species richness, Shannon index, Inverse Simpson index) were equivocal, and therefore are reported per diversity index separately. First, for species number of laboratories SGN and Aarhus University, no significant interaction or 'protocol' effect was detected, while for the factor 'location' significant differences in species number were observed (Table 3). For Aarhus University, the posthoc test showed significant differences between the locations, except between 330 and 840, the two locations with a medium diversity, while for SGN, only a distinction could be made between ZVL and the other locations (ESM Table 6, Fig. 4). For Naturalis, on the other hand, the interaction term 'location*protocol' was significant (Table 3), indicating that the number of detected species depended on the combined effect of protocol and location. Only in the location with high diversity (120), the used protocol impacted the detected number of species, with significantly higher numbers obtained using the fixed GEANS protocol (ESM Table 6, Fig. 4). Second, for the Shannon index, the interaction effect was not significant for any of the laboratories. The factor 'protocol' significantly affected the Shannon index for two laboratories (Table 3), but the own protocol resulted in lower Shannon indices for Naturalis and in higher Shannon values for SGN compared

Table 2
Output of the different ANOVAs for the reproducibility test, one for each diversity estimates (Species richness, Inverse Simpson index and Shannon index). Inverse Simpson and Shannon indices were calculated after a square root transformation on the read numbers.

		df	sum sq	mean sq	F value	pr(<F)
Species Richness	laboratory	3	21.6	7.21	0.3425	0.7947
	location	3	6505.5	2168.49	103.0278	**3.10E-15**
	laboratory:location	9	237.3	26.37	1.2527	0.3048
	residuals	28	589.3	21.05		
InvSimpson_Sqrt	laboratory	3	12.91	4.305	1.5409	0.2258
	location	3	371.54	123.845	44.3302	**9.192E-11**
	laboratory:location	9	29.30	3.255	1.1653	0.3538
	residuals	28	78.22	2.794		
Shannon_Sqrt	laboratory	3	0.1331	0.0444	0.4349	0.7297
	location	3	28.1163	9.3721	91.8726	**1.33E-14**
	laboratory:location	9	0.4636	0.0515	0.5049	0.8584
	residuals	28	2.8563	0.1020		

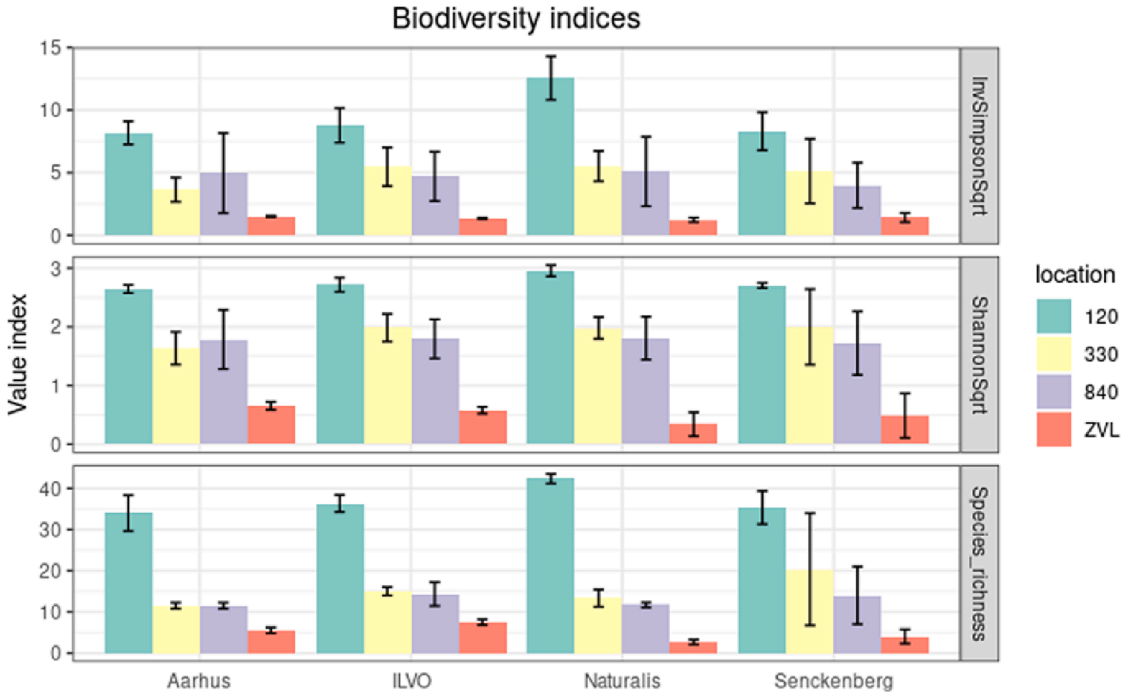


Fig. 2. Mean numbers and standard deviations of the different diversity estimates (Inverse Simpson, Shannon index and Species richness) in each location (120, 330, 840 and ZVL). The Inverse Simpson and Shannon indices were calculated after square root transformation on the read numbers, indicated by "Sqrt". Mean numbers were calculated with three biological replicates, except when samples were removed at the rarifying step, so for 330-Aarhus, 840-Aarhus, ZVL-Aarhus, ZVL-ILVO, only two biological replicates were used.

with the fixed GEANS protocol (ESM Table 6, Fig. 4). The factor 'location' was significant for all laboratories (Table 3), but for Naturalis and Aarhus University, pairwise posthoc tests showed significant differences between the locations, except between 330 and 840, the two locations with a medium diversity, while for SGN, only a distinction could be made between ZVL and the other locations (ESM Table 6, Fig. 4). Last, for the Inverse Simpson index, a significant interaction term was detected in the ANOVA for Naturalis (Table 3). Similar as for the species richness, a significantly higher Inverse Simpson index was observed only in the location with high diversity (120) using the fixed GEANS protocol (ESM Table 6, Fig. 4). The factor 'location' significantly affected the inverse Simpson index for Aarhus University and SGN (Table 3). Post hoc test showed significant differences between ZVL and 840 for Aarhus University, while only differences were seen between ZVL and 120 for SGN (ESM Table 6, Fig. 4). For SGN, also a significant effect of the main factor 'protocol' was observed (Table 3), with significantly higher numbers when using the adapted protocol (ESM Table 6, Fig. 4).

