MBMG
Metabarcoding & Metagenomics

Research Article

Towards harmonization of DNA metabarcoding for monitoring marine macrobenthos: the effect of technical replicates and pooled DNA extractions on species detection

Laure Van den Bulcke[1,2], **Annelies De Backer**[1], **Bart Ampe**[1], **Sara Maes**[1], **Jan Wittoeck**[1], **Willem Waegeman**[2], **Kris Hostens**[1], **Sofie Derycke**[1,2]

1 *Flanders Research Institute for Agriculture, Fisheries and Food, Ostend, Belgium*

2 *Ghent University, Ghent, Belgium*

Corresponding author: Laure Van den Bulcke (laure.vandenbulcke@ilvo.vlaanderen.be)

Abstract

DNA-based monitoring methods are potentially faster and cheaper compared to traditional morphological benthic identification. DNA metabarcoding involves various methodological choices which can introduce bias leading to a different outcome in biodiversity patterns. Therefore, it is important to harmonize DNA metabarcoding protocols to allow comparison across studies and this requires a good understanding of the effect of methodological choices on diversity estimates. This study investigated the impact of DNA and PCR replicates on the detection of macrobenthos species in locations with high, medium and low diversity. Our results show that two to three DNA replicates were needed in locations with a high and medium diversity to detect at least 80% of the species found in the six DNA replicates, while three to four replicates were needed in the location with low diversity. In contrast to general belief, larger body size or higher abundance of the species in a sample did not increase its detection prevalence among DNA replicates. However, rare species were less consistently detected across all DNA replicates of the location with high diversity compared to locations with less diversity. Our results further show that pooling of DNA replicates did not significantly alter diversity patterns, although a small number of rare species was lost. Finally, our results confirm high variation in species detection between PCR replicates, especially for the detection of rare species. These results contribute to create reliable, time and cost efficient metabarcoding protocols for the characterization of macrobenthos.

Key Words

biological replicates, Bulk DNA, diversity, DNA replicates, Macrobenthos, Metabarcoding protocol, North Sea, PCR replicates, pooling, technical replicates

Introduction

Marine and coastal ecosystems are highly valuable environments as they deliver many ecosystem services (Daily et al. 2009; Duncan et al. 2015) and functions (Duncan et al. 2015) to society. Sustainable marine exploitation is thus important to ensure a healthy marine environment, and therefore, evaluation of the potential effects of human activities is needed through environmental impact assessments (EIAs). Such assessments typically use macrobenthic communities as indicators (Van Hoey et al. 2010), because they rapidly respond to environmental changes.

Traditionally, macrobenthic species identification for EIAs is based on morphological characteristics, but potentially faster and cheaper methods, like DNA metabarcoding, have been successfully applied to monitor human impacts in marine environments (Chariton et al. 2010; Bik et al. 2012; Pawlowski et al. 2014; Lejzerowicz et al. 2015; Lobo et al. 2017; Aylagas et al. 2016).

DNA metabarcoding of macrobenthos uses bulk DNA, extracted from the organisms in the samples. Through PCR, a portion of the mitochondrial COI gene is amplified using specific primers and amplification conditions, followed by next generation sequencing (Baird and

PENSOFT.

Hajibabaei 2012). When using Illumina's sequencing by synthesis, the COI PCR products are loaded onto a flow cell and modified nucleotides with a fluorescent tag bind to the PCR template strand during bridge amplification. Millions of copies are amplified by cluster generation, and this fluorescent signal is translated into sequences in a process called base calling. Since multiple species in many samples can be sequenced in a single instrument run, processing time of samples is substantially reduced compared to morphological identification (Aylagas et al. 2018). After data analysis, the obtained sequences are linked to taxonomic names by comparing them to DNA sequences of morphologically identified specimens in private or public reference databases. To be applicable in EIAs, a standardized protocol that allows for reproducible and reliable DNA metabarcoding results is a prerequisite (Darling et al. 2017; Goodwin et al. 2017; Hering et al. 2018; Pawlowski et al. 2018).

A recent literature review on DNA metabarcoding for marine macrobenthos showed the large variation in methodological steps between different studies (van der Loos and Nijland 2020). Specifically, DNA extraction and PCR amplification were pinpointed for possibly introducing bias in macrobenthos diversity estimates. After collection, macrobenthos samples are sieved and organisms are blended to a homogeneous solution (Aylagas et al. 2016). Next, a subsample is taken from the homogenous "soup" for DNA extraction. In theory, this subsample should contain the DNA of all species present in the sample. Yet, variation in taxonomic composition between subsamples is observed for meiofauna in sediment samples, even after sieving to counteract for the heterogeneous distribution (Brannock and Halanych 2015), as well as for deep sea benthos (Lejzerowicz et al. 2014). Therefore, taking subsamples for DNA extractions, from now on referred to as DNA replicates, has been suggested to increase reliability of species detection (Feinstein et al. 2009; Lanzen et al. 2017). Intuitively, communities with higher diversity may require more DNA replicates to detect all species. Often, the separate DNA extractions are generally pooled together for further processing (Aylagas et al. 2016; Hollatz et al. 2017), to decrease costs and processing time of the metabarcoding protocol. When DNA extractions are pooled, less volume is taken from the separate DNA extractions, which decreases the chance to find rare species. This has been demonstrated in a study on detecting fresh water fishes in eDNA samples (Sato et al. 2017), where DNA concentrations are typically very low. The impact of pooling for bulk macrobenthos communities on the detection of species has hitherto not been investigated.

Another important step in the metabarcoding protocol that can introduce bias is PCR amplification. PCR is a stochastic process, because a molecule will be either replicated or not based on whether the primers were able to sufficiently bind during each PCR cycle. The impact of primer choice on species detection is well known (Elbrecht and Leese 2017; Lobo et al. 2017; Braukmann et al. 2019; Derycke et al. 2021), but the number of PCR replicates can also affect the detected diversity (Dopheide et al. 2018). PCR replicates, where the same DNA extract is amplified in separate PCR reactions, can decrease the stochasticity of this process (Leray and Knowlton 2015). Increasing the number of PCR replicates increases the number of detected species, especially rare species, but this also increases work and costs. Taking three PCR replicates in faeces from bats showed that approximately 72% of the species present in the separate PCR replicates were unique (Alberdi et al. 2017). In contrast, PCR replicates did not significantly contribute to the explained variation in community composition of freshwater macrobenthos (Martins et al. 2019). The use of three PCR replicates has been suggested for macrobenthic communities (Bourlat et al. 2016; Leray and Knowlton 2017), but a recent literature review on marine metabarcoding found that 48% of the investigated studies did not perform PCR replicates (van der Loos and Nijland 2020).

In this study, we evaluated the effect of DNA and PCR replicates, as well as the effect of pooling DNA replicates before PCR amplification on alpha and beta diversity of macrobenthos. We considered that the impact of technical replication may be linked to the diversity of the sampled communities and therefore sampled three macrobenthic communities in the Belgian part of the North Sea, characterized by low, medium and high diversity (Breine et al. 2018). First, we investigated bias in species detection introduced by DNA extraction by sequencing six DNA replicates of all biological samples collected in the three macrobenthos communities. Second, we investigated whether pooling the DNA extractions, as well as the number of pooled DNA extracts, would affect alpha and beta diversity patterns. Pooling multiple DNA extractions before PCR reduces the number of samples further in the metabarcoding process and would decrease time and costs. Third, we investigated PCR bias, by sequencing a subset of the DNA extractions in three separate PCR reactions. These results contribute to create time and cost efficient protocols yielding reliable and robust results for macrobenthos monitoring.

Material and methods

Sample collection

Sampling was carried out at three different locations in the Belgian Part of the North Sea (BPNS), covering macrobenthic communities with low, medium and high diversity (Breine et al. 2018) (Suppl. material 1: Fig. S1). Each community is named after the indicator species, i.e. the species that contributed most to within cluster group similarity (Breine et al. 2018). The *Limecola balthica* community in location ZVL represents a low diverse community (around 6 species per sample) with fine muddy sediment and low bioturbation. The *Hesionura elongata* community in location 840 represents a medium diverse community (around 14 species per sample) with an

offshore coarse sandy habitat. The *Abra alba* community in location 120 represents a highly diverse community (around 26 species per sample) with coastal fine muddy sand and a high bioturbation potential. In each location, three Van Veen grabs were taken as biological replicates (A, B and C). The sediment was sieved on a 1 mm sieve and animals remaining on the sieve were fixed in ethanol and stored at -20 °C until further processing.

