Capture by hybridization for full-length barcode-based eukaryotic and prokaryotic biodiversity inventories for deep sea ecosystems

Babett Günther[1], Sophie MARRE[2], Clémence Defois[2], Thomas Merzi[3], Philippe Blanc[3], Pierre PEYRET[2], and Sophie Arnaud-Haond[1]

[1]Ifremer
[2]Clermont Auvergne University
[3]Total SE

April 21, 2021

Abstract

Biodiversity inventory remains limited in marine systems due to unbalanced access to the three ocean dimensions. The use of environmental DNA (eDNA) for metabarcoding allows fast and effective biodiversity inventory and is forecast as a future biodiversity research and biomonitoring tool. However, in poorly understood ecosystems, eDNA results remain difficult to interpret due to large gaps in reference databases and PCR bias limiting the detection of some major phyla. Here, we aimed to circumvent these limitations by avoiding PCR and recollecting larger DNA fragments to improve assignment of detected taxa through phylogenetic reconstruction. We applied capture by hybridization (CBH) to enrich DNA from deep-sea sediment samples and compared the results with those obtained through an up-to-date metabarcoding PCR-based approach (MTB). Originally developed for bacterial communities by targeting 16S rDNA, the CBH approach was applied to 18S rDNA to improve the detection of species forming benthic communities of eukaryotes, with particular focus on metazoans. The results confirmed the possibility of extending CBH to metazoans with two major advantages: i) CBH revealed a broader spectrum of prokaryotic, eukaryotic, and particularly metazoan diversity, and ii) CBH allowed much more robust phylogenetic reconstructions of full-length barcodes with up to 1900 base pairs. This is particularly important for taxa whose assignment is hampered by gaps in reference databases. This study provides a database and probes to apply 18S CBH to diverse marine systems, confirming this promising new tool to improve biodiversity assessments in data-poor ecosystems like those in the deep sea.

1. Introduction

Over two-thirds of the Earth is covered by oceans, likely sheltering a high level of still poorly studied biodiversity, particularly in the deep sea (Costello & Chaudhary, 2017; Costello, Cheung, & De Hauwere, 2010). Today, molecular approaches provide nonintrusive methods to study the diversity of marine environments, even those that are hardly accessible to sampling. The analysis of environmental DNA (eDNA; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012) represents a promising path to inventory biodiversity and sets the ground for the development of molecular biomonitoring protocols (Andruszkiewicz et al., 2017; Apothéloz-Perret-Gentil et al., 2017; Bohan et al., 2017; Cordier et al., 2018; Derocles et al., 2018). Approaches based on eDNA (air, ground, sediment, or water, relatively easy to access and sample) target the genetic material present in the environment (Bohmann et al., 2014; Thomsen & Willerslev, 2015), allowing us to unravel the nature of macro- and microorganisms present in the surrounding habitats. First developed for the uncultivable majority that represents the microbial world (Xu, 2006), metabarcoding approaches, relying on PCR-based amplicon sequence identification combined with high-throughput sequencing, were transferred to eukaryotes early on (Creer et al., 2010; Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011; Taberlet et al., 2012; Valentini, Pompanon, & Taberlet, 2009).

Over the last decade, metabarcoding protocols have been improved, from sampling up to bioinformatic steps, to optimize their resolution and interpretation. Nevertheless, biomonitoring and biodiversity inventory using metabarcoding are challenging (Miya et al., 2015; Yamamoto et al., 2017) for two main reasons. First, this method relies on PCR-based DNA enrichment, suffering biases due to unequal amplification across taxa (PCR bias) and artifacts (PCR errors prominent to sequence errors), leading to biased biodiversity inventories (Acinas, Sarma-Rupavtarm, Klepac-Ceraj, & Polz, 2005; Kanagawa, 2003; Sefc, Payne, & Sorenson, 2007; Smyth et al., 2010). Second, the evolution of high-throughput sequencing technologies available on the market led to higher yield but shorter sequencing fragments, limiting the use of metabarcoding to short fragments. Such short fragments (usually 150 to 450 base pairs) lead to a less reliable assignment of sequences to taxa and hamper the use of data produced for comprehensive phylogenetic reconstructions. This is particularly limiting in ecosystems where biodiversity is poorly described and reference databases contain large gaps, for which many unassigned sequences can correspond to existing undescribed biodiversity, yet teasing them apart from spurious sequences would require phylogenetic reconstruction. The limiting factor for taxonomic assignment of deep sea organisms is the general lack of sequence references in marine systems. Some major groups, such as nematodes, which are the most abundant and diverse benthic metazoan taxa, can rarely be identified genetically (Dell'Anno, Carugati, Corinaldesi, Riccioni, & Danovaro, 2015; Gambi & Danovaro, 2016). Thus, long, high-quality barcode libraries are needed to improve taxonomic identification in general, especially for poorly known groups.

Theoretically, direct metagenomic sequencing (such as shotgun sequencing) could solve these limitations, as these sequences can also be reconstructed from eDNA to obtain a comprehensive overview of the taxonomic diversity of the studied community, free of PCR bias (Porter & Hajibabaei, 2018) and allowing reliable phylogenies based on long fragments. However, the production of metagenomes is still extremely costly, leading to a dominance of prokaryotic sequences, and the*de novo* reconstruction of comprehensive metagenomes is highly time consuming; differentiating between biological differences and sequencing errors is hardly possible and highly limited by gaps in reference databases (Ghurye, Cepeda-Espinoza, & Pop, 2016; Quince, Walker, Simpson, Loman, & Segata, 2017).

As an intermediate, less expensive option, to avoid the two main limitations associated with metabarcoding, two other methods of DNA enrichment are available (Mamanova et al., 2010; Mertes et al., 2011): the molecular inversion probe (MIP) and capture by hybridization (CBH). CBH exists in two variations, "on-array capture" on a solid microarray or "in-solution capture", which takes place within a fluid medium (Gasc, Peyretaillade, & Peyret, 2016). Here, we use the latter, which was first described by Gnirke et al. (2009) for human exome resequencing, whereby hybrid probes are designed to enrich genomic DNA. While the initial cost of this system is high, by multiplexing libraries, efficient sequencing of several samples (up to 96-well plates) has been shown to be highly efficient (Meyer & Kircher, 2010). Moreover, a diversity of probes (single-stranded sequences of DNA) designed in different locations of the target gene regions allows capturing a much wider diversity and recovering long fragments, thereby improving taxonomic assignment and allowing reasonable phylogenetic reconstruction (Denonfoux et al., 2013; Gasc & Peyret, 2018). Furthermore, a low concentration DNA template is sufficient, allowing this method to be successfully used in low biomass environments (such as air or deep-sea biomes) wherein generally lower DNA concentrations are obtained, as in deep oligotrophic aquifers (Ranchou-Peyruse et al., 2017). The first test using eDNA showed that a 100-fold lower concentration can be detected with CBH than with traditional methods (Seeber et al., 2019), while others mentioned reduced tractability for DNA with less than 0.1 ng of total gDNA (Wilcox et al., 2018). Testing within complex prokaryotic communities even allowed the detection of extremely rarely represented members (less than 0.0001%) (Gasc & Peyret, 2018). It has been suggested that the final success of this method depends strongly on the probes (Ribière et al., 2016) rather than on the initial biomass and DNA concentration.

Improved biodiversity assessments can thus be expected using CBH, (i) avoiding PCR steps, yet targeting a broader range of biodiversity in a single reaction by using a comprehensive and versatile set of probes and (ii) reconstructing long fragments for full barcode regions, allowing reliable phylogenetic positioning and reconstruction. In recent years, this methodology has proven to markedly improve microbial diversity inventories with precise taxonomic affiliation at the species level (Gasc & Peyret, 2018). CBH was also applied to recover full-length microbial eukaryotic cDNAs in complex environmental samples (Bragalini et al., 2014) and to directly capture long DNA fragments (Gasc & Peyret, 2017). Additionally, CBH using mitochondrial barcodes for inventories of metazoans in bulk or ethanol-preserved samples resulted in a highly accurate census of species (Gauthier et al., 2020; Shokralla et al., 2016), and similar results were obtained when testing the detection of a broad range of metazoans, including mammals, from aquatic and sediment eDNA samples (Seeber et al., 2019; Wilcox et al., 2018). Additionally, CBH represents a promising path for phylogenetic studies, as recently shown for butterflies (Kawahara et al., 2018).