More than half of the species were shared between the fixed GEANS and own protocols for each institute (59%, 59% and 56% for resp. Aarhus University, Naturalis and SGN) (ESM Fig. 6). High read abundance (>100 reads, summed over the different samples) was seen for these species (54/71 species), in contrast to species uniquely found by one method (10/49 species) (ESM Table 7). The fixed GEANS protocol and the own protocol roughly shared the same percentage of species with the morphological identification (31.7% versus 32.7% for Aarhus University, 36% versus 33% for Naturalis and 32% versus 29% for SGN) (ESM Fig. 6).

3.3. Beta diversity

3.3.1. Reproducibility test

Beta diversity patterns based on the Bray-Curtis dissimilarity index

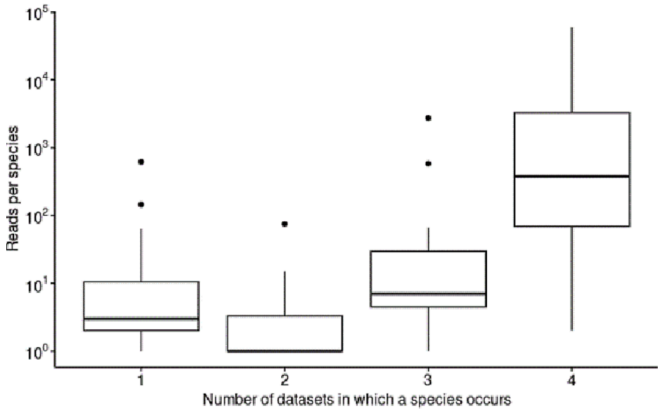


Fig. 3. Read abundance per species for species that occur in one, two, three or to four datasets from the reproducibility test (ILVO, Naturalis, SGN and Aarhus).

were comparable between the four laboratories, as the NMDS plot clearly showed clustering based on the macrobenthic communities, independent of the laboratory that conducted the work (Fig. 5A). This was corroborated by the PERMANOVA results that showed no significant effect of the interaction term laboratory*location ($F = 0.2465$, $p_{Permanova} = 1$) nor of the main factor laboratory ($F = 0.841$, $p_{Permanova} = 0.612$), while for the factor location a significant effect was detected ($F = 31.500$, $p_{Permanova} = 0.0001$). The location explained 64% of the observed variation, while the different laboratories only accounted for 3% of the variation. The residuals –here the biological replicates– and the interaction effect explained the remainder of the variation, resp. 32% and 1% (ESM Table 8).

3.3.2. Robustness test

The effect of the used lab protocol (fixed versus own) on the beta diversity pattern using Bray-Curtis dissimilarity was investigated for the three laboratories separately. For both Aarhus University and Naturalis, neither interaction term 'location*protocol' nor factor 'protocol' was significant (Aarhus University: $F = 0.5325$, $p_{Permanova} = 0.9096$ & $F = 0.2937$, $p_{Permanova} = 0.9272$ and Naturalis: $F = 0.5469$, $p_{Permanova} = 0.8938$ & $F = 1.3943$, $p_{Permanova} = 0.2391$, ESM Table 8). The main factor 'location', however, was significant (Aarhus University: $F = 18.1681$, $p_{Permanova} = 0.0001$; Naturalis: $F = 0.5469$, $p_{Permanova} = 0.0001$) and accounted for 79% of the variation in both laboratories (ESM Table 8). This was confirmed in the NMDS plots of Aarhus university and Naturalis where samples were also mainly discriminated based on location, with the two medium diversity locations (330 and 840) clustering closer together (Fig. 5C-D). Within each cluster, no distinction between both protocols (fixed versus own) could be discerned (Fig. 5C-D). For SGN, slightly different results were observed. Again, clustering in the NMDS plot was mainly based on the different locations of the samples, however, different subclusters can be distinguished according to the protocol used (Fig. 5B). Permanova

Table 3
Output of the different ANOVAs for the robustness test, one for each diversity estimates (Species richness, Inverse Simpson index and Shannon index) and for each laboratory (Aarhus University, SGN, Naturalis) separately. Inverse Simpson and Shannon indices were calculated after a square root transformation on the read numbers.

			df	sum sq	mean sq	F value	pr(<F)
Species richness	Aarhus	protocol	1	11.57	11.57	2.2508	0.1574
		location	3	2776.62	925.54	180.0299	**7.78E-11**
		protocol*location	3	4.22	1.41	0.2735	0.8435
		residuals	13	66.83	5.14		
	Naturalis	protocol	1	192.7	192.67	60.842	**7.69E-07**
		location	3	3169.7	1056.56	333.649	**1.25E-14**
		protocol*location	3	292.3	97.44	30.772	**7.05E-07**
		residuals	16	50.7	3.17		
	Senckenberg	protocol	1	82.26	82.26	1.7975	0.201368
		location	3	1945.82	648.61	14.1735	**0.000159**
		protocol*location	3	110.53	36.84	0.8051	0.511683
		residuals	14	640.67	45.76		
InvSimpson	Aarhus	protocol	1	0.359	0.359	0.1654	0.6908
		location	3	158.917	52.972	24.3877	**1.298E-05**
		protocol*location	3	2.986	0.995	0.4582	0.7161
		residuals	13	28.237	2.172		
	Naturalis	protocol	1	40.084	40.085	17.5035	**0.000702**
		location	3	183.350	61.117	26.6881	**1.825E-06**
		protocol*location	3	45.423	15.141	6.6116	**0.004092**
		residuals	16	36.641	2.290		
	Senckenberg	protocol	2	58.842	58.842	9.1491	**0.009094**
		location	3	87.389	29.130	4.5293	**0.020281**
		protocol*location	3	14.932	4.977	0.7739	0.527690
		residuals	14	90.040	6.431		
Shannon	Aarhus	protocol	1	0.0113	0.0113	0.1555	0.6997
		location	3	12.0881	4.0294	55.3063	**1.167E-07**
		protocol*location	3	0.0868	0.0289	0.3973	0.7572
		residuals	13	0.9471	0.0729		
	Naturalis	protocol	1	0.9951	0.9951	16.4394	**0.0009203**
		location	3	17.2860	5.7620	95.1854	**2.047E-10**
		protocol*location	3	0.2674	0.0891	1.4726	0.2595925
		residuals	16	0.9685	0.0605		
	Senckenberg	protocol	1	2.6252	2.62525	16.1990	**0.0012536**
		location	3	6.8388	2.27961	14.0662	**0.0001654**
		protocol*location	3	1.5082	0.50273	3.1021	0.0609398
		residuals	14	2.2689	0.16206		