Sample processing and morphological identification

The samples were further processed as described in Derycke et al. (2021), following the protocol outlined by Aylagas et al. (2016). Samples were decanted using a 1 mm sieve and tap water until no specimens were recovered from the samples (varying from 6 to 13 times). The specimens on the sieve were stored in ethanol, while the remaining material (e.g. shells) was screened for heavier specimens, which were picked and added to the decanted material in ethanol. To compare with the traditional method based on morphological species identification, one replicate from each location (ZVL-A, 840-C, 120-B) was identified morphologically up to species level, except for juveniles which were identified up to genus level, and specimens belonging to Nemertea, Anthozoa and Oligochaeta, which were identified up to phylum, class and order level, respectively. The collected specimens in ethanol were mixed with a blender or a mortar and pestle for samples with less than 100 ml to obtain a homogenous bulk solution.

Experimental set-up and library preparation

DNA Replicates

Of each biological replicate, six subsamples of 2 ml were taken from the bulk solution for DNA extraction, resulting in 18 subsamples per location (Fig. 1). After centrifuging these 2 ml Eppendorf tubes, the supernatant was removed with a pipet. The remaining ethanol was evaporated by incubating the Eppendorf tubes with open lids at 50 °C for one hour. After transferring the pellet to a 2 ml tube with beads from the DNeasy PowerSoil Kit (QIAGEN, Hilden), 10 µl of Proteinase K (20 mg/ml) was added for the lysing step. The rest of the DNA extraction was done following the protocol provided by the manufacturer. The PCR amplification consisted of two steps. The first step was performed with primers that amplify 313 bp of the Folmer region (Leray et al. 2013), in triplicate. The PCR mix contained 8.5 µl nuclease free water, 12.5 µl 2× KAPA HiFi HotStart ReadyMix, 0.75 µl of the forward primer (MICOIintF with Illumina adaptor 10 µM) and 0.75 µl of the reverse primer (jgHCO2198 with Illumina adaptor 10 µM). The primer sequences can be found in Suppl. material 2: Table S1. After adding 2.5 µL of the DNA extract, this mix was incubated using the following PCR conditions: initial denaturation for 3 min at 95 °C, 35 cycles of denaturation for 30 s at 98 °C, annealing for 30 s at 57 °C and extension for 30 s at 72 °C, and a final extension for 1 min at 72 °C. The three PCR replicates were pooled
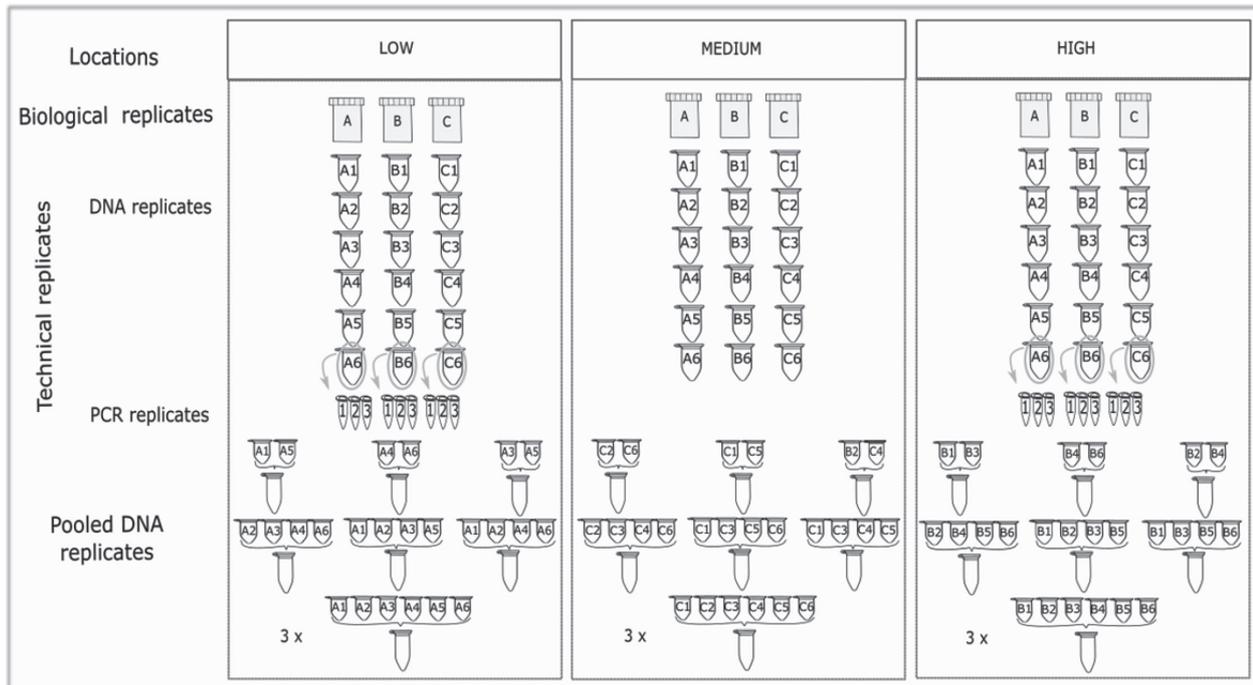

Figure 1. Experimental set-up. Three locations, with a low, medium or high diversity, were sampled and in each location, three Van Veen grab samples (biological replicates A, B and C) were taken. To test the effect of technical replicates on the alpha and beta diversity, six subsamples of each biological replicate were taken for DNA extractions (DNA replicates 1 to 6). In two locations (LOW and HIGH), three separate PCR amplifications were sequenced of one DNA extraction (PCR replicates 1 to 3). To test the effect of pooling multiple DNA replicates, three combinations of two, four or six DNA replicates were pooled before amplification (Pooled DNA replicates for each number of pooled DNA extracts).

together in one tube, of which 25 μl of the pooled PCR product was used for purification using 20 μl Ampure XP beads and eluted in 50 μl TE-buffer. For the second step, the index PCR, 2.5 μl of the purified PCR product was added to a PCR mix containing 5 μl nuclease free water, 12.5 μl 2× KAPA HiFi HotStart ReadyMix and 2.5 μl of each index primer (Nextera XT). PCR conditions for the index PCR were: initial denaturation for 3 min at 95 °C, 8 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 55 °C and extension for 30 s at 72 °C, and a final extension for 3 min at 72 °C. To confirm the success of the index PCR, five samples from PCR1 and the index PCR were loaded on the Qiaxcel (QIAGEN, Hilden). After measuring with the Quantus (Promega, Madison), the indexed PCR products were equimolarly pooled and sent for sequencing using the Illumina Miseq 2*300 bp platform (sequenced by Admera Heath Biopharma Services, New Jersey).

Pooled DNA extractions

Pooling can decrease the processing time of the DNA metabarcoding protocol. To investigate the effect of pooling on the detected diversity with metabarcoding, two, four, and six DNA extracts were pooled before the PCR step. Each level of pooling was performed in triplicate with different randomly chosen DNA extracts from the DNA replicates experiment. This was done for the morphologically identified replicate from each location (ZVL-A, 840-C, 120-B). In total, nine pooled DNA extractions for each location were made (Fig. 1). Pooling of multiple DNA extractions was followed by PCR amplification and sequencing as explained above. The specific combinations of DNA extracts used for pooling can be found in Figure 1.

PCR replicates

To assess variation between PCR replicates, the three PCR products of a subset of samples were processed individually for the index PCR and sequencing, under the same conditions as described above. PCR products of one DNA extract of each biological replicate (A6, B6 and C6) in the lowest and highest diversity locations (ZVL and 120) were used. This resulted in an additional 18 samples (Fig. 1).

The sequencing datasets and corresponding metadata generated for this study can be found as BioProject in the online NCBI repository under accession number PRJ-NA683870.