With this study, we aimed to assess the potential of 16S and 18S rDNA enrichment by CBH coupled with high-throughput sequencing to explore the biodiversity of prokaryotes and eukaryotes, including metazoans, in the deep sea (˜500-2800 m depth). We analyzed eDNA samples extracted from sediment to compare CBH with metabarcoding for the V4 16S rDNA region and the V1-V2 18S rDNA region.

2. Materials and methods

2.1. Samples

The study included ten sediment samples from different deep-sea areas (Appendix 1, Table 1), totaling five stations: four in the Atlantic Ocean and one in the Mediterranean Sea (see Fig. 1, Table 1). Standardized sampling was performed using a Multicorer or Pushcorer (both with cores of 20 cm^2) and slicing under sterile conditions. For this study, only the first two layers of sediment (0-1 cm and 1-3 cm) were used for each station. Samples were frozen at -80°C immediately after recovery and slicing for later extraction. DNA extraction was performed using ˜10 g of substrate with the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) following the manufacturer's protocols. Extraction controls were included (empty sampling bags conditioned onboard and rinsed with water and manufacturer tubes initially filled with sterile water instead of sediment). Final extracts were concentrated three times with 70% ethanol for precipitation reaction and diluted in molecular grade water.

2.2. Metabarcoding

Metabarcoding was performed at the National Center of Sequencing (Genoscope, Paris, France) following protocols published by Brandt et al. (2019). PCR was performed on two barcoding gene regions (Table 2) targeting prokaryotes (V4 16S rDNA region) or eukaryotes (V1-V2 18S rDNA region). Each 25 µl PCR contained Phusion® High-Fidelity PCR 2X Master Mix (GC Buffer for 18S and HF Buffer for 16SV4; New England Biolabs, Ipswich, MA US), 0.5 µM of each primer, and approximately 10-20 ng of DNA template and were filled to volume with molecular water. PCR cycling conditions were 98°C for the 30 s, followed by a 30 (for 18S) or 29 (for 16SV4) cycles of 98°C for 10 s, annealing at 50°C for 45 s, 72°C for 60 s, and final elongation at 72°C for 10 min. Each library was prepared in triplicate PCR with a Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA), and triplicates were pooled for sequencing with the HiSeq technology with the 2x250 bp reading mode.

2.3. Hybridization capture approach

2.3.1. Design of capture probes and in silico tests

Previously, 16S rDNA hybridization capture was successfully enforced to describe the archaeal and bacterial diversity of soil microbial communities. Here, we applied this approach to deep-sea samples by adding 18S rDNA gene capture to specifically target eukaryotes. Probes were designed to target 16S and 18S rDNA genes to ensure specificity against the targeted genes while allowing the detection of sequence variants not yet described in databases by selecting degenerate probes (Table 3). Probes targeting 16S rDNA were previously published (Gasc & Peyret, 2018). In this work, probes targeting 18S rDNA were determined as described in a previous publication. Briefly, probes were designed using KASpOD (Parisot, Denonfoux, Dugat-Bony, Peyret, & Peyretaillade, 2012) and PhylArray algorithms (Militon et al., 2007) and a custom curated database derived from all 18S rDNA sequences from the EMBL and similarly built for PhyloPDb development (Jaziri et al., 2014). Selection of probes distributed over the entire length of the gene and their length (31 to 50 bp) allow specific hybridization with their target even if mismatches are present. For each set of probes, coverage among SSU sequences from the SILVA database (release 132; Quast et al., 2013) was predicted by mapping probes with BBDuk (k=13, copy undefined mode, mm=f; *http://jgi.doe.gov/data-and-tools/bb-tools/*). Finally, adaptors containing the T7 promoters "ATCGCACCAGCGTGT" and "CACTGCGGCTCCTCA" were added to the 5' and 3' ends of probes, respectively, to enable PCR amplification (Ribière et al., 2016). These probes were designated capture probes.

2.3.2. Capture and sequencing

Illumina libraries were constructed from eDNA samples using the Nextera XT (for 9 samples) or TruSeq (for 1A) Kits (Illumina). Biotinylated RNA capture probes were obtained by *in vitro* transcription following the protocol described by Ribière et al. (2016). Hybridization capture was carried out independently for each sample with 16S and 18S rDNA probes at a ratio of 50/50, as described by Gasc and Peyret (2018). In brief, 500 ng of sequencing library mixed with 2.5 µg of salmon sperm DNA was incubated with 500 ng of biotinylated probes in a hybridization buffer (10 × SSPE, Denhardt's 10X solution, 10 mM EDTA and 0.2% SDS) for 24 hours at 65°C. After hybridization, the probe/target heteroduplexes were captured with 500 µg of paramagnetic beads coated with streptavidin (Dynabeads M-280 Streptavidin, Invitrogen). Beads were collected using a magnetic stand (Ambion), washed once with 500 µl of 1x SSC/0.1% SDS buffer, and then three times with 500 µl of 0.1 × SSC/0.1% SDS buffer preheated at 65°C. The captured DNA fragments were eluted with 50 µl of 0.1 M NaOH and transferred to a sterile tube containing 70 µl of 1 M Tris-HCl pH 7.5 buffer. Captured libraries were finally PCR-amplified with 25 cycles using primers fully complementary to Illumina adapters. To increase the enrichment efficiency, the second round of capture was performed from the first-round capture products. Captured libraries were then sequenced with two Illumina MiSeq 2x300 bp runs.

2.4. Sequencing data processing and analysis

Metabarcoding

The bioinformatic pipeline applied is described in Brandt et al. (2019). Primers and leftover adapters were removed with the program Cudadapt (Martin, 2011). The basic pipeline was then based on DADA2 using stringed error correction (Callahan et al., 2016), including fragment size selection with an expected length of 300-500 bp for both 16SV4 and 18SV1V2, followed by a chimera removal step. Sequences were then clustered into operational taxonomic units (OTUs) using the program swarm2 (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2015) with an iterative local threshold (d) of 4 for 18SV1V2 and 1 for 16SV4. The ribosomal sequences were taxonomically assigned against the Silva database (release 132; Pruesse et al., 2007) with the MegaBlast algorithm (Camacho et al., 2009).

A decontamination step was applied after clustering, using extraction and PCR negative controls using decontam (Davis, Proctor, Holmes, Relman, & Callahan, 2018), by the prevalence method with a threshold of 0.8. Finally, all OTU counts were adjusted using an R-based script (Wangensteen, Palacín, Guardiola, & Turon, 2018) to renormalize potential tag switches during library preparations (Schnell, Bohmann, & Gilbert, 2015).

2.4.2. Capture

Raw reads were trimmed for Illumina adapters using Trimmomatic (v 0.38; Bolger, Lohse, & Usadel, 2014) and then quality-filtered with PRINSEQ-lite PERL script (min_qual_mean =25, trim_qual_window=3, trim_qual_step=1, min_len=60; Schmieder & Edwards, 2011). Trimmed reads corresponding to rDNA were extracted using SortMeRNA (v2.1; Kopylova, Noé, & Touzet, 2012) with default parameters. Near-full-length 16S and 18S rDNA sequences were reconstructed using EMIRGE software (v 0.60; Miller, Baker, Thomas, Singer, & Banfield, 2011) and the emirge_amplicon.py script. This tool allows reference-based assembly of reads while allowing the reconstruction of distant variants. The database used was SILVA 132 SSURef NR99, including fragments with lengths from 1200-2000 bp. The parameters used were join_threshold fixed to 1 and 120 iterations. Only sequences longer than 800 bp were kept. Taxonomic affiliation was performed using the plugin “feature-classifier sklearn classifier” from QIIME2 (v. 2019.1; Bokulich et al., 2018; Bolyen et al., 2019) and the full-length SILVA 132 database, with the p-confidence set to 0.7. This type of analysis is further referred to as CBH-long.