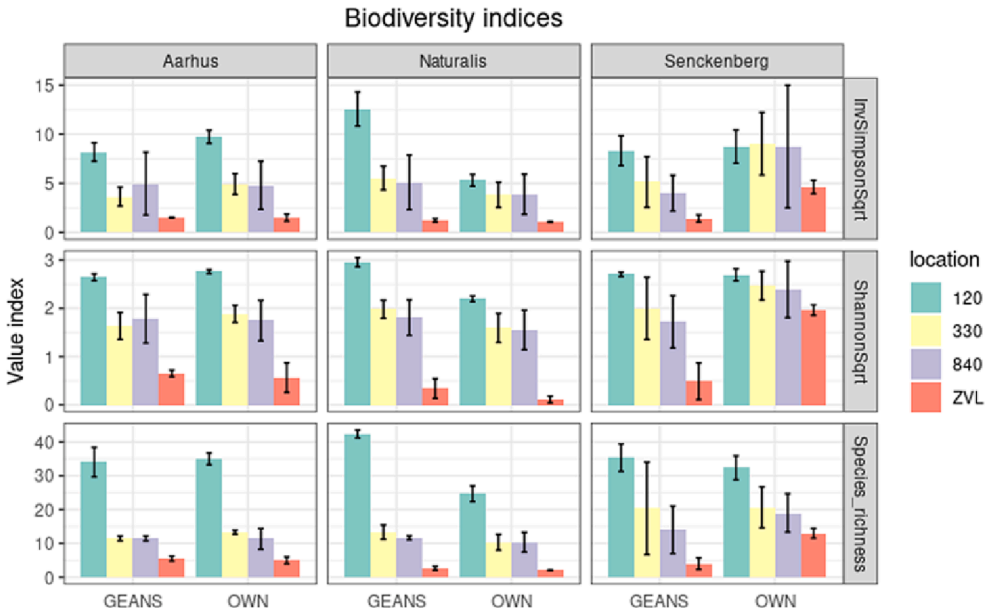


Fig. 4. Mean numbers (of the three biological replicates A, B and C) and standard deviations of the different diversity estimates (Inverse Simpson, Species richness and Shannon index) in each location (120, 330, 840 and ZVL) were calculated in the two methods (GEANS versus OWN protocol) of the robustness test. The Inverse Simpson and Shannon indices were calculated after square root transformation on the read numbers, indicated by adding "Sqrt".

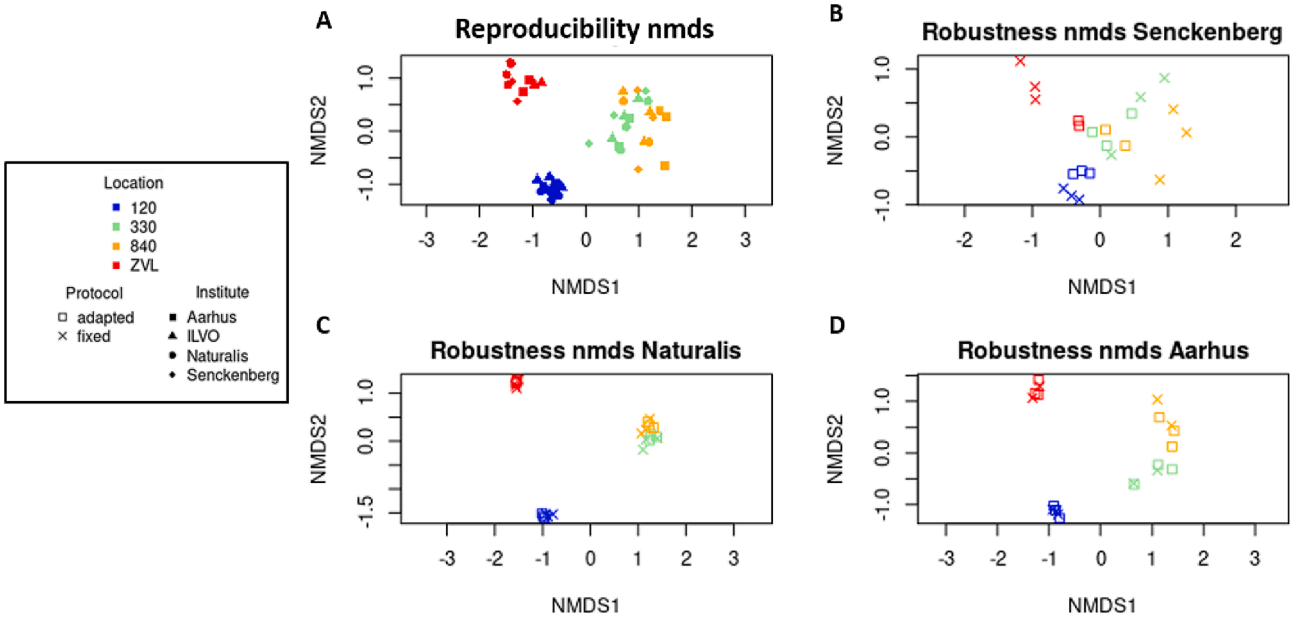


Fig. 5. NMDS plot based on the Bray-Curtis dissimilarity for the reproducibility test (A) and the robustness test, one for each laboratory: Senckenberg (B), Naturalis (C) and Aarhus University (D). The different locations are visualised by different colors (120 = blue, 330 = green, 840 = orange, ZVL = red), while the different institutes (for A: ILVO = triangle, Naturalis = circle, Aarhus = square, Senckenberg = rhombus) and the used protocol (for B, C, D: own = square, fixed = cross) are visualised by different symbols. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

corroborated these visual patterns with significant effects for main factors 'location' ($F = 9.9262$, $p_{Permanova} = 0.0001$) and 'protocol' ($F = 4.1846$, $p_{Permanova} = 0.0018$), resp. explaining 57% and 8% of the variation. The interaction term was not significant ($F = 1.3772$, $p_{Permanova} = 0.1617$) (ESM Table 8).