Processing of raw reads

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 with the DADA2 pipeline in the dada2 v1.17.0 package (Callahan et al. 2016) in R v4.0.2 (R Core Team

2020). The maximum number of errors allowed in a read was set to three, while for all other filtering parameters standard settings were used. Reads were further trimmed by removing parts with a quality score lower than 30. Unique reads were determined, merged and filtered for chimeras for each sample. A reference database was composed of in-house COI sequences of macrobenthos from multiple monitoring campaigns in the Belgian Part of the North Sea, complemented by Bold COI sequences of macrobenthic species living in this area, resulting in 324 COI sequences from 289 species. With the assignTaxonomy function in the dada2 package, this reference database was used for taxonomic assignment of the sequences, by implementing the Ribosomal Database Project (RDP) Classifier (Wang et al. 2007). Standard settings were used, except for the minimum bootstrap confidence parameter, which was set to 80. To compare the taxonomic assignment of ASVs at the different taxonomic levels between the different stations, a barplot was made. The unassigned ASVs were also matched against the NCBI nt database using Blastn. Taxonomic assignments were taken into account when qcov >75 and pident >90. The total number of reads were compared between the different DNA replicates, the pooled DNA extractions and the PCR replicates and visualized by barplots in R.

Alpha diversity of DNA metabarcoding datasets

To take into account the different sequencing depth when comparing alpha diversity, samples of DNA replicates and pooled DNA extractions were rarified at 24,000 reads, while samples of PCR replicates were rarified at 12,000 reads. These numbers were a tradeoff between reaching the plateau of the rarefaction curves and removing a minimum of samples. Samples with lower number of reads were discarded from the dataset, resulting in four removed samples in total: three samples of DNA replicates (ZVL-A-1, ZVL-A-2, 120-C-6) and one sample of PCR replicates (120-C-3).

To assess whether number of ASVs and number of species were significantly different between locations with low, medium and high diversity, nested ANOVA's were conducted with factors diversity (levels: Low, Medium and High) and biological replicate (levels: A, B and C) nested within diversity, separately for the DNA or PCR replicate datasets. Assumptions for the ANOVA test were checked by making plots and performing a Levene test and a Shapiro test for investigating the homogeneity of the variances and the normality of the data, respectively. When the assumption for normality was not met, data were transformed with the reciprocal transformation 1/x. In case the Shapiro test remained significant after transformation, normality was assumed when the values in the center of the Quantile-Quantile plot fitted within the 95% confidence interval.

To investigate the effect of DNA and PCR replicates on ASV and species richness, the ASVs and species uniquely found in one replicate and shared between all replicates were investigated using upset plots with the R

package UpSet R v1.4.0 (Conway et al. 2017). The number of shared and unique ASVs and species per biological replicate in each location, as well as the number shared between two to five replicates were calculated in percentages and visualized in barplots. Using the R package vegan v2.5-6 (Dixon 2003), accumulation curves were constructed to investigate the increase in ASVs and species richness with increasing number of DNA and PCR replicates. Upset plots were made to assess the number of shared and unique species for the morphological and metabarcode datasets (DNA and PCR replicates). For the DNA metabarcode dataset, species detected in all six replicates and uniquely in one replicate were listed. For each species, the corresponding size class was given based on a previous study (Derycke et al. 2021). If the species was also found in the morphologically identified sample, the abundance (counts) was given.

To investigate the effect of pooled DNA extractions on ASV and species richness, differences in the number of ASVs and species between the pooled DNA extractions were tested with a two-way ANOVA with factors diversity (levels: High, Medium and Low) and number of pooled DNA extracts (levels: 2, 4 or 6). Assumptions for the ANOVA test (homogeneity of the variances and normality of the data) were checked by making plots and performing a Levene test and a Shapiro test, respectively. When the assumption for the normality was not met, data were transformed with the reciprocal transformation $1/x$. When this transformation did not give a normal distribution, the tests were still applied if the values fitted within the 95% confidence interval at the center of the Quantile-Quantile plot. Furthermore, we assessed the observed number of species and the mean number of expected species, each for two, four and six pooled DNA extractions. The observed number of species represents the sequencing results of the pooled DNA extractions, while the mean number of expected species is calculated as the mean number of species observed in the different separate DNA extractions that were used for pooling. For each diversity level, a two-way ANOVA with factors number of pooled DNA extracts (levels: 2, 4 or 6) and number of species (levels: Observed or Expected) was conducted. Assumptions for the ANOVA's were checked as explained above. To investigate the unique and shared species between the separate and pooled DNA extractions, upset plots were made using the R package UpSetR v1.4.0. The number of shared and unique species in each location were calculated in percentages and visualized in barplots.

Beta diversity analyses of DNA metabarcoding datasets

To investigate variation in community composition between biological and DNA replicates, non-metric multidimensional scaling (nMDS) plots based on Jaccard (Jaccard 1912) or Bray-Curtis (Edward 1984) dissimilarity indices were constructed, using the R package vegan v2.5-6. To compare the species communities between the different DNA replicates, a nested PERMANOVA was

conducted with diversity (Low, Medium and High) as main effect and biological replicate (A, B and C) nested in diversity. This PERMANOVA was performed with the function Adonis from the R package vegan v2.5-6, using 9999 permutations. A distance dispersion test and permutest were used to test the homogeneity of dispersion in the samples with the R package vegan v2.5-6. To compare the species communities between the different pooled replicates, a two-way PERMANOVA was conducted, with the two main effects diversity (Low, Medium and High) and number of pooled DNA extractions (2, 4 and 6).

Results

Species composition based on morphological identification

One biological replicate from each location (120-B, 840-C, ZVL-A) was identified morphologically up to species level. Respectively 3, 10 and 39 species were identified in the sample (0.1 m²) with low (ZVL-A), medium (840-C) and high (120-B) diversity. A detailed list of all species found in each of the samples is presented in Suppl. material 2: Table S2.

Processing of raw reads

After filtering, the number of reads for the different technical replicates ranged between 215 and 225,722. A detailed table with the number of reads after each filtering step for each replicate can be found in Suppl. material 2: Table S3. The average number of reads per biological replicate ranged between 15,308 and 173,523 (Suppl. material 1: Fig. S3a–c). The mean number of ASVs per biological replicate ranged between 25 and 191 (Suppl. material 1: Fig. S3d–f). After taxonomic assignment, only 8%–26% of the ASVs per biological replicate could be assigned to the species level (Suppl. material 1: Fig. S4). The number of assigned and unassigned ASV per taxon level for the whole dataset can be found in Suppl. material 2: Table S4. After matching the unassigned ASVs against the NCBI database using Blastn, only 8% received reliable taxonomic assignment. To account for variation in read numbers, samples were rarified for further analyses of alpha and beta diversity.

Alpha diversity based on DNA metabarcoding

DNA replicates

The three locations were chosen based on their macrobenthos species diversity. As expected, the nested ANOVA showed that the number of ASVs and the number of assigned species differed significantly between locations with different diversity, and increased from the locations with low (ZVL), medium (840) and high (120) diversity

(Table 1; Suppl. material 1: Fig. S5: ASVs: a; Species: d). In addition to a significant difference between locations, a significant difference between biological replicates within locations for the number of ASVs was also observed (Table 1).

Despite the fact that DNA was extracted from the same biological sample, the majority of ASVs (62%–86%) were uniquely found in one DNA replicate (Fig. 2a). In contrast, only 4%–18% of the ASVs were shared between the six DNA replicates in each biological replicate (Fig. 2a). At the species level, 18%–52% of the detected species were shared between the six DNA replicates (Fig. 2b), and these species were represented by 81% of the reads in the whole dataset. The location with the lowest morphological diversity (ZVL) had the highest number of species that were uniquely found in one DNA replicate (41%–50%, Fig. 2b), but species uniquely found in only one replicate for the different biological replicates were represented by only 14% of the total reads in the dataset.

Irrespective of diversity, species uniquely found in one DNA replicate or shared by all six DNA replicates show similar distribution of the size classes and abundance classes (Suppl. material 2: Table S8; Suppl. material 1: Fig. S6). However, for the different communities separately, smaller species were more often detected uniquely in one replicate than in all six replicates (57% vs 28%, respectively) in the location with high diversity. Furthermore, low abundant species (<5 specimens) were only slightly more found uniquely in one DNA replicate than in all six DNA replicates in the location with high diversity (63% vs 50%; Suppl. material 2: Table S8; Suppl. material 1: Fig. S6). These effects of size and abundance were not observed in the location with medium and low diversity.

