UiT
THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Biosciences, Fisheries and Economics

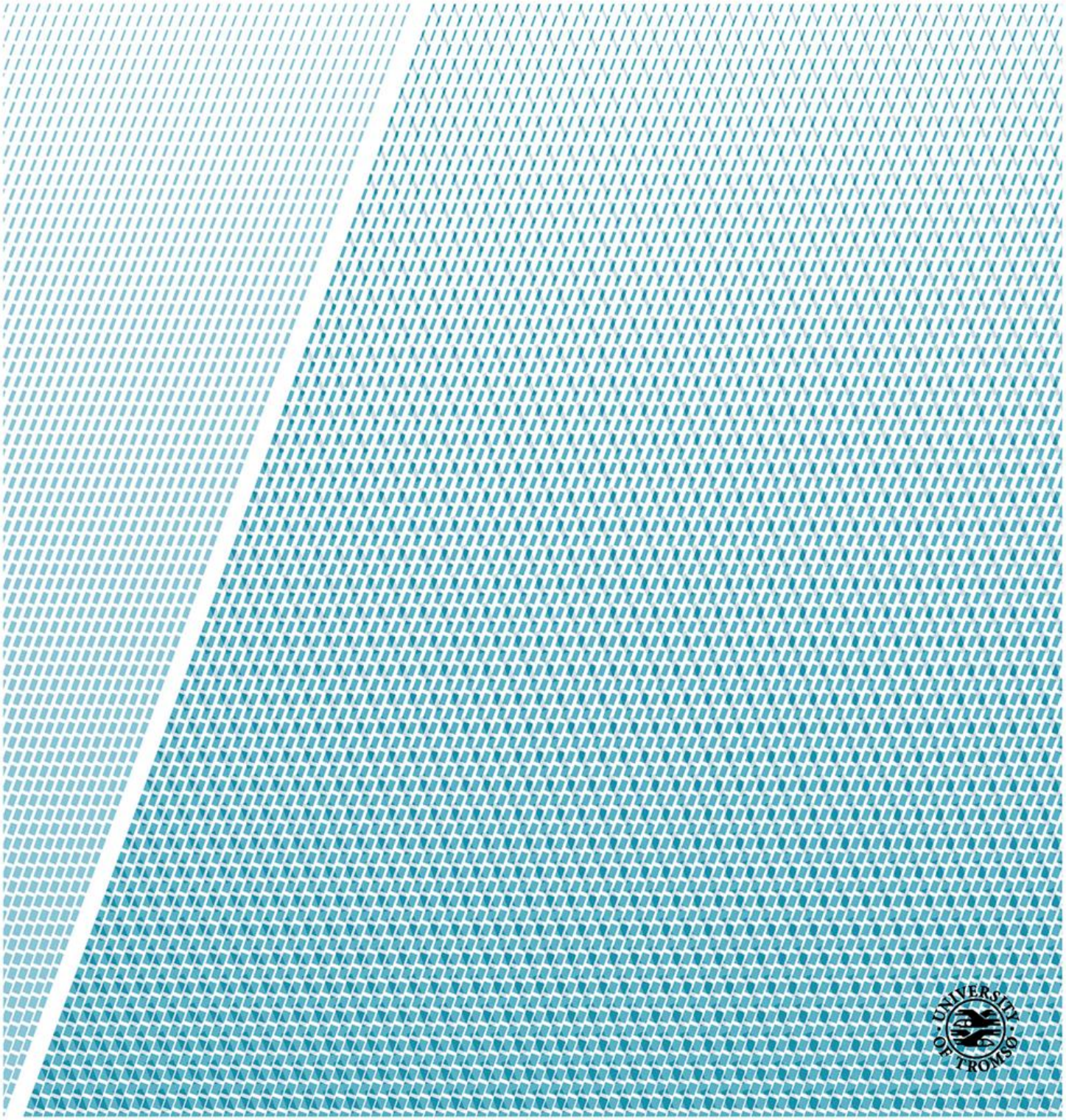
Department of Arctic and Marine Biology

DNA metabarcoding of deep-sea sediment communities using COI: community assessment, spatio-temporal patterns and comparison with the 18S rDNA marker

Sara Atienza Casas

BIO-3950 Master thesis in Biology

June 2019



ABSTRACT

The deep sea is the largest biome on Earth, albeit it is the least studied. Among the complex ecosystems and habitats that form the deep sea, submarine canyons and open slope systems are regarded to be potential hot-spots of biodiversity. The Mediterranean Sea hosts the 8.86% of the inventoried submarine canyons in the global ocean, like the Blanes Canyon, located in its Northwestern section. We assessed spatial (through sediment layers and along a depth gradient) and temporal (in two different seasons) patterns of biodiversity in sediment communities of the Blanes Canyon and its adjacent open slope with eDNA metabarcoding, using a fragment of the mitochondrial gene cytochrome c oxidase subunit I (COI) as a marker. We found a total of 15,318 molecular operational taxonomic units (MOTUs), of which 10,860 could be assigned only to Eukarya. Among those assigned at lower levels, Metazoa, Stramenopiles and Archaeplastida were the dominant taxa. Within metazoans, Arthropoda, Nematoda and Cnidaria were the most diverse among the 28.2% that could be assigned to at least the phylum level. There was a trend towards decreasing diversity in the first few cm (1 to 5) of the sediment, with only 26.3% of the MOTUs shared across sediment layers. Our results show the presence of heterogeneous communities in the studied area, significantly different between zones, depths and seasons. We compared our results with the ones presented in Guardiola et al (2016), obtained using the v7 region of the 18S rRNA gene as genetic marker in the exact same samples. There were remarkable differences in the total number of MOTUs, in the most diverse taxa and in MOTU richness. COI recovered a higher number of MOTUs, but more remained unassigned taxonomically. However, broad spatio-temporal patterns elucidated from both datasets coincided, both markers retrieving the same ecological information. The choice of marker depends on a trade-off between marker variability, primer bias, and completeness of reference databases. Our results showed that COI can be used to accurately characterize the studied communities and to develop high-resolution bioindicators to detect ecological shifts. We also noted that COI reference databases for deep-sea organisms have important gaps, and its completeness is essential in order to successfully apply metabarcoding solutions.

TABLE OF CONTENTS

INTRODUCTION

The world's oceans cover ca. 71% of Earth's surface, and the average depth of the global ocean is 3,688 m, albeit the average depth can vary between regions, ranging from 1,205 m depth found in the Arctic Ocean, the shallowest, to 4,298 m depth, the deepest, found in the North Pacific (Eakins and Sharman, 2010). The deep sea is defined as any environment (including the water column and the seabed) beyond continental shelf depths, that is, below 200 m depth, and it forms the largest biome on Earth. The deep sea covers an area of 360 million km^2, which represents ca. 50%, or even 60%, of the Earth's surface (Costello et al, 2010, Ramirez-Llodra et al, 2011), and the deepest ocean depth recorded is 10,911 m (the Marianas Trench). It is undeniable that the structure of the global ocean is complex and it comprises a wide variety of ecosystems, habitats and environmental conditions. As regards the Mediterranean basin, Danovaro et al (2010) found that it contains numerous unexplored habitats that could be "hot-spots" of biodiversity and they cite, not exhaustively, open slope systems, submarine canyons, deep basins, seamounts, deep-water coral systems, cold seeps and carbonate mounds, hydrothermal vents and permanent anoxic systems. Harris and Whiteway (2011) produced a global inventory of 5,849 separate large submarine canyons, where 8.86% of them (that is, 518) are located in the Mediterranean Sea. That is a large number provided that the Mediterranean Sea only occupies an area and volume of 0.8% and 0.3%, respectively, of the world's oceans (Eakins & Sharman, 2010).

Different types of disturbances have been described to impact the northwestern Mediterranean canyons area: natural events such as eastern wet storms, northern dry storms (Guillén et al, 2006; Bonnin et al, 2008) and episodic dense shelf water cascading (Canals et al, 2006; Heussner et al, 2006; Ulses et al, 2008; Canals et al, 2013). Cold and dry continental northern winds originate dry storms, while wet storms are triggered by eastern winds blowing over the sea surface and absorbing moisture, resulting in the air becoming warm and humid. During northern dry storms, the wind pushes sea water away from the coast, which triggers the appearance of high waves offshore. The scenario during eastern wet storms is the opposite: sea water is pushed towards the coast, high-energy waves erode the coastline and heavy rainfall increases river discharge. In addition, anthropogenic disturbances occur in the area, mainly bottom trawling deep-sea fisheries, where fishing gear is dragged along the ocean floor (Lopez-Fernandez et al, 2013; Puig et al, 2015; Paradis et al, 2018).

According to Lopez-Fernandez et al (2013), northwestern Mediterranean canyons are subdued to higher particle fluxes and sedimentation rates than the adjacent open slope, with particle fluxes being an order of magnitude superior at the same depths (also, Paradis et al, 2018). The stated reasons are the topography of the canyon itself and suspended matter released after bottom trawling of the fishing grounds around the canyon rim and flanks, where also the release of continental and resuspended

material during intense weather-driven events enters the submarine canyon and is carried downwards (Lopez-Fernandez et al, 2013, Paradis et al, 2018). Moreover, lateral inputs are conducted into the canyon by gullies carved in the canyon flanks (Zuñiga et al, 2009). As regards the open slope, Lopez-Fernandez et al (2013) claim that its sediment load is influenced by different factors: recently formed deep waters that are pushed by convection events from the Gulf of Lions and sinking biogenic particles from phytoplankton blooms triggered by the arrival of mineral dust originating from North Africa.

Biodiversity assessment techniques are crucial to understand the structure and function of these important, yet overlooked, marine ecosystems, and to evaluate the impact of the multiple stressors that may be affecting them. Classical techniques based on morphological identification of macrofauna are difficult, time consuming, and prone to misleading results, given the lack of accurate morphological keys for many little known taxonomic groups living in the deep sea. Novel techniques based on molecular identification of sequences detected from environmental DNA (eDNA) samples are a promising tool to monitor these ecosystems, which can potentially provide an accurate and traceable picture of the existing biodiversity (Taberlet et al, 2012). Particularly, obtaining results when using metabarcoding techniques does not depend on neither taxonomic expertise (which is scarce in many taxonomic groups) nor the size of the individuals, the use of this technique saves time (more data can be processed at a time with high throughput sequencing technologies) and resources (samples consist on a small portion of water, soil or sediment), and there is no need for prior taxonomic identification of the organisms because they are identified through their unique DNA sequences.

eDNA comprises DNA from living organisms, including all stages (i.e. eggs, larvae, juveniles, adults), from dead organisms, from gut contents and from extracellular DNA sources such as faeces, exudates, etc. Dell'Anno and Danovaro (2005) found that DNA concentrations in the deep-sea sediments worldwide are 0.31 +/- 0.18 g of DNA m^{-2} in the top cm, which is extremely high, and that DNA content in the uppermost 10 cm of the deep-sea sediments is 0.50 +/- 0.22 Gt, where extracellular DNA accounts for 0.30 to 0.45 of those Gt. Therefore, the deep-sea upper sediment layer is the largest reservoir of DNA in the world ocean.

In this work, we use a novel biodiversity assessment technique based on metabarcoding of eDNA extracted from deep-sea sediments from western Mediterranean canyons and open slope environments, in order to test the feasibility of this technique to characterize the biodiversity present in these deep-sea ecosystems. The targeted genetic marker that we used as a metabarcode for identification purposes was a fragment of the mitochondrial gene cytochrome c oxidase subunit I (COI). There are several reasons why we chose COI. It is a highly conserved gene because it encodes enzymatic elements related to the crucial cell function of aerobic respiration (Richter and Ludwig,

2003, and references therein), so we can expect to find this gene in all eukaryotic organisms (at least, in all organisms which carry mitochondria). It is also subdued to high variation rates and it can incorporate enough variation to distinguish organisms belonging to close species or even to the same species (haplotypes) (Hu et al, 2002). Despite this sequence variability, there are some conserved regions in the COI sequence that can be used as a target for universal primers (Leray et al, 2013; Wangensteen et al, 2018), which can amplify a fragment with the right size to be analyzed in most commonly used sequencers (Wangensteen & Turón 2017). Finally, mitochondrial markers are present in thousands of copies in every cell genome, and mitochondria are protected from degradation by organelle membranes, so mitochondrial markers are expected to be abundant and to constitute a significant fraction of the total eDNA, which are expected to improve their detectability.

The aim of this study was two-fold. First, we wanted to provide an insight of the community profile, to analyze spatial distribution patterns, vertically (through layers of sediment) and horizontally (along a depth gradient), and to seek for seasonal trends (autumn versus spring) of sediment communities inhabiting the Blanes Canyon, one of the main canyons in the described area (Lastras et al, 2011), and the adjacent open slope, both considered potential biodiversity hot-spots (Danovaro et al, 2010). Second, we wanted to compare the information obtained from COI and from 18S rRNA. To this end, we compared our results with those of Guardiola et al (2016), which were obtained analyzing a variable fragment of the v7 region of the nuclear gene coding for the small ribosomal subunit (18S rRNA) as genetic marker. The same samples and DNA extracts as in Guardiola et al (2016) have been used in the present study, so we can verify whether similar patterns and information could be retrieved using different genetic markers.

MATERIALS AND METHODS

Study area and Sampling

The sampling area is located in the Western Mediterranean Sea, NE Iberian Peninsula, where we obtained sediment samples from two different sites: the Blanes canyon (BC) and the southwards adjacent open slope (OS) (Figure 1). The canyon head is incised in the continental shelf less than 4 km offshore, it extends in a N-S direction for about 4 km and then it expands in a NW-SE direction with a meandering course. The canyon grows from 60 m water depth down to 2,600 m depth, where it outflows to the lower Valencia Channel segment, and its width varies from 8 km to a maximum width of 20 km (Amblas et al, 2006; Lastras et al, 2011; Canals et al, 2013).

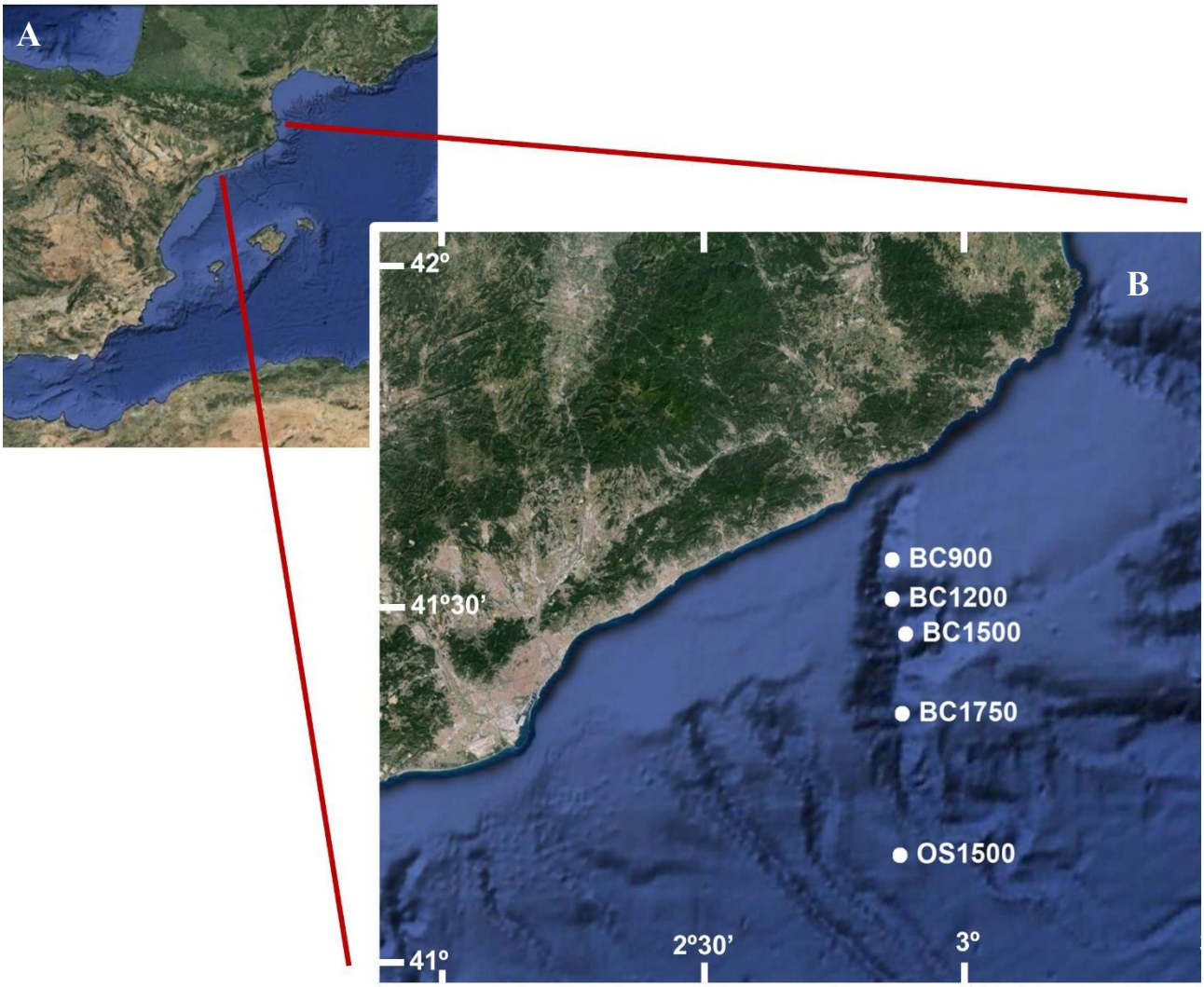


Figure 1. (A) Partial image of the Iberian Peninsula (Google Earth, Landsat) and (B) enlargement showing the Blanes Canyon with indication of the sampling stations (Google Earth, Institut Cartogràfic de Catalunya). Codes as in Table 1 below.