4. Discussion

The implementation of DNA metabarcoding for routine environmental monitoring of marine diversity has not yet been adhered by national and European authorities. A laboratory protocol was developed for macrobenthos metabarcoding in soft sediments within the framework of the Interreg North Sea Region project GEANS (https://www.geans.eu/) that is currently applied by Belgium, The Netherlands, Germany and Denmark (GEANS, 2021). In this interlaboratory study, laboratories validated the reproducibility and robustness of this protocol. An overview table listing the main results of this study can be found in Table 4. First, the reproducibility test demonstrated that when using a fixed lab protocol, metabarcoding results are comparable, especially for high abundant species and species with a large body size. Second, the robustness of DNA metabarcoding was shown since community patterns were comparable between the standardized and adapted protocol for all

Table 4
Summarizing overview table with main results of the reproducibility and robustness test.

		Equation	Reproducibility test (fixed protocol)	Robustness test		
				Aarhus University	Naturalis	SGN
Alpha diversity	**Species richness**	$S = N$With N the number of species in the sample.	• No effect of laboratory, excellent agreement between labs. • Shared species in all four laboratories have high read abundance and/or large body size, while species unique for one lab have low read abundance and were not detected by the morphological analyses. • Distinction between locations with high (120), medium (840 and 330) and low (ZVL) diversity	• No effect of protocol • Distinction between locations with high (120), medium (330 and 840) and low (ZVL) diversity	No effect of protocol in locations with low and medium diversity, but higher species numbers were found with the GEANS protocol in the location with high diversity.	• No effect of protocol • Distinction between ZVL (low diversity) and the other locations
				Shared species between both protocols have high read abundance, while species unique for one method typically have low read abundance.		
	Shannon index	$H = -\sum_{i=1}^{S} p_i \ln p_i$With S the number of species in the sample, p_i the proportion of individuals found in the ith species (Shannon 1948)	• No effect of laboratory, good agreement between labs. • Distinction between locations with high (120), medium (840 and 330) and low (ZVL) diversity	• No effect of protocol • Distinction between the locations, except between 330 and 840, the two locations with medium diversity	• Higher Shannon indices were obtained with the fixed GEANS protocol Distinction between the locations, except between 330 and 840, the two locations with medium diversity	• Higher Shannon indices were detected with the adapted (OWN) protocol. • Distinction between ZVL (low diversity) and the other locations
	Inverse Simpson index	$InvSimpson = \frac{1}{D_{Simpson}} = \frac{1}{\sum_{i=1}^{S} p_i^2}$With S the number of species in the sample, p_i the proportion of individuals found in the ith species (Simpson 1949)	• No effect of laboratory, excellent agreement between labs. • Distinction between locations with high (120), medium (840 and 330) and low (ZVL) diversity	• No effect of protocol. • Distinction between ZVL (low diversity) and 840 (medium diversity) and 120 (high diversity)	No effect of protocol in locations with low and medium diversity, but higher Inverse Simpson indices were found with the GEANS protocol in the location with high diversity.	• Higher Inverse Simpson indices were detected with the adapted (OWN) protocol • Distinction between ZVL (low diversity) and 120 (high diversity)
Beta diversity with Bray-Curtis		$BC_{ij} = 1 - \frac{2*C_{ij}}{S_i + S_j}$With C_{ij} the sum of only the lesser counts for each species found in both sites, S_i, the total number of specimens counted in site i and S_j the total number of specimens counted in site j.	No effect of laboratory on community pattern.	No effect of protocol on community pattern	No effect of protocol on community pattern	Effect of protocol: Main clustering is based on the locations, but within each location, an effect of the protocol can be seen (subclusters fixed GEANS protocol versus OWN protocol)

three institutes, mainly because highly abundant species were detected by all four protocols. However, an effect was seen on alpha diversity, therefore, a standardized metabarcoding protocol should be used when comparing alpha diversity between studies and/or areas. Based on the results of this study, we suggested a few considerations when implementing DNA metabarcoding for marine environmental monitoring.

4.1. DNA metabarcoding shows high reproducibility of diversity measures between laboratories when using a standardized protocol