Accumulation plots, to investigate the number of DNA replicates capturing most of the diversity, illustrate that the number of ASVs was still increasing for nearly all biological replicates, even when six DNA replicates were taken. The plateau was reached in only one location, more precisely

Table 1. Output ANOVA's. Results of the ANOVA's for alpha diversity (ASVs and species level). Crossed factors are indicated with an asterisk, while brackets are used for nested factors. The output of the Levene test and the Shapiro test are shown in the second and third column, respectively.

|	Homogeneity	Normality	Df	SumsOfSqs	F Value	Pr(>F)
ASVs						
DNA replicates						
Diversity			2	73688	60.135	**4.712e-13**
Diversity(Biological_replicate)	0.061	**0.001**	6	52849	14.376	**7.874e-09**
Residuals			42	25733		
Pooled DNA extracts						
Diversity			2	0.0046143	40.297	**2.253e-07**
Nr_pooled			2	0.0000096	0.084	0.920
Diversity*Nr_pooled	0.452	0.060	4	0.0000364	0.159	0.956
Residuals			18	0.0010306		
PCR replicates						
Diversity			1	0.00248999	47.430	**2.632e-05**
Diversity(Biological_replicates)	0.626	0.053	4	0.00037546	1.788	0.202
Residuals			11	0.00057748		
Species						
DNA replicates						
Diversity			2	2955.99	168.022	**<2e-16**
Diversity(Biological_replicate)	0.341	**0.008**	6	115.15	2.182	0.0639
Residuals			42	369.45		
Pooled DNA extracts						
Diversity			2	2064.30	149.828	**6.023e-12**
Nr_pooled			2	2.30	0.167	0.848
Diversity*Nr_pooled	0.526	**1.689e-05**	4	50.37	1.828	0.167
Residuals			18	124.00		
Pooled expected vs observed in HIGH diverse location						
Nr_pooled			2	34.670	1.7451	0.224
Method			1	9.400	2.9597	0.116
Nr_pooled * Method	0.702	**5.816e-04**	2	12.600	0.6342	0.550
Residuals			10	99.333		
Pooled expected vs observed in MEDIUM diverse location						
Nr_pooled			2	0.000188	0.257	0.778
Method			1	0.000227	0.620	0.449
Nr_pooled * Method	0.977	0.0512	2	0.000597	0.814	0.470
Resiudals			10	0.00367		
Pooled expected vs observed in LOW diverse location						
Nr_pooled			2	3.925	1.204	0.344
Method			1	0.453	0.278	0.611
Nr_pooled * Method	0.359	**0.0243**	2	7.630	2.341	0.152
Residuals			10	14.667		
PCR replicates						
Diversity			1	1210.04	339.841	**1.279e-09**
Diversity(Biological_replicates)	0.385	**0.00377**	4	137.26	9.638	0.001
Residuals			11	39.17		
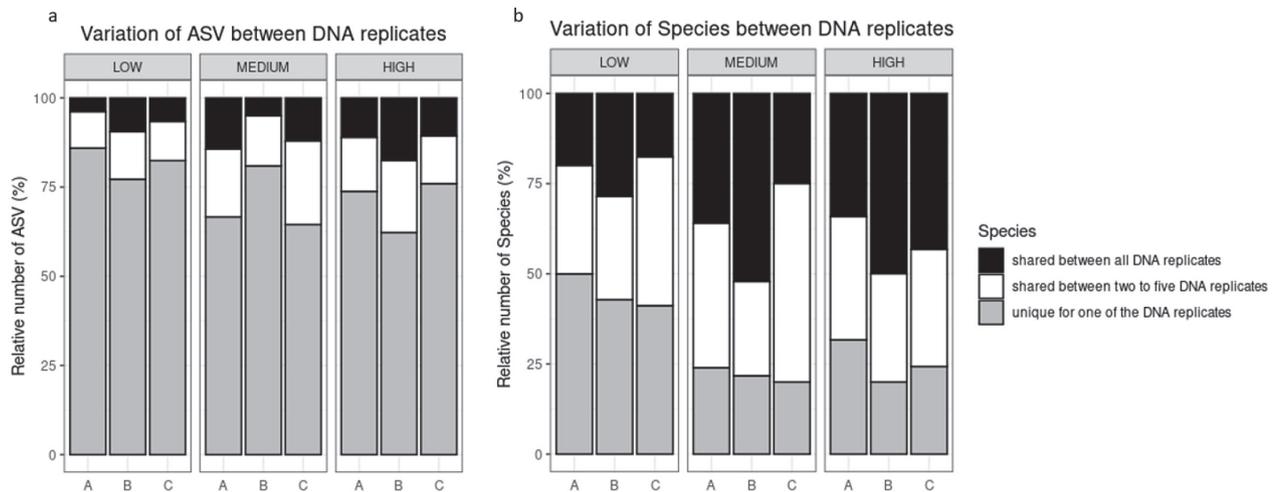
Figure 2. Variation of ASVs and species in DNA replicates. The percentage of unique and shared ASVs (a) or species (b) between all six DNA replicates are shown for locations with low, medium or high diversity. These percentages were calculated based on UpSet plots of the different DNA replicates in biological replicates.

in two biological replicates (A and C) of the location with medium diversity (840) (Suppl. material 1: Fig. S7). In contrast, when looking at the species level, accumulation curves reached a plateau in all biological replicates (Suppl. material 1: Fig. S8). To pick up at least 80% of the species found in the six DNA replicates, two to three replicates were needed in locations with a high (120) and medium (840) diversity, while three to four replicates were needed in the location with low diversity. A detailed table with the predicted number of species when using one, two, three, four, five and six DNA replicates from the species accumulation method can be found in Suppl. material 1: Table S5.

Finally, we assessed the impact of DNA replicates on detecting the number of species that were identified morphologically. For the location with high, medium and low diversity, 31%, 30% and 33% of the morphologically identified species were consistently detected in all six DNA replicates. When looking at the total found species across six DNA replicates (calculated as the union of the six replicates), the detection rate was increased to 49% and 50% of the total number of morphologically identified species for the location with high and medium diversity, while the same percentage was found for the location with low diversity (33%) (Suppl. material 1: Fig. S9). A list with the species only detected with metabarcoding, with indication of the morphological identifications that were only identified up to genus level, can be found in Suppl. material 2: Table S6.

Pooled DNA extractions

To investigate the effect of pooling DNA extracts on the detected diversity with metabarcoding, two, four and six DNA extracts of one biological replicate in the three different locations were pooled before PCR amplification. No significant interaction effect between the diversity and the number of pooled DNA extracts was detected (Table 1). Also no significant difference between the observed number of species or the number of ASVs in one, two, four or six pooled DNA extracts was observed (Table 1). Further-

more, similar as in the DNA replicates, a significant difference between the locations with high, medium and low diversity was found in the number of species and in the number of ASVs (Table 1; Suppl. material 1: Fig. S5b and e). When comparing the mean number of expected species in the two, four or six pooled DNA extracts and the observed number of species, no significant interaction effect between the number of pooled DNA extracts and the number of species (observed vs expected) was detected in the location with low, medium and high diversity (Table 1). For each station, no significant difference between the number of species and between the number of pooled DNA extracts was observed (Table 1; Suppl. material 1: Fig. S10).

Between the separate and the two, four and six pooled DNA extracts, most (68%–88%) of the ASVs were unique (Fig. 3a). Yet, the number of shared species between one and two, four or six pooled DNA extracts was high for the location with high diversity (66%), while 31% and 41% of the species between one and two, four or six DNA extracts were shared in the location with low and medium diversity, respectively (Fig. 3b). More unique species were found in the separate DNA extracts for the location with medium diversity (71%), while 17% and 30% of the species was unique for the location with low and high diversity (Fig. 3c).