Additionally, Kraken2-based analysis (Wood, Lu, & Langmead, 2019; Wood & Salzberg, 2014) was performed starting from paired reads to evaluate all captured diversity (without gene reconstruction), as too low coverage of some taxa could hinder the possibility of reconstructing longer sequences and thus cause the lack of these taxa in the final dataset. The database used was the prepackaged SILVA database provided by Kraken2. We tested the confidential score from 0.0 to 1.0 with 0.1 steps. For the final analyses, a score of 0.7 was retained, ensuring good specificity of taxonomic affiliation. This is in line with a previous report that values from 0.6 up to 0.7 indicated the best results for sensitivity and precision (Wood & Salzberg, 2014). Data related to these analyses are further mentioned as CBH-short.

2.5. Formatting data and statistics

Although the same reference databases were used, the taxonomic ranking system, such as phylum and class, did not deliver comparable formats in terms of taxonomic descriptions. Therefore, each individual identified was assigned to its taxonomic ID lineage based on the NCBI taxonomy database (Sayers et al., 2020) to compare the same taxonomic levels among methods.

Similarity analyses were performed with Vegan (Oksanen et al., 2008) using the Bray-Curtis similarity test and similarities between methods were compared by a Mantel test. To compare relative abundance data, the count of sequences was standardized to the total number of reads per sample (Hellinger, decostand).

3. Results

3.1. Probes for CBH

In addition to the fifteen 16S rDNA probes available, seventeen explorative probes targeting 18S rDNA were designed to recover a broad range of eukaryotes, with a special emphasis on metazoans. Probe positions within 16S and 18S rDNA genes are visualized in Appendix 2. The coverage of the two sets of probes, determined *in silico* , reached 99.99% for 16S rDNA and 99.98% for 18S rDNA of the 369953 and 55145 targeted taxa present in the SILVA SSU NR99 (version 132; see Appendix 3), indicating a high potential sensitivity of probes for the recovery of prokaryotic and eukaryotic diversity.

3.2. Sequencing output and fragment length

CBH resulted in 2,231,139 to 12,799,906 reads per sample. After the first filtering and trimming steps, approximately 7 to 34% were identified as rDNA (general sequencing output given in Appendix 4). For MTB, a total of 288,094 to 1,278,400 sequences were obtained per sample, lading 2231 to 4133 OTUs for 16SV4 and 770 to 2169 for 18SV1V2 per sample. The CBH data were analyzed by two different pipelines, first direct use of the short fragments (hereafter referred to as CBH-short) with Kraken 2, using the unaligned reads with a mean length of 200 to 289 bp, which was shorter than the fragments of up to 450 bp obtained with metabarcoding. Second, EMIRGE was used to reconstruct “full barcodes” (hereafter named CBH-long) allowed the reconstruction of fragments of on average 731 near-full length markers per sample, reaching up to 1200 to 1450 bp for archaea, up to 1600 bp for bacteria and 1200 to 1900 bp for eukaryotes (Fig. 2). However, for a small number of taxa, 60 to 95% of sequences were lost.

3.3. Comparing general diversity

As expected, the taxonomic assignment led to different outcomes depending on the analyses, as metabarcoding results were presented as OTUs, CBH-long results as reconstructed sequences, and CHB-short results as grouped (not clustered) sequences for the same taxa. Interestingly, the detection trends of taxa and phyla among the methods varied across the three analyzed kingdoms, Archaea, Bacteria and Eukaryota (synoptic visualization in Fig 3, sample-wise data listed in Appendix 5).

In the case of Archaea CHB-long, a slightly higher number of taxa was revealed than with CBH-short or

MTB, but with one missing phylum (DPANN group). Most of this difference came from the Tack group, with 419 taxa identified by CBH-short, 869 taxa identified using long sequences and only 166 taxa identified with MTB. For bacteria, CBH shows generally higher diversity detection; CBH-short detected most phyla, with 18 and 11 using CBH-long and 12 with MTB. Some rare phyla were only detected by CBH-long or MTB, but all were detected by CBH-short. Additionally, over 1200 more taxa were revealed via full-length barcodes than with OTUs by MTB. More importantly was the difference in the displayed diversity itself; over 50 percent of bacteria identified with CBH belonged to Proteobacteria (3588 taxa with CBH-long; 10520 taxa with CBH short), and only 25% (1228 taxa) belonged to Proteobacteria with MTB.

In contrast, using CBH for eukaryotes showed opposite trends. Here, most detections of phyla and other taxa were obtained by MTB. Overall, less than 2% of the reconstructed fragments (n=267) were identified compared to those with metabarcoding (OTU n=13794), and only six out of 18 phyla were identified, whereas CBH-short still detected 14 phyla (n=4571). However, these newly designed 18S probes mostly focused on metazoans, as part of Opisthokonta; here, CBH showed a relatively higher percentage of detections. With a total of 2527 taxa identified using metabarcoding, 2089 were CBH short, and 181 were full-length sequences. Finally, the overall identification and displayed biodiversity varied in the amount and abundance of phyla, depending on the kingdom/gene region as well as the analytical method.

3.3. CBH for metazoans

The identification of taxa regarding their phyla in the NCBI taxonomy is presented in Fig. 4. As expected, generally higher diversity was detected by CBH-short (Kraken2), with 21 phyla, compared to 16 with metabarcoding. Some groups, Acanthocephala, Cycliophora, Entoprocta, Gnathostomulida, Micrognathozoa, and Nematomorpha, were only detected with CBH, while only one group was specifically detected by metabarcoding (Ctenophora). The greatest difference was in the number of detected nematodes, with 1074 (MTB) to 82 (CBH-short) taxa. Analyzing sample-specific results showed in more detail the differences but also similarities of the different methods (Figs. 5 and 6). As an example, CBH-short showed an outstanding number of detected sponges in sample 1A, with 20 taxa identified within all four lineages of Porifera, Calcarea, Demospongiae, Hexactinellida, and Homoscleromorpha. A total of 638602 (over 97% of all metazoan sequences in this sample) sequences belonging to Demospongiae were detected, representing 16 differe nt families. Within this sample, 29 different sequences of Demospongiae were reconstructed (CBH-long). Additionally, for the metabarcoding, over 87% of the metazoan sequences identified in this particular sample were Porifera, but only one large cluster (OTU) of Demospongiae and two small calcares were detected. At the species level, 12 demospongies were identified with CBH-long, where one overlap with the metabarcoding approach detected species at the genus level.

3.4. Statistical analyses comparing CBH with metabarcoding

The ten samples chosen for this comparative study were collected at five different sites, including two different ecosystems, seamounts (in the Atlantic Ocean and the Mediterranean) and hydrothermal vents (inactive site, characteristic of deep bathyal sediment systems), to test different biological communities. Based on multidimensional analysis from Bray Curtis (based on the number of sequences) distances of CBH-short and MTB methods for prokaryote communities, the methods clearly appeared to be the dominant driver of differentiation (Appendix 6A). Mantel test showed that all three analyses were significantly different from each other ($p<0.01$). In contrast, for 18S detection, clusters were segregated per location of origin rather than method (Appendix 6B). Mantel test showed no significant difference ($p=0.02$), suggesting less technical bias in the community description of eukaryotes.

4. Discussion

Deep-sea biodiversity assessments are one of the major gaps in marine ecology and evolution and are strongly needed for international policy and management (Costa, Fanelli, Marini, Danovaro, & Aguzzi, 2020). Metagenomics (MTG) allows high-resolution inventories relying on strong phylogenetic reconstructions but is still too expensive and time consuming, as well as biased toward prokaryotes that represent most biomass in sediment samples (Danovaro, Snelgrove, & Tyler, 2014). Although metabarcoding (MTB) is relatively fast and less expensive, it suffers biases and restricts resolution to the species level.

In this study, we show that capture by hybridization (CBH) can offer a good balance between MTG and MTB, allowing the detection of a broader spectrum of phyla than MTB, and the reconstruction of full-length barcodes of up to 2000 bp allows the potential for robust phylogenetic reconstruction to improve the resolution of taxonomic assignments of poorly studied taxa (Wilson, Brandon-Mong, Gan, & Sing, 2018).