The sampling area was accessed using the *R/V García del Cid* from the Spanish Research Council. The device used to obtain the sediment samples was a 6-tube multicorer KC Denmark A/S, with each corer having 9.4 cm of inner diameter and a length of 60 cm. Three hauls (replicates) were made less

than 600 m apart at 4 different depths inside the canyon: 900 m, 1,200 m, 1,500 m and 1,750 m; while at one depth in the adjacent open slope: 1,500 m. The same spots were sampled in two consecutive cruises: autumn 2012 (DOSMARES II) and spring 2013 (DOSMARES III) (Table 1). One of the corers per haul was used for our sediment sampling, and was subsampled on board by taking one mini-corer (PVC tube 3.6 cm in diameter and 5 cm in height). Consecutively, each mini-corer sample was divided into 3 different layers following the sediment profile: A (first cm), B (second cm) and C (third to fifth cm). All 90 samples, 45 per season (5 stations x 3 corers x 3 layers), were preserved in absolute ethanol and used for DNA extraction.

DOSMARES II CRUISE					
Zone	**Date**	**Station**	**Depth (m)**	**Lat N**	**Long E**
Blanes Canyon (BC)	7/10/2012	BC900	874	41°31'17"	02°50'52"
			890	41°34'14"	02°50'47"
			852	41°34'17"	02°50'35"
		BC1200	1,232	41°30'43"	02°50'46"
			1,194	41°30'44"	02°50'35"
	8/10/2012		1,248	41°30'44"	02°50'54"
		BC1500	1,450	41°27'31"	02°52'57"
			1,463	41°27'38"	02°52'46"
			1,457	41°27'29"	02°52'58"
	11/10/2012	BC1750	1,746	41°21'16"	02°52'13"
			1,751	41°21'20"	02°52'13"
			1,727	41°21'38"	02°52'15"
Open Slope (OS)	12/10/2012	OS1500	1,454	41°08'42"	02°53'33"
			1,451	41°08'37"	02°53'32"
			1,480	41°08'30,7"	02°53'48,1"
DOSMARES III CRUISE					
Zone	**Date**	**Station**	**Depth (m)**	**Lat N**	**Long E**
Blanes Canyon (BC)	5/4/2013	BC900	870	41° 34' 19"	02° 50' 59"
			903	41° 34' 19"	02° 50' 56"
			899	41° 34' 19"	02° 50' 53"
	6/4/2013	BC1200	1,227	41° 31' 01"	02° 51' 03"
			1,258	41° 30' 52"	02° 50' 50"
	10/4/2013		1,234	41° 30' 46 "	02° 50' 54"
		BC1500	1,483	42° 27' 15"	02° 52' 57"
	23/04/2013		1,516	41° 27' 03"	02° 52' 53"
	24/04/2013		1,520	41° 27' 02"	02° 52' 50"
	23/04/2013	BC1750	1,730	41° 21' 14"	02° 51' 54"
			1,785	41° 21' 14"	02° 52' 17"
			1,830	41° 21' 06"	02° 52' 20"
Open Slope (OS)	22/04/2013	OS1500	1,507	41° 08' 28"	02° 53' 46"
	23/04/2013		1,480	41° 08' 23"	02° 53' 33"
			1,456	41° 08' 18"	02° 53' 41"

Table 1. Details of the sampling localities where the 90 samples analyzed in this study were collected during the DOSMARES II and DOSMARES III cruises.

DNA extraction, amplification and sequencing

We processed 10 g of sediment per sample for DNA extraction using the PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories, Inc.). As explained before, we targeted COI as genetic marker and amplified the ‘‘Leray fragment’’ of ca. 313 bp, but we adapted the pre-existing primer set by manually checking such set against representative sequences from the Genbank database in order to have a wider eukaryotic coverage. The new highly degenerated set of reverse and forward primers (“Leray-XT”) included the reverse primer jgHCO2198 5’ -TAIACYTCIGGRTGICCRAARAAYCA-3’ (Geller et al, 2013) and a novel forward primer mlCOIintF-XT 5’-GGWACWRGWTGRACWITITAYCCYCC-3’, modified from the mlCOIintF primer described by Leray et al (2013). The modification, developed in Wangensteen et al (2018), consisted on the incorporation of two more degenerate bases and two inosine nucleotides in the most variable positions.

Amplification through Polymerase Chain Reaction (PCR) of the target zone took place in a total volume of 30 µl per sample and the mixture had 6 different components: AmpliTaq® Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA) 5 U/µl (0.24 µl), forward and reverse 8-base tagged primers mix (1.2 µl of 5 µM), buffer 10x3 µl of $MgCl_2$ (3 µl), deoxyribonucleotide triphosphate (dNTP, 2.5 mM each, 2.4 µl), bovine serum albumin (20 mg/ml, 0.24 µl) and DNA template (3 µl). The addition of a different tag per sample, in both the reverse and forward primers, had the purpose of uniquely identifying the amplified sequences that belonged to a particular sample. Tags had a minimum difference of 3 base pairs out of the total 8 between one another, and were created with the program named OligoTag (Boyer et al, 2016).

The PCR procedure consisted of a first denaturation step that lasted 10 min at 95ºC, followed by 45 cycles of denaturation at 95ºC for 30 s, hybridization at 45ºC for 30 s and elongation at 72ºC for 30 s. Three PCR-blanks were run by amplifying the PCR mixture without the DNA template, as well as three negative controls that were run with ultrapure water (Milli-Q System), and were used as template. Quality of amplifications once the PCR step finalized was assessed by electrophoresis in agarose gels. Purification of all PCR products took place by using Minelute PCR purification columns (Qiagen, Valencia, CA, USA) and such products were pooled by marker. FASTERIS (Plan-les-Ouates, Switzerland; https://www.fasteris.com/dna/) was in charge of library preparation and sequencing by doing a complete run on an Illumina MiSeq platform using v3 chemistry (2 x 300 bp paired-ends). The protocol utilized by FASTERIS is called Metafast and it incorporates Illumina adapters using a ligation procedure that does not require any further PCR steps, hence minimizing biases.

Metabarcoding Pipeline and taxonomic assignment

Our metabarcoding pipeline was based on the OBITools v1.01.22 software suite developed by Boyer et al (2016). Initially, the length of the raw reads was trimmed to a median Phred quality score higher than 30, from which we obtained 15,497,229 reads. A second filtering step took place using illuminapairedend, where 14,845,308 reads with paired-end alignment quality scores higher than 40 were demultiplexed using ngsfilter, which also removed primer sequences. Moreover, the 8-base sample tags were used to assign reads to the samples they belonged to. A third filtering step, this time by length (obigrep), was applied to the 10,243,806 assigned reads, resulting on 3,138,541 reads with lengths varying between 309 and 318 bp, and without ambiguous positions. 1,013,295 unique sequences were then grouped and dereplicated by using obiuniq, and chimeric sequences were tracked down and removed utilizing the uchime_denovo algorithm implemented in vsearch v1.10.1 (Rognes et al, 2016), leaving 989,639 sequences.

The sequences were then clustered into molecular operational taxonomic units (MOTUs) using SWARM v2 (Mahé et al, 2015). According to Mahé et al (2015), this high-resolution *de novo* clustering method is especially suitable for unexplored environments that lack quality taxonomic reference databases or that may include rare taxa. Their algorithm uses an *exact-string comparison* approach: for each sequence, the algorithm generates sequences containing one (or the desired number) variant (*d*, local clustering threshold), called *microvariants*, and checks whether those microvariants are present in the pool. This method is faster than the *approximate-string comparison* approach, where an exact pairwise alignment comparison of each sequence is done against all the remaining sequences. This way, microvariants from low abundant MOTUs are cross-checked with microvariants from high-abundant sequences and fused together if deemed necessary.

The parameters we used are d=13 and t=10 (number of computation threads to use). The number of MOTUs obtained after the clustering phase was 45,845. Consecutively, we used obigrep to remove singletons and obtained 32,354 MOTUs. We performed this step after the clustering phase so as to minimize data loss because singletons made up a substantial proportion of the reads (29.43%). According to Wangensteen & Turon (2017), long (> 300 bp) markers are more prone to random point sequencing errors, and an early removal of singletons could end up yielding an excessively pruned dataset.

Taxonomic assignment was done using ecotag (Boyer et al, 2016), which only chooses one representative sequence per MOTU (which is called “seed”) among the microvariants for assignment purposes. Ecotag uses a local reference database as well as a phylogenetic tree-based approach (based on NCBI taxonomy) in order to assign sequences without a perfect match. The ecotag program searches for the best hit of a query sequence in the reference database based on the similarity (fraction

of identity) between the sequences. It then searches all entries in the database as similar or more to the best-hit than the query sequence is. Then, MOTUs are assigned to the most recent common ancestor (contained in the NCBI taxonomy tree) to all the sequences included in the set described above. Taxonomic assignment will therefore depend on the similarity of the sequences as well as on the quality of the reference database. We developed a mixed reference database by fusing sequences from two different sources together: (i) *in silico* ecoPCR against the release 117 of the EMBL nucleotide database; and (ii) sequences obtained from the Barcode of Life Datasystem (Ratnasingham & Hebert, 2007) using a custom R script to pinpoint the Leray fragment. This new database (db_COI_MBPK) included 188,929 reference sequences. As one of the purposes of this project is to assess the completeness of current barcoding databases for Mediterranean deep-sea marine taxa, we purposely resorted to sequences already available from public repositories. MOTUs were classified in accordance with the major super-groups of eukaryotes described in Guillou et al (2013), with only one exception: Ophistokontha was split into Metazoa, Fungi, and Other Opisthokontha.

The last part was the performance of a 4-step final refining of the dataset obtained after the taxonomic assignment. First, we carried out a blank correction process: any MOTUs where the number of reads from blank and negative controls was above 10% of the total number of reads were removed (31,297 MOTUs and 2,746,678 reads remained). Then, the minimal relative abundance filtering took place: the number of reads of a MOTU in a sample was compared to the total number of reads of that sample and set to 0 if it was below 0.00005 of the total (31,297 MOTUs and 2,679,433 reads remained). Third, we manually removed all MOTUs that were not assigned to marine eukaryotes, that is: non-marine organisms, prokaryotes and the root of the Tree of Life (22,248 MOTUs and 2,084,086 reads remained). Finally, we eliminated from the dataset those MOTUs that contained less than 5 reads after applying the previous 3 steps, leaving 15,318 MOTUs and 2,068,065 reads left.

Data analysis

We performed several types of analyses in order to obtain the results detailed below, and the global dataset was divided in subsets accordingly. The first analysis was a spatio-temporal study of the 90 samples gathered from the Blanes canyon and the adjacent open slope (72 and 18 samples, respectively), 45 corresponding to Autumn and 45 to Spring. In order to carry out the COI-18S comparison, we used public results published in Guardiola et al (2016), as well as analyzed raw data from the dataset used by the authors, which was obtained from the exact same samples.

The program used to carry out the analyses was R version 3.2.3. All t-tests were obtained by using the function *t.test*. Mantel tests were performed using the function *mantel* so as to assess the degree of similarity between ordinations obtained with different datasets (COI vs 18S).

We resorted to the vegan 2.5-4 package (Oksanen et al, 2016) in order to draw rarefaction and species accumulation curves, the former by using the function *rarecurve* and the latter, by using *specaccum.* We used presence/absence data (Jaccard index) to build distance matrices in order to assess if the tested factors were significant. Reduced-space graphical representations of our data were obtained with non-metric multidimensional scaling (nMDS) ordinations after installing the package Mass and running the function *isoMDS*, which included those distance matrices and applied 300 iterations. We then tried to elucidate spatio-temporal patterns by carrying out comparisons between seasons (Autumn and Spring), layers (A, B and C), zones (Blanes canyon and the adjacent open slope) and depths (900 m, 1200 m, 1500 m and 1750 m). PERMANOVA (permutational analyses of variance) tests were done with the function *adonis* included in the vegan package. When a factor was found significant, two additional tests were performed. First, PERMDISP (tests of multivariate dispersion) tests, also included in the vegan package under the function *betadisper,* were carried out to determine if such significance was due to a different multivariate mean or to a different heterogeneity of the groups. And second, the execution of permutational pair-wise tests, which needed the installation of the library pairwiseAdonis to use the function *pairwise.adonis,* enabled us to discover where the significant differences were when there were multiple levels inside that factor. We used the Benjamini-Yekutieli (2001) FDR correction to correct p-values for multiple comparisons.

In order to draw ellipses which were proportional to the number of MOTUs of the global dataset per sediment layer, we used the VennDiagram function *draw.triple.venn.* Finally, the environmental fit test was obtained with the function *envfit* included in the vegan package and it was executed by incorporating data related to biotic and abiotic variables from Roman et al (2016) (particularly, data on chlorophyll a (Chl a), chloroplastic pigment equivalents (CPE), chlorophyll a divided by its degradation products (phaeopigments) (Chl a / P), total nitrogen content (TNC), organic carbon content (OCC), molar carbon-nitrogen ratio (C/N), Clay, Silt and Sand), because these data were obtained during the same cruises and at the exact same sampling stations as our data. The aim of this test is to fit environmental vectors onto an ordination (that is, the nMDS), and plot them adjusting the direction of the vectors to the main gradient of change in the environmental variable and the length of the vectors to the correlation between the spatial configuration and the environmental variable.

RESULTS

Community profile

The final dataset obtained after carrying out the aforementioned process consisted of 15,318 clusters, or MOTUs, that contained 2,068,065 reads. The 38.69% of clusters (5,927 MOTUs) were assigned at 80% or more similarity with a sequence included in the reference database, and 0.75% (115 MOTUs) were assigned at 90% or more similarity. The highest assignment was made at 99,68% similarity and the lowest, at 55.31%. Blank and negative controls (6 in total) had on average 520 reads per sample and, cumulatively, 3,178 reads, which is negligible compared to the global mean number of reads per sample (22,978.5) and the total number of MOTUs.