PCR replicates

The number of shared ASVs between the different PCR replicates originating from the same DNA extracts was low and ranged between 9%–38%, while 52%–88% of the ASVs were unique for one of the three PCR replicates (Fig. 4a). At the species level, 40%–76% of the detected species in each biological replicate was shared between the three PCR replicates (Fig. 4b), represented by 80% of the total reads. However, there was still a high number of unique species (17%–50%), but these species were represented by only 1% of the reads in the whole dataset. The number of ASVs and the number of assigned species differed significantly between locations with different diversity (Table 1) and biological replicates (Table 1).
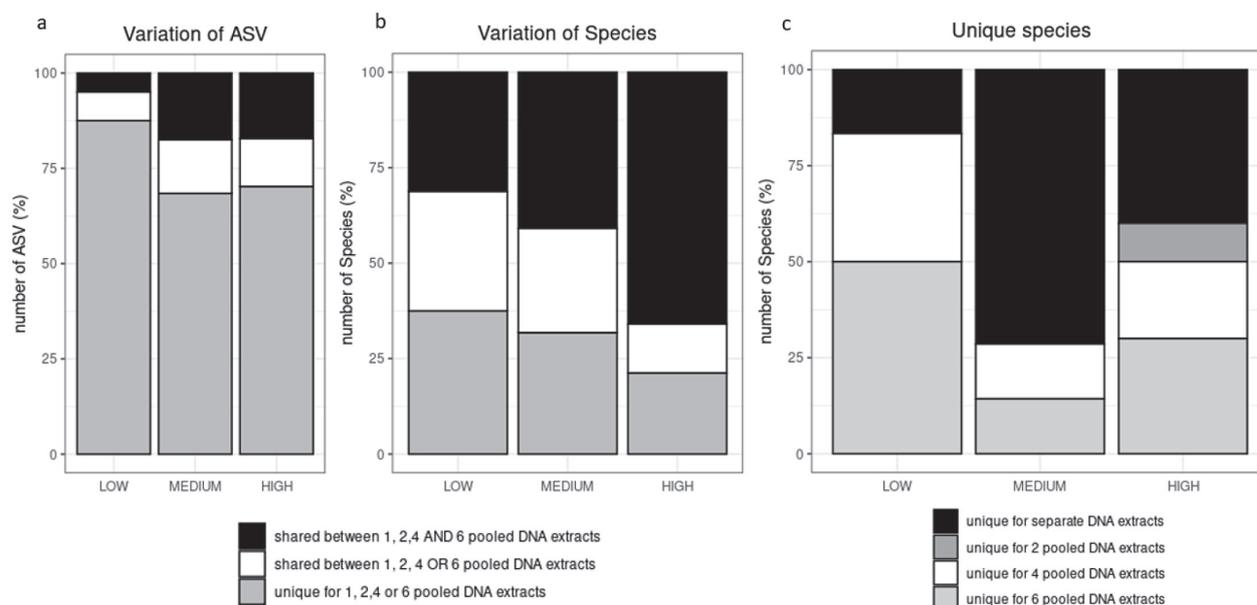
Figure 3. Variation of ASVs and species in number of pooled DNA extracts. The percentage of unique and shared ASVs (a) or species (b) between the separate and pooled DNA extracts are shown for the different locations. These percentages are calculated based on UpSet plots. Therefore, sets with the different replicates were made for the separate DNA extracts (1,2,3,4,5,6) and pooled DNA extracts: two (2a,2b,2c), four (4a,4b,4c) and six (6a,6b,6c). (c) The unique number of species in the separate and pooled DNA extracts (2, 4 or 6) for each location.

Finally, we assessed the effect of taking multiple PCR replicates on the detection of morphologically identified species. In the location with high diversity, 39 species were morphologically identified, of which 17 species were also found with metabarcoding. Most of them, 15 species, were found in each PCR replicate. In the location with low diversity, three species were morphological identified, of which one species was also found with metabarcoding. This species, *Limecola balthica*, was present in all three PCR replicates (Suppl. material 1: Fig. S11).

Accumulation plots show that no plateau was reached in the number of ASVs, even when three PCR replicates were taken (Suppl. material 1: Fig. S12). When looking at the species level, only a small increase is observed when using three PCR replicates instead of two. However, in one biological replicate in the location with high diversity (HIGH-A), the curve is still increasing when using three instead of two replicates (Suppl. material 1: Fig. S13). To pick up at least 80% of the species found in the three PCR replicates, one to two replicates are needed in the location with high diversity, while it varies from two to three replicates in the location with low diversity. A detailed table with the predicted number and percentage of species found in one, two or three PCR replicates can be found in Suppl. material 2: Table S5.

Betadiversity analysis of the DNA metabarcoding datasets

DNA replicates

PERMANOVA shows a significant effect of the diversity, as well as of the biological replicates on the community composition (Table 2). The homogeneity of dispersion

Table 2. Output PERMANOVA's. Results of the PERMANOVA's for beta diversity on species level. Crossed factors are indicated with an asterisk, while brackets are used for nested factors. The output of the permdisp is shown in the second column.

DNA replicates	Permdisp	Df	SumsOfSqs	F.Model	R2	Pr(>F)
Diversity	1e-04	2	5.5435	32.425	0.44624	**1e-04**
Diversity(Biological_ replicate)	4e-04	6	3.2889	6.4125	0.26475	**1e-04**
Residuals		42	3.5902		0.28901	
Total		50	12.4227		1.00000	
Pooled DNA extractions						
Diversity	1e-04	2	6.3483	29.0437	0.59102	**0.0001**
Nr_pooled	0.198	3	0.4361	1.3302	0.04060	0.1959
Diversity*nr_pooled		6	0.5689	0.8677	0.5297	0.6325
Residuals		31	3.3879		0.31541	
Total		42	10.7413		1.00000	

was significantly different between levels of diversity and between levels of biological replicates, but based on the plot, the centroids were clearly distinguishable from each other (Suppl. material 1: Fig. S2). The diversity and biological replicates explains 44% and 26% of the observed variation, respectively. The other 28% is explained by the residuals, here including the DNA replicates (Table 2). The nMDS plot using the Jaccard dissimilarity index illustrates clear clustering of the DNA replicates based on the diversity of the communities (Fig. 5). The Bray-Curtis index yielded identical results (Suppl. material 1: Fig. S14; Suppl. material 2: Table S7).

Pooled DNA extractions

The nMDS plot using the Jaccard dissimilarity index shows no separate clusters for the separate DNA extracts and two, four or six pooled DNA extracts (Fig. 6). The
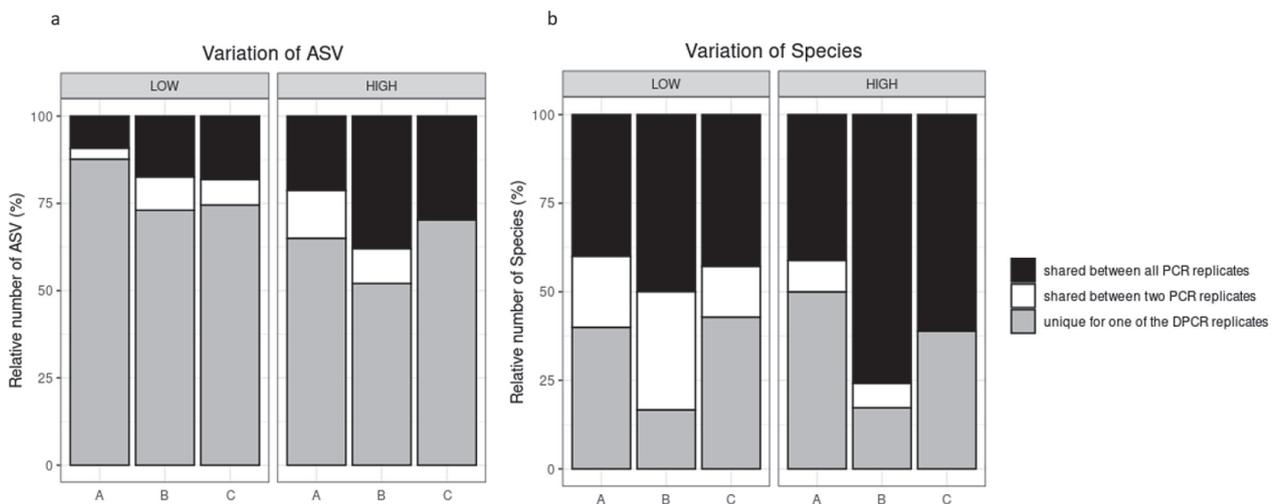
Figure 4. Variation of ASVs and species in PCR replicates. The average percentage of unique and shared ASVs (a) or species (b) between all three PCR replicates are shown for locations with low or high diversity. These percentages were calculated based on UpSet plots of the different PCR replicates in biological replicates and used to calculate the means of the different biological replicates in the different stations.
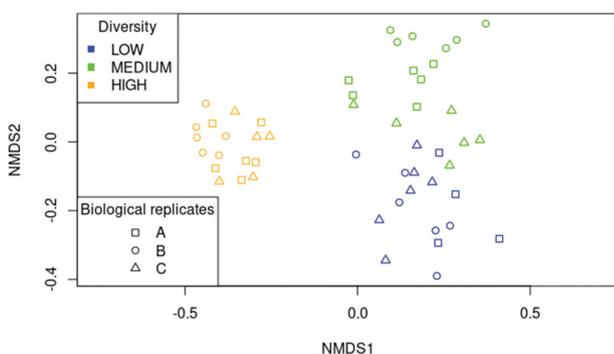

Figure 5. nMDS plot for DNA replicates, based on Jaccard dissimilarity index. Community composition between DNA replicates of the different biological replicates (marked with different symbols) and between the different locations with low, medium or high diversity (marked with different colors) are shown.
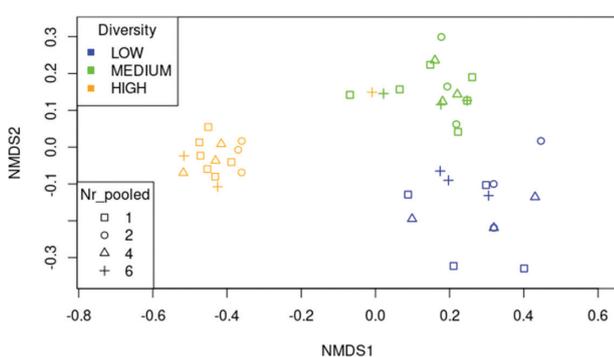
Figure 6. nMDS plot for pooled DNA extractions, based on Jaccard dissimilarity index. Community composition between separate DNA replicates and pooled DNA extractions (marked with different symbols) in the different locations with low, medium or high diversity (marked with different colors) are shown.