The approach applied here, initially developed for prokaryotes by Gasc and Peyret (2018), confirms that CBH can be adapted to describe bacterial communities in the marine realm. Furthermore, the first assay of transfer of this methodology to deep-sea metazoan communities proved successful, with five more phyla (upon 16) detected with CBH than MTB. This is in line with other recent CBH studies targeting more restrictive sets of metazoan phyla from bulk DNA, their conservative ethanol or eDNA samples (here freshwater and sediment; Gauthier et al., 2020; Shokralla et al., 2016; Wilcox et al., 2018). All these studies also demonstrated the ability of CBH to overcome biases due to PCR steps in MTB and enhance detection rates, leading not only to a more reliable representation of the species present in the samples but also of their relative biomasses, highlighting possibilities for quantitative analyses. Here, CBH results confirmed the improvement of biodiversity inventories allowed by direct taxonomic identification using Kraken2, which is based on exact k-mer matches and clustering of detected taxa to the lowest common ancestor (Wood & Salzberg, 2014). The sequence reads included different fragments within the 18S gene and were not merged. One must thus keep in mind this may contribute to a larger amount of taxa detected with CBH than with MTB, where sequences are merged pairwise, filtered, clustered and are all from identical positions, yet this clearly does not explain the broadest spectra of diversity with 5 (>30%) more phyla detected with CBH than MTB.

The second aim of this study was the reconstruction of full barcodes to detect a broader spectrum of diversity and to enable robust phylogenetic reconstruction with long sequences. This has been proven feasible and underlines that a much greater sequencing depth than the one used in this first assay will be required. An example of successful reconstruction of long fragments and increased resolution is the case of Porifera taxa in sample 1A. While MTB detected only 3 taxa, 20 identifications assigned to distinct taxa were recorded with CBH short, resulting in 29 long sequences with CBH long, assigned to 12 clearly identified species. The different hypervariable regions of 18S used for MTB vary in their resolution for the different phyla and nucleotide positions (Machida & Knowlton, 2012); using a full barcode thus clearly improves the number and quality of detected taxa, as exemplified in sample 1A for *Porifera* . The full 18S should thus be targeted to obtain the best out of the two objectives of CBH: i) increase the breadth of phyla detected and improve their taxonomic identification and ii) enable several robust phylogenetic and phylogeographic approaches to describe and unravel the biodiversity of deep-sea communities. In addition, focusing on longer fragments will limit sources of "contamination" and focus on contemporary communities by limiting the contribution of extracellular DNA from marine sediments (Torti, Lever, & Jørgensen, 2015).

4.1. Perspectives on improvement

Taking full advantage of the potential of CBH will require several important improvements to focus on biodiversity inventories based on long and resolutive fragments: adapting sequencing depth and testing of*de novo* DNA reconstruction.

In fact, the reconstruction was done here by mapping sequences to the existing SILVA databases using the program EMIRGE (Miller et al., 2011). This program developed for 16S has been shown to be extremely efficient, with 2224 full barcodes reconstructed, leading to 6132 detections in ten samples. Theoretically, a similar result could be expected for eukaryotic 18S, yet the numbers of identifications with EMIRGE were less than 6% compared to those with short CBH with Kraken2. To exclude the possibility that EMIRGE is not sufficient for 18S reconstruction, we also tested the program MAFFT for reconstruction (Katoh, Misawa, Kuma, & Miyata, 2002), which was also shown to be suitable, with similar results (data not presented). Thus, enhancing the sequencing depth is the main solution to improving the reconstruction of long fragments based on CBH.

For eDNA analysis in general, high sequencing depth is crucial (Singer, Fahner, Barnes, McCarthy, & Hajibabaei, 2019) and will allow *de novo* assembly instead of mapping algorithms, improving species identification (Deiner et al., 2017) and the detection of taxa or groups absent from the reference dataset. In this first study, both gene regions were pooled for sequencing, and the results suggest i) that the 16S probes were more efficient than the newly designed 18S probes in uncovering community richness, ii) the dominance of prokaryotic biomass was reflected, or iii) the method suffered from both limitations. The unbalanced biomass could be fixed by sequencing libraries from each rDNA separately. However, most bacterial diversity may be revealed with lower sequencing depth for these unicellular organisms, while the higher variations in body size and of the number of rDNA 18S copies among metazoans may result in a more uneven distribution of the number of fragments among taxa. In any case, a higher sequencing depth will be needed to unravel the diversity of the communities they form through full barcode reconstruction. Setting a generic number of sequences would not be realistic because the optimal number strongly depends on the diversity and biomass of the standing stock in the sediment. When studying new areas, pilot metabarcode studies may help tune the sequencing depth, allowing the optimal inventory of full-length metabarcodes in the studied ecosystems.

We tested here for the first time a new set of DNA probes that proved useful and efficient, yet the method still needs improvement to capture several taxa, such as nematodes and some other meiofauna taxa with low levels of detection. In fact, meiofauna of deep-sea sediments are generally dominated by nematodes and copepods in terms of biomass, abundance, and species richness (Zeppilli et al., 2018). Neither CBH nor MTB consistently reflects this, suggesting that not only sequencing depth but also the versatility of the set of probes needs to be enhanced.

Finally, the huge gaps in knowledge of marine biodiversity, magnified in the deep sea, result in a paucity of deep-sea sequences in nucleotide reference databases (Sinniger et al., 2016), inhibiting the efficiency of reference-based bioinformatic reconstructions (Mendoza, Sicheritz-Pontén, & Gilbert, 2014). This likely explains the very uneven reconstruction success across the phyla observed, with rather good results for Porifera, Annelida, and some Arthropods, while for other phyla such as Deuterostomia, Molluska, Nematoda, Cnidaria, and Platyhelminthes, no reconstruction could be obtained despite numerous identified sequences.

4.2. Cost factor

One of the reasons for developing metabarcoding as a future biomonitoring tool is the decreased costs (both in terms of time and money) compared to full metagenome reconstruction through shotgun sequencing or classical morphological approaches. Based on recent calculations for marine biomonitoring, the metabarcoding approach would represent only approximately half of the costs and less than 70% of the time compared to morphological assessments (Aylagas, Borja, Muxika, & Rodríguez-Ezpeleta, 2018). We found that for CHB, the laboratory time effort is nearly the same, but the costs are at present still higher than those of metabarcoding, although considerably lower than those of metagenome reconstruction. The major factor is the preparation of sequencing libraries per sample for CBH compared to metabarcoding. In addition, although multiplexing library protocols help increase cost efficiency (for example, Förster et al., 2018; Meyer & Kircher, 2010), the results presented here show that full metabarcodes must be targeted to obtain satisfying read lengths and taxonomic assignments, and multiplexing cannot encompass as many libraries in a sequencing lane as MTB, with over 100 samples. The computational power of these methods is also comparable: reconstruction can be demanding, yet error correction for high-quality metabarcoding is also time- and resource-consuming. Nevertheless, as in other recent comparable studies (Liu et al., 2016; Seeber et al., 2019), we see the advantages of overcoming the slightly higher prices in the long term. In particular, using CBH for baseline studies can improve local reference databases thanks to improved taxonomic inferences through phylogenetic reconstruction, which will be particularly important for less described areas and taxa. These extended libraries will in turn contribute to improving the output of massive metabarcoding screenings.

4.3. Conclusion

This study presents the potential advantages of using capture-based target gene enrichment for biodiversity assessment of prokaryotic and eukaryotic communities, with a specific application in poorly known deep-sea benthic ecosystems. The results showed that capture-based target gene enrichment has the potential for considerable added value compared to metabarcoding in performing referenced biodiversity inventories and phylogenetic reconstruction, improving knowledge for both biomonitoring and management purposes. CBH showed the ability to reach the two main goals initially established in this study: (i) revealing a broader spectrum of metazoan diversity and (ii) reconstructing full-length barcode regions (up to 1900 bp) allowing better phylogenetic reconstruction, thus improving taxonomic assignments. Consequently, provided sequencing depth is sufficient to allow *de novo* reconstruction, CBH can be applied at slightly higher costs than MTB for i) identifying taxa that are not well represented (nor any close relatives) in databases and ii) performing phylogenetic and phylogeographic studies.