This study showed no significant effect of the laboratory that processed the samples on tested diversity indices (species richness, Shannon index and Inverse Simpson index, Table 4). The four reproducibility datasets shared all species with high abundance while most of the species unique to one dataset had low abundance in the metabarcoding datasets (<100 reads). This is in agreement with Buchner et al. (2021) who found high reproducibility, except for rare and small species when repeating the metabarcoding process for freshwater macrobenthos bulk DNA samples with an automated liquid handling machine. Deeper sequencing of samples is suggested to counteract for this loss of rare species (Smith and Peay, 2014). In this study, only five additional species were detected in the non-rarified Naturalis dataset, which showed higher read numbers compared to the other datasets of the reproducibility test. Together with the rarefaction curve (ESM Fig. 3A), this suggests that the datasets in this study had enough reads to detect all species and it is very unlikely that the lower sequencing depth can explain the loss of these low abundant species. Another possible scenario could be the heterogeneity of the samples from which the subsamples were taken, after mixing the bulk samples with a blender for metabarcoding studies (Antich et al., 2021; Aylagas et al., 2018). Although it is assumed that tissue of all species in the sample is present in the subsample (Duarte et al., 2021; van der Loos and Nijland, 2021), studies showed this is not always the case (Lejzerowicz et al., 2014; Van den Bulcke et al., 2021), and, therefore it has been suggested to increase the volume of the subsample or take multiple subsamples for DNA extraction. In the fixed protocol for this study, three subsamples were taken to accommodate for this variability. Comparison between the fixed (with three DNA replicates) and own protocol of Aarhus University (with one replicate) in the robustness test showed no significant differences in the number of detected species. Nevertheless, considering that all unique species have very low read numbers, and as such have lower chance to be present in the subsample, the heterogeneity of the samples is a plausible explanation. Next to these false negatives for some laboratories, also false positives are possible (Yang et al., 2020) and probably, the high species richness in the biological replicate 330B processed by SGN can be explained by false positives. Many species detected in this sample were also detected in 120A, 120B and 120C, in contrast to the other biological replicates of location 330. Furthermore, these species mostly showed low abundance in 330B, but had high read numbers in the replicates of the location 120. Therefore, we suspect tag jumping between the samples should be taken in consideration (Jia et al., 2022). A lab protocol can be adjusted to minimize tag jumping, for example by using all tags only once in each library, taking PCR replicates, minimizing the handlings with tagged amplicons, taking negative controls at each step of the protocol (Schnell et al., 2015) and using a positive control (a sample with known mock community or a spiked sample where DNA sequences with known concentrations were added) (van der Loos and Nijland, 2021), or correcting the possible cross-over in the bioinformatic pipeline using control samples (Beentjes et al., 2019; Davis et al., 2018). Regardless of the possible false positives/negatives, each laboratory was able to distinguish between the different locations, with all three diversity indices. The Shannon and Inverse Simpson indices are based on both the number of species and their abundance. For metabarcoding datasets, relative read abundance can be used as a proxy for species abundances (Cahill et al, 2018). However, often no correlation has been observed between read abundance and the number of specimens in a sample (Hollatz et al., 2017; Lamb et al., 2019; Leray and Knowlton, 2017). In contrast, a (weak) positive correlation has been seen between relative read abundance and biomass (Elbrecht and Leese, 2017; Lamb et al., 2019). Next to the biomass/size of the specimens, also other factors like the presence/absence of a exoskeleton (Derycke et al., 2021) and the PCR/primer bias (Nichols et al., 2018) can affect the relative read abundance. As such, much higher reads for one species than for another species are obtained. With a square root transformation, these differences in high versus low read numbers between species become smaller, resulting in a higher evenness. The Shannon and Inverse Simpson indices are impacted by the species richness and the evenness, so after transformation (resulting in a higher evenness), higher index values were obtained. Without transformation, these indices were only able to distinguish between the locations with extreme diversity (high versus low) (ESM Fig. 7). Next to alpha diversity, each laboratory was also able to detect similar community patterns discriminating the different macrobenthic communities with high, medium or low diversity. In a cross-laboratory experiment on biofouling samples from a broad geographical scale (Australia, Canada, New Zealand and the USA), a similar distinction between communities was observed with DNA metabarcoding, which could be expected since these communities were very different from each other (Zaiko et al., 2021). In this study, we show that, even at smaller regional scale (Belgian part of the North Sea), bulk DNA metabarcoding reflects community composition in a reproducible way.

4.2. Adapting the lab protocol affected alpha diversity, but had no effect on beta diversity patterns

The robustness test compared samples processed in three different laboratories using their own metabarcoding protocol versus a fixed one. Aarhus University reduced the number of DNA replicates to one, changed the polymerase and the DNA template volume, but no significant differences in alpha or beta diversity were seen between both protocols (Table 4). This is in contrast to the recommendation to use multiple replicates for DNA extraction in literature (Lejzerowicz et al., 2014; Van den Bulcke et al., 2021). In the Aarhus protocol, one DNA replicate seemed sufficient to cover the biodiversity displayed by the fixed protocol. Moreover, a cross laboratory study (Zaiko et al., 2021) showed that the DNA template and polymerase can have an impact on the metabarcoding output. Naturalis changed the polymerase, the DNA template volume and the clean-up, resulting in significant lower values of the number of species and the InvSimpson and Shannon indices compared to the GEANS protocol (Table 4). The used PCR clean-up kit is a bead based kit, similar like the kit used by the fixed GEANS protocol. Furthermore, using a higher DNA template volume, the chance of picking up species with low read abundance increases. Therefore, these changes are unlikely to explain the lower number of species detected by the own protocol. Naturalis used a polymerase with a 2× higher fidelity than Taq DNA polymerase versus a 100× higher fidelity for the KAPA HIFI polymerase, meaning the polymerase of Naturalis is the one with the lowest fidelity of our experiment. More errors may lead to inefficient primer binding or inaccurate taxonomic assignment by RDP, which may explain the lower number of species found with the Naturalis own protocol. SGN changed the DNA extraction kit, the volume of the DNA replicates, the DNA polymerase, the PCR cycling conditions, the template DNA volume and the PCR clean-up products. Here, no significant effect on the number of species was seen, but a significantly higher InvSimpson and Shannon index was calculated with their own protocol and also the beta diversity pattern was affected by the protocol (Table 4). As the same template DNA volume and a DNA polymerase with comparable fidelity were used as the Aarhus University lab, where no effect was observed, it is unlikely that these changes explain the differences. Furthermore, limited impact of the used DNA extraction kit is expected, since no difference in operational taxonomic unit richness has been observed between five different DNA extraction methods

(Vasselon et al., 2017). The Shannon and invSimpson are indices taking abundance into account, as is beta diversity, where a clear significant (based on Permanova) sub clustering for the factor protocol was observed in the NMDS plot. Since read numbers are largely affected by PCR biases (Leray and Knowlton, 2017), the observed differences are likely caused by the adapted PCR cycling conditions. Caution is however needed since with this experimental design, we cannot distinguish between the impacts of the different changes that may have cumulative effects (or not).