PERMANOVA test shows no significant interaction effect between diversity and pooled DNA extracts and a significant effect of diversity (Table 2). A significant difference

in dispersion between the levels of diversity (and between levels of number of pooled DNA extractions) was found, but the centroids on the plot were clearly distinguishable from each other, indicating that significant PERMANOVA results for this factor were not caused by dispersion effects (Suppl. material 1: Fig. S2).

Furthermore, 59% of the variation is explained by the stations, and only 4% by the number of pooled DNA extracts (Table 2). As such, there is a significant impact of the stations on the community composition, but not of the number of pooled DNA extracts (Table 2). The output of the Bray-Curtis index can be found in Suppl. material 1: Fig. S13 and Suppl. material 2: Table S7.

Discussion

DNA metabarcoding is a promising method to study diversity, providing that the impact of methodological choices on detection of macrobenthic species is known. Here, we provide empirical evidence on the importance of DNA and PCR replication for species detection in bulk DNA samples to contribute to a metabarcoding protocol that is robust and reliable, but also time and cost-efficient. First, our data showed that a huge portion of the amplicon sequence variants (ASVs) were unique for a replicate, but most of these unique ASVs did not receive taxonomic assignment and represented very few reads, indicating that these ASVs did not reflect the genetic diversity in the sample, but were most likely generated during the metabarcoding process. Second, our results showed that there was substantial variation between DNA and PCR replicates, but this variation was lower than the differences between biological replicates. Finally, when pooling two, four or six DNA extractions, a similar number of species and similar species composition were detected as in the separate DNA extractions.

High number of unique ASVs per sample does not show the genetic diversity, but reflects PCR or sequencing errors

In this study, the dada2 protocol was used for processing of the raw reads, generating ASVs. A high number of unique ASVs was observed for a specific sample, even when these ASVs were obtained from the same homogenous 'soup' (DNA replicates) or DNA extract (PCR replicates). Most of these unique ASVs did not get a good taxonomic assignment with the used reference database, as just a small part (13%–41% for the different biological replicates) was assigned at phylum or higher level. Aylagas et al. (2018) noted a high number of non-metazoan taxonomic assignments, but in this study, even when these unassigned ASVs were matched against the NCBI database using Blastn, only an additional 8% of the unique unassigned ASVs gave a reliable taxonomic species name. Similar observations were observed in another metabarcoding study for macrobenthos using various primer sets (Derycke et al. 2021), where the unassigned ASVs were explained as unknown diversity, genetic noise like PCR or sequencing errors, nuclear pseudogenes or aspecific amplification of other genomic regions.

Our study allows to rule out some of these explanations. First, if the unassigned ASVs would represent unknown biological diversity, we would expect to find these ASVs in multiple DNA replicates and especially in multiple PCR replicates. Between DNA replicates, a high number of ASVs (62%–86%) were uniquely found in one replicate (Fig. 2a). With our reference database, between 69% and 94% of these unique ASVs were not assigned at phylum level. Also for PCR replicates, the majority (52%–88%) of the ASVs were unique for one of the three PCR replicates from the same DNA extract (Fig. 4a). Between 55% and 92% of these unique ASVs were not assigned at phylum level with our reference database. After matching these ASVs against the NCBI database, still 83% of the unassigned unique ASVs did not receive a phylum level assignment. Second, if pseudogenes or other genomic regions would be amplified, we would expect this amplification to occur in more than one DNA replicate and in more than one PCR replicate, since the starting DNA pools are expected to be very similar for these technical replicates. We therefore conclude that most of these ASVs are probably randomly generated during the PCR or sequencing process, for example due to base substitution errors and DNA damage. These results indicate that using ASVs without taxonomic assignment leads to inadequate alpha diversity measurements and highlight the importance of taxonomic assignment for biologically meaningful diversity estimates using DNA metabarcoding.

DNA replicates from bulk DNA samples increase macrobenthos species detection

The need for DNA replication in sediment samples has already been demonstrated for several groups (Feinstein et al. 2009; Lejzerowicz et al. 2014; Brannock and Halanych 2015; Lanzen et al. 2017). In bulk samples, DNA of all species in the homogenous soup is assumed to be present in the subsample for DNA extraction (van der Loos and Nijland 2020). In our study, roughly 75% of species were detected in multiple DNA replicates, and therefore at least 25% of species are in danger of being missed in studies that do not use replicates (Fig. 2b). This could potentially be linked to an inefficient homogenization of the bulk organisms. However, macrobenthic specimens were homogenized with a mixer or a mortar and pestle; the sample was thoroughly shaken before taking subsamples for DNA extractions and a thick 'soup' was observed. Despite this homogenizing step, our results showed the need to take several DNA replicates when the aim is to detect as many species as possible.

Next to homogenization, larger and more abundant species have been assumed to be more easily detected due to the higher contribution of tissue in the 'soup', in contrast to smaller and rarer species (Deagle et al. 2018). If this is true, we expect that species that are consistently detected in all six DNA replicates would be large or more abundant than species found in only one replicate. This appeared not to be the case, as the distribution of size classes as well as of abundance classes were very similar between the two categories (Suppl. material 2: Table S8). Furthermore, this bias is thought to appear more in samples with higher diversity (Hollatz et al. 2017). In locations with high diversity, smaller species (size class < 21 mm) were indeed more frequently found uniquely in one replicate than in all six replicates (57% vs 28%, respectively), in contrast to the locations with medium (20% vs 67%, respectively) and low diversity (50% vs 83%, respectively). However, 50% of the species found in all DNA replicates were low abundant species (<5 specimens), while 63% of the species uniquely found in one DNA replicate had low abundances. This suggest that there is no clear tendency towards detecting more abundant species in more species rich communities.

To detect as many species as possible in the homogenous soup, multiple DNA replicates should be taken. One DNA replicate contained 45%–70% of the observed diversity, in line with the results of Lejzerowicz et al. (2014), where approximately 58% of total diversity was present in one replicate. In locations with high diversity, multiple replicates are needed to counteract for the bias in smaller animals, while in locations with low diversity, each species present has a high contribution to the total diversity, therefore, missing one species in this location can have a high impact.

Pooling DNA replicates does not change alpha and beta diversity patterns

The results of this study illustrate the need for DNA replicates, but if these replicates could be pooled, time and costs could be decreased. Our statistical results show that the number of species as well as the community composi-

tion observed in pooled extractions compared to separate DNA extractions were not significantly different, illustrating that pooling does not alter alpha and beta diversity patterns. However, a small number of species (9%–38%) is no longer detected in the pooled extractions compared to separate extracts (Fig. 3b). First, this can be explained by the number of sequencing reads per sample. A higher sequencing depth can improve the detection of targeted organisms and estimates of alpha diversity (Smith and Peay 2014; Alberdi et al. 2017). For example, in a study on six freshwater macroinvertebrate target taxa, five of them were detected with 10,000 reads, while at least 40,000 reads were needed to detect all six taxa (Nichols et al. 2019). In our study, we compared the species found in the pooled samples (with approximately 20,000 reads per pooled sample) and species found in the six separate samples (each with 20,000 reads, so 120,000 reads in total). With the higher sequencing depth for the species found in all the separate samples, there is an increased chance to detect even the rare species. Indeed, 77% of species that are not picked up in the pooled DNA extractions have a low abundance (<50 reads), so the low abundance of species detected only in separate DNA extracts can explain the variation between replicates. Second, as we always used the same volume for the PCR amplification, less volume is taken from the separate DNA extracts when pooling, resulting in lower DNA amounts of the rare species present in one of the DNA extracts.