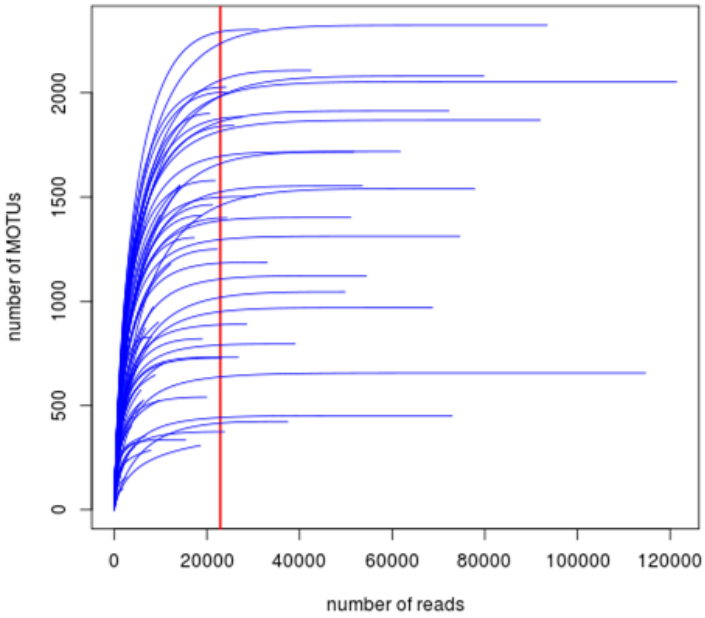


Figure 2. Rarefaction curves of the number of MOTUs obtained at increasing number of reads per sample.

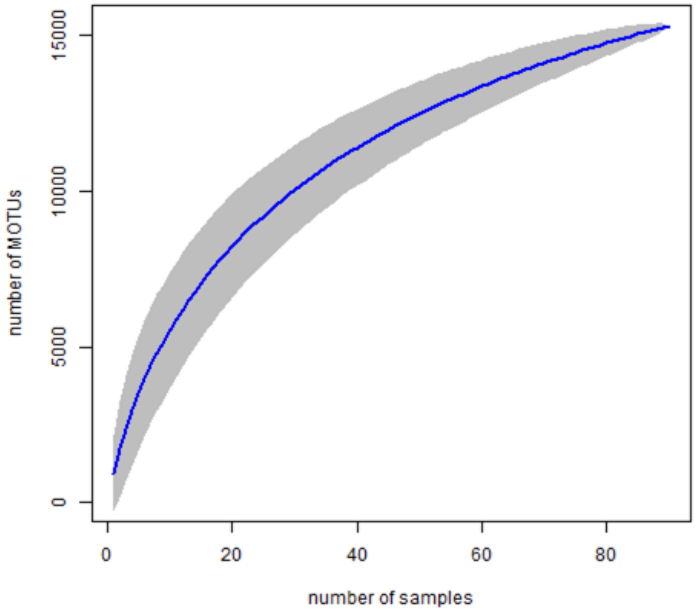


Figure 3. Species accumulation curve of the number of new MOTUs obtained after adding increasing number of samples to the analysis, with the confidence interval (95%) represented.

We performed rarefaction curves in order to check if the sequencing depth was appropriate to capture MOTU richness in our samples. Figure 2 shows that, in general, a plateau in the number of MOTUs per sample was reached at ca. 20,000 reads. Therefore, taking into account that the mean number of reads per sample in the global dataset was 22,978.5, it indicates that this sequencing depth captures most of the MOTUs present in our samples. We then performed a species accumulation curve to visualize the accumulation rate of new MOTUs with increasing sampling effort and to take account of between sample heterogeneity. We can see in Figure 3 that the curve does not reach an asymptote after processing 90 samples, yet it begins to plateau. Consequently, carrying out additional sampling would incorporate more MOTUs to the dataset.

The clusters were distributed in 11 super-groups: Alveolata, Rhizaria, Metazoa, Stramenopiles, Amoebozoa, Hacrobia, Fungi, Archaeplastida, Apusozoa, Opisthokonta (other than Metazoa and Fungi) and Excavata. However, there were 10,860 MOTUs (70.9% of the total) assigned to the broad category Eukarya, which means that only the remaining 4,458 MOTUs found a lower-level hit in the reference database. Of those 4,458 MOTUs, 2,998 (67.25%) were assigned at 80% or more similarity. Metazoa was the super-group with the highest number of MOTUs, and it accounted for the 49.53% (2,208) of clusters assigned below the Eukarya category. Stramenopiles and Archaeplastida were the second and third super-groups with the highest number of MOTUs (751 and 640, respectively). Figure 4 shows the total number of MOTUs per super-group, with (left) and without (right) MOTUs assigned to the broad category Eukarya.

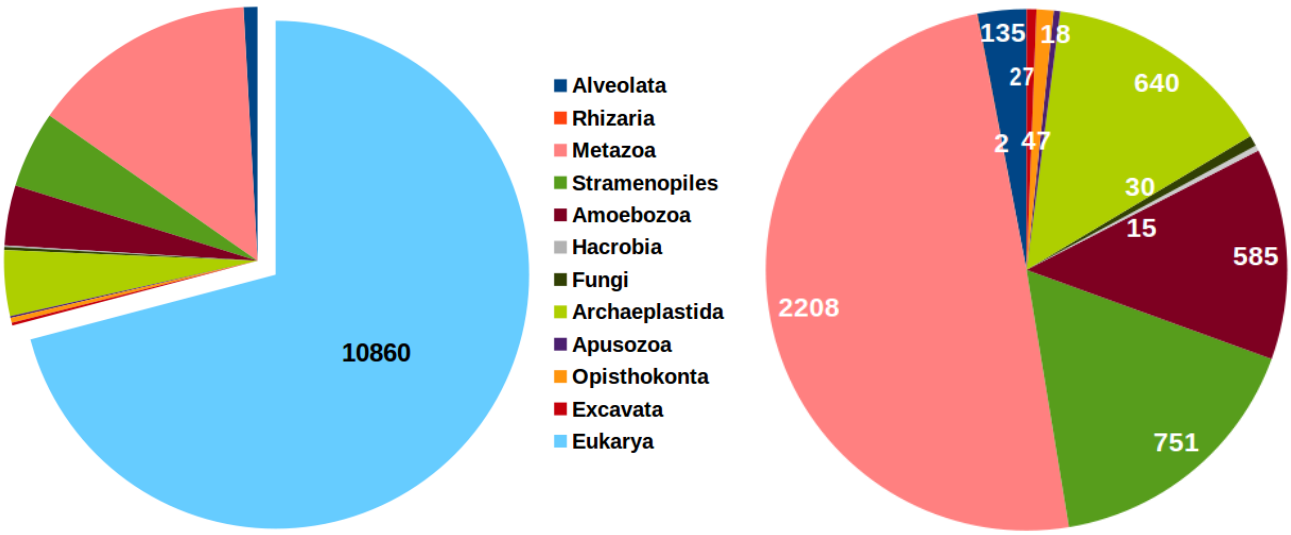


Figure 4. Number of MOTUs belonging to each of the 11 super-groups, with MOTUs assigned to Eukarya (left) and without (right). Numbers correspond to the total number of MOTUs per group.

We identified 13 phyla among metazoans: Annelida, Arthropoda, Bryozoa, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Nemertea, Platyhelminthes, Porifera, Rotifera and Xenacoelomorpha. However, it is worth noting that phyla which could potentially be present in our samples such as Chaetognatha, Ctenophora, Gastrotricha, Hemichordata, Kinorhyncha, Loricifera or Priapulida were not detected in our dataset. Additionally, it should be taken into account that metazoans from the phylum Chordata, order Salpida, were unintentionally excluded from our analysis due to the length of their COI region (304 bp) compared to the length range selected when performing the metabarcoding pipeline (309-318 bp). A relevant proportion of metazoan phyla, precisely the 71.83%, remained without further assignment. Arthropoda, Nematoda and Cnidaria were the metazoan phyla with the highest number of MOTUs (196, 109 and 88, respectively). Figure 5 shows the total number of MOTUs per metazoan phyla, with (left) and without (right) the group "Unassigned" (i.e., metazoan MOTUs that could not be assigned at the phylum or lower level).

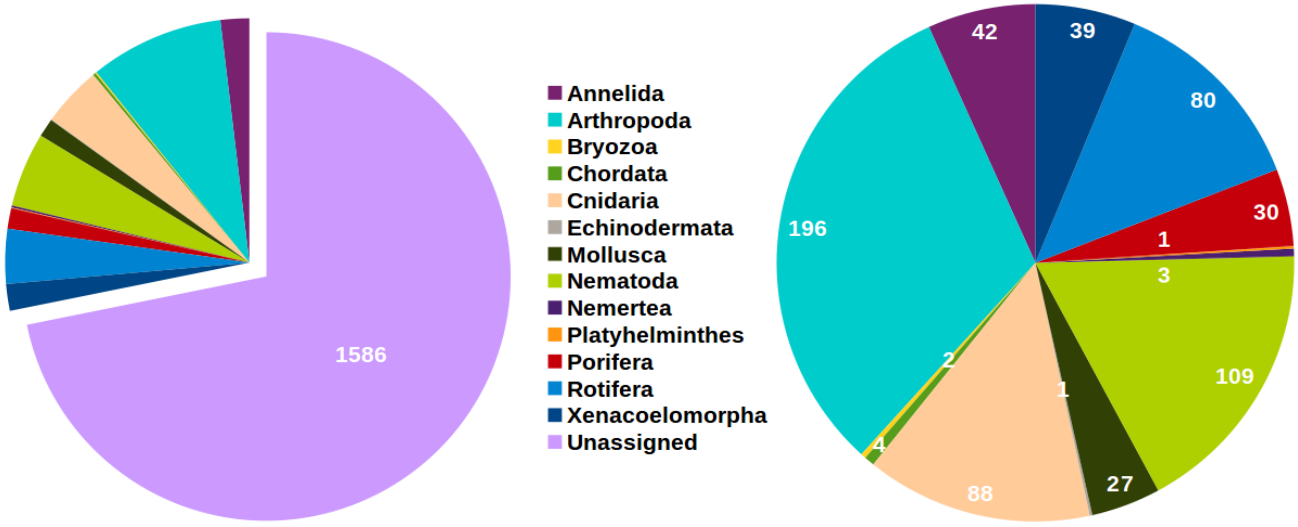


Figure 5. Proportion of MOTUs belonging to each of the 13 metazoan phyla, with (left) and without (right) further-unassigned MOTUs. Numbers correspond to the total number of MOTUs per group.

Frequencies of the taxonomic ranks at which MOTUs in our dataset were assigned by the ecotag procedure (without MOTUs assigned to Eukarya) are shown in Figure 6, left. The 3 most frequent categories were, in this order, super-group (1,917), species (841) and class (762), which collectively represent the 79% of assignments. Only one fifth of the MOTUs were assigned at the lowest taxonomic categories, that is, species and genus. Figure 6, right, shows the distribution of the 841 super-group MOTUs (Opisthokonta did not have any) assigned to species, where 52.32% of the MOTUs assigned to this rank belonged to Amoebozoa. It is interesting to point out that only 6% of metazoan MOTUs were assigned at species level, whereas 75% of the Amoebozoan were so.

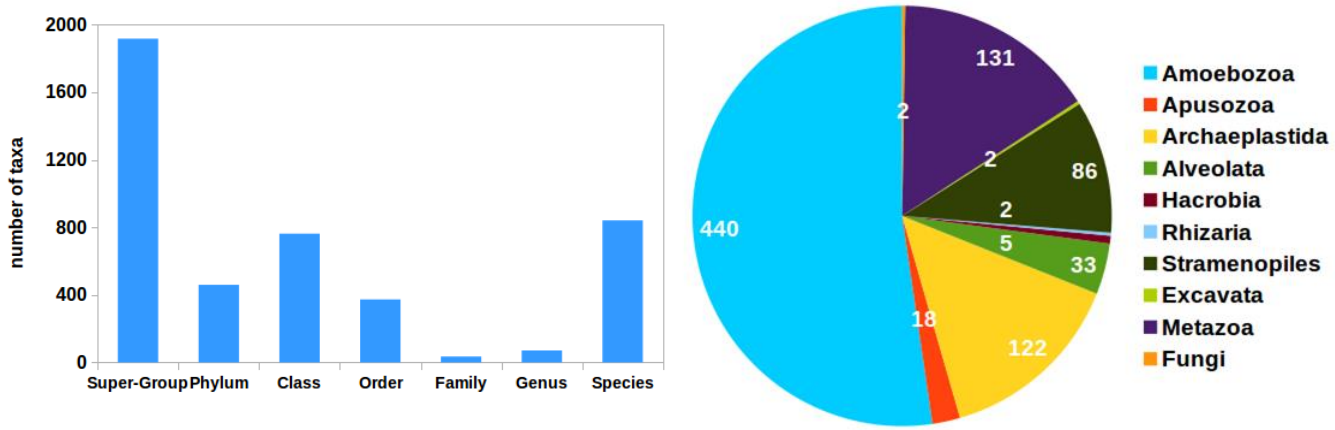


Figure 6. Left, assignment frequencies of MOTUs corresponding to the 11 super-groups by the ecotag procedure to the following taxonomic categories: super-group, phylum, class, order, family, genus and species. Right, proportion of MOTUs of the 11 super-groups assigned to the species category. Numbers correspond to the total number of MOTUs of each group.

Frequencies of the 3 most diverse super-groups are shown in Figure 7, left. Stramenopiles had 56.86% of its MOTUs assigned to class, while Archaeplastida had 55.16% assigned to phylum. For the metazoans assigned at phylum or lower rank, order was the most frequent rank, followed by species. When looking at the 3 main metazoan groups (Figure 7, right), order was the most frequent rank (57.14% in Arthropoda, 96.33% in Nematoda and 79.45% in Cnidaria).

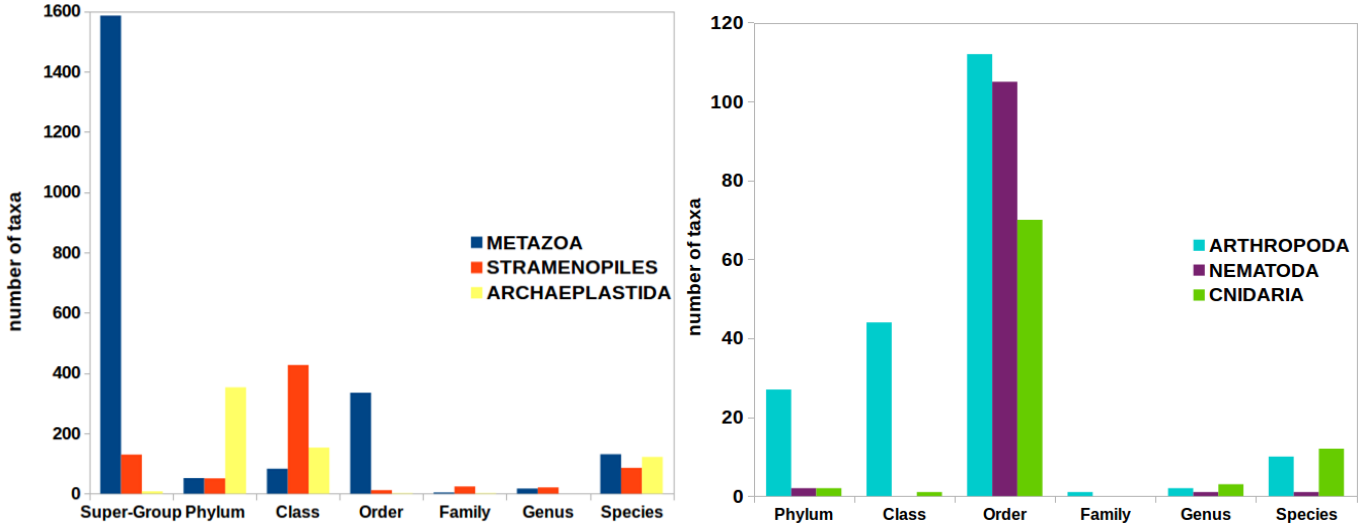


Figure 7. Left, assignment frequencies of MOTUs corresponding to the 3 most diverse super-groups by the ecotag procedure to the following taxonomic categories: super-group, phylum, class, order, family, genus and species. Right, assignment frequencies of MOTUs corresponding to the 3 most diverse metazoan phyla by the ecotag procedure to the following taxonomic categories: phylum, class, order, family, Genus and species.