To summarize, the impact of changing a few steps in the metabarcoding protocol on community patterns (beta diversity) is limited, since for all three labs the main clustering is based on the locations. In contrast, for alpha diversity equivocal results are observed between the used protocols. To counteract for these differences, we emphasize the need for standardisation when looking at alpha diversity. We suggest that at least the fidelity of the DNA polymerase and the PCR cycling conditions need to be standardised in the metabarcoding protocol when comparing alpha diversity of metabarcoding results across countries and studies.

4.3. Considerations for implementing DNA metabarcoding for marine environmental monitoring

Different case studies showed that DNA metabarcoding can be used to assess biodiversity (Aylagas et al., 2016; Lejzerowicz et al., 2015; Pawlowski et al., 2014), but the lack of standardisation is an important drawback to routinely implement this method in monitoring programs (Darling et al., 2017; Goodwin et al., 2017). Based on our results of the reproducibility and robustness tests using bulk DNA metabarcoding, some statements can be made on the use of DNA metabarcoding for marine environmental monitoring. Studies focusing on beta diversity patterns or changes in community patterns across space and/or time can rely on bulkDNA metabarcoding since results are comparable to patterns observed in morphological studies (Cahill et al., 2018), reproducible and with the protocol differences tested here, quite robust as well. However, it is pertinent which changes are made, as for example the chosen DNA source (Derycke et al., 2021), primer pair (Braukmann et al., 2019; Elbrecht and Leese, 2017; Lobo et al., 2017) and PCR replicates (Van den Bulcke et al., 2021) already showed significant differences, and therefore, these steps in the metabarcoding protocol cannot be changed without implications for diversity assessments. In contrast, using another DNA polymerase with similar fidelity or another PCR clean-up kit (based on the same principle) yield similar ecological patterns and can therefore be changed without impacting bulkDNA metabarcoding results. When a study aims to focus on alpha diversity i.e. species richness and other biodiversity indices such as e.g. Shannon or Inverse Simpson, a fully standardised protocol is of high importance. Similar to other metabarcoding studies (Alberdi et al., 2017; Brannock and Halanych, 2015; Dopheide et al., 2018), our study showed that changes in the protocol can result in significant differences in these indices. It is therefore important to adhere to a predefined workflow and when not possible, to at least provide a detailed description of the technical details of the protocol used. This insight becomes definitely important when comparing alpha diversity results over time or between regions that are the result of different metabarcoding studies. In this case observed differences in e.g. number of species might be caused by methodological changes instead of being ecological relevant differences.

In some studies, the species list is important, Aylagas et al. (2018) already highlighted the importance of a reliable and well-curated sequence reference database to assign correct species names to sequences for this purpose. The last decade, genetic data repositories (e.g. BOLD or GenBank) containing DNA barcodes are growing, but this is a long-term project and effort is made in the curation of these reference databases (Radulovici et al., 2021). Still, many species are missing a reference barcode. In 2019, only 22–48% of European marine species were present in BOLD (Weigand et al., 2019). Our study corroborated this since of the 24 morphologically identified species not detected by metabarcoding, 17 species were not present in our reference database.

Finally, this study showed that species with high read numbers and large body size were more easily detected. For low abundant and smaller species, this is not always the case, as some of these were only detected by one laboratory. Furthermore, we saw that species uniquely found with the morphology and with a reference sequence in the reference database, mostly belonged to the phyla “Polychaeta”, a class with higher variation in the COI gene and most likely a lower primer efficiency (Carr et al., 2011). On the other hand, bulkDNA metabarcoding will pick up species that are not detected using traditional methods as well (Aylagas et al., 2016). Therefore, DNA metabarcoding and the morphological identification are seen as complementary methods to assess diversity (Kelly et al., 2017). Despite the discrepancies in detected species, both the morphology-based and genetic datasets were able to detect the difference in diversity with the three indices (species richness and inverse Simpson and Shannon index after square root transformation of the reads).

5. Conclusion

There is a trend towards using genetic methods for marine environmental monitoring, also referred to as Biomonitoring 2.0 (Baird and Hajibabaei, 2012), but therefore metabarcoding data should be comparable and reproducible across countries and studies. We show that when using a standardized protocol the detected alpha diversity was very similar for the different laboratories, especially for high abundant species and species with a large body size. In addition, an almost identical clustering in macrobenthic community composition based on the different locations was observed. Second, minor effects were seen on the community composition when slightly modifying the lab protocol, while alpha diversity was significantly impacted. Consequently, a standardized protocol allows a better comparison between metabarcoding results from different studies. In the absence of an agreed standardized protocol across countries, we stress the importance to provide a detailed description of the lab protocol used to obtain the metabarcoding data to allow a correct interpretation of metabarcoding results across studies.

6. Data accessibility statement

The sequencing datasets and corresponding metadata generated for this study will become available in the online system Marine Data Archive (MDA) https://marinedataarchive.org/.

CRediT authorship contribution statement

Laure Van den Bulcke: Formal analysis, Visualization, Writing – original draft. **Annelies De Backer:** Conceptualization, Writing – review & editing, Project administration, Funding acquisition. **Jan Wittoeck:** Investigation. **Kevin Beentjes:** Investigation, Writing – review & editing. **Sara Maes:** Investigation, Resources. **Magdalini Christodoulou:** Conceptualization, Investigation, Writing – review & editing. **Pedro Martinez Arbizu:** Conceptualization. **Rumakanta Sapkota:** Investigation. **Berry Van der Hoorn:** Conceptualization, Writing – review & editing. **Anne Winding:** Writing – review & editing. **Kris Hostens:** Writing – review & editing, Funding acquisition. **Sofie Derycke:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The sequencing datasets and corresponding metadata generated for this study is available in the online system Marine Data Archive (MDA) https://mda.vliz.be/archive.php?folder=8705

Acknowledgements

We thank Sahar Khodami for her contribution to the lab work at Senckenberg am Meer, and Hans Hillewaert for his contribution to the reference database. The input of Bart Ampe and Joran Vanhollebeke on the statistics and the bioinformatics pipeline, respectively, was highly appreciated. Ship Time RV Belgica was provided by BELSPO and RBINS-OD Nature. We thank the crew for the logistic support during the sampling campaign. Funding for this research was received through the Sand Fund of the Federal Public Service Economy and GEANS – Genetic tools for Ecosystem health Assessment in the North Sea region, an Interreg project supported by the North Sea Programme of the European Regional Development Fund of the European Union.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2023.110207.