Acknowledgements

We are grateful to Projects Pourquoi Pas les Abysses, Atlas and Total for funding. We would also like to acknowledge the teams and sampling conducted during the cruises of Medwaves (project ref) and MarMine (MarMine Project; No. 247626/O30). We acknowledge Sébastien Ybert & Chantal Cahu for help in managing the project. We thank Miriam Brandt, and Diane Bailleul for their work in the laboratory work and early steps of this study.

Funding

Founded by Total SE, Contract no. 9194 TOTAL FR0000.

References

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology, 71* (12), 8966–8969. doi: 10.1128/aem.71.12.8966-8969.2005

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

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., & Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources, 17* (6), 1231–1242. doi: 10.1111/1755-0998.12668

Aylagas, E., Borja, Á., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators, 95* , 194–202. doi: 10.1016/j.ecolind.2018.07.044

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., . . . Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature, 392* (6671), 71–75. doi: 10.1038/32160

Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution, 32* (7), 477–487. doi: 10.1016/j.tree.2017.03.001

Bohmann, K., Evans, A., Gilbert, M. T., Carvalho, G. R., Creer, S., Knapp, M., . . . de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution, 29* (6), 358–367. doi: 10.1016/j.tree.2014.04.003

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., . . . Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome, 6* (1), 90. doi: 10.1186/s40168-018-0470-z

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

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37* (8), 852–857. doi: 10.1038/s41587-019-0209-9

Bragalini, C., Ribière, C., Parisot, N., Vallon, L., Prudent, E., Peyretaillade, E., . . . Luis, P. (2014). Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Research, 21* (6), 685–694. doi: 10.1093/dnares/dsu030

Brandt, M. I., Trouche, B., Quintric, L., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2019). A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding.*bioRxiv* , 717355. doi: 10.1101/717355

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods, 13* (7), 581–583. doi: 10.1038/nmeth.3869

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics, 10* , 421. doi: 10.1186/1471-2105-10-421

Cordier, T., Forster, D., Dufresne, Y., Martins, C. I. M., Stoeck, T., & Pawlowski, J. (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring.*Molecular Ecology Resources, 18* (6), 1381–1391. doi: 10.1111/1755-0998.12926

Costa, C., Fanelli, E., Marini, S., Danovaro, R., & Aguzzi, J. (2020). Global deep-sea biodiversity research trends highlighted by science mapping approach. *Frontiers in Marine Science, 7* , 384. doi: 10.3389/fmars.2020.00384

Costello, M. J., & Chaudhary, C. (2017). Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Current Biology, 27* (11), R511–R527. doi: 10.1016/j.cub.2017.04.060

Costello, M. J., Cheung, A., & De Hauwere, N. (2010). Surface area and the seabed area, volume, depth, slope, and topographic variation for the world's seas, oceans, and countries. *Environmental Science & Technology, 44* (23), 8821–8828. doi: 10.1021/es1012752

Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-Davis, R. M., Sung, W., Power, D. M., . . . Thomas, W. K. (2010). Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology, 19 Suppl 1* , 4–20. doi: 10.1111/j.1365-294X.2009.04473.x

Danovaro, R., Snelgrove, P. V., & Tyler, P. (2014). Challenging the paradigms of deep-sea ecology. *Trends in Ecology & Evolution, 29* (8), 465–475. doi: 10.1016/j.tree.2014.06.002

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.*Microbiome, 6* (1), 226. doi: 10.1186/s40168-018-0605-2

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution, 8* (12), 1888–1898. doi: 10.1111/2041-210X.12836

Dell'Anno, A., Carugati, L., Corinaldesi, C., Riccioni, G., & Danovaro, R. (2015). Unveiling the biodiversity of deep-sea nematodes through metabarcoding: Are we ready to bypass the classical taxonomy? *PLoS One, 10* (12), e0144928. doi: 10.1371/journal.pone.0144928

Denonfoux, J., Parisot, N., Dugat-Bony, E., Biderre-Petit, C., Boucher, D., Morgavi, D. P., . . . Peyret, P. (2013). Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Research, 20* (2), 185–196. doi: 10.1093/dnares/dst001

Derocles, S. A. P., Bohan, D. A., Dumbrell, A. J., Kitson, J. J. N., Massol, F., Pauvert, C., . . . Evans, D. M. (2018). Biomonitoring for the 21st century: Integrating next-generation sequencing into ecological network analysis. In D. A. Bohan, A. J. Dumbrell, G. Woodward, & M. Jackson (Eds.), *Advances in ecological research* (Vol. 58, pp. 1–62). San Diego, CA: Academic Press.

Förster, D. W., Bull, J. K., Lenz, D., Autenrieth, M., Paijmans, J. L. A., Kraus, R. H. S., . . . Fickel, J. (2018). Targeted resequencing of coding DNA sequences for SNP discovery in nonmodel species.*Molecular Ecology Resources, 18* (6), 1356–1373. doi: 10.1111/1755-0998.12924

Gambi, C., & Danovaro, R. (2016). Biodiversity and life strategies of deep-sea meiofauna and nematode assemblages in the Whittard Canyon (Celtic margin, NE Atlantic Ocean). *Deep Sea Research Part I: Oceanographic Research Papers, 108* , 13–22. doi: 10.1016/j.dsr.2015.12.001

Gasc, C., & Peyret, P. (2017). Revealing large metagenomic regions through long DNA fragment hybridization capture. *Microbiome, 5* (1), 33. doi: 10.1186/s40168-017-0251-0

Gasc, C., & Peyret, P. (2018). Hybridization capture reveals microbial diversity missed using current profiling methods. *Microbiome, 6* (1), 61. doi: 10.1186/s40168-018-0442-3

Gasc, C., Peyretaillade, E., & Peyret, P. (2016). Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Research, 44* (10), 4504–4518. doi: 10.1093/nar/gkw309

Gauthier, M., Konecny-Dupré, L., Nguyen, A., Elbrecht, V., Datry, T., Douady, C., & Lefébure, T. (2020). Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Molecular Ecology Resources, 20* (1), 79–96. doi: 10.1111/1755-0998.13088

Ghurye, J. S., Cepeda-Espinoza, V., & Pop, M. (2016). Metagenomic assembly: Overview, challenges and applications. *Yale Journal of Biology and Medicine, 89* (3), 353–362.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., . . . Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.*Nature Biotechnology, 27* (2), 182–189. doi: 10.1038/nbt.1523

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One, 6* (4), e17497. doi: 10.1371/journal.pone.0017497

Jaziri, F., Parisot, N., Abid, A., Denonfoux, J., Ribière, C., Gasc, C., . . . Peyret, P. (2014). PhylOPDb: A 16S rRNA oligonucleotide probe database for prokaryotic identification. *Database (Oxford), 2014* (0), bau036. doi: 10.1093/database/bau036

Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering, 96* (4), 317–323. doi: 10.1016/s1389-1723(03)90130-7

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research, 30* (14), 3059–3066. doi: 10.1093/nar/gkf436

Kawahara, A. Y., Breinholt, J. W., Espeland, M., Storer, C., Plotkin, D., Dexter, K. M., . . . Lohman, D. J. (2018). Phylogenetics of moth-like butterflies (Papilionoidea: Hedylidae) based on a new 13-locus target capture probe set. *Molecular Phylogenetics and Evolution, 127* , 600–605. doi: 10.1016/j.ympev.2018.06.002

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.*Bioinformatics, 28* (24), 3211–3217. doi: 10.1093/bioinformatics/bts611

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., . . . Zhou, X. (2016). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources, 16* (2), 470–479. doi: 10.1111/1755-0998.12472

Machida, R. J., & Knowlton, N. (2012). PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *PLoS One, 7* (9), e46180. doi: 10.1371/journal.pone.0046180

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: Highly-scalable and high-resolution amplicon clustering.*PeerJ, 3* , e1420. doi: 10.7717/peerj.1420

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., . . . Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods, 7* (2), 111–118. doi: 10.1038/nmeth.1419

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal, 17* (1), 10. doi: 10.14806/ej.17.1.200

Mendoza, M. L. Z., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2014). Environmental genes and genomes: Understanding the differences and challenges in the approaches and software for their analyses.*Briefings in Bioinformatics, 16* (5), 745–758. doi: 10.1093/bib/bbv001

Mertes, F., Elsharawy, A., Sauer, S., van Helvoort, J. M. L. M., van der Zaag, P. J., Franke, A., . . . Brookes, A. J. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing.*Briefings in Functional Genomics, 10* (6), 374–386. doi: 10.1093/bfgp/elr033

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing.*Cold Spring Harbor Protocols, 2010* (6). doi: 10.1101/pdb.prot5448

Militon, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D., . . . Peyret, P. (2007). PhylArray: Phylogenetic probe design algorithm for microarray. *Bioinformatics, 23* (19), 2550–2557. doi: 10.1093/bioinformatics/btm392

Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W., & Banfield, J. F. (2011). EMIRGE: Reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology, 12* (5), R44–R44. doi: 10.1186/gb-2011-12-5-r44

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

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P. M., . . . Wagner, H. (2008). *The vegan package.*Community Ecology Package.