Regarding the percentage of similarity between the clusters generated by the metabarcoding pipeline and the best hit in the customized reference database, Fungi holds the highest mean percentage among super-groups, whereas Excavata holds the lowest (86.49% and 73.82%, respectively). It is interesting to highlight that in Metazoa, the super-group with the highest number of MOTUs, assignments are

made at a mean percentage of similarity of 78.52%. This percentage is below 80% similarity, which is the minimum threshold for assignments to be considered reliable. There are two metazoan phyla with assignments made at mean percentages above 95% similarity: Chordata and Echinodermata. However, the number of organisms included in each taxonomic group are 4 and 1, respectively. The percentage of similarity in the most abundant phylum, Arthropoda, falls also below the 80% similarity threshold (79.85%), whereas the second and third, Nematoda and Cnidaria, are above it (81.67% and 84.41%).

Spatio-temporal patterns

As explained before, we processed 90 samples, 45 from the Autumn period and 45 from the Spring one, obtained in 2 consecutive cruises. The samples yielded 15,318 eukaryotic MOTUs in total, and only 4,458 of them were assigned to the 11 super-groups. When focusing on seasonal patterns, Autumn had slightly more MOTUS than Spring (3,461 compared to 3,369 MOTUs). Furthermore, after comparing the proportions of each group per season (Figure 8) (Autumn, outer circle; Spring, inner circle) no major differences were found in the resulting seasonal communities, picturing a stable community over time.

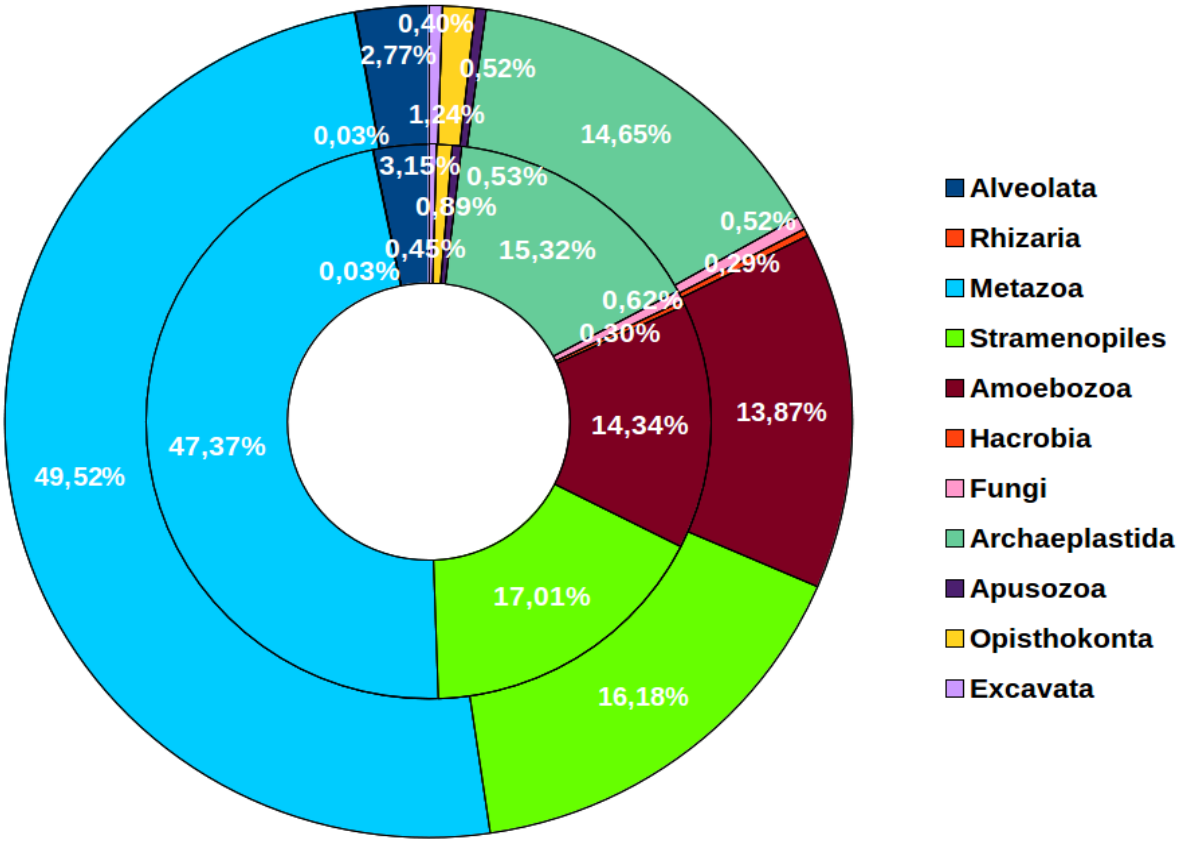


Figure 8. Proportion of MOTUs assigned to the 11 super-groups in Autumn (outer circle) and in Spring (inner circle). Numbers correspond to the percentage that each group represents in the resulting seasonal community.

Regarding Metazoa, Autumn was, as well, the season with the highest number of MOTUs compared to Spring (1,714 vs 1,596) and the difference was not significant either. In Figure 9 we observe the same pattern as with the super-groups: the proportion of the different phyla (Autumn, outer circle; Spring, inner circle) in the community remains fairly constant over time. However, there was one phylum where the number of MOTUs notably increased in Spring: Nematoda (from 68 to 89). By contrast, in the phylum Mollusca, the number of MOTUs decreased by half in Spring (from 25 to 12). Platyhelminthes were only represented in Spring, and Echinodermata, in Autumn, albeit with only 1 MOTU each.

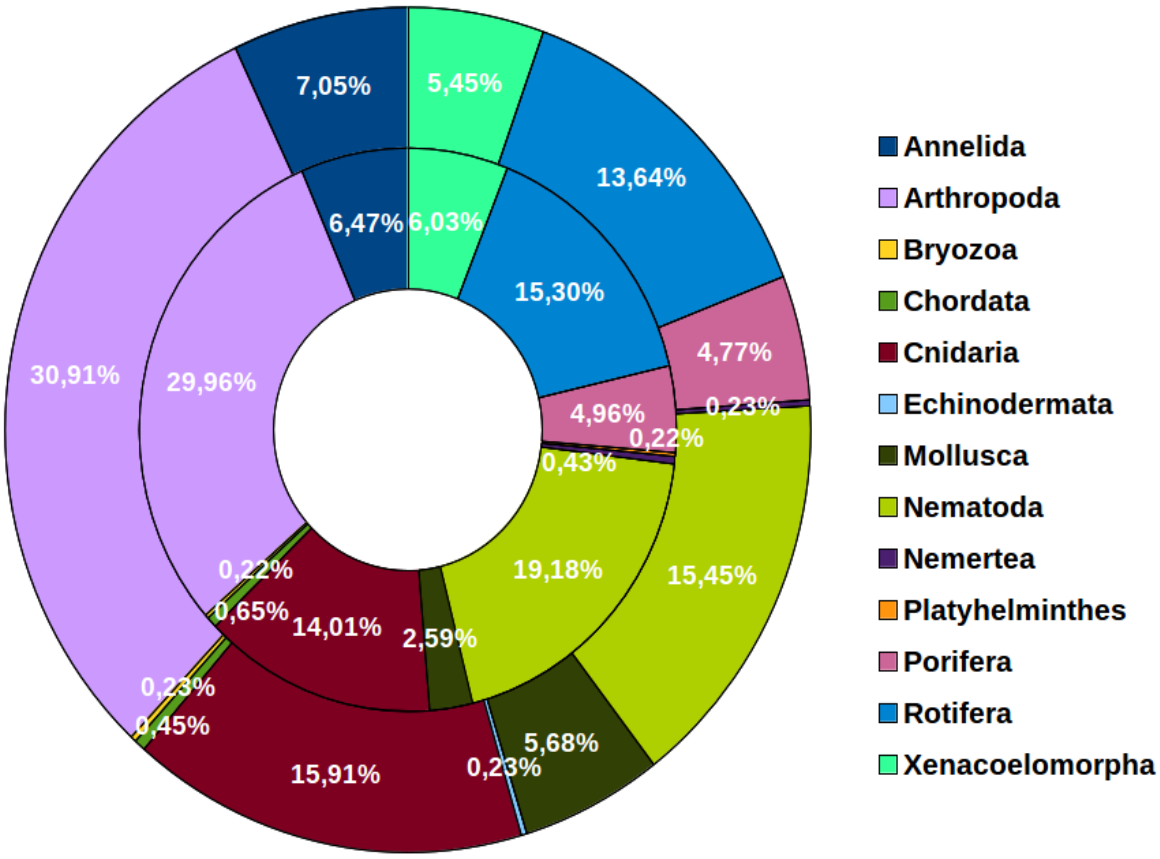


Figure 9. Proportion of MOTUs assigned to the 13 metazoan phyla in Autumn (outer circle) and in Spring (inner circle). Numbers correspond to the percentage that each group represents in the resulting seasonal community.

Figure 10 shows the seasonal patterns of MOTU richness per sample for the different super-groups, where we can observe certain trends. Overall, MOTU richness was higher in Spring samples in all super-groups except for Opisthokonta, Rhizaria, Hacrobia and Excavata. Metazoa was the most diverse group per sample, followed by Amoebozoa and Archaeplastida in Autumn, and by Amoebozoa and Stramenopiles in Spring. Nevertheless, statistical comparisons (t-tests) showed that seasonal differences were only significant in Alveolata, Stramenopiles and Apusozoa.

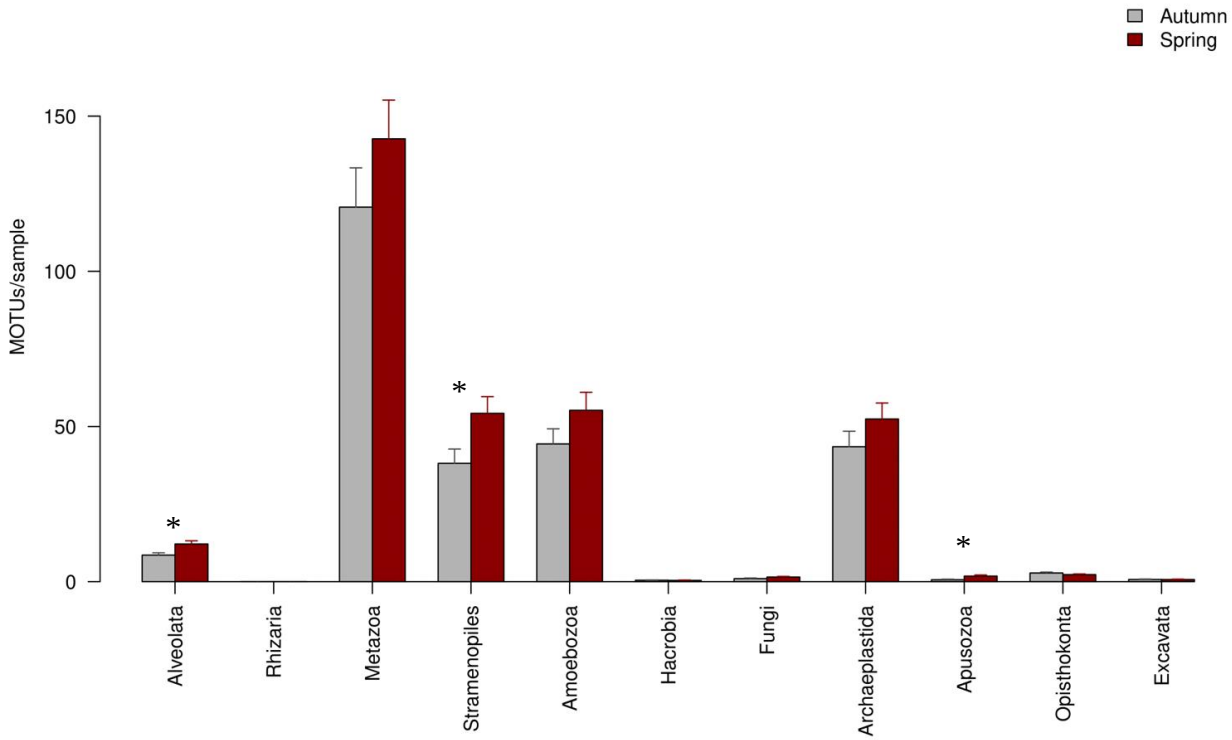


Figure 10. Mean number of MOTUs per sample and per season for the different super-groups. (*: significant differences between seasons assessed by t-tests). Bars represent standard errors.

A parallel study by phylum was carried out on the metazoans. Figure 11, up, illustrates the outstanding presence of unassigned metazoan MOTUs per sample in our data. We then provide the same figure (Figure 11, down) with only the metazoans assigned to the 13 phyla. Again, MOTU richness per sample was higher in Spring samples, with the exception of the phylum Mollusca. Arthropoda was the most MOTU-rich phylum, followed by Cnidaria and Rotifera in Autumn, and by Rotifera and Nematoda in Spring. Temporal differences were particularly remarkable in the two latter because both doubled their MOTU richness per sample in Spring. Statistical comparisons (t-tests) revealed that the only significant seasonal differences in the number of MOTUs per sample appeared in Chordata, Nematoda, Porifera and Rotifera.

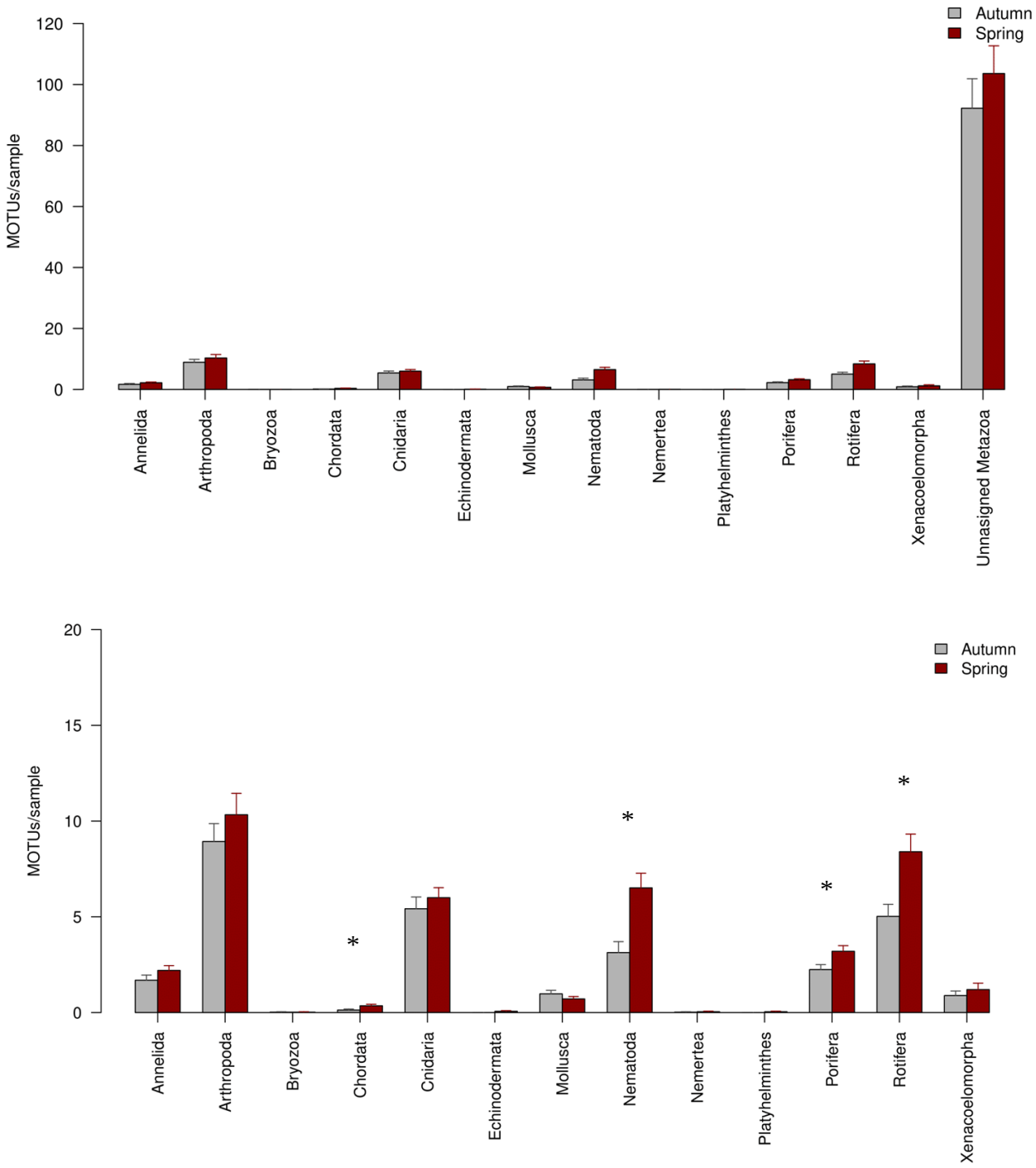


Figure 11. Mean number of MOTUs per sample and per season for the different metazoan phyla, with (up) and without (down) unassigned MOTUs. (*: significant differences between seasons assessed by t-tests). Bars represent standard errors.