Despite the small differences observed when pooling DNA extracts, pooling provides more robust outcomes and can compensate for inefficient DNA extraction. The metabarcoding protocol and the decision of whether or not to pool depends on the research question. For example, in environmental impact assessments, the most abundant species are important to detect, as these have the highest impact on the community patterns. In this case, pooling is suggested as it can decrease time and costs. When the protocol would be used for mapping the diversity in a certain area, the rare species are also important to detect. In this case, DNA replicates are better analyzed separately or when pooled should be sequenced with sufficient depth, to increase the chance of detecting as many species as possible.

PCR replicates from bulk DNA samples substantially differ in the detection of rare species

In this study, a high variation between the PCR replicates was detected for bulk DNA samples. Several plateaus were reached in the different species accumulation curves, but moderate numbers of species were shared between all PCR replicates. A first explanation for the variation between the PCR replicates could be the low abundance of multiple species, as most species (91%) found in only one replicate had low abundances (<50 reads). PCR amplification can introduce bias in the metabarcoding protocol, mainly due to the stochasticity of the process (Leray and Knowlton 2017). The first two to three cycles are the most important, meaning rare sequences have a lower chance to be picked up (Kebschull and Zador 2015), as observed in eDNA studies (Sato et al. 2017). Occupancy models, assessing the detection probability, showed that at least eight PCR replicates should be used for taxa with low detection probability in eDNA samples (Ficetola et al. 2015). However, these eDNA samples, which typically have low DNA amounts, are more vulnerable to stochasticity during the PCR amplification. In bulk DNA samples, a higher sequencing depth instead of more PCR replicates was suggested to increase the repeatability (Smith and Peay 2014). However, other studies on bulk DNA samples suggested a minimum of three PCR replicates (Bourlat et al. 2016; Leray and Knowlton 2017).

A second possibility why these species are not found in all three PCR replicates could be the primer efficiency. However, the used primer, the Leray primers set, has been shown to amplify most macrobenthic species of the North Sea (Derycke et al. 2021). When looking at the taxonomy of species present in only one replicate, some classes are highly present, like Polychaeta (43% of the species that are present in only one replicate). Because primer efficiency is low on sequences of Polychaeta, these sequences do not always get amplified (Carr et al. 2011).

Third, the PCR amplification conditions are important. Since the same cycling conditions are used for all PCR replicates, it seems unlikely that this would introduce variation between PCR replicates. Also the same PCR volume (25 µl) was used for the separate PCR replicates, but using small or bigger volumes can have an impact. Typically, 10 µL to 50 µL is used (Aylagas et al. 2014; Clarke et al. 2017; Elbrecht and Leese 2017). Increased reaction volumes are suggested with high DNA concentrations, for example DNA extracted from bulk samples as in this study, to minimize inhibition (Elbrecht and Leese 2017), resulting in higher costs. Although lower PCR volumes can increase pipetting errors, volumes of 6 µL have been successfully applied (Braukmann et al. 2019). For bacterial 16S (Minich et al. 2018), as well as for metazoan bulk samples (Buchner et al. 2021), no difference in species richness was observed between PCR volumes. However, using smaller PCR volumes can have an impact on the PCR stochasticity, resulting in a lower detection rate of rare species, as explained above.

In this study, variation between PCR replicates was observed, particularly due to the stochasticity and primer bias of the PCR amplification, similar as found in a study on marine artificial mock communities by Leray and Knowlton (2017). To detect as many species as possible, at least three PCR replicates should be used to counteract PCR stochasticity. However, the number of species uniquely found in one replicate (17%–50%) were represented by only 1% of the reads in the whole dataset. This indicated that for environmental impact assessments, which focus on the most abundant species, fewer PCR replicates could be used. To reduce the bias introduced by the PCR process, primer free approaches such as mitogenome enrichment followed by direct sequencing or

capture-by-hybridization probes have high potential as there is no impact of primer efficiency (Liu et al. 2016; Giebner et al. 2020) or PCR amplification bias, but these techniques are currently still too expensive for routine use in monitoring.

Comparison between morphological identification and the metabarcoding method

Metabarcoding and morphological identification showed comparable results for the most abundant species. High abundances were found for 52% of the species detected by both methods. However, approximately only half the number of species detected by morphological identification were also picked up by metabarcoding. First, this can be explained by the different resolution of both methods. With the morphological identification, 15 specimens did not have a taxonomic assignment to species level, but only to genus, family or class level, while the genetic method was able to assign eight out of these 15 specimens to species level (Suppl. material 2: Table S6). Second, often an incomplete reference database is said to be responsible for the decreased detectability (Aylagas et al. 2018), as often a high number of sequences lack a taxonomic assignment (Leray and Knowlton 2015; Aylagas et al. 2018). Of the 26 species that were only found in the morphological samples, only six were not present in the reference database, highlighting that most of these missed morphologically identified species are not only due to an incomplete reference database. Third, in a previous study, we suggested that low abundances and primer efficiency are also part of the reason that some morphological species are not detected in the metabarcoding dataset (Derycke et al. 2021). Indeed, most of the species only found by the morphological method have low abundances (<3 counts). Yet, some species, mainly within Polychaeta, have higher abundances, e.g. *Notomastus latericeus* was counted 17 times. It has been shown that primer efficiency is low for Polychaeta, due to the high variation in the COI gene, and therefore these sequences do not always get amplified (Carr et al. 2011).

Conclusion

This study highlighted the importance of technical replicates. First, taking into account the financial constraints and additional time needed with every DNA extraction, our data show that limited gain is achieved when conducting more than three DNA replicates for macrobenthic metabarcoding studies of medium and high diverse locations. However, when sampling in a low diverse environment, taking more replicates should be considered. Second, for environmental impact assessments, the protocol can be made more time and cost efficient by pooling the separate DNA extracts without a valuable loss in species detection. Third, substantial variation between PCR replicates was observed in this study. As numerous papers still lack PCR replicates, we want to emphasize the importance of taking at least three PCR replicates, the maximum tested replicates in this study, when investigating species diversity. These results contribute to create time and cost efficient metabarcoding protocols for environmental impact assessments, so yielding reliable results.

Data accessibility statement

The sequencing datasets and corresponding metadata generated for this study can be found as BioProject under accession number PRJNA683870 in the online NCBI repository https://www.ncbi.nlm.nih.gov/. The R script to reproduce the findings of this study can be found at https://gitlab.com/Lvandenbulcke/dna-metabarcoding-technical-replicates-and-pooling.

References

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K, Mahon A (2017) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9: 134–147. https://doi.org/10.1111/2041-210X.12849

Aylagas E, Borja Á, Irigoien X, Rodríguez-Ezpeleta N (2016) Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. Frontiers in Marine Science 3: e96. 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. Ecological Indicators 95: 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044

Aylagas E, Borja Á, Rodríguez-Ezpeleta N (2014) Environmental Status Assessment Using DNA Metabarcoding: Towards a Genetics Based Marine Biotic Index (gAMBI). PLoS ONE 9: e90529. https://doi.org/10.1371/journal.pone.0090529

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

Bik HM, Halanych KM, Sharma J, Thomas WK (2012) Dramatic Shifts in Benthic Microbial Eukaryote Communities following the Deepwater Horizon Oil Spill. PLoS ONE 7(6): e38550. https://doi.org/10.1371/journal.pone.0038550

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

Bourlat SJ, Q. H, Finnman J, Leray M (2016) Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Dual-PCR Method. In: Bourlat S (Ed.) Marine Genomics Methods in Molecular Biology. Humana Press, New York, NY, 197–207. https://doi.org/10.1007/978-1-4939-3774-5_13

Brannock PM, Halanych KM (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 TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, de Waard JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources 19: 711–727. https://doi.org/10.1111/1755-0998.13008

Breine NT, 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. Estuarine, Coastal and Shelf Science 214: 173–184. https://doi.org/10.1016/j.ecss.2018.09.012