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples.*Environmental Microbiology, 18* (5), 1403–1414. doi: 10.1111/1462-2920.13023

Parisot, N., Denonfoux, J., Dugat-Bony, E., Peyret, P., & Peyretaillade, E. (2012). KASpOD—A web service for highly specific and explorative oligonucleotide design. *Bioinformatics, 28* (23), 3161–3162. doi: 10.1093/bioinformatics/bts597

Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis.*Molecular Ecology, 27* (2), 313–338. doi: 10.1111/mec.14478

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research, 35* (21), 7188–7196. doi: 10.1093/nar/gkm864

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research, 41* (Database issue), D590–D596. doi: 10.1093/nar/gks1219

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Erratum: Corrigendum: Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology, 35* (12), 1211. doi: 10.1038/nbt1217-1211b

Ranchou-Peyruse, M., Gasc, C., Guignard, M., Aüllo, T., Dequidt, D., Peyret, P., & Ranchou-Peyruse, A. (2017). The sequence capture by hybridization: A new approach for revealing the potential of mono-aromatic hydrocarbons bioattenuation in a deep oligotrophic aquifer. *Microbial Biotechnology, 10* (2), 469–479. doi: 10.1111/1751-7915.12426

Ribière, C., Beugnot, R., Parisot, N., Gasc, C., Defois, C., Denonfoux, J., . . . Peyret, P. (2016). Targeted gene capture by hybridization to illuminate ecosystem functioning. In F. Martin & S. Uroz (Eds.),*Microbial environmental genomics (MEG)* (pp. 167–182). New York, NY: Springer.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2020). GenBank. *Nucleic Acids Research, 48* (D1), D84–D86. doi: 10.1093/nar/gkz956

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets*. Bioinformatics* , 27(6), 863-864.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources, 15* (6), 1289–1303. doi: 10.1111/1755-0998.12402

Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., & Greenwood, A. D. (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Molecular Ecology Resources, 19* (6), 1486–1496. doi: 10.1111/1755-0998.13069

Sefc, K. M., Payne, R. B., & Sorenson, M. D. (2007). Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics, 8* (4), 879–884. doi: 10.1007/s10592-006-9240-8

Shokralla, S., Gibson, J. F., King, I., Baird, D. J., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2016). Environmental DNA barcode sequence capture: Targeted, PCR-free sequence capture for biodiversity analysis from bulk environmental samples. *bioRxiv* , 087437. doi: 10.1101/087437

Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., & Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: A case study of eDNA metabarcoding seawater. *Scientific Reports, 9* (1), 5991. doi: 10.1038/s41598-019-42455-9

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A. J., Yamamoto, H., Chevaldonné, P., . . . Creer, S. (2016). Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. *Frontiers in Marine Science, 3* , 92. doi: 10.3389/fmars.2016.00092

Smyth, R. P., Schlub, T. E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., . . . Mak, J. (2010). Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene, 469* (1-2), 45–51. doi: 10.1016/j.gene.2010.08.009

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology, 21* (8), 1789–1793. doi: 10.1111/j.1365-294x.2012.05542.x

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation, 183* , 4–18. doi: 10.1016/j.biocon.2014.11.019

Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments.*Marine Genomics, 24* , 185–196. doi: 10.1016/j.margen.2015.08.007

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution, 24* (2), 110–117. doi: 10.1016/j.tree.2008.09.011

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ, 6* , e4705. doi: 10.7717/peerj.4705

Wilcox, T. M., Zarn, K. E., Piggott, M. P., Young, M. K., McKelvey, K. S., & Schwartz, M. K. (2018). Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources, 18* (6), 1392–1401. doi: 10.1111/1755-0998.12928

Wilson, J. J., Brandon-Mong, G. J., Gan, H. M., & Sing, K. W. (2018). High-throughput terrestrial biodiversity assessments: Mitochondrial metabarcoding, metagenomics or metatranscriptomics? *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis, 30* (1), 60–67. doi: 10.1080/24701394.2018.1455189

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20* (1), 257. doi: 10.1186/s13059-019-1891-0

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology, 15* (3), R46. doi: 10.1186/gb-2014-15-3-r46

Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology, 15* (7), 1713–1731. doi: 10.1111/j.1365-294x.2006.02882.x

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., . . . Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports, 7* , 40368. doi: 10.1038/srep40368

Zeppilli, D., Leduc, D., Fontanier, C., Fontaneto, D., Fuchs, S., Gooday, A. J., . . . Fernandes, D. (2018). Characteristics of meiofauna in extreme marine ecosystems: A review. *Marine Biodiversity, 48* (1), 35–71. doi: 10.1007/s12526-017-0815-z

Data accessibility

All original fastq.gz sequencing files were deposited in the NCBI open-access sequence read archive (SRA) under accession numbers ERS3168383, ERS3168384, ERS3168414, ERS3168415, ERS3168569, ERS3168570, ERS3168460, ERS3168461, ERS3168762, ERS3168763 (MTB); and SRR13482776-SRR13482785 (CBH). All metadata associated with the samples, including detailed accession numbers, are given in Appendix 1 and are available at https://www.pangaea.de/ and BioSamples.

Author contributions

S.AH. and P.P. designed and directed the project. T.M and P.B. funded and helped supervise the project. C.D., S.M. and B.G. performed the experiments and primary data analyses. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Tables

Table 1. Samples, stations and positions displayed in Fig. 1. Included are the station, depth of sediment from the core (slice), seafloor depth, and topographic features. Appendix 1 includes all further information on the samples and samplings.

Table 2. Primers used for PCR enrichment. The barcode region, amplicon length, sequence, and first publication reference were included in detail.

Name	Direction	Region	A. size bp	Primer sequence 5´-3´	Publication
SSUF04	Forward	18S V1-V2	356	GCTTGTCTCAAAGATTAAGCC	(Blaxter et al., 1998)
SSURmod	Reverse	18S V1-V2		CCTGCTGCCTTCCTTRGA	(Sinniger et al., 2016)
515F-Y	Forward	16S V4	411	GTGYCAGCMGCCGCGGTAA	(Parada, Needham, & Fuhrman, 20
926R	Reverse	16S V4		CCGYCAATTYMTTTRAGTTT	(Parada et al., 2016)

Table 3. Probes used in this study; the 16S probe was designed by Gasc and Peyret (2018), and the 18S probe was designed in this study.