The seasonal proportion of MOTUs of the main super-groups and metazoan phyla grouped by sampling zone (canyon and slope) and depth is shown in Figure 12. The proportion of super-group MOTUs appeared stable, but there was more variance in metazoan phyla.

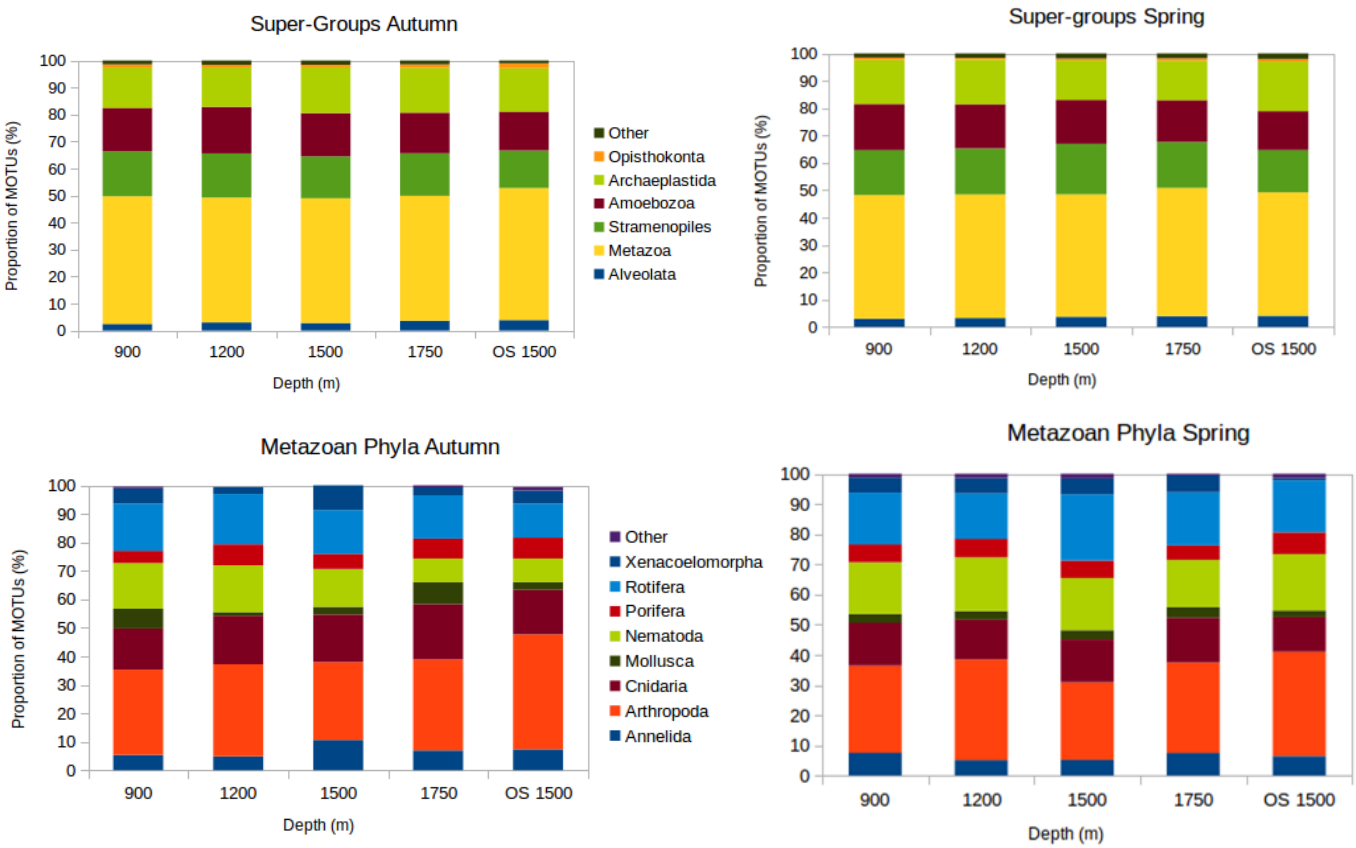


Figure 12. Proportion of MOTUs of the different super-groups (up) and metazoan phyla (down) per sampling station and season.

Regarding the vertical distribution of taxa of the global dataset by sediment layer, diversity decreased following the sediment profile. This way, the highest diversity was found in the most superficial layer (layer A, the first cm of sediment), followed by the intermediate layer (layer B, second cm of the sediment) and, finally, by the deepest one (third to fifth cm of the sediment), with 12,399, 9,025 and 6,495 MOTUs, respectively. Therefore, as shown in Figure 13, ca. 80% of all MOTUs were present in layer A; ca. 60% in layer B; and ca. 40% in layer C. It is interesting to note that, as well, the number of exclusive taxa per layer notably decreases as we move deeper in the sediment profile, from 29.1%, to 9.4%, ending in 5.6%. The percentage of MOTUs shared by layer A and B (19.1%) is higher than the one shared by layers A and C (6.4%) and by B and C (4.05%). The 26.3% of the 15,318 MOTUs were shared by the 3 sediment layers.

The number of MOTUs per sediment layer of the 11 super-groups is 3,609, 2,674 and 1,955, respectively. In metazoans, the total number of MOTUs also decreases from layer A to C (500, 347 and 218 MOTUs, respectively).

As regards the horizontal distribution, in the canyon, the number of MOTUs in the 11 super-groups decreased with depth, from the shallow stations located at 900 m depth to the deepest ones at 1,750 m depth (2,429, 2,304, 2,083 and 1,837 MOTUs, in that order), as well as of the 13 metazoan phyla

(318, 284, 263 and 231 MOTUs, in that order). However, it is interesting to point out that the lowest number of MOTUs corresponded to the stations located in the open slope, at 1,500 m depth (1,373 super-group MOTUs and 181 metazoan MOTUs).

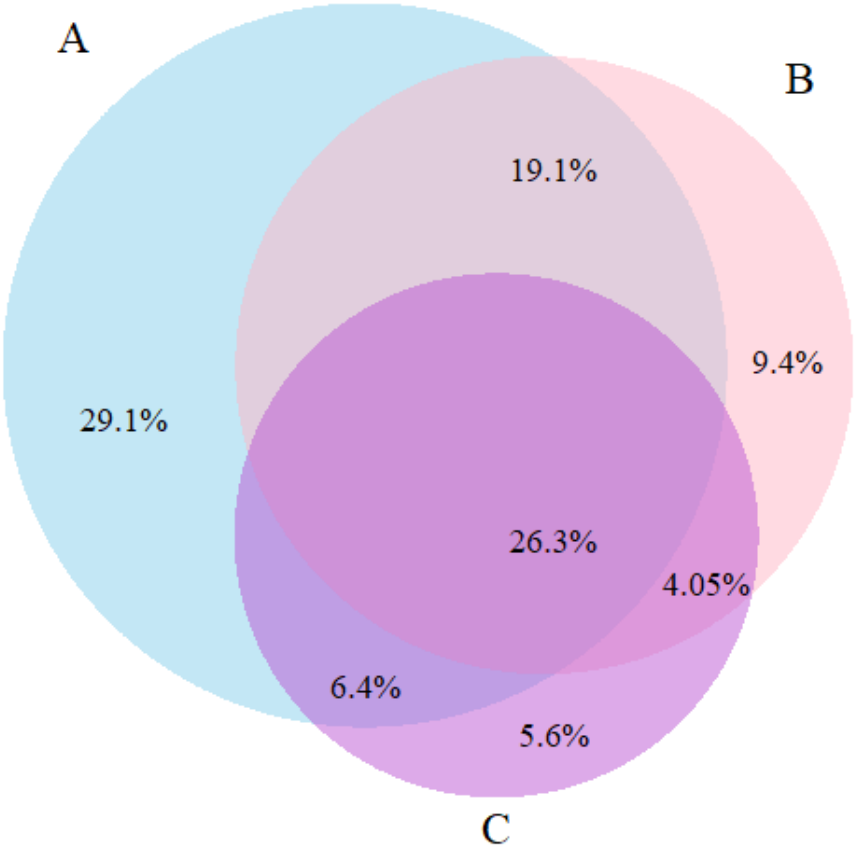


Figure 13. Venn diagram of the number of MOTUs found in the three sediment layers studied. Numbers correspond to the percentage of MOTUs.

A non-metric multidimensional scaling (or nMDS) analysis based on presence/absence data (Jaccard index) of the three sediment layers (with the three replicates of each locality pooled together) is shown in Figure 14. It shows that our data tended to group by layer, following the structure of the original sediment profile, even though there was a certain overlap between the inertia ellipses of consecutive layers. We can also observe differences among the extension of the inertia ellipses: samples in layer A were the least dispersed, followed by layer B, and the extension of the ellipse in layer C was the broadest. It is interesting to note that, in each layer, the samples from the open slope (represented by triangles) appeared separated from the samples from the canyon and closer to the samples from the open slope of the other layers. The stress value (9.33%), reflecting the disagreement between the 2-d configuration and the structure of the distance matrix, was below 10%, which is regarded to be an accurate representation.

A PERMANOVA analysis of the factors layer (fixed) and zone (random) showed significant effects of both factors in structuring the community. Both PERMDISP tests were not significant and,

therefore, revealed that the effects were not due to differences in the dispersion levels. The interaction between layer and zone was not significant either. As the factor layer had a significant effect, we performed pairwise tests between the 3 layers, and the only significant difference was found between layers A and C (all three analyses are shown in table 2).

PERMANOVA (LAYERS)	df	SS	Pseudo-F	P-value	Permdisp
Layer	2	0.973	2.200	0.002 *	0.221
Zone	1	1.307	5.914	0.001 *	0.094
Layer*Zone	2	0.543	1.229	0.127	
Residuals	24	5.305			

Pairwise comparisons	df	SS	Pseudo-F	P-adjusted
A - B	1	0.359	1.418	0.250
A - C	1	0.763	2.939	0.006 *
B - C	1	0.337	1.194	0.328

Table 2. PERMANOVA and PERMDISP tests of the factors Layer and Zone for the Jaccard index (*: significance of the factor). The results for permutational Pairwise tests of levels of the factor Layer are also provided (*: significant outcome after FDR correction).

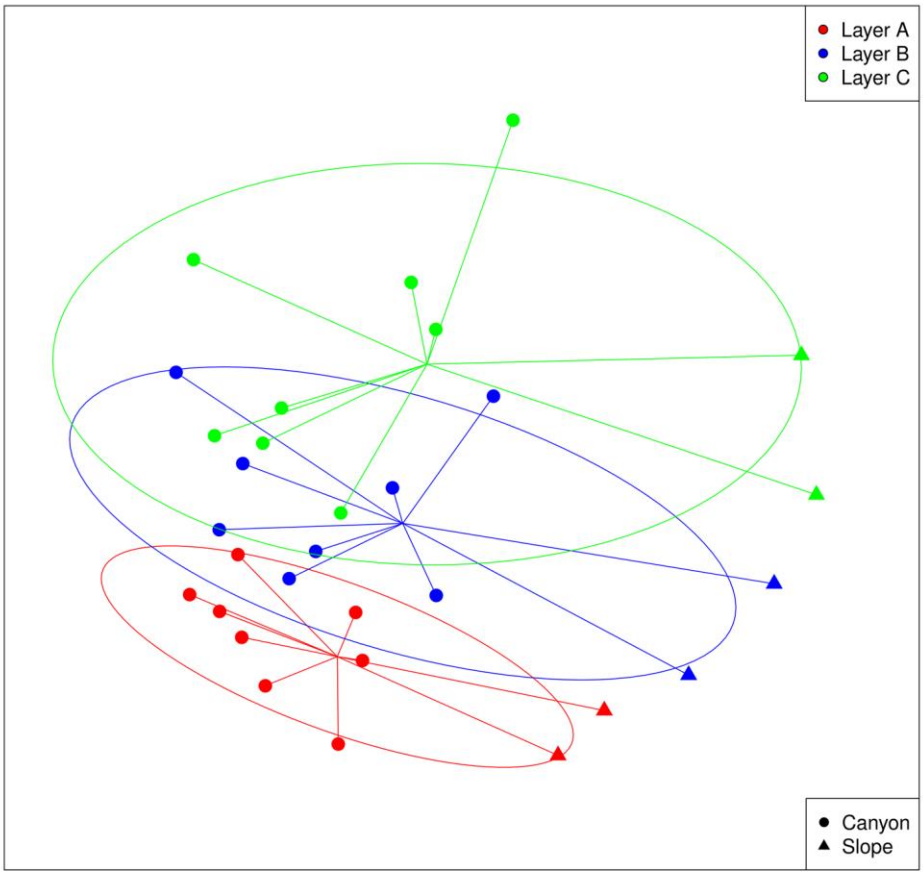


Figure 14. nMDS representation of the samples by layers (the three replicates per locality merged). The stress value of the final ordination is 9.33%.

A second nMDS was carried out in order to analyze spatio-temporal patterns between both localities and seasons (in this case, the three layers of each sample were pooled), and is displayed in Figure 15. The most evident characteristic of this ordination was the clear separation between samples from the

Blanes canyon and the ones belonging to the open slope, visualized as the lack of overlap in the inertia ellipses, which means that the community structures captured were clearly different. When analyzing the data per season, we can see that, even though the centroids of the inertia ellipses per season were separated, there was an overlap in their areas. The autumn samples in both zones reflected a noticeable higher dispersion than the spring ones, which can be interpreted as a higher heterogeneity in their community composition.