Buchner D, Beermann AJ, Leese F, Weiss M (2021) Cooking small and large portions of "biodiversity-soup": Miniaturized DNA metabarcoding PCRs perform as good as large-volume PCRs. Ecology and Evolution 11(13): 9092–9099. https://doi.org/10.1002/ece3.7753

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Carr CM, Hardy SM, Brown TM, Macdonald TA, Hebert PD (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

Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Frontiers in Ecology and the Environment 8: 233–238. https://doi.org/10.1890/090115

Clarke LJ, Beard JM, Swadling KM, Deagle BE (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution 7: 873–883. https://doi.org/10.1002/ece3.2667

Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33: 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Daily GC, Polasky S, Goldstein J, Kareiva PM, Mooney HA, Pejchar L, Ricketts TH, Salzman J, Shallenberger R (2009) Ecosystem services in decision making: time to deliver. Frontiers in Ecology and the Environment 7: 21–28. https://doi.org/10.1890/080025

Darling JA, Galil BS, Carvalho GR, Rius M, Viard F, Piraino S (2017) Recommendations for developing and applyng genetic tools to assess and manage biological invasions in marine ecoystems. Marine Policy 85: 54–64. https://doi.org/10.1016/j.marpol.2017.08.014

Deagle BE, Clarke LJ, Kitchener JA, Polanowski AM, Davidson AT (2018) Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. Molecular Ecology Resources 18: 391–406. https://doi.org/10.1111/1755-0998.12740

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. Frontiers in Marine Science 8: 637858. https://doi.org/10.3389/fmars.2021.637858

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

Dopheide A, Xie D, Buckley TR, Drummond AJ, Newcomb RD, Bunce M (2018) Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution 10: 120–133. https://doi.org/10.1111/2041-210X.13086

Duncan C, Thompson JR, Pettorelli N (2015) The quest for a mechanistic understanding of biodiversity-ecosystem services relationships. Proceedings of the Royal Society B – Biology Sciences 282: 20151348. https://doi.org/10.1098/rspb.2015.1348

Edward WB (1984) Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. Advances in Ecological Research 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. Frontiers in Environmental Science 5: e11. 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 LM, Sul WJ, Blackwood CB (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

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Raye G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol Ecol Resour 15(3): 543–556. https://doi.org/10.1111/1755-0998.12338

Giebner H, Langen K, Bourlat SJ, Kukowka S, Mayer C, Astrin JJ, Misof B, Fonseca VG (2020) Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. Molecular Ecology Resources 20: 1333–1345. https://doi.org/10.1111/1755-0998.13201

Goodwin KD, Thompson LR, Duarte B, Kahlke T, Thompson AR, Marques JC, Cacador I (2017) DNA Sequencing as a Tool to Monitor Marine Ecological Status. Frontiers in Marine Science 4: e107. https://doi.org/10.3389/fmars.2017.00107

Hering D, Borja A, Jones JI, 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 op-

tions for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

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

Jaccard P (1912) The distribution of the flora in the alpine zone.1. New Phytologist 11: 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Kebschull JM, Zador AM (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Research 43(21): e143. https://doi.org/10.1093/nar/gkv717

Lanzen A, Lekang K, Jonassen I, Thompson EM, 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 Research Part II: Topical Studies in Oceanography 108: 17–26. https://doi.org/10.1016/j.dsr2.2014.07.018

Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J (2015) High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Reports 5: e13932. https://doi.org/10.1038/srep13932

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences of the United States of America 112: 2076–2081. https://doi.org/10.1073/pnas.1424997112

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 JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology 10: e14. https://doi.org/10.1186/1742-9994-10-34

Liu S, Wang X, Xie L, Tan M, Li Z, Su X, Zhang H, Misof B, Kjer KM, Tang M, Niehuis O, Jiang H, Zhou X (2016) Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. Molecular Ecology Resources 16: 470–479. https://doi.org/10.1111/1755-0998.12472

Lobo J, Shokralla S, Costa MH, Hajibabaei M, Costa FO (2017) DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. Scientific Reports 7: e15618. https://doi.org/10.1038/s41598-017-15823-6

Martins FMS, Galhardo M, Filipe AF, Teixeira A, Pinheiro P, Pauperio J, Alves PC, Beja P (2019) Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples

for freshwater biomonitoring. Molecular Ecology Resources 19: 863–876. https://doi.org/10.1111/1755-0998.13012

Minich JJ, Humphrey G, Benitez RAS, Sanders J, Swofford A, Allen EE, Knight R (2018) High-Throughput Miniaturized 16S rRNA Amplicon Library Preparation Reduces Costs while Preserving Microbiome Integrity. Msystems 3: e00166-18. https://doi.org/10.1128/mSystems.00166-18

Nichols SJ, Kefford BJ, Campbell CD, Bylemans J, Chandler E, Bray JP, Shackleton M, Robinson KL, Carew ME, Furlan EM (2019) Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. Freshwater Biology 65: 607–620. https://doi.org/10.1111/fwb.13443

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014) Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecology Resources 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 MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Jones JI, Sagova-Mareckova M, Moritz C, Barquin J, Piggott JJ, 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 (2018) The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. Science of the Total Environment 637: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Sato H, Sogo Y, Doi H, Yamanaka H (2017) Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports 7: e14860. https://doi.org/10.1038/s41598-017-14978-6

Smith DP, Peay KG (2014) Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. PLoS ONE 9: e90234. https://doi.org/10.1371/journal.pone.0090234

van der Loos LM, Nijland R (2020) Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology 30(13): 3270–3288. https://doi.org/10.1111/mec.15592

Van Hoey G, Borja A, Birchenough S, Buhl-Mortensen L, Degraer S, Fleischer D, Kerckhof F, Magni P, Muxika I, Reiss H, Schroder A, Zettler ML (2010) The use of benthic indicators in Europe: from the Water Framework Directive to the Marine Strategy Framework Directive. Marine Pollution Bulletin 60(12): 2187–2196. https://doi.org/10.1016/j.marpolbul.2010.09.015

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73: 5261–5267. https://doi.org/10.1128/AEM.00062-07

Supplementary material 1
Figures S1–S14

Author: Laure Van den Bulcke, Annelies De Backer, Bart Ampe, Sara Maes, Jan Wittoeck, Willem Waegeman, Kris Hostens, Sofie Derycke

Data type: zip. archiv (PNG. and JPG. images)

Explanation note: ESM Figure S1. Map of the Belgian Part of the North Sea (BPNS) with the three sampling locations. ESM Figure S2. Multivariate homogeneity of group dispersions (variances). ESM Figure S3. Mean number of reads and ASVs. ESM Figure S4. Taxonomic assignment of ASVs. ESM Figure S5. Mean number of ASVs and species. ESM Figure S6. Size and abundance distribution. ESM Figure S7. Accumulation plots for DNA replicates on ASV level. ESM Figure S8. Accumulation plots for DNA replicates on species level. ESM Figure S9. UpSet plots for comparison of morphological identified samples for DNA replicates. ESM Figure S10. Observed vs expected number of species in different number of pooled DNA extracts. ESM Figure S11. Upset plots for comparison of morphological identified samples for PCR replicates. ESM Figure S12. accumulation plots based on ASVs for PCR replicates. ESM Figure S13. accumulation curves based on species for PCR replicates. ESM Figure S14. nMDS plot for DNA replicates (a) and pooled DNA extractions (b) based on Bray-Curtis.

Link: https://doi.org/10.3897/mbmg.5.71107.suppl1

Supplementary material 2
Tables S1–S8

Author: Laure Van den Bulcke, Annelies De Backer, Bart Ampe, Sara Maes, Jan Wittoeck, Willem Waegeman, Kris Hostens, Sofie Derycke

Data type: zip. archiv (PDF. files)

Explanation note: ESM Table S1. Used primerset. ESM Table S2. Abundanties Morphological identified sample. ESM Table S3. Read numbers. ESM Table S4. Unassigned ASVs. ESM Table S5. Output UpSetPlots. Prediction of the number of species when using x number of replicates from the species accumulation method. This finds the expected richness following Coleman et al. 1982. ESM Table S6. Unique species genetic identified sample, with comparable species from the morphological identified sample. ESM Table S7. Table with results of PERMANOVA for DNA replicates and pooled DNA extractions with the Bray-Curtis index. Table S8. The species picked up in all six DNA replicates (A) and unique for one of the DNA replicates (B) are listed. The corresponding size class is given in column F. If the species is also found in the morphological identified sample, the abundance (counts) is given.

Link: https://doi.org/10.3897/mbmg.5.71107.suppl2