Name	Sequence 3'-5'
16S_1	CCAGACTCCTACGGGAGGCAGCAGTGGGGAA
16S_2	AAACTCCTACGGGAGGCAGCAGTGGGGAATCT
16S_3	CRAACSGGATTAGATACCCSGGTAGTCC
16S_4	AACAGGATTAGATACCCTGGTAGTCCACGCC
16S_5	GGGAGCAAACAGGATTAGATACCCTGGTAGT
16S_6	AACAGGATTAGATACCYTGGYAGTCCACGC
16S_7	AACAGGATWAGATACCCKGGYAGTCCAYRC
16S_8	ACTCAAAGGAATTGACGGGGGCCCGCACAAG
16S_9	CACAAGCGGTGGAGCATGTGGTTTAATTCGA
16S_10	CGCAAGDRTGAAACTTAAAGGAATTGGCGGGGGAGCAC
16S_11	GTTGGGTTAAGTCCCGCAACGAGCGCAACCC
16S_12	GAGAGGWGGTGCATGGCCGYCGYCAGYTCGT
16S_13	CATGGTTGTCGTCAGCTCGTGTCGTGAGATG
16S_14	TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCSS
16S_15	TCGTCAGCTCGTGTYGTGAGRTGTTSGGTTAAGTCC
18S_1	AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAA
18S_2	TCTGGTGCCAGCAGCCGCGGTAATTCCAGCT
18S_3	TGCCAGCAGCCGCGGTAATTCCAGCTCCAAT
18S_4	CGCGGTAATTCCAGCTCCAATAGCGTATATT
18S_5	GAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGT
18S_6	GTCCCTGCCCTTTGTACACACCGCCCGTCGC
18S_7	GATTACGTCCCTGCCCTTTGTACACACCGCC
18S_8	TTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTA
18S_9	GAGCCTGCGGCTTAATTTGACTCAACACGGG
18S_10	AAGGAATTGACGGAAGGGCACCACCAGGAGT
18S_11	GGAAGGGCACCACCAGGAGTGGAGCCTCGGCTTAATTTGACTCAACACGG
18S_12	TGGTGGTGCATGGCCGTTCTTAGTTGGTGGA
18S_13	TGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCT
18S_14	GCAATAACAGGTCTGTGATGCCCTTAGATGT
18S_15	AAACTTAAAGGAATTGACGGAAGGGCACCAC
18S_16	GGGGGAGTATGGTCGCAAGGCTGAAACTTAA
18S_17	GTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCA

Figures

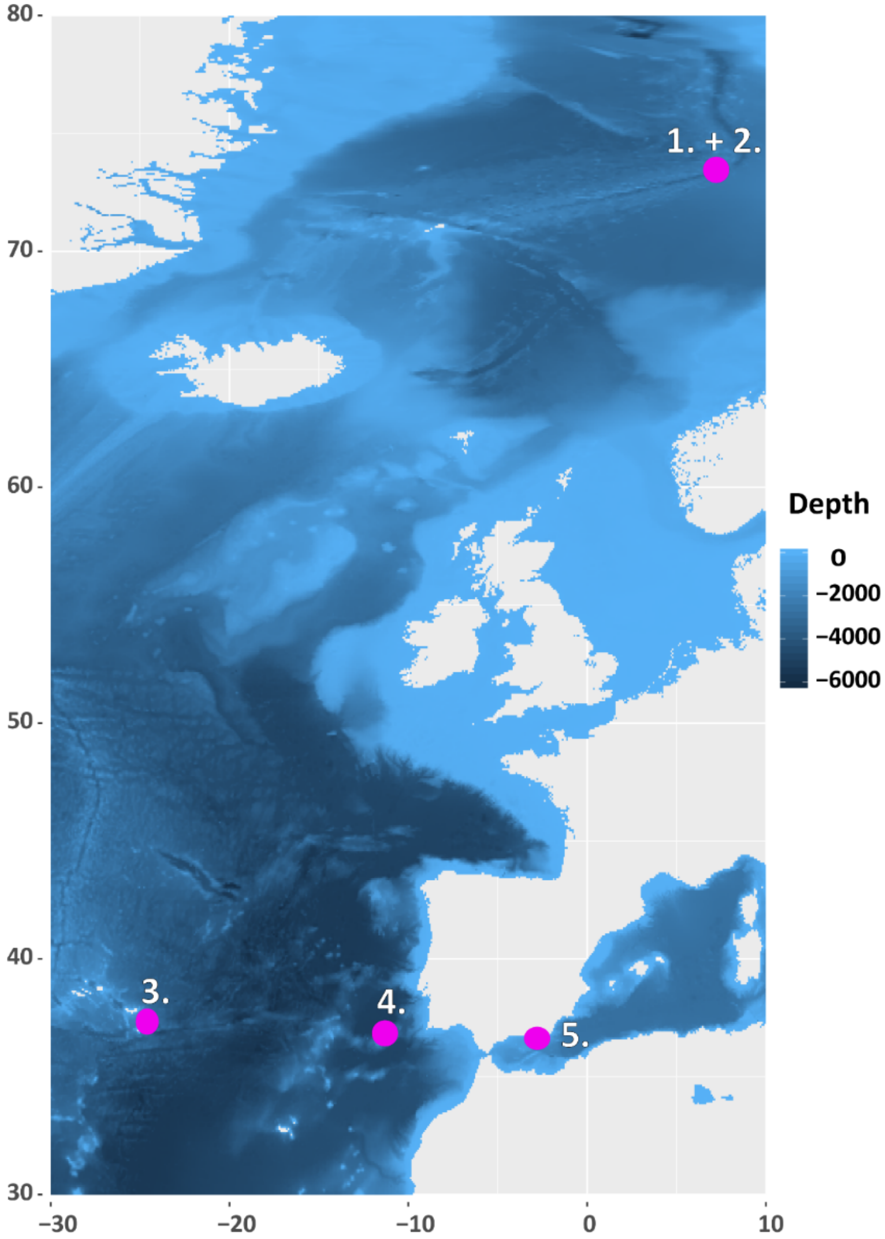

Fig. **1.** Map showing the five stations used for sampling in this study, with hydrothermal vents at stations 1 and 2 as well as seamounts at the others. The information for the stations and samples is presented in Table 1.

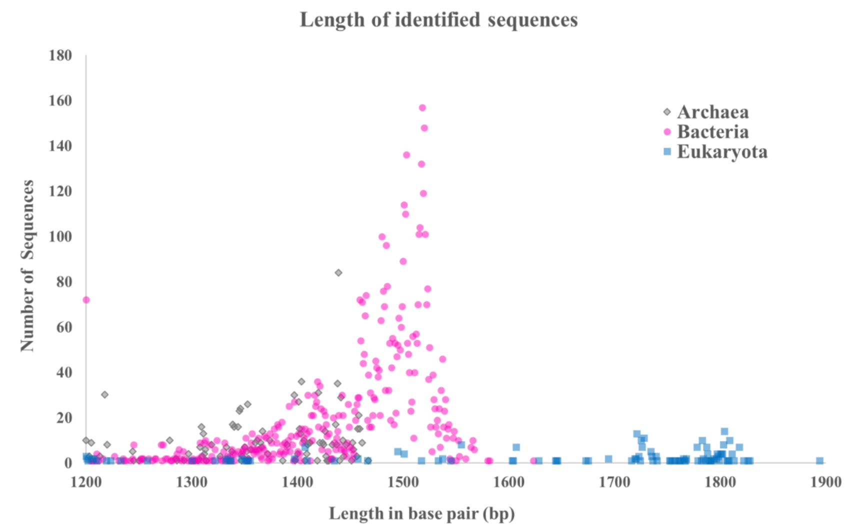

Fig. 2. Query length of identified sequences with the CBH analyzed using the EMIRGE pipeline (CBH-long). Metabarcoding of the same genes ranked between 313-411 base pairs (bp).