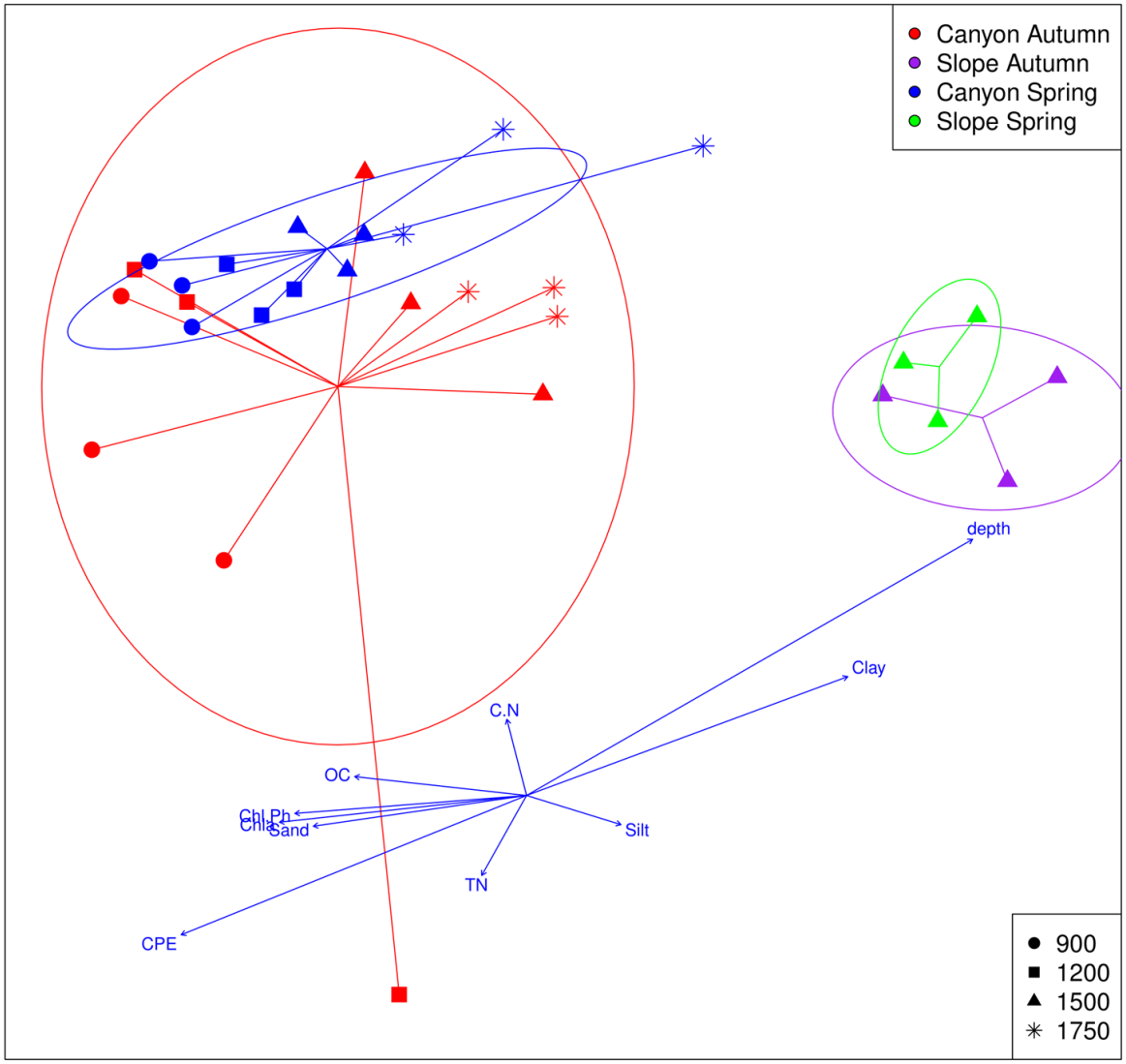


Figure 15. nMDS representation of the samples of the spatio-temporal study (the three samples per layer merged). The stress value of the final ordination is 9.17%. The fitted vectors Depth, Clay, Silt, TN, CPE, Sand, Chl a, Chl a/Phaeo, OC and C/N are added in the ordination, but displaced from the center for clarity.

PERMANOVA tests performed using season and zone (canyon, open slope) as factors attested that both factors had a significant effect, especially zone, in structuring the community, but their

interaction was not significant. When testing for dispersion with PERMDISP tests, the effect of Zone could be attributed to the dispersion of the samples (table 3).

PERMANOVA (BC+OS)	df	SS	Pseudo-F	P-value	Permdisp
Season	1	0.435	1.407	0.029 *	0.177
Zone	1	1.194	3.859	0.001 *	0.009 *
Season*Zone	1	0.307	0.992	0.409	
Residuals	26	8.043			

Table 3. PERMANOVA and PERMDISP tests of the effect of zone (canyon and slope) and season (autumn and spring) for the Jaccard index. The three layers of each sample were pooled. (*: significance of the factor).

We then proceeded with only samples from the canyon and performed a new PERMANOVA test with season and depth as factors. Both factors had a significant effect in structuring the analyzed community, but their interaction was not significant. The effect of both factors was not due to the dispersion of their samples, as the corresponding PERMDISP tests reflected. The results of the pairwise tests between depths revealed statistically significant differences between all depths except for 900m vs 1,200m and 1500m vs 1750 m (all three analyses are shown in table 4).

PERMANOVA (CANYON)	df	SS	Pseudo-F	P-value	Permdisp
Season	1	0.445	1.553	0.012 *	0.092
Depth	3	1.378	1.604	0.001 *	0.999
Season*Depth	3	0.986	1.147	0.054	
Residuals	16	4.584			

Pairwise comparisons	df	SS	Pseudo-F	P-adjusted
900-1200	1	0.297	0.976	1
900-1500	1	0.492	1.673	0.015 *
900-1750	1	0.644	2.168	0.02 *
1200-1500	1	0.418	1.375	0.029 *
1200-1750	1	0.555	1.805	0.015 *
1500-1750	1	0.349	1.178	0.25

Table 4. PERMANOVA and PERMDISP tests of the effects of season and depth on the samples from the Blanes Canyon for the Jaccard index (*: significance of the factor). The results for permutational pairwise tests of levels of the factor depth are also provided (*: significant outcome after FDR correction). The three layers of each sample were pooled.

The last test that we performed was the environmental fit by incorporating data related to biotic and abiotic variables from Roman et al (2016) (table 5). The vectors resulting after applying the test on the global dataset (their 2-d representation is incorporated in figure 15) showed that 2 variables were significantly related to the ordination: depth and CPE, with depth having the highest correlation (r^2=0.393). However, when analyzing data from the canyon only, depth was the only significant

variable. This result corroborates the spatial distribution of the canyon samples in the nMDS, distributed from left to right following the depth gradient.

ENVIRONMENTAL FIT	r2	Pr(>r)
Depth	0.393	0.001 *
Clay	0.16	0.138
Silt	0.013	0.839
Sand	0.059	0.464
CPE	0.191	0.044 *
Chl a	0.078	0.25
Chl a / Phaeo	0.068	0.416
OC	0.038	0.58
TN	0.017	0.846
C/N	0.013	0.968

Table 5. Environmental fit test to test the relationship of biotic and abiotic variables with the ordination presented in figure 15 for the Jaccard index (*: significance of the variable).

COI – 18S comparison

We compared our results with the ones obtained in Guardiola et al (2016) using DNA from the nuclear ribosomal gene coding for the small subunit (18S rRNA gene). Their final dataset consisted of 4,953 MOTUs and 4,685,457 reads after applying a similar metabarcoding pipeline, but with two main differences: they excluded MOTUs (i) not assigned to at least the category super-group; and (ii) that did not have at least 80% similarity with the best hit in the reference database. In our case, when excluding MOTUs assigned to Eukarya, the dataset consisted of 4,458 MOTUs and 880,696 reads; and the dataset would be reduced to only 3.016 MOTUs and 571,744 reads if we also excluded MOTUs assigned below the 80% threshold.

The mean number of reads per sample in Guardiola et al (2016) was higher (35,362), and rarefaction curves proved that their sequencing depth was also adequate to capture the diversity in the samples, with curves plateauing at ca. 20.000 reads. Nevertheless, the species accumulation curves shown in Figure 16 give a different perspective of the sampling effort needed to characterize the studied community. In this regard, the 18S curve leveled off with the sampling effort applied at values of number of MOTUs below 5,000, whereas the COI curve suggests that more sampling is needed in order to fully characterize the community, with an expected number of MOTUs well above 15,000.

Contrarily to what we found, Spring had a higher number of MOTUs than Autumn. Metazoa appeared as the most diverse super-group in both datasets, but the second and third positions differed: Alveolata and Rhizaria in the 18S dataset, and Stramenopiles and Archaeplastida in the COI dataset. It is interesting to note that Alveolata and Rhizaria occupy the fifth and eleventh position (the latter with

only 2 MOTUs) in the COI dataset and Stramenopiles and Archaeplastida were the fourth and seventh most diverse taxa in the 18S dataset.

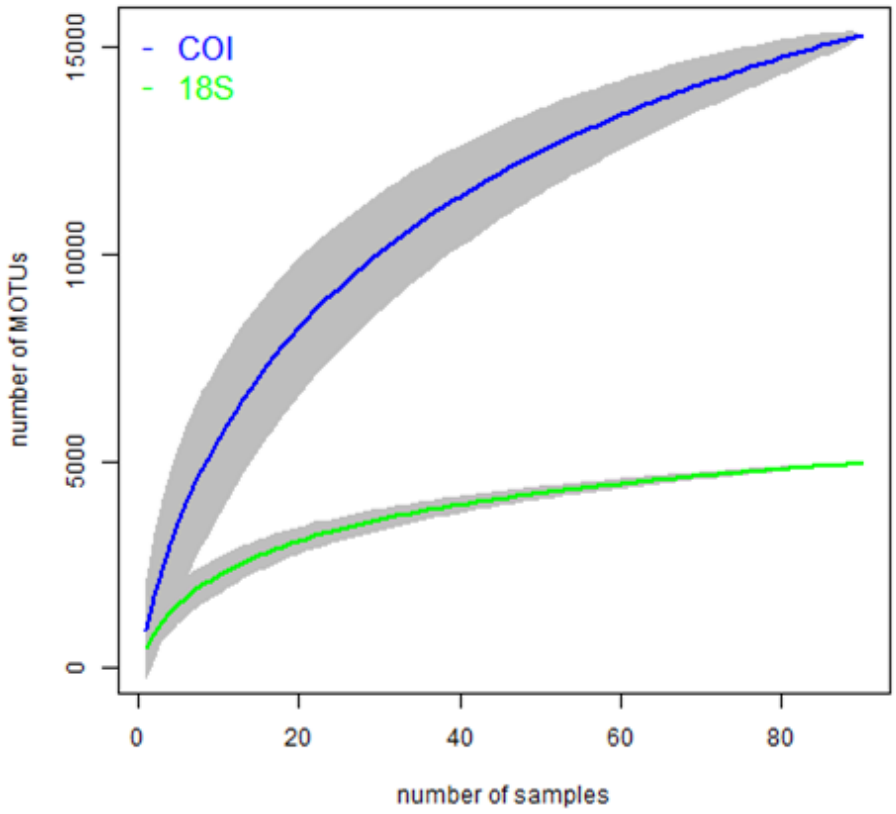


Figure 16. Species accumulation curves of the number of new MOTUs obtained after adding increasing number of samples to the analysis, with the confidence interval (95%) represented, for the COI and the 18S datasets.

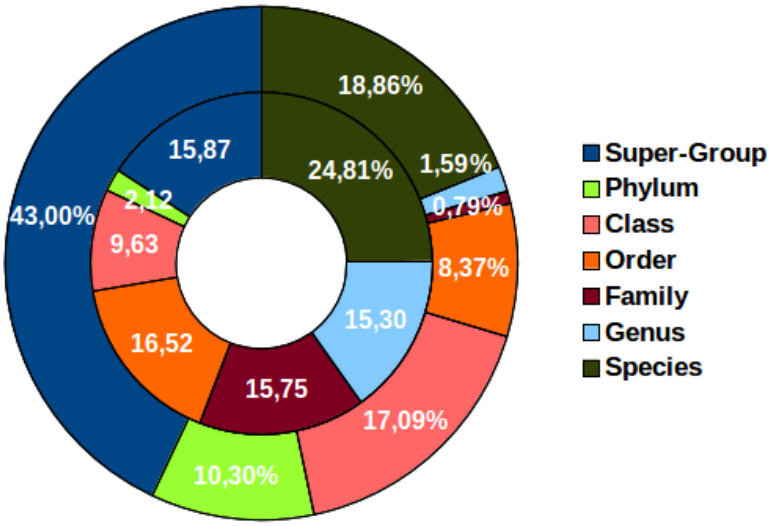


Figure 17. Proportion of MOTUs assigned to the following taxonomic categories: super-group, phylum, class, order, family, genus and species, in the COI dataset (outer circle) and in the 18S dataset (inner circle). Numbers correspond to the percentage of MOTUs that each group represents in each dataset.

Figure 17 shows the proportion of eukaryotic MOTUs assigned to each rank with COI (outer circle) and with 18S (inner circle). When using 18S, species is the most frequent rank, with the rest of MOTUs distributed quite proportionally among the other categories. However, with COI, there is an outstanding presence of MOTUs assigned to super-group, which is almost 3 times higher than with 18S. The second most frequent rank is species.

The comparison of MOTU richness per sample of the super-groups obtained with COI and with 18S (Figure 18) also implied noteworthy differences. In fact, statistical analyses (t-tests) revealed that only MOTU richness captured in 2 of the 11 super-groups, precisely Metazoa and Excavata, were not significantly different. Particularly noticeable are the differences obtained in Alveolata and Rhizaria: they are the first and third most MOTU-rich super-groups per sample in the 18S dataset while they have low abundance or are practically absent in the COI dataset. Moreover, the 18S dataset contains, in general, more MOTUs per sample than the COI one.

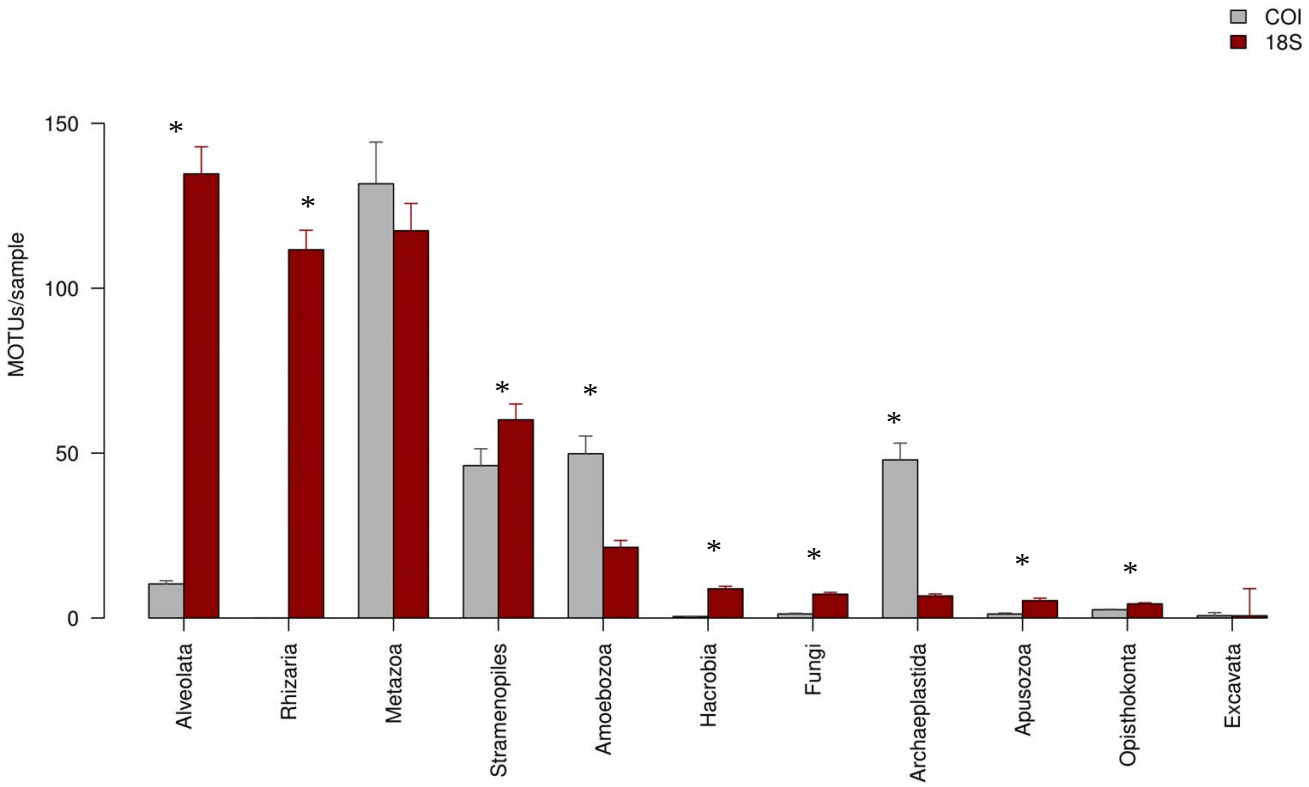


Figure 18. Mean number of MOTUs per sample and per genetic marker for the different super-groups. (*: significant differences between genetic markers assessed by t-tests). Bars represent standard errors.