References

Alberdi, A., Aizpurua, O., Gilbert, M.T.P., Bohmann, K., Mahon, A., 2017. Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods Ecol. Evol. 9, 134–147. https://doi.org/10.1111/2041-210x.12849.

Antich, A., Palacín, C., Cebrian, E., Golo, R., Wangensteen, O.S., Turon, X., 2021. Marine biomonitoring with eDNA: Can metabarcoding of water samples cut it as a tool for surveying benthic communities? Mol. Ecol. 30, 3175–3188. https://doi.org/10.1111/mec.15641.

Aubry, A., Elliott, M., 2006. The use of environmental integrative indicators to assess seabed disturbance in estuaries and coasts: Application to the Humber Estuary, UK. Mar. Pollut. Bull. 53, 175–185. https://doi.org/10.1016/j.marpolbul.2005.09.021.

Aylagas, E., Borja, Á., Irigoien, X., Rodríguez-Ezpeleta, N., 2016. Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. Front. Mar. Sci. 3 https://doi.org/10.3389/fmars.2016.00096.

Aylagas, E., Borja, A., Muxika, I., Rodriguez-Ezpeleta, N., 2018. Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. Ecol. Ind. 95, 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044.

Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol. Ecol. 21, 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x.

Beentjes, K.K., Speksnijder, A.G.C.L., Schilthuizen, M., Hoogeveen, M., Pastoor, R., van der Hoorn, B.B., 2019. Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. PLoS One 14, e0226527. https://doi.org/10.1371/journal.pone.0226527.

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

Brannock, P.M., Halanych, K.M., 2015. Meiofaunal community analysis by high-throughput sequencing: comparison of extraction, quality filtering, and clustering methods. Mar. Genomics 23, 67–75. https://doi.org/10.1016/j.margen.2015.05.007.

Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J.R., Sones, J.E., Zakharov, E.V., Hebert, P.D.N., 2019. Metabarcoding a diverse arthropod mock community. Mol. Ecol. Resour. 19, 711–727. https://doi.org/10.1111/1755-0998.13008.

Breine, N.T., De Backer, A., Van Colen, C., Moens, T., Hostens, K., Van Hoey, G., 2018. Structural and functional diversity of soft-bottom macrobenthic communities in the Southern North Sea. Estuar. Coast. Shelf Sci. 214, 173–184. https://doi.org/10.1016/j.ecss.2018.09.012.

Buchner, D., Macher, T.H., Beermann, A.J., Werner, M.T., Leese, F., 2021. Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. Environ. Sci. Ecotechnol. 8, 100122 https://doi.org/10.1016/j.ese.2021.100122.

Cahill, A.E., Pearman, J.K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J.-P., Olenin, S., Šiaulys, A., Somerfield, P.J., Trayanova, A., Uyarra, M.C., Chenuil, A., 2018. A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. Ecol. Evol. 8, 8908–8920. https://doi.org/10.1002/ece3.4283.

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 (7), 581–583.

Carr, C.M., Hardy, S.M., Brown, T.M., Macdonald, T.A., Hebert, P.D., 2011. A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. PLoS One 6, e22232. https://doi.org/10.1371/journal.pone.0022232.

Daily, G.C., Polasky, S., Goldstein, J., Kareiva, P.M., Mooney, H.A., Pejchar, L., Ricketts, T.H., Salzman, J., Shallenberger, R., 2009. Ecosystem services in decision making: time to deliver. Front. Ecol. Environ. 7, 21–28. https://doi.org/10.1890/080025.

Darling, J.A., Galil, B.S., Carvalho, G.R., Rius, M., Viard, F., Piraino, S., 2017. Recommendations for developing and applyng genetic tools to assess and manage biological invasions in marine ecosystems. Mar. Policy 85, 54–64. https://doi.org/10.1016/j.marpol.2017.08.014.

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

Derycke, S., Maes, S., Van den Bulcke, L., Vanhollebeke, J., Wittoeck, J., Hillewaert, H., Ampe, B., Haegeman, A., Hostens, K., De Backer, A., 2021. Detection of macrobenthos species with metabarcoding is consistent in bulkDNA but dependent on body size and sclerotization in eDNA from the ethanol preservative. Front. Mar. Sci. https://doi.org/10.3389/fmars.2021.637858.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

Dopheide, A., Xie, D., Buckley, T.R., Drummond, A.J., Newcomb, R.D., Bunce, M., 2018. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods Ecol. Evol. 10, 120–133. https://doi.org/10.1111/2041-210x.13086.

Duarte, S., Leite, B.R., Feio, M.J., Costa, F.O., Filipe, A.F., 2021. Integration of DNA-based approaches in aquatic ecological assessment using benthic macroinvertebrates. Water 13, 331.

Duncan, C., Thompson, J.R., Pettorelli, N., 2015. The quest for a mechanistic understanding of biodiversity-ecosystem services relationships. Proc. Biol. Sci. 282, 20151348. https://doi.org/10.1098/rspb.2015.1348.

Edward, W.B., 1984. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. Adv. Ecol. Res. 14, 1–55. https://doi.org/10.1016/S0065-2504(08)60168-3.

Elbrecht, V., Leese, F., 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Front. Environ. Sci. 5, 11. https://doi.org/10.3389/fenvs.2017.00011.

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

Feinstein, L.M., Sul, W.J., Blackwood, C.B., 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. Appl. Environ. Microbiol. 75, 5428–5433. https://doi.org/10.1128/AEM.00120-09.

GEANS. 2021. Laboratory protocol for DNA metabarcoding of soft sediment macrobenthos from the North Sea.