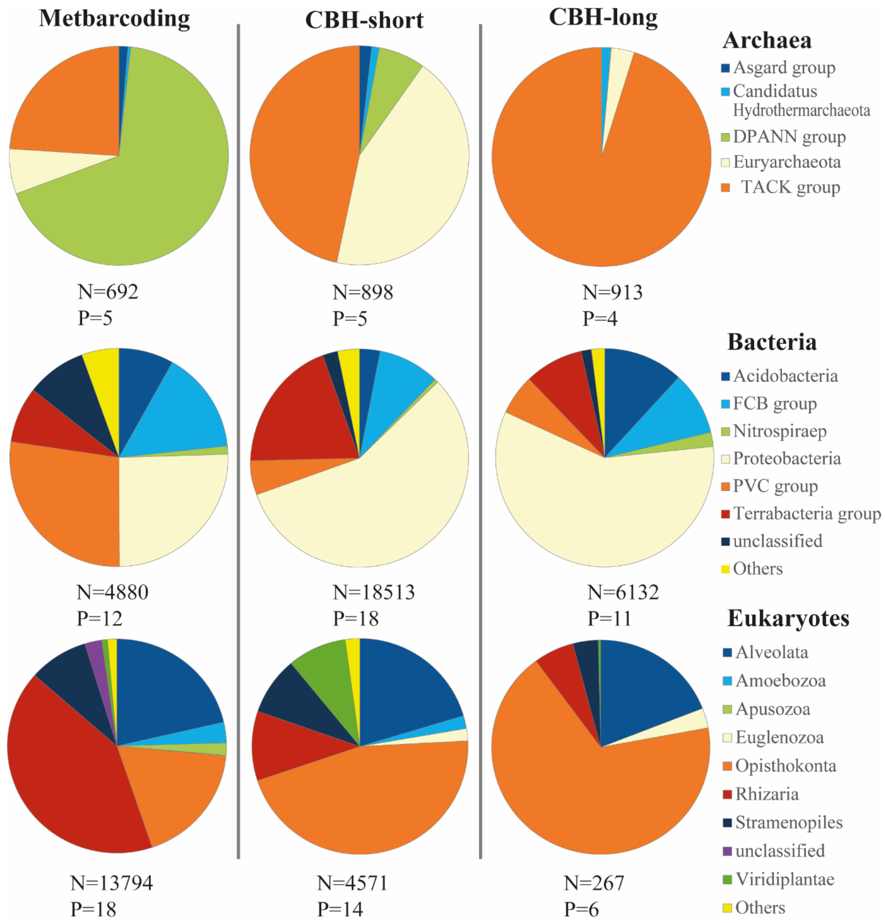

Fig. 3. Comparison of high-level taxonomy of all ten samples using metabarcoding versus CBH with two distinct pipelines: CBH-long (EMIRGE) and CBH-short (Kraken2). "N" indicates the number of identified taxa, and "P" the number of phyla, while phyla with less than 2% occurrence were counted together as "Others".

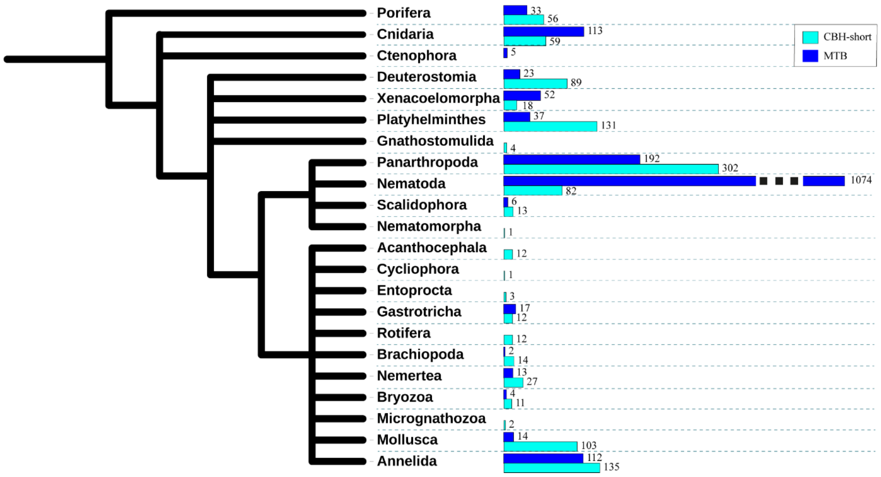

Fig. 4. Phylogenetic tree based on NCBI taxonomy, created by iTol. Indicated are the counted taxa per phylum, while metabarcoding is indicated with dark blue and CBH-short with light blue.

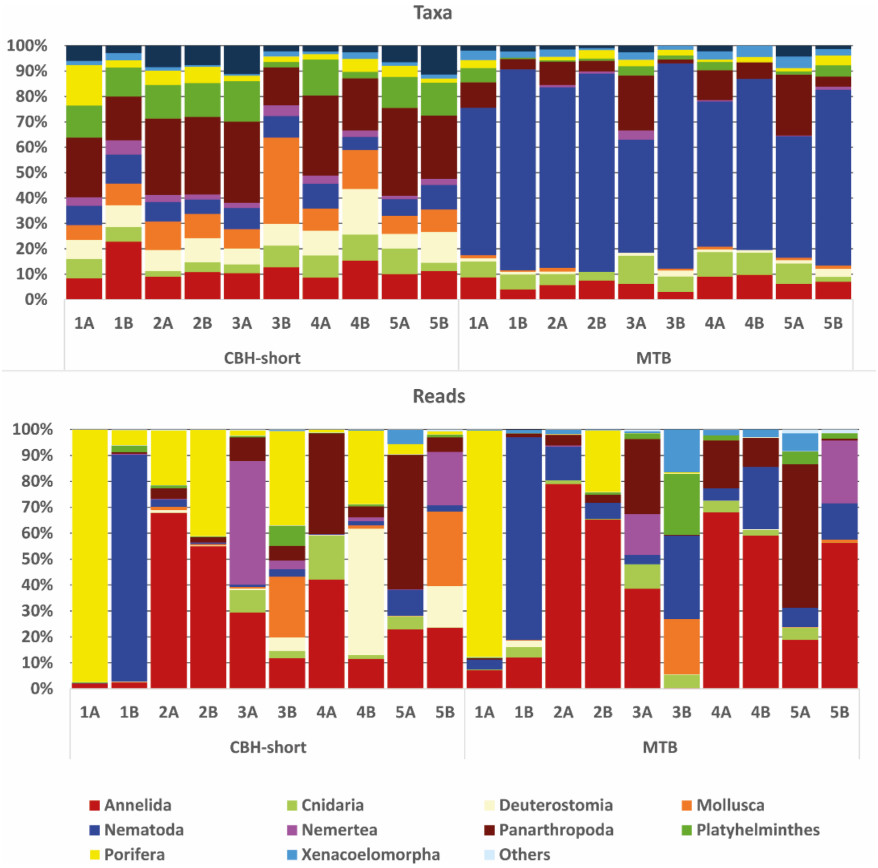

Fig. 5. Sample-wise detection with 18S analyses from CBH-short and metabarcoding for all detected metazoan phyla. The upper graph shows the percentage of detected taxa per phylum, and the lower shows the percentage of reads per phylum.

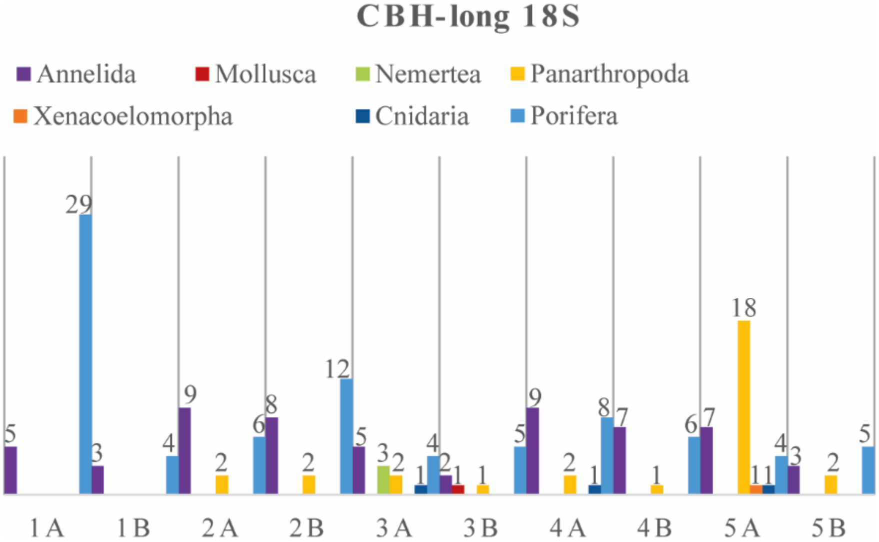

Fig. 6. Numbers of taxa per sample using CBH-long.