Despite the fact that the 18S dataset from Guardiola et al (2016) contained sequences from 20 metazoan phyla and our dataset, only from 13 phyla, there are remarkable similarities. The 5 most diverse phyla in the 18S dataset were, in that order, Nematoda, Arthropoda, Annelida, Platyhelminthes and Cnidaria, whereas in the COI dataset were, respectively, Arthropoda, Nematoda, Cnidaria, Annelida and Xenacoelomorpha. It is interesting to highlight that targeting the 18S rRNA gene enabled the assignment of 651 MOTUs to Nematoda, whereas only 109 could be assigned to

the same taxon when targeting the COI gene. Moreover, it is striking that Platyhelminthes, the fourth most abundant phylum in the 18S dataset, was represented by only 1 MOTU in the COI dataset.

Regarding MOTU richness, in Figure 19 we observed a highly significant difference in unassigned MOTUs per sample: whereas the COI dataset contained around one hundred unassigned metazoan MOTUs per sample, the 18S dataset had around 10. When focusing on the 13 metazoan phyla (Figure 20), all differences were statistically significant. Again, MOTU richness is, in general, higher in the 18S dataset, with especially noteworthy differences in 2 phyla: (i) Nematoda, which had the highest number of MOTUs per sample in the 18S dataset, was the fourth most MOTU-rich phylum in the COI dataset; and (ii) whereas Chordata was the third most MOTU-rich phylum in the 18S dataset, was one of the least MOTU-rich phylum in the COI dataset.

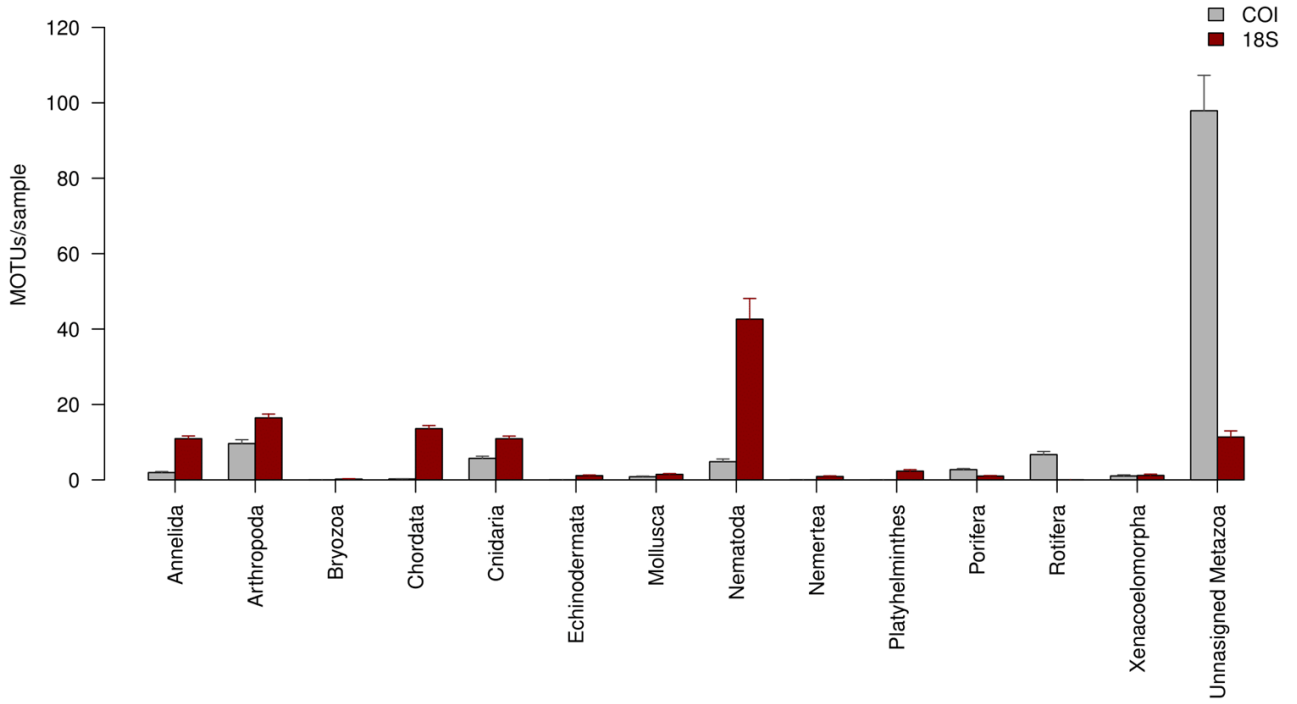


Figure 19. Mean number of MOTUs per sample and per genetic marker for the different metazoan phyla, including unassigned MOTUs. Bars represent standard errors.

Coincidentally, the diversity in the 3 layers of the sediment analyzed with the 18S gene also decreased following the depth gradient, as well as the number of exclusive MOTUs per layer. As regards MOTUs shared by the 3 layers, whereas in our dataset, 26.3% of MOTUs were shared, the 38.95% were shared in the 18S dataset. Consequently, our dataset detected a more heterogeneous community.

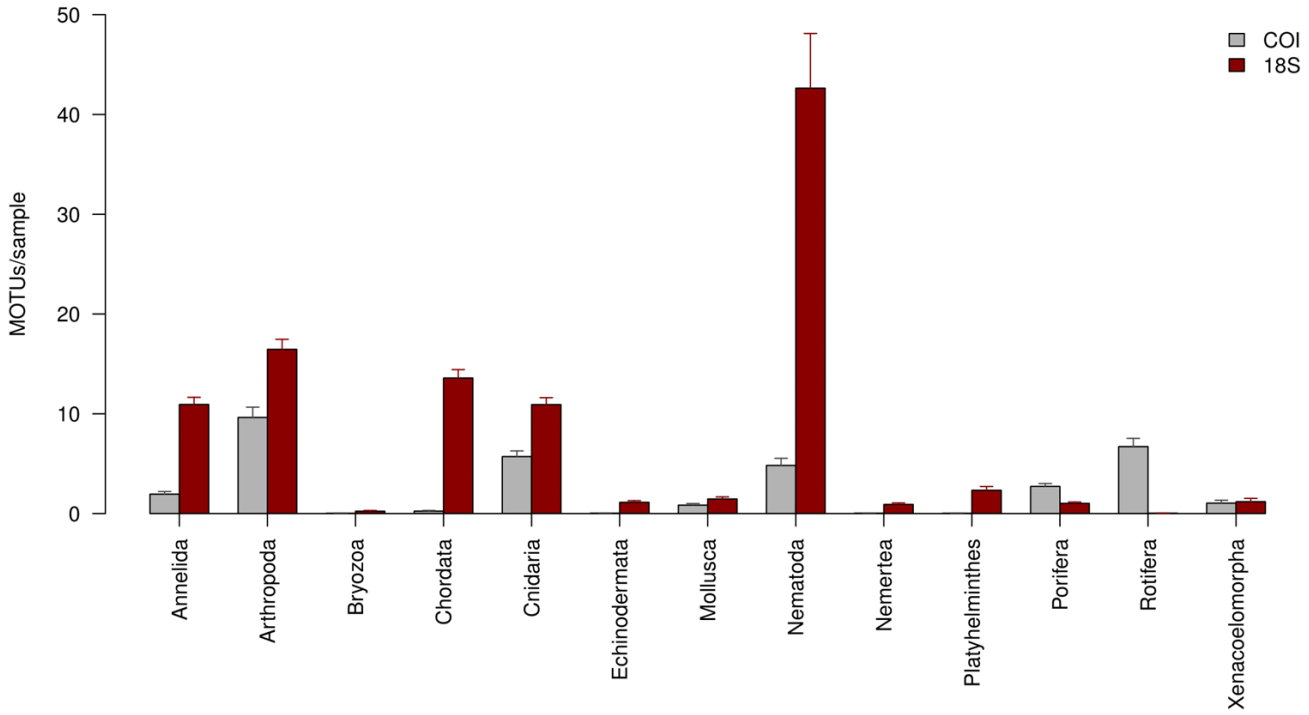


Figure 20. Mean number of MOTUs per sample and per genetic marker for the different metazoan phyla. Bars represent standard errors.

When comparing the nMDS based on presence/absence data for the 3 layers of sediment (Fig. 14 and Fig. 4 in Guardiola et al 2016), we observed that, disregarding the type of marker, samples were grouped following the gradient in layer depth, and samples from the open slope were also set apart. Nevertheless, there were differences among the inertia ellipses: while in the 18S nMDS the inertia ellipse of layer A was the broadest, reflecting more heterogeneity in the community structure, followed by layer B and layer C, in that order, in the COI nMDS we found the reverse pattern, the inertia ellipse of layer A reflects a more homogeneous layer compared to B and C, the latter being the most heterogeneous.

The ordinations provided by both markers in the second nMDS (Fig. 15 and Fig. 5 in Guardiola et al 2016) had in common the clear distinction between samples from the open slope and from the canyon (even though it was clearer in the COI ordination). Focusing on the canyon, there was a coincidence in both ordinations: the centroids of the inertia ellipses representing samples from different seasons appeared separated, but the inertia ellipses overlapped, and there appeared to be a higher heterogeneity in samples from Autumn in both cases. Similarly, depth happened to be an environmental factor structuring both datasets, but it should be noted that the 18S ordination did not include the analysis of the other 9 environmental factors tested in the COI one.

We performed two Mantel Tests in order to compare the distance matrices used to obtain the MDS ordinations of both markers. In both cases the correlation coefficients were high (0.699 for the layer ordination, 0.845 for the spatio-temporal ordination) and significant ($p<0.001$).

DISCUSSION

The deep-sea represents the largest biome, covering more than 50%, of the Earth's surface. However, it is the least studied environment, mainly due to its vast extension, to the variability of habitats and environmental conditions and to its remoteness, which precludes an easy access of human resources and the necessary devices to explore it and carry out sampling campaigns. Even though it is said that species richness decreases with depth as a norm, the topographic and oceanographic characteristics of the Western Mediterranean make scientists claim that it is a biodiversity hot-spot (Danovaro et al 2010). The main aim of this thesis was to provide an insight of the community profile and of the seasonal (autumn versus spring), vertical (by layers of sediment) and horizontal (by depth) distribution patterns of sediment communities located in a large submarine canyon and on the adjacent open slope in the Western Mediterranean Sea.

Among the available tools to carry out these analyses, we chose metabarcoding for the reasons stated earlier, even though the resolution of the data obtained after applying a metabarcoding pipeline should be put in perspective and analyzed in light of the marker used and the reference database available related to that marker. There are two main gene barcodes that can target the whole tree of life: COI and 18S rRNA, and in this thesis we amplified a fragment of the hypervariable marker COI in our samples, but were able to compare our results with the results obtained after the amplification of a variable fragment of the 18S rRNA from the very same samples. That gave us a unique opportunity to compare the similarities and differences between the datasets retrieved using both markers.

Moreover, it is important to note that recovering DNA directly from sediments without any sieving includes more extracellular DNA, but it has the risk of incorporating non-benthic organisms. Additionally, selecting large DNA fragments, such as > 300 bp, will favour the amplification of DNA from living (or recently dead) organisms as compared to the shorter (ca. 130 bp) fragment of the 18S rRNA used by Guardiola et al 2016, because DNA molecules tend to fragment and degrade over time.

The amplification of the COI region (and performance of the described bioinformatic pipeline) yielded a high number of MOTUs (15,318), yet the number of MOTUs that received assignments below the eukarya category was relatively small (4,458). Therefore, 70.9% of the eukaryotic clusters found by the Swarm algorithm (and that survived the refinement process) remained unknown (figure 4). There was also a generally low similarity between sequences in our dataset and their best hit in the reference database, which was ca. 80% on weighted average for all the 11 super-groups together and ca. 82% on weighted average for all the 13 metazoan phyla together. In fact, less than 1% of MOTUs were assigned at similarities above 90%. The similarity of MOTUs to their respective best-matches in the reference database can provide an idea of the degree of completeness of the reference database, together with the number of MOTUs that remain with unknown taxonomic identification

and assigned only to higher ranks by the ecotag procedure. Furthermore, assignments below the eukarya category were made at high taxonomic levels (figure 17, outer circle), with the 53.3% of assignments corresponding to the super-group and phylum levels, and only 20% at the species and genus level. At the species level, 52.32% of assignments belonged to organisms of only one super-group, Amoebozoa. In the metazoans assigned below the kingdom level, order was the most frequent rank (53,86%).

All this evidence led us to conclude that there are major gaps in reference databases that use COI as marker regarding sediment communities, or even regarding marine organisms in general, located in the Western Mediterranean Sea. The importance of these gaps is increased by the natural variability of the marker. Thus, the lower variability of 18S makes it possible to identify a sequence to a lower taxonomic rank whenever a related sequence, even if it comes from a far relative, is present in the reference database. Conversely, identification with the hypervariable marker COI needs a closely related sequence to be present in the reference database, which is not the case for many organisms living in Mediterranean deep-sea ecosystems, particularly the small ones comprising the meio- and micro-organisms present in the sediments.

Rarefaction curves, as a function of number of reads per sample, reached in general a plateau (figure 2), which can be understood as an adequate sequencing depth in our samples that reached, or was close to, MOTU saturation. That is, that the vast majority of MOTUs present in every sample was retrieved during its processing, providing reliable MOTU richness values. Nevertheless, the rate of accumulation of new MOTUs (or organisms) as more samples are analyzed (figure 3) was far from acquiring an asymptotic shape. Consequently, sampling effort should improve by increasing the number of replicates per locality in order to carry out exhaustive community inventories, with the aim of acquiring a more complete view of the biodiversity present, even more in allegedly biodiversity hot-spots like submarine canyons and open slopes in the Western Mediterranean Sea.

Focusing on MOTU diversity, even though we found representatives of the 11 super-groups, only 4 groups, Metazoa, Stramenopiles Archaeplastida and Amoebozoa, accounted for 93.85% of the diversity. As regards the metazoans, 13 phyla were represented in our dataset. Diversity was more spread among the different phyla, with the 4 main groups (Arthropoda, Nematoda, Cnidaria and Rotifera) accounting for 76.05% of the MOTUs. Notwithstanding, it is important to note that the community profiling described herein could suffer modifications in the future if the eukaryotic sequences and clusters that remained unassigned (70.9% of all MOTUs) found a hit in more complete reference databases in future updates.

An important decision that should be taken when designing a monitoring program based on metabarcoding is the type of information that one wants to obtain: an exhaustive list of the organisms living in an ecosystem or habitat, or broad ecological information about spatio-temporal patterns.

The number of MOTUs per season were not significantly different, neither in the 11 super-groups nor in the 13 metazoan phyla, albeit in Autumn the dataset was slightly more diverse than the Spring one. According to figures 8 and 9, the composition of the sediment communities remained mainly homogeneous at the super-group and metazoan phyla levels over time. MOTU richness per sample followed a different pattern: Spring samples contained a higher number of MOTUs than Autumn samples in 7 out of the 11 super-groups and in 12 out of the 13 metazoan phyla, yet the only statistically significant seasonal differences were found in 3 super-groups and 4 metazoan phyla. This apparent discordance means that, even though the list of taxa was slightly larger in Autumn as a whole, the number of MOTUs contained in individual Spring samples was higher than the Autumn samples.