Goodwin, K.D., Thompson, L.R., Duarte, B., Kahlke, T., Thompson, A.R., Marques, J.C., Cacador, I., 2017. DNA sequencing as a tool to monitor marine ecological status. Front. Mar. Sci. 4 https://doi.org/10.3389/fmars.2017.00107.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hanfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., Kelly, M., 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Res. 138, 192–205. https://doi.org/10.1016/j.watres.2018.03.003.

Hollatz, C., Leite, B.R., Lobo, J., Froufe, H., Egas, C., Costa, F.O., Chain, F., 2017. Priming of a DNA metabarcoding approach for species identification and inventory in marine macrobenthic communities. Genome 60 (3), 260–271. https://doi.org/10.1139/gen-2015-0220.

Jia, Y., Zhao, S., Guo, W., Peng, L., Zhao, F., Wang, L., Fan, G., Zhu, Y., Xu, D., Liu, G., Wang, R., Fang, X., Zhang, H., Kristiansen, K., Zhang, W., Chen, J., 2022. Sequencing introduced false positive rare taxa lead to biased microbial community diversity, assembly, and interaction interpretation in amplicon studies. Environ. Microbiome 17, 43. https://doi.org/10.1186/s40793-022-00436-y.

Kelly, R.P., Closek, C.J., O'Donnell, J.L., Kralj, J.E., Shelton, A.O., Samhouri, J.F., 2017. Genetic and manual survey methods yield different and complementary views of an ecosystem. Front. Mar. Sci. 3 https://doi.org/10.3389/fmars.2016.00283.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G., Taylor, M.I., 2019. How quantitative is metabarcoding: A meta-analytical approach. Mol. Ecol. 28, 420–430. https://doi.org/10.1111/mec.14920.

Lanzen, A., Lekang, K., Jonassen, I., Thompson, E.M., Troedsson, C., 2017. DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. PLoS One 12, e0179443. https://doi.org/10.1371/journal.pone.0179443.

Lejzerowicz, F., Esling, P., Pawlowski, J., 2014. Patchiness of deep-sea benthic Foraminifera across the Southern Ocean: Insights from high-throughput DNA sequencing. Deep Sea Res. Part II: Top. Stud. Oceanogr. 108, 17–26. https://doi.org/10.1016/j.dsr2.2014.07.018.

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T.A., Black, K.D., Pawlowski, J., 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Sci. Rep. 5 https://doi.org/10.1038/srep13932.

Leray, M., Knowlton, N., 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5, e3006. https://doi.org/10.7717/peerj.3006.

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, 14. https://doi.org/10.1186/1742-9994-10-34.

Lobo, J., Shokralla, S., Costa, M.H., Hajibabaei, M., Costa, F.O., 2017. DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. Sci. Rep. 7, 13. https://doi.org/10.1038/s41598-017-15823-6.

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E., Shapiro, B., 2018. Minimizing polymerase biases in metabarcoding. Mol. Ecol. Resour. 18, 927–939. https://doi.org/10.1111/1755-0998.12895.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., Wilding, T.A., 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Mol. Ecol. Resour. 14, 1129–1140. https://doi.org/10.1111/1755-0998.12261.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., Barquin, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018a. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637–638, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Jones, J.I., Sagova-Mareckova, M., Moritz, C., Barquin, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018b. The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.

Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. Nature 405, 212–219. https://doi.org/10.1038/35012221.

Radulovici, A.E., Vieira, P.E., Duarte, S., Teixeira, M.A.L., Borges, L.M.S., Deagle, B., Majaneva, S., Redmond, N., Schultz, J.A., Costa, F.O., 2021. Revision and annotation of DNA barcode records for marine invertebrates: report of the 8th iBOL conference hackathon. bioRxiv: 2021.2003.2007.434272. doi:10.1101/2021.03.07.434272.

Schnell, I.B., Bohmann, K., Gilbert, M.T.P., 2015. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol. Ecol. Resour. 15, 1289–1303. https://doi.org/10.1111/1755-0998.12402.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Technol. J. 27, 379–423.

Simpson, E. 1949. Measurement of diversity. Nature 163: 688. doi: 10.1038/163688a0.

Smith, D.P., Peay, K.G. 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS One 9: e90234. doi: 10.1371/journal.pone.0090234.

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

van der Loos, L.M., Nijland, R., 2021. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Mol. Ecol. 30 (13), 3270–3288.

Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M., Bouchez, A., 2017. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Sci. 36, 162–177. https://doi.org/10.1086/690649.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/aem.00062-07.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., Dijkstra, K.-D., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Leese, F., Macher, J.N., Mamos, T., Paz, G., Pešić, V., Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Várbíró, G., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Sci. Total Environ. 678, 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247.

Yang, C., Bohmann, K., Wang, X., Cai, W., Wales, N., Ding, Z., Gopalakrishnan, S., Yu, D. W. 2020. Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. bioRxiv: 2020.2007.2007.187666. doi:10.1101/2020.07.07.187666.

Yoder, M., Tandingan De Ley, I., King, I., Mundo-Ocampo, M., Mann, J., Blaxter, M., Poiras, L., Ley, P., 2006. DESS: A versatile solution for preserving morphology and extractable DNA of nematodes. Nematology 8, 367–376. https://doi.org/10.1163/156854106778493448.

Zaiko, A., Greenfield, P., Abbott, C., von Ammon, U., Bilewitch, J., Bunce, M., Cristescu, M.E., Chariton, A., Dowle, E., Geller, J., Ardura Gutierrez, A., Hajibabaei, M., Haggard, E., Inglis, G.J., Lavery, S.D., Samuiloviene, A., Simpson, T., Stat, M., Stephenson, S., Sutherland, J., Thakur, V., Westfall, K., Wood, S.A., Wright, M., Zhang, G., Pochon, X., 2021. Towards reproducible metabarcoding data: Lessons from an international cross-laboratory experiment. Mol. Ecol. Resour. 22, 519–538. https://doi.org/10.1111/1755-0998.13485.

Zinger, L., Bonin, A., Alsos, I.G., Balint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding-Need for robust experimental designs to draw sound ecological conclusions. Mol. Ecol. 28, 1857–1862. https://doi.org/10.1111/mec.15060.