Taxonomic richness decreased following the vertical structure of the sediment, from the 1st cm down to the fraction that comprised the 3rd to the 5th cm, which is consistent with the known decreasing abundance of meiofauna in the first centimeters of sediment (Romano et al, 2013, Kalogeropoulou et al, 2016). Overall, layer A hosted almost twice the MOTUs found in layer C and 27.21% more MOTUs than layer B, and only 26.3% of all MOTUs inhabited (or were present in) the three layers. Each layer seemed to harbor distinct ommunities, albeit only layers A and C differed significantly in a PERMANOVA analysis, and the latter was the most heterogeneous layer. PERMANOVA analysis also showed that layer and zone (canyon or open slope) were significant factors structuring the studied community, and that these differences were due to community composition, and not to different variance (dispersion). The horizontal structure of the sediment also experienced a downfall in taxonomic richness by depth, which is also consistent with the findings presented in the review by Costello & Chaudhary (2017).

The second nMDS showed that the community found in the canyon was significantly different from the community found in the adjacent open slope, and that the communities resulting from samples from the same localities could be well differentiated by season, being the composition of the Autumn community far more heterogeneous, and revealing a strong temporal component of variation. However, in order to formulate better conclusions in this regard, it would be essential to analyze the annual cycle, as well as to introduce more replicates. A separate analysis of the samples from the canyon revealed that depth was also a significant factor structuring the community, and that the communities found at a given depth are significantly different from communities found separated by two or more depth levels, except for communities from 1,200 m and 1,500m depth, which are not

significantly different. Out of a list of 10, 2 environmental factors were significantly correlated with the global ordination: depth and CPE. However, only depth was significant when analyzing separately the samples from the canyon. CPE were the sum of Chl a and phaeopigments (its degradation products) and was used to estimate the organic matter produced by surface primary productivity. That is, it was a proxy for food signal in the sediment together with Chl a. The fact that CPE was an important structuring factor is in line with what was found in Romano et al (2016). The authors found a decrease in sedimentary food sources along the depth gradient and its values were much lower at the slope than in the canyon, which was taken by the authors as an argument supporting the important role played by canyons in channeling organic inputs.

Moreover, even though the number of useful reads obtained when targeting the COI region was less than that obtained when targeting 18S rRNA (2,068,065 and 4,685,457, respectively), the number of MOTUs was 3 times higher in the COI dataset (15,318 and 4,953, respectively). This is coherent with previous findings (Tang et al 2012, Cowart et al 2015, Wangensteen et al 2018), and is likely due to the much lower variability of the 18S gene, that underestimates the true number of species (Tang et al 2012). We can also point out that Guardiola et al (2016) used a different clustering algorithm (CROP), and that might have influenced the number of MOTUs recovered. Additionally, the 18S marker is known to have low taxonomic resolution for many eukaryotic groups. This means that sequences from different organisms can be identical. In short, with 18S we trade the capacity for assigning more MOTUs successfully thanks to its low variability with the low taxonomic resolution achieved (we cannot rely on low-level taxonomic assignments as several taxa can share the same sequence). With COI we detect a much higher number of MOTUs (often more than nominal species, Tang et al 2012) at the cost of many of them remaining anonymous (but when a low level assignment is made at high percent similarity, we can be reasonably sure that the correct species has been assigned). This differential taxonomic resolution of both markers is also related to the divergent species accumulation curves obtained (figure 16), where the number of MOTUs found when targeting the COI region do not seem to capture the diversity present in the whole area.

Metazoa was, in both datasets, the most diverse super-group, even though in the COI dataset it represented the 49.53% of the MOTUs that could be assigned to super-groups, whereas it represented the 33.49% in the 18S dataset. In addition, the 18S dataset included 7 additional phyla not present in the COI dataset. The five most MOTU-rich phyla were practically the same in both datasets, but in different order and in different proportions. Besides, MOTU richness per sample obtained with each dataset supported the differences in the total number of MOTUs per group because almost all pairwise comparisons were significantly different (figures 18 and 20). Therefore, one would obtain a completely different taxonomic community profile depending on the choice of marker to be

amplified. Primer bias, or the differential capacity of the metabarcoding primers to bind to sequences of some groups, can also be a concern when choosing a marker, as it has been pointed out that COI lacks true universal primers due to an excess of variability (Deagle et al 2014), although this has been counteracted with the development of new sets of degenerate primers such as the one used here targeting the second half of the standard barcode fragment (Wangensteen et al 2018). For 18S, several primer pairs exist that target several variable regions of the gene and that are reasonably universal (Hadziavdic et al 2014).

Notwithstanding the above, spatio-temporal patterns rendered by each marker did match, in broad terms. First of all, diversity decreased following a layer depth gradient, and factors layer and zone (canyon and open slope) were key in structuring the community, albeit each marker provided a different perspective regarding the degree of heterogeneity of the community inhabiting each layer. Additionally, there was a clear distinction between communities from the canyon and the open slope, and within each zone, from autumn and spring, being the former communities more heterogeneous. Finally, depth resulted to be an important factor when trying to explain community structure, with communities gradually changing and being significantly different at lags of two or more depth levels.

In summary, our results suggest that, despite the many expected differences in the performance of both markers, COI can retrieve broadly the same ecological information than the more frequently used marker 18S from eDNA samples of Mediterranean deep-sea ecosystems. The number of MOTUs retrieved is much higher with COI, but the taxonomic assignment can be considered poor compared to that obtained from 18S, which is mainly attributable to a higher natural variability of the former, coupled with significant gaps in current reference databases. The higher variability of COI makes it a more promising marker for developing high-resolution bioindicators of particular ecological conditions, compared to 18S. These developments should rely, for the time being, in taxonomy-free methods, which would allow to detect ecological shifts, even if the indicator sequences do not yet have a taxonomic identification. Finally, the applications of COI are being hampered by significant gaps in the reference databases for deep-sea organisms, which would need to be filled in future barcoding projects, if taxonomy-based applications are to be developed.

ACKNOWLEDGEMENTS

It has been an extremely challenging journey, but the fulfilment of this master's thesis project gave me the opportunity to acquire previously non-existent bioinformatic skills, to appreciate the value of eDNA metabarcoding as an essential tool to bring some light to the still scarce knowledge of deep-sea sediment communities, and to discover that I am capable of overcoming even the biggest challenges.

I would like to express my gratitude to my internal supervisors, Owen Wangensteen and Kim Præbel, for giving me the opportunity to execute this utterly interesting master's thesis project and to work remotely during a personal moment when I needed it the most.

My infinite gratitude goes to my external supervisor, Xavier Turón, who has been an amazing support and has provided extremely valuable help and knowledge in key moments until the very end.

A big thank you to Adrià Antich, who was of inestimable help and had an endless patience.

But my special gratitude goes to my family:

- To my parents, Antonio and Esther, for always being there, for backing me up when I decided to take a new path in life and for instilling in me the mindset to take action for one's beliefs.
- To my brother Alex, for always pushing me to be a better version of myself and to always give the best of me.
- To my beloved grandmother Fernanda, for always being an example of strong woman even when fighting against the demons of dementia.
- And to my dogs, Pippa and Romy, for their unconditional love.

Funding statement: The research reported was funded by projects DOSMARES (CTM2010-21810), CHALLENGEN (CTM2013-48163) and PopCOmics (CTM2017-88080) of the Spanish Government.

REFERENCES

Amblas, D., Canals, M., Urgeles, R., Lastras, G., Liquete, C., Hughes-Clarke, J. E., Calafat, A. M., 2006. Morphogenetic mesoscale analysis of the northeastern Iberian margin, NW Mediterranean Basin. *Marine Geology,* 234(1-4):3–20. DOI 10.1016/j.margeo.2006.09.009.

Benjamini, Y-, Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165-1188

Bonnin, J., Heussner, S., Calafat, A., Fabres, J., Palanques, A., Durrieu de Madron, X., Canals, M., Puig, P., Avril, J. & Delsaut, N., 2008. Comparison of horizontal and downward particle fluxes across canyons of the Gulf of Lions (NW Mediterranean): Meteorological and hydrodynamical forcing. *Continental Shelf Research,* 28–1957:1970. DOI 10.1016/j.csr.2008.06.004.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* 16:176–182 DOI 10.1111/1755-0998.12428.

Canals, M., Company, J.B., Martín, D., Sànchez-Vidal, A., & Ramírez-Llodrà, E., 2013. Integrated study of Mediterranean deep canyons: Novel results and future challenges. *Progress in Oceanography,* 118:1–27. DOI 10.1016/j.pocean.2013.09.004.

Canals, M., Puig, P., Durrieu de Madron, X., Heussner, S., Palanques, A. & Fabres, J., 2006. Flushing submarine canyons. *Nature,* 444:354–7. DOI 10.1038/nature05271.

Costello, M.J., Cheung, A., and De Hauwere, N., 2010. Topography statistics for the surface and seabed area, volume, depth and slope, of the world's seas, oceans and countries. *Environ. Sci. Technol*. 44, 8821–8828.

Costello, M.J. & Chaudhary, C., 2017. Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current Biology,* 27, R511–R527.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE.* 10:e0117562 DOI 10.1371/journal.pone.0117562.

Danovaro, R., Company, J.B., Corinaldesi, C., D'Onghia, G., Galil, B., Gambi, C., Gooday, A.J., Lampadariou, N., Luna, G.M., Morigi, C., Olu, K., Polymenakou, P., Ramirez-Llodra, Eva, Sabbatini, A., Sardà, F., Sibuet, M., Tselepides, A., 2010. Deep-Sea Biodiversity in the Mediterranean Sea: The Known, the Unknown, and the Unknowable. *PLOS ONE*, 5(8): e11832. https://doi.org/10.1371/journal.pone.0011832

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10(9), 20140562-20140562. DOI 10.1098/RSBL.2014.0562.

Dell'Anno, A. & Danovaro, R., 2005. Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science*, 309: 2179.

Eakins, B.W. & Sharman, G.F., 2010. Volumes of the World's Oceans from ETOPO1, NOAA National Geophysical Data Center, Boulder, CO.
https://www.ngdc.noaa.gov/mgg/global/etopo1_ocean_volumes.html

Geller, J., Meyer, C., Parker, M., Hawk, H., 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* 13:851–86. DOI 10.1111/1755-0998.12138.

Guardiola, M., Wangensteen, O.S., Taberlet, P., Coissac, E., Uriz, M.J. & Turon, X., 2016. Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ* 4:e2807. DOI 10.7717/peerj.2807.

Guillén, J., Bourrin, F., Palanques, A., Durrieu de Madron, X., Puig, P., & Buscail, R., 2006. Sediment dynamics during wet and dry storm events on the Têt inner shelf (SW Gulf of Lions). *Marine Geology,* 234(1-4), 129–142. DOI 10.1016/j.margeo.2006.09.018.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41:D597–D604. DOI 10.1093/nar/gks1160.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., Troedsson, C., 2014. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE*, 9(2): e87624. DOI 10.1371/journal.pone.0087624

Heussner, S., Durrieu de Madron, X., Calafat, A., Canals, M., Carbonne, J., Delsaut, N., & Saragoni, G., 2006. Spatial and temporal variability of downward particle fluxes on a continental slope: Lessons from an 8-yr experiment in the Gulf of Lions (NW Mediterranean). *Marine Geology,* 234:63–92. DOI 10.1016/j.margeo.2006.09.003.

Hu, M., Chilton, N. B., Zhu, X., & Gasser, R. B., 2002. Single-strand conformation polymorphism-based analysis of mitochondrial cytochrome c oxidase subunit 1 reveals significant substructuring in hookworm populations. *Electrophoresis*, 23:27–34. DOI 10.1002/1522-2683(200201)23:1<27::aid-elps27>3.0.co;2-7.

Kalogeropoulou, V., Bett, B.J., Gooday, A.J., Lampadariou, N., Martinez Arbizu, P., Vanreuselm A., 2016. Temporal changes (1989–1999) in deep-sea metazoan meiofaunal assemblages on the Porcupine Abyssal Plain, NE Atlantic. *Deep-Sea Research II,* 57:1383–1395. DOI 10.1016/j.dsr2.2009.02.002.

Lastras, G., Canals, M., Amblas, D., Lavoie, C., Church, I., De Mol, B., Duran, R., Calafat, A.M., Hughes-Clarke, J.E., Smith, C.J., Heussner, S., 2011. Understanding sediment dynamics of two large submarine valleys from seafloor data: Blanes and La Fonera canyons, northwestern Mediterranean Sea, *Marine Geology*, 280(1–4):20–39. DOI 10.1016/j.margeo.2010.11.005.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J., Machida, R.J., 2013. A new versatil primer set targeting a short fragment of the mithocondrial COI region for metabarcoding metazoan diversity:application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10:13. DOI https://doi.org/10.1186/1742-9994-10-34.

Lopez-Fernandez, P., Calafat, A., Sanchez-Vidal, A., Canals, M., Mar Flexas, M., Cateura, J., & Company, J. B., 2013. Multiple drivers of particle fluxes in the Blanes submarine canyon and southern open slope: Results of a year round experiment. *Progress in Oceanography,* 118:95–107. DOI 10.1016/j.pocean.2013.07.029.

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

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2016. Vegan: community ecology package. R-package version 2.3-4. Available at https:// cran.r-project.org/ package=vegan.

Paradis, S., Puig, P., Sanchez-Vidal, A., Masqué, P., Garcia-Orellana, J., Calafat, A., Canals, M., 2018. Spatial distribution of sedimentation-rate increases in Blanes Canyon caused by technification of bottom trawling fleet. *Progress in Oceanography,* 169:241–252. DOI 10.1016/j.pocean.2018.07.001.

Puig, P., Martín, J., Masqué, P. & Palanques, A., 2015. Increasing sediment accumulation rates in La Fonera (Palamós) submarine canyon axis and their relationship with bottom trawling activities. *Geophysical Research Letters*, 42:8106–8113. DOI 10.1002/2015GL065052.

Ramirez-Llodra, E., Tyler, P. A., Baker, M. C., Bergstad, O. A., Clark, M. R., Escobar, E., et al,2011. Man and the last great wilderness: human impact on the deep sea. *PLoS ONE,* 6:e22588. doi: 10.1371/journal.pone.0022588

Richter, O.-M. H., & Ludwig, B. (n.d.), 2003. Cytochrome c oxidase — structure, function, and physiology of a redox-driven molecular machine. *Reviews of Physiology, Biochemistry and Pharmacology*, 47–74. DOI 10.1007/s10254-003-0006-0

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

Román, S., Vanreusel, A., Romano, C,. Ingels, J., Puig, P., Company, J.B., Martin, D., High spatiotemporal variability in meiofaunal assembleges in Blanes Canyon (NW Mediterranean) subject to anthropogenic and natural disturbances. *Deep-Sea Research I*, 117:70-83.

Romano, C., Coenjaerts, J., Flexas, M.M., Zúniga, D., Vanreusel, A., Company, J.B., Martín, D., 2013. Spatial and temporal variability of meiobenthic density in the Blanes submarine canyon (NW Mediterranean). *Progress in Oceanography,* 118:144–158. DOI 10.1016/j.pocean.2013.07.026.

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

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America,* 109:16208–16212. DOI 10.1073/pnas.1209160109.

Ulses, C., Estournel, C., Puig, P., Durrieu de Madron, X., & Marsaleix, P., 2008. Dense shelf water cascading in the northwestern Mediterranean during the cold winter 2005: Quantification of the export through the Gulf of Lion and the Catalan margin. *Geophysical Research Letters,* 35(7), L07610. DOI10.1029/2008gl033257

Wangensteen, O.S., Palacín, C., Guardiola, M. and 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

Wangensteen, O.S. & Turon, X., 2017. Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In: Rossi S., Bramanti L., Gori A., Orejas C. (eds) Marine Animal Forests. Springer, Cham.

Zúñiga, D., Flexas, M. M., Sanchez-Vidal, A., Coenjaerts, J., Calafat, A., Jordà, G. & Company, J. B., 2009. Particle fluxes dynamics in Blanes submarine canyon (Northwestern Mediterranean). *Progress in Oceanography,* 82(4):239–251. DOI 10.1016/j.pocean.2*009.07.002.*